Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 6350–6354, July 1993
Genetics

Analysis of the murine *All-1* gene reveals conserved domains with human *ALL-1* and identifies a motif shared with DNA methyltransferases

QING MA, HANSJUERG ALDER, KELLY K. NELSON, DEVJANI CHATTERJEE, YANSONG GU, TATSUYA NAKAMURA, ELI CANAANI, CARLO M. CROCE, LINDA D. SIRACUSA, AND ARTHUR M. BUCHBERG*

Jefferson Cancer Institute and Department of Microbiology and Immunology, Jefferson Medical College, 233 South 10th Street, Philadelphia, PA 19107-5541

Communicated by Peter C. Nowell, March 31, 1993

ABSTRACT **A series of translocation break points found in a subset of human acute leukemias have one of the breaks on human chromosome 11q23. This region has recently been cloned and a large gene, *ALL-1*, with homology to the *Drosophila* trithorax gene has been identified. This paper describes the cloning, sequencing, and mapping of the mouse homolog of *ALL-1*. We have found a motif present in *All-1* that shows homology to the zinc-binding domain of DNA (cytosine-5) methyltransferases (EC 2.1.1.63). Sequence analysis of the murine *All-1* gene has identified distinct regions of homology with the human *ALL-1* gene; these highly conserved domains may define regions of functional significance in mammals. In addition, we have identified alternatively spliced forms of *All-1* within one of the zinc-finger domains, suggesting that there may be different targets and/or functions for All-1 proteins. Finally, we report that *All-1* resides in the proximal portion of mouse chromosome 9 and is a candidate for a mutation that results in skeletal transformations during embryonic development.**

The molecular basis of cancer is steadily being uncovered due to investigations focusing on genes whose altered expression leads to abnormal cellular differentiation and/or proliferation. Identification of the genes responsible has been possible due to numerous chromosomal rearrangements found in specific types of hematopoietic tumors (for review, see refs. 1 and 2). These rearrangements provide cytogenetic landmarks to follow in studies directed at identifying and cloning oncogenes. Chromosomal translocations play a role in tumorigenesis by activating cellular protooncogenes or by resulting in the production of chimeric genes capable of transforming hematopoietic cells (for review, see refs. 1 and 3).

Chromosomal rearrangements involving human chromosome (chr) 11q23 have been found in acute lymphocytic leukemia, acute myeloid leukemia, and acute monoblastic and myelomonocytic leukemia (for review, see ref. 2). Reciprocal translocations have been observed in leukemic cells between chr 11q23 and chr 1, 2, 4, 6, 9, 10, 15, 17, 19, or X. The most common rearrangement is a reciprocal translocation between chr 4q21 and chr 11q23; it is found in ≈10% of patients with acute lymphocytic leukemia, most frequently in infants. These observations suggested that a gene at or near the chr 11q23 break point is involved in development or differentiation of hematopoietic lineages and that altered expression of this gene leads to leukemia.

Chr 11q23 translocation break points were determined to lie between the *CD3* and *PBGD* genes, based on results from somatic-cell hybrid and fluorescent *in situ* hybridization studies (4–7). Yeast artificial chromosome libraries were screened and clones containing the *CD3D* and *CD3G* genes were identified (7–9). The isolated yeast artificial chromosome clones spanned the translocation break points, which were clustered in a small region only a few kilobases in length, indicating that the protooncogene being sought resided at or very near the site of the break-point fusions (8, 9).

The search for the gene led to the cloning of a gene called *ALL-1* (8–10), *HRX* (11), *Htrx1* (12), or *MLL* (13). The *ALL-1* gene was found to contain a single long open reading frame able to code for 3962 aa (10, 11). Three regions of the *ALL-1* gene demonstrated strong homology with the *Drosophila* trithorax (*trx*) gene (14). These regions are cysteine-rich and two of them contain zinc-finger-like domains (10, 11). This homology suggests that *ALL-1* is the mammalian homolog of *trx* and may function as a factor that interacts with DNA or DNA–protein complexes to govern developmental processes. We report here the sequencing and mapping of the murine *All-1* gene.† The results reveal four regions of high homology between the mouse and human genes (indicating four functional domains) and also identify a motif within *All-1* implicated in protein–DNA interactions that may discriminate between methylated and unmethylated DNA. Finally, the chromosomal location of *All-1* in the mouse suggests that it may be a candidate gene for the luxoid (*lu*) mutation, which affects skeletal morphology and limb development (15).

MATERIALS AND METHODS

Screening of Libraries and Sequencing. Mouse WEHI-3 cell line and C57BL/6 spleen and B6/CBA lung cDNA libraries (Stratagene) were screened (16) with human cDNA clones (V1, SKV3, V8, and V26) spanning the *ALL-1* gene (10). The libraries were then rescreened with the mouse cDNA clones as probes to generate a contig of overlapping cDNA sequences. Genomic clones were isolated from a 129/Sv cosmid library (Stratagene) by using cDNA clones as probes. Plasmid inserts were sequenced using the Applied Biosystems model 373A DNA sequencing system. Sequence analysis was performed using the software package from the Genetics Computer Group (Madison, WI) (17).

Southern Blot Analyses. The interspecific backcross of [(AEJ/Gn × *Mus spretus*)F_1 × AEJ/Gn] mice was as described (18). Genomic DNA extractions, restriction digestions, gel electrophoresis, Southern blot transfers, and washes were as described (19); hybridization conditions were as described (18). The *All-1* probe was ^{32}P-labeled by random priming; the *d*, *Odc–rs14*, and *Tpi–rs4* probes were ^{32}P-labeled using nick-translation (Boehringer Mannheim).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: cM, centimorgan(s); chr, chromosome; SSLP, simple sequence length polymorphism; MTase, methyltransferase; GTE, glycine, threonine, and glutamic acid; RFLP, restriction fragment length polymorphism.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. L17069).

Conditions for PCR Analyses. DNA oligomers for sequencing and simple sequence length polymorphism (SSLP) mapping were made using an Applied Biosystems model 394 DNA synthesizer. To detect *D9Mit2* and *Cd3d*, genomic DNA from each N_2 progeny was amplified with primer pairs, *Taq* DNA polymerase, and buffer (Boehringer Mannheim). The thermo cycler protocol (Perkin–Elmer/Cetus) was an initial denaturation at 94°C for 4 min, followed by 40 cycles of 94°C for 30 sec, 54°C for 60 sec, and 72°C for 90 sec. PCR products were visualized by ethidium bromide staining of 3% agarose gels.

RESULTS

Sequence Analysis of the Murine *All-1* Gene. Human cDNA clones were used to probe murine cDNA libraries to identify the murine *All-1* gene. We identified 22 cDNA clones that formed a contig of 14.67 kb, spanning most of the predicted length of the *All-1* transcript. The nucleotide sequence terminates in a poly(A) tract, representing the 3′ end of the mRNA. An AATAA sequence 20 bases from the 3′ end represents a partial consensus motif for poly(A) addition (AATAAA) (for review, see ref. 21).

The 3′ untranslated region is 3264 nt long and contains four repeats of ATTTA (Fig. 1*B*). This motif is found in mRNAs that are degraded rapidly, including mRNAs encoding growth factors and protooncogenes, and is believed to play a direct role in governing mRNA turnover (22). Three of these ATTTA motifs are conserved with the human *ALL-1* sequence, whereas the region surrounding the fourth ATTTA motif has not yet been sequenced in humans. The conservation of these motifs in the untranslated region of *All-1* between mouse and humans suggests a functional importance and indicates that All-1 mRNA may be degraded rapidly.

Characteristics of the *All-1* Predicted Open Reading Frame. The deduced amino acid sequence of All-1 is shown in Fig. 1*A*. The predicted mouse protein is highly conserved with the human ALL-1 protein (see below). Consistent with the human sequence, we find three regions exhibiting sequence conservation with the trx protein of *Drosophila* (Fig. 1*A* and ref. 4). These regions include two cysteine-rich regions (aa 1333–1533 and 1775–1882) in the middle of the protein believed to be involved in the formation of DNA binding zinc fingers (trx zinc finger), as well as the C-terminal end (aa 3655–3870), which shows high homology to the C terminus of trx. Additionally, we find the same domains that are related to the consensus sequence involved in binding in the minor groove of A·T-rich DNA (23). These "A·T-hook" domains are localized in the N-terminal end of All-1 (Fig. 1*A* see Fig. 4, region A); additionally, the first A·T hook also contains a potential cdc2 kinase phosphorylation site (24). Unlike human ALL-1, the second potential cdc2 kinase phosphorylation site is not present in the murine All-1 protein.

All-1 Contains a Cysteine-Rich Region Conserved with Mammalian DNA Methyltransferase (MTase). Sequence analysis of the predicted murine open reading frame revealed another conserved sequence motif. This region is located at aa 1053–1119 and possesses a high cysteine content (Fig. 1, see Fig. 4, region B). This domain is homologous to a region in the mammalian DNA MTase (Fig. 2) (25–27). Two adjacent

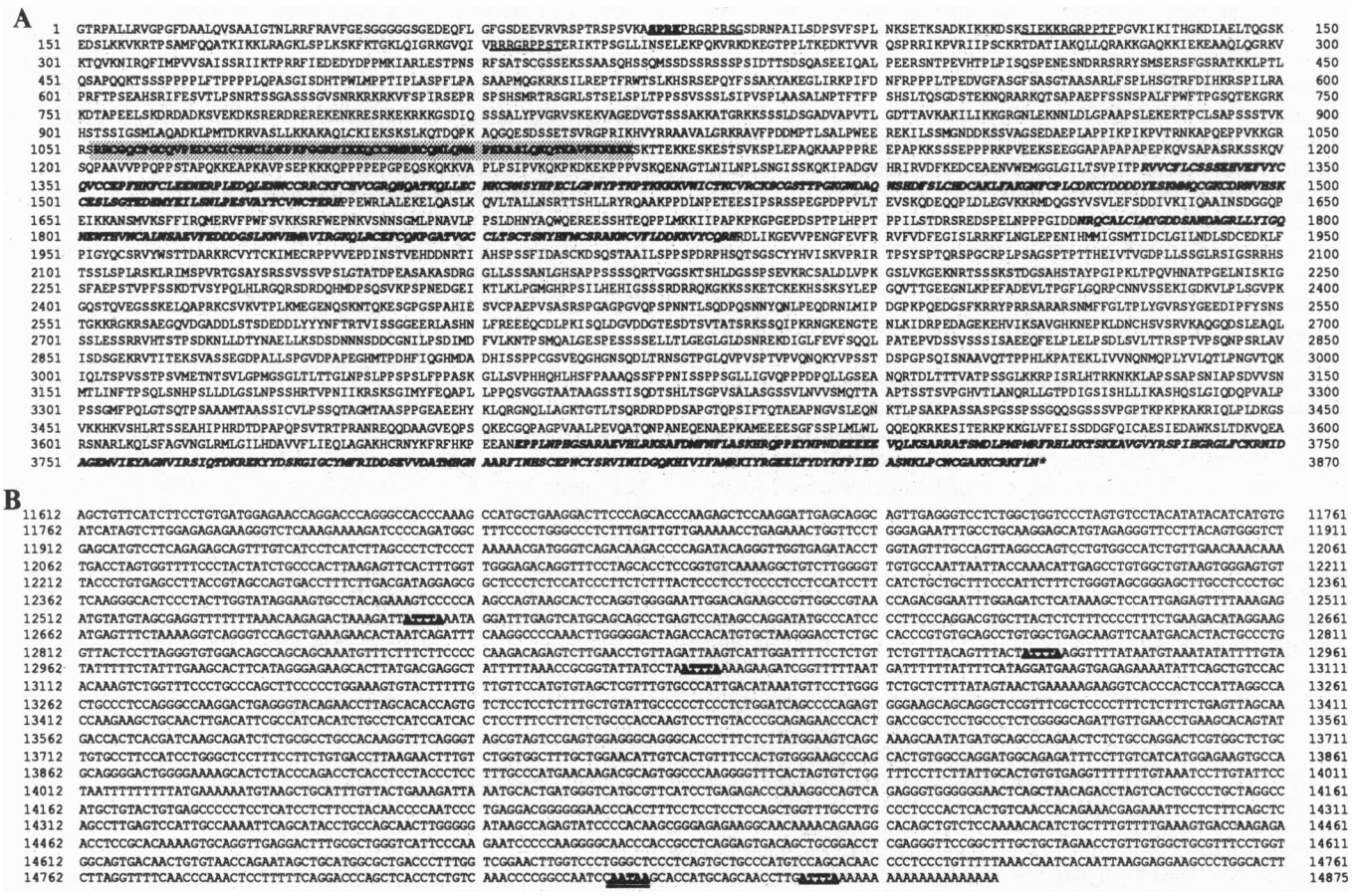

FIG. 1. Predicted amino acid sequence of the murine *All-1* gene and 3′ untranslated nucleotide sequence. (*A*) Open reading frame. Boldface italicized type represents regions homologous to the *Drosophila trx* gene. Underlined residues (aa 72–210) identify the region showing homology to the A·T-hook motif. Boldface shaded type identifies the region showing homology to mammalian DNA MTases. (*B*) A single underline in the 3′ untranslated region identifies "ATTTA" repeats; a double underline identifies the potential poly(A) addition sequence "AATAA." Additional 5′ sequences have not yet been characterized.

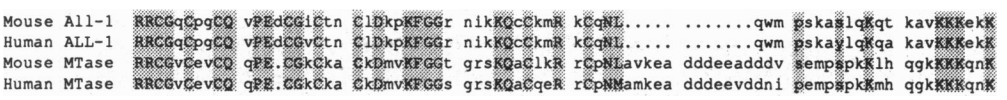

FIG. 2. Alignment of All-1 with mammalian DNA MTases. The aligned regions of mouse All-1, human ALL-1, mouse DNA MTase, and human DNA MTase are shown. The shaded areas identify regions of identity in at least three of the four aligned sequences. The region shown from the mouse All-1 sequence is aa 1053–1119. The dots indicate gaps inserted to optimize alignment. Amino acids identical in all four proteins are shown in uppercase type.

areas exhibit this homology: the first is the cysteine-rich domain showing high sequence identity with the DNA MTase (25 out of 45 aa are identical, including all the cysteine residues) and the second, located just to the C-terminal side of this domain, is a region rich in basic residues. These regions are conserved in both the mouse and human *All-1* genes and between the mouse and human MTases (Fig. 2). The DNA MTases are a group of enzymes involved in methylating cytosines in CpG sites (26, 27); the presence of this motif in *All-1* has significant implications for the proposed functions of *All-1* (see *Discussion*).

Alternative Forms of *All-1* Are Due to Alternative Splicing. We identified two *All-1* cDNA clones from the spleen cDNA library that were identical except for the presence or absence of 9 nt. The presence of these 9 nt results in the addition of glycine, threonine, and glutamic acid (GTE) to the protein. The area of the sequence difference is located 3 aa from the C-terminal end of the fourth trx zinc finger domain as proposed by Gu *et al.* (10), or alternatively, within the sixth trx zinc finger domain as proposed by Tkachuk *et al.* (11). Interestingly, when comparing the sequence with the known sequence of *ALL-1*, the sequence reported by Gu *et al.* (10) contains the GTE tripeptide, whereas the sequence reported by Tkachuk *et al.* (11) lacks the GTE tripeptide. The results from cloning and sequencing the mouse *All-1* gene suggest that the alternative forms (with and without GTE) are a result of alternative splicing. To confirm that the sequence difference could be attributed to alternative splicing, the genomic region flanking the GTE tripeptide was entirely sequenced from a murine *All-1* cosmid clone (Fig. 3). The extra 9 nt are located at the 3′ end of the exon encoding the fourth trx zinc finger (Fig. 3). Both putative splice donor sites (with and without GTE) conform to the consensus splice donor site (for review, see ref. 28). Reverse transcription/PCR analysis of total RNA from spleen confirmed that both forms of the mRNA are present (data not shown).

Similarity of the Predicted All-1 Protein Between Mouse and Human. Fig. 4 shows a comparison of the murine and human All-1 proteins. Overall there is a 90.8% identity of the predicted amino acid sequences. This homology appears to be divided into four larger regions of higher sequence identity, disrupted by three smaller regions of lower sequence identity (Fig. 4), suggesting the presence of four domains in mammalian All-1. These domains include the N-terminal region containing the A·T hook and DNA MTase zinc-binding domains (region 1, A and B), the middle region exhibiting homology to the zinc fingers of trx (region 2, C and D), and the C-terminal domain also showing high sequence identity with trx (region 4, E). In addition, the region between the trx zinc fingers and C terminus also exhibits a high degree of sequence identity between mouse and human (region 3), suggesting the presence of a fourth structural domain important in All-1 normal functions.

The *All-1* Gene Maps to Mouse Chr 9. To further characterize the *All-1* locus, its murine chromosomal location was determined using interspecific backcross analyses. Genomic DNA from the parents of the interspecific backcross (AEJ/Gn and *M. spretus*) was digested with several restriction endonucleases and analyzed by Southern blot hybridization using the *All-1* probe to identify restriction fragment length polymorphisms (RFLPs) useful for establishing the map location of *All-1*. The segregation pattern of the *M. spretus*-specific *Bcl* I fragment was followed in the N_2 progeny and compared to the segregation patterns of known loci previously mapped using RFLPs or SSLPs (Table 1 and Fig. 5). Table 1 lists the molecular probes and PCR oligomer pairs along with their corresponding RFLPs and SSLPs for the loci used to position *All-1* on mouse chr 9. Gene order was resolved by minimizing the number of multiple recombinants along the length of the chromosome. The order of the loci and the ratio of the number of recombinants to the total number of N_2 offspring examined are as follows: *Odc–rs14*–(10/146)–*Tpi–rs4*–(4/87)–*D9Mit2*–(5/107)–[*All-1, Cd3d*]–(16/106)–*d*. The genetic distances between the loci in centimorgans (cM ± SE) are as follows: *Odc–rs14*–(6.8 ± 2.1 cM)–*Tpi–rs4*–(4.6 ± 2.2 cM)–*D9Mit2*–(4.7 ± 2.0 cM)–[*All-1, Cd3d*]–(15.8 ± 3.6 cM)–*d*. Placement of these genes on the linkage map of mouse chr 9 is shown in Fig. 5. No recombinants were detected between *All-1* and *Cd3d* in 104 N_2 progeny, indicating that these loci are tightly linked and must lie <2.9 cM apart (upper 95% confidence limit).

DISCUSSION

The murine *All-1* gene has been cloned and sequenced. Analysis of the sequence has shown that the predicted protein is at least 3923 aa long. Comparison of the murine *All-1* gene with the human *ALL-1* gene identifies four distinct domains of high (>94%) sequence identity, suggesting a functional importance for each region.

The motif reported here in the mammalian *All-1* genes, which shows homology to the DNA MTases, helps formulate a comprehensive hypothesis concerning the functions of mammalian *All-1*. The preferred substrate of the mammalian

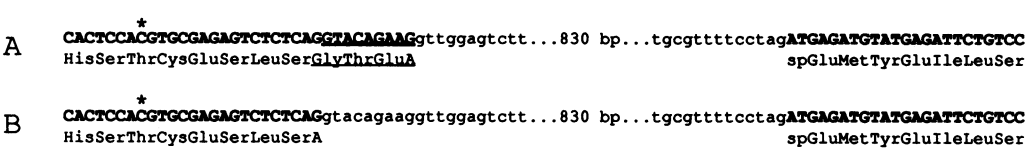

FIG. 3. Exon–intron sequence in the fourth *trx*-like zinc-finger region. Oligonucleotides derived from the cDNA sequence flanking the GTE tripeptide were used to amplify the relevant region of the genomic clone by PCR. The nucleotide and predicted amino acid sequences around the splice site are shown. Exon sequences are in boldface uppercase type; intron sequences are in lowercase type; most of the intron is not shown. (*A*) Sequence of the spliced mRNA +GTE; underlined nucleotides and amino acids represent the region encoding the GTE. (*B*) Nucleotide and amino acid sequences of the −GTE form. The starred triplet represents a coding difference between the cDNA and genomic sequences. In the cDNA sequence shown in Fig. 1*A*, aa 1500 is Lys (AAG), whereas the corresponding genomic sequence encodes a Thr (ACG); this difference may be due to a polymorphism between C57BL/6 (origin of cDNA clones) and 129/Sv (origin of genomic clones) or to *Taq* DNA polymerase error during PCR amplification of the genomic sequence.

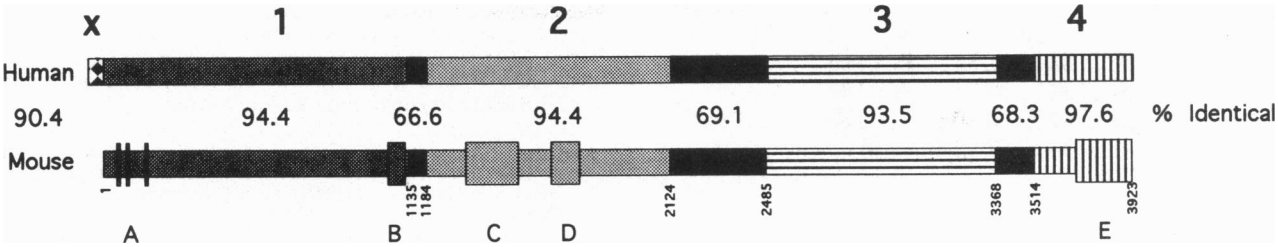

FIG. 4. Comparison between mouse and human *All-1*-deduced protein sequences. Schematic representation of human ALL-1 (*Upper*) and murine All-1 (*Lower*) proteins. The different shaded regions indicate the different domains identified on the basis of homology. The percent sequence identity is shown between the two lines. The top numbers represent four domains of high sequence identity; X represents the portion of human ALL-1 not yet analyzed in mouse. The relative position of the various motifs identified in All-1 are indicated by the enlarged boxes: A, A·T hook region; B, DNA MTase domain; C, the first *trx* conserved zinc-finger domain; D, the second *trx* conserved zinc-finger domain; E, the C-terminal *trx* conserved region. Numbers beneath the mouse sequence represent the amino acid position for the border of each domain.

MTases is hemimethylated DNA; the action of these enzymes produces a fully methylated double-stranded DNA molecule (25). The DNA MTase is a composite protein, with its C-terminal 500 aa encoding the catalytic domain and has homology with bacterial type II DNA MTases. The 1000-residue N-terminal domain contains the cysteine-rich zinc-binding region. Cleavage of this N-terminal domain from the C-terminal domain results in the promiscuous methylation of unmethylated substrates (33). These results suggest that the N-terminal region of DNA MTase differentiates between unmethylated and hemimethylated DNA. The DNA MTase cysteine-rich region binds zinc and may be capable of forming a zinc-finger-like domain (33). These observations suggest that the putative zinc finger of DNA MTase is capable of distinguishing hemimethylated from unmethylated CpG dinucleotides. It has been proposed that the N-terminal region interacts with its substrate (hemimethylated DNA) via a methylation-dependent alteration in the conformation of the DNA molecule, rather than direct contact in the major groove (33). The high degree of homology of the All-1 cysteine-rich domain (Fig. 4, motif B) with the domain of DNA MTases suggests that the All-1 protein has the capacity to distinguish between methylated and unmethylated DNA.

The inclusion of the DNA MTase zinc-binding domain (Fig. 4, motif B) with the trx conserved domains (Fig. 4, motifs C–E) suggests that the function of *All-1* has evolved to take into consideration the presence of methylated cytosine sequences in mammalian DNA. Thus, the state of methylation of a given target gene may influence the ability of *All-1* to regulate its expression. Tissue-specific methylation of *All-1* target genes would thus provide one level of control governing *All-1* in its ability to spatially and temporally regulate the expression of its target genes.

The initial analysis of murine *All-1* revealed the presence of alternative splicing. This alternative splicing is due to a choice of two distinct splice donor sites, with the additional 9 nt occurring at the end of an exon. The Wilms tumor zinc-finger protein, WT1, is a sequence-specific transcription factor that is also subject to alternative use of splice donor sites (34, 35). Specifically, between zinc fingers 3 and 4 of WT1, the alternative form has an additional 3 aa. These additional 9 nt are localized at the 3′ end of the exon, analogous to what we have identified for murine *All-1*. Moreover, it was shown that the alternative forms of WT1 result in proteins with different recognition sequences (36, 37). By analogy, these results suggest that alternative All-1 forms have distinct target specificity as well.

The alternative splicing occurs in the first zinc-finger domain homologous to *Drosophila trx*. Interestingly, the +GTE form disrupts this conserved domain. The −GTE form is conserved between mammals and *Drosophila*, suggesting a functional importance for this spliced form. However, the +GTE form is only conserved between mammals (no evidence exists for its presence in *Drosophila*), suggesting a distinct role for the +GTE form of the All-1 protein in the mammalian lineage.

The *All-1* locus was found to map near the *Cd3d* locus on mouse chr 9. This finding was not unexpected, since a probe for the human homolog of *Cd3* was used to isolate a yeast artificial chromosome clone that contained the *ALL-1* gene in humans (8, 9, 38). *All-1* maps near two spontaneous mouse mutations, luxoid (*lu*) and rough (*ruf*) (15, 39). The *ruf* mutation causes a rough coat and mild hyperkeratosis of the skin (39), whereas the *lu* mutation results in preaxial polydactyly of the hindfeet in heterozygotes and in preaxial polydactyly of both forefeet and hindfeet as well as tibial hemimelia and an increased number of vertebrae, ribs, and

Table 1. Listing of RFLPs and SSLPs used for mapping *All-1* in the mouse

Locus	Probe or primer	Gene name	Enzyme	Size, kb		Ref.
				AEJ/Gn	*M. spretus*	
All-1	All-1	Acute lymphocytic leukemia	*Bcl* I	6.0	14.5	9
d	D46	Dilute coat color locus	*Bgl* I	15.5	13.0	29
Odc–rs14	pCR6	Ornithine decarboxylase-related sequence 14	*Kpn* I	—	7.7	30
Tpi–rs4	pHTPI-5A	Triose phosphate isomerase-related sequence 4	*Hin*dIII	—	9.7	30
D9Mit2	GTGGTCTGCCCTCTTCACAT CAAAGCCAGTCCAACTCCAA	DNA segment MIT-2		180	160	31
Cd3d	AAGAAGTTTCCATGACATCATGAA AGAAGAAAATTCTTGACAGCTCTG	CD3 antigen, δ polypeptide		210	340	31

Underlined fragments identify the segregating *M. spretus* alleles followed in the N_2 progeny. The *All-1* probe used was a 1-kb *Eco*RI fragment derived from one end of the cos20 clone (9); the probe covers a portion of the 3′ end of intron 2 and the 5′ end of exon 3 (10). The D46 probe contains ≈4.2 kb of the 3′ end of the dilute cDNA (29). In AEJ/Gn at the *Odc–rs14* and *Tpi–rs4* loci, multiple fragments were observed; only the fragments identifying *Odc–rs14* and *Tpi–rs4* are given. *Cd3d* has been alternatively referred to as *T3d* and *D9Mit23*.

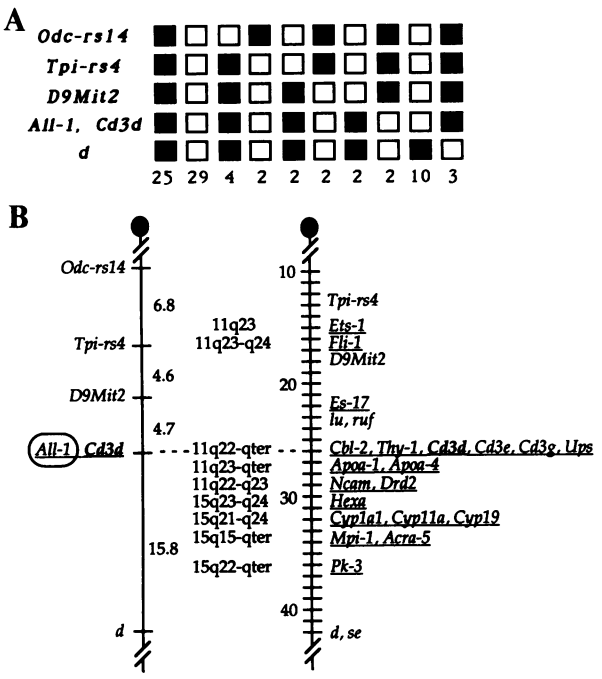

FIG. 5. Chromosomal location of *All-1* in the mouse. (*A*) Haplotype analysis of 81 N_2 progeny from the backcross. Loci followed are listed to the left. Each column represents the chromosome identified in the N_2 offspring inherited from the (AEJ/Gn × *M. spretus*)F_1 parent. Solid squares represent the AEJ/Gn allele; open squares represent the *M. spretus* allele. The number of N_2 progeny carrying each haplotype is listed at the bottom. (*B*) Genetic linkage map showing the location of *All-1* (circled). The left chromosome shows the loci typed in the backcross (Table 1), with distances between loci given in cM. The right chromosome shows a partial version of the consensus linkage map of mouse chr 9 (32). The locus in boldface type and the dashed line shows the reference locus used to align the maps. Loci mapped in humans are underlined; gene locations on human chromosomes are shown in the middle.

sternebrae in homozygotes (15, 40, 41). Based on phenotype and the sequence relationship between *All-1* and the *Drosophila trx* gene, *All-1* is a reasonable candidate gene for the *lu* mutation. The *Drosophila trx* gene is known be involved in maintenance and possibly activation of the expression of homeobox genes in the Antennapedia and Bithorax complexes; mutations in the *trx* gene can result in homeotic transformations (42–44). The abnormal expression of murine *Hox* genes (20) and the *lu* mutation result in similar skeletal defects in mice. Thus, these observations suggest that the *lu* mutation may result from alterations in *All-1* expression.

The predicted amino acid sequence of the ALL-1 protein suggests that ALL-1 could represent a multifunctional protein, whose motifs suggest that it plays a role in transcriptional regulation during development and that juxtaposition of several of these domains in other contexts results in the neoplastic transformation of hematopoietic cells. In addition, the presence of alternatively spliced forms indicates that different All-1 proteins may influence the expression of distinct sets of genes. Further exploration of the end products of the *All-1* locus will provide insight into the multiple roles this gene plays in mammalian development and tumorigenesis. Future studies will lead to an understanding of the *ALL-1* gene and provide mouse model systems useful for testing therapeutic approaches for acute leukemias in humans.

We thank A. M. Mazo and L. Mizrokhi for valuable discussions and criticisms of the manuscript and T. Bestor and M. Darby for valuable discussions. This research was supported in part by National Cancer Institute Grant CA58586 (awarded to A.M.B.) and National Cancer Institute Outstanding Investigator Grant CA39860 (awarded to C.M.C.). L.D.S. was the recipient of an American Cancer Society Junior Faculty Research Award.

1. Croce, C. M. (1991) in *Origins of Human Cancer: A Comprehensive Review* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 527–542.
2. Heim, S. & Mitelman, F. (1987) *Cancer Cytogenetics* (Liss, New York).
3. Cleary, M. L. (1991) *Cell* **66**, 619–622.
4. Savage, P. D., Jones, K. C., Silver, J., Geurts Von Kessel, A. H. M., Gonzales-Sarmiento, R., Palm, L., Hanson, C. A. & Kersey, J. H. (1988) *Cytogenet. Cell Genet.* **49**, 289–292.
5. Yunis, J. J., Jones, C., Madden, M. T., Lu, D. & Mayer, M. G. (1989) *Genomics* **5**, 84–90.
6. Wei, S., Rocchi, M., Archidiacon, N., Sacchi, N., Romeo, G. & Gatti, R. A. (1990) *Cancer Genet. Cytogenet.* **46**, 1–8.
7. Rowley, J. D., Diaz, M. O., Espinosa, R., III, Patel, Y. D., van Melle, E., Ziemin, S., Taillon-Miller, P., Lichter, P., Evans, G. A., Kersey, J. H., Ward, D. C., Domer, P. H. & Le Beau, M. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9358–9362.
8. Cimino, G., Moir, D. T., Canaani, O., Williams, K., Crist, W. M., Katzav, S., Cannizzaro, L., Lange, B., Nowell, P. C., Croce, C. M. & Canaani, E. (1991) *Cancer Res.* **51**, 6712–6714.
9. Cimino, G., Nakamura, T., Gu, Y., Canaani, O., Prasad, R., Crist, W. M., Carrol, A. J., Baer, M., Bloomfield, C. D., Nowell, P. C., Croce, C. M. & Canaani, E. (1992) *Cancer Res.* **52**, 3811–3813.
10. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M. & Canaani, E. (1992) *Cell* **71**, 701–708.
11. Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71**, 691–700.
12. Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nat. Genet.* **2**, 113–118.
13. McCabe, N. R., Burnett, R. C., Gill, H. J., Thirman, M. J., Mbangkollo, D., Kipiniak, M., van Melle, E., Ziemin-van der Poel, S., Rowley, J. D. & Diaz, M. O. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11794–11798.
14. Mazo, A. M., Huang, D.-H., Mozer, B. A. & Dawid, I. B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2112–2116.
15. Green, M. C. (1955) *J. Hered.* **46**, 91–99.
16. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1991) *Current Protocols in Molecular Biology* (Wiley, New York).
17. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
18. Marini, J. C., Nelson, K. K., Battey, J. & Siracusa, L. D. (1993) *Genomics* **15**, 200–202.
19. Siracusa, L. D., Buchberg, A. M., Copeland, N. G. & Jenkins, N. A. (1989) *Genetics* **122**, 669–679.
20. LeMouellic, H., Lallemand, Y. & Brulet, P. (1992) *Cell* **69**, 251–264.
21. Wahle, E. & Keller, W. (1992) *Annu. Rev. Biochem.* **61**, 419–440.
22. Shaw, G. & Kamen, R. (1986) *Cell* **46**, 659–667.
23. Reeves, R. & Nissen, M. S. (1990) *J. Biol. Chem.* **265**, 8573–8582.
24. Reeves, R., Langan, T. A. & Nissen, M. S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1671–1675.
25. Bestor, T. H. (1990) *Philos. Trans. R. Soc. London* **326**, 179–187.
26. Bestor, T. H., Laudano, A., Mattaliano, R. & Ingram, V. (1988) *J. Mol. Biol.* **203**, 971–983.
27. Yen, R.-W. C., Vertino, P. M., Nelkin, B. D., Yu, J. J., El-Deiry, W., Cumaraswamy, A., Lennon, G. G., Trask, B. J., Celano, P. & Baylin, S. B. (1992) *Nucleic Acids Res.* **20**, 2287–2291.
28. Aebi, M. & Weissmann, C. (1987) *Trends Genet.* **3**, 102–107.
29. Mercer, J. A., Seperack, P. K., Strobel, M. C., Copeland, N. G. & Jenkins, N. A. (1991) *Nature (London)* **349**, 709–713.
30. Siracusa, L. D., Jenkins, N. A. & Copeland, N. G. (1991) *Genetics* **127**, 169–179.
31. Dietrich, W., Katz, H., Lincoln, S. E., Shin, H.-S., Friedman, J., Dracopoli, N. C. & Lander, E. S. (1992) *Genetics* **131**, 423–447.
32. Kingsley, D. M. (1992) *Mamm. Genome* **3**, S136–S152.
33. Bestor, T. H. (1992) *EMBO J.* **11**, 2611–2617.
34. Haber, D. A., Sohn, R. L., Buckler, A. J., Pelletier, J., Call, K. M. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9618–9622.
35. Gessler, M., Konig, A. & Bruns, G. A. P. (1992) *Genomics* **12**, 807–813.
36. Rauscher, F. J., III, Morris, J. F., Tournay, O. E., Cook, D. M. & Curran, T. (1990) *Science* **250**, 1259–1262.
37. Bickmore, W. A., Oghene, K., Little, M. H., Seawright, A., van Heyningen, V. & Hastie, N. D. (1992) *Science* **257**, 235–237.
38. Ziemin-van der Poel, S., McCabe, N. R., Gill, H. J., Espinosa, R., III, Patel, Y., Harden, A., Rubinelli, P., Smith, S. D., Le Beau, M. M., Rowley, J. D. & Diaz, M. O. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10735–10739.
39. Sweet, H. O., Oda, S.-I., Taylor, B. A., Rowe, L., Davisson, M. T., Cook, S., Spencer, C. & Sundberg, J. (1990) *Mouse Genome* **86**, 236–237.
40. Forsthoefel, P. (1958) *J. Morphol.* **102**, 247–287.
41. Forsthoefel, P. (1959) *J. Morphol.* **104**, 89–141.
42. Ingham, P. W. (1983) *Nature (London)* **306**, 591–593.
43. Ingham, P. W. (1985) *J. Embryol. Exp. Morphol.* **89**, 349–365.
44. Capdevila, M. P. & Garcia-Bellido, A. (1981) *Roux's Arch. Dev. Biol.* **190**, 339–350.