Molecular Characterization of the Mouse *In(10)17Rk* Inversion and Identification of a Novel Muscle-Specific Gene at the Proximal Breakpoint

Kathleen F. Benson[1] and Kiran Chada

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Manuscript received July 19, 2001
Accepted for publication October 8, 2001

ABSTRACT

Chromosomal rearrangements provide an important resource for molecular characterization of mutations in the mouse. *In(10)17Rk* mice contain a paracentric inversion of ~50 Mb on chromosome 10. Homozygous *In(10)17Rk* mice exhibit a *pygmy* phenotype, suggesting that the distal inversion breakpoint is within the *pygmy* locus. The *pygmy* mutation, originally isolated in 1944, is an autosomal recessive trait causing a dwarf phenotype in homozygous mice and has been mapped to the distal region of chromosome 10. The *pygmy* phenotype has subsequently been shown to result from disruption of the *Hmgi-c* gene. To identify the *In(10)17Rk* distal inversion breakpoint, *In(10)17Rk* DNA was subjected to RFLP analysis with single copy sequences derived from the wild-type *pygmy* locus. This analysis localized the *In(10)17Rk* distal inversion breakpoint to intron 3 of *Hmgi-c* and further study determined that a fusion transcript between novel 5′ sequence and exons 4 and 5 of *Hmgi-c* is created. We employed 5′ RACE to isolate the 5′ end of the fusion transcript and this sequence was localized to the proximal end of chromosome 10 between markers *Cni-rs2* and *Mtap7*. Northern blot analysis of individual tissues of wild-type mice determined that the gene at the *In(10)17Rk* proximal inversion breakpoint is a novel muscle-specific gene and its disruption does not lead to a readily observable phenotype.

THE *pygmy* (*pg*) locus on mouse chromosome 10 is an important regulator of body size. *Pygmy* mice originally appeared in 1944 as undersized segregants in a strain of mice selected for small size (MacArthur 1944). Previously in our laboratory, a transgenic insertional mutant that was shown to be allelic to the spontaneous mutant *pygmy* was isolated (Xiang *et al.* 1990). The founder transgenic mouse harbored two different transgene integration events at the *pygmy* locus, which were segregated in the progeny, forming two distinct lines, designated $pg^{TgN40ACha}$ (A) and $pg^{TgN40BCha}$ (B). Subsequent characterization of the A, B, and *pg* mutations determined that sizable deletions had occurred in each of these mutants (Xiang *et al.* 1990; Zhou *et al.* 1995; K. F. Benson and K. Chada, unpublished results).

The creation of new mutations at the *pygmy* locus by transgene insertion resulted in the identification of the *Hmgi-c* gene at the *pygmy* locus and a targeted disruption demonstrated that its loss of expression was the cause of the *pygmy* phenotype (Zhou *et al.* 1995). Translocations involving human *Hmgi-c* that create fusion transcripts composed of exons 1 through 3 of *Hmgi-c* and novel 3′ partners have been shown to cause tumors of mesenchymal origin (Ashar *et al.* 1995), which is consistent with the wild-type expression pattern of murine *Hmgi-c* (Zhou *et al.* 1995, 1996). While *Hmgi-c* is a developmentally regulated gene with highest levels of expression in the embryo (Zhou *et al.* 1995, 1996), recent studies have also demonstrated the importance of *Hmgi-c* expression for expansion of adipocytes in white adipose tissue of the adult mouse (Anand and Chada 2000).

Another line of mice with a *pygmy* phenotype arose as a consequence of exposure to the chemical mutagen triethylenemelamine during a large scale mutagenesis experiment conducted at Jackson Laboratory. *In(10)17Rk* mice contain a paracentric inversion of approximately 50 Mb within mouse chromosome 10 between proximal region A4 and distal region D2 (Roderick 1983). Homozygous *In(10)17Rk* mice exhibit a *pygmy* phenotype and the *pygmy* locus maps to the D2 region of mouse chromosome 10 (Lyon and Searle 1989), suggesting that the distal inversion breakpoint may be within the *pygmy* locus. Ongoing studies in our laboratory to identify the *pygmy* gene by positional cloning methods provided an array of probes along the *pygmy* locus for investigating the precise location of the *In(10)17Rk* distal inversion breakpoint. This study describes the molecular characterization of both the proximal and distal *In(10)17Rk* inversion breakpoints and the consequent identification of a novel muscle-specific gene disrupted by the proximal breakpoint.

[1]*Corresponding author:* Markey Molecular Medicine Center, Division of Medical Genetics, School of Medicine, University of Washington, Box 357720, 1705 NE Pacific St., Seattle, WA 98195.
E-mail: kfbenson@u.washington.edu

MATERIALS AND METHODS

Genotyping: The genotypes were established for mice in line $pg^{TgN40ACha}$ and the spontaneous mutant *pg* as previously described (Xiang *et al.* 1990), and mice containing the *In(10)17Rk* inversion were detected by a PCR-based restriction fragment length polymorphism as described in Cherath *et al.* (1995).

DNA preparation and blot hybridization: High-molecular-weight genomic DNA was isolated from mouse livers using standard procedures (Ausubel *et al.* 1988). Genomic DNA (10 μg) of each genotype was digested with various restriction enzymes, electrophoresed in 0.8% agarose gels, and transferred to Duralon (Stratagene, La Jolla, CA). DNA hybridizations were performed as described in Ashar *et al.* (1994).

Determination of the 5′ end of the fusion transcript: We performed 5′ rapid amplification of cDNA ends (RACE) on RNA from homozygous *In(10)17Rk* embryos using antisense primer 4862, (5′-ATGGATCCCTAATCCTCCTCTGC-3′) from the beginning of exon 5 of *Hmgi-c* for first-strand cDNA synthesis. For reverse transcription, 20 pmol antisense oligonucleotides were used to prime cDNA synthesis in a 20-μl reaction containing 1 μg of total RNA from 12.5-days post-coitum (dpc) homozygous *In(10)17Rk* embryos. First-strand cDNAs were A-tailed and then used for subsequent PCR in a 50-μl reaction containing 1× PCR buffer (0.05 M KCl, 0.01 M Tris-HCl, pH 8.3, 0.015 M $MgCl_2$, 0.001% gelatin), 50 pmol sense oligo(T) primer 4993, (5′-GCAATACGACTCACTATAGTTTTTTTT TTTTT-3′), 50 pmol antisense primer 4862, 40 nmol dNTP mixture (Pharmacia, Piscataway, NJ), 10% DMSO, and 2.5 units Taq DNA polymerase (GIBCO-BRL, Bethesda, MD). After initial denaturation at 94° for 5 min, the cycling conditions were as follows: 94° for 30 sec, 55° for 1 min and 72° for 1 min for 30 cycles. The PCR product was cloned in pBluescript KS+ (Stratagene) and double-strand sequencing was performed with the Sequenase Version 2.0 sequencing kit (USB Corporation, Cleveland).

Chromosome mapping: Chromosome mapping was performed using the Jackson Laboratory BSS backcross panel with progeny from matings of (C57BL/6JEi × SPRET/Ei)F1 × SPRET/Ei mice (Rowe *et al.* 1994). A 219-bp probe was amplified from the 5′ RACE product by designing primers to the novel sequence. These are sense primer 6905 (5′-TCCTGAT CAGCTCTCAGTGG-3′) and antisense primer 6876 (5′-CTG GCAGCAGACCCATACG-3′) corresponding to nucleotides 171–190 and 371–389 of the *In(10)17Rk-p* cDNA, respectively. PCR conditions were as described for 5′ RACE. Southern blot analysis of *Hin*dIII-digested C57BL/6JEi and SPRET/Ei genomic DNAs with the 219-bp sequence detected single 3.2- and 5.0-kb bands, respectively. DNAs of the progeny were digested with *Hin*dIII and analyzed by Southern blot, using the 219-bp probe to determine the presence of C57BL/6JEi and/or SPRET/Ei alleles in each sample. The data were analyzed by the Jackson Laboratory.

Northern hybridization: Total RNA from whole embryos and individual newborn and adult mouse tissues was isolated by the CsCl-guanidinium thiocyanate method (Chirgwin *et al.* 1979). Of each RNA sample, 5 μg was resolved on a 1.2% agarose formaldehyde-denaturing gel, transferred to a nylon membrane (Duralon, Stratagene), immobilized by UV cross-linking, and hybridized with ^{32}P-labeled probes using standard procedures. The filters were rehybridized with a ^{32}P-labeled oligonucleotide complementary to 28S rRNA (Barbu and Dautry 1989) to ensure that equal amounts of RNA were present in each lane.

RT-PCR: A total of 20 pmol of antisense primer 4862, corresponding to the beginning of exon 5 of murine *Hmgi-c* was used to prime cDNA synthesis of 1 μg of RNA from 12.5-dpc embryos of the genotypes C57BL/6J, hemizygous *In(10)17Rk*, and homozygous *In(10)17Rk*. PCR was performed on first-strand cDNAs as described above with antisense primer 4862 and sense primer 6809, (5′-GGATGTTGACTATGAAGAAACC-3′), corresponding to nucleotides 456–477 of the *In(10)17Rk-p* cDNA.

Isolation of full-length transcripts: We performed 3′ RACE on C57BL/6J adult heart RNA with the Gene Racer kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. The gene-specific primers derived from the novel sequence of the fusion transcript (Figure 5A) are sense primer Rod 5′-3 (5′-CTCCCTCCTTGTAAGAACTCAGC-3′; bp 15–37) and nested sense primer 6826 (5′-TGTAAGAACTCAGCCC TGCTGC-3′; bp 24–45). PCR products were cloned into the pGEM-T Easy vector system (Promega, Madison, WI) and double-strand sequencing was performed with the Applied Biosystems (Foster City, CA) PRISM Big Dye terminator chemistry on an Applied Biosystems 310 machine. The full-length 2.5-kb cDNA sequence (accession no. AF422244) and the full-length 4.4-kb cDNA sequence (accession no. AF422245) have been deposited in the National Center for Biotechnology Information (NCBI) GenBank database.

TABLE 1

Manifestation of allelism between *In(10)17Rk* and *pygmy* as expressed by the production of small progeny

Cross	No. of mice	
	Normal	Small
Rk/+ × *pg*/+	14	3
Rk/+ × *A*/+	18	4

Mating was established between hemizygous *In(10)17Rk* (*Rk*/+) mice and mice heterozygous for either the spontaneous *pygmy* mutation (*pg*/+) or $pg^{TgN40ACha}$ transgene insertion (A/+).

RESULTS

***In(10)17Rk* represents a fourth *pg* allele:** *In(10)17Rk* mice contain a paracentric inversion of ~50 Mb within mouse chromosome 10, localized between proximal region A4 and distal region D2 (Roderick 1983). Although it has been documented that homozygous *In(10)17Rk* mice exhibit a *pygmy* phenotype (Roderick 1983), no allelism studies with *pygmy* mice have been published. Therefore matings were established between hemizygous *In(10)17Rk* mice (Rk/+; mice with a *pygmy* phenotype are infertile) and either heterozygous spontaneous *pygmy* (*pg*/+) or heterozygous transgenic insertional mutant A ($pg^{TgN40ACha}$) mice (A/+). Of 17 mice born to *In(10)17Rk*/+ × *pg*/+ matings, 3 were small and of the 22 progeny resulting from *In(10)17Rk*/+ × A/+ matings, 4 exhibited the *pygmy* phenotype (Table 1). All small mice resulting from these matings were subsequently shown to be hemizygous for the *In(10)17Rk* inversion as well as heterozygous for either *pg* or $pg^{TgN40ACha}$, respectively (data not shown).

Restriction fragment length polymorphism analysis localizes *In(10)17Rk-d* to intron 3 of *Hmgi-c*: Since inver-

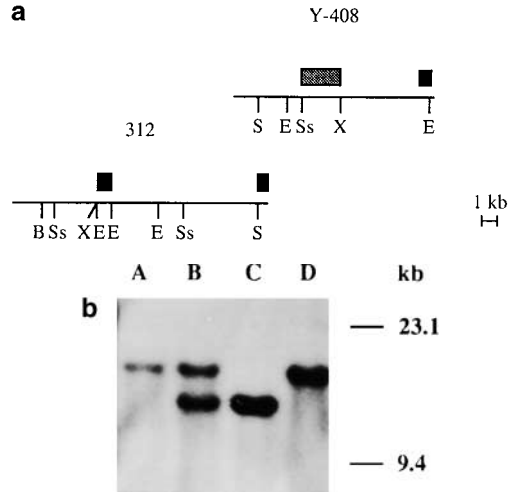

FIGURE 1.—(a) Restriction enzyme maps of overlapping λ-clones 312 and Y-408 from the *pg* locus. Single copy sequences (represented by solid boxes) were used for RFLP analysis of homozygous *In(10)17Rk* DNA. The shaded box indicates a 2.5-kb fragment that identified a polymorphic *Xba*I band in the *In(10)17Rk* allele. B, *Bam*HI; E, *Eco*RI; S, *Sal*I; Ss, *Sst*I; X, *Xba*I. (b) Southern blot of C57BL/6J (lane A), hemizygous *In(10)17Rk* (lane B), homozygous *In(10)17Rk* (lane C), or DBA/2J (lane D) adult genomic DNA digested with *Xba*I and hybridized with a 2.5-kb *Sst*I/*Xba*I fragment derived from λ-clone Y-408. Molecular weight standards derived from *Hin*dIII digestion of λDNA (GIBCO-BRL) are indicated on the right.

sions predominantly result in retention of all genetic material (BULTMAN *et al.* 1992; QUADERI *et al.* 1997; PERRY *et al.* 1998), breakpoints can be identified only by probes that hybridize to restriction fragments bridging the inversion breakpoint. It was reasoned that if the distal inversion breakpoint was within the *pygmy* locus, it would be identified as a restriction fragment length polymorphism (RFLP) between the wild-type and *In(10)17Rk* alleles. Single copy sequences were isolated by restriction enzyme analysis of λ-clones derived from a yeast artificial chromosome clone containing the *pygmy* locus (ZHOU *et al.* 1995) and hybridized to Southern blots of DNA from C57BL/6J, hemizygous *In(10)17Rk*, and homozygous *In(10)17Rk* mice digested with *Eco*RI and *Bam*HI. A large (>30 kb) *Bam*HI genomic fragment migrated faster in *In(10)17Rk* DNA (data not shown) and was investigated further. This region of the pygmy locus is encompassed by λ-clones Y-408 and 312 (Figure 1a). Further analysis of these clones identified a 2.5-kb *Sst*I/*Xba*I single copy fragment that, when used as a probe, detected a polymorphism for the restriction enzyme *Xba*I (Figure 1b). DNA from the mouse strain DBA/2J was also included in the analysis since this is the original strain on which the *In(10)17Rk* inversion was induced (RODERICK 1983). Southern blot analysis with the restriction enzymes *Xba*I, *Sst*I, and *Eco*RI and single copy sequences from λ-clones 312 and Y-408 (Figure 1a) localized the *In(10)17Rk* distal inversion breakpoint to a 6.5-kb *Eco*RI/*Sal*I fragment present in clone 312. This 6.5-kb fragment was cloned into pBluescript KS+ (Stratagene) and further analyzed by restriction enzyme digestion (Figure 2A).

A comparison between the Southern blot results (including the additional enzyme *Pvu*II) and the restriction enzyme sites within the 6.5-kb sequence determined that the *In(10)17Rk* distal inversion breakpoint was located within a 0.6-kb *Pvu*II/*Sst*I fragment (Figure 2A). Southern blots were then prepared from DNA of C57BL/6J, hemizygous *In(10)17Rk*, homozygous *In(10)17Rk*, and DBA/2J mice digested with the restriction enzymes *Hin*dIII, *Pst*I, *Eco*RI, and *Xba*I (Figure 2B) as well as the additional enzymes *Bam*HI, *Hin*cII, *Kpn*I, and *Pvu*II (data not shown) and hybridized with the 0.6-kb *Pvu*II/*Sst*I sequence. For all enzymes analyzed, the homozygous *In(10)17Rk* DNA produced a novel hybridization pattern consisting of two unique bands that differed from the C57BL/6J and DBA/2J lanes, while hemizygous *In(10)17Rk* DNA produced a compound hybridization pattern reflecting the presence of both a wild-type and *In(10)17Rk* inversion allele (Figure 2B). This places the distal inversion breakpoint within intron 3 of *Hmgi-c* (Figure 3).

The *In(10)17Rk* inversion disrupts *Hmgi-c* expression and creates a fusion transcript: *Hmgi-c* is encoded by five exons that span ~110 kb (ZHOU *et al.* 1996). To investigate the status of *Hmgi-c* transcripts within *In(10)17Rk* mice, Northern blots were prepared from total RNA of 12.5-dpc embryos of the genotypes C57BL/6J, hemizygous *In(10)17Rk*, and homozygous *In(10)17Rk* and hybridized with *Hmgi-c* exonic sequences that were either 5′ (exons 2 and 3) or 3′ (exon 5) to the distal inversion breakpoint (Figure 3). The 5′ probe failed to detect any transcripts in homozygous *In(10)17Rk* embryos (Figure 4). The identification of a transcript in homozygous *In(10)17Rk* mice by the exon 5 probe, which is similar in size to *Hmgi-c* although at greatly reduced levels (Figure 4), was observed only upon overexposure of the Northern blot. On the basis of the size of this transcript (3.7 kb) and the positioning of the *In(10)17Rk* distal inversion breakpoint within intron 3 of *Hmgi-c*, we postulated that the RNA species detected was a fusion transcript consisting of exons 4 and 5 of *Hmgi-c* (3.1 kb) fused to an ~600-bp novel 5′ sequence.

5′ RACE isolates novel sequence from fusion transcript: Identification of a transcript from the *In(10)17Rk* allele provided an opportunity to isolate the 5′ sequence of this fusion transcript that is not derived from *Hmgi-c*. Performing 5′ RACE on RNA from homozygous *In(10)17Rk* embryos using a specific primer from exon 5 of *Hmgi-c* resulted in the isolation of a 670-bp product. When this product was cloned and sequenced, it demonstrated a novel 586-bp 5′ sequence followed by the 33-bp exon 4 of *Hmgi-c* spliced to the beginning of exon 5 of *Hmgi-c* (Figure 5A).

RT-PCR was performed to confirm the origin of the

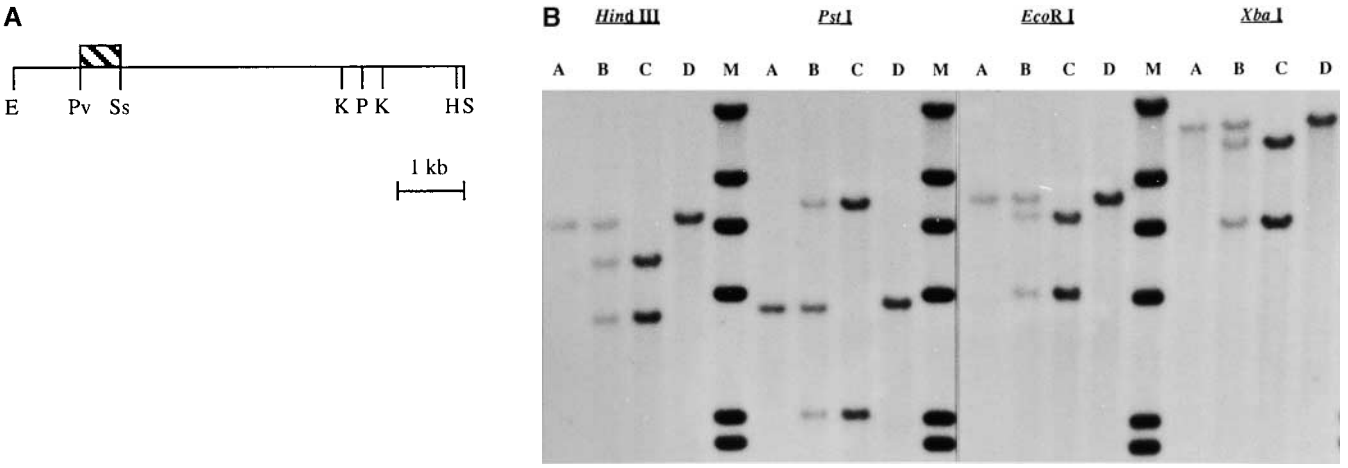

FIGURE 2.—(A) Detailed restriction enzyme map of a 6.5-kb *Eco*RI/*Sal*I fragment derived from λ-clone 312. The striped box indicates a 0.6-kb single copy sequence that spans the *In(10)17Rk* distal inversion breakpoint. E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; Ss, *Sst*I. (B) Southern blot of C57BL/6J (lanes A), hemizygous *In(10)17Rk* (lanes B), homozygous *In(10)17Rk* (lanes C) or DBA/2J (lanes D) adult genomic DNA digested with *Hin*dIII, *Pst*I, *Eco*RI or *Xba*I. Each sample was hybridized with a 0.6-kb *Pvu*II/*Sst*I sequence derived from a 6.5-kb *Eco*RI/*Sal*I fragment from λ-clone 312. M, λ *Hin*dIII molecular weight marker (GIBCO-BRL).

fusion transcript from the *In(10)17Rk* allele. Using the combination of a primer from exon 5 of *Hmgi-c* and a primer derived from the novel sequence of the 5′ RACE product, a predicted product of 216 bp was amplified upon RT-PCR of hemizygous and homozygous *In(10)17Rk* RNA samples (Figure 5B). RT-PCR of wild-type RNA did not produce an amplification product.

Chromosomal localization of *In(10)17Rk-p*: A 219-bp sequence was PCR amplified from the novel 5′ sequence of the fusion transcript (see MATERIALS AND METHODS). Southern blot analysis of the parental mouse strains C57BL/6JEi and SPRET/Ei with the restriction enzyme *Hin*dIII identified fragments of 3.2 and 5.0 kb, respectively. This polymorphism was utilized for analysis of 94 progeny from the Jackson BSS backcross mapping panel (ROWE *et al.* 1994). The data place *In(10)17Rk-p* in the proximal end of chromosome 10 between the markers *Cni-rs2* and *Mtap7* (Figure 6A).

Tissue distribution of *In(10)17Rk-p* transcripts: The 219-bp sequence derived from the *In(10)17Rk-p* fusion transcript was used for Northern blot analysis of wild-type tissues. This identified transcripts of 2.5 and 4.4 kb in adult heart, femur muscle, and diaphragm (Figure 7A). Transcripts were not detected in skin, white adipose tissue, spleen, thymus, or kidney (Figure 7A) or in the additional 10 adult tissues of liver, lung, adrenal, brain, pituitary, stomach, small and large intestine, pancreas, and testes (data not shown). When 11 newborn tissues (brain, heart, lung, liver, kidney, stomach, small and large intestine, spleen, pancreas, and skin) were analyzed, only heart exhibited expression, at levels lower than those of adult (data not shown).

Next the presence of transcripts was assessed in heart and femur muscle of adult hemizygous and homozygous *In(10)17Rk* mice. Compared to C57BL/6J, expression levels are reduced in hemizygous *In(10)17Rk* muscle tissue and completely absent in homozygous mice (Figure 7B). Hemizygous *In(10)17Rk* RNA contains both wild-type and fusion transcripts (lanes 3–4) while only the fusion transcript is detected in homozygous *In(10)17Rk* RNA (lanes 5–6).

Isolation of the full-length 2.5- and 4.4-kb transcripts: Gene-specific primers were designed from the novel sequence of the fusion transcript and used for 3′ RACE of adult C57BL/6J heart RNA (see MATERIALS AND METHODS). This resulted in the isolation of both the

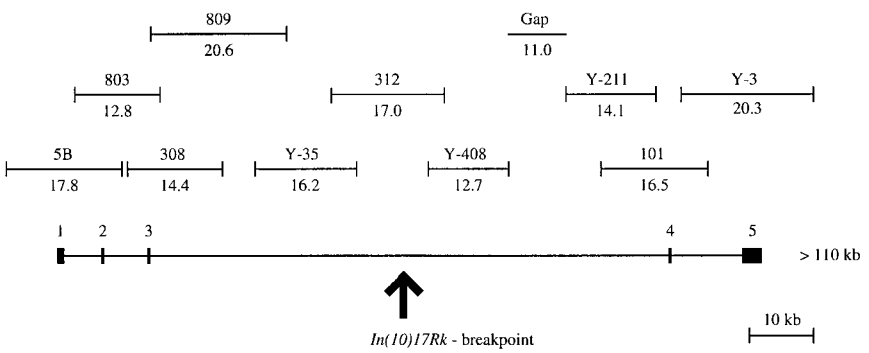

FIGURE 3.—λ-clones spanning the murine *Hmgi-c* locus. Gap refers to an 11-kb unclonable region of the locus (K. PRZYBYSZ, K. F. BENSON and K. CHADA, unpublished data). Exons are represented as solid boxes and are not drawn to scale. An arrow indicates the *In(10)17Rk* distal breakpoint within intron 3.

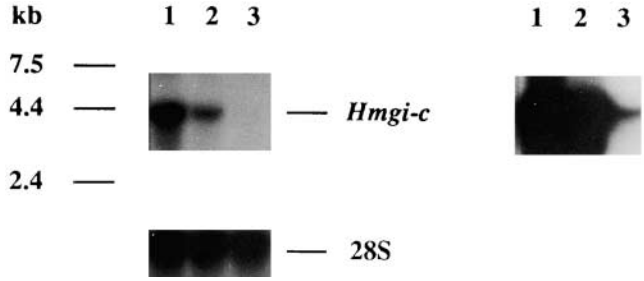

FIGURE 4.—Northern blot of 5 μg total RNA from C57BL/6J (lane 1), hemizygous *In(10)17Rk* (lane 2), or homozygous *In(10)17Rk* (lane 3) 12.5-dpc embryos probed with either exons 2 and 3 (left) or exon 5 (right) of murine *Hmgi-c*. The blots were exposed to film at –70° for 4 and 48 hr, respectively, and reprobed with an oligonucleotide complementary to murine 28S ribosomal RNA (BARBU and DAUTRY 1989) to ensure equal loading of samples.

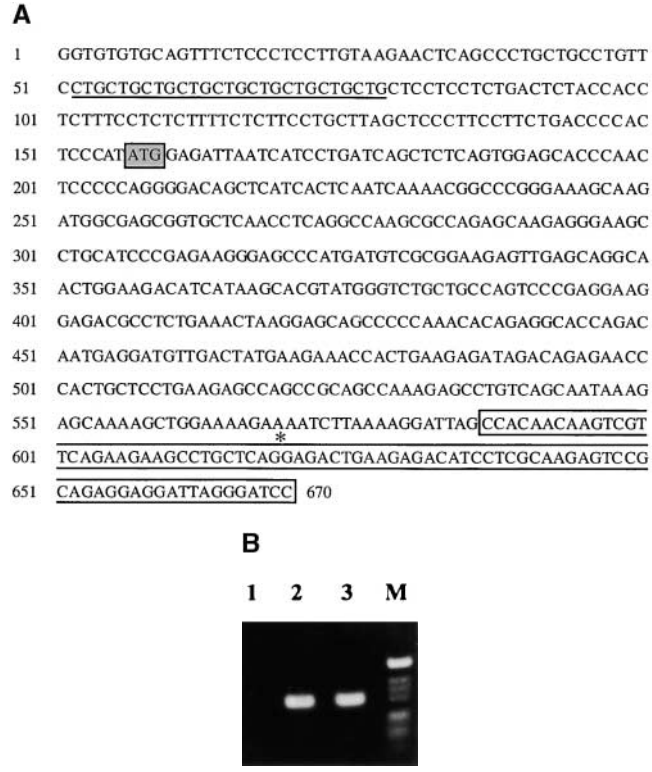

FIGURE 5.—(A) Partial cDNA sequence of the 5′ RACE product showing the 5′ end of the fusion transcript. The boxed region defines the beginning of the sequence derived from murine *Hmgi-c* while an asterisk indicates the splice junction between exons 4 and 5 of murine *Hmgi-c*. (B) RT-PCR product derived from the fusion transcript. One microgram of total RNA from 12.5-dpc embryos of the genotypes C57BL/6J (lane 1), hemizygous *In(10)17Rk* (lane 2), or homozygous *In(10)17Rk* (lane 3) was reverse transcribed using a primer from exon 5 of *Hmgi-c* as described in MATERIALS AND METHODS. PCR was performed with a primer from exon 5 of *Hmgi-c* as well as a primer designed from the novel sequence of the fusion transcript. M, 1-kb ladder molecular weight marker (GIBCO-BRL).

2.5- and 4.4-kb full-length cDNAs. The 2.5-kb cDNA is 2435 bp in length (GenBank accession no. AF22244) and the 4.4-kb cDNA is 4394 bp in length (accession no. AF22245). Both RNAs encode a 684-amino-acid protein with the 4.4-kb RNA containing an additional 1959 bp of 3′ untranslated sequence.

Gene disrupted by *In(10)17Rk-p* is homologous to a chicken muscle-specific gene: BLAST searches of the NCBI databases with the 2.5- and 4.4-kb transcript sequences returned significant homology to a chicken muscle-specific protein called MDP77 (UYEDA *et al.* 2000; GenBank accession no. BAA94755). There is 74% identity and 85% similarity over 463 residues beginning with the initiating methionine (Figure 8). This homology is revealed when the mouse sequences are translated in the +1 frame. There are a few gaps in alignment in the amino-terminal end of the proteins with the chicken MDP77 gene encoding a 676-amino-acid protein and the two mouse transcripts encoding an identical 684-amino-acid protein. The carboxy-terminal one-third of the chicken and mouse proteins completely diverges (data not shown). No other similarities were detected at the protein level.

Currently the mouse genomic sequence from the region of chromosome 10 containing the MDP77 gene is not available in the NCBI sequence databases. However, homology searches performed at the nucleotide level returned matches to the chicken MDP77 gene as well as to separate but consecutive regions of a genomic clone derived from human chromosome 6 (clone RP3-522B19, GenBank accession no. AL158850). This was seen only with the regions of the 2.5- and 4.4-kb cDNAs that are 3′ to the sequence present in the fusion transcript. These regions of punctuated homology between the mouse cDNA sequence and human chromosome 6 are presumed to represent individual exons, indicating that the 2.5-kb mouse RNA is encoded by a minimum of 11 exons (data not shown).

The PCR amplification and sequencing of a 3.5-kb fragment from mouse genomic DNA using primers derived from the 5′ (primer 4826) and 3′ (primer 6876) regions of the novel 586-bp sequence of the fusion transcript (Figure 5A) revealed the presence of a single intron (data not shown). This information determines that the fusion transcript is composed of exons 1 and 2 of mouse MDP77 fused to exons 4 and 5 of *Hmgi-c* and places the *In(10)17Rk-p* inversion breakpoint within intron 2 of the mouse MDP77 gene. Figure 8 indicates the position of the *In(10)17Rk-p* inversion breakpoint within the protein separating the first 143 amino acids (encoded by exons 1 and 2) from the remainder of the protein.

DISCUSSION

Allelism studies confirm *In(10)17Rk* as another *pg* allele: The *pygmy* mutation arose spontaneously in 1944 (MACARTHUR 1944) as an autosomal recessive trait causing a dwarf phenotype and was mapped to mouse chromosome 10

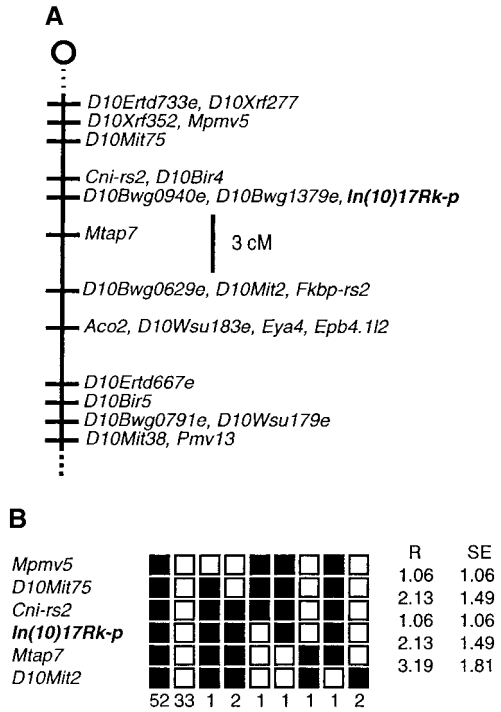

FIGURE 6.—(A) Map figure showing the proximal end of Jackson BSS chromosome 10 with loci linked to *In(10)17Rk-p*. The map is depicted with the centromere toward the top. A 3-cm scale bar is shown to the right. Loci mapping to the same position are listed in alphabetical order. (B) Haplotypes of the region surrounding the *In(10)17Rk-p* inversion breakpoint on chromosome 10 in 94 progeny from the Jackson BSS cross. Solid boxes represent the C57BL6/JEi allele; open boxes represent SPRET/Ei alleles. The number of animals used to derive the data is given at the bottom of each column of boxes. The percentage recombination (R) between adjacent loci is given to the right, with the standard error (SE) for each R. Missing typings were inferred from surrounding data when assignment was unambiguous. Raw data from the Jackson Laboratory were obtained from http://www.jax.org/resources/documents/cmdata.

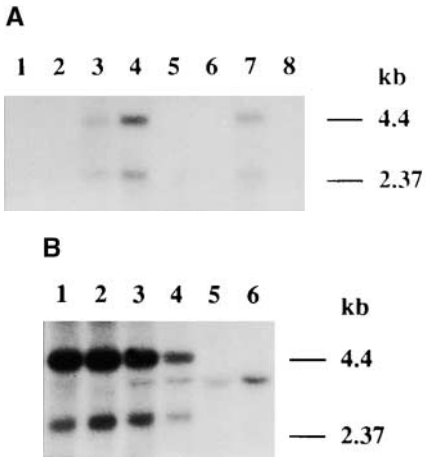

FIGURE 7.—(A) Northern blot of eight wild-type adult mouse tissues hybridized with a 219-bp sequence derived from the 5′ end of the *In(10)17Rk-p* fusion transcript (see MATERIALS AND METHODS). Lanes 1–8: skin, white adipose tissue, diaphragm, femur muscle, spleen, thymus, heart, and kidney. (B) Northern blot of total RNA isolated from heart and femur muscle of adult C57BL/6J (lanes 1–2), hemizygous *In(10)17Rk* (lanes 3–4), and homozygous *In(10)17Rk* (lanes 5–6) mice hybridized with the *In(10)17Rk-p* 219-bp probe.

(LYON and SEARLE 1989); however, a molecular probe was not available to identify the gene whose disruption gave rise to the phenotype. Remutation of the *pygmy* locus by transgene integration (XIANG *et al.* 1990) facilitated cloning of the locus and the subsequent identification of the *pygmy* gene as *Hmgi-c* (ZHOU *et al.* 1995). During the molecular analysis of the transgenic insertional mutant we initiated studies of *In(10)17Rk* mice that contain an ~50-Mb inversion in chromosome 10 and also exhibit a *pygmy* phenotype (RODERICK 1983). The allelism studies confirm the hypothesis that *In(10)17Rk* represents another allele of *pygmy*. Both *Rk*/+ × *pg*/+ and *Rk*/+ × A/+ matings produced small progeny in proportions consistent with an autosomal recessive trait and all small mice were shown to be compound heterozygous for the *In(10)17Rk* inversion and either the *pg* or A mutations, respectively.

***In(10)17Rk-d* localized to intron 3 of *Hmgi-c*:** Next a detailed RFLP analysis of the *pygmy* locus was undertaken to identify the *In(10)17Rk* distal inversion breakpoint. The identification of an *Xba*I polymorphism by Southern analysis with a single copy sequence from λ-clone 312 led to the final localization of the *In(10)17Rk-d* inversion breakpoint to a 0.6-kb *Pvu*II/*Sst*I fragment within intron 3 of *Hmgi-c*. The relatively minimal loss of genetic material found in the analysis of some inversions is well illustrated by rearrangements occurring at the agouti (*a*) locus on mouse chromosome 2. Characterization of the distal inversion breakpoint of the *Is1Gso* mutant detected a deletion of only 29 bp (BULTMAN *et al.* 1992), while characterization of the proximal and distal inversion breakpoints in the *a18H* mouse identified deletions of only 18 and 20 bp, respectively (PERRY *et al.* 1998). Since we found large deletions in the transgenic insertional and spontaneous *pygmy* mutants (XIANG *et al.* 1990; ZHOU *et al.* 1995; K. F. BENSON and K. CHADA, unpublished results), analysis of the *In(10)17Rk* inversion has proven crucial for narrowing down the location of the *pygmy* gene.

The localization of the *In(10)17Rk* distal inversion breakpoint to intron 3 of *Hmgi-c* provides a molecular explanation for the disruption of *Hmgi-c* expression in *In(10)17Rk* mice resulting in the *pygmy* phenotype. Interestingly, intron 3 of human *Hmgi-c* is the most common site of chromosomal rearrangement in human tumors (KAZMIERCZAK *et al.* 1998), suggesting that intron 3 of *Hmgi-c* may contain some sequence element such as a low copy repeat (SHAFFER and LUPSKI 2000) that confers susceptibility to rearrangement. Evidence that this element could be conserved in intron 3 of mouse *Hmgi-c* is the relative absence of repetitive sequences in this region (K. F. BENSON and K. CHADA, unpublished results).

FIGURE 8.—Alignment of homologous segments of the predicted protein sequence of the mouse *In(10)17Rk-p* muscle-specific gene and chicken MDP77. Amino acid identity is indicated by reiteration of the one-letter amino acid convention on the middle line of each row while amino acid similarity is denoted by +. Gaps are represented by dashes. The mouse sequence has an additional six amino acids in the N terminus compared with MDP77. The position at which the *In(10)17Rk-p* inversion breakpoint disrupts the mouse protein is indicated by an asterisk. A solid bar demarcates the beginning of a leucine zipper motif in MDP77 that is conserved in the mouse sequence.

Fusion transcript created by inversion facilitates mapping of *In(10)17Rk-p*: Northern blot analysis of homozygous *In(10)17Rk* embryonic RNA with exon 5 of *Hmgi-c* identified a transcript of ~3.7 kb. Since the *In(10)17Rk* distal inversion breakpoint is within intron 3 of *Hmgi-c*, resulting in the separation of exons 1–3 from 4 and 5, this transcript was hypothesized to consist of novel 5′ sequence and *Hmgi-c* exons 4 and 5. A 586-bp novel sequence was identified by 5′ RACE and further utilized for interspecific backcross mapping studies. The placement of *In(10)17Rk-p* in the proximal end of chromosome 10 between markers *Cni-rs2* and *Mtap7* determines that the *In(10)17Rk* inversion encompasses ~60 cM of chromosome 10.

Northern analysis with exons 2 and 3 of *Hmgi-c* failed to detect transcripts in homozygous *In(10)17Rk* embryos when levels of *Hmgi-c* are normally high. Therefore, it is unlikely that the *In(10)17Rk* inversion creates a reciprocal fusion transcript composed of *Hmgi-c* exons 1–3 and novel 3′ sequences or results in truncated *Hmgi-c* consisting of exons 1–3. This is supported by the findings that both *Hmgi-c* truncations and fusion with novel 3′ partners result in tumors (ASHAR *et al.* 1995; FEDELE *et al.* 1998; BATTISTA *et al.* 1999), which were not seen in *In(10)17Rk* mice (K. F. BENSON and K. CHADA, unpublished observations). Additionally, 3′ RACE experiments performed with a specific primer from exon 3 of *Hmgi-c* were unsuccessful in detecting any *Hmgi-c* containing transcripts from homozygous *In(10)17Rk* embryos (data not shown).

***In(10)17Rk-p* disrupts a muscle-specific gene:** We investigated the wild-type expression pattern of the novel portion of the *In(10)17Rk-p* fusion transcript. Northern blot analysis of 18 adult and 11 newborn tissues with the 219-bp probe detected transcripts of 2.5- and 4.4-kb in adult heart, femur muscle, and diaphragm and newborn heart. All other tissues were negative for expression. The finding that the disrupted gene is homologous to a chicken muscle-specific gene that encodes a 77-kD protein called MDP77 (UYEDA *et al.* 2000) supports our conclusion that the *In(10)17Rk-p* inversion breakpoint

disrupts a muscle-specific gene. MDP77 was isolated from a cDNA library derived from chicken denervated crus muscle and Northern blotting and *in situ* hybridization localized expression to cardiac and skeletal muscle (Uyeda *et al.* 2000). The high level of amino acid conservation and similar expression pattern suggest that the gene residing at the *In(10)17Rk-p* inversion breakpoint is the murine homolog of MDP77.

Expression of murine MDP77 is completely absent in homozygous *In(10)17Rk* mice as determined by Northern blot analysis of adult heart and femur muscle. Nevertheless, homozygous *In(10)17Rk* mice do not exhibit any obvious phenotype in addition to that of *pygmy*. Based upon translation of the complete mouse cDNA sequence and comparison to chicken MDP77, the majority of the protein is deleted, including a putative leucine zipper (see Figure 8). Therefore it is unlikely that the fusion transcript retains function of the muscle-specific gene. Instead, lack of an additional phenotype may reflect gene redundancy as exemplified by the muscle-specific genes dystrophin and utrophin (Grady *et al.* 1997).

There is a $(CTG)_9$ $(CTC)_3$ repeat in the 5′ end of the mouse cDNA that is 69 bp preceding the translational initiation codon (position –105 to –69) and interestingly, there is a $(CTG)_3$ repeat at a similar position in the chicken MDP77 cDNA, position –79 to –70 (GenBank accession no. D89999). Recently a number of CUG repeat-containing RNAs have been identified (Lu *et al.* 1999; Miller *et al.* 2000; Ladd *et al.* 2001). These appear to be expressed predominantly in muscle tissue and the CUG repeats are important for the appropriate processing of these RNAs by RNA-binding proteins (Philips *et al.* 1998; Miller *et al.* 2000; Ladd *et al.* 2001). One such example is the DMPK protein kinase gene that contains a CTG repeat in the 3′ untranslated region, expansion of which leads to the human illness myotonic dystrophy (Brook *et al.* 1992; Fu *et al.* 1992; Mahadevan *et al.* 1992). Evidence suggests that the disease occurs as a consequence of the deregulation of processing of multiple CUG repeat-containing RNAs (Philips *et al.* 1998; Lu *et al.* 1999; Mankodi *et al.* 2000). This is supported by the recent report by Liquori *et al.* (2001), identifying the expansion of a CCTG repeat in the *ZNF9* gene as the genetic cause of myotonic dystrophy type 2 (DM2). The underlying pathogenic mechanism common to these two similar diseases is the production of RNAs containing an expanded CUG repeat (DM1) or CCUG repeat (DM2).

One feature of myotonic dystrophy is cardiomyopathy, which has been suggested to reflect the degree of CTG expansion (Finsterer *et al.* 2001). Based on loci flanking the *In(10)17Rk-p* inversion breakpoint, the human syntenic region is chromosome 6q22-25 (Mouse Genome Database; http://www.informatics.jax.org). Intriguingly, two forms of autosomal dominant dilated cardiomyopathy have been localized here. These are dilated cardiomyopathy CMD1F, mapping to 6q23 [Online Mendelian Inheritance in Man (OMIM) entry 602067; Messina *et al.* 1997], and dilated cardiomyopathy CMD1J, mapping to 6q23-24 (OMIM 605362; Schonberger *et al.* 2000). The detection of homology at the nucleotide level between the mouse MDP77 cDNA sequence and the human chromosome 6-derived genomic clone RP3-522B19 strongly suggests the existence of a human MDP77 gene in this region. Additionally, clone RP3-522B19 has been placed within a larger contiguous sequence of human chromosome 6 (GenBank accession no. NT_025741) and is flanked by some of the same microsatellite markers used to map the CMD1F and CMD1J loci (Messina *et al.* 1997; Schonberger *et al.* 2000), further emphasizing the suitability of human MDP77 as a candidate gene for these diseases.

Characterization of the proximal and distal *In(10)17Rk* inversion breakpoints has provided both an explanation for the *pygmy* phenotype of *In(10)17Rk* mice and identified a novel muscle-specific gene in mouse. A homolog of this gene has been reported in chicken and a human version is likely to be present in the syntenic region of human 6q22-25. The mapping of two forms of autosomal dominant dilated cardiomyopathy to this chromosomal region makes MDP77 a candidate gene for these diseases.

This work was supported by a grant from the National Institutes of Health (CA77929).

LITERATURE CITED

Anand, A., and K. Chada, 2000 *In vivo* modulation of *Hmgic* reduces obesity. Nat. Genet. **24:** 377–380.

Ashar, H. R., K. F. Benson, N. A. Jenkins, D. G. Gilbert, N. G. Copeland *et al.*, 1994 *Ifg*, *Gli*, *Mdm1*, *Mdm2* and *Mdm3*: candidate genes for the mouse *pg* locus. Mamm. Genome **5:** 608–611.

Ashar, H. R., M. Schoenberg Fejzo, A. Tkachenko, X. Zhou, J. A. Fletcher *et al.*, 1995 Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains. Cell **82:** 57–65.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.*, 1988 *Current Protocols in Molecular Biology*, Vol. 2. Wiley, New York.

Barbu, V., and F. Dautry, 1989 Northern blot normalization with a 28S oligonucleotide probe. Nucleic Acids Res. **17:** 7115.

Battista, S., V. Fidanza, M. Fedele, A. J. P. Klein-Szanto, E. Outwater *et al.*, 1999 The expression of a truncated *HMGI-C* gene induces gigantism associated lipomatosis. Cancer Res. **59:** 4793–4797.

Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church *et al.*, 1992 Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell **68:** 799–808.

Bultman, S. J., E. J. Michaud and R. P. Woychik, 1992 Molecular characterization of the mouse agouti locus. Cell **71:** 1195–1204.

Cherath, L., K. F. Benson and K. Chada, 1995 Identification and characterization of a novel conserved DNA repeat. Mamm. Genome **6:** 611–616.

Chirgwin, J. M., A. E. Przbyla, R. J. MacDonald and W. J. Rutter, 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:** 5294–5299.

Fedele, M., M. T. Berlingieri, S. Scala, L. Chiariotti, G. Viglietto *et al.*, 1998 Truncated and chimeric HMGI-C genes induce neo-

plastic transformation of NIH3T3 murine fibroblasts. Oncogene **17:** 413–418.

Finsterer, J., E. Gharehbaghi-Schnell, C. Stollberger, K. Fheodoroff and A. Seiser, 2001 Relation of cardiac abnormalities and CTG-repeat size in myotonic dystrophy. Clin. Genet. **59:** 350–355.

Fu, Y. H., A. Pizzuti, R. G. Fenwick, J. King, S. Rajnarayan *et al.*, 1992 An unstable triplet repeat in a gene related to myotonic dystrophy. Science **255:** 1256–1258.

Grady, R. M., H. Teng, M. C. Nichol, J. C. Cunningham, R. S. Wilkinson *et al.*, 1997 Skeletal and cardiac myopathies in mice lacking utrophin and dystrophin: a model for Duchenne muscular dystrophy. Cell **90:** 729–738.

Kazmierczak, B., J. Bullerdiek, K. H. Pham, S. Bartnitzke and H. Wiesner, 1998 Intron 3 of HMGIC is the most frequent target of chromosomal aberrations in human tumors and has been conserved basically for at least 30 million years. Cancer Genet. Cytogenet. **103:** 175–177.

Ladd, A. N., N. Charlet and T. A. Cooper, 2001 The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol. Cell. Biol. **21:** 1285–1296.

Liquori, C. L., K. Ricker, M. L. Moseley, J. F. Jacobsen, W. Kress *et al.*, 2001 Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science **293:** 864–867.

Lu, X., N. A. Timchenko and L. T. Timchenko, 1999 Cardiac *elav*-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. Hum. Mol. Genet. **8:** 53–60.

Lyon, M. F., and A. G. Searle (Editors), 1989 *Genetic Variants and Strains of the Laboratory Mouse.* Oxford University Press, Oxford.

MacArthur, J. W., 1944 Genetics of body size and related characters. Am. Nat. **78:** 142–157.

Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya *et al.*, 1992 Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science **255:** 1253–1255.

Mankodi, A., E. Logigian, L. Callahan, C. McClain, R. White *et al.*, 2000 Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science **289:** 1769–1772.

Messina, D. N., M. C. Speer, M. A. Pericak-Vance and E. M. McNally, 1997 Linkage of familial dilated cardiomyopathy with conduction defect and muscular dystrophy to chromosome 6q23. Am. J. Hum. Genet. **61:** 909–917.

Miller, J. W., C. R. Urbinati, P. Teng-umnuay, M. G. Stenberg, B. J. Byrne *et al.*, 2000 Recruitment of human muscleblind proteins to $(CUG)_n$ expansions associated with myotonic dystrophy. EMBO J. **19:** 4439–4448.

Perry, W. L., C. M. Hustad, D. A. Swing, T. N. O'Sullivan, N. A. Jenkins *et al.*, 1998 The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a^{18H} mice. Nat. Genet. **18:** 143–146.

Philips, A. V., L. T. Timchenko and T. A. Cooper, 1998 Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science **280:** 737–741.

Quaderi, N. A., S. Schweiger, K. Gaudenz, B. Franco, E. I. Rugarli *et al.*, 1997 Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. Nat. Genet. **17:** 285–291.

Roderick, T. H., 1983 Using inversions to detect and study recessive lethals and detrimentals in mice, pp. 135–167 in *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk*, edited by F. J. de Serres and W. Sheridan. Plenum Press, New York.

Rowe, L. B., J. H. Nadeau, R. Turner, W. N. Frankel, V. A. Letts *et al.*, 1994 Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. Mamm. Genome **5:** 253–274.

Schonberger, J., H. Levy, E. Grunig, S. Sangwatanaroj, D. Fatkin *et al.*, 2000 Dilated cardiomyopathy and sensorineural hearing loss: a heritable syndrome that maps to 6q23–24. Circulation **101:** 1812–1818.

Shaffer, L. G., and J. R. Lupski, 2000 Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu. Rev. Genet. **34:** 297–329.

Uyeda, A., I. Fukui, K. Fujimori, K. Kiyosue, H. Nishimune *et al.*, 2000 MDP77: a novel neurite-outgrowth-promoting protein predominantly expressed in chick muscle. Biochem. Biophys. Res. Commun. **269:** 564–569.

Xiang, X., K. F. Benson and K. Chada, 1990 Mini-mouse: disruption of the pygmy locus in a transgenic insertional mutant. Science **247:** 967–969.

Zhou, X., K. F. Benson, H. R. Ashar and K. Chada, 1995 Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature **376:** 771–774.

Zhou, X., K. F. Benson, K. Przybysz, J. Liu, Y. Hou *et al.*, 1996 Genomic structure and expression of the murine *Hmgi-c* gene. Nucleic Acids Res. **24:** 4071–4077.

Communicating editor: N. A. Jenkins