

## Identification and characterization of a novel, evolutionarily conserved gene disrupted by the murine *H $\beta$ 58* embryonic lethal transgene insertion

JAMES J. LEE, GLENN RADICE\*, CHRISTOPHER P. PERKINS and FRANK COSTANTINI†

Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, 701 W 168<sup>th</sup> Street, New York, NY 10032, USA

\*Present address Department of Biophysics, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

†To whom correspondence should be addressed

### Summary

The *H $\beta$ 58* transgenic mouse line carries a recessive insertional mutation that results in developmental abnormalities beginning at day 7.5 *p.c.* and embryonic arrest at about day 9.5. In this paper, we describe the characterization of a novel gene encoded at the *H $\beta$ 58* locus, whose disruption appears to be responsible for the mutant phenotype. The wild-type *H $\beta$ 58* gene encodes a single 2.7 kb mRNA during embryonic and fetal development, and in many adult somatic tissues. In the mutant locus, this transcription unit is split by the transgene insertion, and one of its coding exons is deleted. Consistent with the physical disruption of the gene, the level of the *H $\beta$ 58* mRNA in heterozygous mutant mouse tissues was half the normal level, indicating that the mutant allele fails to encode a stable mRNA. *In situ* hybridization studies revealed that expression of the wild-type *H $\beta$ 58* gene begins in the oocyte, and continues throughout pre- and post-implantation embryogenesis, despite the fact that homozygous

mutant embryos develop successfully through the egg cylinder stage (day 6.5 *p.c.*). In the early post-implantation embryo, expression of the normal *H $\beta$ 58* gene is relatively low in the embryonic ectoderm, the tissue displaying the earliest phenotypic effects of the mutation, and highest in the visceral endoderm. We therefore propose that the effects of the mutation on the embryonic ectoderm may be exerted indirectly, via the visceral endoderm. Sequence analysis of *H $\beta$ 58* cDNA clones revealed no homology between the 38×10<sup>3</sup> M<sub>r</sub> *H $\beta$ 58* protein and other known proteins. However, the *H $\beta$ 58* gene displayed extremely strong conservation between mammals and birds (>96% amino acid identity), although it appeared less conserved in amphibians and invertebrates.

Key words: mouse embryo, *H $\beta$ 58*, insertional mutation, *in situ* hybridization, transgenic mice, evolutionary conservation

### Introduction

The normal course of mammalian embryogenesis depends on the correct temporal and spatial regulation of a large number of genes, as well as on complex regulative interactions between tissues. One strategy to identify genes that are expressed in the early embryo and, concurrently, to gain insight into their function in embryogenesis is to generate and characterize mutations that interfere with early murine development (reviewed by Magnuson, 1986; Gridley et al., 1987; Rossant and Joyner, 1989; Reith and Bernstein, 1991; Gridley, 1991). When retroviruses or transgenes are introduced into the mouse germ line, they appear to integrate in a largely random manner and they occasionally disrupt a host gene at the locus of insertion. Consequently, 5-10% of retroviral or transgene insertions are associated with recessive mutations with discernable developmental consequences (Pal-

miter and Brinster, 1986; Gridley et al., 1987). Although many such mutations have been identified, only a handful so far have been successfully characterized at a molecular level, leading to the identification of the mutant gene (*Mov-13*, Schnieke et al., 1983; *limb deformity*, Woychik et al., 1985, 1990; Maas et al., 1990; *Mov-34*, Soriano et al., 1987; Gridley et al., 1991; *Mpv-17*, Weiher et al., 1990; *dilute*, Mercer et al., 1991). With the exception of *limb deformity*, those cases in which a mutant gene has been identified involved a retroviral insertion event (*dilute*, *Mov 13*, *Mov-34*, *Mpv-17*).

We have recently described a recessive lethal mutation in transgenic line *H $\beta$ 58* (Costantini et al., 1985, 1989), which was caused by the integration of 10-20 tandem copies of an 8 kb transgene and an accompanying 2-3 kb deletion of host DNA on mouse chromosome 10 (Radice et al., 1991). Homozygous mutant *H $\beta$ 58* embryos developed normally through the

egg cylinder stage, day 6.5 *post coitum* (*p.c.*). However, at day 7.5 *p.c.*, the primitive streak stage, the homozygotes began to display reproducible abnormalities, the earliest of which was a severe reduction in the size of the embryonic ectoderm, the progenitor of all future embryonic tissues. Homozygotes subsequently became increasingly abnormal and, while forming recognizable derivatives of all germ layers, they never exceeded the size of a normal embryo at an early somite stage (8.5 days *p.c.*).

Among the extraembryonic lineages, the amnion and chorion of 7.5 day homozygous mutant embryos sometimes displayed abnormal folding, suggesting over-proliferation of these tissues, although at 8.5 or 9.5 days the amnion appeared normal. The allantois developed normally through day 8.5, but it failed to fuse with the chorion to form the chorioallantoic placenta. The visceral endoderm also appeared normal from its inception through the subsequent development of the visceral yolk sac (Radice et al., 1991). Thus, based on the mutant phenotype, we concluded that the gene(s) inactivated by the *Hβ58* insertion was not required for development prior to the egg cylinder stage, and that its earliest function was likely to involve the development of the embryonic ectoderm.

Using molecular probes derived from this locus, we detected the presence of a transcription unit active during embryonic development (Radice et al., 1991). In this paper, we describe the identification of a previously unknown gene that has been physically disrupted by the *Hβ58* insertion, and we characterize the normal pattern of expression of this gene during embryogenesis, as well as its evolutionary conservation. We discuss the possible relationship between the developmental expression pattern of the *Hβ58* gene and the phenotype of mutant embryos lacking the gene.

## Materials and methods

### Isolation and analysis of cDNA clones

Genomic probe C (Fig. 1c) was used to isolate the cDNA clone λE30.1 from an 8.5 day *p.c.* mouse embryo library (Fahrner et al., 1987). The cDNA insert was subcloned in the plasmid pBluescript KS(+) (Stratagene), to yield clone pE30.1. Clone pE30.1 was used as a probe to screen a chick d13 embryo brain cDNA library (a gift of Barbara Ranscht, La Jolla Cancer Research Foundation), and the longest chicken cDNA, pCK4.4 (2493 nt), was subcloned in pBluescript KS(+).

### Preparation of probes

Southern and Northern blots were hybridized with the insert DNA from cDNA clone pE30.1, which was isolated from agarose gels by electroelution and <sup>32</sup>P-labelled by nick translation to specific activity  $\sim 6 \times 10^8$  disintegrations/min/μg. For molecular titration assays, the 1.2 kb *Eco*RI fragment of cDNA clone pE30.1, containing the 3' untranslated region, was subcloned into pBluescript KS(+) to produce clone pE1.2. This DNA was linearized with *Hind*III and transcribed using T<sub>7</sub> RNA polymerase (Lee and Costlow, 1987), yielding a <sup>32</sup>P-labelled RNA probe of  $\sim 6 \times 10^8$  disintegrations/min/μg. *In situ* hybridizations were performed using <sup>35</sup>S-labelled RNA

probes ( $\sim 2 \times 10^9$  disintegrations/min/μg) derived from cDNA clone pE30.1. The probes were synthesized using T<sub>7</sub> or T<sub>3</sub> RNA polymerase (Pharmacia) from templates linearized with *Hind*III (anti-sense probe) or *Bam*HI (sense probe), and were reduced in size to 100-200 nucleotides by limited alkaline hydrolysis (Cox et al., 1984).

### RNA isolation

RNA was prepared from staged embryos (after removal of the visceral and parietal yolk sacs) and adult mouse tissues by the method of Auffray and Rougeon (1980). Poly(A)<sup>+</sup> RNA was isolated by affinity chromatography (Aviv and Leder, 1972).

### Northern blot analysis

Northern blot analyses were performed essentially as described by Posakony et al. (1983). Briefly, 10 μg of poly(A)<sup>+</sup> RNA was fractionated by gel electrophoresis on 1.2% agarose gels containing formaldehyde, and transferred to Gene-Screen-plus (NEN-Dupont). The filters were hybridized in a buffer containing 50% formamide, 5× SET, 1× Denhardt's, 20 mM PB, 1% SDS, 100 μg/ml salmon sperm DNA, and 3-5 ng/ml of nick-translated probe, at 42°C overnight. The filters were washed at 65°C, first for 3 × 20 min in 4× SET, 0.2% SDS, then for 3 × 20 min in 2× SET, 0.2% SDS, and finally for 3 × 20 min in 0.2× SET, 0.2% SDS. Hybridization was visualized by autoradiography using Kodak AR-5 X-ray film.

### Molecular titration of *Hβ58* transcripts

Transcript prevalence in tissue RNAs from wild-type and *Hβ58* heterozygous mice was determined using a molecular titration assay with an RNA probe (see above), according to the procedures described by Lee and Costlow (1987). Briefly, hybridizations were carried out at 50°C, 0.40 M [Na<sup>+</sup>] in 50% formamide. The quantity of RNA-RNA hybrid in each sample was assayed by digestion with RNase A and RNase T<sub>1</sub> in 0.375 M [Na<sup>+</sup>] followed by precipitation in 5% trichloroacetic acid.

### DNA sequence analysis

DNA fragments derived from the murine cDNA pE30.1, and from chicken cDNA pCK4.4, were subcloned in pBluescript KS(+) Single- or double-stranded clones were employed in sequencing reactions (Sanger et al., 1977) using Sequenase (US Biochemical). The sequence data were analyzed using the program PRONUC (Devereux et al., 1984).

### Sample preparation for *in situ* hybridization

Adult ovaries and testes were obtained from C57BL/6J mice, and embryos from natural matings between inbred FVB/N mice. Midday on the day of the vaginal plug was considered day 0.5 *p.c.* Embryos or tissue samples were fixed at 4°C overnight in freshly prepared 4% paraformaldehyde-1× PBS. The fixed samples were dehydrated in an ethanol series, cleared with xylene, and embedded in paraffin. Sections were cut at 6-8 μm, collected on TESPA-treated slides (Berger, 1986), air dried, and stored at 4°C. Blastocysts, too small to be processed independently, were inserted into the ampulla dissected from a pregnant female (0.5 day *p.c.*) before fixation and embedding. Pretreatment, hybridization, and post-hybridization washes of the slides were performed as described (Toth et al., 1987; Wilkinson and Green, 1990). For autoradiography, the slides were coated with Kodak NTB 2 emulsion, exposed for 1-4 days at 4°C, developed in Kodak D19 and stained with H and E.

### Cross-species genomic blot hybridization

Genomic DNAs were digested with *Eco*RI, fractionated on a

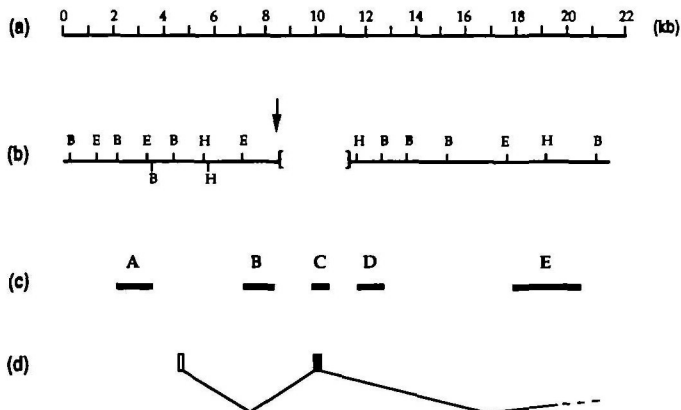
0.7% agarose gel, and transferred to nitrocellulose (Schleicher and Schuell). The blot was hybridized overnight with 10 ng/ml of probe, at 37°C, in a buffer containing 43% formamide, 5× SSC, 20 mM phosphate buffer, 1× Denhardt's, 0.2% SDS, 10% dextran sulphate, 100 µg/ml sonicated, denatured salmon sperm DNA (McGinnis et al., 1984). The blot was washed at 50°C for 4 × 30 min in 2× SSC, 0.2% SDS.

## Results

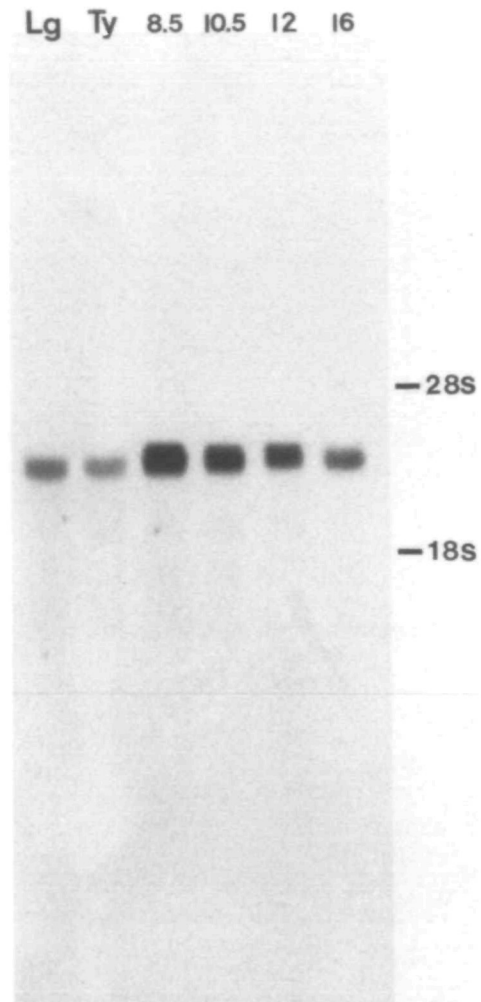
### *cDNA cloning of a mRNA encoded at the Hβ58 locus*

In order to identify and characterize the gene disrupted by the *Hβ58* insertion event (Fig. 1b), we used a genomic fragment from the locus (probe C, Fig. 1c) to screen a cDNA library prepared from 8.5 day *p.c.* mouse embryos (Fahrner et al., 1987). This probe had been previously shown to hybridize to a 2.7 kb mRNA in 8.5 day *p.c.* embryos, while four other probes from the locus, (A,B,D,E in Fig. 1c) failed to hybridize (Radice et al., 1991). Several overlapping cDNA clones were isolated, the longest of which was the nearly full-length (2.6 kb) clone, pE30.1. Hybridization of clone pE30.1 to restriction digests of genomic λ-clones from this region, as well as sequence analysis of the cDNA and genomic clones, indicated that the cDNA included sequences both 5' and 3' to the inserted transgene array, as well as within the deleted region (Fig. 1d, data not shown). Therefore, the cDNA represented a gene that had been structurally disrupted by the *Hβ58* mutation, which we will refer to as the *Hβ58* gene.

To examine initially the expression of this gene during normal murine embryogenesis, we used cDNA pE30.1 to probe Northern blots containing poly(A)<sup>+</sup> RNA from whole, normal mouse embryos at days 8.5, 10.5, 12.5 and 16.5 *p.c.*, and several adult tissues. A single band of 2.7 kb was observed, whose absolute



**Fig. 1.** Organization of the *Hβ58* locus and transcription unit. (a) Scale in kb. (b) Restriction enzyme map of the mutant *Hβ58* locus, showing the site of the transgene insertion (arrow) and the approximate extent of a deletion (brackets). B, *Bam*HI; E, *Eco*RI; H, *Hind*III (c) Single copy probes used to search for a transcript encoded at the locus (Radice et al., 1991) (d) Splicing patterns of the 2.7 kb mRNA, indicating the 5' untranslated regions (open box) and partial coding region (solid box).

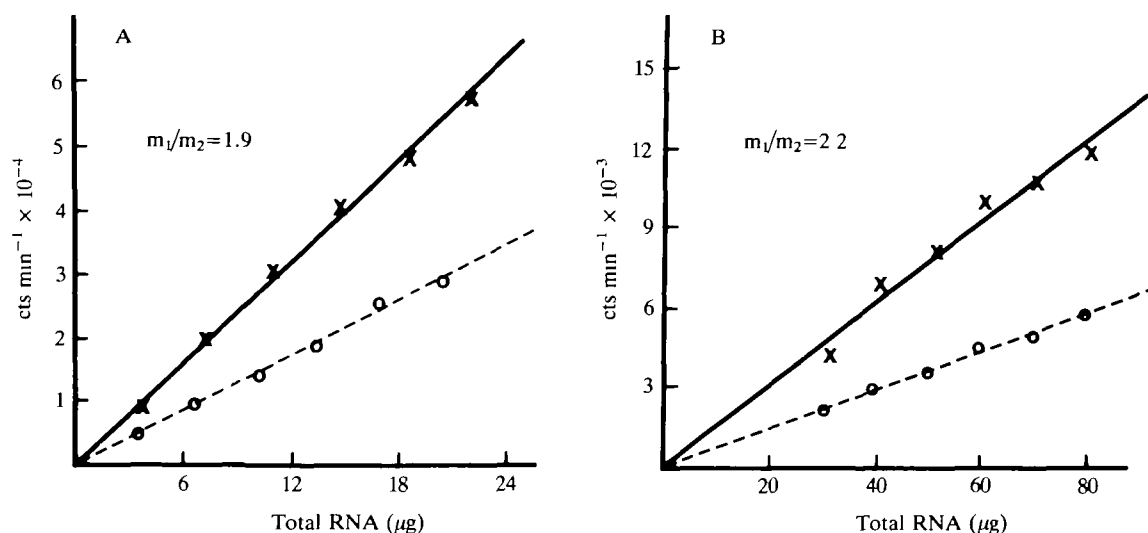


**Fig. 2.** Northern blot analysis of *Hβ58* gene expression in adult tissues and post-gastrulation embryos/fetuses. Analysis of poly(A)<sup>+</sup> RNA from whole embryos at 8.5, 10.5, 12.5 and 16.5 days *p.c.*, as well as lung (Lu) and thymus (Ty) of adult mice. 10 µg of poly(A)<sup>+</sup> RNA was used in each lane, and equal loading of RNAs was verified by ethidium bromide staining (data not shown). The probe was the <sup>32</sup>P-labeled insert DNA of cDNA pE30.1. The positions of 18S and 28S ribosomal RNAs are indicated.

level steadily decreased during fetal development (Fig. 2), but which continued to be present in all somatic tissues of adult mice (Fig. 2 and data not shown).

### *A two-fold reduction in the level of Hβ58 transcripts in mice heterozygous for the insertional mutation*

To confirm at the RNA level that the expression of the *Hβ58* gene was blocked by the insertional mutation, we measured the abundance of the mRNA in several tissues of heterozygous *Hβ58* mice compared to wild-type mice. We expected that the prevalence of *Hβ58* transcripts in tissues of heterozygous mutant mice would be approximately half the wild-type level, in the absence of compensation for the loss of one allele. Probe excess hybridizations (Lee et al., 1986) were carried out using an anti-sense RNA probe derived



**Fig. 3.** Comparison of the level of *Hβ58* transcripts in wild-type (+/+) vs. heterozygous (*Hβ58*/+) mouse tissues by RNA titration analysis. An RNA probe representing the 3'-untranslated region (see Materials and methods) was utilized in probe excess titrations with total testis RNA (A) or kidney RNA (B), from wild-type mice (x, solid lines) or mice heterozygous for the *Hβ58* insertion (o, dashed lines). The probe was hybridized with increasing amounts of wild-type or heterozygous RNA (abscissa), treated with RNase, and the amount of RNase-resistant probe, measured in cts/min, is plotted. The slopes of the lines were derived by linear least-squares regression analysis

from a subclone of the cDNA pE30.1, and varying amounts of total RNA from adult kidney or testis. The data are plotted in Fig. 3 as RNase-resistant probe cts/min vs. input RNA, where the slope of each line is proportional to the abundance of the *Hβ58* transcripts. For each tissue, the *Hβ58* RNA was approximately two-fold less abundant in heterozygous than in wild-type mice (testis: 1.9; kidney: 2.1). This quantitative reduction in the level of *Hβ58* transcripts implied that the structural alterations at the *Hβ58* locus either blocked transcription or interfered with normal post-transcriptional processing, resulting in the failure of the mutant allele to produce significant levels of the mature mRNA.

#### *Sequence of embryonic cDNA pE30.1 and its potential protein*

Having concluded that cDNA clone pE30.1 represented a gene that had been disrupted by the *Hβ58* insertional mutation, we determined the nucleotide sequence of the 2600 bp cDNA insert (Fig. 4). The sequence contained an open reading frame of 1073 nt (327 amino acids) ending with a COOH-terminal methionine, and followed by multiple stop codons in all reading frames. To verify that we had identified the correct open reading frame, we synthesized *in vitro* a sense-strand RNA, translated it *in vitro* and analyzed the products by SDS-PAGE. The size of the major polypeptide produced was consistent with the predicted size of  $38 \times 10^3 M_r$  (data not shown).

Secondary structure and hydropathicity analysis of the prospective protein showed no identifiable structures such as signal peptides or transmembrane regions. Homology searches of the Genbank and NBRF data bases with the nucleotide and predicted amino acid

sequences of the pE30.1 cDNA clone did not reveal any significant similarities with known genes. In addition, searches for specific conserved motifs characteristic of known classes of proteins (e.g. DNA-binding domains, ligand binding sites, etc.) were negative. We concluded that the protein potentially encoded by the embryonic *Hβ58* mRNA was a novel protein, lacking recognizable structures that might provide insight into its biochemical properties or subcellular location.

#### *Detection of Hβ58 transcripts in ovarian oocytes, eggs and pre-implantation embryos*

While Northern blot analysis (Fig. 2) established that the *Hβ58* gene was active as early as 8.5 days *p.c.*, these data did not indicate when the gene was first expressed, or determine its spatial pattern of expression in the embryo. To address these issues, we performed *in situ* hybridization studies, using an <sup>35</sup>S-labelled probe representing the complete cDNA pE30.1 (Fig. 5A). Analysis of adult ovaries (Fig. 5B-C) revealed that the *Hβ58* mRNA was present at high levels in oocytes of all stages, while the ovarian stroma and follicle cells contained only low, uniform levels of the mRNA. The RNA was still present at high levels in the mature oocyte (Fig. 5C), as well as in the ovulated egg (Fig. 5D). In the 3.5 day *p.c.* blastocyst (Fig. 5E-F), *Hβ58* transcripts were readily detected in both the inner cell mass and trophectoderm. While we cannot rule out the possibility that *Hβ58* transcripts in the blastocyst include maternal mRNA inherited from the egg, they most likely represent newly synthesized mRNA (see Discussion). If true, this would indicate that the *Hβ58* gene is already transcriptionally active during pre-implantation development.

```

1041 - GAC TCG CAG GCC TCT GCG GAG CAG CCT GAG ATG TAA      gcaggtgggcagagcag      - 327
      Asp Ser Glu Ala Ser Ala Glu Gln Pro Glu Met ---
1094 - caagtgcgaagaccggggccagagagcagctcgctggcgtagcagcagctgcagcgtgcggacatcc
1161 - caactccacgtctgctagctctccttatcctagagtgctcggttattttacagtatgattaagggt
1228 - ttgtgtataacattctaaagaagtctcttggaaacctggagaattctcaagtcgcttttttg
1295 - tttcttttttactgcacattgaagcactcagaagcataaacattaaaggattttctagtgtagtag
1362 - actcgttatttgttaattttgctttctttatgtcataatgttgaaaaacctttctcgtttttcatgcta
1429 - atgattaccctctattctctatgtgtcaaaaaataaaatttttggtaagaataatgaaaaaac
1496 - tgagcagcccttatgttccatacaagttttaaggcatggctcctcagtatttttgagaatttgtgttta
1563 - ttttacctgttgatcagaacaaacctgggtgttacctatggttaggcattctgggtctaccattcaca
1630 - gctcctgaatttatactcttgggaactctcgtaggattatgaagaatgtgcaccacaggatgacatg
1697 - gtttgtcactgctggccctcttgcgccctatttgaccacgcttggctatgactctctgtatcatcatgta
1764 - gttctgcagtggtggaagtgcacattggcattttagaactgtttaaacctgtatcatttttttcc
1831 - ataaagactttgaaacatgtatatattttttaatttaaggactactttccatgtgtatgc aaatcgt
1898 - cacctaggaatgagagagcttggggagaaccgcggaagtctccatagccggcaaatctctgttgac
1965 - tggaaatttgtgctagttcacaactgacttcacgtacgttctatttataaaaccgcaaccaacacgtga
2032 - taaaaacactaaattcataccattctctgaaaagaaggctgcagtcagctagcacagacacacaac
2099 - tgaggcgctatggggcagaggaagcctgtgtgggggccactctcactcagggcttggtgggttctgtg
2166 - accagttgttctctgataccactcagtcacaggtttgtcagtttaacttactttcaggaagtctaca
2233 - acccaactctatgtttcttttaaacctatttttctactatgtggcaatttgataaaatggccagcaacc
2300 - agccctggccaaagggttcacataccacgtctcggagcagctatgtcgaattgtgggttctctaaggagc
2367 - gccatgtttctccttcatgtaagtctcctgttcagagctctcaaggtttcaaaatgccaaatattttcat
2434 - gttcacttgcagtgtataatctaggaatattcagagagatttggaanaacaaattgattttaaccaggc
2501 - ctcaagttgtgtcacaacctgatataataaataaataattggaacagaaaaaataaaaaa

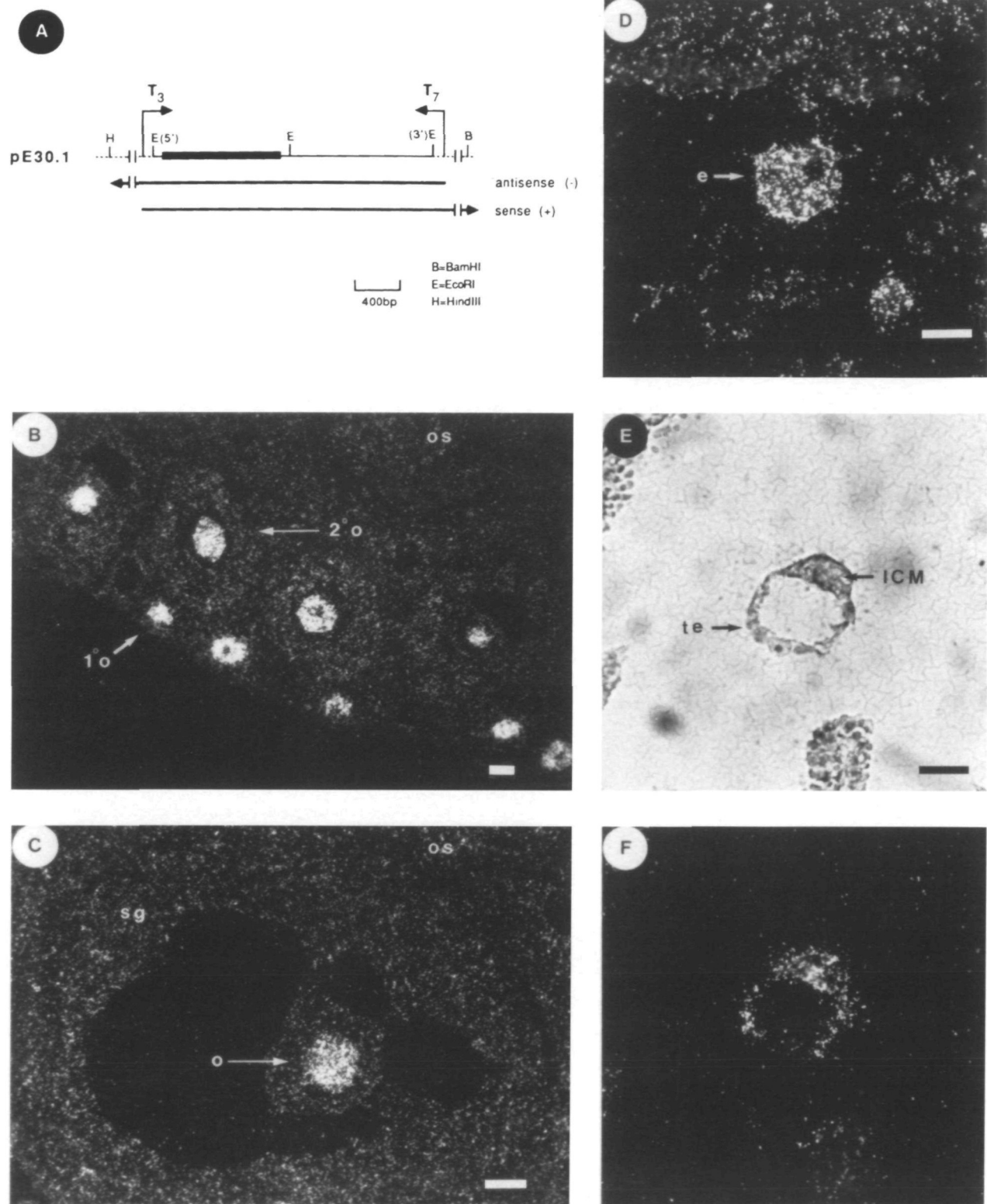
```

**Fig. 4.** Nucleotide and deduced amino acid sequence of the murine cDNA pE30.1. The 5' and 3'-untranslated regions are shown in lower case letters, and the open reading frame in upper case letters, with the inferred amino acid sequence below. The 5' to 3' orientation was determined by solution hybridization with single-stranded RNA probes representing either strand of DNA (data not shown). Nucleotides are numbered on the left and amino acids on the right. The poly(A) addition signal, AATAAA, is underlined.

### *Spatial distribution of H $\beta$ 58 transcripts in post-implantation embryos*

Figs 6 and 7 illustrate the accumulation of *Hβ58*

transcripts in the embryonic and extraembryonic lineages during post-implantation stages. At the egg cylinder stage (day 6.0 and day 6.5 *p.c.*, Fig. 6A-D), the



embryonic ectoderm contained a low but detectable level of *Hβ58* RNA, and the extraembryonic ectoderm displayed a slightly higher level. The RNA level in the embryonic ectoderm remained low and increased only slightly during the gastrulation stage (day 7.0 and 7.5, Fig. 6E-H). However, progressively higher levels of *Hβ58* mRNA were seen in several extraembryonic

tissues, including the amniotic fold at day 7.0, the chorion and the ectoplacental cone at day 7.0-7.5, and most dramatically in the visceral endoderm from day 6.5 onward.

At the early somite stage (day 8.5 *p.c.*), all three germ layers of the embryo proper displayed a relatively low and spatially uniform level of *Hβ58* mRNA (Fig.



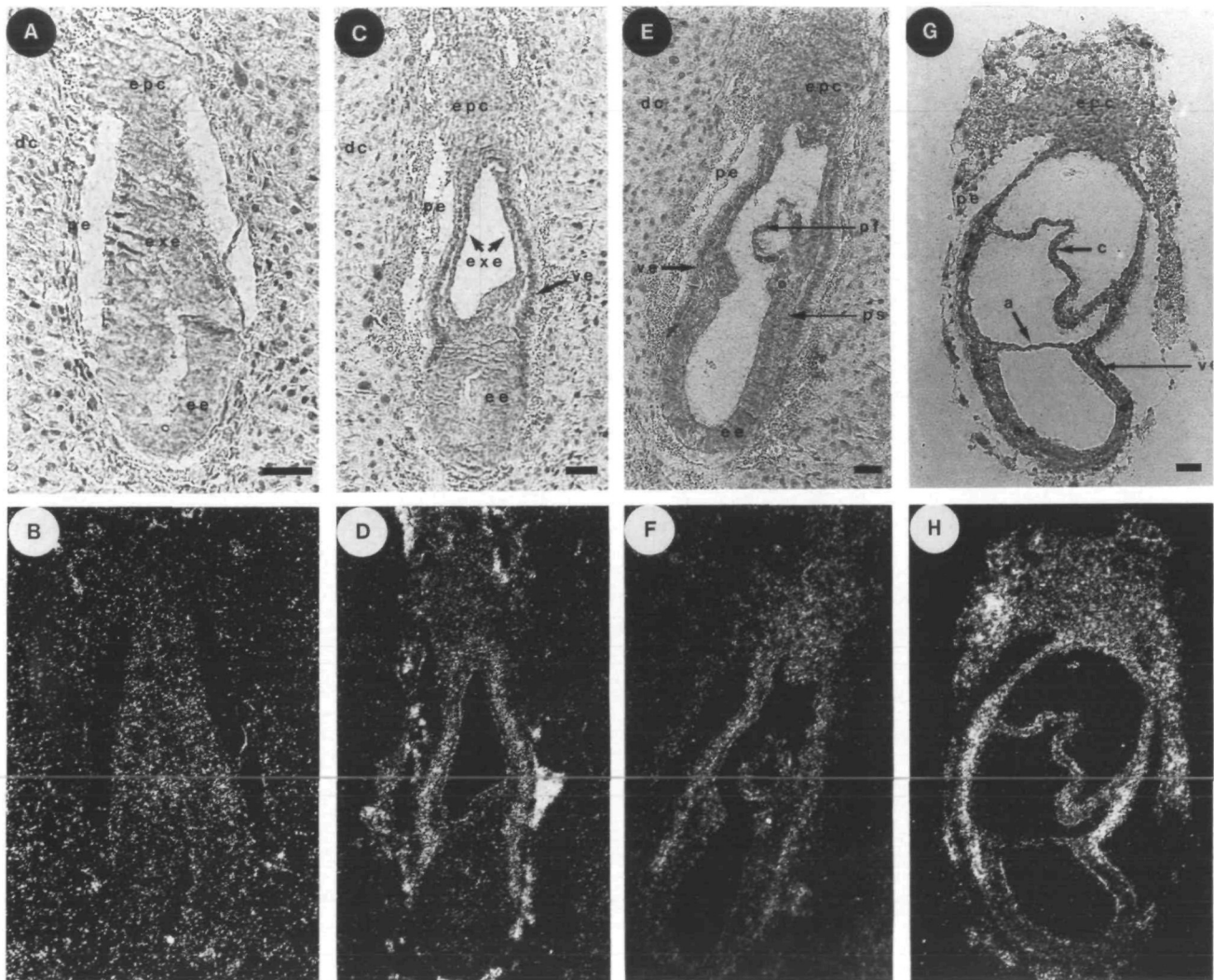
**Fig. 5.** Localization of *Hβ58* transcripts in developing oocytes and preimplantation embryos. (A) Derivation of RNA probes from the *Hβ58* cDNA clone pE30 1, which includes the entire 3'-untranslated and partial 5'-untranslated regions (thin lines) and the coding region (solid box) of the *Hβ58* mRNA sequence. Dashed lines denote vector sequences. The anti-sense probe, synthesized from the T<sub>7</sub> promoter (see Materials and Methods), was used for all *in situ* hybridizations shown. No specific signals were detected when the corresponding sense probe (synthesized from the T<sub>3</sub> promoter) was hybridized to

adjacent sections (data not shown). (B) *In situ* hybridization to a section of adult ovary, showing ovarian stroma (os), primary (1°o) and secondary (2°o) oocytes. (C) Higher magnification view of a mature follicle showing mature ovum (o) and stratum granulosum (sg). (D) Section through an ovulated egg (e) in the ampulla. (E and F) The same section through a day 3.5 blastocyst, showing inner cell mass (ICM) and trophoctoderm (te). Panels B, C, D and F were photographed under dark-field, and panel E under bright-field illumination. Bars, 60 μm.

7A-B), as did most extraembryonic tissues including the allantois (Fig. 7A-B), amnion (Fig. 7C-D), and blood islands (Fig. 7C-D). However, the visceral yolk sac (previously the visceral endoderm) and the ectoplacental cone continued to accumulate relatively high levels of *Hβ58* transcripts (Fig. 7A-D).

During organogenesis (day 9.5 *p.c.*), distinct quanti-

tative modulations of *Hβ58* expression became visible in both embryonic and extraembryonic lineages (Fig. 7E-F). In addition to the visceral yolk sac, the region of the chorioallantoic placenta immediately adjacent to the proliferating allantois displayed a very high level of *Hβ58* transcripts. The embryonic tissues displayed an increase in the average level of the mRNA, and



**Fig. 6.** Distribution of *Hβ58* transcripts in the early post-implantation embryo. Sagittal sections through embryos at day 6.0 (A and B), day 6.5 (C and D), and day 7.0 *p.c.* (E and F). (G and H) Transverse section through an embryo at day 7.5 *p.c.*; (A, C, E, and G) bright-field illumination; (B, D, F, and H) dark-field illumination. dc, decidual tissue; epc, ectoplacental cone; pe, parietal endoderm; exe, extraembryonic ectoderm; ve, visceral endoderm; ps, primitive streak; a, amnion; ee, embryonic ectoderm; pf, proamniotic fold; c, chorion. Bar, 60 μm.

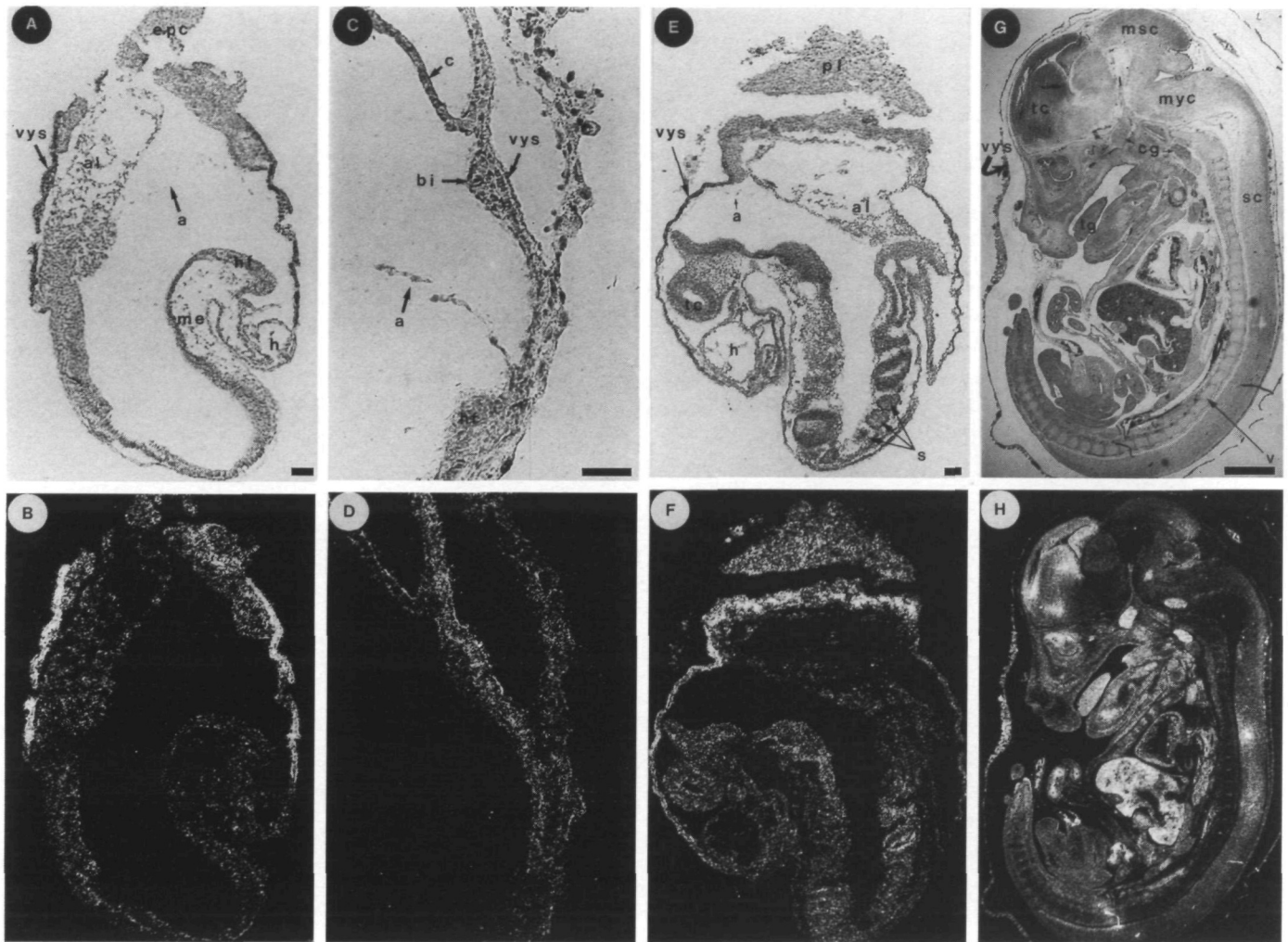


Fig. 7. Distribution of *Hβ58* transcripts during early organogenesis and mid-gestation development (A and B) Mid-sagittal section through an 8.5 day *p.c.* embryo. (C and D) Transverse section through the anterior portion of an 8.5 day embryo. (E and F) Mid-sagittal section through a 9.5 day embryo. (G and H) Mid-sagittal section through a 15.5 day fetus. epc, ectoplacental cone; vys, visceral yolk sac; al, allantois; a, amnion; me, mesenchyme; hf, head-fold; h, heart; c, chorion; bi, blood island; s, somite; pl, placenta; tc, telencephalon; v, vertebra; sc, spinal chord; msc, mesencephalon, myc, myelencephalon. Bars in A, C, E, 70  $\mu\text{m}$ ; Bar in G, 1000  $\mu\text{m}$ .

quantitative regional differences could be seen for the first time.

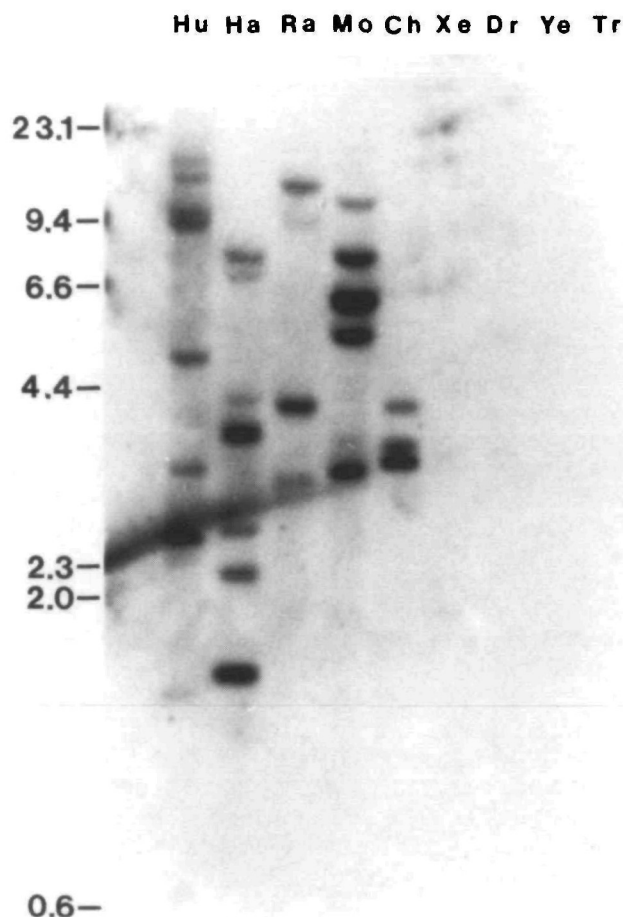
At a later stage of fetal development (day 15.5 *p.c.*, Fig. 7G-H), the *Hβ58* gene appeared to be expressed ubiquitously, although the mRNA level varied considerably between different tissues. The highest levels were seen in several organs derived from diverse embryonic cell lineages, including the telencephalon, cranial ganglia, tongue, liver, and gut.

#### *The Hβ58 protein is highly conserved between mammals and birds*

To assess the degree to which the *Hβ58* gene had been conserved in evolution, we used the murine cDNA as a probe in low-stringency Southern hybridization exper-

iments with genomic DNAs of diverse species, representing four classes of vertebrata as well as invertebrate and single-cell species. As shown in Fig. 8, strong cross-hybridization was detected with DNA from all mammals examined (human, hamster, and rat), as well as the avian (chicken). However, the hybridization signal fell considerably in the amphibian, *Xenopus laevis*, and no reproducible signal was observed with any of the invertebrate or single-cell species. To further define the pattern of sequence conservation between mammals and birds, we cloned and sequenced a cDNA representing a chicken homolog of the murine *Hβ58* gene. The predicted chicken amino acid sequence was remarkably similar to the murine sequence, with identity at 314 out of 326 residues, and conservative substitutions at most other residues (Fig. 9).





**Fig. 8.** Phylogenetic conservation of *Hβ58* cDNA sequences. Sequence homology between the murine cDNA pE30.1 and genomic DNA of different organisms was assayed by Southern blot hybridization. The amount of DNA from each organism was adjusted to correct for genome size, and contained the genomic equivalent of 10  $\mu$ g of mouse DNA. The organisms, chosen from diverse phylogenetic lineages, were human (Hu), hamster (Ha), rat (Ra), mouse (Mo), chicken (Ch), *Xenopus laevis* (Xe), *Drosophila melanogaster* (Dr), *Saccharomyces cerevisiae* (Ye), and *Trypanosoma brucei* (Tr). The hybridization and washing conditions were adjusted to allow cross-reaction with weakly homologous sequences (McGinnis et al., 1984; see Materials and Methods). The sizes in kb are indicated at left.

## Discussion

Mouse embryos homozygous for the *Hβ58* transgene insertion display a characteristic set of abnormalities beginning at the primitive streak stage (day 7.5 *p.c.*), they undergo limited development to the early somite stage, and die shortly thereafter. We have proposed that these developmental defects are due to the disruption of an essential gene at the locus of the insertion on chromosome 10 (Radice et al., 1991). In this paper, we have described the identification of a gene at the *Hβ58* locus, which is normally expressed

1	MSFLGGFFGPVCEIDVILNDAETRKTAEMKTEDGKVEKHFLFYDGESVSG	50
	MSFLGGFFGPICEIDVALNDGETRKMARMKTEDGKVEKHFLFYDGESVSG	
51	KVNVS-KQQGKRLEHQGIRIEFVGQIELFNDKSNTHFVNLVKELALPGE	99
	KVNLAFAKQPGKRLEHQGIRIEFVGQIELFNDKSNTHFVNLVKELALPGE	
100	LTQSRSDYDFEFMQVEKPYESYIGANVRLRYFLKVTIVRRLSDIVKEYDLI	149
	LTQSRSDYDFEFMQVEKPYESYIGANVRLRYFLKVTIVRRLTDLVKEYDLI	
150	VHQLATYDPVNNISIKMEVGIEDCLHIEFEYNKSKYHLKDVIVGKIYPLL	199
	VHQLATYDPVNNISIKMEVGIEDCLHIEFEYNKSKYHLKDVIVGKIYPLL	
200	RIKIQHMELQLIKKEITGIGPSTTTTETETIAKYEIMDGAPVKGESIPRL	249
	RIKIQHMELQLIKKEITGIGPSTTTTETETIAKYEIMDGAPVKGESIPRL	
250	FLAGYDPTPTMRDVNKKFSVRVFLNLVLVDEEDRRYFKQQEITLWRKAPE	299
	FLAGYDPTPTMRDVNKKFSVRVFLNLVLVDEEDRRYFKQQEITLWRKAPE	
300	KLRKQRTNPHQRFPESPDQSASAEQPEM	326
	KLRKQRTNPHQRFPESPDQSASAEQPEM	

**Fig. 9.** Comparison of the deduced amino acid sequences of the mouse embryo cDNA pE30.1 (top) and a homologous chick brain cDNA (bottom). The numbers shown refer to the mouse amino acid sequence. Amino acid identity between the two sequences is shown as a vertical line, while double dots, single dots, and no dots between corresponding amino acids represent decreasing levels of conservation, determined by the algorithm of Gribskov and Burgess (1986). At the nucleic acid level, the chicken and mouse cDNA sequences were 85% identical in the coding region and 40% identical in the 3' untranslated region (data not shown), while the complete 5' untranslated region of the chicken cDNA was not available for comparison.

during embryogenesis but has been physically and functionally disrupted by the transgene insertion. Sequence analysis of cDNA clones revealed that the wild-type *Hβ58* gene encodes a previously unidentified  $38 \times 10^5 M_r$  protein that is not obviously related to any known protein, but which has been strongly conserved in evolution among mammals and birds. In order to compare the normal expression pattern of the *Hβ58* gene with the phenotypic defects seen in mutant embryos, we have characterized the spatial and temporal pattern of *Hβ58* mRNA accumulation during normal mouse development. Although the earliest phenotypic effects of the mutation are seen in the embryonic ectoderm of the early post-implantation embryo, *in situ* hybridization analysis showed that the normal *Hβ58* gene is expressed throughout pre- as well as post-implantation development, and that the highest levels of the mRNA are found not in the embryonic ectoderm but in the visceral endoderm and other extraembryonic lineages.

Several lines of evidence establish that the gene that we identified has been inactivated by the transgene insertion. First, the transcription unit encoding the *Hβ58* mRNA spans the insertion site, with its 5' untranslated region located upstream of the insertion, its first coding exon deleted from the mutant chromosome, and its remaining exons located downstream from the insertion site (Fig. 1; Lee et al., manuscript in

preparation). Thus, the mutant *Hβ58* gene is incapable of encoding the normal 2.7 kb mRNA and  $38 \times 10^3 M_r$  protein product. To confirm the disruption of the *Hβ58* gene at the RNA level, we measured the mRNA abundance in tissues of heterozygous mutant mice (which are phenotypically normal) compared with those of wild-type animals, and found the expected two-fold reduction. In addition, we have observed that embryonic stem cell lines homozygous for the *Hβ58* mutation (Robertson et al., 1992) contain no detectable transcripts from the *Hβ58* gene, although the 2.7 kb mRNA is abundant in wild-type ES cells (unpublished data). We therefore conclude that the synthesis of stable mRNA from the *Hβ58* gene has been extinguished by the insertional mutation, and that this gene is likely to be responsible for the mutant phenotype.

The *Hβ58* gene was found to be expressed in normal embryos, at least at the RNA level, well before any abnormalities become apparent in homozygous mutant embryos, where the gene product is absent. The *Hβ58* mRNA detected in normal blastocysts and egg cylinder stage embryos is unlikely to represent maternal mRNA, most of which is destroyed at the two-cell stage (Piko and Clegg, 1982). While the possibility cannot yet be excluded that the normal development of homozygous mutant embryos for the first 6-7 days of embryogenesis is dependent on *Hβ58* mRNA or protein inherited from the oocyte, such maternal rescue seems unlikely due to relatively large size and cell number of the 6.5 day embryo (~1000 cells; Snow, 1977) compared to the oocyte. Therefore, our results may imply that the *Hβ58* gene product is redundant with other gene products (zygotic or maternal in origin) at early stages, but unique and therefore essential after the egg cylinder stage.

As the earliest defect seen in the homozygous mutant embryo was a retardation in growth of the embryonic ectoderm beginning at day 7.5 *p.c.*, it was surprising to find that the normal *Hβ58* gene was expressed at much lower levels in this cell lineage than in several extraembryonic lineages, from day 6.5 onward. It also appeared paradoxical that the visceral endoderm/visceral yolk sac lineage, which in normal 6.5-8.5 day embryos contained the highest levels of *Hβ58* mRNA, was unaffected morphologically by the mutation (Radice et al., 1991). However, given the established nutritive role, and possible inductive role, of the visceral endoderm (Beck et al., 1967; Azar and Eyal-Giladi, 1981; Mitrani et al., 1983), it is interesting to speculate that the normal expression of the *Hβ58* gene product in this extraembryonic tissue may be required for proper development of the embryonic ectoderm. This would be easiest to imagine if the *Hβ58* gene product were a secreted protein, although it lacks a recognizable signal sequence. Alternatively, the *Hβ58* protein may function at an earlier step in the synthesis, transport, or secretion of some other molecule, and thus indirectly affect the ability of the visceral endoderm to support embryonic growth.

While the lack of correspondence between the major embryonic sites of *Hβ58* gene expression and the

earliest phenotypic effects of the mutation was surprising, there have recently been several reports of highly tissue-specific phenotypes resulting from mutations in genes that are rather widely expressed (e.g., *c-src*, Soriano et al., 1991; *c-abl*, Schwartzberg et al., 1989; *Wnt-1*, McMahon and Bradley, 1990). Nevertheless, it is important to consider the evidence that the *Hβ58* gene is a single copy gene, and that the hybridization probes that we have used are specific for transcripts from the *Hβ58* locus. First, Southern hybridization experiments using genomic probes (Radice et al., 1991) or cDNA probes (Fig. 8 and unpublished data) provided no indication that there are closely related sequences elsewhere in the genome. Similarly, genetic mapping studies using a cDNA fragment as a probe indicated a unique locus on chromosome 10 of *Mus musculus* (Justice et al., 1990). Second, Northern blot experiments detected only a single 2.7 kb mRNA in adult somatic tissues and in whole embryos (Fig. 2 and unpublished data), and this transcript was absent in homozygous mutant embryonic stem cells (unpublished data). Therefore, it seems highly unlikely that the pattern of expression defined by our *in situ* hybridization experiments includes transcripts from a different but related gene.

While the widespread expression of the *Hβ58* gene in many differentiated tissues of fetal and adult mice argues that the gene does not function solely in early embryogenesis, neither is the data consistent with a "housekeeping" role for the gene. Homozygous mutant embryos survive for 10-11 days *p.c.*, developing a variety of organized tissues, and several terminally differentiated cell types (Radice et al., 1991). Furthermore, embryonic stem cell lines derived from homozygous *Hβ58* embryos grow well in culture, and can contribute extensively to the tissues of chimeric fetuses (Robertson et al., 1992). Therefore, the *Hβ58* gene product is apparently not required for indefinite cell proliferation *in vitro* or extensive cell differentiation *in vivo*. We have also observed that the *Hβ58* gene is highly expressed in male germ cells, where it uses an alternative promoter and alternative splicing to produce a larger mRNA, potentially encoding a  $42 \times 10^3 M_r$  protein (Lee et al., manuscript in preparation). Therefore, this alternate form of the *Hβ58* gene product may play a role in spermatogenesis.

Cross-hybridization studies revealed sequences closely related to the murine *Hβ58* gene in the DNA of other mammals as well as the chicken. The hybridization signal, however, decreased considerably using DNA from the amphibian *X. laevis*, and was not detectable in invertebrates. The analysis of a homologous chicken cDNA revealed a remarkable degree of conservation (>96% identity) between the murine and chicken amino acid sequences. In contrast, most proteins show only 60-80% similarity between mouse and chicken (R. Doolittle, personal communication). High levels of sequence conservation, however, are typically seen in structural proteins such as actins (Vandekerckhove and Weber, 1984), tubulins (Alexandraki and Ruderman, 1983) and histones (Isenberg,

1979). Whereas the extreme conservation of structural proteins extends throughout the animal kingdom, the H $\beta$ 58 protein is likely to be far less conserved in the amphibian *X. laevis* based on the low extent of DNA cross-hybridization with the mouse cDNA. It is of interest that both avians and mammals are amniotes, whose embryos utilize extraembryonic membranes to interact with the external environment (Gauthier et al., 1988), while amphibians are non-amniotes, whose embryos interact directly with their environment. Given the observation that the highest levels of expression of the H $\beta$ 58 gene were in the extraembryonic tissues of the 6.5 to 8.5 day *p.c.* conceptus, as well as the abnormalities of the amnion and chorion of several homozygous mutant 7.5 day embryos (Radice et al., 1991), we suggest that the strong evolutionary conservation of the gene among amniotes may reflect a role of the protein in the functions of extraembryonic membranes.

We thank E. Robertson and C. Perez-Stable for their comments on the manuscript. This work was supported by a Postdoctoral Fellowship (to J. J. L.) and a Faculty Research Award (to F. C.) from the American Cancer Society, and by grants from the National Foundation March of Dimes and the NIH (to F. C.).

## References

- Alexandraki, D. and Ruderman, J. V. (1983). Evolution of  $\alpha$ - and  $\beta$ -tubulin genes as inferred by the nucleotide sequences of sea urchin cDNA clones *J Mol Evol* 19, 397-410.
- Auffray, C. and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. *Eur J Biochem* 107, 303-314.
- Aviv, H. and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid cellulose *Proc natn Acad Sci U S A* 69, 1408-1412.
- Azar, Y. and Eyal-Giladi, H. (1981). Interaction of epiblast and hypoblast in the formation of the primitive streak and the embryonic axis in chick, as revealed by hypoblast rotation experiments *J Embryol Exp Morphol* 61, 133-144.
- Beck, R., Lloyd, J. B. and Griffiths, A. (1967). A histochemical and biochemical study of some aspects of placental function in the rat using maternal injection of horseradish peroxidase *J. Anat* 101, 461-478.
- Berger, C. N. (1986). *In situ* hybridization of immunoglobulin-specific RNA in single cells of the B lymphocyte lineage with radiolabelled DNA probes *EMBO J* 5, 85-93.
- Costantini, F., Radice, G., Lee, J. J., Chada, K. K., Perry, W. and Son, H. J. (1989). Insertional mutations in transgenic mice *Prog Nucl Acids Res Molec Biol* 36, 159-169.
- Costantini, F., Radice, G., Magram, J., Stamatoyannopoulos, G., Papayannopoulou, Th. and Chada, K. (1985). Developmental regulation of human globin genes in transgenic mice. *Cold Spring Harbor Symp Quant Biol* 50, 361-370.
- Cox, K. H., DeLeon, D. V., Angerer, L. M. and Angerer, R. C. (1984). Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes *Dev Biol* 101, 485-502.
- Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acid Res.* 12, 387-395.
- Fahrner, K., Hogan, B. L. M. and Flavell, R. A. (1987). Transcription of H-2 and Qa genes in embryonic and adult mice *EMBO J.* 6, 1265-1271.
- Gauthier, J. A., Kluge, A. G. and Rave, T. (1988). Amniote phylogeny and the importance of fossils *Cladistics* 4, 2-18.
- Gribskov, M. and Burgess, R. R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T<sub>4</sub> are homologous proteins *Nucleic Acid Res* 14, 6745-6763.
- Gridley, T. (1991). Insertional versus targeted mutagenesis in mice *The New Biologist* 3, 1025-1034.
- Gridley, T., Jaenisch, R. and Gendron-Maguire, M. (1991). The murine *Mov-34* gene: full-length cDNA and genomic organization *Genomics*, in press.
- Gridley, T., Soriano, P. and Jaenisch, R. (1987). Insertional mutagenesis in mice *Trends Genet* 3, 162-166.
- Isenberg, I. (1979). Histones *Ann Rev. Biochem.* 48, 159-191.
- Justice, M. J., Siracusa, L. D., Sillan, C. M., Copeland, N. G. and Jenkins, N. A. (1990). A multilocus genetic linkage map of mouse chromosome 10 from interspecific backcross analysis *Genetics* 125, 855-866.
- Lee, J. J., Calzone, F. J., Britten, R. J., Angerer, R. C. and Davidson, E. H. (1986). Activation of sea urchin actin genes during embryogenesis. Measurement of transcript accumulation from five different genes in *Strongylocentrotus purpuratus* *J Mol Biol* 188, 173-183.
- Lee, J. J. and Costlow, N. (1987). A molecular titration assay to measure transcript prevalence levels *Meth Enzymology* 152, 633-648.
- Maas, R. L., Zeller, R., Woychik, R. P., Voght, T. F. and Leder, P. (1990). Disruption of formin-encoding transcripts in two mutant limb deformity alleles *Nature* 346, 853-855.
- Magnuson, T. (1986). Mutations and chromosomal abnormalities: How are they useful for studying genetic control of early mammalian development. In *Experimental Approaches to Early Embryonic Development* (Rossant, J. and R. A. Pederson, eds.), Cambridge University Press.
- McGinnis, W., Hart, C. P., Gehring, W. J. and Ruddle, F. H. (1984). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of drosophila *Cell* 38, 675-680.
- McMahon, A. P. and Bradley, A. (1990). The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain *Cell* 62, 1073-1085.
- Mercer, J. A., Seperack, P. K., Strobel, M. C., Copeland, N. G. and Jenkins, J. A. (1991). Novel myosin heavy chain encoded by murine *dilute* coat color locus *Nature* 349, 709-713.
- Mitrani, E., Shimoni, Y., and Eyal-Giladi, H. (1983). Nature of the hypoblastic influence on the chick embryo epiblast. *J Embryol Exp Morphol* 75, 21-30.
- Palmiter, R. D. and Brinster, R. L. (1986). Germ-line transformation of mice *Ann Rev Genet* 20, 465-499.
- Piko, L. and Clegg, K. B. (1982). Quantitative changes in total RNA, total poly (A), and ribosomes in early mouse embryos. *Dev Biol* 89, 362-378.
- Posakony, J. W., Flytzanis, C. N., Britten, R. J. and Davidson, E. H. (1983). Interspersed sequence organization and developmental representation of cloned poly (A) RNAs from sea urchin eggs *J Mol Biol* 167, 361-389.
- Radice, G., Lee, J. J. and Costantini, F. (1991). H $\beta$ 58, an insertional mutation affecting early postimplantation development of the mouse embryo *Development* 111, 801-811.
- Reith, A. D. and Bernstein, A. (1991). Molecular basis of mouse developmental mutants *Genes Dev* 5, 115-123.
- Robertson, E. J., Conlon, F. L., Barth, K. S., Costantini, F. and Lee, J. J. (1992). Using ES cells to study mutations affecting postimplantation development in the mouse. In: *CIBA Symposium #165 "Postimplantation Development in the Mouse"*, in press.
- Rossant, J. and Joyner, A. L. (1989). Towards a molecular-genetic analysis of mammalian development *Trends Genet* 5, 277-283.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors *Proc natn Acad Sci U S A* 74, 5463-5467.
- Schnieke, A., Harbers, K. and Jaenisch, R. (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the  $\alpha 1$  (I) collagen gene *Nature* 304, 315-320.
- Schwartzberg, P. L., Goff, S. P. and Robertson, E. J. (1989). Germ-line transmission of a *c-abl* mutation produced by targeted gene disruption in ES cells. *Science* 246, 799-803.
- Snow, M. H. L. (1977). Gastrulation in the mouse. Growth and

- regionalization of the epiblast. *J Embryol Exp Morphol* **42**, 293-303
- Soriano, P., Gridley, T. and Jaenisch, R.** (1987) Retroviruses and insertional mutagenesis in mice. Proviral integration at the *Mov-34* locus leads to early embryonic death. *Genes Dev* **1**, 366-375
- Soriano, P., Montgomery, C., Geske, R. and Bradley, A.** (1991) Targeted disruption of the *c-src* proto-oncogene leads to osteopetrosis in mice. *Cell* **64**, 693-702.
- Toth, L. E., Slawin, K. L., Pinter, J. E. and Nguyen-Huu, M. C.** (1987). Region-specific expression of mouse homeobox genes in the embryonic mesoderm and central nervous system. *Proc natl Acad Sci U S A* **84**, 6790-6794
- Vandekerckhove, J. and Weber, K.** (1984). Chordate muscle actins differ distinctly from invertebrate muscle actins. *J Mol Biol.* **179**, 391-413
- Weiher, H., Noda, T., Gray, D. A., Sharpe, A. H. and Jaenisch, R.** (1990) Transgenic mouse model of kidney disease. Insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. *Cell* **62**, 425-434.
- Wilkinson, D. G. and Green, J.** (1990) *In situ* hybridization and the three-dimensional reconstruction of serial sections. In *Postimplantation Mammalian Embryos: A Practical Approach* (Copp, A. J. and Cockcroft, D. L. eds) pp 155-171, Oxford University Press, Oxford
- Woychik, R. P., Maas, R. L., Zeller, R., Vogt, T. F. and Leder, P.** (1990). "Formins". Proteins deduced from the alternative transcripts of the *limb deformity* gene. *Nature* **346**, 850-853.
- Woychik, R. P., Stewart, T. A., Lewis, L. G., D'Eustachio, P. and Leder, P.** (1985) An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. *Nature* **318**, 36-40

(Accepted 13 December 1991)