To the Graduate Council:

I am submitting herewith a dissertation written by Gwenn M. Hansen entitled "Genomic Analysis of a Region Implicated in Lymphoma Development: Viral Insertion Mutations at *Lvis1* Disrupt the Expression of Multiple Genes." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

Monica J. Justice

Monica J. Justice, Major Professor

We have read this dissertation
and recommend its acceptance:

Michael L. Mucenski

Richard J. Mural

Dubeny K. Jolen

Mary Ann Handel

Accepted for the Council:

C.W. Minkel

Associate Vice Chancellor and
Dean of The Graduate School

GENOMIC ANALYSIS OF A REGION IMPLICATED IN LYMPHOMA DEVELOPMENT: VIRAL INSERTION MUTATIONS AT *LVIS1* DISRUPT THE EXPRESSION OF MULTIPLE GENES

A Dissertation

Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Gwenn M. Hansen

August, 1998

ABSTRACT

Cancer is one of the leading causes of death in the human population. Effective treatment of cancer, as with any disease, depends upon the extent to which the etiology of the disease process is known. The main objective of this research is to further our understanding of the etiology of lymphoma by studying the molecular genetic basis of the development of lymphoma in mice. The identification of new loci that are implicated in causing murine lymphoma will facilitate our understanding of hematopoietic disease in humans.

Retroviruses and other transposable elements are often found as causative agents in mouse cancers. One way these elements can cause cancer is to integrate near and alter the expression of normal somatic cellular proto-oncogenes. Retroviruses act as insertional mutagens and can serve as useful molecular tags to clone new genes causally associated with tumorigenesis. Nearly all strains that make up the AKXD recombinant inbred (RI) series of mice have a high incidence of lymphoma caused by the expression and subsequent somatic integration of murine leukemia retroviruses (MuLVs). The lymphomas of AKXD mice, therefore, can be used to identify and clone proto-oncogenes associated with lymphoma development using the MuLV sequence as a molecular tag.

Conventional insertion site cloning requires the construction of genomic libraries which are screened by hybridization to isolate viral/genomic junction fragments. To eliminate the need for hybridization-based cloning, I adapted universal PCR to clone a viral insertion site in the mouse. Using a highly degenerate primer containing a short anchor sequence, coupled with a viral specific primer, we have amplified genomic sequences flanking a somatically acquired retrovirus in DNA isolated from a mouse B-cell lymphoma. This locus was designated *lymphoid viral integration site 1*, (*Lvis1*). *Lvis1* was mapped to distal mouse chromosome 19 using

interspecific backcross analysis. A detailed molecular genetic linkage analysis of this region, including a number of human gene markers known to map to human chromosome 10 revealed a significant region of conserved synteny between distal mouse chromosome 19 and human chromosome 10q23-q26.

Analysis of *Lvis1* in 270 AKXD tumors revealed that this locus is frequently altered by viral integration. Alterations at this locus have been observed in 9.6% of AKXD lymphomas, accounting for 22% of lymphomas classified as B-cell in origin. Notably, integrations at *Lvis1* occur in a small number of T-cell tumors, but have not been observed in myeloid tumors. This locus is disrupted by retroviral integration in lymphomas from 14 AKXD strains. Random sequencing of BAC clones spanning >110 kb identified four genes near the site of viral insertion: *Hex*, *mEg5*, *msec15*, and *L35a*. Expression analysis of these genes in tumors with *Lvis1* alterations revealed that viral insertions differentially effect neighboring gene regulation. *Hex*, an orphan homeobox gene involved in embryonic developmental patterning and hematopoiesis, is significantly upregulated in *Lvis1* tumors, whereas no *Hex* expression is detected in tumors lacking *Lvis1* alterations. Both *mEg5* and *msec15* show a slight upregulation in *Lvis1* tumors, while *L35a* expression appears unaffected. These data provide the first genetic evidence for the possible role of these genes in lymphoma development, and highlight the use of genomic approaches for the study of position effect mutations.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	OVERVIEW AND INTRODUCTION	1
	Organizational Overview	1
	Insertion Mutation Analysis	2
	Advances in Mammalian Germline Insertional Mutagenesis	3
	Somatic Mutation and Cancer Development	5
	Retroviral Insertion Mutations are Associated with the Development of Cancer	5
	Loci Disrupted by Viral Integration Identify Proto-Oncogenes and Tumor Suppressor Genes Involved in Both Mouse and Human Lymphoma Development	7
	AKXD Recombinant Inbred Strains as a Model System for the Study of Lymphoma Development	7
	Research Focus and Goal	10
II.	USING UNIVERSAL PCR TO CLONE VIRAL INTEGRATION SITES IN THE MOUSE	11
	Abstract	11
	Introduction	11
	Materials and Methods	13
	Results	16
	Discussion	18
III.	MOLECULAR GENETIC LINKAGE ANALYSIS OF MOUSE CHROMOSOME 19	20
	Introduction	20

III. A. *Pten*, A CANDIDATE TUMOR SUPPRESSOR GENE, MAPS TO MOUSE CHROMOSOME 19 23
Methods and Results 24
Discussion 26
Acknowledgments 27
III. B. REFINEMENT OF THE GENETIC MAP OF MOUSE CHROMOSOME 19 - HUMAN CHROMOSOME 10Q23-26 HOMOLOGOUS REGION 28
Abstract 28
Introduction 28
Material and Methods 30
Results 31
Discussion 32
Acknowledgments 36
IV. VIRAL INSERTION MUTATIONS AT *Lvis1* DISRUPT LOCAL GENE EXPRESSION AND IMPLICATE MULTIPLE GENES IN LYMPHOMA DEVELOPMENT 37
Abstract 37
Introduction 38
Methods 40
Results 44
Discussion 54
V. SUMMARY AND DISCUSSION OF SIGNIFICANT FINDINGS AND FUTURE PROJECT DIRECTIONS 59
Summary of Experimental Findings 59
Universal PCR: Current Uses and Future Applications 61

Random Sequencing as a Tool For Gene Identification..... 62

Limitations of Available Database Information..... 66

Future Experiments to Test Candidate Genes at *Lvis1*..... 68

Additional Gene Function Analysis..... 69

Identification of a Potential Gene Regulatory Region..... 71

Conclusion..... 72

REFERENCES..... 73

APPENDIX..... 95

VITA..... 141

LIST OF TABLES

TABLE PAGE

CHAPTER I.

I-1. Viral integration sites in AKXD and BXH-2 strains..... 97

CHAPTER III.

III-1. Probes/Amplimers used for mapping studies..... 107

CHAPTER IV.

IV-1. AKXD tumor cell lineage distribution of *Lvis1* integrations..... 115

IV-2. Frequency of viral alterations in AKXD neoplasias.. 116

IV-3. Virally induced rearrangements at *Lvis1* in AKXD tumors..... 117

IV-4. Blastn search results of shotgun sequencing clone.... 123

LIST OF FIGURES

FIGURE PAGE

CHAPTER I.

I-1. Identifying viral integration sites in the mouse..... 96

CHAPTER II.

II-1. Schematic representation of universal PCR using MuLV LTR specific primers..... 98 - 99

II-2. Southern blot analysis of universal PCR products ... 100 - 101

II-3. Comparison of amplified viral junction fragment sequence with sequence of the germline allele..... 102

II-4. Southern blot analysis of brain and tumor DNA..... 103 - 104

CHAPTER III.

PART A.

III A-1. Segregation of alleles in 124 (SB/Le x *Mus spretus*) F_1 x SB/Le interspecific backcross progeny..... 105 - 106

PART B

III B-1. Segregation of alleles in 130 (SB/Le x *Mus spretus*)F_1 x SB/Le interspecific backcross progeny..... 108 - 109

III B-2. Molecular genetic linkage maps of mouse chromosome 19..... 110 - 111

III B-3. Localization of *Hex-rs* to mouse chromosome 3..... 112 - 113

CHAPTER IV.

IV-1. Southern blot analysis of AKXD tumor DNA..... 114

IV-2. Chromosomal localization of *Lvis1* in the mouse.... 118

IV-3. Location and orientation of proviruses as *Lvis1* in AKXD tumors 119

IV-4. Physical map of BAC clones at *Lvis1* 120

IV-5. DNA fiber FISH mapping of BAC clones ..... 121 - 122

IV-6. Integrated physical map of genes and mutations at *Lvis1* ..... 124 - 125

IV-7. Nucleotide sequence of the *mEg5* cDNA and the predicted amino acid sequence ..... 126 - 129

IV-8. Comparison of the amino acid similarity between motor domains of bim-C family proteins 130

IV-9. Alignment of potential $p34^{cdc2}$ phosphorylation sites in bim-C related proteins 131

IV-10. Expression of *L35a* in AKXD-18 spleen and lymph node tumors ..... 132 - 133

IV-11. Expression of *Hex* in AKXD-18 spleen and lymph node tumors ..... 134 - 135

IV-12. Expression of *mEg5* in AKXD-18 spleen and lymph node tumors ..... 136 - 137

IV-13. Expression of *msec15* in AKXD-18 spleen and lymph node tumors ..... 138 - 139

CHAPTER V.

1. Mapping of *mEg5-2* 140

LIST OF ABBREVIATIONS

BAC	Bacterial artificial chromosome
bp	Basepair
cDNA	Complementary DNA synthesized by reverse Transcription of messenger RNA
cM	Centimorgan
DNA	Deoxynucleic Acid
dNTPs	Deoxynucleotide tri-phosphate
EST	Expressed sequence tag
IB	Interspecific Backcross
kb	Kilobase
L	Liter
ml	Milliliter (10^{-3} liter)
mM	Milimolar (10^{-3} mole/liter)
µg	Microgram (10^{-6} gram)
µl	Microliter (10^{-6} liter)
n	Sample number
N_2	Second generation of offspring
ng	Nanogram (10^{-9} gram)
nt	Nucleotide
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
pmole	Picomole (10^{-12} mole)
RFLP	Restriction fragment length polymorphism

SDS	Sodium dodecyl sulfate
SSLP	Simple sequence length polymorphism
SHSM	Split-Hand Split Foot Malformation
Dac	Dactylpolyplasia
F1	First generaltion of offspring
RNA	Deoxyribonucleic Acid

I. OVERVIEW AND INTRODUCTION

ORGANIZATIONAL OVERVIEW

This work is composed of three major experimental portions, organized in a series of logical steps toward the cloning and analysis of a locus causally associated with the development of lymphoma. A chapter of general introduction precedes three chapters of experimental work, which are followed by a chapter of discussion, conclusions, and future experiments. The background and introduction, materials and methods, results, and discussion pertaining to each phase of experimentation are written separately, but all references are combined and placed at the end of the dissertation. Chapter I is a general introduction and review of insertional mutation, with an emphasis on the specific use of this technique in the study of the molecular basis of cancer development in mice. Chapter II is a manuscript in preparation for submission to *BioTechniques* which describes the development and use of a PCR technique for the amplification and cloning of viral insertion sites in the mouse. Chapter III includes two manuscripts. The first is a paper published in *Mammalian Genome* (Hansen and Justice, 1998), and the second is a manuscript in preparation for submission to *Genomics*. These two papers describe mapping experiments used to further develop and define the mouse chromosome 19 genetic linkage map. These papers are preceded by a brief overview of the uses of gene mapping techniques that is not intended for publication, but is useful for a complete understanding of the value of these particular experiments in the overall goal of this research. Chapter IV is a manuscript in preparation for submission to *Nature Genetics* describing the cloning and genomic analysis of a viral integration site frequently altered in B-cell lymphomas of AKXD mice. A final summary and discussion of the significant results presented in

this report are included in a final chapter, Chapter V. This chapter also discusses ongoing and future experiments. All tables and figures for each chapter have been pooled and placed in an appendix at the end of the dissertation. Figures are preceded by a figure legend page where necessary.

INSERTION MUTATION ANALYSIS

Insertion mutations result from incorporation of foreign DNA sequences into the genomic structure. The phenotypic consequences of insertional mutation vary. Insertion events can disrupt, activate, or have no effect on local gene expression, while in other cases, these events can alter the expression of genes at a great distance from the insertion site (Favor and Morawetz, 1992; Lazo *et al.*, 1990; Bedell *et al.*, 1996). For purposes of gene function analysis, insertional mutagens are valuable resources because they act not only to induce mutation, but also to mark the site of mutation for subsequent cloning and study. This approach to gene function analysis was first used in *Drosophila melanogaster* to clone the *white* gene (Bingham *et al.*, 1981), and has since been used in multiple genetic systems from plants to mice to identify and study gene function.

Insertional mutagenesis studies in *D. melanogaster* have been particularly useful due to the fact that transposable elements in this organism, such as the P element, transpose with high frequency. In addition, significant modifications of the original P element have allowed the creation of novel P element plasmids (Robertson *et al.*, 1988; Wilson *et al.*, 1989) which are more stable than the original element, contain selectable markers, and which are designed such that genomic DNA flanking the site of insertion can be quickly isolated by plasmid rescue (Bellen *et al.*, 1989). These new P element plasmids are ideal for large scale mutagenesis screens and have been used to created thousands of mutant lines (Cooley *et al.*, 1989). Because of the ease of generating and

analyzing insertional mutations in the *Drosophila* system, this approach to the study of gene function has been undertaken in a logical progression from genome sequence analysis, and has yielded a comprehensive gene-disruption library with insertions lying within or near most *Drosophila* genes (Spradling *et al.*, 1995).

ADVANCES IN MAMMALIAN GERMLINE INSERTIONAL MUTAGENESIS

Insertional mutagenesis approaches in the mouse have proven less successful. A number of examples of spontaneous insertional mutation causing classical recessive mouse mutations have been studied (Favor and Morawetz, 1992 for review). However, it is estimated that spontaneous insertional mutations in the mouse account for at least 5% of mutations (Stoye *et al.*, 1988), while in *Drosophila*, the percentage is estimated to be between 50 and 86% (Ramel, 1989; Sobels, 1989). Aside from the availability of spontaneous mutations, a number of reasons account for the fact that mammalian model organisms are not particularly suited for large scale insertional analysis. As opposed to experimental systems such as *D. melanogaster* and *Caenorhabditis elegans,* mammalian systems are cumbersome due to the size of their genomes, their long gestation times and reproductive cycles, and the significant costs associated with the care and housing of animals. In addition, many of the mutation screening techniques used to identity genes that function in embryonic development in *D. melanogaster and C. elegans* can not be employed in mammalian systems where the majority of embryonic development takes place *in utero.* Therefore, the experimental study of mammalian gene function has focused mainly on characterization of single gene function using available physical methods rather than using large scale mutation screening. Currently, however, this approach is changing. Improvements in large scale sequencing techniques have lead to significant advances in genome study and

characterization, such that soon sequence information will be available for the majority of human genes (Gibbs, 1995). Unfortunately, available techniques such as gene knock-out and transgenic technology are limited, and will not yield complete functional analysis of these genes and their interactions (Davis and Justice, 1998).

To address this concern, novel approaches for the study of mammalian genes *in vivo* are being developed. One such method involves the use of insertional mutagenesis in mouse embryonic stem (ES) cells (Friedrich and Soriano, 1991, Skarnes *et al.,* 1995). In general, this method is used to generate null alleles, where a promoter-less selectable marker is used to mutagenize ES cells. Mutation events in which the targeting vector inserts into an expressed cellular gene, such that expression of the marker construct is activated by the host gene promoter elements, can be easily identified, and large numbers of insertion mutations can be generated within the cell culture system. *In vivo* analysis of these mutations, however, must still be carried out on an individual basis. To further refine this strategy for generation and processing of an even greater number of mutations, sequenced-based screening has been developed (Zambrowicz *et al.*, 1998, Hicks *et al.,* 1997). This process involves sequencing a short segment of DNA flanking the site of targeting vector insertion. Sequence information from each mutant is then used to search the available nucleotide databases to quickly determine the identity of the mutagenized locus, or mark the site of a yet unidentified expressed gene. Individual clones are then saved for later processing by interested investigators. While providing valuable resources for gene function analysis, this method takes advantage of the wealth of genome resources, and makes insertional mutation screening feasible on a large scale in the mouse.

SOMATIC MUTATION AND CANCER DEVELOPMENT

Insertional mutagenesis is a powerful genetic and molecular tool that is used in most model organisms primarily for germline mutagenesis. Fortunately, insertional mutation analysis need not be limited to the study of germline mutations, for it is also of great value in the study of gene function through the analysis of somatic mutations. Cancer is essentially a genetic disease brought about by somatic mutation, and this concept is primary to our understanding of lymphoma development. Cancer results from an accumulation of multiple genetic alterations that ultimately disrupt the control of normal cell growth and differentiation and lead to the monoclonal expansion of transformed cells. In human hematopoietic disease, these genetic alterations are commonly chromosomal translocations. One of the classic examples is the translocation known as the Philadelphia chromosome, t[(9;22)(q24;q11)]. This translocation joins the *BCR* gene on human chromosome 22 with the *Abl* proto-oncogene on chromosome 9, resulting in the activation of the ABL tyrosine kinase which is causally associated with human chronic myelogenous leukemia (CML) (reviewed in Rabbitts and Boehm, 1991). It is important to note, however, that the t(9;22)(q24;q11) translocation alone is not sufficient for the development of CML, and further it may not be the first step in CML development (Vogelstein & Kinzler, 1993).

RETROVIRAL INSERTION MUTATIONS ARE ASSOCIATED WITH THE DEVELOPMENT OF CANCER

In the mouse, the most common models of tumorigenesis are those in which hematopoietic cancers develop as a consequence of viral insertion mutations. Many inbred strains of mice have a high incidence of spontaneous retrovirally induced

leukemias and lymphomas. Such strains harbor type C retroviruses, such as murine leukemia retroviruses (MuLVs), within their genome, and develop neoplastic disease following a long latency period (Bishop, 1987). Unlike acute-transforming retroviruses which contain oncogenic sequences, MuLVs and other slow-transforming retroviruses contain no oncogenic sequences, and are thought to induce or potentiate tumorigenesis by integrating near and altering the expression of cellular proto-oncogenes (Nusse, 1986; van Lohuizen and Berns, 1990). A requisite part of the life cycle of the retrovirus requires insertion of a DNA copy of the retroviral genome into the genomic DNA of the host (Varmus, 1982). Although most integration events are not detrimental, proviral integrations at or near coding regions can alter normal gene activity. When a virus integrates near a cellular proto-oncogene, regulatory control elements present within the viral long terminal repeat (LTR) can affect gene expression, which, in somatic cells, can result in altered growth regulation leading to neoplastic disease. Therefore, MuLV sequences can be used as molecular tags to identify genes causally associated with disease. Moreover, this approach may prove useful in the identification of genes important for cellular growth and differentiation.

In practice, potential proto-oncogenes are initially identified as sites of retroviral integration occurring in multiple independent tumors. Because retroviruses can integrate into many sites in the host genome, integration at the same location in multiple independent tumors provides strong indication that the locus encodes a gene that can, when altered by viral integration, provide the cell a growth advantage. These sites are identified by cloning cellular DNA/proviral DNA junction fragments from tumor DNA. Unique-sequence cellular DNA probes that flank the site of integration are then used to screen DNA from retrovirally induced tumors to determine whether the locus has been disrupted in multiple tumors (Fig. I-1).*

* All tables and figures may be found in the appendix.

LOCI DISRUPTED BY VIRAL INTEGRATION IDENTIFY PROTO-ONCOGENES AND TUMOR SUPPRESSOR GENES INVOLVED IN BOTH MOUSE AND HUMAN LYMPHOMA DEVELOPMENT

Over the years a number of genes at or near sites of viral integration have been implicated in the disease process. Growth factors such as *IL2, IL3, Int1, Int2, Fgf,* and *Cfsm*; growth factor receptors such as *IL2Rα, Csfmr,* and *c-erbB*; signal transduction loci such as *Lck, Pim1, Mos* and *Hras;* and tumor suppressor loci such as *p53* and *Nf1* have been disrupted by viral integration, leading to neoplastic transformation (van Lohuizen & Berns, 1990, and Lazo & Tsichlis, 1990, for reviews). More recently, *Meis1*, a novel *PBX1*-related homeobox gene was identified as a common site of retroviral integration in myeloid tumors (Moskow *et al.,* 1995). *PBX1* is implicated in human pre-B-cell acute lymphoblastic leukemia in patients who carry the t(1;19) chromosomal translocation, which links the majority of the *PBX1* gene, including the homeobox, to the 5' region of *E2A* (Kamps *et al.*, 1990; Nourse *et al.,* 1900). In mice, viral integrations at *Meis1* have been shown to increase gene expression. Further analysis has shown that many tumors with integrations at *Meis1* also have integrations in either *Hoxa7* or *Hoxa9*, implying that these genes cooperate with *Meis1* in tumor formation (Nakamura *et al.*, 1996).

AKXD RECOMBINANT INBRED STRAINS AS A MODEL SYSTEM FOR THE STUDY OF LYMPHOMA DEVELOPMENT

Many recombinant inbred (RI) strains of mice have been used as model systems for identifying and studying genes associated with hematopoietic diseases (Justice *et al.,* 1992). One series of RI strains that is particularly useful for the

study of lymphoma development is the AKXD RI series of strains, which develop a variety of virally induced hematopoietic tumors. These strains are derived from a cross between two inbred strains, AKR/J and DBA/2J, that differ significantly in their lymphoma incidence (Mucenski *et al.*, 1986). The AKR/J parent strain has a high incidence of T-cell lymphoma, causally associated with the expression of two endogenous ecotropic MuLVs, *Emv11* and *Emv14* (Jenkins *et al.*, 1982). The T-cell lymphoma phenotype characteristic of AKR mice is also associated with the presence of recombinant viruses known as mink cell focus-forming (MCF) viruses, found in both pre-leukemic and leukemic thymuses of this mouse strain (Herr and Gilbert, 1983). MCF viruses are not encoded within the germline of the mouse, but are formed by recombination between one ecotropic virus and at least two non-ecotropic viruses. While ecotropic viruses are weakly leukemogenic and primarily associated with tumors of B-cell and myeloid lineages (Zijlstra *et al.*, 1986), most MCF viruses are highly leukemogenic. In addition, MCF viruses are thought to infect and replicate selectively in immature thymocytes present in the thymic cortex (Cloyd, 1983). Since thymectomized AKR mice do not exhibit a high incidence of lymphoma, it is likely that the recombinant MCF viruses are the primary agent of T-cell disease in these strains (Copeland *et al.*, 1985). In contrast, the DBA/2J strain has a low incidence of lymphoma, carrying a single MuLV locus, *Emv3*, which is replication defective.

The AKXD RI strains differ from both parental strains in lymphoma susceptibility, lymphoma cell type, average age of disease onset, and somatic viral content. This suggests that in addition to replication competent viruses, a number of other loci that influence disease type are present in the parental strains, and have segregated during inbreeding becoming fixed in a variety of different combinations in the resulting RI strains. Twenty-one of the twenty-three RI strains develop

lymphoma. Each of these twenty-one strains have inherited at least one of the replication competent viruses from the AKR/J parental strain (*Emv11* or *Emv14*). Neither of the two strains that do not develop lymphoma carry *Emv11* or *Emv14*. The AKXD lymphomas are predominantly B-cell in origin, where seven strains develop B-lineage lymphomas, six develop T-lineage lymphomas, seven develop both T- and B-lineage lymphomas, and one strain develops myeloid leukemias (Gilbert *et al.*, 1993; Mucenski *et al.*, 1988). Analysis of the proviral content of the AKXD tumors is consistent with the induction of the these tumors by viral integration. All tumors contain somatically acquired proviruses and most are monoclonal in origin. Coincident with previous studies of viral insertions in B-lineage lymphomas (Zijlstra *et al.*, 1986), most of the somatically acquired viruses detected in AKXD B-lineage lymphomas were ecotropic viruses (Gilbert *et al.*, 1993; Mucenski *et al.*, 1988).

To date, a number of viral integration sites have been cloned and analyzed to identify genes that contribute to the development of hematopoietic cancer in the AKXD RI mouse strains, as well as in the BXH-2 RI strain (Bedigian, 1981) (Table I-1). The majority of these loci have been found to be associated primarily with myeloid leukemias. Studies of the B-cell lymphomas from these strains have shown that known proto-oncogenes are not the preferred sites of viral integration these tumors. Many loci implicated in hematopoietic disease, including *c-myc, N-myc, Pim1, Pim2, Pvt1, c-myb, Evi1, Evi2, Evi5, Gfi1, Fim1, Mlvi2, p53, Bcl2, Cbl, c-abl, Rag1, Gata2*, and *Tpl2* have been analyzed for rearrangements in tumors from the AKXD RI strains (Mucenski *et al.*, 1988; Liao *et al.*, 1995; G.M. Hansen and M. J. Justice, unpublished results). No rearrangements were found at any loci known to be involved in B-cell differentiation and disease (*Bcl2, Cbl, c-abl, Rag1, Gata2);* however, rearrangements at *Pim1*, *Evi1, Evi5*, and *Gfi1* were found, accounting for approximately 13% of the B-

cell lymphomas from these strains. These data indicate that the majority of genes responsible for lymphoma development in these strains are yet to be identified.

RESEARCH FOCUS AND GOAL

The primary goal of this project is to understand the molecular genetic basis of the development of B-cell lymphomas in mice by identifying additional genes that contribute to the disease process. The mouse is an ideal model system in which to study the development of hematopoietic disease. Inbred strains of mice provide a defined genetic background to study loci involved in the disease process. In addition, mouse hematopoietic tumors have both clinical and immunopathologic similarities to human lymphomas and related leukemias (Pattengale, 1994). Therefore, identification of new loci that are implicated in causing murine B-cell tumors should facilitate our understanding of the development of hematopoietic disease in humans.

II. USING UNIVERSAL PCR TO CLONE VIRAL INTEGRATION SITES IN THE MOUSE

ABSTRACT

Murine leukemia retroviruses (MuLVs) induce leukemia and lymphoma in susceptible strains of mice as a result of insertional activation of cellular proto-oncogenes or inactivation of tumor suppressor genes. Therefore, cloning of viral insertion sites is used to identify disease genes. Conventional insertion site cloning requires the construction of genomic libraries which are screened by hybridization to isolate viral/genomic junction fragments. To eliminate the need for hybridization-based cloning, we adapted universal PCR to clone a viral insertion site in the mouse. Using a highly degenerate primer containing a short anchor sequence, coupled with a viral-specific primer, we have amplified genomic sequences flanking a somatically acquired retrovirus in DNA isolated from a mouse B-cell lymphoma. This technique provides a simple alternative to hybridization-based techniques and can be adapted for rapid cloning of large numbers of insertion sites in the mouse.

INTRODUCTION

AKXD recombinant inbred strains of mice develop a variety of hematopoietic cancers as a consequence of somatic viral insertions that alter the expression of cellular proto-oncogenes and tumor suppressor genes. The process of identifying these genes begins with cloning genomic sequence flanking the site of insertion in tumor tissue. Because not all viral insertion events contribute to the disease process, only those sites disrupted by viral insertion in multiple independent tumors are analyzed. Experimental

evidence from our laboratory, as well as from others (Nakamura *et al.*, 1996) indicates that less than half of all randomly cloned viral insertion mutations identify frequently altered sites. Thus, to clone disease-associated sites, it is necessary to clone and analyze a large number of insertion sites. Conventionally, viral insertion sites are cloned by constructing genomic libraries and screening by hybridization to isolate fragments containing viral sequences. Although effective, this approach is labor-intensive and designed for processing samples on an individual basis. Alternatively, PCR techniques are extremely suitable for high throughput cloning. To develop a PCR-based approach to clone viral insertion sites in the mouse, we adapted a technique designed to amplify transposon insertion junctions from *Woot*, a *gypsy*-class transposable element found in *Tribolium castaneum* (Beeman and Stauth, 1997). This technique combines a primer specific for the long terminal repeat (LTR) of *Woot* with a highly degenerate primer containing a short 3' anchor sequence (Sarkar *et al.*, 1993). Two to three rounds of PCR from unmodified template DNA were used to successfully amplify and sequence multiple transposon junctions fragments (Beeman and Stauth, 1997). To adapt this approach for cloning viral insertion sites in the mouse, we substitute MuLV specific LTR primers for *Woot* specific primers and further increase our specificity by size fractionating the genomic template DNA and increasing the stringency of the second round of PCR amplification.

Using this technique we have amplified the genomic sequences flanking a somatically acquired retrovirus in DNA isolated from B-cell lymphoma infiltrated spleen tissue from an AKXD mouse. This locus, designated *Lymphoid viral integration site 1 (Lvis1)* detects a tumor-specific restriction fragment length alteration, and potentially identifies a gene(s) involved in cancer development.

MATERIALS AND METHODS

Overview of universal PCR. Sequences flanking the viral insertion were amplified using two rounds of PCR. The initial round was performed on 200 - 500 ng genomic DNA using a universal primer (UP1 or UP2) and a MuLV LTR specific primer (SP1). The second round combined a single microliter of first round reaction as template, with a nested LTR specific primer (SP2) and the M13F primer which is included as an adapter in the universal primer design. Because the MuLV is flanked on both ends by identical LTR elements, LTR specific primers will bind and amplify from both ends of the MuLV. This will produce two classes of products: those that represent viral sequences internal to the downstream LTR, and those that represent mouse DNA flanking the upstream LTR (Fig. II-1A). To evaluate the specificity of amplification, second round reactions were analyzed by Southern hybridization with viral probes, and only those products thought to represent virus/genomic junction fragments were chosen for further analysis.

Template preparation. DNA (50 μg) isolated from spleen tissue with lymphoma cell infiltration from AKXD-18 animal no. 108 was digested to completion with *Xba*I and fractionated by electrophoresis through a 0.8% SeaKem® LE agarose (FMC Bioproducts, Rockland, ME) gel in 1xTAE buffer (40 mM Tris-acetate, 20 mM sodium acetate, 1 mM EDTA). DNA fragments corresponding in size to the tumor-specific proviral DNA/genomic DNA junction fragment (12-17 kb) were purified from gel slices, and identified by Southern blot hybridization using the *eco* probe, a probe specific for the envelope gene region of the AKV ecotropic MuLV(Chattopadhyay, 1980).

Universal primer design. MuLV LTR specific primers were designed to amplify the upstream insertion junction. The nested MuLV forward primers were SP1 5'-

CTGAGAACATCAGCTCTG-3', and SP2 5'-CTGGCTAAGCCTTATGAAGGGGTCTTTC-3'. For increased specificity, the SP2 oligo was designed as a 28-mer, such that the annealing temperature of the second round reaction could be increased to 60°C. These primers bind 71 bp and 2 bp from the 5' LTR terminus, respectively. The universal primers consisted of three regions: a 3' anchor sequence of 6 nucleotides (nt) (based on restriction enzyme recognition sequences), a fully degenerate region of 8 nt, and a 5' M13F linker. Amplification of mouse DNA upstream of the integrated virus is dependent upon the proximity of the universal primer binding site to the viral LTR specific primer binding site. Thus, to increase the likelihood of a binding site near the integrated virus, we chose anchor sequences that are common in mouse genomic DNA. In addition, we tried to minimize the likelihood of internal virus amplification by choosing one anchor sequence, *Pst*I, that is not present in ecotropic viral genomic DNA internal to the downstream LTR (Chattopadhyay *et al.*, 1980). The two universal primers used for this experiment were UP1, containing a *Pst*I anchor sequence, 5'-GGGTTTTCCCAGTCACGACNNNNNNNNCTGCAG-3' and UP2, containing a *Xba*I anchor sequence, 5'-GGGTTTTCCCAGTCACGACNNNNNNNNTCTAGA-3'.

Reaction conditions. PCR was performed using a MJ Research thermal controller with the cycling parameters: 94°C for 3 min, 30 cycles of 94°C for 30 s, 50°C for 2 min, and 72°C for 2 min, with a final 7 min incubation at 72°C. The initial round of PCR contained 2 µl of the selected subgenomic DNA fraction, 2.5 pmol SP1, 25 pmol UP1, 1.5 mM MgCl, 0.2 mM each dNTP, 2.5 µl 10x PCR buffer (Perkin Elmer-Cetus, Foster City, CA), and 0.25 µl AmpliTaq DNA polymerase (Perkin Elmer-Cetus, Foster City, CA) in a volume of 25 µl. One microliter of this reaction was used as template in a nested reaction with the following changes: 25 pmol of both SP2 and

M13F were used as primers, and the annealing temperature was increased to 60°C. One half of each reaction was run on a 1.5% SeaKem® LE agarose gel in 1xTAE. Products of interest were excised from the gel, purified using Wizard™ PCR Preps (Promega, Madison, WI), and cloned into the pGEM® -T vector (Promega, Madison, WI). The remaining reaction volume was run on a 1% SeaKem® LE agarose gel in 1xTPE, photographed, and transferred to duplicate pieces of MSI® nylon membrane using standard Southern blotting procedures (Sambrook *et al.*, 1989).

DNA extraction and Southern blot hybridizations. All molecular biology techniques were performed according to standard procedures (Sambrook *et al.*, 1989). High molecular weight genomic DNA was extracted from frozen normal tissues and tissues with lymphoma cell infiltration. Southern blots were hybridized at 65°C (Church and Gilbert, 1984), and were washed three times for 30 min each at 65°C with a solution of 0.2X SSCP/ 0.1% SDS.

Hybridization Probes. The SP2 oligonucleotide used as a probe for Southern analysis was end-labeled with [γ-^{32}P]. End-labeling was performed in a total volume of 10 µl, and contained 5 pmol SP2, 5 U T4 Polynucleotide Kinase (PNK), 5 pmol gamma ^{32}P and 1 µl 10x T4 PNK buffer (Life Technologies, Inc., Gaithersburg, MD). The reaction was incubated at 37°C for 45 min, and the enzyme was inactivated by incubation at 70°C for 10 min. Labeled oligonucleotides were separated from unincorporated nucleotides using a NucTrap® Probe Purification Column (Stratagene, La Jolla, CA). The viral control probe used to detect internal virus sequence amplification is a 322 bp fragment from nucleotide position 6504 - 7825 of the MuLV (Genbank accession no. J01998). Probe pL1 was generated by PCR using primers 10842-o 5'-AGAACTTGAGGCAAAGTTGC-3' and 10842-i 5'-GGTCCCATCCGTCTGAATTC-3' in the following reaction: 50 pmol p10842, 100

ng each primer, 2.5 µl 10x PCR buffer (Life Technologies, Inc., Gaithersburg, MD), 1.5 mM MgCl, 0.2 mM each dNTP and 0.25 µl Taq DNA polymerase (Life Technologies, Inc., Gaithersburg, MD) in a 25 µl reaction volume. The reaction was cycled using the parameters: 94°C for 3 min, then 30 cycles of 94°C for 30 sec, 55°C for 2 min, 72°C for 2 min followed by a 7 min incubation at 72°C. pL1 was labeled using the Prime-It® II Random Primer Labeling Kit (Stratagene, La Jolla, CA) with the following changes: 200 ng of both 10842-o and 10842-i were substituted for the random primers in the annealing reaction. The *eco* probe is a 305 bp PCR fragment representing the envelope gene region of the AKV ecotropic MuLV (position 6271-6566 of Genbank accession no. J01998) amplified from pECO (Chattopadhyay, 1980) using primers (forward) 1A 5'-ATGTGGTGGTCCAGAATC-3' and (reverse) 2A 5'-AGTCTGTAATTTTAAGTCGG-3'.

DNA sequencing. DNA was sequenced using the Sequenase™ Version 2.0 DNA Sequencing Kit (United States Biochemical, Cleveland, OH).

RESULTS

As shown in the schematic in Figure II-1A, MuLV specific primers designed against the LTR element can anneal and amplify from either the upstream (5') or downstream (3') LTR, generating either internal virus amplification products or the desired viral/genomic junction fragments. To minimize the likelihood of internal virus sequence amplification resulting from universal primer binding within the virus genome, we incorporated a *PstI* anchor sequence into UP1. The ecotropic proviral genome contains two *PstI* sites, which are located 35 bp from the 5' terminus of each viral LTR. Therefore, because there are no binding sites for UP1 within the viral genome internal to the binding site for SP2 in the 3' LTR, the UP1 primer should not

generate internal virus amplification products (Fig. II-1B). Alternatively, the UP2 primer was designed with a *Xba*I recognition sequence. The ecotropic viral genome contains a *Xba*I cut site at position 7360 (Chattopadhyay *et al.*, 1980). This site lies approximately 500 bps internal of the 3' viral LTR, and thus UP2 would be expected to amplify a 523 bp internal virus product. To determine whether universal PCR using UP1 or UP2 generated viral and/or target amplification, Southern analysis of aliquots of universal PCR reactions were hybridized with either the SP2 oligo, for detection of all virus priming, or the VC1 internal virus control probe. As shown in Figure 2, the expected 523 bp virus product was amplified with UP2. In addition, a virus product of ~ 600 bp was amplified with the UP1 primer, as well as multiple smaller products that did not hybridize to the internal virus control probe VC1. Any of the products not recognized by VC1 potentially represent the desired viral/genomic junction fragment. One of these products, a 186 bp fragment generated with UP1, was cloned and sequenced. Figure II-3 shows a manual alignment of the sequence of this fragment with mouse genomic sequence from this locus (Chapter IV, and unpublished data). Sequence comparison reveals that the UP1 primer bound specifically to a *Pst*I sequence present in mouse genomic DNA from the unrearranged locus.

Using a probe containing only the mouse genomic DNA portion of the universal PCR product, designated pL1, Southern hybridization was performed on brain and tumor DNA from AKXD-18, animal 108. This analysis revealed that probe pL1 detects a tumor-specific restriction fragment length polymorphism, consistent in size with that of the somatically acquired retrovirus in the tumor sample (Fig. II-4), and therefore represents the locus disrupted by viral integration in this tumor.

DISCUSSION

The intent of our project is the identification of genes involved in cancer development through the specific amplification of genomic sequences flanking somatically acquired MuLVs in mouse lymphomas. Therefore, we chose to modify available universal PCR protocols to allow amplification from MuLV LTR sequences. In addition, we increased the specificity of our reactions by size fractionating the template DNA to enrich for the desired viral/genomic restriction fragment, as well as increasing the stringency of primer annealing in subsequent PCR reactions. These changes allowed efficient amplification of a MuLV insertion site from lymphoma tissue, and provide data for the development of an effective PCR-based cloning strategy for the isolation of insertion sites from large numbers of tumor samples.

As shown in Figure II-1, LTR specific primers can amplify both target and non-target sequences. A known binding site for the UP2 primer in the ecotropic MuLV predicted amplification of a 523 bp non-target amplification product, which was detected by hybridization with the VC1 probe (Fig. II-3). Although no *Pst*I sites are present upstream of the SP1 or SP2 binding sites in the ecotropic MuLV, a number of other viral species present in inbred mouse strains do contain *Pst*I sites (Chattopadhyay *et al.*, 1980) which are consistent with generation of a ~600 bp viral product using UP1, as detected by Southern analysis (Fig. II-3) (Chattopadhyay *et al.*, 1980). Thus, it should be possible to accurately predict the size of internal virus primed products, or non-target products, using available restriction enzyme mapping data from MuLV sequences. Alternatively, it may be possible to choose anchor sequences not present upstream of the 3' LTR in any known MuLV, such that no internal virus products are amplified, thus completely eliminating the need for hybridization screening of PCR products to identify target amplification.

The value of the universal PCR technique applied in this report stems from fact that it can be easily modified for essentially any application that requires amplification of unknown sequence flanking known sequence. By substituting nested gene-specific primers, it could be used to amplify and clone exon/intron boundaries. Using nested vector primers, it could be used to clone the ends of large genomic clones for physical mapping studies. In addition, it may be used to isolate sequences flanking other inserted elements, such as transgenic constructs or other viral or transposable elements in the mouse or other organisms.

III. MOLECULAR GENETIC LINKAGE ANALYSIS OF MOUSE CHROMOSOME 19

INTRODUCTION

The first evidence of genetic linkage in the mouse was observed between two visible coat color phenotypes, albino and pink-eyed dilution (Haldane *et al.*, 1915). Since then, a great deal of scientific effort has been focused on the assignment of genetic markers to individual chromosomes. Viewed simply as a genetic tool, linkage maps provide the means necessary to identify relationships between molecular markers and mouse mutations causing biologically interesting phenotypes. Molecular linkage maps, however, serve a number of other purposes. Mapping data and resources provide tools for assembling and confirming physical maps and for comparative genomic studies. Because the mouse serves as an excellent model for the study of human disease and disease genes, the comparison of molecular and phenotype mapping data from mouse and human serves to identify candidate genes for human disease phenotypes. Alternatively, in situations where candidate genes are yet unidentified, mapping studies provide molecular markers that can serve as points of entry for positional cloning strategies to identify the disease gene.

To map molecular markers it is first necessary to identify detectable sequence polymorphisms between the strains used for segregation analysis. For gene markers, this polymorphism is often a restriction fragment length polymorphism (RFLP) detectable by Southern blot analysis of genomic DNA. In addition to hybridization-based mapping, simple repeat sequences such as dinucleotide repeats are commonly used as molecular markers for mapping studies because the lengths of these repeat sequences vary greatly between species. These markers are typed by PCR using

primers flanking the repeat to detect simple sequence length polymorphisms (SSLPs). Because of the prevalence of dinucleotide repeat sequences in the mouse genome, hundreds of such markers are available for each chromosome. The power of this technique is evidenced by advances in experimental approaches such as quantitative trait analysis which requires large numbers of molecular markers to establish linkage with often subtle phenotypes. Prior to the advent of SSLP mapping, these experiments were often not feasible.

Due to the high degree of sequence variation between wild mouse strains and common inbred laboratory strains, the use of interspecific backcrosses (IB) between the wild mouse species *Mus spretus* and inbred laboratory stains provides an excellent system in which to identify and follow allele segregation to determine chromosome location. IB segregation patterns reflect a single meiosis, as opposed to a more than twenty for recombinant inbred strain panels. In addition, more statistical power can be achieved for IB panels because large numbers of animals can be collected for analysis, whereas the number of RI strains is often limited by the difficulty in developing and maintaining individual strains through more than 20 generations. IB mapping has been used extensively to build linkage maps of all mouse chromosomes (Copeland *et al.*, 1993), and these data, along with mapping data from other sources, are compiled and available for use from the Mouse Genome Database (http://www.informatics.jax.org/).

This chapter combines two manuscripts describing the use of IB mapping to generate molecular genetic linkage maps of mouse chromosome 19. The first paper was published in *Mammalian Genome* (Hansen and Justice, 1998) and presents data mapping a recently identified tumor suppressor gene, *Pten.* The second is a manuscript in preparation for submission to *Genomics.* This paper describes a detailed molecular linkage map of distal mouse chromosome 19 that refines and extends the molecular

linkage map of mouse chromosome 19, as well as the human chromosome 10q23-26 conserved syntenic region.

The purpose of these experiments was two-fold. First, *Lvis1*, the viral integration site identified in AKXD B-cell lymphoma (Chapter II), maps to distal mouse chromosome 19 (unpublished data, see Chapter IV). Thus, to provide additional information which help to determine whether *Lvis1* represents a novel or previously identified locus, I mapped known viral insertion sites, tumor suppressor genes and proto-oncogenes to mouse chromosome 19. In building further the chromosome 19 linkage map for positioning of *Lvis1*, I have mapped a number of genes to mouse chromosome 19 in collaboration with Dr. Charles Schwartz (Greenwood Genetics Center, Greenwood, SC). Another goal of these particular experiments was to identify candidate genes for the human split-hand/split-foot malformation that maps to human chromosome 10q24-25 (Nunes *et al.*, 1995). These analyses contribute toward a detailed molecular linkage map of mouse chromosome 19 for defining human-mouse conserved synteny, identifying candidate genes for known classical mouse mutations and human disorders, and for mapping *Lvis1*, our original goal.

Published in: *Manmmalian Genome*

III A. *Pten*, A CANDIDATE TUMOR SUPPRESSOR GENE, MAPS TO MOUSE CHROMOSOME 19

Gwenn M. Hansen and Monica J. Justice

Life Sciences Division, Oak Ridge National Laboratory,

Oak Ridge, Tennessee 37831, USA

Oak Ridge National Laboratory managed by Lockheed Martin Energy Research Corporation for the U.S. Department of Energy under contract No. DE-AC05-96OR22464.

METHODS AND RESULTS

Species Mouse

Locus name Phosphatase and tensin homolog

Locus symbol *Pten*

Map positions. *Pten* is located on central Chromosome (Chr) 19. The chromosomal order for the loci (with map distances in centimorgans ± the standard error, n=124) is: centromere-*D19Mit14*-17.7 ± 3.4 cM-*Pten*-2.4 ± 1.4 cM-*D19Mit12*-5.6 ± 2.1 cM-*Nkx2-3*-3.2 ± 1.6 cM-*Tlx1*-12.9 ± 3.0 cM-*D19Mit1*.

Method of mapping: Each of the loci in this study were mapped on an existing DNA panel of N_2 progeny from the (SB/Le x *M. spretus*)F_1 x SB/Le interspecific backcross (Justice *et al.*, 1990). SB/Le and *M. spretus* parental DNA was digested with several restriction enzymes and analyzed by Southern blot hybridization to identify at least one restriction fragment length polymorphism (RFLP) for each of the gene probes. To map Chr 19 specific simple sequence length polymorphisms (SSLPs), Mouse Map Pairs (Research Genetics, Huntsville, AL) identifying the Chr 19 specific SSLPs below were used to amplify genomic DNA from SB/Le, *M. spretus*, and N_2 backcross mice.

Molecular reagents used for mapping: The *Pten* probe used for hybridization experiments was a 896-bp cDNA fragment generated by RT-PCR using primers (forward) 5'-ATATCAAGAGGATGGATTCG-3' and (reverse) 5'-CCTTGTCATTATCTGCACGC-3' (Genbank Accession No. U92437). Total RNA from mouse liver was prepared using RNAZol B (Tel-Test, Inc., Friendswood, TX), and used to generate first strand cDNA template using the Gibco BRL SuperscriptII Preamplification System (Life Technologies, Inc., Gaithersburg, MD). PCR was performed using a MJ Research thermal controller, with the following cycling

parameters: 94°C for 3 min., followed by 30 cycles of 94°C for 30 sec., 60°C for 2 min., and 72°C for 2 min., with a final 7 min. incubation at 72°C. Each reaction contained 1 ul of first stand cDNA template, 200 ng of each primer, 1.5 mM MgCl, 0.2 mM each dNTP, 2.5 ul 10x PCR buffer (Gibco BRL), and 0.2 ul Taq DNA polymerase (Gibco BRL) in a volume of 25 ul. The *Nkx2-3* probe is a 1.3-kb genomic *Pst*I fragment cloned into pGEM3Zf(-) containing the *Nkx2-3* homeobox domain. The *Tlx1* probe is a 200-bp fragment amplified from human genomic DNA using primers (forward) 5'-CAGGCGACAGACAGCAGAGGAACG-3' and (reverse) 5'-TGACGCTGGTGATTTTGGTGGAGT-3'. This fragment contains exon 3 of the human HOX11 gene (Genbank Accession No. S38742).

Allele detection. The *Pten* probe identified a *Pvu*II polymorphism between SB/Le (1.8, 3.5 and 7.5 kb) and *M. spretus* (1.8, 3.5, and 4.2 kb). The probe for *Nkx2-3* identified a *Pvu*II polymorphism between SB/Le (5.4 kb) and *M. spretus* (4.4 kb). The probe for *Tlx1* identified a *Taq*I polymorphism between SB/Le (0.6, 1.8 kb) and *M. spretus* (0.6, 1.0 kb). Genomic DNA from N_2 backcross progeny was digested with the appropriate enzyme and analyzed by Southern blot hybridization with each of the above probes. All probes used for RFLP analysis were gel purified and labeled with [α-^{32}P]dCTP using a random prime labeling kit (Stratagene, La Jolla, CA) using standard procedures, with the exception of the *Tlx1* probe, for which 400 ng of each gene specific primer was substituted for the random primers of the labeling kit. Southern hybridizations were carried out at 65°C for the *Pten* and *Nkx2-3* probes, and at 55°C for the *Tlx1* probe (Church and Gilbert, 1984) with the following changes: Southern blots hybridized with the *Pten* and *Nkx2-3* probes were washed 3 times at 65°C with 0.2X SSCP/ 0.1% SDS for 30 min., Southern blots hybridized with the *Tlx1* probe were washed 2 times at 55°C with 2X SSCP/ 0.1% SDS, 1 time at 55°C with 1X SSCP / 0.1% SDS, and 1 time at 65°C with 1X SSCP / 0.1% SDS for 30

min. *D19Mit* markers were mapped using a previously published silver stain protocol (Liu *et al.*, 1994). *D19Mit1* amplified a 140-bp product from SB/Le parental DNA and a 155-bp product from *M. spretus* parental DNA, *D19Mit12* amplified a 150-bp fragment from SB/Le parental DNA and a 120-bp fragment from *M. spretus* parental DNA, and *D19Mit14* amplified a 140-bp product from SB/Le parental DNA and a 115-bp product from *M. spretus* DNA. Informative RFLPs or SSLPs in *M. spretus* DNA were monitored in 124 N_2 progeny (Fig. III A-1).

Previously identified homology: The human homolog of *Pten* has been mapped to human Chr 10q23.3 (Li, *et al.*, 1997; Steck *et al.*, 1997).

DISCUSSION.

We present data mapping *Pten* to mouse Chr 19 relative to two previously mapped gene loci, *Nkx2-3* and *Tlx1*, and three anonymous microsatellite markers in a single mouse interspecific backcross. *PTEN* is a protein tyrosine phosphatase recently identified as a putative tumor suppressor gene mutated in brain, breast, and prostate cancer (Li, *et al.*, 1997; Steck *et al.*, 1997). In addition, mutations in *PTEN* have been identified in patients with Cowden syndrome (CS) (Liaw *et al.*, 1997), a rare autosomal dominant disorder characterized by a predisposition to a variety of neoplasms and hamartomas, a benign disorganized mass of cells that arises within the tissue of origin (Lloyd and Dennis, 1963). With the identification of *PTEN* mutations in CS, it has been hypothesized that *PTEN* may be acting in the process of cellular organization. It is interesting to note that two classical mouse mutations that map to mouse Chr 19, aphakia (*ak*) and asebia (*ab*), display pathologies of cellular disorganization and benign proliferation (MGD, 1997). *ak* is a disorder of the developing eye in which the epithelium of the lens rudiment becomes disorganized due to abnormal proliferation of

lens cells, resulting in a lack of lens development (Zwaan *et al.*, 1975). *ab* is a disorder of the skin and hair, characterized by hyperplasia of the cellular layers of the skin and abnormalities of cells of the inner root sheath (Josofowicz, 1978). Additional study of *Pten* function would be aided by the identification of a mouse strain carrying a mutation in *Pten*. The similarities in pathology between *ak*, *ab*, and CS, combined with the localization of *Pten* to mouse Chr 19 make these classical mouse mutations candidates for *Pten* analysis.

ACKNOWLEDGMENTS.

The *Tlx1* probe was kindly provided by Dr. Charles Schwartz, Greenwood Genetics, Greenwood, SC. The *Nkx2-3* probe was kindly provided by Dr. Richard Harvey, The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia. This work is supported by a grant from the National Cancer Institute (R29CA63229), and by an American Cancer Society Junior Faculty Research Award (JFRA0553) to M.J.J.

III B. REFINEMENT OF THE GENETIC MAP OF THE MOUSE CHROMOSOME 19 - HUMAN CHROMOSOME 10Q23-26 HOMOLOGOUS REGION

ABSTRACT

Using an existing interspecific backcross between SB/Le and *M. spretus* mice, we have constructed a molecular genetic linkage map of mouse chromosome 19 using 23 molecular markers, six of which have not been mapped in the mouse: *Hex, msec15, Nfkb2, Nlz, Su(fu),* and *Wnt8b*. A number of markers known to map to human chromosome 10 were mapped in this study, and these data confirm and extend a significant region of conserved synteny between distal mouse chromosome 19 and human chromosome 10q23-q26. These data further refine available mouse and human linkage data and identify potential candidate genes for mouse and human phenotypes associated with this region.

INTRODUCTION

Many mouse chromosomes share significant blocks of homology with human chromosomes. These regions are extremely useful for the comparative mapping and identification of candidate genes for phenotypes in human and mouse. For example, one such conserved syntenic tract is shared between mouse chromosome 19 and human chromosome 10q23-26. This region contains many classical mouse mutations, of which at least seven have been localized within the region of mouse 19 - human 10q homology (MGD, June 1998). Many human disease phenotypes also map within this

region (OMIM, June 1998). A good example of the usefulness of comparative mapping in this region is shown in the case of Hermansky-Pudlak syndrome. This syndrome is reminiscent of Chediak-Higashi syndrome and is characterized by reduced pigmentation, platelet storage dysfunction, and kidney lysosome function disorders (MIM 203300). This disorder was mapped to human chromosome 10q23.1-q23.2 (Fukai *et al.*, 1995), and has been found to be the result of a mutation in a novel transmembrane protein, HPS (Oh *et al.*, 1996). Six classical mouse mutations exhibit both reduced pigmentation and platelet storage dysfunction; two of these mutations map to mouse chromosome 19. Candidate gene analysis of *Hps* in the *pale ear* (*ep*) *mouse* successfully identified an intracisternal A particle insertion within the coding region of the gene, which is proposed to be responsible for the mutant phenotype (Gardner *et al.*, 1997).

To further define the mouse chromosome 19 molecular linkage map, we mapped 23 loci on a single interspecific backcross. Of these, six had previously not been mapped in the mouse. These include: Hematopoietically expressed homeobox gene (*Hex*), the mouse homolog of yeast *sec15* (*msec15*), Nuclear factor of kappa-B subunit 2 (*Nfkb2*), the mouse homolog of the *Drosophila* gene *neuralized* (*Nlz*), Suppressor of fused (*Su(fu)*), and Wingless related MMTV integration site 8b (*Wnt8b*). The human homologs of *Nlz* (*h-neu*) and *Nfkb2* (*NFKB2*) have been implicated in oncogenesis (Nakamura *et al.*, 1998; Neri *et al.*, 1991); *Hex* and *Wnt8b* are thought to function in embryonic development and patterning (Thomas *et al.*, 1998; Lako *et al.*, 1998); *Su(fu)* is involved in the Hedgehog signaling pathway, and *msec15* is a subunit of the exocyst protein complex and is involved in exocytosis (Kee *et al.*, 1997). In addition, these data refine gene order for a number markers previously localized to mouse chromosome 19, and generate a detailed molecular linkage map of the mouse chromosome 19 - human chromosome 10q23-26 homologous region.

Mice. The interspecific backcross (SB/Le x *M. spretus*)F_1 x SB/Le has been described (Justice *et al.*, 1990).

Probes. Probes for *Nkx2-3*, *Pten,* and *Tlx1* have been described (Hansen and Justice, 1998).

The *Chuk, Fas*, and *Fratl* probes were generated by RT-PCR using primers listed in Table III B-1. Total RNA from mouse liver was prepared using RNAZol B (Tel-Test, Inc., Friendswood, TX), and used to generate first strand cDNA template using the Gibco BRL Superscript Preamplification System (Life Technologies, Inc., Gaithersburg, MD). PCR was performed using an MJ Research thermal controller, with cycling parameters: 94°C for 3 min, 30 cycles at 94°C for 30 sec, 60°C for 2 min, and 72°C for 2 min, with a final 7 min incubation at 72°C. Each reaction contained 1 µl of first stand cDNA template, 200 ng of each primer, 1.5 mM MgCl, 0.2 mM of each dNTP, 2.5 µl 10x PCR buffer (Life Technologies, Inc., Gaithersburg, MD), and 0.2 µl Taq DNA polymerase (Life Technologies, Inc., Gaithersburg, MD) in a final volume of 25 µl.

The pD1.1/14 probe is a 1.2 kb *Hin*dIII genomic DNA representing the *His2* locus (Askew *et al.,* 1991). The *Evi1* probe (SP) is a 400 bp *Sst*I-*Pst*I genomic fragment (Mucenski *et al.,* 1991). The *Fgf8* probe is a 0.8 kb *Xba*I-*Bam*HI cDNA fragment representing variant 4 (Crossley and Martin, 1995). The *Hex* and *PitX3* probes were amplified from B6 genomic DNA using the primers listed in Table III B-1. The *Notch2* probe is a 557 bp cDNA probe amplified from a 12.5d embryo cDNA library (Stratagene, La Jolla, CA) using the primers listed in Table III B-1. All other probes were mouse or human expressed sequence tags obtained from Research Genetics (Huntsville, AL) (Table III B-1).

Allele detection. Genomic DNA from N_2 backcross progeny was digested with the appropriate enzyme and analyzed by Southern blot hybridization with the probes listed in Table III B-1. All probes used for RFLP analysis were gel purified and labeled with [α-^{32}P] dCTP using the Prime-It® II Random Prime Labeling Kit (Stratagene, La Jolla, CA). Hybridizations were carried out at 65°C for all mouse DNA probes, and at 55°C for human probes (Church and Gilbert, 1984), with the following wash conditions: Southern blots hybridized with mouse probes were washed three times at 65°C with 0.2X SSCP/ 0.1% SDS for 30 min, Southern blots hybridized with human probes were washed twice at 55°C with 2X SSCP/ 0.1% SDS, once at 55°C with 1X SSCP / 0.1% SDS, and once at 65°C with 1X SSCP / 0.1% SDS for 30 min. *D19Mit* PCR products were visualized using a silver stain protocol (Liu *et al.*, 1994). Allele sizes for each marker are listed in Table III B-1. Informative RFLPs or SSLPs in *M. spretus* DNA were monitored in 130 N_2 progeny for markers mapped to chromosome 19 (Fig. III B-1). Chromosome 3 RFLPs were monitored in 70 N_2 progeny (Fig. III-3).

RESULTS

The loci in Table III B-1 were mapped on an existing DNA panel of N_2 progeny from the (SB/Le X *M. spretus*)F_1 x SB/Le interspecific backcross (Justice *et al.*, 1990). SB/Le and *M. spretus* parental DNA was digested with several restriction enzymes and analyzed by Southern blot hybridization to identify at least one restriction fragment length polymorphism (RFLP) for each of the 18 gene loci. SB/Le and *M. spretus* parental DNAs were amplified using PCR to identify simple sequence length polymorphisms (SSLP) for each of the seven anonymous microsatellite markers. Informative RFLPs or SSLPs, listed in Table III B-1, were monitored in 130 N_2 progeny for all chromosome 19 markers (Fig. III B-1). The most likely chromosomal

order for these loci (with map distances in centiMorgans ± the standard error, n=130) is: centromere-*D19Mit14*-14.6 ± 3.1 cM- *D19Mit13*- 3.1 ± 1.5 cM-*Pten*-1.5 ± 1.1 cM-*Fas*-0.8 ± 0.8 cM-[*His2, D19Mit12,*]-0.8 ± 0.8 cM-[*Hex, msec15, D19Mit19, D19Mit20*]-4.6 ± 1.8 cM-*Fratl*-0.8 ± 0.8 cM-*Nkx2-3*-0.8 ± 0.8 cM-*Chuk*-1.5 ± 1.1 cM-[*Pax2, Wnt8b, D19Mit27*]-1.5 ± 1.1 cM-*Tlx1*-0.8 ± 0.8 cM-*Fgf8*-0.8 ± 0.8 cM-[*Nfkb2, Pitx3, Su(fu)*]-0.8 ± 0.8 cM-*Neuralized*-10.8 ± 2.7 cM-*D19Mit1*. Brackets indicate loci that are not separated by crossovers (Fig. III B-2).

The *Hex* probe detected a second RFLP that did not cosegregate with markers on chromosome 19. This RFLP maps to mouse chromosome 3, between *Evi1 and Notch2*. The most likely chromosomal order for these loci (with map distances in centiMorgans ± the standard error, n=70) is: *Evi1*-27 ± 5.3 cM-*Hex-rs*-4.3 ± 2.4 cM-*Notch2*(Fig. III B-3).

DISCUSSION

We used an existing interspecific backcross between SB/Le and *M. spretus* mice to further define the molecular genetic linkage map of chromosome 19 and the mouse-human comparative map with 23 molecular markers, six of which have not been mapped in the mouse: *Hex, msec15, Nfkb2, Nlz, Su(fu),* and *Wnt8b*. Four gene markers localized in our study have not been mapped in the human: *msec15, Fratl, Nkx2-3,* and *Su(fu)*. The significant degree of conserved synteny observed between mouse chromosome 19 and human chromosome 10q implies that these genes map to human chromosome 10q23-26. In addition, although *Hex* has been mapped to human chromosome 10 using somatic cell hybrids (Hromas *et al.*, 1993), the specific chromosomal location had not yet been established. Our data indicate that *HEX* likely maps to human chromosome 10q24.

Comparison with the MGD Web Map

The gene order and map distances determined in our linkage study are in approximate agreement with the chromosome 19 linkage data available from the Mouse Genome Database (MGD; June 1998). However, some differences were identified. *Pten* was previously mapped relative to five markers included in this study (*D19Mit14, D19Mit12, Nkx2-3, Tlx1* and *D19Mit1*), but not relative to *Fas* (Hansen and Justice, 1998). Integration of this data into the MGD Web Map (http://www.informatics.jax.org/) placed the *Pten* locus distal to *Fas* (Fig. III B-2). This placement does not agree with our data, which locate *Pten* 1.5 cM proximal of *Fas.* Thus *Pten* is the most proximal mouse chromosome 19 marker known to map to human chromosome 10.

No crossovers were detected between *His2* and *D19Mit12* on our backcross, however these markers are placed 10 map units apart, at positions 36 and 26 respectively (MGD; June 1998). In addition, *D19Mit13* maps 3.1 cM proximal to *Pten* on our backcross, yet is located over seven map units distal of *Pten* (position 33) in the MGD Web Map (Fig. III B-2). A detailed mapping study of the classical mutation *bm* found no recombination between this mutation and the microsatellite marker *D19Mit13* (Rusiniak, *et al.,* 1996). The map location of *D19Mit13* determined in our study implies that *bm* maps proximal to *Pten* . Therefore it is possible that the human homolog of the gene responsible for the *bm* phenotype may not map to human chromosome 10q, as would be assumed from the MGD map location (MGD, June 1998).

Gene order (but not map distance) of all markers distal of *D19Mit20* was consistent with the placement determined in our study, except for the *Chuk* locus, which maps between *Nkx2-3* and *Pax2* on our backcross (Fig. III B-2).

Markers detecting multiple loci.

As noted in Table III B-1, the *Chuk* probe detected two RFLPs that did not cosegregate. These results are consistent with previous reports of a *Chuk* related locus, *Chuk-rs1*, mapping to mouse chromosome 16 (Mock *et al.*, 1995). The *Hex* probe also detected a second RFLP that segregated independently of the chromosome 19 locus. This RFLP cosegregated with markers on chromosome 3 (Fig. III B-3), indicating the presence of a *Hex* related sequence, *Hex-rs,* on mouse chromosome 3. Alternatively, the presence of a *Hex* pseudogene in the mouse has been reported, although it was not thought to be present in the *M. spretus* genome (Bedford *et al.*, 1993). Therefore our results could be interpreted as evidence supporting the presence of a *Hex* pseudogene in the *M. spretus* genome from which our backcross was derived.

Candidate genes for know mouse mutations.

Of the 16 known mouse mutations that map to chromosome 19 (MGD, June 1998), only two, *ep* and *lpr*, have been molecularly cloned (Adachi *et al.*, 1993; Wu *et al.*, 1993; Gardner *et al.*, 1997). The markers mapped in this study localize to the distal portion of chromosome 19. This region contains five classical mutations for which affected genes have not been identified: *aphakia* (*ak*), *brachymorphic* (*bm*), *asebia* (*ab*), *Dactylaplasia* (*Dac*), *and ruby-eye* (*ru*) (MGD; June 1998). *ak* is a recessive mutation which is characterized by homozygotes with small eyes lacking lens structure (Varnum *et al.*, 1968). *PitX3* is expressed in the developing lens and is mutated in both anterior segment mesenchymal dysgenesis (ASMD) and congenital cataracts in

humans (Semina *et al.*, 1998), and thus has been proposed as a candidate for the *ak* mutation (Semina, *et al.*, 1997).

ab is a spontaneous mutation that affects the skin and hair (Josefowicz *et al.*, 1978). Notably, hair growth in *ab* can be stimulated by transplantation of skin from a normal animal, suggesting that a diffusable factor is missing in *ab* mice (Gates *et al.*, 1965). *msec15* is the mouse homolog of rat *rsec15,* a component of the exocyst protein complex which functions in exocytosis (Kee *et al.*, 1997). A defect in exocytosis is consistent with the *ab* phenotype. The *ru* mutant is characterized by reduced pigmentation and platelet storage pool dysfunctions (Novak *et al.*, 1984), which is also consistent with a defect in exocytosis. However, the map location of *msec15* is not consistent with its involvement in *ru* (Fig. III B-2; MGD, June 1998). Alternatively, other *rsec* proteins have been localized to nerve terminals, and are thought to mediate neurotransmitter release (Hsu, *et al.*, 1996). Linkage to markers on human chromosome 10q23.3-24.1 has been reported for partial epilepsy (Ottman, *et al.*, 1995). An epilepsy phenotype could be associated with improper neuron signaling due to defects in neurotransmitter release, and thus *msec15* could be considered a candidate for this disorder.

Candidate gene analysis within this region has successfully uncovered the relationship of a number of previously unassociated syndromes mapped to human chromosome 10q24. The *PTEN* tumor suppressor gene, found to be deleted in a glioblastoma cell lines, has subsequently been found to be mutated in disorders such as Cowden disease and Bannayan-Zonana syndrome, as well as juvenile polyposis (Smith and Ashworth, 1998; Eng, 1998; and Parsons, 1998 for reviews). Further candidate gene analysis of existing and newly mapped loci will likely be useful for identifying additional disease genes in this region.

ACKNOWLEDGMENTS

Sequence of the *PitX3* primers was kindly provided by Dr. Jeffrey Murray, The University of Iowa, Iowa City, IA. This work is supported by a grant from the National Cancer Institute (R29CA63229), and by an American Cancer Society Junior Faculty Research Award (JFRA0553) to M.J.J.

IV. VIRAL INSERTION MUTATIONS AT *Lvis1* DISRUPT LOCAL GENE EXPRESSION AND IMPLICATE MULTIPLE GENES IN LYMPHOMA DEVELOPMENT

ABSTRACT

AKXD recombinant inbred mice develop a variety of leukemias and lymphomas due to retrovirally mediated insertional activation of cellular proto-oncogenes. These mice serve as valuable tools for the identification of genes involved in cancer development, as the retroviral element acts not only as a mutagen, but also as a molecular tag to provide ready access to the disrupted locus. Proviral tagging can be used to screen for sites disrupted by viral integration in multiple, independently arising neoplasias, thus identifying genes involved in the disease process. A locus disrupted by viral integration in B-cell lymphomas of AKXD mice, *Lvis1*, was cloned using universal PCR. *Lvis1* maps to distal mouse chromosome 19, in a region conserved with human chromosome 10q. This locus is disrupted by retroviral integration in lymphomas from 14 AKXD strains, and occurs in a total of 22% of B-cell lymphomas in AKXD mice. Random sequencing of BAC clones spanning >110 kb identified four genes near the site of viral insertion: *Hex*, *mEg5*, *msec15,* and *L35a*. Expression analysis of these genes in tumors with *Lvis1* alterations revealed that viral insertions differentially affect regulation of neighboring genes. *Hex*, an orphan homeobox gene involved in embryonic developmental patterning and hematopoiesis, is significantly upregulated in *Lvis1* tumors, whereas no *Hex* expression is detected in tumors lacking *Lvis1* alterations. Both *mEg5* and *msec15* show a slight upregulation in *Lvis1* tumors, while *L35a* expression appears unaffected. These data provide the first genetic

evidence for the possible role of these genes in lymphoma development, and highlight the use of genomic approaches for the study of position effect mutations.

INTRODUCTION

Insertional mutagenesis is a powerful genetic and molecular tool for the analysis of gene function. It is employed in both plant and animal model systems to evaluate the phenotypic effects of genetic alteration. In most model organisms, this approach is used primarily for germline mutagenesis. However, insertional mutation analysis is not limited to the study of germline mutations, but also is applicable to the study of somatic mutations. The most obvious genetic disease brought about by somatic mutation is cancer. Cancer results from an accumulation of multiple genetic alterations that ultimately disrupt processes that monitor and control cell growth. Much of the molecular study of human cancer is focused on the identification of genes altered by translocation or deletion events frequently observed in human tumors. With the rapid accumulation of genomic and expressed sequence data as a result of genome sequencing efforts, the positional cloning strategies required to identify genes disrupted by translocation or deletion events are becoming less laborious and more successful. However, model systems that allow direct access to disrupted loci simplify the process of disease gene identification.

Mouse models of cancer development have proven an effective tool for the molecular analysis of processes that also occur in human cancer. In the mouse, mobile DNA elements such as murine retroviruses induce somatic insertion mutations that contribute to the development of both hematopoietic and mammary cancers (Ihle *et al.*, 1990; van Lohuizen and Berns, 1990). By using the retrovirus as a molecular tag, it is possible to identify sites in the genome that are frequently altered in murine tumors.

The study of these sites has led to the identification of many proto-oncogenes as well as tumor suppressor genes. Not surprisingly, many genes found to be responsible for cancer development in the mouse have also been shown to play a role in human disease: *EVI1* (Morishita *et al.* 1992, Levy *et al.*, 1994, and Suzukawa *et al.*, 1994), *NF1* (Bader and Miller, 1978), and *HOXA9* (Nakamura *et al.*, 1996b and Borrow *et al.*, 1996).

Notably, although proviral tagging studies have identified many hematopoietic disease genes, few genes known to be involved in B-cell lymphoma development have been identified (van Lohuizen and Berns, 1990). Seven AKXD recombinant inbred (RI) mouse strains develop predominantly B-cell lymphoma (Mucenski *et al.*, 1986; Gilbert *et al.*, 1993). Of the 19 known oncogenes and/or previously identified viral integration sites that have been analyzed for lymphoma-specific alterations in AKXD B-lineage lymphomas, only four loci, *Pim1*, *Evi1*, *Evi5*, and *Gfi1*, show lymphoma-specific alterations (Mucenski *et al.*, 1988; Liao *et al.*, 1995; G.M. Hansen and M. J. Justice, unpublished results). No rearrangements have been found at loci known to be involved specifically in B-cell differentiation and disease (*Bcl2, Cbl, c-abl, Rag1, Gata2).* The insertion mutations detected at *Pim1*, *Evi1*, *Evi5*, and *Gfi1* occur in only 13% of the AKXD B-cell lymphoma samples analyzed, suggesting that the majority of loci that contribute to B-cell lymphoma development in these strains remains to be identified.

The first common site of retroviral integration to be identified by proviral tagging in mouse AKXD B-cell lymphomas was *Evi3* (Justice *et al.*, 1994). Integrations at *Evi3* are B-cell specific, and occur in 9% of the pre-B and B-cell lymphomas in the AKXD strains. *Evi3* integrations are observed almost exclusively in pre-B cell lymphomas from the AKXD-27 strain, and transcripts from this locus have been shown to be activated by viral insertion at *Evi3* (Justice *et al.*, 1994).

Our study reports the identification of a second site of retroviral integration in B-cell lymphoma, *Lvis1*. Proviral insertions at this locus have been detected in 14 AKXD RI strains, and in 22% of AKXD B-cell lymphomas. Thus, alterations at *Lvis1* likely play a significant role in B-cell lymphoma development.

METHODS

AKXD mice. The AKXD RI strains of mice were derived by Benjamin A. Taylor (The Jackson Laboratory, Bar Harbor, Maine). AKXD mice used to collect the panel of 270 AKXD lymphomas have been previously described (Mucenski *et al.*, 1986; Gilbert *et al.*, 1993). Additional AKXD mice used in this study were maintained in our colony at Oak Ridge National Laboratory.

DNA extraction and Southern blot hybridizations. High molecular weight genomic DNAs were extracted from frozen normal tissues, and tissues with lymphoma cell infiltration. Unless otherwise specified, bacteriophage and plasmid DNAs were purified by standard procedures (Sambrook *et al.*, 1989).

RNA isolation and Northern blot hybridizations. RNA was extracted from normal and lymphoma cell infiltrated tissues using RNAZol B (TelTest). Total RNA (10 μg) was fractionated by electrophoresis in 1.0% agarose gels containing 7% formaldehyde and transferred to Magna Charge™ nylon membranes (Micron Separations Inc.). The membranes were prehybridized and hybridized as previously described (Church and Gilbert, 1984).

Genomic clone isolation. Overlapping λ phage containing the *Lvis1* locus were isolated from a Lambda Dash II mouse genomic library prepared by partial digestion of 129/Sv DNA with *Sau*3A (Stratagene). A screen of 10^6 plaques gave three positive clones which were purified to generate a contig of > 30 kb of genomic DNA. BAC clones containing the *Lvis1* locus were obtained by screening Mouse BAC library filters (Research Genetics) by hybridization with probe pL1. Five positive clones representing four unique BAC clones were identified. Two additional BAC clones were isolated using a probe derived from random sequence clone G5 from BAC 595k13 (see shotgun library construction below). Insert length was determined for each clone by digestion with *Not*I followed by PFGE on a 1% SeaKem® GTG agarose gel in 0.5X TBEX (68 mM Tris base, 23 mM boric acid, 1.3 mM Na_2EDTA), with switch times of 6.8s - 13.6s. BAC ends were amplified using a linker-ligation PCR procedure (Mueller and Wold, 1989; Kere *et al.*, 1992), using T7 and/or Sp6 primers. Amplified fragments were cloned into the pCRII vector (Invitrogen) and sequenced using M13F and M13R primers.

Fiber FISH mapping. BAC clones 371o8, 102o18 and 174j3 were mapped using high resolution fiber FISH analysis by SeeDNA Biotech, Inc., Toronto, Canada (Heng, 1998). Briefly, lymphocytes were isolated from mouse spleen and cultured for 68 hr. at 37°C in RPMI 1640 medium supplemented with 15% fetal calf serum, 3 μg/ml concanavalin A, 10 μg/ml lipopolysaccharide and $5x10^{-5}$ M mercaptoethanol. Cells were hypotonically treated, fixed, and dropped onto slides. Slides were soaked in PBS for 1 min. and treated with 200 ul of alkaline solution. DNA fibers were stretched by pulling the edge of a cover slide across the surface of the cells. Stretched fibers were treated with one drop of methanol, soaked in a series of 70, 95, and 100% ethanol, and air dried. BAC clone 371o8 was biotinylated with dATP using the BRL BioNick

labeling kit (15°C, 2hr) (Heng *et al.*, 1992). BAC clones 102o18 and 174j3 were labeled using digoxigenin (Heng and Tsui, 1994). The procedure for FISH detection has been described (Heng *et al.*, 1992, Heng and Tsui, 1993).

Chromosomal mapping. Each of the loci in this study were mapped on an existing DNA panel of N_2 progeny from the (SB/Le x *M. spretus*) F_1 x SB/Le interspecific backcross (Justice *et al.*, 1990). Genomic DNA from N_2 backcross progeny was digested with the appropriate enzyme and analyzed by Southern blot hybridization with each of the probes used in this study. Probes for *Chuk, Fas, Fratl, His2, Nkx2-3, and Pten* have been described (Hansen and Justice, 1998, and Chapter III). The pL1 probe is a 126 bp fragment representing the *Lvis1* locus (Chapter II). Probes were gel purified and labeled with [α-^{32}P] dCTP using the Prime-It® II Random Prime Labeling Kit (Stratagene, La Jolla, CA), except for pL1, which was labeled using gene specific primers (Chapter II). Hybridizations were carried out at 65°C for all probes (Church and Gilbert, 1984), with the following wash conditions: Southern blots were washed three times at 65°C with 0.2X SSCP/ 0.1% SDS for 30 min. *D19Mit* PCR products were visualized using a silver stain protocol (Liu *et al.*, 1994). The pL1 probe identified a *Hin*dIII polymorphism between SB/Le (14 kb) and *M. spretus* (10 kb). Allele sizes for all other markers have been described (Chapter III). Informative RFLPs or SSLPs in *M. spretus* DNA were monitored in 130 N_2 progeny.

Shotgun library construction and plasmid template preparation. BAC DNA used for library construction was prepared by large-scale alkaline lysis, followed by CsCl gradient purification using standard procedures. DNA (20 µg) was sonicated in a 10% glycerol solution, and ethanol precipitated. Fragments were resuspended in 1X ligation buffer and combined with 2 µl of a 1 mM dNTP solution, 1.5 µl Klenow

and 1.5 µl T4 DNA Polymerase (GIBCO-BRL). The reaction was incubated at room temperature for 30 min., and inactivated by incubation at 70°C for 10 min. Samples were run on a 1.0% SeaKem® LE agarose gel (FMC Bioproducts) for size selection. Fragments of 1-3 kb were gel purified and blunt end ligated into the *Eco*RV cloning site of pBSKS+ (Stratagene). Ligation products were transformed into DH5α library efficiency competent cells (GIBCO-BRL). Sequence-ready plasmid DNA templates were prepared using the QIAprep 8 Turbo Miniprep Kit (Qiagen) on the BioRobot™ 9600 (Qiagen).

DNA sequencing. DNA sequencing was carried out using the PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit (Perkin Elmer) on the ABI Model 373 DNA Sequencer (Applied Biosystems). Sequence primers were either T3 or T7 sequencing primers, or synthetic oligomers derived from previously determined sequence. Each sequence was edited to remove vector and ambiguous sequence to a final length of no more than 600 bps with Sequence Navigator or Factura™, and was then used to search NCBI databases for nucleotide, EST, and protein similarity. In addition, each sequence was analyzed by GRAIL to detect potential coding sequences. The final sequence data set was analyzed using AutoAssembler™ 1.4.0 to identify sequence overlaps. Edited contigs were re-evaluated for known sequences by searching NCBI databases, and were also reanalyzed using GRAIL.

RESULTS

A technique for rapid cloning of viral insertion sites

Proviral tagging studies facilitate the identification of disease genes by marking them with viral sequences detectable by hybridization. Cloning the sequences flanking the integrated virus, however, requires the construction of genomic or subgenomic libraries, followed by multiple rounds of library screening to isolate the DNA fragment of interest. To eliminate the need for hybridization-based cloning we adapted a universal PCR technique to clone viral integration sites in the mouse. Universal PCR utilizes a highly degenerate primer containing a short anchor sequence based on restriction enzyme recognition sequences (Sarkar *et al.*, 1993; Beeman and Stauth, 1998). When coupled with a specific primer, in our case a primer specific for viral LTR sequences, the universal primer allows amplification of unknown mouse genomic sequence flanking the site of viral insertion. Using this technique we amplified the genomic sequences flanking a somatically acquired retrovirus in DNA isolated from B-cell lymphoma infiltrated spleen tissue from an AKXD-18 animal (Chapter II). A single copy probe (pL1) representing the site of viral insertion in this lymphoma was used to screen a panel of DNA from 270 AKXD lymphomas to identify alterations in other independently derived tumors. Brain and lymphoma-infiltrated tissue DNAs digested with *Xba*I and hybridized with pL1 revealed lymphoma-specific DNA rearrangements in 23 AKXD lymphomas (Fig. IV-1, and data not shown). Further Southern blot analysis using *Sac*I revealed three additional samples with tumor-specific rearrangements. Thus, the *Lvis1* locus is disrupted in 26 out of 270 samples, or 9.6% of AKXD neoplasias, and therefore defines a common site of retroviral integration.

Lvis1 most frequently altered site in AKXD B-cell lymphomas

Viral alterations at *Lvis1* were identified almost exclusively in lymphomas classified as B-cell in origin (Table IV-1). When compared to alterations detected at other known common sites of retroviral alteration in AKXD tumors, *Lvis1* is the most frequently disrupted locus in B-cell lymphomas (Table IV- 2). *Lvis1* mutations were also detected in a small number of other AKXD neoplasias (T-cell, and mixed T- and B- cell neoplasias), and are observed in neoplasias from 14 of 21 lymphomatous AKXD RI strains (Table IV- 3). Strains AKXD-13 and AKXD-18 show a high incidence of *Lvis1* alterations, where four out of six AKXD-18 B-cell lymphomas and six out of ten AKXD-13 B-cell lymphomas carry viral insertions at *Lvis1*. Viral insertion site strain preference has also been observed for mutations at *Evi1* and *Evi3* (Mucenski *et al.*, 1988, Justice *et al.*, 1994). These data indicate that individual AKXD RI strains are more susceptible to certain genetic alterations, suggesting that the AKR/J and DBA/2J parental strains carry lymphoma susceptibility alleles which have segregated in the resulting RI strains. In addition, *Lvis1* mutations show a lineage bias, as they are observed almost exclusively in B-cell lineage lymphomas. In the human, lineage specificity is frequently associated with specific genetic lesions in hematopoietic tumors and bone and soft tissue tumors (Rabbitts, 1994; Barr, 1998). In most instances, these specific genetic lesions involve translocation events where at least one of the two gene products altered as a result of the translocation is known to function in a lineage-specific developmental pathway (Barr, 1998). With respect to viral insertion mutations at *Lvis1*, this principle suggests that a gene at this locus may function in a B-cell lineage-specific developmental pathway.

Chromosomal location of *Lvis1*

The mouse chromosomal location of *Lvis1* was determined by interspecific backcross (IB) mapping using an existing panel of N_2 progeny from the (SB/Le x *Mus Spretus*)F1 x SB/Le backcross (Justice, *et al.*, 1990). An informative *Hin*dIII polymorphism between SB/Le and *M. spretus* was identified using the pL1 probe, and the segregation pattern of the *M. spretus* allele was monitored in 130 N_2 animals. The *Lvis1* segregation pattern was compared with segregation patterns of gene and microsatellite markers that scan the entire mouse genome. This analysis reveals that *Lvis1* maps to mouse chromosome 19. A number of genes with characteristics consistent with that of proto-oncogenes also map in this region (MGD, June 1998). We map *Lvis1* between two previously identified viral integration sites: *His2*, identified in myeloid cell lines (Askew, *et al.*, 1991), and *Frat1*, a viral integration site altered in advanced T-cell lymphomas (Jonkers *et al.*, 1996) (Fig. IV-2 and Chapter III-B). No other genes or loci known to be involved in cancer development have been mapped within this interval in the mouse (MGD, June 1998); thus *Lvis1* identifies a novel site of retroviral integration.

Mapping proviral insertion mutations at *Lvis1*

The pL1 probe was used to screen a genomic lambda phage library to obtain clones representing the normal, unrearranged *Lvis1* locus. These clones were mapped by restriction enzyme analysis, and partial sequence data was collected from single copy subclones to determine the precise location and orientation of provirus insertion in tumor DNA from animal 108 (Chapter II and Fig. IV-3). The remaining proviral insertions were mapped by Southern blot analysis and are shown in Figure 3. This analysis was done by comparing the sizes of restriction fragment length alterations detected with probe pL1 to the known restriction maps of MuLVs (Chattopadhyay *et*

al., 1982) . Similar to previous studies of B-cell lymphomas (Zijlstra *et al.*, 1986), the majority of inserted viruses were ecotropic MuLVs (Fig. IV-3). However, a number of inserted proviruses had restriction map patterns consistent with that of MCF viruses (Chattopadhyay *et al.*, 1982). These viruses are not encoded within the germline of the mouse, but are formed through recombination events between one ecotropic virus and at least two non-ecotropic viruses. Thus, the sequences and restriction maps of these viruses vary. In general, however, two basic restriction patterns are observed, reflecting either those viruses categorized as MCF class I viruses (known to infect thymic cells and accelerate AKR leukemia) or MCF class II viruses (not thymotropic or leukemogenic) (Cloyd *et al.*, 1980). The identification of both MCF class I and class II viruses at *Lvis1* indicates that both of these virus types can infect AKXD B-cells. Thus, MCF class I viruses are not exclusively thymotropic.

Provirus mapping studies revealed a slight bias with respect to direction of proviral gene transcription; however, viral insertions are spread across a region of approximately 6 kb, and lie in both transcriptional orientations (Fig. IV-3). Viral insertion mutations thought to alter neighboring gene expression through the action of the viral promoter elements would be expected to lie in the same transcriptional orientation and within the 5' UTR or coding region of the affected gene. The varied distribution and orientation of proviral sequences at *Lvis1* is similar to that observed at viral insertion sites thought to disrupt local gene expression by an enhancer activation mechanism (van Lohuizen and Berns, 1990).

In addition to the restriction mapping studies described above, each of the genomic clones representing the *Lvis1* locus was also analyzed to identify genes near the site of viral insertions. Single copy DNA fragments within the cloned region were identified and used as probes to hybridize Southern blots of genomic DNA representing a variety of animal species (Zoo blot). These experiments detect evolutionary sequence

conservation, which is a common characteristic of expressed sequences. Although multiple unique sequence probes were identified and tested within the cloned region (approx. 30 kb), none of the probes contained sequences conserved in multiple species (data not shown). These results suggest that insertion mutations at *Lvis1* may not lie within or immediately near coding sequences. This hypothesis is supported by the fact that viral integrations at *Lvis1* show a varied distribution indicative of an enhancer mechanism of gene disruption. Viral enhancer sequences have been shown to influence gene transcription bidirectionally and over large distances (Varmus, 1982; Bedell *et al.*, 1996), thus it is possible that the gene(s) affected by viral integrations at *Lvis1* lies at a distance to the site of insertions and is outside of the cloned genomic region. Alternatively, it is also possible that the cloned genomic region contains expressed sequences that we failed to identify as a consequence of the hybridization-based screening procedure. To address both of these possibilities we took advantage of available genomic techniques to scan a region of over 100 kb surrounding the site of *Lvis1* mutations by sample sequence analysis.

Physical mapping of *Lvis1* region

A larger genomic region representing the *Lvis1* locus was cloned by screening a BAC genomic library. Four unique clones at *Lvis1* were isolated using pL1, and two additional clones were isolated using a unique sequence probe from BAC clone 595k13. The size of each clone was determined by *Not*I digestion and PFGE analysis. The physical map was assembled by clone overlap as detected by Southern hybridization using single copy probes from BAC ends, as well as by high-resolution fiber FISH (Fig. IV-4). Restriction mapping with *Not*I and *Cla*I predicted that the cloned region spans 280-300 kb. This agrees with distance calculations from fiber

FISH mapping of clone 371o8 and 174j3, estimating the cloned distance to be between 260 - 280 kb (Fig. IV-5).

Sample sequencing method reveals candidate proto-oncogenes at *Lvis1*

The sample sequencing strategy used to identify genes at *Lvis1* was based on a proposed streamlined shotgun strategy for the identification of candidate genes over a large genomic region (Claverie, 1994). Initially, two BAC clones isolated with probe pL1 were chosen for library construction (595k13 and 542i7). Following PFGE analysis and Southern blot mapping of unique sequence probes from the ends of these clones, which localized each clone with respect to *Lvis1*, we focused the majority of our sequencing effort on BAC 542i7 (Fig. IV-4). Purified DNA from each BAC clone was randomly fragmented by sonication, and fragments in a range from 1-3 kb were subcloned into pBSKS+. Sequence was collected from both ends of each insert using the T7 and T3 primers for 50 random clones from BAC 595k13 and 240 random clones from BAC 542i7. Sequences were edited to a length of no more than 600 bps, to yield more than 2-fold sequence redundancy for the 110 kb region surrounding *Lvis1,* (BAC 542i7). Each sequence was searched individually against nucleotide and protein databases using the BLAST algorithms BLAST-nr, BLASTN-dbEST, and BLASTX-nr (Gish and States, 1993) to identify significant sequence similarity. Sequences were then assembled into contigs, some of which were placed on the physical map by Southern blot and PFGE analysis. Excluding matches with repetitive sequences, initial sequence searches identified significant similarity to database entries representing three genes: hematopoietically expressed homeobox gene *(Hex)*, human kinesin-related spindle protein *(HKSP),* and ribosomal protein L35a *(L35a)* (Table IV- 4).

The first candidate gene identified at *Lvis1* was *Hex.* Sequence from multiple random clones from both BAC 595k13 and 542i7 showed complete identity to

nucleotide database sequences for the *Hex* gene (Table IV- 4, and data not shown). Sequence from the first intron and exon 2 of the *Hex* gene aligned with sequence collected from the end of BAC 542i7. This overlap predicts that *Lvis1* lies downstream of the *Hex* gene (Fig. IV-6). The distance between *Hex* and *Lvis1* was determined by restriction enzyme mapping to be approximately 50 kb. This distance was also measured by fiber FISH mapping using clones 371o8 and 174j3. The pL1 probe which represents the *Lvis1* locus lies < 5 kb from the end of BAC clone 371o8 (Fig. IV-4, and data not shown). Clone 174j3 does not contain *Hex* gene sequences, such that the distance between this clone and clone 371o8 represents the distance between *Hex* and *Lvis1*, including the entire genomic length of the *Hex* gene. The average distance between these clones was determined to be 2.25 μm, or 52-53 kb, based on 10 measurements using the length of both clones as a standard (Fig. IV-5). These measurements are in agreement with the distance determined by restriction enzyme mapping.

Although the existing database entries for *Hex* contain only the coding region of the gene, our sequence analysis identified additional 3' UTR exon sequence by similarity to 3' UTR exons of the human homolog *PRH* (Genbank accession no. X67235) and to three mouse ESTs (Genbank accession no. AA985962, AA163146, and AA170521) (data not shown).

A second gene at *Lvis1*, designated *mEg5*, was identified by similarity to the human genes *HKSP* (Genbank accession no. U37426) and *KRP* (Genbank accession no. X85137). Exon sequence from a single random clone from BAC 595k13 was 85% and 81% identical over 168 bps to *HKSP* and *KRP*, respectively (database entries for HKSP and KRP are >99% identical, and likely denote the same gene). Genomic sequence from this random clone aligned with sequence collected from the end of BAC 44g13. This overlap, confirmed by both Southern and PFGE analysis, places *mEg5*

upstream of the *Hex* gene in the same transcriptional orientation (Fig. IV-6), approximately 75-80 kb from the site of viral insertions.

Similarity to mRNA sequence for mouse ribosomal protein L35a (Genbank accession no. Y16430) was detected with sequence from a single subclone from BAC 542i7. Sequence from this clone is 97% identical to the database entry for mRNA sequence of mouse *L35a* over the entire coding region, and is also highly similar to mRNA sequence from the rat *L35a* gene, as well as DNA sequence from two rat *L35a* pseudogenes (Genbank accession no. X05704 and X05705) (Fig. IV-6, and data not shown). Because the genomic sequence corresponding to the Genbank entry Y16430 is not available, we can not exclude the possibility that the mouse *L35a* gene is intron-less. Thus we propose that this clone represents *L35a* or a *L35a*-related gene. Alternatively, because this genomic sequence is not completely identical to the database entry for mouse *L35a* and does not contain intron sequence, it may represent a pseudogene of *L35a* in the mouse. This clone maps between *Hex* and *Lvis1* on the physical contig (Fig. IV-6).

To determine whether the region surrounding *Lvis1* contains additional genes not represented in the nucleotide, protein, or EST databases, the entire set of random sequences was also analyzed using GRAIL (Uberbacher and Mural, 1991). This analysis identified a number of sequences predicted to have an excellent likelihood of being transcribed according to either GRAIL 1, 1a, or 2. To further evaluate the significance of GRAIL positive clones, the entire set of sequence files was assembled using the AutoAssembler program and edited contigs were again analyzed by GRAIL. All edited contigs containing GRAIL predicted exons with a score of excellent were analyzed for unique sequence probes, which were used for Zoo blot and Northern blot hybridization experiments. Although a subset of these were conserved across species (data not shown), only one detected transcripts by Northern blot analysis. This contig

is present on BAC clones 371o8 and 542i7, placing it downstream of the integrated viruses (Fig. IV-6). Subsequent database searches with this exon sequence have found that it is 97% identical to the rat mRNA sequence for *rsec15* over 139 bps (corresponding to position 359-497 of Genbank accession no. AF032668).

Sequence of *mEg5*

To further evaluate *mEg5* as a candidate proto-oncogene at *Lvis1*, we determined the complete cDNA sequence of this gene. Initial Southern blot analysis revealed that probes containing *mEg5* exons detect two distinct loci in the mouse. This was confirmed by IB mapping experiments that verified the chromosome 19 map location, and localized a second, less intensely hybridizing locus to proximal mouse chromosome 6 (Chapter V, and data not shown). Thus, to determine the cDNA sequence of the chromosome 19 copy of *mEg5*, we used RACE analysis with primers designed from the two exons identified by shotgun sequence analysis (these exons correspond to positions 2090-2256 of the human HKSP mRNA sequence, and lie within the tail region of the predicted kinesin related protein). Cloning and sequencing of the largest 5' and 3' race products allowed the assemblage of a 4,728 bp cDNA (Fig. IV-7). The first in frame ATG codon is found at position 246 within a good Kozak translation initiation sequence, and is preceded by three in-frame TGA stop codons (positions 27, 274, and 208). The open reading frame predicts a 1,052 amino acid protein showing 77% identity to *HEg5*. *HEg5* is a member of the bim-C subfamily of kinesin related proteins. Members of this group of proteins are characterized by a highly conserved motor domain and are thought to be necessary for spindle formation during cell division. The motor domain of a potential mouse homolog of *HEg5* has been previously identified (Nakagawa *et al.*, 1997). This sequence, designated *Kif11*, is highly similar, but not identical to the mRNA sequence for the motor domain of

mEg5. In addition, the predicted amino acid sequence of the motor domain of *mEg5* is 98% identical to both *HEg5* and rat *KRP6*, whereas it is 97% identical to *Kif11* (Fig. IV-8). We therefore speculate that the *Kif11* motor domain sequence represents the chromosome 6 locus detected with our *mEg5* probe. Confirmation of this will await the cloning and sequencing of the complete *mEg5*- related gene on mouse chromosome 6.

Comparison of the tail region of *mEg5* with database sequences for other members of the bim-C subfamily of kinesin related proteins revealed that *mEg5* contains an evolutionarily conserved, threonine-containing cdc2 consensus phosphorylation site in the C-terminal tail region of the predicted protein (Fig. IV-9). Studies of *HEg5* have shown that this site is phosphorylated specifically during mitosis (Blangy *et al.*, 1995). This phosphorylation event regulates the proper localization of *HEg5* to the mitotic spindle during metaphase, and is mediated *in vitro* by $p34^{cdc2}$/cyclin B. Conservation of the cdc2 phosphorylation site within the *mEg5* sequence suggests possible functional homology between *HEg5* and *mEg5*. It should be noted, however, that this domain is not essential for spindle localization or protein function in all members of the bim-C family. At least one member of the bim-C subfamily, *cut7* (*S. pombe*; Hagan and Yanagida, 1990), shares the cdc2 phosphorylation site, yet neither mutational inactivation of residues essential for phosphorylation, nor deletion of the entire carboxy-terminal domain appears to affect spindle association or protein function (Drummond and Hagan, 1998). This result suggests that protein function predictions can not be based solely on conservation of functional domains within the bim-C family.

Viral integrations at *Lvis1* disrupt local gene expression differentially

To determine whether neighboring gene expression was altered as a result of *Lvis1* mutation, Northern analysis was performed using AKXD-18 lymphomas for

which RNAs were available. Northern analysis using the genomic clone containing *L35a* detected a single transcript of ~400 bps in spleen RNA, however no difference was detected in the expression of *L35a* in tumors with or without viral integrations at *Lvis1* (Fig. IV-10). Therefore, we eliminated this gene as a candidate for the *Lvis1* associated phenotype.

High levels of *Hex* expression are evident in AKXD lymphomas carrying *Lvis1* alterations, whereas no expression is detected in lymphomas lacking *Lvis1* alterations (Fig. IV-11). Analysis of *msec15* expression revealed two transcripts of 2.7 and 3.5 kb (Fig. IV-12), consistent with the sizes detected in rat tissues (Kee, *et al.*, 1997). These transcripts are expressed in tumor tissue at relatively low levels, but appear to be expressed at slightly higher levels in tumors with *Lvis1* alterations. A similar result was observed with probes representing the 3' UTR of *mEg5* (Fig. IV-13). This indicates that viral insertions at *Lvis1* differentially alter the expression of neighboring genes. *Hex,* which lies ~50 kb upstream of the viral insertions is markedly upregulated in tumors with *Lvis1* alterations. *msec15*, directly downstream of viral insertions, is only slightly upregulated, as is *mEg5*, which lies upstream of *Hex.* Notably, *L35a* expression does not seem to be affected.

DISCUSSION

We have employed a genomic approach to the analysis of viral insertion mutations associated with the development of B-cell lymphoma in mice. Using sample sequence screening, we have identified four genes in the genomic region surrounding *Lvis1*. These genes were physically mapped in relation to the site of viral integration and analyzed to determine the effect of virus insertion at *Lvis1* on gene expression. Interestingly, viral insertions did not alter neighboring gene expression equally.

Furthermore, expression of genes both upstream and downstream of integrated viruses was influenced. Expression of the *msec15* gene, downstream of the integrated viruses, appeared to be slightly upregulated. Genes involved in exocytosis have not been implicated in cancer, thus upregulation of *msec15* may not contribute to the observed lymphoma phenotype.

Kinesin proteins are molecular motors that are responsible for trafficking molecules within the cell. Members of the bim-C family of kinesin-related proteins are thought to play a role in spindle formation and stabilization during mitosis (Barton and Goldstein, 1996 for review). Members of this family of kinesin proteins are known to localize to the mitotic spindle, and share a highly conserved motor domain as well as a consensus phosphorylation site in the tail domain of the protein. *In vitro* studies of *HEg5* have shown that this site is phosphorylated by $p34^{cdc2}$ in a cell-cycle specific manner (Blangy *et al.*, 1995). Evidence from null alleles of *Drosophila klp61F*, and immunodepletion studies *of XEg5* have shown that members of this family are necessary for spindle separation during cell division (Heck *et al.*, 1993; Sawin *et al.*, 1992). Thus bim-C kinesin-related proteins are essential molecules in the cell-cycle pathway. To date, no experimental evidence has been found to implicate kinesin-related proteins in cancer development. However, because bim-C kinesin-related proteins function during cell division, the proximity of virally induced alterations to *mEg5* along with the resultant activation of expression of *mEg5* in B-cell lymphomas warrants further study of the effects of upregulation of kinesin motor proteins within hematopoietic cells.

The most striking effect on gene expression was observed for the *Hex* gene. *Hex* is an orphan homeobox gene that was identified in screens for homeobox genes expressed in hematopoietic cell lineages (Bedford *et al.*, 1993; Crompton *et al.*, 1992). *Hex* shows varied expression in certain hematopoietic cells, as it is not expressed in T-

cells, and appears to be downregulated in terminally differentiated B-cells (Manfioletti *et al.*, 1995). These data were used to support the hypothesis that the *Hex* gene plays a role in the processes of cellular differentiation in certain hematopoietic cell lineages. Recent studies, however, suggest that this gene may also play a role in patterning during early embryonic development (Thomas *et al.*, 1998). *Hex* gene expression is first detected in the primitive endoderm of the blastocyst, where it appears to be evenly distributed. However, as development proceeds, *Hex* expression becomes restricted to a small portion of the visceral endoderm at the distal tip of the egg cylinder. At the onset of gastrulation, 6.5 days post coitum, the subset of visceral endoderm cells expressing *Hex* appear to migrate unilaterally to assume an anterior position in the embryo, in direct opposition to *Brachyury* expression and the developing primitive steak. Thus, it has been proposed that *Hex* expression shows the earliest molecular anteroposterior asymmetry in the mouse embryo (Thomas *et al.*, 1998). In addition to this early asymmetric pattern of expression, *Hex* is also expressed in the thyroid primordium, is transiently expressed in the nascent blood islands of the visceral yolk sac, and later is expressed in embryonic angioblasts and endocardium (Thomas *et al.*, 1998).

Transcription factor activation is one of the most common genetic anomalies associated with cancer development. The *Hex* gene is most closely related to a second orphan homeobox gene, *HOX11*. Notably, this gene has been linked to lymphoma development as a consequence of its involvement in the chromosomal translocations t[(10;14)(q24;q11)] and t[(7;10)(q35;q24)] which are observed in human T-cell acute lymphoblastic lymphoma. These anomalies juxtapose the promoter elements from the T-cell receptor α or β genes with *HOX11* gene sequences, thereby activating the expression of *HOX11* (Dube, *et al.*, 1991; Hatano *et al.*, 1991; Kennedy *et al.*, 1991). Although studies have shown that *Hex* expression is elevated in certain human tumor

samples and in some hematopoietic cell lines derived from tumor tissue (Manfioletti, *et al.,* 1995), no direct evidence for the role of this gene in cancer has previously been reported.

Proviral tagging studies have also identified viral integration sites near other homeobox genes. The viral integration sites *Evi6* and *Evi7* are found within the *Hoxa* gene cluster, and activate the expression of *Hoxa9,* and *Hoxa7*, respectively (Nakamura *et al.*, 1996a). Further evidence for the role of *HOXA9* in cancer has been found from the analysis of t[(7;11)(p15;p15)] chromosomal translocations in acute myeloid leukemia. These translocations fuse the *HOXA9* gene with the nucleoporin gene *NUP98* (Nakamura *et al.*, 1996b; Borrow *et al.*, 1996). It is interesting to interpret the expression analysis of lymphomas carrying viral integrations at these sites in light of the genomic expression analysis presented in this paper. Specifically, expression analysis of RNA from leukemias harboring viral insertions at *Evi7* show that although viral insertions, in both transcriptional orientations, occur directly between the two closely linked genes *Hoxa7* and *Hoxa9*, viral sequences activate the expression of only one gene, *Hoxa7*. While it is reasonable to suggest that viral insertional mutations that act over long distances through an enhancer mechanism will upregulate all neighboring genes, this has not been borne out in our study or in the *Hoxa* viral insertion study. The classically proposed mechanism for the activation of genes through the action of viral LTR enhancer sequences predicts that the viral enhancer sequences within the LTR will act on the nearest promoter element. This dictates that such insertions should lie either in the same transcriptional orientation at the 3' end of the gene, or in the opposite transcriptional orientation at the 5' end of the gene to avoid positioning viral promoter elements between the viral enhancer and the host gene promoter (Nusse *et al.*, 1984; Wasylyk *et al.*, 1983). Clearly our data do not follow this classical enhancer activation mechanism, and imply that the action of

inserted LTR elements within a genomic context is more complex. The significant upregulation of *Hex* gene expression, as opposed to other neighboring genes, could indicate that the viruses have inserted within an enhancer region that controls *Hex* expression. This interpretation is supported by the fact that sequences 5' of the *Hex* gene do not contain endoderm-specific *Hex* gene enhancer elements (Tristan Rodriguiz and Rosa Beddington, personal communication). Alternatively, the viral insertion mutations at *Lvis1* may disrupt repressor sequences, thus allowing increased *Hex* expression. Analysis of potential gene regulatory sequences near *Lvis1* should help resolve these questions.

In summary, viral insertion mutations at *Lvis1* alter the expression of multiple genes. Although viral insertions appear to upregulate the expression of the *Hex* gene preferentially, it is possible that more than one gene at this locus may contribute to lymphoma development. Future studies evaluating the effects of ectopic expression of these genes individually, or in combination, will be necessary to determine whether or not unscheduled expression of *Hex, mEg5,* and/or *msec15* contributes to cancer development.

V. SUMMARY AND DISCUSSION OF SIGNIFICANT FINDINGS AND FUTURE PROJECT DIRECTIONS

SUMMARY OF EXPERIMENTAL FINDINGS

This work describes the identification and cloning of *Lvis1*, a site of retroviral integration in AKXD B-cell lymphomas. Analysis of *Lvis1* in 270 independently derived AKXD tumors revealed that this locus is frequently disrupted by viral integration. Alterations at *Lvis1* were detected in 26 tumor samples, accounting for 22% of AKXD B-cell lymphomas, and occurring in 14 of 21 lymphomatous AKXD strains. These results are significant, as *Lvis1* is the most frequently disrupted site in AKXD lymphomas (see Chapter IV, Table IV- 2). Previous studies have shown that sites frequently altered by viral integration in tumor tissue harbor genes that contribute to the development of cancer (van Lohuizen & Berns, 1990 and Lazo & Tsichlis, 1990 for reviews). In addition, many of the genes altered by viral integration in mouse cancers have been found to be disrupted by translocation and deletion events in human cancer (Morishita *et al.* 1992; Levy *et al.*, 1994; Suzukawa *et al.*, 1994; Bader and Miller, 1978; Nakamura *et al.*, 1996b; and Borrow *et al.*, 1996).

Lvis1 was mapped to mouse chromosome 19 using interspecific backcross analysis (Chapter IV). A detailed molecular genetic map of this region was constructed, allowing the positioning of *Lvis1* between two previously identified viral integration sites: *His2* and *Fratl*. These studies revealed that *Lvis1* is closely linked to two microsatellite markers, *D19Mit19* and *D19Mit20.* No genes known to be involved in cell growth, differentiation, or cancer have been mapped to the interval between *His2* and *Fratl* (MGD, June, 1998). These data suggested that the gene(s) altered by viral insertion mutations at *Lvis1* may represent a previously unidentified proto-oncogene.

Therefore, to identify potential candidate genes near *Lvis1* that may be responsible for the associated lymphoma phenotype, genomic sequences at this locus were cloned and analyzed using a random sample sequencing protocol (Claverie, 1994). Four genes were identified and placed on the physical map in relation to viral insertion mutations at *Lvis1*: *Hex, mEg5, L35a,* and *msec15.* The expression patterns these genes were evaluated in tumor samples with and without viral insertions at *Lvis1*, showing that viral insertions altered gene expression differentially. The *Hex* gene appears significantly upregulated in tumors carrying *Lvis1* integrations; the *mEg5* and *msec15* genes are slightly upregulated, while expression of *L35a* appears unaffected. These results identify *Hex, mEg5,* and, *msec15* as candidate proto-oncogenes. The role of transcription factors, similar to *Hex,* in cancer development has been well established (Rabbitts and Boehm, 1991; Barr *et al.,* 1998). *Hex* is an orphan homeobox gene most closely related to (mouse/human) *Tlx1/HOX11.* The *HOX11* gene has been implicated in the development of T-cell leukemia, as it is upregulated as a result of the t(10;14) translocation common in T-cell acute lymphoblastic leukemia (Dube *et al.,* 1991). Similarly, ectopic expression of *Tlx1* in mice causes T-cell acute lymphoblastic leukemia (Hawley *et al.,* 1997). Our studies suggest that *Hex* disruption should be evaluated in human leukemias and lymphomas.

A potential role for *mEg5* and *msec15* in cancer development has not been established. *mEg5* is a kinesin-related gene similar to members of the bim-C family. These genes are known to be involved in spindle formation during mitosis. However, the possible effects of upregulation of bim-C kinesin-related genes are unknown. *msec15* is a protein component in the exocyst protein complex, which is involved in processes of exocytosis. Genes involved in exocytosis have not been implicated in cancer. Future studies evaluating the effects of ectopic expression of these genes will

be necessary to determine whether or not unscheduled expression of one or more of these genes contributes to cancer development.

UNIVERSAL PCR: CURRENT USES AND FUTURE APPLICATIONS

The *Lvis1* locus was cloned using universal PCR. The design of this technique allows amplification of unknown genomic sequences flanking an integrated MuLV. It was developed to overcome the cumbersome physical methods classically used for insertion site cloning in the mouse. The experimental results outlined in chapter II revealed that the universal primers used in these experiments bind and amplify specifically from their intended sequence targets (restriction enzyme recognition sites). This is significant because one of the limiting factors in the use of this technique for amplifying and cloning tumor associated viral insertion sites in the mouse relates to the fact that MuLVs are flanked on both ends by LTR elements. Therefore, primers specific for an LTR element will generate products representing both flanking mouse genomic DNA and internal virus sequence (see Chapter II). Because we have demonstrated that universal primers amplify specifically from intended sequence targets, it will be possible to eliminate all internal virus amplification (non-target products) by designing a universal primer that contains an anchor sequence not present within the virus genome. In addition, as PCR techniques are extremely suitable for high throughput cloning, we can utilize this technique to amplify and clone most, if not all, viral insertion sites in the AKXD tumor-bank. These experiments will be essential for identifying the complete set of genes that are responsible for lymphoma development in these strains.

The universal PCR technique is also a valuable tool for other applications. We use the universal primer in combination with viral specific primers to amplify sequences

flanking viral insertion sites. The universal primers described in this report, however, can be used in combination with any set of nested specific primers to amplify unknown sequence flanking known sequence. For example, it could be used to clone sequences flanking an inserted transgene construct, to clone and sequence exon/intron boundaries, or to isolate the ends of large genomic clones for physical map construction. This technique is thus a very general tool that will likely prove very valuable for a variety of cloning applications.

RANDOM SEQUENCING AS A TOOL FOR GENE IDENTIFICATION

The experiments described in Chapter IV present data collected using random sample sequencing. This method of gene identification has developed as a consequence of recent advances in sequencing techniques and equipment, as well as concomitant advances in computer-based sequence analysis programs. In addition, the success of these methods is a direct reflection of the extent of available gene and protein sequence data. To clearly interpret the implications of the data that have been collected in this study, it is necessary to understand exactly how the data were generated, as well as to understand the theoretical basis of random sequencing methods.

A number of methods for large-scale genome sequencing have been developed. These methods commonly employ a random sequencing strategy, where the target region to be sequenced, be it an entire genome or an individual genomic clone (YAC, BAC, cosmid, etc...), is fractionated into hundreds or thousands of small (0.5 to 2.0 kb) overlapping fragments from which single sequence runs are collected. To determine the complete sequence of the target, the individual sequences are analyzed using computer-based sequence comparison programs to identify significant overlaps. When a sufficient number of sequence runs have been collected, the detectable overlaps

allow assemblage of contiguous sequence. This strategy is useful for large-scale sequencing projects, as the majority of steps necessary for random sequencing can be automated. However, these methods require the analysis of a large number of fragments (or subclones) to generate the necessary sequence overlap required for sequence assembly. The degree of sequence redundancy, or coverage, is determined by the size of the target sequence. For example, a 6-fold sequence redundancy is necessary to generate the completed sequence of a 20 kb target region. This means that 120 kb of actual sequence data must be collected from a library of random clones from this 20 kb target before contiguous sequence can be assembled. For larger target regions, the degree of redundancy necessary for complete sequence assembly increases linearly. Resultantly, the degree of redundancy necessary for the assembly of very large target regions becomes experimentally prohibitive. Therefore, most sequencing projects begin by collecting 6-8 fold sequence redundancy, and then employ either a primer directed method for the final stages of sequencing, or a second round of random sequencing carried out on smaller subclones of the target region containing those sequences not represented in the original assembly (these subsequent steps are referred to as "gap closure").

The methods described above have been used successfully to determine the entire genomic sequence of many organisms, and eventually both the human and mouse genomes will also be determined. Currently, however, these projects are not complete. Therefore, research projects requiring the knowledge of all genes within a particular region in either the mouse or human must incorporate available physical methods for gene identification. A number of techniques are used, including exon trapping, cDNA selection, and direct cDNA library screening using large genomic clones. All of these methods allow the identification of genes. However, no single method has proven effective in consistently identifying all genes within a target region. Because of the

advances in DNA sequencing technologies, many of the resources available to large-scale sequencing facilities are also available, in some degree, to individual research labs. Therefore, an additional method for the identification of genes within a genomic region has become available. This method is essentially a scaled-down version of genomic sequencing. It takes advantage not only of the technical methods used for large-scale sequencing, but also utilizes the current sequence analysis programs for evaluation of collected sequence.

As previously mentioned, determining the complete sequence of any genomic target region requires the researcher to collect a large amount of redundant sequence data. In addition, the degree of redundancy needed for assembly of contiguous sequence increases linearly with the size of the target, eventually becoming experimentally prohibitive. If the experimental goal is simply the identification of genes within a 20 kb genomic region, it is hardly necessary to collect the 120 kb of random sequence necessary for completed sequence. This fact becomes even more apparent when the distribution of sequence file overlaps within a project is analyzed. Remember that these methods require random fragmentation and cloning of hundreds of DNA fragments from the target region. These fragments are then partially sequenced, by generating a single sequence file of either one or both ends of the fragment, where the length of each sequence file is between 500 and 600 bps (the capacity of most available fluorescent sequencers). If a 6-fold redundancy is collected from a 20 kb region, then on average the entire target region should be represented six times in the set of sequences. Obviously the actual data does not contain six overlapping files of sequence for every part of the target. Instead, certain regions are represented many more than six times, whereas other regions may be represented only once. These data may allow the assembly of an accurate finished sequence, however, if the goal of the experiment is simply to identify all genes within the region, collecting tens of sequencing runs which

simply represent identical copies of previously collected sequence is both unnecessary and inefficient. Therefore, genomic sequencing methods are modified for the purposes of gene identification. Instead of collecting highly redundant data sets, these methods collect only enough sequence such that more than 90% of the target region is sampled. This quantity can be determined theoretically by plotting the fraction of target sequence sampled as a function of redundancy (Claverie, 1994). For targets of essentially any length, collection of 2- to 2.5-fold sequence redundancy allows the determination of 90 - 95% of the target sequence. An adequate sampling of sequence within a target region is therefore achieved with a low level of sequence redundancy.

In this study, I describe the use of this low redundancy random sequencing method to identify genes near *Lvis1*. Random sequences representing more than a 2-fold sequence redundancy were collected from a target region of 110 kb. This degree of redundancy, as outlined above, predicts that sequence information has been collected from 90 - 95% of the target region. These sequence files do not contain the necessary degree of sequence overlap for the assembly of contiguous genomic sequence. Completed genomic sequence of the *Lvis1* target region would require a total of 8-fold redundancy (more than 600 kb of additional random sequence) as well as an additional method for final gap closure. For purposes of gene identification, this additional sequencing is not justified. It has been clearly demonstrated that sampling and analysis of 90 - 95 % of target region sequence is more than sufficient for the identification of all genes within a target region (Claverie, 1994). However, although nearly all of the sequence contained in the genomic clone 542i7 has been sampled, it is possible that additional genes are present within BAC 542i7. Therefore, for added specificity, we combined our sequence analysis approach with hybridization based methods to increase the likelihood that an expressed sequence within our set of clones would be identified. This was carried out by analyzing all clones containing GRAIL-predicted exons for

evolutionary sequence conservation. Each clone was also used to screen Northern blots of tumor tissue RNA or RNA from a variety of adult mouse tissues. These additional experiments allowed the identification of the *msec15* gene prior to its deposit within the nucleotide database (see chapter IV), and therefore show that both GRAIL analysis and hybridization-based analyses can and should be used to complement sequencing strategies for a more complete genomic gene screen. Furthermore, as described in Chapter IV, sequences contained within the original lambda clone contig (see Chapter IV, Fig. IV-4 and Fig. IV-6) were screened using a hybridization-based strategy to identify potential genes within the immediate genomic region surrounding *Lvis1*. This region has not been found to contain any expressed sequence. As we have screened for genes within this region by two independent means, it appears highly unlikely that additional genes are present within this interval.

LIMITATIONS OF AVAILABLE DATABASE INFORMATION

As mentioned above, it is important to keep in mind the potential limitations of the nucleotide and protein databases. Although a number of entire genomes have been completely sequenced, many human and mouse genes have not yet been identified. Certainly not all genes or gene families are represented in the nucleotide and protein databases. Thus, the sequence of a gene may be represented within the sample sequencing data set, yet not be identified during database similarity searches. In addition, although a large number of genes not yet represented in the nucleotide and protein databases are present within the expressed sequence databases (dbEST), the majority of these cDNA clones are represented by single sequence files containing only 3' or 5' UTR sequence. For example, although we easily identified the *Hex* gene by similarity to nucleotide database entries, we would have been unable to correctly

identify this gene using only information from the dbEST database. The dbEST database contains three mouse ESTs representing the *Hex* gene. Each of these EST clones are represented in the database by approximately 300 bps of sequence from the extreme 3' UTR of *Hex*. As none of the nucleotide database entries for the *Hex* gene contain 3' UTR sequences, these EST clones are not identified as *Hex* ESTs in dbEST. Fortunately, the set of sequence files we collected during sample sequencing at *Lvis1* contained portions of every coding exon of the *Hex* gene, as well as 3' UTR sequence. Notably, no other EST clones were detected during database searches with sequences generated in this project.

It is obvious that even when the EST databases are essentially complete, they may not be particularly useful for gene identification strategies using random sample sequence analysis unless the entire cDNA sequences are collected and deposited in the database. Remember that low redundancy sample sequencing strategies sample 90 - 95% of a genomic region. Therefore the random sequence files will likely contain one or more exons from any gene within the region, but not all exons. Because EST databases contain sequences collected nearly exclusively from the 3' and/or 5' end of cDNA clones, successful gene identification can only be achieved if these sequences are sampled. A random sample sequencing strategy can not guarantee this. Thus, these strategies must incorporate at least one analysis program that allows gene or exon identification by sequence attributes rather than by similarity to known gene sequences. For this purpose we analyzed our sequences using GRAIL (Uberbacher and Mural, 1991), and then further by hybridization. Combining these methods has allowed a very thorough genomic analysis of this region that takes advantage of all available sequencing and sequence analysis techniques.

FUTURE EXPERIMENTS TO TEST CANDIDATE GENES AT *LVIS1*

The level of expression of each of the genes identified near the *Lvis1* locus was determined in hematopoietic tumors with or without viral insertions at *Lvis1*. These experiments revealed that viral insertion mutations at *Lvis1* are associated with increases in the level of expression of neighboring genes. To directly determine whether unscheduled expression of any of these candidate genes affects the processes of cellular growth and/or differentiation, we have begun transgenic studies of both *Hex* and *mEg5*. I have developed transgenic constructs for both the *Hex* and *mEg5* genes to express each gene ubiquitously, using the human beta-actin gene promoter, or specifically within hematopoietic cell lineages, using the immunoglobulin gene enhancer sequences . This will allow us to evaluate each gene's ability to influence cell growth within all cell lineages as opposed to the lineages in which *Lvis1* mutations were identified. We would expect that if either *Hex* or *mEg5* contributes to cancer development, transgenic lines over-expressing these genes would have an increased incidence of lymphoma development. It is also possible that these lines may develop other types of cancer. By determining the cell lineage(s) of potential tumors, we may be able to determine whether the observed B-cell lineage bias associated with *Lvis1* mutations is a consequence of functional restrictions of *Hex* or *mEg5*. It will also be necessary to test *msec15*. Although I have not made transgenic constructs to evaluate the role of this gene in cancer development, these experiments will also be necessary to determine the possible role of this gene in lymphoma development. *msec15* transgenics may also be useful in a more basic sense to simply determine more about the function of this protein.

Transgenic studies are commonly used to test candidate oncogenes. By upregulating the level of expression of genes such as *Myc*, *Ras*, *Pim1*, and *Tlx1* in

transgenic mice, each of these genes have been shown to potentiate tumor development. Furthermore, these transgenic lines are now being used to identify cooperating oncogenes by crossing transgenic lines or infecting these lines with oncogene-bearing retroviruses (van Lohuizen and Berns, 1991 for review). Therefore our transgenic studies will not only allow us to determine if upregulation of candidate genes at *Lvis1* potentiate lymphoma development, but by crossing these transgenic lines we will be able to determine if *Hex, mEg5* and/or *msec15* cooperate in oncogenesis. In addition, they also will be useful in future experiments to identify other potential synergizing oncogenes.

ADDITIONAL GENE FUNCTION ANALYSES

Identifying genes that contribute to cancer is the first step toward an understanding of this disease. Once an oncogene has been identified, it then becomes necessary to determine the normal function of the gene, such that a more detailed understanding of the processes that contribute to disease progression can be determined. For example, identifying genes that are activated by transcription factors such as *Hex* may identify additional molecules that contribute to the disease process. Furthermore, these studies may identify possible targets for molecular assays of disease diagnosis or for therapeutic strategies for disease intervention.

Most gene function analyses begin with the generation of a null allele. This work is currently underway for the *Hex* gene, and is being carried out in the laboratory of Dr. Rosa Beddington. Although the null phenotype has not yet been reported, this mutation will likely be embryonic lethal. This hypothesis is based on the fact that *Hex* is expressed in a distinctive pattern very early during embryonic development (Thomas *et al.*, 1998). Although expression patterns do not dictate function, they do determine

where a gene can function. If this gene has a significant function within the embryonic cells in which it is expressed, these cells will be disrupted due to a lack of *Hex* expression, producing an embryonic phenotype. It is also very important to note, however, that during our mapping studies, we identified a second locus in the mouse that could be detected with *Hex* exon probes during Southern analysis (Chapter III B). This locus was mapped to mouse chromosome 3. If this is a *Hex*-related gene, it is possible that it may be able to compensate functionally for the *Hex* gene. This would result in a normal, or less severe *Hex-/Hex-* phenotype. Such a result would be misleading if it is taken to mean that the *Hex* gene does not have a significant function. Regardless, it will be necessary to determine the characteristics of the locus on chromosome 3 for proper interpretation of *Hex* knock-out experiments. If the *Hex* null phenotype is embryonic lethal, studies that allow analysis of gene function using less severe alleles (point mutations, conditional knock-outs, etc...) may become extremely valuable for evaluating this gene's function(s).

Studies have been done to simulate a null phenotype for *HEg5* using immunodepletion (Sawin *et al.*, 1992). These studies have shown that this gene is essential for cell division. Therefore, knock-out studies of the *mEg5* gene would be expected to produce a lethal phenotype. However, this assumption is plagued by the fact that a second highly related locus is present in the mouse, designated *mEg5-2* (Chapter VI). We have mapped this locus to mouse chromosome 6, near a classical mouse mutation named *hop*(Fig. V-1). In addition, previous reports of a mouse homologue of *HEg5* (Nakagawa *et al.*, 1997) as well as Northern blot analysis carried out in our laboratory (data not shown) indicate that this locus encodes an expressed gene that is very similar, but not identical to *mEg5*. This *mEg5*-related gene could possibly have an identical or similar function, and thus could potentially compensate for *mEg5* gene function in a null mutant. Notably, the mRNA transcripts representing

mEg5-2 are smaller than those representing *mEg5*, the locus identified near *Lvis1*. This may indicate that *mEg5-2* has a different protein structure, or lacks some of the domains present in *mEg5*. Taken together, these data suggest that cloning and analysis of *mEg5-2* will be necessary before functional studies for either of these genes can be proposed. This work is currently underway in our laboratory. It also will be interesting to evaluate *mEg5-2* in the *hop* mutant. *hop* homozygotes walk with a characteristic hopping gait, moving their hind legs simultaneously (MGD, June 1998). In addition to a number of other abnormalities, analysis of sperm cells from homozygous male animals has shown that these cells fail to complete the second meiotic division (Johnson *et al.*, 1971). Because kinesin related proteins are known to be involved in spindle formation and stabilization during cell division, *mEg5-2* appears to be a likely candidate gene for this mutation. Analysis of *mEg5-2* as a candidate gene will be possible as soon as cloning and expression studies of this gene have been completed.

IDENTIFICATION OF A POTENTIAL GENE REGULATORY REGION

Physical mapping studies localizing *Lvis1* mutations in relation to neighboring genes revealed that insertion mutations lie approximately 50 kb downstream of the *Hex* gene. This is interesting, as these mutations appear to cause a significant upregulation of *Hex* gene expression. This could be explained if viral elements have integrated near and are acting to amplify endogenous *Hex* enhancer elements. To test this hypothesis, we could evaluate sequences at and near the site of *Lvis1* mutations for their ability to drive the expression of a reporter gene. This would be done by cloning genomic DNA fragments from this region into a reporter gene construct containing a minimal promoter. Expression of the reporter gene could be evaluated in cells in culture, or in

the whole animal. Using the latter approach, an elegant study of the regulatory regions of the *Bmp5* gene was conducted (DiLeone *et al.*, 1998). This study revealed that *Bmp5* has a complex expression pattern that is mediated by multiple enhancer elements found within a region of over 100 kb 3' of both the coding and non-coding exons of this gene (DiLeone *et al.*, 1998). We would therefore expect that a similar study of regions 3' of *Hex* would reveal whether *Hex* gene enhancer sequences are present near the site of *Lvis1* insertions.

CONCLUSION

In final conclusion, this dissertation focuses on the genomic analysis of a region implicated in lymphoma development. Our results suggest, in part, that genomic analysis allows a more thorough evaluation of mouse viral insertion sites and that several previously identified sites may benefit from a "wider" genomic view (especially those at which genes have not been identified). In our case, a simple evaluation of only those sequences immediately flanking the site of insertion would have failed to uncover any of the genes within this region. We therefore would not have identified the *Hex* gene, and concomitantly we would not have discovered that viral insertions at *Lvis1* significantly increase the expression of *Hex*. By using a genomic approach we identified multiple genes as well as a potential gene regulatory region. Further, this work has implicated *Hex* and possibly *mEg5* in cancer for the first time.

Ongoing studies will focus on determining how these genes function in the whole organism, how they operate in developmental pathways, and whether misregulation of these genes contributes to the development of cancer.

REFERENCES

Adachi, M., Watanabe-Fukunaga, R., and Nagata, S. (1993). Aberrant transcription caused by the insertion of an early transposable element in an intron of the *Fas* antigen gene of *lpr* mice. Proc Natl Acad Sci 90, 1756-1760.

Askew, D., Bartholomew, C., Buchberg, A., Valentine, M., Jenkins, N., Copeland, N., and Ihle, J. (1991). *His-1* and *His-2*: identification and chromosomal mapping of two commonly rearranged sites of viral integration in a myeloid leukemia. Oncogene 6, 2041-2047.

Bader, J., and Miller, R. (1978). Neurofibromatosis and childhood leukemia. J Pediatr 92, 925-929.

Barr, F. (1998). Translocations, cancer and the puzzle of specificity. Nat Genet 19, 121-124.

Barton, N., and Goldstein, L. (1996). Going mobile: microtubule motors and chromosome segregation. Proc Natl Acad Sci 93, 1735-1742.

Bedell, M., Jenkins, N., and Copeland, N. (1996). Good genes in bad neighbourhoods. Nat Genet 12, 229-232.

Bedford, F., Ashworth, A., Enver, T., and Wiedemann, L. (1993). *HEX*: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human. Nucleic Acids Res 21, 1245-9.

Bedigian, H., Taylor, B., and Meier, H. (1981). Expression of murine leukemia viruses in the highly lymphomatous BXH-2 recombinant inbred mouse strain. J Virol 39, 632-640.

Beeman, R., and Stauth, D. (1997). Rapid cloning of insect transposon insertion junctions using 'universal' PCR. Insect Mol Biol 6, 83-88.

Berns, A. (1988). Provirus tagging as an instrument to identify oncogenes and to establish synergism between oncogenes. Arch Virol 102, 1-18.

Berns, A. (1991). Tumorigenesis in transgenic mice: identification and characterization of synergizing oncogenes. J Cell Biochem 47, 130-135.

Bingham, P., Levis, R., and Rubin, G. (1981). Cloning of DNA sequences from the white locus of *D. melanogaster* by a novel and general method. Cell 25, 693-704.

Bishop, J. (1987). The molecular genetics of cancer. Science 235, 305-311.

Blangy, A., Lane, H., d'Herin, P., Harper, M., Kress, M., and Nigg, E. (1995). Phosphorylation by p34cdc2 regulates spindle association of human Eg5, a kinesin-related motor essential for bipolar spindle formation in vivo. Cell 83, 1159-1169.

Borrow, J., Shearman, A., Stanton, V. J., Becher, R., Collins, T., Williams, A., Dube, I., Katz, F., Kwong, Y., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J., and Housman, D. (1996). The t(7;11)(p15;p15) translocation in acute myeloid

leukaemia fuses the genes for nucleoporin *NUP98* and class I homeoprotein *HOXA9*. Nat Genet 12, 159-167.

Chattopadhyay, S., Lander, M., Rands, E., and Lowy, D. (1980). Structure of endogenous murine leukemia virus DNA in mouse genomes. Proc Natl Acad Sci 77, 5774-5778.

Chattopadhyay, S., Lander, M., Rands, E., and Lowy, D. (1980). Structure of endogenous murine leukemia virus DNA in mouse genomes. Proc Natl Acad Sci 77, 5774-5778.

Church, G., and Gilbert, W. (1984). Genomic Sequencing. Proc Natl Acad Sci 81, 1991-1995.

Claverie, J. (1994). A streamlined random sequencing strategy for finding coding exons. Genomics 23, 575-581.

Cleary, M., Chao, J., Warnke, R., and Sklar, J. (1984). Immunoglobulin gene rearrangement as a diagnostic criterion of B-cell lymphoma Proc Natl Acad Sci 81, 593-597.

Cloyd, M., Hartley, J., and Rowe, W. (1980). Lymphomagenicity of recombinant mink cell focus-inducing murine leukemia viruses. J Exp Med 151, 542-552.

Cloyd, M. (1983). Characterization of target cells for MCF viruses in AKR mice. Cell 32, 217-225.

Coffin, J. (1992). Retroviral DNA integration. Dev Biol Stand 76, 141-151.

Cooley, L., Kelley, R., and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P elements. Science 239, 1121-1128.

Copeland, N., Mucenski, M., Taylor, B., and Jenkins, N. (1985). Molecular genetic approaches to the study of murine lymphoma. In Banbury Report 20: Genetic manipulation of the early mammalian embryo (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), pp. 79-91.

Copeland, N., Jenkins, N., Gilbert, D., Eppig, J., Maltais, L., Miller, J., Dietrich, W., Weaver, A., Lincoln, S., and Steen, R. (1993). A genetic linkage map of the mouse: current applications and future prospects. Science 262, 57-66.

Copeland, N., Jenkins, N., and Harvey, R. (1994). The murine homeobox genes *Nkx2.3* and *Nkx2.6* are located on chromosomes 19 and 14, respectively. Genomics 22, 655-656.

Crompton, M., Bartlett, T., MacGregor, A., Manfioletti, G., Buratti, E., Giancotti, V., and Goodwin, G. (1992). Identification of a novel vertebrate homeobox gene expressed in haematopoietic cells. Nucleic Acids Res 20, 5661-5667.

Crossley, P., and Martin, G. (1995). The mouse *Fgf8* gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. Development 121, 439-451.

Davis, A., and Justice, M. (1998). An Oak Ridge legacy: the specific locus test and its role in mouse mutagenesis. Genetics 148, 7-12.

DiLeone, R., Russell, L., and Kingsley, D. (1998). An extensive 3' regulatory region controls expression of *Bmp5* in specific anatomical structures of the mouse embryo. Genetics 148, 401-408.

Drummond, D., and Hagan, I. (1998). Mutations in the bimC box of *cut7* indicate divergence of regulation within the bimC family of kinesin related proteins. J Cell Sci 111, 853-865.

Dube, I., Kamel-Reid, S., Yuan, C., Lu, M., Wu, X., Corpus, G., Raimondi, S., Crist, W., Carroll, A., and Minowada, J. (1991). A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid neoplasias with chromosomal translocation t(10;14). Blood 78, 2996-3003.

Eng, C. (1998). Genetics of Cowden syndrome: through the looking glass of oncology. Int J Oncol 12, 701-710.

Favor, J., and Morawetz, C. (1992). Insertional mutations in mammals and mammalian cells. Mutat Res 284, 53-74.

Feldman, B., Reid, T., and Cleary, M. (1997). *Pim1* cooperates with *E2a-Pbx1* to facilitate the progression of thymic lymphomas in transgenic mice. Oncogene 15, 2735-2742.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev 5, 1513-1523.

Fukai, K., Oh, J., Frenk, E., Almodovar, C., and Spritz, R. (1995). Linkage disequilibrium mapping of the gene for Hermansky-Pudlak syndrome to chromosome 10q23.1-q23.3. Hum Mol Genet 4, 1665-1669.

Gardner, J., Wildenberg, S., Keiper, N., Novak, E., Rusiniak, M., Swank, R., Puri, N., JN, F., Hagiwara, N., Lehman, A., Gales, T., Bayer, M., King, R., and Brilliant, M. (1997). The mouse pale ear (*ep*) mutation is the homologue of human Hermansky-Pudlak syndrome. Proc Natl Acad Sci 94, 9238-9243.

Gates, A., and Karasek, M. (1965). Hereditary absence of sebaceous glands in the mouse. Science 148, 1471-1473.

Gibbs, R. (1995). Pressing ahead with human genome sequencing. Nat Genet 11, 121-125.

Gilbert, D., Neumann, P., Taylor, B., Jenkins, N., and Copeland, N. (1993). Susceptibility of AKXD recombinant inbred mouse strains to lymphomas. J Virol 67, 2083-2090.

Gish, W., and States, D. (1993). Identification of protein coding regions by database similarity search. Nat Genet 3, 266-272.

Gray, D. (1991). Insertional mutagenesis: neoplasia arising from retroviral integration. Cancer Invest 9, 295-304.

Hagan, I., and Yanagida, M. (1990). Novel potential mitotic motor protein encoded by the fission yeast *cut7+* gene. Nature 347, 563-566.

Haldane, J., Sprunt, A., and Haldane, N. (1915). Reduplication in mice. J Genet 5, 133-135.

Hansen, G., and MJ, J. (1998). *Pten*, a candidate tumor suppressor gene, maps to mouse chromosome 19. Mammal Genet 9, 88-90.

Hatano, M., Roberts, C., Minden, M., Crist, W., and Korsmeyer, S. (1991). Deregulation of a homeobox gene, *HOX11*, by the t(10;14) in T cell leukemia. Science 253, 79-82.

Haupt, Y., Harris, A., and Adams, J. (1992). Retroviral infection accelerates T lymphomagenesis in *E mu-N-ras* transgenic mice by activating *c-myc* or *N-myc*. Oncogene 7, 981-986.

Heck, M., Pereira, A., Pesavento, P., Yannoni, Y., Spradling, A., and Goldstein, L. (1993). The kinesin-like protein KLP61F is essential for mitosis in Drosophila. J Cell Biol 123, 665-679.

Heng, H., Squire, J., and Tsui, L. (1992). High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. Pro Natl Acad Sci 89, 9509-9513.

Heng, H., and Tsui, L. (1993). Modes of DAPI banding and simultaneous in situ hybridization. Chromosoma 102, 325-332.

Heng, H., and Tsui, L. (1994). FISH detection on DAPI-banded chromosomes. Methods Mol Biol 33, 35-49.

Heng, H., and Tsui, L. (1998). High resolution free chromatin/DNA fiber fluorescent in situ hybridization. J Chromatogr A 806, 219-229.

Herr, W., and Gilbert, W. (1983). Somatically acquired recombinant murine leukemia proviruses in thymic leukemias of AKR/J mice. J Virol 46, 70-82.

Hicks, G., Shi, E., Li, X., Li, C., Pawlak, M., and Ruley, H. (1997). Functional genomics in mice by tagged sequence mutagenesis. Nat Genet 16, 338-344.

Hromas, R., Radich, J., and Collins, S. (1993). PCR cloning of an orphan homeobox gene (*PRH*) preferentially expressed in myeloid and liver cells. Biochem Biophys Res Commun 195, 976-83.

Hsu, S., Ting, A., Hazuka, C., Davanger, S., Kenny, J., Kee, Y., and Scheller, R. (1996). The mammalian brain *rsec6/8* complex. Neuron 17, 1209-1219.

Ihle, J., Morishita, K., Matsugi, T., and Bartholomew, C. (1990). Insertional mutagenesis and transformation of hematopoietic stem cells. Prog Clin Biol Res 352, 329-337.

Jenkins, N., Copeland, N., Taylor, B., and Lee, B. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. J Virol 43, 26-36.

Johnson, D., and Hunt, D. (1971). Hop-sterile, a mutant gene affecting sperm tail development in the mouse. J Embryol Exp Morphol 25, 223-236.

Jonkers, J., and Berns, A. (1996). Retroviral insertional mutagenesis as a strategy to identify cancer genes. Biochim Biophys Acta 1287, 29-57.

Jonkers, J., Korswagen, H., Acton, D., Breuer, M., and Berns, A. (1997). Activation of a novel proto-oncogene, *Fratl*, contributes to progression of mouse T-cell lymphomas. EMBO J 16, 441-450.

Josefowicz, W., and Hardy, M. (1978). The expression of the gene asebia in the laboratory mouse. I. Epidermis and dermis. Genet Res 31, 53-65.

Justice, M., Silan, C., Ceci, J., Buchberg, A., Copeland, N., and Jenkins, N. (1990). A molecular genetic linkage map of mouse chromosome 13 anchored by the beige (*bg*) and satin (*sa*) loci. Genomics 6, 341-351.

Justice, M., Jenkins, N., and Copeland, N. (1992). Recombinant inbred mouse strains: models for disease study. Trends Biotechnol 10, 120-126.

Justice, M., Morse, H. r., Jenkins, N., and Copeland, N. (1994). Identification of *Evi-3*, a novel common site of retroviral integration in mouse AKXD B-cell lymphomas. J Virol 68, 1293-1300.

Kamps, M., Murre, C., Sun, X., and Baltimore, D. (1990). A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. Cell 60, 547-555.

Kee, Y., Yoo, J., Hazuka, C., Peterson, K., Hsu, S., and Scheller, R. (1997). Subunit structure of the mammalian exocyst complex. Proc Natl Acad Sci 94, 14438-14443.

Keller, S., Jones, J., Boyle, A., Barrow, L., Killen, P., Green, D., Kapousta, N., Hitchcock, P., Swank, R., and Meisler, M. (1994). Kidney and retinal defects (*Krd*), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the *Pax2* locus. Genomics 23, 309-320.

Kennedy, M., Gonzalez-Sarmiento, R., Kees, U., Lampert, F., Dear, N., Boehm, T., and Rabbitts, T. (1991). *HOX11*, a homeobox-containing T-cell oncogene on human chromosome 10q24. Proc Natl Acad Sci 88, 8900-8904.

Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M., and Schlessinger, D. (1992). Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. Genomics 14, 241-248.

Lako, M., Lindsay, S., Bullen, P., Wilson, D., Robson, S., and Strachan, T. (1998). A novel mammalian wnt gene, *WNT8B*, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. Hum Mol Genet 7, 813-822.

Lazo, P., Lee, J., and Tsichlis, P. (1990). Long-distance activation of the *Myc* protooncogene by provirus insertion in *Mlvi-1* or *Mlvi-4* in rat T-cell lymphomas. Proc Natl Acad Sci 87, 170-173.

Levy, E., Parganas, E., Morishita, K., Fichelson, S., James, L., Oscier, D., Gisselbrecht, S., Ihle, J., and Buckle, V. (1994). DNA rearrangements proximal to the *EVI1* locus associated with the 3q21q26 syndrome. Blood 83, 1348-1354.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S., Giovanella, B., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M., and Parsons, R. (1997). *PTEN*, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275, 1943-7.

Liao, X., Buchberg, A., Jenkins, N., and Copeland, N. (1995). *Evi-5*, a common site of retroviral integration in AKXD T-cell lymphomas, maps near *Gfi-1* on mouse chromosome 5. J Virol 69, 7132-7137.

Liao, X., Du, Y., Morse, H. r., Jenkins, N., and Copeland, N. (1997). Proviral integrations at the *Evi5* locus disrupt a novel 90 kDa protein with homology to the *Tre2* oncogene and cell-cycle regulatory proteins. Oncogene 14, 1023-1029.

Liaw, D., Marsh, D., Li, J., Dahia, P., Wang, S., Zheng, Z., Bose, S., Call, K., Tsou, H., Peacocke, M., Eng, C., and Parsons, R. (1997). Germline mutations of the *PTEN* gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 16, 64-67.

Liu, C., Zhu, G., Converse, R., Kao, C., Nakamura, H., Tseng, S., Mui, M., Seyer, J., Justice, M., Stech, M., Hansen, G., and Kao, W. (1994). Characterization and chromosomal localization of the cornea-specific murine keratin gene *Krt1.12.* J Biol Chem 269, 24627-24636.

Lloyd, K., and Dennis, M. (1963). Cowden's disease. A possible new symptom complex with multiple system involvement. Ann Intern Med 58, 136-142.

Manfioletti, G., Gattei, V., Buratti, E., Rustighi, A., De, I. A., Aldinucci, D., Goodwin, G., and Pinto, A. (1995). Differential expression of a novel proline-rich homeobox gene (*PRH*) in human hematolymphopoietic cells. Blood 85, 1237-45.

Milot, E., Belmaaza, A., Rassart, E., and Chartrand, P. (1994). Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology 201, 408-412.

Mock, B., Connelly, M., McBride, O., Kozak, C., and Marcu, K. (1995). *CHUK*, a conserved helix-loop-helix ubiquitous kinase, maps to human chromosome10 and mouse chromosome 19. Genomics 27, 348-351.

Monnier, V., Dussillol, F., Alves, G., Lamour-Isnard, C., and Plessis, A. (1998). *Suppressor of fused* links *fused* and *cubitus interruptus* on the hedgehog signalling pathway. Curr Biol 8, 583-586.

Morishita, K., Parganas, E., William, C., Whittaker, M., Drabkin, H., Oval, J., Taetle, R., Valentine, M., and Ihle, J. (1992). Activation of *EVI1* gene expression in human acute myelogenous leukemias by translocations spanning 300-400 kilobases on chromosome band 3q26. Proc Natl Acad Sci 89, 3937-3941.

Moskow, J., Bullrich, F., Huebner, K., Daar, I., and Buchberg, A. (1995). *Meis1*, a *PBX1*-related homeobox gene involved in myeloid leukemia in BXH-2 mice. Mol Cell Biol 15, 5434-5443.

Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. World Wide Web (URL: http://www.informatics.jax.org/). (June, 1998).

Mucenski, M., Taylor, B., Jenkins, N., and Copeland, N. (1986). AKXD recombinant inbred strains: models for studying the molecular genetic basis of murine lymphomas. Mol Cell Biol 6, 4236-4243.

Mucenski, M., Taylor, B., Copeland, N., and Jenkins, N. (1987). Characterization of somatically acquired ecotropic and mink cell focus-forming viruses in lymphomas of AKXD recombinant inbred mice. J Virol 61, 2929-2933.

Mucenski, M., Taylor, B., Ihle, J., Hartley, J., Morse, H. d., Jenkins, N., and Copeland, N. (1988). Identification of a common ecotropic viral integration site, *Evi-1*, in the DNA of AKXD murine myeloid tumors. Mol Cell Biol 8, 301-308.

Mucenski, M., McLain, K., Kier, A., Swerdlow, S., Schreiner, C., Miller, T., Pietryga, D., Scott, W. J., and Potter, S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell 65, 677-689.

Mueller, P., and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246, 780-786.

Nakagawa, T., Tanaka, Y., Matsuoka, E., Kondo, S., Okada, Y., Noda, Y., Kanai, Y., and Hirokawa, N. (1997). Identification and classification of 16 new kinesin superfamily (KIF) proteins in mouse genome. Proc Natl Acad Sci 94, 9654-9659.

Nakamura, T., Largaespada, D., Shaughnessy, J. J., Jenkins, N., and Copeland, N. (1996a). Cooperative activation of *Hoxa* and *Pbx1*-related genes in murine myeloid leukaemias. Nat Genet 12, 149-153.

Nakamura, T., Largaespada, D., Lee, M., Johnson, L., Ohyashiki, K., Toyama, K., Chen, S., Willman, C., Chen, I., Feinberg, A., Jenkins, N., Copeland, N., and Shaughnessy, J. J. (1996b). Fusion of the nucleoporin gene *NUP98* to *HOXA9* by the

chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. Nat Genet 12, 154-158.

Nakamura, H., Yoshida, M., Tsuiki, H., Ito, K., Ueno, M., Nakao, M., Oka, K., Tada, M., Kochi, M., Kuratsu, J., Ushio, Y., and Saya, H. (1998). Identification of a human homolog of the Drosophila *neuralized* gene within the 10q25.1 malignant astrocytoma deletion region. Oncogene 16, 1009-1019.

Neiman, P. (1994). Retrovirus-induced B cell neoplasia in the bursa of Fabricius. Adv Immunol 56, 467-487.

Neri, A., Fracchiolla, N., Roscetti, E., Garatti, S., Trecca, D., Boletini, A., Perletti, L., Baldini, L., Maiolo, A., and Berti, E. (1995). Molecular analysis of cutaneous B- and T-cell lymphomas. Blood 86, 3160-3172.

Nourse, J., Mellentin, J., Galili, N., Wilkinson, J., Standbridge, E., Smith, S., and Cleary, M. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. Cell 60, 535-545.

Novak, E., Hui, S., and Swank, R. (1984). Platelet storage pool deficiency in mouse pigment mutations associated with seven distinct genetic loci. Blood 63, 536-544.

Nunes, M., Schutt, G., Kapur, R., Luthardt, F., Kukolich, M., Byers, P., and Evans, J. (1995). A second autosomal split hand/split foot locus maps to chromosome10q24-q25. Hum Mol Genet 24, 2165-2170.

Nusse, R., van Ooyen, A., Cox, D., Fung, Y., and Varmus, H. (1984). Mode of proviral activation of a putative mammary oncogene (*int-1*) on mouse chromosome 15. Nature 307, 131-136.

Nusse, R. (1986). The activation of cellular oncogenes by retroviral insertion. Trends in Genet 2, 244-247.

Oh, J., Bailin, T., Fukai, K., Feng, G., Ho, L., Mao, J., Frenk, E., Tamura, N., and Spritz, R. (1996). Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. Nature Genetics 14, 300-306.

Online Mendelian Inheritance in Man, OMIM (TM). Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 1997. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

Ottman, R., Risch, N., Hauser, W., Pedley, T., Lee, J., Barker-Cummings, C., Lustenberger, A., Nagle, K., Lee, K., Scheuer, M., and al, e. (1995). Localization of a gene for partial epilepsy to chromosome 10q. Nat Genet 10.

Parsons, R. (1998). Phosphatases and tumorigenesis. Curr Opin Oncol 10, 88-91.

Pattengale, P. (1994). Tumours of the lymphohaematopoietic system. IARC Sci Publ 111, 651-670.

Rabbitts, T., and Boehm, T. (1991). Structural and functional chimerism results from chromosomal translocation in lymphoid tumors. Adv Immunol 50, 119-146.

Ramel, C. (1989). The nature of spontaneous mutations. Mutat Res 212, 33-42.

Robertson, H., Preston, C., Phillis, R., Johnson-Schlitz, D., Benz, W., and Engels, W. (1988). A stable genomic source of P element transposase in Drosophila melanogaster. Genetics 118, 461-470.

Rusiniak, M., O'Brien, E., Novak, E., Barone, S., McGarry, M., Reddington, M., and Swank, R. (1996). Molecular markers near the mouse brachymorphic (*bm*) gene, which affects connective tissues and bleeding time. Mamm Genome 7, 98-102.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning: a laboratory manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sarkar, G., Turner, R., and Bolander, M. (1993). Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. PCR Methods Appl 2, 318-322.

Sawin, K., LeGuellec, K., Philippe, M., and Mitchison, T. (1992). Mitotic spindle organization by a plus-end-directed microtubule motor. Nature 359, 540-543.

Semina, E., Reiter, R., and Murray, J. (1997). Isolation of a new homeobox gene belonging to the *Pitx/Rieg* family: expression during lens development and mapping to the aphakia region on mouse chromosome 19. Hum Mol Genet 6, 2109-2116.

Semina, E., Ferrell, R., Mintz-Hittner, H., Bitoun, P., Alward, W., Reiter, R., Funkhauser, C., Daack-Hirsch, S., and Murray, J. (1998). A novel homeobox gene *PITX3* is mutated in families with autosomal-dominant cataracts and ASMD. Nat Genet 19, 167-170.

Skarnes, W., Moss, J., Hurtley, S., and Beddington, R. (1995). Capturing genes encoding membrane and secreted proteins important for mouse development. Proc Natl Acad Sci 92, 6592-6596.

Smith, A., and Ashworth, A. (1998). Cancer predisposition: where's the phosphate? Cur Biol 8, R241-R243.

Sobels, F. (1989). Models and assumptions underlying genetic risk assessment. Mutat Res 212, 77-89.

Spradling, A., Stern, D., Kiss, I., Roote, J., Laverty, T., and Rubin, G. (1995). Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci 92, 10824-10830.

Steck, P., Pershouse, M., Jasser, S., Yung, W., Lin, H., Ligon, A., Langford, L., Baumgard, M., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D., and Tavtigian, S. (1997). Identification of a candidate tumour suppressor gene, *MMAC1*, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 15, 356-62.

Stewart, M., Terry, A., O'Hara, M., Cameron, E., Onions, D., and Neil, J. (1996). *til-1*: a novel proviral insertion locus for Moloney murine leukaemia virus in lymphomas of *CD2-myc* transgenic mice. J Gen Virol 77, 443-446.

Stoye, J., Fenner, S., Greenoak, G., Moran, C., and Coffin, J. (1988). Role of endogenous retroviruses as mutagens: the hairless mutation of mice. Cell 54, 383-391.

Suzukawa, K., Parganas, E., Gajjar, A., Abe, T., Takahashi, S., Tani, K., Asano, S., Asou, H., Kamada, N., and Yokota, J. (1994). Identification of a breakpoint cluster region 3' of the ribophorin I gene at 3q21 associated with the transcriptional activation of the *EVI1* gene in acute myelogenous leukemias with inv(3)(q21q26). Blood 84, 2681-2688.

Tang, J., Ho, F., Chan, A., and Srivastava, G. (1998). Clonality of lymphomas at multiple sites in SJL mice. Lab Invest 78, 205-212.

Thomas, P., Brown, A., and Beddington, R. (1998). *Hex*: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. Development 125, 85-94.

Uberbacher, E., and Mural, R. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci 88, 11261-11265.

van Lohuizen, M., and Berns, A. (1990). Tumorigenesis by slow-transforming retroviruses--an update. Biochim Biophys Acta 1032, 213-235.

van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. (1991). Identification of cooperating oncogenes in *E mu-myc* transgenic mice by provirus tagging. Cell 65, 737-752.

Varmus, H. (1982). Form and function of retroviral proviruses. Science 216, 812-820.

Varnum, D., and Stevens, L. (1968). Aphakia, a new mutation in the mouse. J Hered 59, 147-150.

Vogelstein, B., and Kinzler, K. (1993). The multistep nature of cancer. Trends Genet 9, 138-141.

Wasylyk, B., Wasylyk, C., Augereau, P., and Chambon, P. (1983). The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-514.

Wilson, C., Pearson, R., Bellen, H., O'Kane, C., Grossniklaus, U., and Gehring, W. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. Genes Dev 3, 1301-1313.

Wu, J., Zhou, T., He, J., and Mountz, J. (1993). Autoimmune disease in mice due to integration of an endogenous retrovirus in an apoptosis gene. J Exper Med 178, 461-468.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D., Guillemot, F., Gasca, S., Cado, D., and Auerbach, A. (1995). A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. Genetics 139, 889-899.

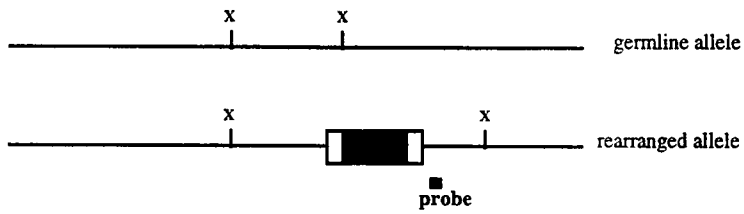
Zambrowicz, B., Friedrich, G., Buxton, E., Lilleberg, S., Person, C., and Sands, A. (1998). Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392, 608-611.

Zijlstra, M., Quint, W., Cuypers, T., Radaszkiewicz, T., Schoenmakers, H., de Goede, R., and Melief, C. (1986). Ecotropic and mink cell focus-forming murine leukemia viruses integrate in mouse T, B, and non-T/non-B cell lymphoma DNA. J Virol 57, 1037-1047.

Zwaan, J., and Kirkland, B. (1975). Malorientation of mitotic figures in the early lens rudiment of aphakia mouse embryos. Anat Rec 182, 345-354.

APPENDIX

A.



B.

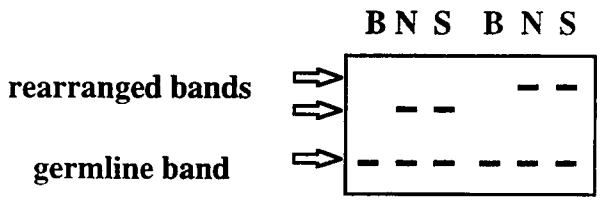


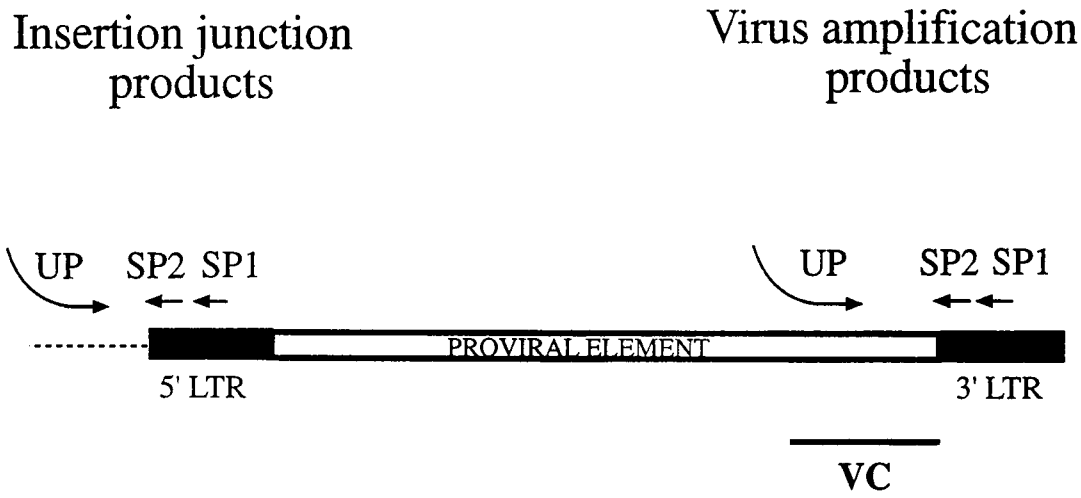
Fig. I - 1. Identifying viral integration sites in the mouse. A. Schematic representation of a germline and rearranged allele. The horizontal lines depict genomic DNA. An X is used to indicate the location of theoretical *XbaI* restriction fragment sites. The striped box represents a proviral element. A probe flanking the site of the inserted provirus is indicated below the line representing the rearranged allele. B. Representative diagram showing possible Southern blot analysis using the probe shown in part A. Symbols are used to show the tissues analyzed: brain (B), lymph node tumor (N), and spleen tumor (S). A probe representing a viral viral integration site would be expected to detect a normal allele (germline band) and a tumor-specific rearranged allele (rearranged bands). Brain is used as a control because it is rarely infiltrated by hematopoietic tumor tissue.

Table I - 1. Viral integration sites in AKXD and BXH-2 mouse strains.

Insertion Site	Tumor Subset	Possible Gene	Tumor Type	Mechanism / site of insertion	Gene Type
Evi1	AKXD	*Evi1*	Myeloid	gene upregulation	transcription factor
Evi2	BXH-2	*Nf1*	Myeloid	gene inactivation/ intron insertion	tumor suppressor
Evi3	AKXD	*Evi3*	Stem cell, Pre-B cell	unknown	unknown
Evi5	AKXD	*Evi5*	T cell	unknown/ 3' coding region	tcb box, cell cycle gene
Evi6	BXH-2	*Hoxa9*	Myeloid	gene upregulation/ upstream of 5' UTR	homeobox transcription factor
Evi7	BXH-2	*Hoxa7*	Myeloid	gene upregulation/ upstream of 5' UTR	homeobox transcription factor
Evi8	BXH-2	*Meis1*	Myeloid	gene upregulation/ downstream of 3' UTR	PBX1-transcription factor
Evi9	BXH-2	unknown	Myeloid	unknown	unknown
Evi10	BXH-2	*Meis1*	Myeloid	gene upregulation/ upstream of 5' UTR	PBX1-transcription factor
Meis1	BXH-2	*Hoxa7*	Myeloid	gene upregulation/ 5' and 3' of coding region	PBX1-transcription factor
Lvis1	AKXD	*Hex*? *msec15*? *mEg5*?	B cell	gene upregulation/ LTR enhancer effects?	homeobox transcription factor exocyst complex protein kinesin-related protein

Fig. II - 1. Schematic representation of universal PCR using MuLV LTR specific primers. A. Schematic representation of a proviral element showing the approximate binding sites for LTR specific primers SP1 and SP2. Amplification from either the 5' or 3' LTR element is directed by SP1 and SP2, producing amplification products representing both mouse genomic DNA flanking the integrated virus (from the 5' LTR element) and virus genomic DNA (from the 3' LTR). The universal primer is depicted by a curved arrow. The curved portion of the arrow represents the M13F linker sequence, which is not expected to bind to mouse DNA. Dashed lines denote mouse genomic DNA. The control probe used to detect amplification of internal virus-specific products is shown as a line below the proviral element, and is labeled VC. B. Schematic map of the *ecotropic* provirus, showing *Xba*I and *Pst*I restriction sites. Two rounds of PCR amplification using the UP2 universal primer should generate a 523 bp *ecotropic* virus product, shown at the bottom of the figure. The UP1 primer, which binds specifically at *Pst*I sites, will not amplify *ecotropic* virus genomic products.

A.


Insertion junction products
Virus amplification products
UP
SP2 SP1
UP
SP2 SP1
PROVIRAL ELEMENT
5' LTR
3' LTR
VC

B.

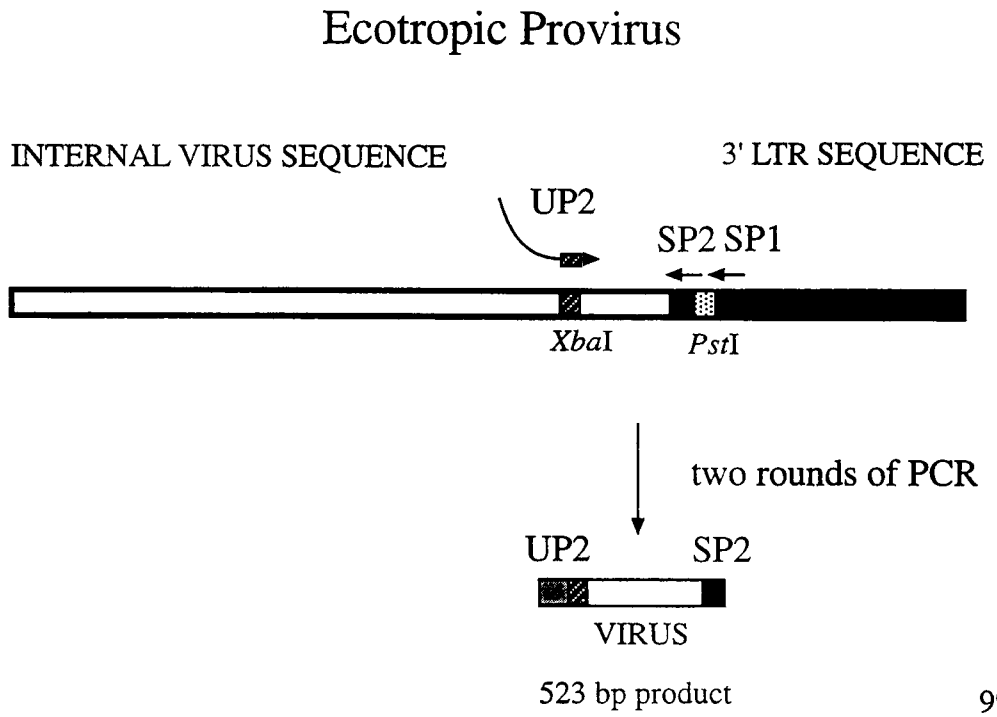
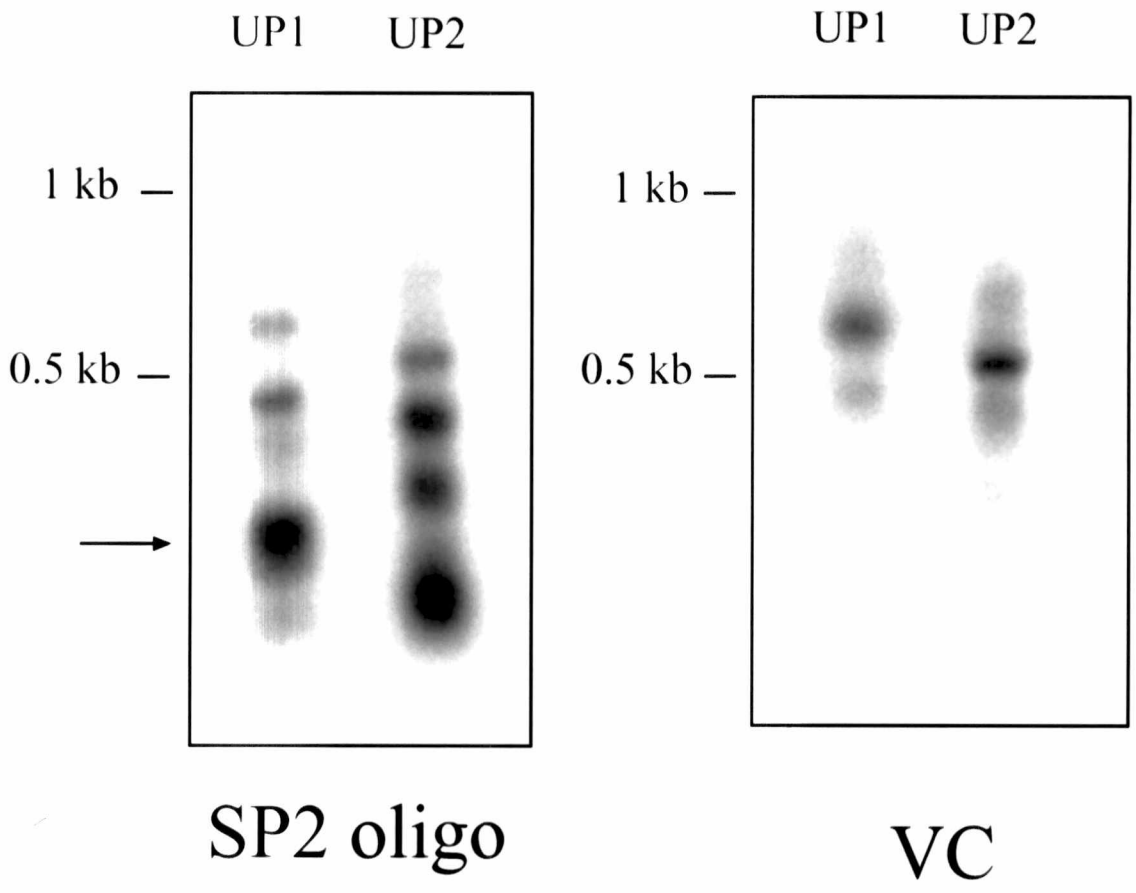

Ecotropic Provirus
INTERNAL VIRUS SEQUENCE
3' LTR SEQUENCE
UP2
SP2 SP1
XbaI
PstI
two rounds of PCR
UP2
SP2
VIRUS
523 bp product

Fig. II - 2. Southern blot analysis of universal PCR products. The SP2 probe was used to detect all virus primed products, while the VC probe was used to identify amplification of internal virus sequences. Hybridization of universal PCR products generated using either the UP1 or UP2 primer revealed amplification of multiple products that did not hybridize to the VC probe. These fragments likely represent viral insertion junctions, and should contain mouse genomic DNA flanking an inserted virus. The position of size markers are shown to the left of each panel. An arrow is used to identify the 186 bp amplification product that was cloned and sequenced.


UP1
UP2
1 kb
0.5 kb
SP2 oligo
UP1
UP2
1 kb
0.5 kb
VC

1 GGG TTTTC CCAGT CACGA CNGGG CAAG**C TGCAG** CCTCG GTCCC ATCCG TCTGA

AAG TTAAG AACAC CACCT TGCCA GTTC**C TGCAG** CCTCG GTCCC ATCCG TCTGA

54 ATTCT TCCTA CATCT CGGTG AAGTT TATGG GGCCA AGAAG TTACC TGAGA CTCAG

ATTCT TGCTA CATCT CGGTG AAGTT TATGG GGCCA AGAAG TTACC TGAGA CTCAG

↓

109 GGCTT TGATA TCATT AGGAG AGAAA TCTTG AGCAA CTTTG CCTCA AGTTC TGAAG

GGCTT TGATA TCATT AGGAG AGAAA TCTTG AGCAA CTTTG CCTCA AGTTC AAATC

164 ACCCC AACTA TTCCG AATCG GTC

AAGGC ATACA TGAAA TCACC ACA

Figure II - 3. Comparison of amplified viral junction fragment sequence with sequence of the germline allele. The top line of each pair of sequences represents sequence from clone pL1, including the universal primer sequence (UP1, double underline) and viral specific primer sequence (SP2, single underline). Nucleotide numbers are listed on the left. The bottom line represents the unrearranged locus. The *Pst*I site is shown in bold, and an arrow is used to mark the site of viral insertion.

Fig. II - 4. Southern blot analysis of brain and tumor DNA. Southern hybridization of DNA from brain (non-tumor) and tumor tissue, digested with either *Xba*I (panel A), or *Eco*RI (panel B). The *eco* probe, which is specific to the envelop gene region of AKV ecotropic MuLV (Chattopadhyay *et al.*, 1980), detects two endogenous ecotropic proviral elements with *Xba*I, and three with *Eco*RI. In addition, the *eco* probe detects a single tumor-specific restriction fragment (arrow) using either enzyme. The pL1 probe identifies a germline *Xba*I restriction fragment of 6.5 kb, as well as a tumor-specific restriction fragment of ~15 kb (arrow), which is consistent in size with the *Xba*I fragment detected with the *eco* probe. The pL1 probe also identifies a tumor-specific restriction fragment of 9.3 kb (arrow) using *Eco*RI digested DNA, consistent in size with the tumor-specific *Eco*RI fragment detected with the *eco* probe. The germline *Eco*RI restriction fragment is < 0.5 kb, and was not detected by hybridization with the pL1 probe.

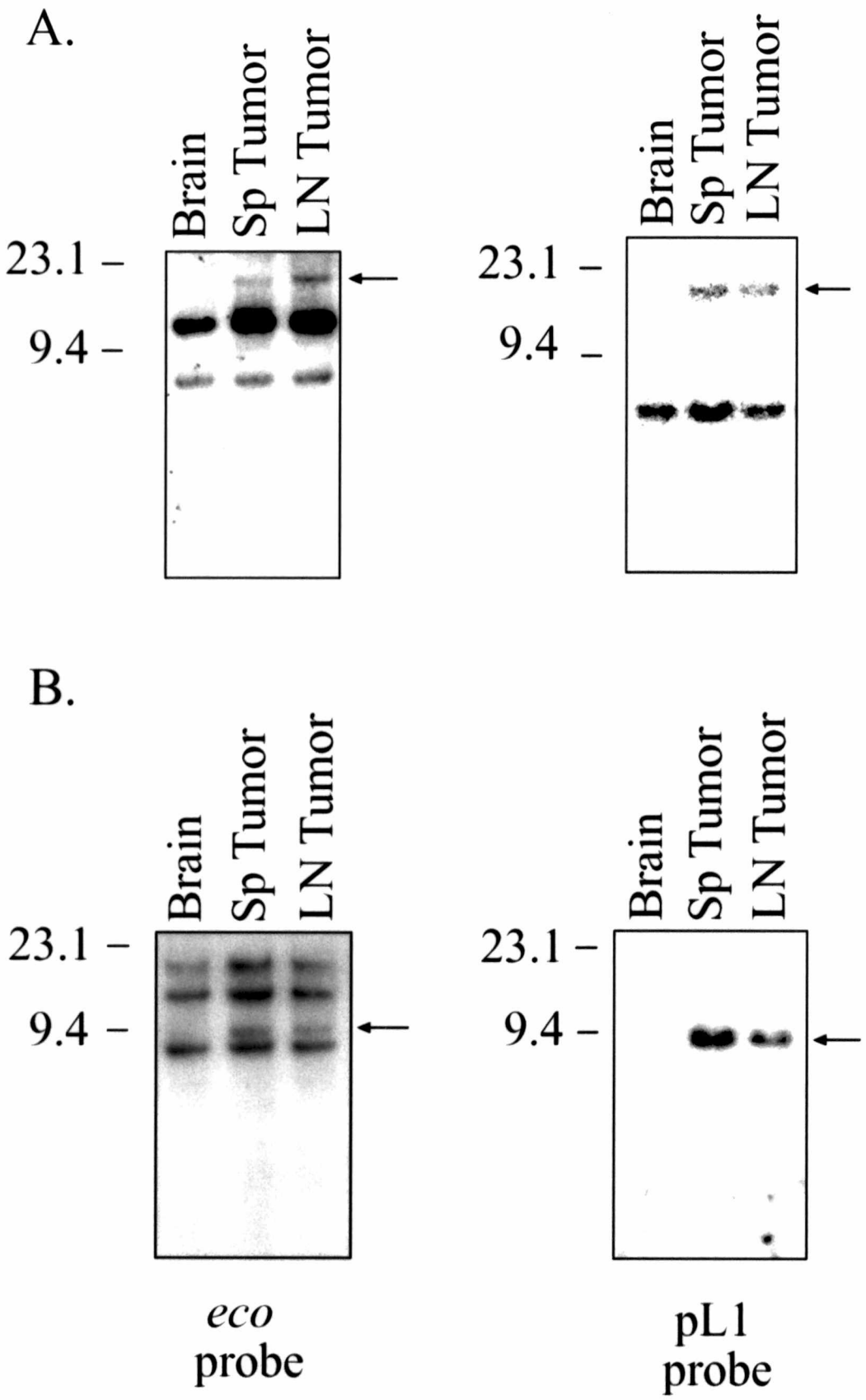
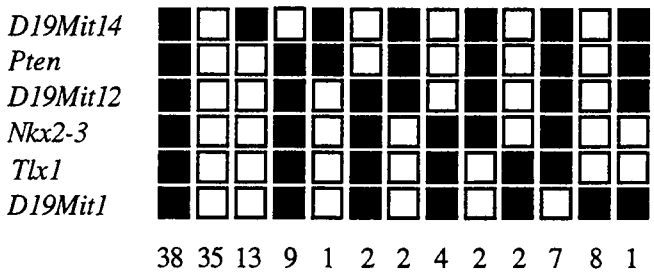

A.
Brain
Sp Tumor
LN Tumor
23.1 –
9.4 –
Brain
Sp Tumor
LN Tumor
23.1 –
9.4 –
B.
Brain
Sp Tumor
LN Tumor
23.1 –
9.4 –
Brain
Sp Tumor
LN Tumor
23.1 –
9.4 –
eco
probe
pL1
probe

Fig. III A - 1. Segregation of alleles in 124 (SB/Le x *Mus spretus*)F_1 x SB/Le interspecific backcross progeny. (A) Each column represents the chromosome identified in the backcross progeny that was inherited from the (SB/Le x *M.spretus*)F_1 parent. The white boxes represent the presence of an *M. spretus* allele, whereas black boxes represent the presence of an SB/Le allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. (B) Chr 19 linkage map showing the location of markers mapped in this study. Distances between loci are given in centimorgans and are listed on the left. Loci that have been mapped in the human are underlined, and their human chromosomal location is listed on the right.

A


D19Mit14
Pten
D19Mit12
Nkx2-3
Tlx1
D19Mit1
38 35 13 9 1 2 2 4 2 2 7 8 1

B

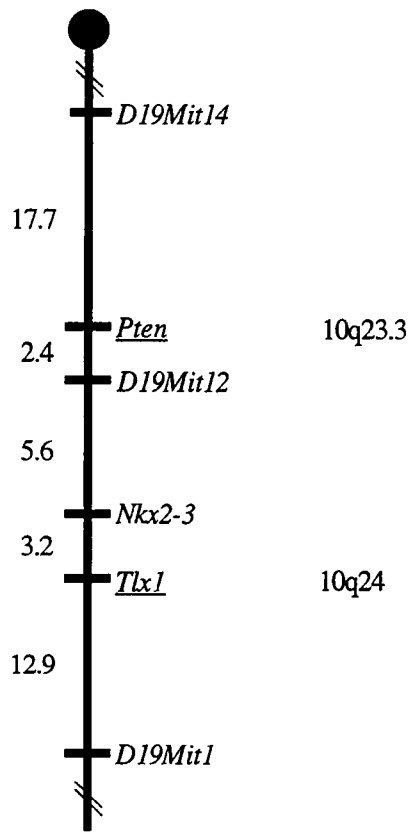

D19Mit14
17.7
Pten
10q23.3
2.4
D19Mit12
5.6
Nkx2-3
3.2
Tlx1
10q24
12.9
D19Mit1

TABLE III - 1. Probes/Amplimers used for mapping studies

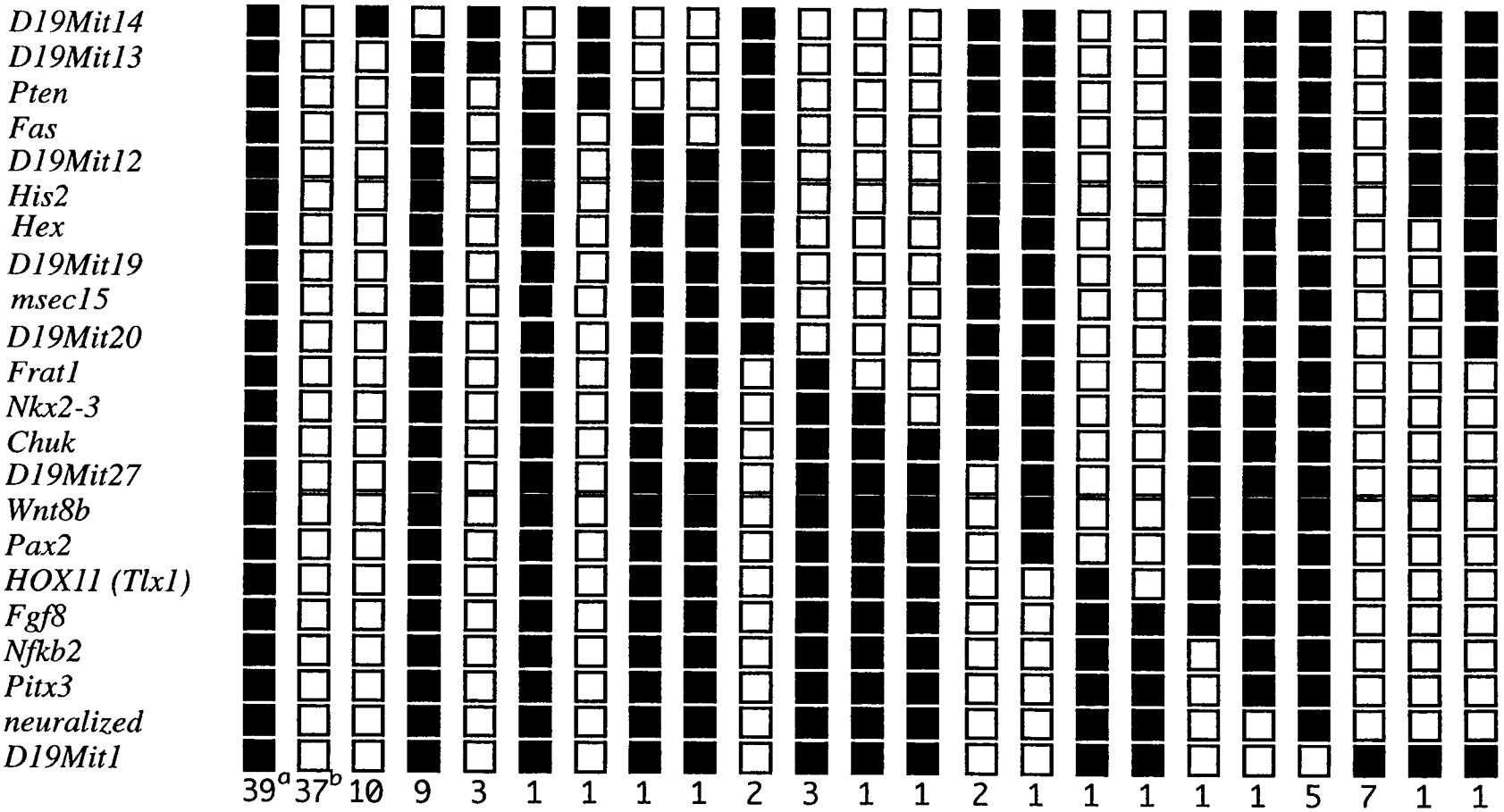
Gene or Segment Name	Locus	Probe or Primers	Enzyme	Size (kb unless noted)	
				SB/Le	*M. spretus* [a]
Conserved helix-loop-helix ubiquitous kinase genbank accession No. U12473	*Chuk*	gaacgtcagtctgtaccagc cagcatcaattggagccagc Template: mouse liver cDNA Probe size: 760bps	*Bgl* II	1.5, 2.6, 3.9	1.2[b], 2.6, 5
Ecotropic viral integration site 1	*Evi1*	SP	*Pst* I	0.9	0.7
Fas antigen genbank accession No. E05335	*Fas*	aggtactaatagcatctccg ctaaggttctgcgacattcg Template: mouse liver cDNA Probe size: 812bps	*Pvu* II	1, 6, 9	1.5, 2.7, 9
Fibroblast growth factor 8	*Fgf8*	Fgf8 cDNA, variant 4	*EcoR* I	15	9, 6
Frequently rearranged in advanced T-cell lymphoma genbank accession No. U58974	*Frat1*	gctacagttacacgcaaagc ctcaacactctgctaacagc Template: mouse liver cDNA Probe size: 847 bps	*EcoR* I	5.8	3, 1.8
Hematopoietic insertion site 2	*His2*	pD1.1/14	*Pvu* II	7.3	9
Hematopoietically expressed homeobox genbank accession No. AA985962	*Hex*	tatacatgttatagacacag ggcattcttaacattgctgc Template: B6 genomic DNA Probe size: 279 bps	*Pst* I	2.5, 5.4	3[c], 7.5
mouse homolog of yeast *sec15*	*msec15*	mouse EST clone dbEST ID 606262	*BamH* I	23, 10, 9, 5.6, 1.8	14, 6.1, 3.6
mouse homolog of *Drosophila neuralized* (D-neu) protein	*Nlz*	human EST clone dbEST ID 352633	*Bgl* II	16	6
mouse homolog of *Drosophila Notch*, 2 genbank accession No. D32210	*Notch2*	acccgaatggctatgaatgc agtacagtttccgtggatgc Template: 12.5d embryo cDNA Probe size: 567 bps	*Pvu* II	4.6, 2.0	5.0, 1.2
Nuclear factor kappa-B, subunit 2	*Nfkb2*	mouse EST clone dbEST ID 352633	*BamH* I	6.8	3.5
Drosophila NK2 transcription factor related locus 3	*Nkx2-3*	pgMlx29-Pl	*Pvu* II	5.4	4.4
Paired box gene 2	*Pax2*	human EST clone dbEST ID: 1231511	*BamH* I	12.5, 4.7, 2.3, 1.9	12.5, 4.7, 2.6, 1.9
Paired-like homeobox transcription factor 3 genbank accession No. AF005772	*PitX3*	gtgtggttcaagaaccggc ttgaccgagttgaaggcgaa Template: mouse B6 DNA Probe size: 211 bps	*EcoR* I	1.2	0.7
Phosphatase and tensin homolog (mutated in multiple advanced cancers 1) genbank accession No. U92437	*Pten* (*Mmac1*)	atatcaagaggatggattcg ccttgtcattatctgcacgc Template: mouse liver cDNA Probe size: 896bps	*Pvu* II	1.8, 3.5, 7.5	1.8, 3.5, 4.2
T cell leukemia homeobox	*Tlx1*	HOX11ex3	*Taq* I	0.6, 1.8	0.6, 1
Suppressor of *fused*	*Su(fu)*	mouse EST clone dbEST ID: 687394	*EcoR* I	16, 8.5, 3.4 2.8, 2.0, 1.1	23, 8.5, 5.5 2.8, 2.0, 1.1
Wingless related MMTV integration site 8b	*Wnt8b*	mouse EST clone dbEST ID: 811439	*EcoR* I	12.5, 5.4	15
DNA segment, Ch.19, Mit-1	*D19Mit1*	-	-	140 bps	155 bps
DNA segment, Ch.19, Mit-12	*D19Mit12*	-	-	150 bps	120 bps
DNA segment, Ch.19, Mit-13	*D19Mit13*	-	-	245 bps	285 bps
DNA segment, Ch.19, Mit-14	*D19Mit14*	-	-	140 bps	115 bps
DNA segment, Ch.19, Mit-19	*D19Mit19*	-	-	115 bps	140 bps
DNA segment, Ch.19, Mit-20	*D19Mit20*	-	-	230 bps	180 bps
DNA segment, Ch.19, Mit-27	*D19Mit27*	-	-	170 bps	130 bps

[a] The underlined restriction fragments identify the segregating *M. spretus* alleles monitored in the N2 progeny.

[b] This RFLP cosegregated with markers on chromosome 16.

[c] This RFLP cosegregated with markers on chromosome 3.

Fig III B - 1. Segregation of alleles in 130 (SB/Le x *Mus spretus*)F_1 x SB/Le interspecific backcross progeny. Loci analyzed are listed on the left. Each column represents the chromosome identified in the N_2 offspring inherited from the (SB/Le x *Mus spretus*)F_1 parent. Filled squares represent the SB/Le allele; open squares represent the *M. spretus* allele. The number of N_2 offspring inheriting each type of chromosome is given at the bottom of each column.



a The genotype of one of these animals was not determined for *Pax2* or *Tlx1*.
b The genotype of one of these animals was not determined for *Nkx2-3*.

Fig. III B - 2. Molecular genetic linkage maps of mouse chromosome 19. The left chromosome shows the loci mapped by interspecific backcross analysis in the current study. Recombination distances (in centiMorgans) are listed to the left. Loci mapped in humans are underlined; gene locations on human chromosomes are shown in the middle. The right chromosome shows a partial version of the consensus linkage map of mouse chromosome 19 (MGD, June 1998). The two maps were arbitrarily aligned at *Fratl* (dashed lines).

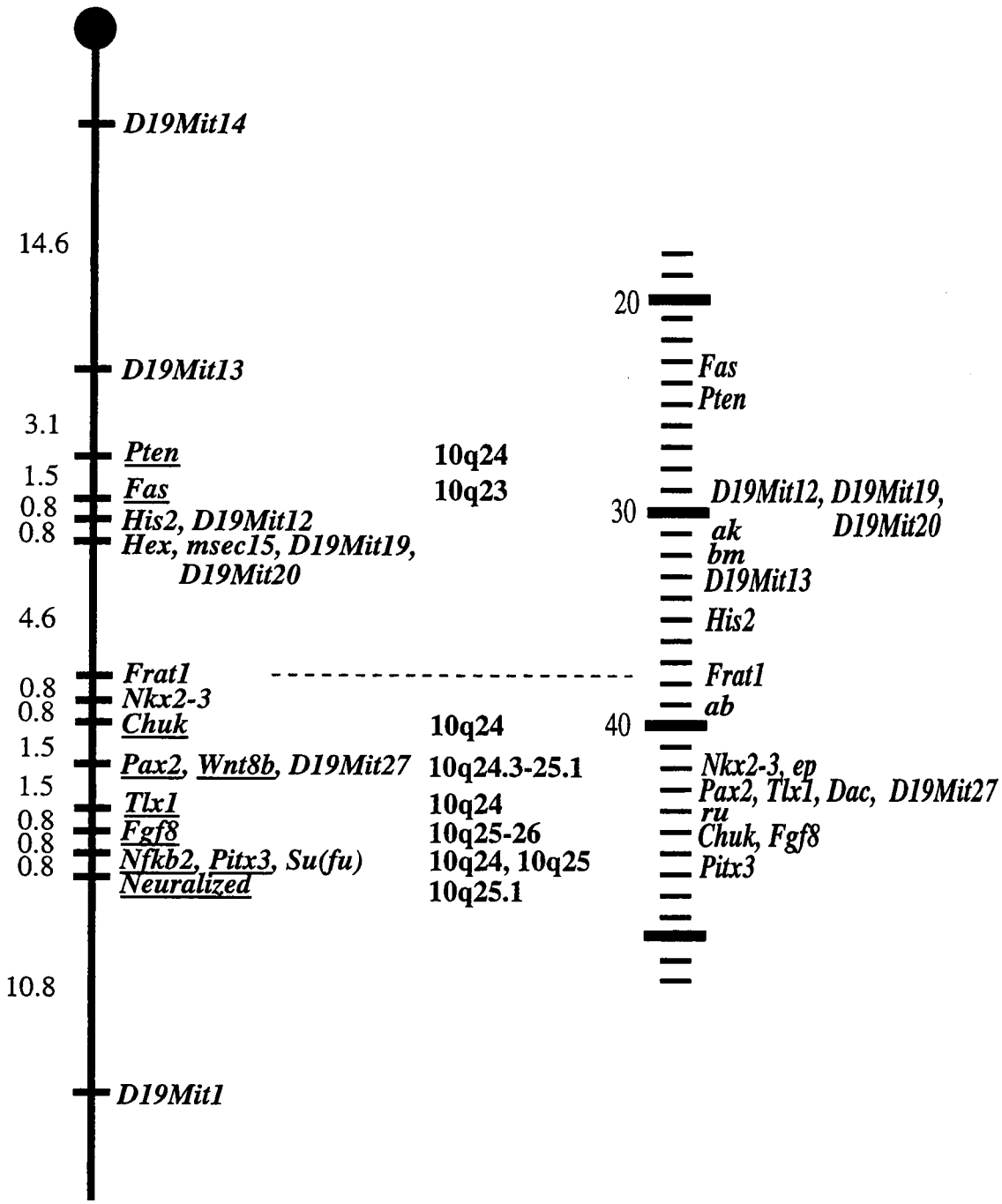
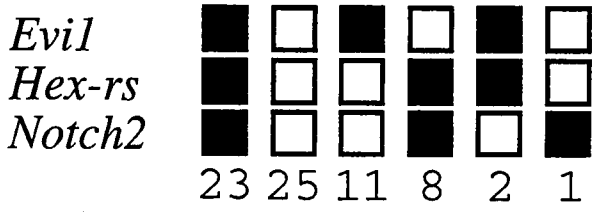
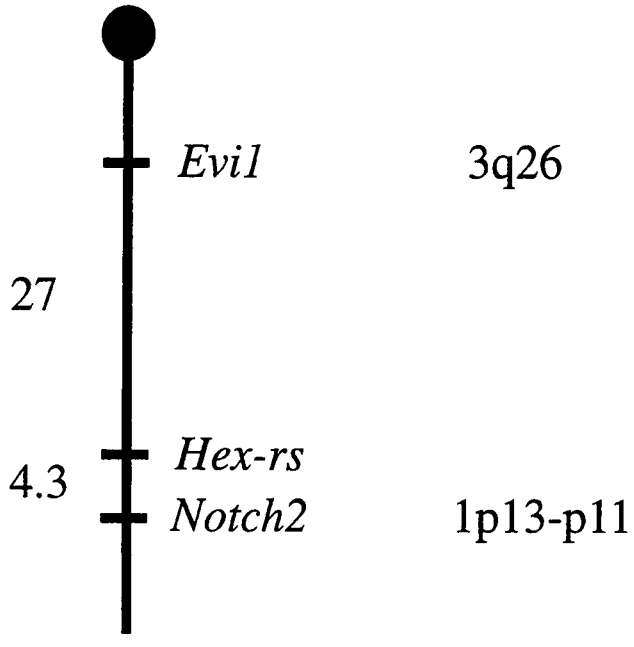

D19Mit14
14.6
D19Mit13
3.1
Pten
10q24
1.5
Fas
10q23
0.8
His2, D19Mit12
0.8
Hex, msec15, D19Mit19,
D19Mit20
4.6
Fratl
0.8
Nkx2-3
0.8
Chuk
10q24
1.5
Pax2, Wnt8b, D19Mit27
10q24.3-25.1
1.5
Tlx1
10q24
0.8
Fgf8
10q25-26
0.8
Nfkb2, Pitx3, Su(fu)
10q24, 10q25
0.8
Neuralized
10q25.1
10.8
D19Mit1
20
Fas
Pten
30
D19Mit12, D19Mit19,
D19Mit20
ak
bm
D19Mit13
His2
Fratl
ab
40
Nkx2-3, ep
Pax2, Tlx1, Dac, D19Mit27
ru
Chuk, Fgf8
Pitx3

Fig. III B - 3. Localization of *Hex-rs* to mouse chromosome 3. The upper figure shows the segregation of alleles in 70 (SB/Le x *Mus spretus*)F_1 x SB/Le interspecific backcross progeny. Each column represents the chromosome identified in the N_2 offspring inherited from the (SB/Le x *Mus spretus*)F_1 parent. Filled squares represent the SB/Le allele; open squares represent the *M. spretus* allele. The number of N_2 offspring inheriting each type of chromosome is given at the bottom of each column. The lower figure shows the chromosome 3 linkage map; recombination distances are given in centiMorgans to the left of the chromosome, and human map locations are shown to the right.


Evi1
Hex-rs
Notch2
23 25 11 8 2 1


Evi1
3q26
27
4.3
Hex-rs
Notch2
1p13-p11

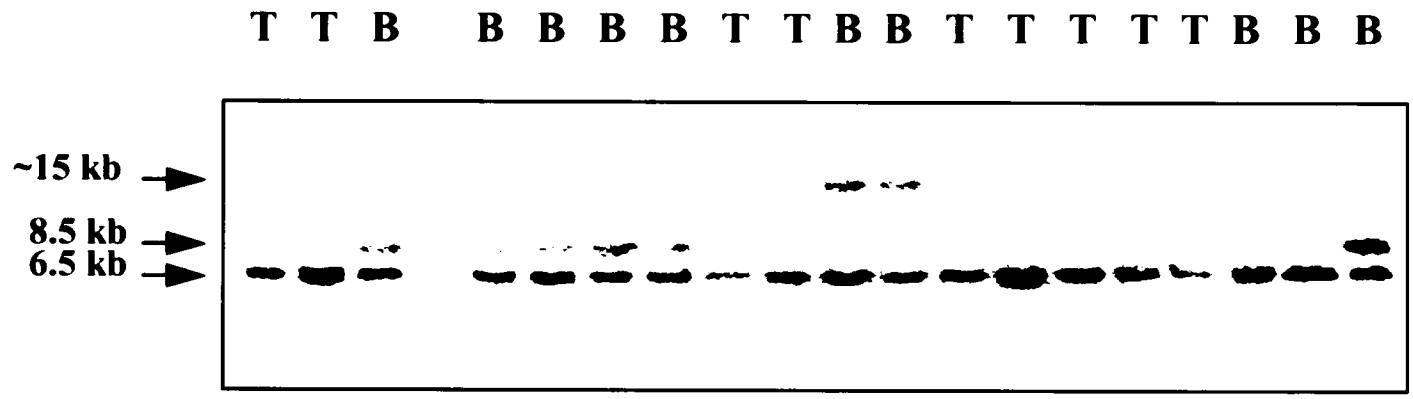


Fig. IV-1. Southern blot analysis of AKXD tumor DNA. A representative Southern blot of tumor DNA from AKXD animals hybridized with pL1 is shown. DNA was digested with *Xba*I and run on a 0.8% agarose gel. The cell type of each tumor is indicated above the lanes: T-cell (T), B-cell (B). The size of each restriction fragment detected with pL1 is listed to the left. A 6.5 kb fragment representing the germline allele is seen in each lane. Restriction fragments of 8.5 kb or 15 kb can be seen in most of the tumors classified as B-cell in origin.

Table IV - 1. AKXD tumor cell lineage distribution of *Lvis1* integrations

Tumor Type	Total Tumors Analyzed	*Lvis1* rearrangements
B cell	90	20
T cell	111	5
Mixed T and B	11	1
Pre B cell	29	0
Stem cell	6	0
Myeloid	9	0

Table IV - 2. Frequency of viral alterations in AKXD neoplasias

Commom site of viral integration	Incidence[a]	
	All tumors	B-cell tumor subset
Evi1	9.3%	6.7%
Evi2	1.5%	0
Evi3	4.1%	5.6%
Evi5	3.7%	1.1%
Gfi1	5.2%	2.2%
Lvis1	**9.6%**	**22.2%**
Mlvi1	1.9%	0
Mlvi2	0.7%	0
Pim1	7.0%	3.3%
Pim2	1.1%	0

[a] The percentage of leukemias/lymphomas with alterations at each locus is based on analysis of 270 AKXD tumor samples, of which 90 represent B-cell lymphomas.

Table IV - 3. Virally induced rearrangements at *Lvis1* in AKXD tumors

Recombinant Inbred Strain	Tumor Number	Tumor Type	Number of somatically acquired viruses (eco or MCF)	Other viral integrations
AKXD-2	381	B	4	*Evi1*
AKXD-7	195	T	8	*c-Myc*
AKXD-10	309	B	3	*Pim1*
AKXD-11	413	B	4	
AKXD-12	292	T	2	
AKXD-13	60	T	7	*Pim2*
AKXD-13	85	B	4	
AKXD-13	103	B	2	
AKXD-13	105	B	3	
AKXD-13	110	B	5	
AKXD-13	116	B	4	
AKXD-13	136	B	6	
AKXD-14	58	T/B	3	
AKXD-15	138	B	3	
AKXD-15	194	B	2	
AKXD-16	405	B	6	
AKXD-18	44	B	2	
AKXD-18	108	B	1	
AKXD-18	141	B	4	
AKXD-18	159	B	3	
AKXD-21	281	T	10	
AKXD-21	290	T	7	*Pvt1, Evi5*
AKXD-22	186	B	1	
AKXD-22	187	B	2	
AKXD-23	170	B	4	
AKXD-26	297	B	7	

—

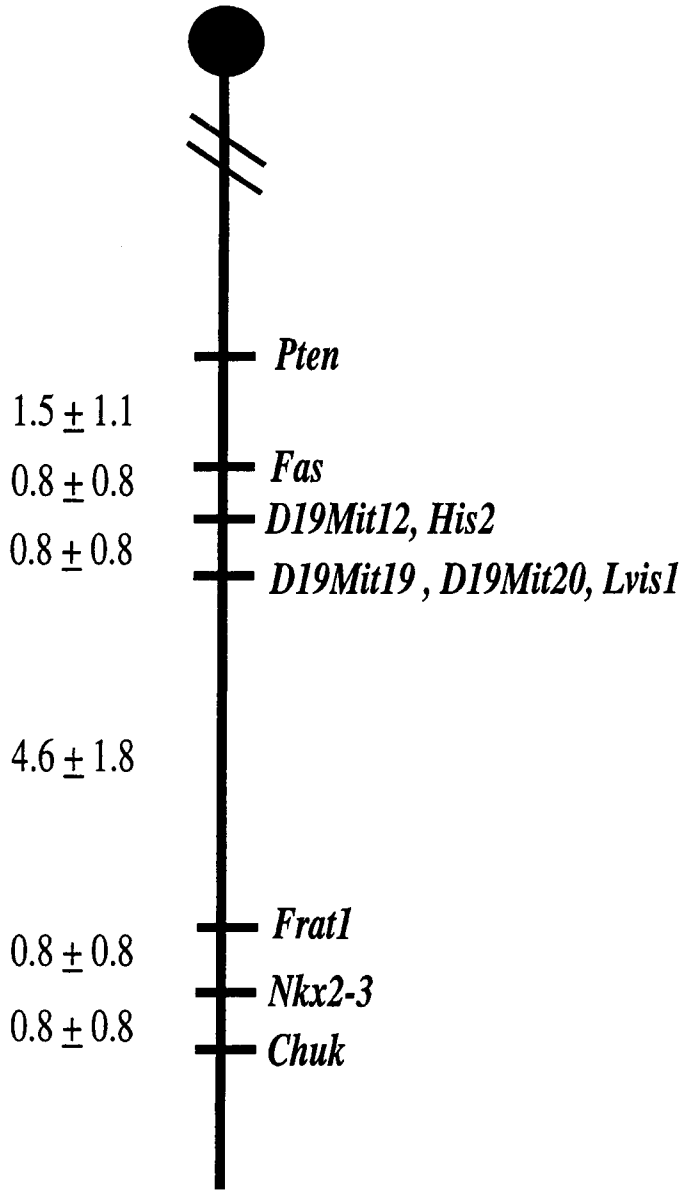


Fig. IV - 2. Chromosomal localization of *Lvis1* in the mouse. A partial molecular genetic linkage map of chromosome 19 depicts the map location of *Lvis1*. A total of 130 animals were analyzed for each marker, and the calculated recombination frequencies between loci expressed in centiMorgans ± the standard error are indicated to the left of the map.

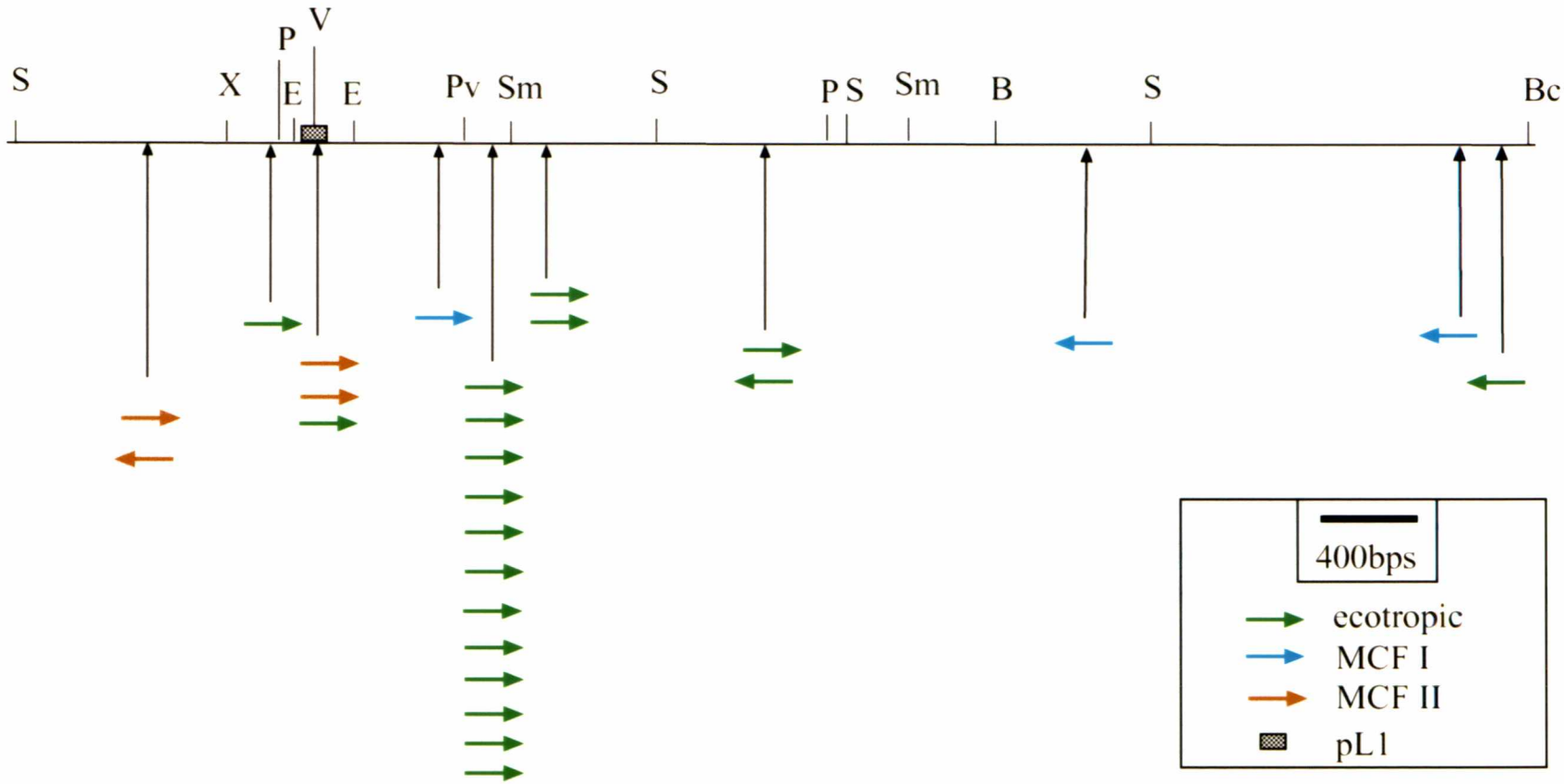


Fig. IV-3. Location and orientation of proviruses at *Lvis1* in AKXD tumors. Somatically acquired proviruses at *Lvis*1 were identified in tumor DNA from 26 AKXD animals. The location, transcriptional orientation, and type of each provirus was determined by restriction enzyme analysis and Southern blot hybridization of tumor and brain DNA. Colored arrows indicate the type of virus identified in each tumor, and the direction of these arrows designate the transcriptional orientation of each provirus. Enzyme designations include: *BamH*1 (B), *Bcl*I (Bc), *Sac*I (S), *Xba*I (X), *Pvu*II (Pv), *EcoR*I (E), *EcoR*V (V), *Pst*I (P), and *Sma*I (Sm).

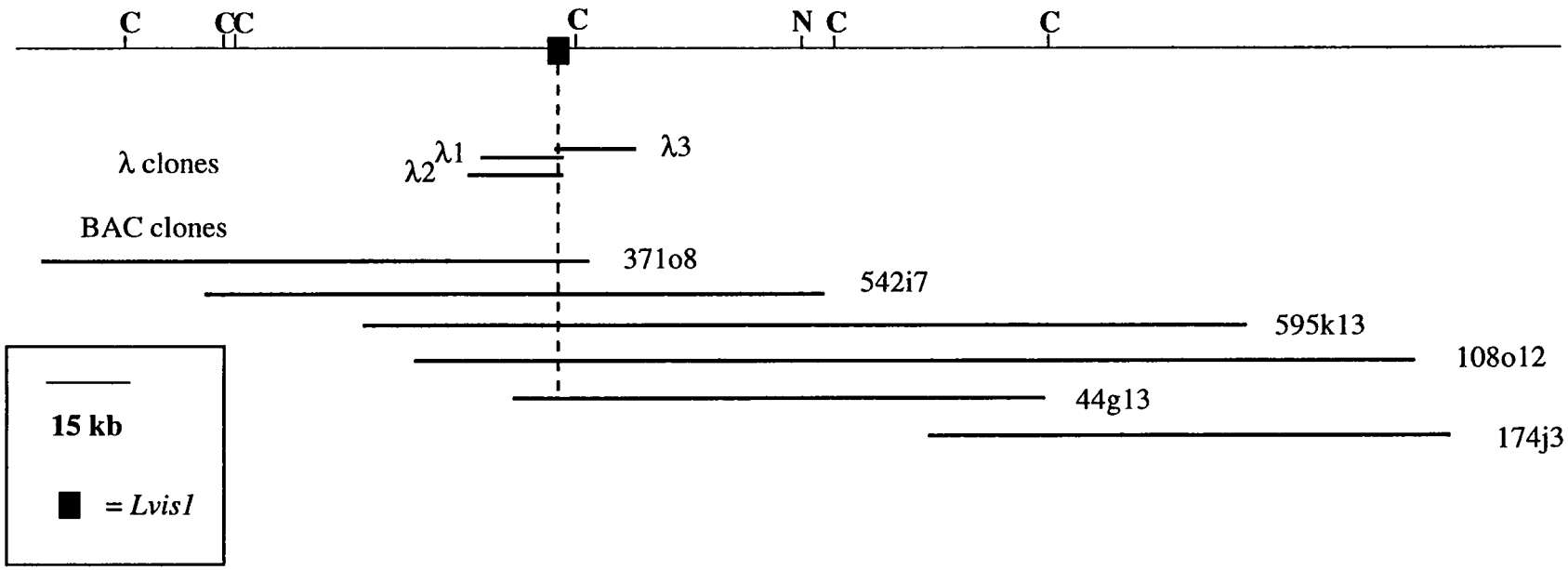


Fig. IV - 4. Physical map of BAC clones at *Lvis1*. Six BAC clones spanning approximately 280 kb were mapped relative to *Lvis1* by restriction enzyme mapping and high-resolution fiber FISH analysis. Clones 542i7 and 595k13 were used for shotgun sequencing to identify genes near the site of viral insertion. Restriction enzymes used: C, *Cla*I and N, *Not*I.

Fig. IV - 5. DNA fiber FISH mapping of BAC clones. BAC clone 371o8 was mapped relative to both clones 102o18 and 174j3 using fiber FISH. A) Example of DNA fiber FISH using clones 371o8 and 102o18. B) and C) Examples of DNA fiber FISH using clones 371o8 and 174j3. Green color represents clone 371o8 labeled with biotin and detected by FITC, while clones 102o18 and 174j3 were labeled with Dig and detected by Rodamine, and are shown in red. The area of clone overlap in panel A appears yellow-orange.

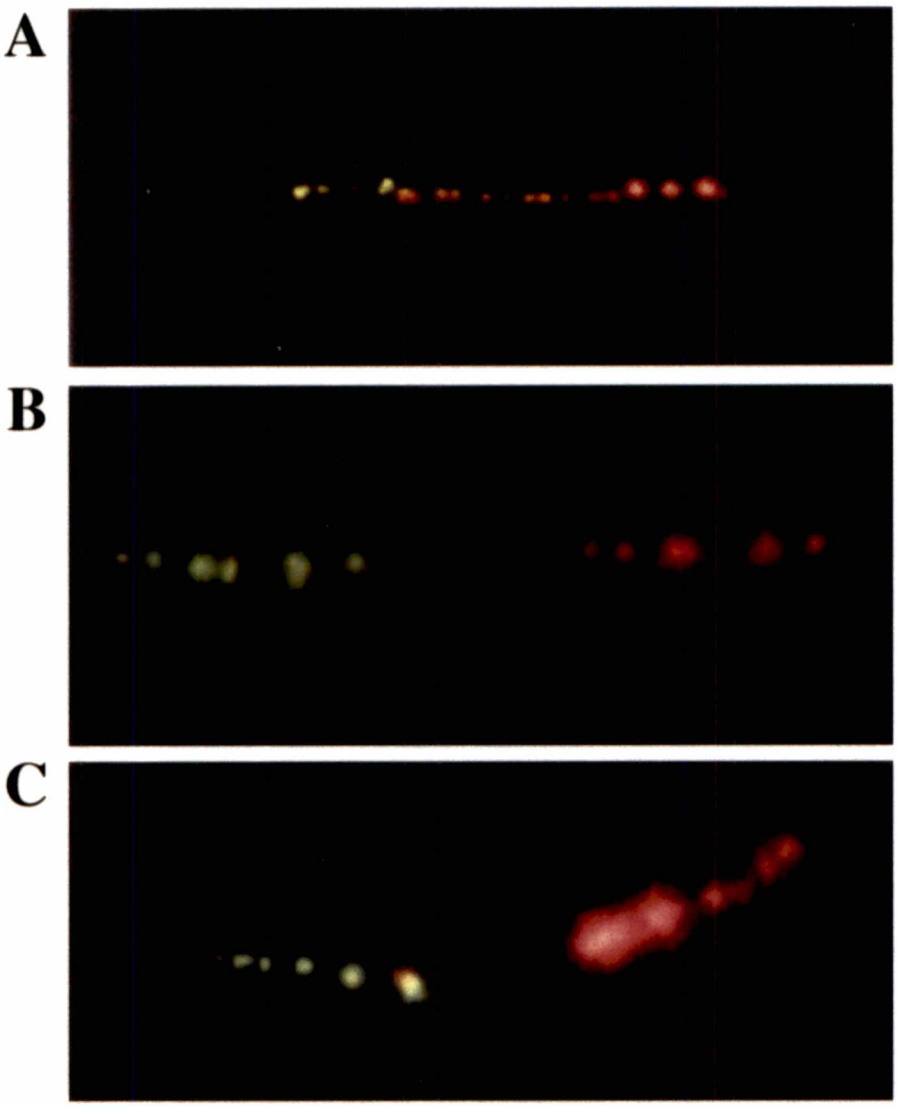

A
B
C

Table IV - 4. Blastn seach results of shotgun sequenceing clones.

Clone	Matching Gene	Acces. No.	Species	Score
G131	Hex	Z21524	mouse	5.1e-34
	Probox protein	X64711	chicken	5.7e-17
	PRH	X67235	human	8.4e-17
	PRH	L16499	human	8.4e-17
G5	HKSP	U37426	human	1.8e-41
	KRP	X85137	human	3.7e-36
H444	L35a	Y16430	mouse	3.7e-144
	L35a pseudogene	X05704	rat	2.5e-137
	L35a pseudogene	X05705	rat	2.5e-137
	L35a	X03475	rat	2.3e-126

Fig. IV - 6. Integrated physical map of genes and mutations at *Lvis1*. Each proviral insertion mutation is depicted by an arrow in the inset at the top of the figure. This mutation map was oriented to the BAC physical map by sequence overlap between this region of insertion (collected from subclones of genomic lambda clones) and the end clone from BAC. Each gene identified is shown as a black box. This box approximates the genomic region spanned by each gene, except for *msec15*, for which this distance has not yet been determined. The direction of transcription is shown for *Hex* and *mEg5*. Enzymes used are *Cla*I, C and *Not*I, N.

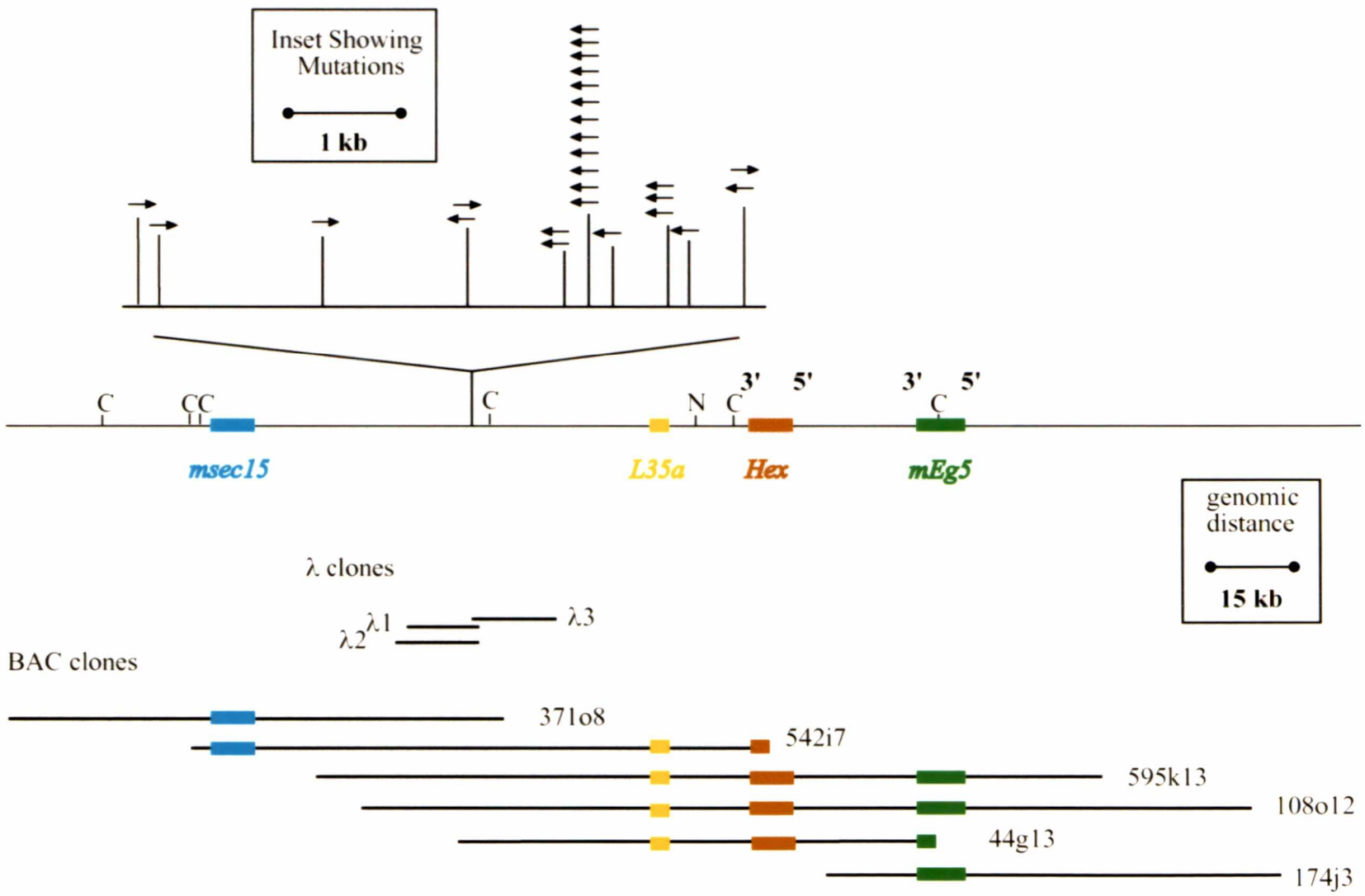

Inset Showing Mutations
1 kb
C
CC
C
N
C
3'
5'
3'
5'
C
msec15
L35a
Hex
mEg5
genomic distance
15 kb
λ clones
λ1
λ2
λ3
BAC clones
371o8
542i7
595k13
108o12
44g13
174j3

Fig. IV - 7. Nucleotide sequence of the *mEg5* cDNA and the predicted amino acid sequence. The nucleotide sequence is numbered to the left. The amino acid sequence is numbered to the right. The consensus cdc2 phosphorylation site is underlined. The locations of the polyA tail sequences of the two shorter transcripts identified using RACE are designated by an asterisk. Possible polyadenylation signal sequences are underlined.

```
   1  GGTCAGTCGTCTCCATTTTTCAAAATTGATGCGCGGAATCGTTGCTGATTTCGGAGCAGA
  61  GAgGAGTTCGTCCTGTCCGGCTTTTGCGGGGCGGCGGCCACgGCCAAGAGcCTGCGTGGA
 121  CTCGGGGACGCCGAGCTGCGAGTCTCGGTCCTCGTGGCCTTGGCAGCACCGGGTGAGGAG
 181  AgGCTGCTCCCGGTTCTCACTGTGTCTGAGTCTCCGCTAGGCCGGCAGGTTTTGGCTCGA
 241  CCGTCATGGCGTCCCAGCCGAGTTCTTTGAAGAAGAAAGAGGAAAAGGGCAGGAACATCC
          M  A  S  Q  P  S  S  L  K  K  K  E  E  K  G  R  N  I  Q    19
 301  AGGTGGTGGTGAGATGCAGACCATTTAATCTGGCAGAGCGGAAAGCTAATGCCCACTCAG
        V  V  V  R  C  R  P  F  N  L  A  E  R  K  A  N  A  H  S  V    39
 361  TAGTGGAATGTGACCATGCACGGAAAGAAGTCAGTGTACGGACTGCAGGGTTGACGGACA
        V  E  C  D  H  A  R  K  E  V  S  V  R  T  A  G  L  T  D  K    59
 421  AGACCTCCAAGAAAACATACACGTTTGATATGGTGTTTGGAGCATCTACAAAACAAATTG
        T  S  K  K  T  Y  T  F  D  M  V  F  G  A  S  T  K  Q  I  D    79
 481  ATGTTTAcCGAAGTGTTGTTTGTCCAATTCTAGATGAAGTTATTATGGGCTATAATTGCA
        V  Y  R  S  V  V  C  P  I  L  D  E  V  I  M  G  Y  N  C  T    99
 541  CCATCTTCGCATATGGTCAGACTGGCACTGGAAAAACtTTTACAATGGAAgGTGAAAGGT
        I  F  A  Y  G  Q  T  G  T  G  K  T  F  T  M  E  G  E  R  S   119
 601  CAcCTAATGAAGTATATACCTGGGAGGAGGATCCTCTGGCTGGTATAATTCCACGCACTC
        P  N  E  V  Y  T  W  E  E  D  P  L  A  G  I  I  P  R  T  L   139
 661  TTCATCAAATTTTTGAGAAACTTACTGATAATGGCACTGAGTTTTCAGTTAAAGTGTCCC
        H  Q  I  F  E  K  L  T  D  N  G  T  E  F  S  V  K  V  S  L   159
 721  TATTGGAAATCTATAATGAGGAGCTTTTTGATCTTCTTAGTCCATCTTCTGATGTTTCTG
        L  E  I  Y  N  E  E  L  F  D  L  L  S  P  S  S  D  V  S  E   179
 781  AAAGGCTGCAGATGTTTGATGATCCCCGGAACAAGAGAgGAGTGATAATCAAAGGCTTAG
        R  L  Q  M  F  D  D  P  R  N  K  R  G  V  I  I  K  G  L  E   199
 841  AGGAAATCACAGTACACAATAAAGATGAAGTCTACCAAATCTTAGAGAAGGGAGCAGCAA
        E  I  T  V  H  N  K  D  E  V  Y  Q  I  L  E  K  G  A  A  K   219
 901  AAAGGACAACTGCAGCAACCTTGATGAATGCTTACTCTAGTCGTTCACACTCAGTTTTTT
        R  T  T  A  A  T  L  M  N  A  Y  S  S  R  S  H  S  V  F  S   239
 961  CTGTTACGATACACATGAAAGAAACAACAATTGATGGAGAAGAGCTTGTTAAAATTGGAA
        V  T  I  H  M  K  E  T  T  I  D  G  E  E  L  V  K  I  G  K   259
1021  AGTTGAATTTGGTTGATCTTGCAGGAAGTGAAAATATTGGGCGTTCTGGAGCTGTTGACA
        L  N  L  V  D  L  A  G  S  E  N  I  G  R  S  G  A  V  D  K   279
1081  AGAgGGCCCGGGAAgCTGGAAATATCAACCAATCCcTCTTGACTCTGGGAAGAGTTATTA
        R  A  R  E  A  G  N  I  N  Q  S  L  L  T  L  G  R  V  I  T   299
1141  CTGCTCTTGTGGAAAGAACACCTCATATTCCTTATCGAGAATCTAAACTAACTAGAATCC
        A  L  V  E  R  T  P  H  I  P  Y  R  E  S  K  L  T  R  I  L   319
1201  TGCAAGATTCTCTTgGGGGACGTACAAGAACATCTATAATTGCAACCATTTCCCCTGCAT
        Q  D  S  L  G  G  R  T  R  T  S  I  I  A  T  I  S  P  A  S   339
1261  CTTTCAATCTTGAGGAAACTCTGAGTACATTGGAATATGCTCACAGAGCAAAGAACATAA
        F  N  L  E  E  T  L  S  T  L  E  Y  A  H  R  A  K  N  I  M   359
1321  TGAATAAGCCTGAAGTTAATCAAAAACTCACCAAAAAAGCTCTTATTAAGGAGTATACAG
        N  K  P  E  V  N  Q  K  L  T  K  K  A  L  I  K  E  Y  T  E   379
1381  AAGAGATAGAGCGTTTGAAGCGAGATCTTGCAGCAGCTCGTGAGAAAAATGGAGTGTACA
        E  I  E  R  L  K  R  D  L  A  A  A  R  E  K  N  G  V  Y  I   399
1441  TCTCTGAAGAAAGTTTTAGAGCCATGAATGGAAAGGTAACTGTTCAGGAGGAACAAATTG
        S  E  E  S  F  R  A  M  N  G  K  V  T  V  Q  E  E  Q  I  V   419
1501  TTGAGTTGGTTGAAAAAATCGCTGTTCTTGAGGAGGAGCTCAGTAAGGCTACAGAGTTAT
        E  L  V  E  K  I  A  V  L  E  E  E  L  S  K  A  T  E  L  F   439
1561  TTATGGATAGTAAGAACGAACTTGACCAGTGTAAATCTGACCTGCAAACCAAGACACAGG
        M  D  S  K  N  E  L  D  Q  C  K  S  D  L  Q  T  K  T  Q  E   459
1621  AACTTGAAACCACTCAGAAACATTTGCAAGAAACAAAATTACAACTGGTTAAAGAGGAAT
        L  E  T  T  Q  K  H  L  Q  E  T  K  L  Q  L  V  K  E  E  Y   479
1681  ATGTCTCTTCAGCCTTGGAAAGAACCGAGAAGACACTGCATGACACGGCCAGCAAGTTGC
        V  S  S  A  L  E  R  T  E  K  T  L  H  D  T  A  S  K  L  L   499
1741  TTAACACGGTTAAAGAAACCACCAGGGCTGTATCTGGTCTACATTCTAAACTGGACCGCA
```

```
          N  T  V  K  E  T  T  R  A  V  S  G  L  H  S  K  L  D  R  K    519
1801  AGAGAGCAATCGATGAGCACAACGCTGAAGCTCAGGAGAGCTTTGGCAAAAACCTCAACA
          R  A  I  D  E  H  N  A  E  A  Q  E  S  F  G  K  N  L  N  S    539
1861  GTCTGTTTAATAATATGGAAGAATTGATTAAGGATGGCAGTGCGAaACAAAAGgCCATGC
          L  F  N  N  M  E  E  L  I  K  D  G  S  A  K  Q  K  A  M  L    559
1921  TAGACGTTCATAAGACACTGTTTGGTAACCTGATGTCTTCTAGTGTCTCTGCATTAGACA
          D  V  H  K  T  L  F  G  N  L  M  S  S  S  V  S  A  L  D  T    579
1981  CCATTACCACGACAGCACTTGAATCTCTCGTGTCTATTCCAGAAAATGTGTCCGCTCGTG
          I  T  T  T  A  L  E  S  L  V  S  I  P  E  N  V  S  A  R  V    599
2041  TTTCTCAGATTTCTGATATGATATTGGAAGAGCaATCGTTAGCAGCACAAAGTaAAAGTG
          S  Q  I  S  D  M  I  L  E  E  Q  S  L  A  A  Q  S  K  S  V    619
2101  TTCTGCAAGGATTGATTGATGAACTTGTGACCGAcCTTTTCACTTCCCTGAAGACCATCG
          L  Q  G  L  I  D  E  L  V  T  D  L  F  T  S  L  K  T  I  V    639
2161  TAGCCCCTAGTGTGGTTTCCATCTTGAACATAAATAAGCAGCTACAGCATATTTTCAGGG
          A  P  S  V  V  S  I  L  N  I  N  K  Q  L  Q  H  I  F  R  A    659
2221  CTTCATCGACAGTGGCTGAAAAGGTAGAAGATCAAAAAAGAGAAATAGACAGTTTTCTCA
          S  S  T  V  A  E  K  V  E  D  Q  K  R  E  I  D  S  F  L  S    679
2281  GCATATTGTGTAACAATTTACATGAACTCCGAGAAAACaCAGTTTCTTCCTTGGTTGAAT
          I  L  C  N  N  L  H  E  L  R  E  N  T  V  S  S  L  V  E  S    699
2341  CACAAAAGCTTTGTGGAGACCTAACTGAAGACCTGAAGACAATAAAGGAAACCCATTCAC
          Q  K  L  C  G  D  L  T  E  D  L  K  T  I  K  E  T  H  S  Q    719
2401  AGGAACTTTGCCAGTTAAGCAGTAGTTGGGCAGAGAGATTCTGTGCTTTGGAGAAGAAGT
          E  L  C  Q  L  S  S  S  W  A  E  R  F  C  A  L  E  K  K  Y    739
2461  ATGAAAACATCCAGAAACCACTGAACAGTATTCAAGAAAATACAGAGCGGAGGTCTACTG
          E  N  I  Q  K  P  L  N  S  I  Q  E  N  T  E  R  R  S  T  D    759
2521  ATATAATCAATAAAACAACAGTTCACAGTAAGAAAATTCTTGCTGAATCTGATGGATTAT
          I  I  N  K  T  T  V  H  S  K  K  I  L  A  E  S  D  G  L  L    779
2581  TACAAGAACTCAGACACTTTAACCAAGAAGGCACACAGCTGGTTGAAGAGTCTGTAGGAC
          Q  E  L  R  H  F  N  Q  E  G  T  Q  L  V  E  E  S  V  G  H    799
2641  ACTGCAGTTCACTCAACAGCAACCTGGAGACTGTATCCCAAGAGATCACCCAGAAGTGTG
          C  S  S  L  N  S  N  L  E  T  V  S  Q  E  I  T  Q  K  C  G    819
2701  GGAcCCTGAACACAAGCACAGTTCATTTCTCTGATCAGTGGGCATCCTGCCTAAGCAAGA
          T  L  N  T  S  T  V  H  F  S  D  Q  W  A  S  C  L  S  K  R    839
2761  GAAAGGAAGAACTTGAGAATTTAATGGAGTTTGTAAATGGCTGTTGTAAAGCTTCAAGTT
          K  E  E  L  E  N  L  M  E  F  V  N  G  C  C  K  A  S  S  S    859
2821  CGGAGATCACTAAGAAAGTAAGAGAACAGAGCGCAGCTGTTGCGAACCAGCACAGCTCCT
          E  I  T  K  K  V  R  E  Q  S  A  A  V  A  N  Q  H  S  S  F    879
2881  TTGTTGCTCAGATGACTTCCGATGAAGAAAGCTGTAAAGCAGGAAGCCTGGAGCTTGATA
          V  A  Q  M  T  S  D  E  E  S  C  K  A  G  S  L  E  L  D  K    899
2941  AAACTATAAAGACTGGGTTAACAAAGCTGAATTGCTTTCTGAAACAGGATCTGAAACTAG
          T  I  K  T  G  L  T  K  L  N  C  F  L  K  Q  D  L  K  L  D    919
3001  ATATCCCAACAGGTATGACACCAGAGAGGAAAAAATATTTATATCCAACAACACTTGTGA
          I  P  T  G  M  T  P  E  R  K  K  Y  L  Y  P  T  T  L  V  R    939
3061  GAACTGAACCACGAGAGCAGCTCCTTGATCAGCTGCAAAAGAAACAACCACCAATGATGC
          T  E  P  R  E  Q  L  L  D  Q  L  Q  K  K  Q  P  P  M  M  L    959
3121  TAAACAGCTCAGAAGCCAGCAAGGAGACCAGTCAGGACATGGATGAAGAGAGGGAGGCTC
          N  S  S  E  A  S  K  E  T  S  Q  D  M  D  E  E  R  E  A  L    979
3181  TGGAGCAGTGTACTGAGGAACTTGTAAGTCCAGAGACAACTGAACATCCCAGTGCAGATT
          E  Q  C  T  E  E  L  V  S  P  E  T  T  E  H  P  S  A  D  C    999
3241  GCTCTTCCAGCAGAGGTCTTCCATTTTTCCAGCGAAAAAAGCCACATGGAAAAGACAAAG
          S  S  S  R  G  L  P  F  F  Q  R  K  K  P  H  G  K  D  K  E   1019
3301  AAAACAGAGGCCTTAACCCGGTGGAGAAGTATAAAGTGGAAGAGGCCTCGGATCTcTCCA
          N  R  G  L  N  P  V  E  K  Y  K  V  E  E  A  S  D  L  S  I   1039
3361  TCTCCAAGAGCAGACTGCCGCTTCACACCTCCATAAACCTCTAGCTGATCTGAGGCTTAG
          S  K  S  R  L  P  L  H  T  S  I  N  L                       1052
3421  GGTGTCATCTTTAAAATACAACCTGAAACTCCAGAGTCTGAAGCTATGTACArATGAAAA
3481  GGGGACTGCTGTGTGAGGGCCCCACAGTAACTGTAGTTGAACTGAAAGTCTTTTTTATAA
```

3541 TCCCTGTAGTCCAAGGATGTAGTAAGCTGGGTATCATTTGGGATTTACATTGAATATGTG
3601 TGTGTGTTTCTAGCTTTTTATATAAAGAAGCTCTTCTGTAACAAGTAAGTATTTTTCTTG
3661 TATATAATTAAATACCAAATATATGGAAATCATTGTTCCAGGTTTAGACTTGTATTGGTG
3721 AATGCCATCTCCTTTGCTGTCTGGCCAAgGCTGTTTCCCTACCTCTAACCAGCCTTTTCT
3781 AGTTTGTCTTCGACTCCTGTCTCCCTTTTCCTGTCACTTAACCTTTTGTAGCCTACAGAA
3841 AGGTTTCTTTAGTATGAGAAACGCAGGGTTTTACCTGGAACTCTTCTATCTCACTGATTA
3901 CCCTTACAATCACTGTCAAAACACCTGACCCTGGAAGGACCCTTCCTTTTGGGTCCTTCA
3961 TTTGTTCTCGCTGATATGCACACATCTCATTCCTTCCTGAGGTCTGACACATCACCAGCC
4021 TCCCTGTAGTGTCGGCTCCTAATGCAGGAGACCTTGAGGCCCCACTTAAAAGGTCTTTAA
4081 GTCTCTTAGAAATGATAAGATTGTTTTCAAACTACCAATTATGACTATAACTTCTATTTT
4141 GTTTATCTTTTAAAACTTTATTCCTTGTATATAATAAAAATGTATATATAC*ATTCCCTCT
4201 GTTGAAATATTAACCCTCGCTCTTTCGCTTTGACAACGCTGTTTCTGGAAGTTGGCCATT
4261 TTAAGATGAACCAGAGCACCTACCTAAAAGTGTATGCAGAGAGTAAGTGTGGTCACGTCA
4321 AGTTCATGTATACGGTTCAGAGTGATGTGGTTTTGGCACAGTCTCACTATGGAGCCTGAT
4381 CTACAAAGTGAGTGCCAGGACAGCTAGGGCTACAGAGAAACCCTGTCTCAAAAAAACAAA
4441 AGAAATAGAACTTTGTAACAAAGAGCTAACCCCT*AATAGAAAATTAGAAAAAAATGCTAC
4501 TTTAAGAAATTTATTTTCTTGTACTATTTCTCCAACTCATTGTAGAAATAGAAACGTTAT
4561 TTTAATAAAAATAATGCATAAATTTCTCTGCAAACAGATGCCAACCTAATCTGTGGATAT
4621 TATCTAATTTGCTTTTGGTAAATCTTTAATTGGTAATTGTTTTTGAATTTTTTAAATTAA
4681 ATTATTTTAATTTTTTGTTCGTAAAAAAAAAAAAAAAAAAAAAAAAAA

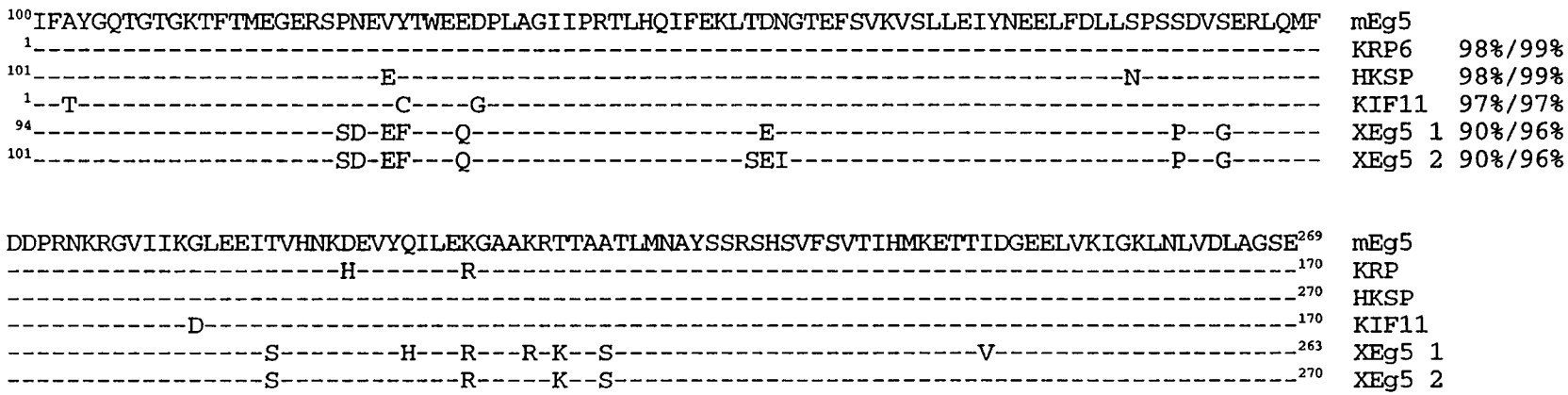


Fig. IV - 8. Comparison of the amino acid similarity between motor domains of bim-C family proteins.
Amino acid sequence comparison of mEg5 (this paper), Kif11 (Genbank acc. no. 2443270), HKSP (Genbank acc. no. 1171153), KRP6 (Genbank acc. no. 2674187), XEG5 1 (Genbank acc. no. 119217), and XEG5 2 (Genbank acc. no. 2497521). The top line lists the amino acid sequence of the motor domain of mEg5. Identical amino acids in the other motor domain sequences are depicted with a dash (-). The identity of each sequence is listed to the right, along with the percent identity/similarity between each sequence and mEg5.

QDLKLDIPTGM**TP**E**RK**KYLYPTTLVRTEPREQLLD mEg5
QDLKLDIPTGT**TP**Q**RK**SYLYPSTLVRTEPREHLLD HEg5
EELRNDVPTGT**TP**Q**RR**DYVYPSLLIKTKPRDVLLE XEg5-2
EELRNDVPTGT**TP**Q**RR**DYAYPSLLVKTKPRDVLLE XEg5-1
SELKTYAPTGT**TP**S**KR**DFVYPRTLVATSPHQEIVR KRP-130
RSLEEYVATGV**TP**K**KR**KYDYISVLPSTESHEVLKS bim-C
TSLVKLETTGD**TP**S**KR**ELPATPSWTRDSSLIKETT Cut7
QKYMDYEPTGA**TP**I**R**SEPDVPSKVTIESLRAMPME TKRP

Fig. IV - 9. Alignment of potential $p34^{cdc2}$ phosphorylation sites in bimC related proteins. Alignment of amino acids 1000-1034 from Cut7 (Genbank acc. no. X57513), 995-1029 from bim-C (Genbank acc. no. M32075), 922-956 from KRP-130 (Swissprot. no. P46863), 914-948 from mEg5, 916-950 from HEg5 (Genbank acc. no. x85137), 926-960 from XEg5-2 (Swissprot. no. Q91783) 919-953 from XEg5-1 (Swissprot. no. P28025) and 937-971 from TKRP (Genbank acc. no. D83711). The consensus site for $p34^{cdc2}$ kinase is S/TPxK/R, and is shown in black bold letters and amino acids which are identical in at least 5 sequences are shown in plain black letters.

Fig. IV - 10. Expression of *L35a* in AKXD-18 spleen and lymph node tumors. Random shotgun sequence clone H444 was used as a probe for Northern analysis of tumor RNA. Each lane contains ~10 μg total RNA, where ethidium bromide staining of 28S RNA is shown in the lower panel as a control for loading. The presence or absence of a retroviral integration at *Lvis1* is summarized as (+) or (-) at the top of the figure. The size of the transcripts detected in each blot are indicated on the left in kilobases.

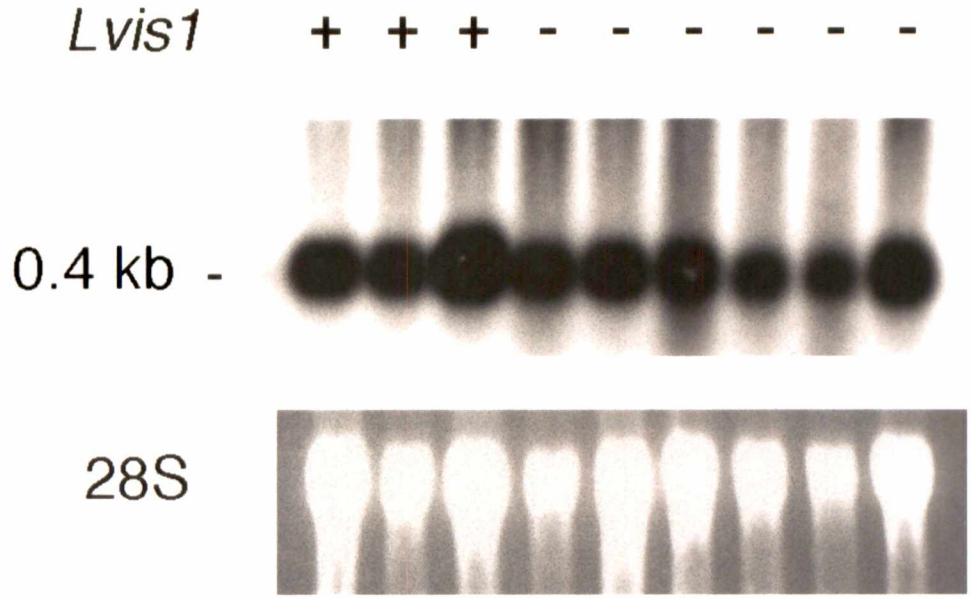

Lvis1
+ + + - - - - - -
0.4 kb -
28S

Fig. IV - 11. Expression of *Hex* in AKXD-18 spleen and lymph node tumors. A cDNA probe containing the first three exons of the *Hex* gene was used as a probe for Northern analysis of tumor RNA. Each lane contains ~10 μg total RNA, where ethidium bromide staining of 28S RNA is shown in the lower panel as a control for loading. The presence or absence of a retroviral integration at *Lvis1* is summarized as (+) or (-) at the top of the figure. The size of the transcripts detected in each blot are indicated on the left in kilobases.

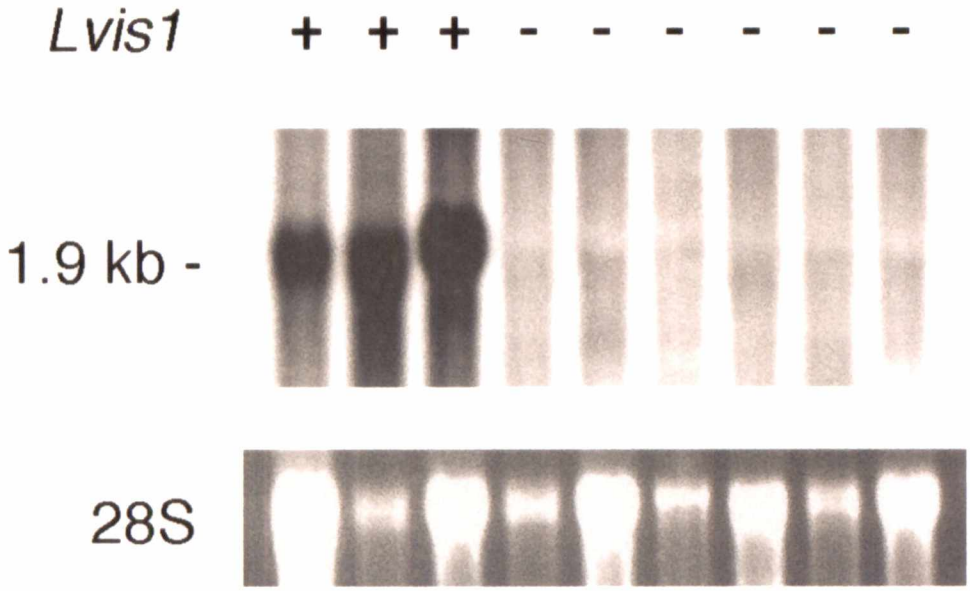

Lvis1
+ + + - - - - - -
1.9 kb -
28S

Fig. IV - 12. Expression of *msec15* in AKXD-18 spleen and lymph node tumors. A mouse EST clone representing *msec15* (dbEST ID no. 606262) was used as a probe for Northern analysis of tumor RNA. Each lane contains ~10 μg total RNA, where ethidium bromide staining of 28S RNA is shown in the lower panel as a control for loading. The presence or absence of a retroviral integration at *Lvis1* is summarized as (+) or (-) at the top of the figure. The size of the transcripts detected in each blot are indicated on the left in kilobases..

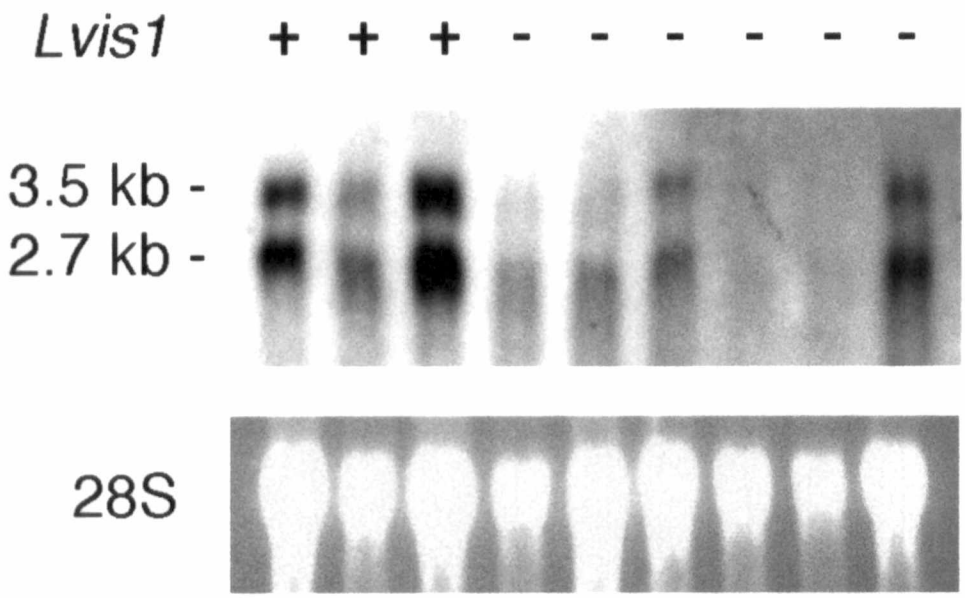
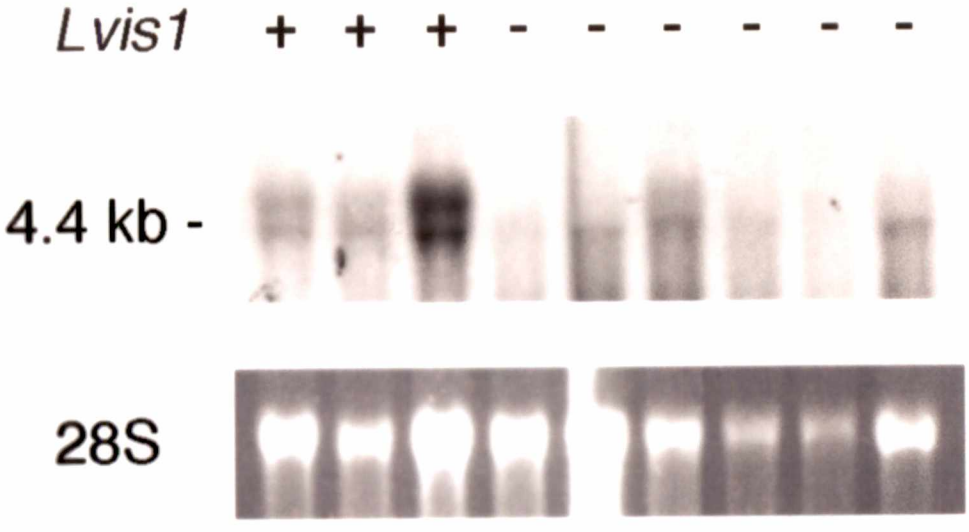

Lvis1
+ + + - - - - - -
3.5 kb -
2.7 kb -
28S

Fig. IV - 13. Expression of *mEg5* in AKXD-18 spleen and lymph node tumors. A 903 bp cDNA probe containing *mEg5* 3' UTR sequence (nucleotide 3397 - 4300) was used as a probe for Northern analysis of tumor RNA. Each lane contains ~10 μg total RNA, where ethidium bromide staining of 28S RNA is shown in the lower panel as a control for loading. The presence or absence of a retroviral integration at *Lvis1* is summarized as (+) or (-) at the top of the figure. The size of the transcripts detected in each blot are indicated on the left in kilobases


Lvis1
+ + + - - - - - -
4.4 kb -
28S

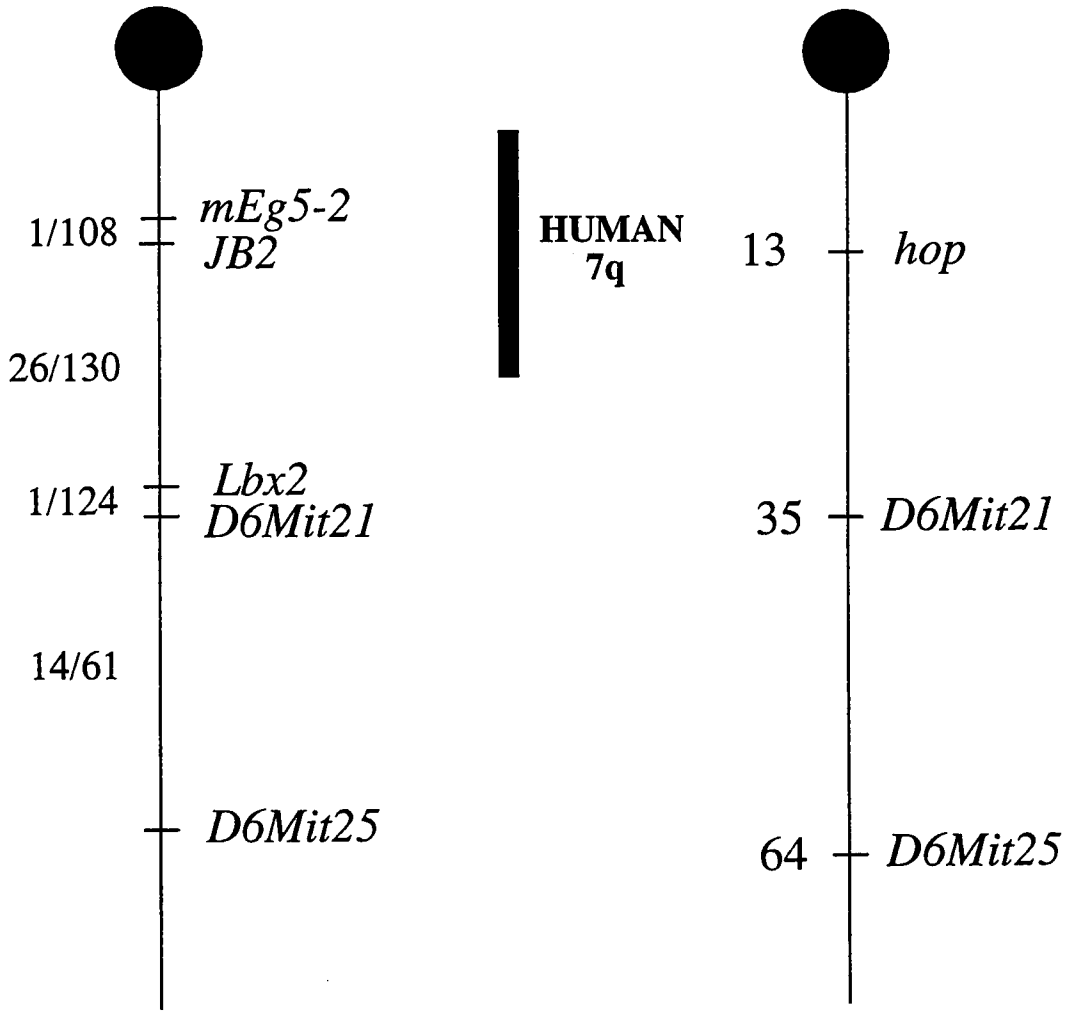


Fig. V-1. Mapping of *mEg5-2*. Molecular genetic linkage maps of mouse chromosome 6. The map to the right is a partial linkage map from the Mouse Genome Database (MDG, June 1998). Numbers to the right of this map indicated map position. The map to the left is a molecular genetic linkage map generated using the (SB/Le x M. spretus)F1 x SB/Le backcross. Numbers to the right of this map indicate the number of crossovers/number animals analyzed. The human homologous region is indicated between the maps.

VITA

Gwenn M. Hansen was born in Virginia, Minnesota on October 11, 1970. She attended public school in Cook, Minnesota, graduating from Cook High School in May, 1989. She entered Gustavus Adolphus College, St. Peter, Minnesota in August of 1989 where in May, 1993 she received a Bachelor of Arts degree in Biology and Psychology. She began graduate school in the Division of Biology at Kansas State University in August of 1993. During her first semester, she chose to work in the laboratory of Dr. Monica J. Justice. She began her dissertation research project in January, 1994, studying the molecular genetic aspects of lymphoma development in mice. During her studies at Kansas State University, she taught laboratory classes in Immunology, Microbiology and General Biology. In August, 1995, she moved with Dr. Justice to Oak Ridge National Laboratory to continue her graduate studies. At this time she transferred to the University of Tennessee - Oak Ridge Graduate School of Biomedical Sciences. As a Ph. D. student, she had the honor of being awarded membership in the Phi Kappa Phi Honor Society in 1996. In July of 1998 she successfully defended her graduate dissertation and was awarded a Doctor of Philosophy degree in biomedical sciences in August, 1998.

Gwenn is currently a postdoctoral fellow in the Department of Molecular and Human Genetics at Baylor College of Medicine in Houston, Texas.