Mamm Genome (2007) 18:852–860
DOI 10.1007/s00335-007-9074-9

An intronic sequence mutated in *flexed-tail* mice regulates splicing of *Smad5*

Shailaja Hegde · Laurie E. Lenox · Andrew Lariviere · Prashanth Porayette · John M. Perry · Michele Yon · Robert F. Paulson

Received: 10 July 2007 / Accepted: 10 October 2007 / Published online: 30 November 2007

Abstract Recent work has identified a growing body of evidence that subtle changes in noncoding sequences can result in significant pathology. These mutations, which would have been called silent polymorphisms in the past, affect gene transcription and mRNA splicing and lead to drastic changes in gene expression. Previous work from our lab has characterized the murine *flexed-tail* (*f*) mutation, which encodes *Smad5*, a transcription factor that functions downstream of the receptors for bone morphogenetic proteins (BMPs). *f/f* mice are unable to rapidly respond to acute anemia. Our analysis of these mice led to the development of a new model for stress erythropoiesis, where BMP4 expression in the spleen leads to the *Smad5*-dependent expansion of a specialized population of stress erythroid progenitors during the recovery from acute anemia. *f/f* mutant mice exhibit a defect in *Smad5* mRNA splicing in the spleen such that the majority of *Smad5* transcripts are two misspliced mRNAs. One of these mRNAs encodes a truncated form of *Smad5* that inhibits BMP4 signaling when overexpressed. Here we show that a mutation in a poly(T) element in intron 4 causes the splicing defect in *f/f* mutant mice. This subtle mutation (loss of 1 or 2 Ts in a 16-T element) results in defects in splicing throughout the *Smad5* gene. Furthermore, we show that this mutation results in tissue-specific splicing defects, which may explain why *f/f* mice are viable when *Smad5*-/- mice are embryonic lethal.

S. Hegde · M. Yon · R. F. Paulson (✉)
Department of Veterinary and Biomedical Sciences, Pennsylvania State University, The Center for Molecular Immunology and Infectious Disease, 115 Henning Building, University Park, Pennsylvania 16802, USA
e-mail: rfp5@psu.edu

R. F. Paulson
The Center for Molecular Immunology and Infectious Disease, Pennsylvania State University, University Park, Pennsylvania 16802, USA

L. E. Lenox · A. Lariviere · R. F. Paulson
The Graduate Program in Biochemistry, Molecular and Microbiology and the Huck Institutes of the Life Sciences Pennsylvania State University, University Park, Pennsylvania 16802, USA

P. Porayette · R. F. Paulson
Department of Molecular Medicine, Pennsylvania State University, University Park, Pennsylvania 16802, USA

J. M. Perry · R. F. Paulson
Department of Cell and Developmental Biology Options, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Present Address:
L. E. Lenox
Abramson Family Cancer Institute, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Present Address:
J. M. Perry
Stowers Institute for Medical Research, Kansas City, Missouri, USA

Introduction

Tissue hypoxia induces a physiologic response designed to increase oxygen delivery to the tissues. One component of this response is increased erythropoiesis. At times of acute need, large numbers of new erythrocytes must be rapidly generated (Socolovsky 2007). This process utilizes progenitors and signals that are distinct from those used by steady-state erythropoiesis (Lenox et al. 2005; Perry et al.

2007). Previous work from our lab has shown that BMP4 acts in concert with stem cell factor (SCF) and hypoxia to induce the expansion of differentiation of a specialized population of stress erythroid progenitors in the spleen (Perry et al. 2007). These progenitors exhibit a greater capacity to generate erythrocytes and differentiate faster than steady-state erythroid progenitors.

The key role for BMP4 in this process was first demonstrated by the analysis of the murine *flexed-tail* (*f*) mutation. *f/f* mice are unable to rapidly generate erythrocytes at times of acute need (Russell 1979). During embryogenesis *f/f* embryos exhibit a fetal neonatal anemia that resolves two weeks after birth, which coincides with the time that the bone marrow becomes the primary site of steady-state erythropoiesis (Gruneberg 1942a, b; Mixter and Hunt 1933; Palis and Segel 1998). As adults, *f/f* mice exhibit normal steady-state blood parameters (Gregory et al. 1975). However, when challenged with an acute anemia, *f/f* mice are significantly slower to recover (Coleman et al. 1969; Lenox et al. 2005). We cloned the *f* locus and showed that it encoded Smad5, a receptor-activated Smad that functions downstream of the receptors for BMP4. *f/f* mice have a defect in splicing such that the majority of the *Smad5* mRNA in the *f/f* mutant spleen is misspliced (Lenox et al. 2005). Two misspliced messages are observed. The common feature of these messages is the deletion of exon 2 which contains the initiator ATG. Overexpression of one of these mRNAs in a BMP4-responsive osteoblast cell line showed that it was capable of inhibiting BMP4 signaling. These data and experiments using *f/Smad5* transheterozygotes demonstrated that the severity of the *f* mutant phenotype depended on the dose of misspliced *Smad5* mRNA (Lenox et al. 2005). In our initial report we were unable to identify the molecular lesion responsible for the splicing defect. In this report we show that *f/f* mice have a mutation in a poly(T) element in intron 4. Wild-type mice have 16 Ts while *f/f* mice have 14 or 15. This subtle change, when introduced into a BAC clone containing the *Smad5* gene, is sufficient to cause a splicing defect. In addition, the splicing defect caused by the mutation appears to be tissue-specific in that the severity of aberrant splicing is variable depending on the tissue and different tissues exhibit different misspliced *Smad5* mRNAs. These data explain why *f/f* mice are viable, while *Smad5*-/- mice are embryonic lethal (Chang et al. 1999; Yang et al. 1999).

Materials and methods

Sequencing the *f* locus

Genomic DNA was isolated from C57BL/6 and C57BL/6-*f/f* mice using standard methods. The ten fragments spanning the *Smad5* locus were generated by PCR using the following primers. The PCR fragments were cloned into PCR XL-Topo vector according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). Multiple independent clones were sequenced for each fragment.

Primers used for amplifying the segments of the *Smad5* locus:

Fragment 1
Forward 5′ AAGTCAAAGGATTTGGAGG 3′
Reverse 5′ TCAGCAGAATTCACTCTCTACCTTG CTA 3′

Fragment 2
Forward 5′ GGGTAGATGATGTGTGTGTTGCCTTT 3′
Reverse 5′ CTCCAAGAACGTATATCATCAGGGC 3′

Fragment 3
Forward 5′ GTTAGCCAGGTACAGTAACGCATGCTT TC 3′
Reverse 5′ GCTGACCAGTTATCACTGTCACA-GAGTGATC 3′

Fragment 4
Forward 5′ GTGGTGGCTGTTATATCAGCTTTG TT 3′
Reverse 5′ ACAAGCATGAGCCAGCACATCTTAT 3′

Fragment 5
Forward 5′ TTCAGGAGAAGAGAGCAAGAAGCAA 3′
Reverse 5′ ACAAGCATGAGCCAGCACATCTTAT 3′

Fragment 6
Forward 5′ TTCAGGAGAAGAGAGCAAGAAGCAA 3′
Reverse 5′ GCAATATATAACATGGGGCAAGCCT 3′

Fragment 7
Forward 5′ CACGGCACATAGCTTCCATAACAC 3′
Reverse 5′ GGAAGGAGGATAGGGGCTGTTAG-GAG 3′

Fragment 8
Forward 5′ TTAATCACCACCCCTTAA AGGTTCC 3′
Reverse 5′ GATTAACATTTGACAACAATCCCAGG 3′

Fragment 9
Forward 5′ TGTTCAGCCTGTCGCCTAT 3′
Reverse 5′ CATACACAGCCTCGAATCCA 3′

Fragment 10
Forward 5′ ACTTTCACCATGGCTTCCATCCCAC 3′
Reverse 5′ CCACAGCACGCCTAGTTACAGCCCT 3′

Gap-repair cloning of intron 4 from *Smad5*

A plasmid containing fragment 8 of the *Smad5* locus was digested with *Nhe*I according to the manufacturer's instructions (Restriction enzyme, New England Biolabs, Ipswich, MA, USA). The digested plasmid was purified by

agarose gel electrophoresis. This digest generates a linear plasmid containing the 5′ 263 bp and 3′ 1779 bp of fragment 8, which was used as a cloning vector. Genomic DNA was isolated from C57BL/6 control and C57BL/6-*f/f* mutant mice. In addition, we used genomic DNA isolated from C58, BALB/c, and 129SvJ to rule out the possibility that the poly(T) mutation was a sequence polymorphism. The genomic DNA was sheared by sonication to generate smaller fragments of DNA. The λ phage recombineering strain DY380 was grown at 30°C until optical density (OD) = 0.4 and then the culture was heat shocked at 42°C for 15 min to induce the expression of the λ phage recombination proteins as previously described (Hegde and Paulson 2004; Yu et al. 2000, 2003). The induced DY380 cells were made electroporation-competent by washing in ice cold sterile water. The cloning vector (5 μg) and genomic DNA (24 μg) were electroporated into induced DY380 bacteria. Recombinant colonies were selected by kanamycin resistance (Kan^R). Full-length recombinants were identified by PCR of plasmid DNA using fragment 8 primers. For each mouse, two independent recombination reactions were done and multiple full-length recombinant clones were sequenced.

Generation of mutations in the poly(T) sequence of the *Smad5* BAC

A *Smad5*-containing BAC clone (RP24 267-I21) from a C57BL/6 BAC library was obtained from the BACPAC resources center. The generation of ΔT and ΔT:poly(T)15 mutants was done by recombineering techniques. We initially constructed the ΔT mutant using overlapping oligonucleotides (Yu et al. 2003). We used the galK coselection technique to identify recombination events, which screens for the targeting of Kan^R to the galK locus on the *E. coli* chromosome (Hegde and Paulson 2004). Thirty nanograms of ΔT forward and ΔT reverse oligonucleotides and 80 ng of a PCR fragment containing the Kan^R gene flanked by homology to galK were electroporated into *Smad5* BAC DY380 bacteria that had been heat shocked at 42°C to induce the λ phage recombination system as described above. Kan^R colonies that formed white galK- colonies when grown on MacConkey galactose indicator agar were selected and the ΔT mutation was confirmed by sequencing. The ΔT mutation BAC was retransformed into DY380 and the ΔT:poly(T)15 mutation was induced using the same procedure except that different oligonucleotides were used.

ΔT mutation primers:
ΔT Forward 5′ GGTTTTTTGTTTGCTTGGAACCTAT 3′
ΔT Reverse 5′ ATAGCCAAAAAgaacaaataggtt 3′

ΔT:poly(T)15 primers:
ΔT:poly(T)15 Forward 5′ TTGTTTGCTTGGTTTTTTTTTTTTTT 3′
ΔT:poly(T)15 Reverse 5′ AAATAGCCAAAAAGAACAATAGGTT 3′

Introduction of mutated and control *Smad5* BACs into MSS31 cells by infection

The initial step in these experiments was to generate BAC clones that contained the selectable markers neomycin/kanamycin resistance (neo^R) and EGFP integrated into BAC vector sequences. Oligos that contained homology to the chloramphenicol resistance (Cm^R) of the BAC vector (Frengen et al. 2000) at their 5′ ends and homology to neo^R/EGFP cassette from pEGFPN3 (Clontech Laboratories Inc., Mountain View, CA, USA) were used to PCR amplify a fragment that contained the neo^R/EGFP cassette flanked by 50 bp of BAC vector homology. This fragment was electroporated into DY380 bacteria which had been heat shocked to induce the expression of λ phage recombination proteins as described above. BAC DNA was isolated from Kan^R, $Cm^{sensitive}$ colonies and assayed for the correct insertion into the BAC vector by PCR.

Primers for neo^R/EGFP recombineering:
Forward 5′ TTCCGGTCACACCACATACGTTCCGCCATTCCTATGCGATGC
ACATGCTGTATGCCGGTATAACCGTATTCACGCCATGC 3′
Reverse 5′AGACTTCCGTTGAACTGATGGACTTATGTCCCATCAGGCTTTGCA
GAACTTTCAGCGGTAACGACCCAACACCGTGCGTT 3′

We next inserted the *invasin* (*inv*) gene from *Yersinia pseudotuberculosis* into the *aspartate β-semialdehyde* (*asd*) locus of each *Smad5*-BAC-containing DY380 strain (Narayanan and Warburton 2003). The invasin plasmid pRI203 was a kind gift of Dr. Ralph Isberg (Isberg et al. 1987). The *invasin* gene was PCR amplified from the plasmid using primers that contained homology to the *asd* locus at their 5′ ends. The fragment generated was purified on an agarose gel. A second fragment containing the Cm^R gene flanked by 50 bp of homology to the galK gene was generated so that coselection for Cm^R and galK- could be used as previously described. DY380 strains containing *Smad5* BAC clones that had been fitted with neo^R/EGFP cassettes were heat shocked at 42°C to induce the λ phage recombination system as described above. The inv fragment (600 ng) and the galK-Cm^R fragment (100 ng) were electroporated into the induced bacteria.

The bacteria were allowed to recover for 90 min at 37°C in 1 ml of LB medium containing 0.5 M diaminopimelic acid (DAP) and bacteria were plated on LB agar containing 0.5 M DAP and kanamycin and chormphenicol. Kan^R, Cm^R colonies were tested on MacConkey galactose indicator agar and those that formed white galK- colonies were selected. The insertion of *inv* into *asd* was then confirmed by PCR.

Primers used to generate *inv* fragment with homology to *asd* at each end:

5′ GAGACCGGCACATTTATACAGCACACATC TTTGCAGGAAAAAAACGCTATGGTTTTCCAGC CAATCGAGTTT 3′

5′ CTCCTGTATTACGCACTAACAGGGGCGGCA TCGCGCCCCAGATTTAATTTATATTGACAGCG-CACAGAGCGGGAACGC 3′

Primers used to verify the insertion of *inv* into *asd*:

5′ TGGTGAAGATGTGCCAAGA 3′

5′ TCCCGGTAAATCATGAAACA 3′

The infection of MSS31 spleen stromal cells with *inv* + *asd*- *Smad5*-BAC-containing DY380 bacteria was done as previously described (Narayanan and Warburton 2003). In brief, overnight cultures (OD = 1.2) were grown in brain-heart infusion broth supplemented with 0.5 M DAP. The bacteria were spun down and resuspended in 1 ml of RPMI 1640 media. The bacteria were overlaid onto monolayer cultures of MSS31 cells (1×10^5 cells/well) at an multiplicity of infection (MOI) of 4500. The plates were centrifuged at 1000 rpm for 10 min. After 2 h at 37°C, the cells were washed with media and then incubated in media containing G418 (0.5 mg/ml). $G418^R$ colonies were pooled after two weeks of selection and analyzed for Smad5 splicing defects.

RT-PCR analysis of *Smad5* expression

Total RNA was isolated from MSS31 cells and the BAC-containing MSS31 cells using the TRIzol reagent (Invitrogen) according to the manufacturer's instructions. cDNA was generated and *Smad5* cDNA was amplified using 5′-GGGGCCGAGCTGCTAAT-3′ and 5′-CTATGAAACA-GAAGAAATGGGG-3′ primers (Lenox et al. 2005). For the analysis of Smad5 splicing in other tissues, these tissues were isolated from C57BL/6-*f/f* and C57BL/6 control mice. The tissues were homogenized in TRIzol reagent according to the manufacturer's instructions and total RNA was isolated. cDNA was generated and *Smad5* expression analyzed by RT-PCR as described for MSS31 cells. For both MSS31 cells and primary tissues, RT-PCR reactions were separated on agarose gels and the major bands were cloned into the PCR XL-TOPO vector. Multiple independent clones were sequenced to determine the alterations in splicing.

Results

Our initial analysis of the splice donor and acceptor sites in the *Smad5* gene failed to identify any mutations that could explain the aberrant mRNA splicing observed in *f/f* mice (Lenox et al. 2005). However, these sequences are only a small part of the *cis*-acting sequences that can regulate mRNA splicing. A growing body of work has identified a large number of both intronic and exonic sequences that can affect splicing (Baralle and Baralle 2005; Buratti et al. 2006; Faustino and Cooper 2003). To determine whether *f* mice had a mutation in such a sequence, we sequenced the entire *Smad5* locus from *f/f* mice. As shown in Fig. 1A, the locus was subdivided into ten fragments 3–5 kb in length,

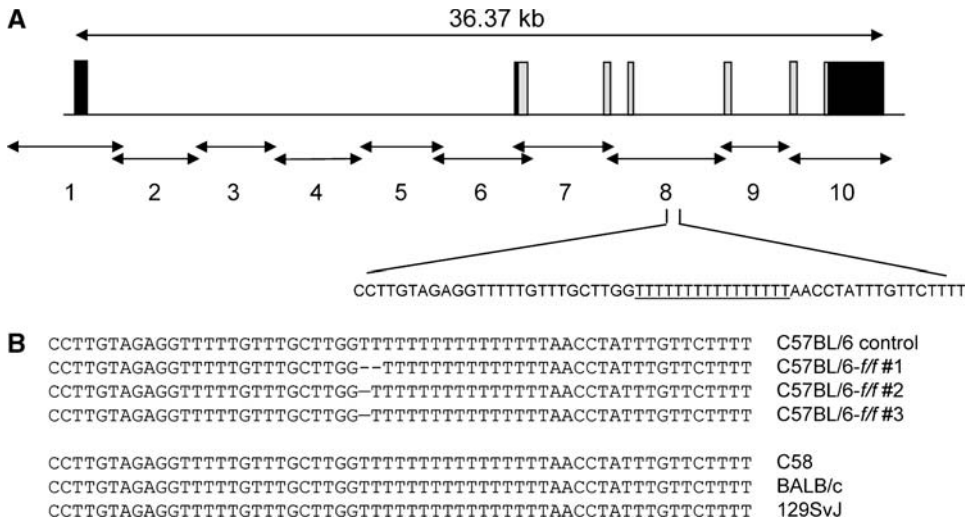

Fig. 1 Sequencing of the *Smad5* locus identifies a mutation in a poly(T) element in intron 4. **A** Schematic of the *Smad5* locus showing the position of the ten PCR fragments used to clone the mutation in *f* mice. Black exons denote 5′ and 3′ UTRs and gray exons coding sequences. Below fragment 8 is the sequence of the poly(T) element. **B** Sequence of the poly(T) element from three C57BL/6 control, three C57BL/6-*f/f*, C58, BALB/c, and 129SvJ mice derived from gap-repair cloning. The three control C57BL/6 mice all showed identical sequence

which were amplified by PCR and cloned. The fragments were sequenced and compared to the C57BL/6 reference sequence. Over the entire 36.4-kb region, we identified only one consistent change in sequence. In intron 4 we identified a homopolymer run of 16 T residues that was reduced to 14 in *f/f* mice (Fig. 1). PCR amplification of homopolymer sequences can be problematic due to polymerase slippage. To circumvent this problem, we cloned this region by gap-repair cloning using the λ phage recombineering system (Copeland et al. 2001; Lee et al. 2001). Starting with a plasmid that contained fragment 8, we generated a cloning vector by digesting the plasmid with *Nhe*I to generate a linear plasmid that contained the 5′ 263 bp and the 3′ 1779 bp. Genomic DNA was isolated from C57BL/6 control and C57BL/6-*f/f* mice. The DNA was sheared to generate smaller DNA fragments. The genomic DNA and the linear fragment 8 cloning vector were electroporated into the λ phage recombineering strain DY380 (Yu et al. 2000). Kan^R recombinants were identified and the isolated plasmids sequenced. The analysis of three independent *f/f* mice showed that they contained either 14 or 15 T residues, while the analysis of control C57BL/6 mice showed that they all contained 16 T residues (Fig. 1B). A difference of one or two T residues is a minor change and could represent a polymorphism in the C57BL/6-*f/f* strain. We analyzed this sequence in other strains to determine whether this change represented a polymorphic sequence or a potential mutation. Using gap-repair cloning we isolated this region of intron 4 from the C57BL/6-related strain C58 and two unrelated strains, BALB/c and 129SvJ. Analysis of the sequence of this region in these strains showed that all of the strains had 16 Ts like the wild-type C57BL/6 mice. This observation suggests that 16 Ts is the normal sequence and the change observed in C57BL/6-*f/f* mice may represent a mutation. We also compared this sequence in the mouse with the sequences of intron 4 in rat and human to determine whether this element is conserved in other species. We observed no similar sequences any where in human or rat *Smad5*, which suggests that this element is mouse-specific (data not shown). The loss of a single T residue would seem unlikely to result in a mutant phenotype as severe as the *f/f* phenotype; however, previous work on several genes, including *CFTR* (Niksic et al. 1999; Nissim-Rafinia et al. 2000; Zuccato et al. 2004) and *MRE11* (Giannini et al. 2004), has shown that changes in poly(T) sequences can have a profound effect on splicing.

To demonstrate that this change in the poly(T) sequence was the cause of the splicing defect, we mutated this sequence in a BAC clone that contained the entire *Smad5* locus and tested whether these mutants affected *Smad5* mRNA splicing. Our rationale for using BAC clones was that the defect that we observed in *f/f* mice did not occur in the splicing of the exons flanking the mutation, but rather affected upstream and downstream sites. Thus, we needed to study the mutation in the context of the entire locus. We used λ phage recombineering to first delete the entire poly(T) sequence to generate the ΔT allele. We then used this allele to reintroduce the poly(T)15 mutant allele into the BAC, which generated ΔT:poly(T)15 allele (Fig. 2A). Transfection of BAC DNA into mammalian cells can be difficult and poor transfection efficiencies are common. An alternative strategy to introduce BACs into tissue culture cells uses bacterial infection to deliver the BAC DNA. Ectopic expression of the *invasin* (*inv*) gene from *Yersinia pseudotuberculosis* makes bacteria competent to invade mammalian cells (Narayanan and Warburton 2003). Inv binds β-integrin on the target cell, which then facilitates the uptake of bacteria (Isberg and Leong 1990). Coupling *inv* expression with mutation of the *aspartate β-semialdehyde* (*asd*) locus, which causes a cell wall defect due to auxotrophy for diaminopimelic acid, leads to efficient uptake of bacteria into cells where they lyse upon division and deliver the BAC DNA into the cells (Narayanan and Warburton 2003) (Fig. 2B).

A selectable marker, neomycin resistance/EGFP (neo^R/EGFP), was inserted into the BAC vector sequences of the control, ΔT, and ΔT:poly(T)15 *Smad5* BAC clones by recombineering. *Inv* was inserted into the *asd* locus by recombineering in *E. coli* DY380 to generate DY380 *inv* + *asd*- strains that contained control or mutant BAC clones. The three BAC strains were incubated with the spleen stromal cell line MSS31. These cells express endogenous *Smad5*. The cells were washed with PBS to remove nonadherent bacteria and then cultured in new media containing G418 to select for cells that have integrated the BAC clone. Pools of neo^R/EGFP+ colonies for each BAC clone were expanded and the expression of *Smad5* was determined by RT-PCR. Figure 3 shows that uninfected MSS31 and those infected with bacteria containing the control BAC clone express wild-type *Smad5* mRNA. However, the cells that received the ΔT exhibited a new *Smad5* mRNA. Sequencing this RT-PCR product revealed that it encoded a misspliced transcript that lacked exon 4 and included an insertion of 15 nucleotides between exons 6 and 7. This insertion was observed previously in one of the misspliced mRNAs that we observed in the spleen of C57BL/6-*f/f* mice (Lenox et al. 2005). However, that aberrant mRNA had a deletion of exon 2 in addition to the deletion of exon 4. In addition, we observed several small RT-PCR products, which contained portions of exon 4 and intron 4. Reintroduction of the mutated poly(T) sequence in the ΔT:poly(T)15 *Smad5* BAC clone resulted in the appearance of the same larger misspliced mRNA as observed with the ΔT BAC. However, the smaller RT-PCR products were not observed. Based on these observations,

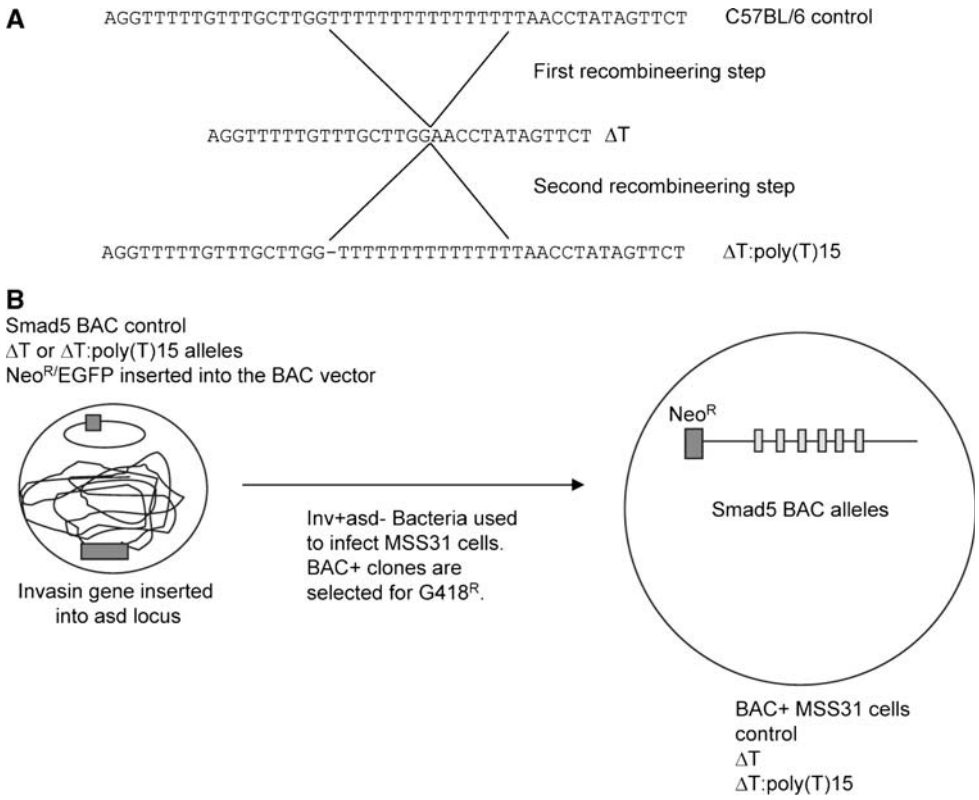

Fig. 2 Generation of mutations in the poly(T) element in *Smad5* BAC clones and the introduction of mutated BACs into MSS31 cells by infection. **A** Schematic of the mutations introduced into the *Smad5* BAC clone by recombineering to generate the ΔT and ΔT:poly(T)15 mutants. **B** A schematic of the strategy used to introduce BAC clones into MSS31 cells by infection

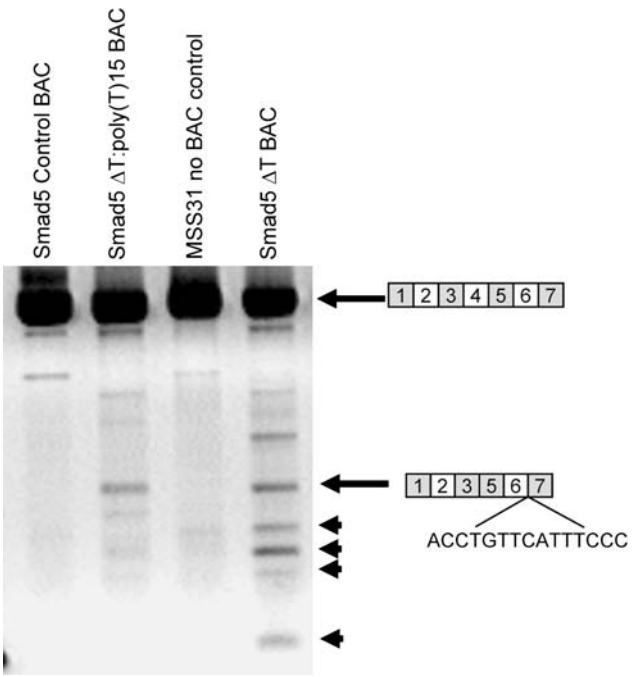

Fig. 3 Mutated *Smad5* BAC clones generate misspliced mRNAs. *Smad5* mRNA was analyzed by RT-PCR using RNA isolated from MSS31 control cells and MSS31 cells that contained mutant and control *Smad5* BAC clones. Arrows indicate major *Smad5* mRNAs and their exon structure. Arrowheads indicate small aberrantly spliced mRNAs that contained variable portions of exon4 and intron 4. All other bands were nonspecific in that they were not consistently observed and did not contain sequences from the *Smad5* locus

we conclude that the poly(T) element in intron 4 of the *Smad5* gene plays a key role in the proper regulation of mRNA splicing.

The analysis of MSS31 cells containing different *Smad5* BAC clones showed that mutation of the poly(T) sequence caused defects in splicing similar but not identical to what we observed in the spleen erythroid progenitors. These data suggest that *in vivo* the poly(T)14/15 mutation may have different effects in different tissues. We tested this possibility by isolating mRNA from spleen, bone marrow, kidney, and liver of C57BL/6-*f/f* and control mice. The different *Smad5* mRNA isoforms were analyzed by RT-PCR and sequence analysis. Figure 4 shows that altered splicing is observed in all of the tissues tested. However, the severity of the splicing defect was variable. Spleen showed an almost complete lack of full-length wild-type *Smad5* mRNA as we previously reported (Lenox et al. 2005). In contrast, bone marrow, kidney, and liver all exhibited substantial amounts of wild-type *Smad5* mRNA and lower levels of aberrantly spliced *Smad5* mRNAs. Similar to our observations *in vitro* with MSS31 cells expressing mutant BAC clones, the sequence of the aberrantly spliced mRNA in bone marrow, kidney, and liver was distinct from the misspliced mRNAs we observed in the spleen and also distinct from those observed in MSS31 cells. This mRNA showed some similarity to those cloned from *f/f* spleen in that exon 1 is spliced into the same cryptic splice

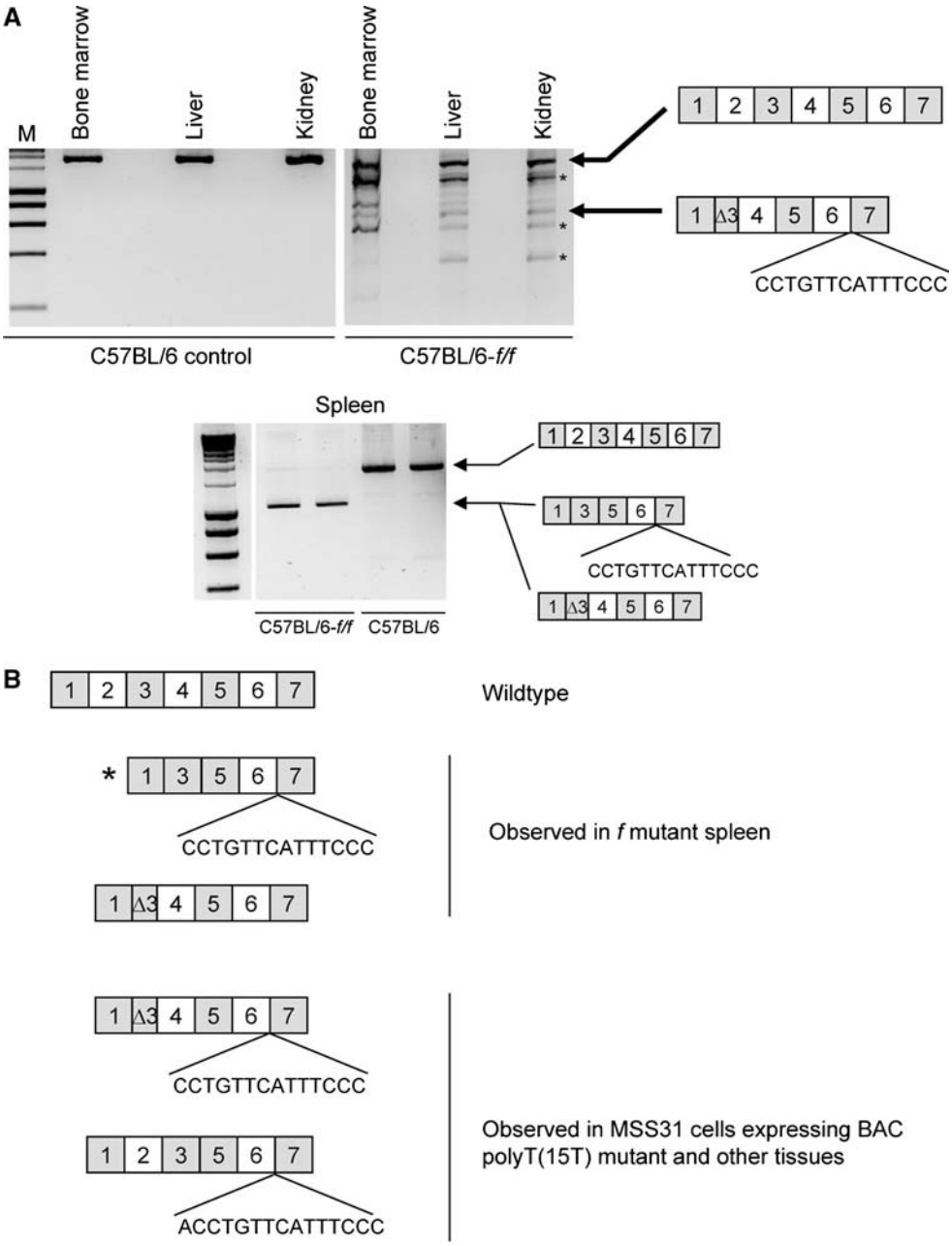

Fig. 4 Analysis of *Smad5* mRNA splicing in other adult tissues. **A** RNA was isolated from bone marrow, kidney, and liver (top) and spleen (bottom) of C57BL/6 and C57BL/6-*f/f* mice. *Smad5* mRNA expression was measured by RT-PCR. The major *Smad5* mRNAs are indicated by the arrows and their exon structure is given. Bands indicated by * are nonspecific in that they are not consistently observed and do not contain sequences from the *Smad5* locus. **B** Schematic of the aberrantly spliced forms of *Smad5* observed in the spleen (Lenox et al. 2005) or in this study. * indicates pathogenic *Smad5* mRNA isoform

acceptor in the middle of exon 3. In addition, this mRNA also contains the 15-bp insert, which corresponds to sequences in intron 6. Taken together, these data support a model where the mutation of the poly(T) sequence in *f/f* mice leads to defects in *Smad5* mRNA splicing, which vary from tissue to tissue.

Discussion

Our analysis of the *Smad5* locus in *f/f* mice has identified a poly(T) element that regulates mRNA splicing. Mutation of this element leads to aberrant use of splice sites and activation of cryptic splice sites, which leads to the skipping of whole or parts of exons and the inclusion of intronic sequences in the mature *Smad5* mRNA. The regulation of splicing is complex. Only in recent years has work expanded on the relatively minimal sequence motifs initially used to characterize splice sites and identified intronic and exonic enhancers and suppressors of splicing (for a review see Baralle and Baralle 2005; Buratti et al. 2006; Faustino and Cooper 2003). The defects associated with the poly(T) mutation in *f/f* mice suggest that this element may function as a negative regulator of splicing that prevents the inappropriate use of cryptic splice sites. Furthermore, these data underscore the idea that mutations

or polymorphisms, previously thought to be silent because they did not affect coding sequences, can have profound effects on splicing and result in significant pathology.

These data also address an important question concerning the relationship between the *f* allele of *Smad5* and the role of *Smad5* during development. The phenotype of the *Smad5*-/- mouse is embryonic lethal between embryonic day 10.5 and 11.5 (E10.5–11.5) (Chang et al. 1999; Yang et al. 1999). The cause of the lethality in these embryos is complex. *Smad5*-/- embryos exhibit defects in angiogenesis and have defective yolk sac vasculature (Chang et al. 1999; Yang et al. 1999). In addition, one group noted that the mutant embryos exhibited a high level of mesenchymal apoptosis and suggested that the interactions between endothelial and mesenchymal cells were disrupted in the mutants (Yang et al. 1999). *f/f* mutant mice do not exhibit embryonic lethality. The primary developmental phenotype is severe fetal anemia, which resolves by two weeks after birth (Gruneberg 1942a, b; Mixter and Hunt 1933; Russell 1979). The effects of the *Smad5* null mutation on fetal erythropoiesis are unknown because the embryos die before the onset of fetal liver erythropoiesis. How might these differences in the phenotypes be reconciled? Our data suggest that the variable splicing defects associated with the poly(T) sequence mutation lead to multiple *Smad5* mRNA isoforms. Our previous data showed that in the spleen only one of the aberrant *Smad5* mRNAs can inhibit BMP4 signaling when overexpressed in a BMP4 responsive cell line (Lenox et al. 2005). This mRNA contained a deletion of exon 2 and an insertion of the same 15 nucleotides between exons 6 and 7. This insertion was also observed in the MSS31 cells that contained the mutated *Smad5* BAC clones. The other predominant misspliced mRNA in the spleen of *f/f* mice had no effect on BMP4 signaling. Its sequence is more similar to the aberrant mRNAs observed in other tissues in *f/f* mice and MSS31:*Smad5* ΔT BAC cells (Fig. 4B). It is important to remember that *f* is a recessive mutation. Heterozygotes do not exhibit a phenotype (Lenox et al. 2005). However, our previous genetic analysis showed that *f* is not a *Smad5* loss-of-function allele in that *f/Smad5*- mice do not exhibit a more severe phenotype than *f/f* (Lenox et al. 2005). The increased level of mutant protein in *f/f* mice coupled with the low levels of wild-type *Smad5* mRNA in the spleen results in a defect in BMP4 signaling in spleen progenitor cells. Based on these data we propose that the specific phenotypes in *f/f* mice are caused by an almost complete absence of wild-type *Smad5* mRNA and the expression of an inhibitory Smad5 protein in BMP4-dependent components of the stress erythroid lineage which leads to fetal anemia and defects in the response to acute anemia in adult mice. *f/f* embryos survive because the splicing defect in *f/f* embryos is not severe and does not lead to the expression of inhibitory proteins in critical tissues during embryogenesis.

f/f mice also exhibit two other phenotypes, namely, tail flexures from which the mutant receives its name and white belly spots (Hunt and Premar 1928; Kamenoff 1935; Mixter and Hunt 1933). Both of these phenotypes can be traced to BMP4-dependent pathways (Christiansen et al. 2000; Murtaugh et al. 1999; Sela-Donenfeld and Kalcheim 1999). In C57BL/6-*f/f* mice tail flexures are not observed. Similarly, the white belly spots are incompletely penetrant. These observations suggest that the splicing defect in *f/f* can be affected by differences in strain background and indicates that splicing control elements may be regulated by allelic differences in *trans*-acting factors that segregate in inbred strains of mice.

The feature that distinguishes the mutation in the poly(T) element of *Smad5* from other examples of poly(T) sequences affecting splicing is that the poly(T) mutation in *Smad5* does not exclusively affect the splicing of the two flanking exons. Other examples such as *CFTR* and *Mre11* exhibit skipping of the exon downstream of the homopolymer sequence. These elements are part of the polypyrimidine stretch that marks the branch point splice acceptor region of the intron (Giannini et al. 2004; Niksic et al. 1999; Nissim-Rafinia et al. 2000). Analysis of these mutations has suggested that the fewer T residues the greater the probability that the downstream exon will be skipped (Niksic et al. 1999; Nissim-Rafinia et al. 2000; Zuccato et al. 2004). In contrast, the *f* poly(T) element appears to regulate the use of a number of cryptic and normal splice sites.

In summary, we have identified a mutation in a poly(T) element in intron 4 of the *Smad5* gene in *f/f*. This mutation leads to a tissue-specific splicing defect, which varies in its severity and the aberrant mRNAs produced. We propose that the variability in the tissue distribution of the splicing defect is responsible for the differences in the phenotype of *f/f* and *Smad5*-/- mice.

Acknowledgements This work was supported by grants from NIH/NHLBI to RFP. The authors thank Don Court and Ralph Isberg for providing reagents.

References

Baralle D, Baralle M (2005) Splicing in action: assessing disease-causing sequence changes. J Med Genet 42:737–748

Buratti E, Baralle M, Baralle FE (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. Nucleic Acids Res 34:3494–3510

Chang H, Huylebroeck D, Verschueren K, Guo Q, Matzuk MM, et al. (1999) Smad5 knockout mice die at mid-gestation due to multiple embryonic and extraembryonic defects. Development 126:1631–1642

Christiansen J, Coles E, Wilkinson D (2000) Molecular control of neural crest formation, migration and differentiation. Curr Opin Cell Biol 12:719–724

Coleman D, Russell E, Levin E (1969) Enzymatic studies of the hemopoietic defect in flexed mice. Genetics 61:631–642

Copeland NG, Jenkins NA, Court DL (2001) Recombineering: a powerful new tool for mouse functional genomics. Nat Rev 2:769–779

Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419–437

Frengen E, Zhao B, Howe S, Weichenhan D, Osoegawa K, et al. (2000) Modular bacterial artificial chromosome vectors for transfer of large inserts into mammalian cells. Genomics 68:118–126

Giannini G, Rinaldi C, Ristori E, Ambrosini MI, Cerignoli F, et al. (2004) Mutations of an intronic repeat induce impaired MRE11 expression in primary human cancer with microsatellite instability. Oncogene 23:2640–2647

Gregory C, McCulloch E, Till J (1975) The cellular basis for the defect in haematopoiesis in flexed-tail mice. III. Restriction of the defect to erythropoietic progenitors capable of transient colony formation *in vivo*. Br J Haematol 30:401–410

Gruneberg H (1942a) The anaemia of the flexed-tail mouse (*Mus musculus* L.) I. Static and dynamic haematology. J Genet 43:45–68

Gruneberg H (1942b) The anaemia of the flexed-tail mice (*Mus musculus* L.) II. Siderocytes. J Genet 44:246–271

Hegde S, Paulson RF (2004) Co-targeting a selectable marker to the *Escherichia coli* chromosome improves the recovery rate for mutations induced in BAC clones by homologous recombination. BioTechniques 36:936–938, 940

Hunt H, Premar D (1928) Flexed-tail a mutation in the house mouse. Anat Rec 41:117

Isberg RR, Leong JM (1990) Multiple beta 1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. Cell 60:861–871

Isberg RR, Voorhis DL, Falkow S (1987) Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells. Cell 50:769–778

Kamenoff R (1935) Effects of the flexed-tail gene on the development of the house mouse. J Morphol 58:117–155

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, et al. (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73:56–65

Lenox L, Perry J, Paulson R (2005) BMP4 and Smad5 regulate the erythroid response to acute anemia. Blood 105:2741–2748

Mixter R, Hunt H (1933) Anemia in the flexed tailed mouse, *Mus musculus*. Genetics 18:367–387

Murtaugh L, Chyung J, Lassar A (1999) Sonic Hedgehog promotes somitic chondrogenesis by altering the cellular response to BMP signaling. Genes Dev 13:225–237

Narayanan K, Warburton PE (2003) DNA modification and functional delivery into human cells using *Escherichia coli* DH10B. Nucleic Acids Res 31:e51

Niksic M, Romano M, Buratti E, Pagani F, Baralle FE (1999) Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. Hum Mol Genet 8:2339–2349

Nissim-Rafinia M, Chiba-Falek O, Sharon G, Boss A, Kerem B (2000) Cellular and viral splicing factors can modify the splicing pattern of CFTR transcripts carrying splicing mutations. Hum Mol Genet 9:1771–1778

Palis J, Segel G (1998) Developmental biology of erythropoiesis. Blood Rev 12:106–114

Perry JM, Harandi OF, Paulson RF (2007) BMP4, SCF, and hypoxia cooperatively regulate the expansion of murine stress erythroid progenitors. Blood 109:4494–4502

Russell E (1979) Hereditary Anemias of the mouse: a review for geneticists. Adv Genet 20:357–459

Sela-Donenfeld D, Kalcheim C (1999) Regulation of the onset of neural crest migration by coordinated activity of BMP4 and Noggin in the dorsal tube. Development 126:4749–4762

Socolovsky M (2007) Molecular insights into stress erythropoiesis. Curr Opin Hematol 14:215–224

Yang X, Castilla LH, Xu X, Li C, Gotay J, et al. (1999) Angiogenesis defects and mesenchymal apoptosis in mice lacking SMAD5. Development 126:1571–1580

Yu D, Ellis H, Lee E-C, Jenkins N, Copeland N, et al. (2000) An efficient recombination system for chromosomal engineering in *Escherichia coli*. Proc Natl Acad Sci U S A 97:5978–5983

Yu D, Sawitzke J, Ellis H, Court D (2003) Recombineering with overlapping single stranded DNA oligonucleotides: testing a recombination intermediate. Proc Natl Acad Sci U S A 100:7207–7212

Zuccato E, Buratti E, Stuani C, Baralle FE, Pagani F (2004) An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. J Biol Chem 279:16980–16988