DEVELOPMENTAL DYNAMICS 212:304–317 (1998)

# *Aquarius*, A Novel Gene Isolated by Gene Trapping With an RNA-Dependent RNA Polymerase Motif

MEHRAN SAM,[1,2] WOLFGANG WURST,[3] MICHAEL KLÜPPEL,[1,2] OU JIN,[1,2] HENRY HENG,[4] AND ALAN BERNSTEIN[1,2]*

[1]*The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada*
[2]*Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada*
[3]*GSF-Research Centre, Institute for Genetics, Munich, Germany*
[4]*Department of Biology, York University, Ontario, Canada*

***ABSTRACT*** **In a retinoic acid (RA) gene-trap screen of mouse embryonic stem (ES) cells, a novel gene, named *Aquarius* (*Aqr*), was identified and characterized. The promoterless *lac*Z marker was used to trap the genomic locus and to determine the expression pattern of the gene. *Aqr* transcripts are strongly induced in response to RA in vitro. During embryogenesis, *Aqr* is expressed in mesoderm, in the neural crest and its target tissues, and in neuroepithelium. Expression was first detected at 8.5 days postcoitum, when neural crest cells are visible at the lateral ridges of the neural plate. The gene-trapped *Aqr* locus was transmitted through the mouse germ line in three genetic backgrounds. In the F2 generation, the expected mendelian ratio of 1:2:1 was observed in all backgrounds, indicating that homozygous mice are viable. Homozygotes are normal in size and weight and breed normally. The gene trap insertion, however, does not seem to generate a null mutation, because *Aqr* transcripts are still present in the homozygous mutant animals. The *Aqr* open reading frame has weak homology to RNA-dependent RNA polymerases (RRPs) of the murine hepatitis viruses and contains an RRP motif. *Aqr* was mapped to mouse chromosome 2 between regions E5 through F2 by using fluorescence in situ hybridization analysis.** ***Dev. Dyn. 1998;212:304–317.*** 

**Key words: mouse embryonic stem cell; gene trap; neural crest; mesoderm; Expressed Sequence Tag database**

## INTRODUCTION

Over the next decade, the complete DNA sequence of a mammal will become available. This information will serve as a resource with which to decipher the genetic instructions required for embryological development, organogenesis, and the functioning of various cell types in the adult. The challenge for biology will be to characterize functionally all of the genes involved in these processes and in disease. Over the past few years, various strategies have been developed and refined to clone the genes corresponding to existing mutant or disease phenotypes or to isolate genes that are the mammalian homologs of genes in other organisms that are more amenable to genetic analysis. These approaches can also be coupled with gene targeting in embryonic stem (ES) cells to generate mice with mutations in any desired gene. However, these procedures are expensive, slow, and labor intensive; hence, they are not yet sufficiently robust to apply on a genome-wide level.

Insertional mutagenesis, based on a stochastic insertion of exogenous DNA into the genome of ES cells, provides another approach to generate novel mutations in the mouse germ line. Furthermore, if the insertional construct includes features such as a promoterless *lac*Z cassette, then the inserted DNA provides an easy readout of the normal expression pattern of the "trapped" gene, and partial sequence information of flanking exons can be derived by various polymerase chain reaction (PCR)-based amplification strategies. This process of gene trapping offers a feasible and potentially robust strategy to mutagenize and characterize the entire mouse genome.

Given the very large size of the mammalian genome and large numbers of genes, pseudogenes, and noncoding regions, many of the ES clones that are isolated by a totally nonselective gene trapping strategy will harbor insertions in regions of the genome that are never expressed. Furthermore, large portions of the remaining insertional events are likely to be in genes that may not be of immediate interest. For example, in a recent test for feasibility of large-scale gene trap mutagenesis, Wurst et al. (1995) recovered 279 gene trapped clones in a screen for developmentally regulated genes. Thirty-six clones (13%) exhibited restricted patterns of gene expression in embryonic and extraembryonic tissues, 88 (32%) showed widespread expression, and 155 (55%) failed to show detectable levels of expression at 8.5 days of gestation.

Grant sponsor: Bristol-Myers Squibb; Grant sponsor: National Cancer Institute of Canada (Terry Fox Program Project Grant).

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number U90333.

*Correspondence to: Alan Bernstein, The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Room 982, Toronto, Ontario M5G 1X5, Canada.

Received 13 August 1997; Accepted 4 December 1997

Two approaches have been described that are designed to enrich for particular subsets of trapping events in vitro, prior to the generation of mice. Skarnes et al. (1995) developed a gene-trapping construct that was designed to trap genes that encode secreted and membrane-spanning proteins, such as receptor tyrosine kinases, phosphatases, cadherin, and laminin. The screening strategy was based on an in-frame insertion of the CD4 transmembrane domain with the βgeo reporter gene as the trapping vector, thereby capturing any insertion occurring downstream of a signal sequence.

In another approach, we previously reported identification and preliminary characterization of 20 gene trap integrations in ES cells that were screened in vitro for response to exogenous retinoic acid (RA; Forrester et al., 1996). All but one of these integrations subsequently showed unique and tissue-specific patterns of expression during embryogenesis. Sequence analysis from six integrations revealed five novel genes and one previously identified gene, the protooncogene c-*fyn*. Germ-line transmission and breeding uncovered one homozygous embryonic lethal insertion and three homozygous viable insertions. A novel family of RA-responsive, repetitive sequences was also identified in this RA screen (Sam et al., 1996).

To better understand B-cell development at the molecular level, the genetic response of B-lineage cells to bacterial lipopolysaccharide (LPS; a potent stimulator of B cells) was also examined in a gene trap approach. Novel LPS-responsive genes were identified and shown to have restricted expression within the B-lymphoid lineage (Kerr et al., 1996).

In other gene trap studies, novel genes with developmentally restricted expression patterns have been identified, and it has been shown that their expression is not disturbed by the insertion of the gene trap vector. Loss-of-function phenotype of homozygous mutant embryos has also been characterized. Takeuchi et al. (1995) generated a gene trap mouse mutation, *jumonji*, that leads to defective neural tube closure and embryonic lethality before day 15.5 of gestation. The gene is expressed predominantly in the cerebellum and at the midbrain-hind brain boundary, and its deduced amino acid sequence shares a portion of significant homology with human retinoblastoma-binding protein, RBP-2. Chen et al. (1996) inactivated the murine *Eck* receptor tyrosine kinase by a retroviral gene trap insertion and showed that the expression of the *Eck* promoter was not affected by provirus integration. The function of α-*E-catenin* in mouse preimplantation development was also defined by using a gene trap approach (Torres et al., 1997). It was shown that a loss-of-function mutation in α-*E-catenin* results in disruption of the trophoblast epithelium and a subsequent developmental block at the blastocyst stage.

The experiments described above demonstrate the power of gene trapping to identify and analyze novel genes and their expression and function in the mouse. In this report, we describe the isolation and characterization of the mammalian *Aquarius* gene, which has been identified by using the RA-induction gene trap strategy described above and by using Internet-based database access and resources. We believe that the combined gene trap and database approach has considerable potential in increasing the pace of gene discovery and functional analyses, particularly as the sequence of the mouse genome comes on line.

## RESULTS

The *Aquarius* gene described in this report was identified in a gene trap screen of totipotent mouse ES cells that was designed to capture genes responsive to exogenous RA (Forrester et al., 1996). In this approach, the gene trap construct carrying a splice acceptor site 5′ of the promoterless *lac*Z was integrated randomly into the mouse genome by electroporation. Twenty gene-trapped ES clones were isolated that were either induced or repressed after 48 hr of exposure to RA. This paper describes the cloning of the *Aqr* gene from one such clone, I193. This ES clone was of particular interest to us, because its *lac*Z was strongly induced in response to RA. RA and its derivatives regulate important biological processes, including cell proliferation, differentiation, and morphogenesis, in a variety of developmental systems (Mangelsdorf et al., 1995).

### Cloning of the RA-Responsive *Aquarius* Gene

To examine RA responsivity, I193 cells were grown in the presence or absence of leukemia inhibitory factor (LIF), which maintains ES cells in an undifferentiated state, and in the absence of LIF and in the presence of RA for 48 hrs. The cells were then stained for *lac*Z activity. Figure 1A–C shows that the *lac*Z reporter gene was strongly induced in the I193 cell line in response to RA, indicating the RA responsivity of the trapped locus. To determine RA response in vivo, we fed pregnant mothers with RA and stained embryos for *lac*Z activity (see Experimental Procedures). We did not detect any reproducibly significant change in the extent of staining in the embryos from RA-treated mothers versus those from untreated mothers. It is possible that metabolic conversion of RA in mothers prevents exposure of embryos to an adequate dose of RA.

To obtain sequence information about the trapped gene, 5′-rapid amplification of cDNA ends-PCR (RACE-PCR) was carried out twice in two independent reactions on total RNA isolated from RA-treated I193 cells. Both experiments yielded an identical 0.4-Kb cDNA, which was cloned and sequenced. One hundred and fifty bases of the cDNA were derived from the gene trap vector. The 5′ half of the RACE product was the Dr repeat element that we have characterized previously. We have published their sequences (Sam et al., 1996), and Dr sequences have been submitted to the GenBank (accession nos. U51725-U51727). The 3′ half of the RACE product was novel, with an open reading frame

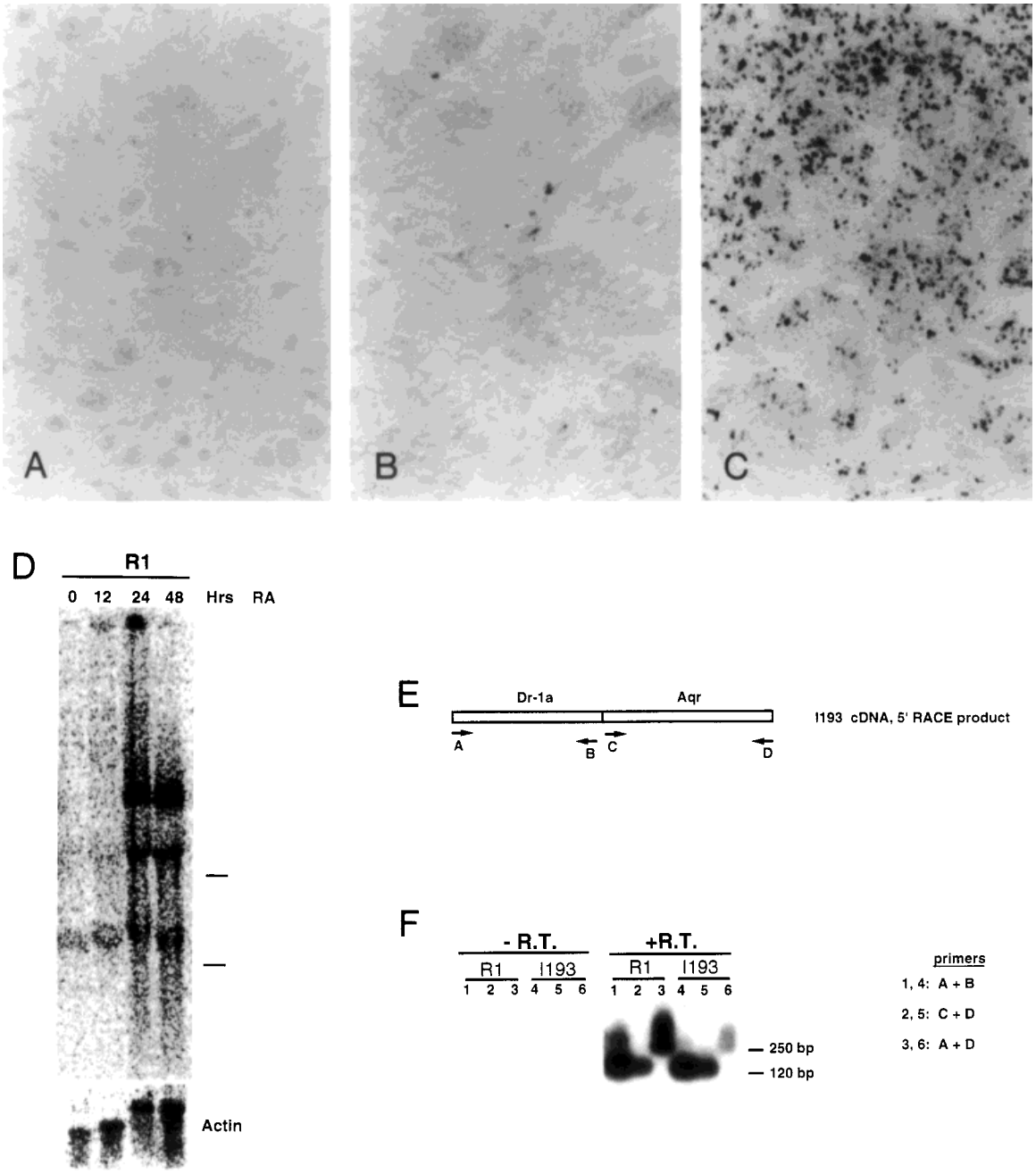


Fig. 1. **A–F:** *lac*Z and *Aquarius* (*Aqr*) expression are induced in vitro in response to retinoic acid (RA). I193 gene-trapped embryonic stem (ES) cells stained for *lac*Z activity, in the presence of LIF and in the absence of RA (A), in the absence of LIF and the absence of RA (B), and in the absence of LIF and the presence of RA (C) for 48 hr. D: RA induction of *Aqr* expression. Wild-type R1 ES cells were treated with RA for 0–48 hr. Northern blot analyses were performed on ten micrograms of total cytoplasmic RNA by using radiolabelled *Aqr* probe and then a mouse actin probe. The migrations of ribosomal RNA (1.9 Kb and 4.9 Kb) are indicated as size markers. E: Diagrammatic representation of the *Aqr* transcript and Dr-1a repeat sequence deduced from the rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) cDNA product. Primers A–D were used in reverse transcriptase-polymerase chain reaction (RT-PCR) experiments to confirm the presence of Dr-1a sequences on the *Aqr* transcript (F).

(ORF) from a 5′ exon of a novel gene, which we have designated *Aquarius* (*Aqr*).

We used the 5′ exon as a probe on Northern blot analysis of RNA isolated at different time points from wild-type R1 ES cells treated with RA (Fig. 1D). Three mRNA isoforms of 2.5, 6.0, and 8.5 Kb were all induced after 24 hr of exposure to RA. The 24-hr delay in induction may suggest that *Aqr* is farther downstream

of an RA signaling pathway; however, such long kinetics of transcription activation by RA are not unusual (Simeone et al., 1990; Tini et al., 1993).

To confirm the inclusion of the Dr repeat element within *Aqr* transcripts, we designed two pairs of PCR primers, one from the Dr element and one from the *Aqr* ORF (Fig. 1E), and we used them in reverse transcriptase (RT)-PCR reactions on total RNA from the parental R1 and the gene-trapped I193 lines (Fig. 1F). Amplification of the expected 250-base-pair (bp) fragment with A and D primers confirmed the presence of a Dr element on the *Aqr* transcript. We have previously reported the inclusion of Dr repetitive elements on many RA-responsive transcripts (Sam et al., 1996).

### *Aquarius* Is Similar to RNA-Dependent RNA Polymerases

To determine the sequence of the coding region corresponding to the *Aqr* ORF, we screened a 12.5-day postcoitum (dpc) embryonic cDNA library using the *Aqr* 5′ exon as a probe. A single 2.4-Kb cDNA was obtained and sequenced. This cDNA contained a novel 552 amino acid ORF with weak homology to RNA-dependent RNA polymerases (RRP) of the murine hepatitis and avian bronchitis viruses (GenBank accession nos. VFIHJH, S15760, and VFIHB2). There is 25% identity between *Aqr* and the viral RRP genes over the entire deduced amino acid sequence of *Aqr* (Fig. 2A). This degree of sequence homology is likely to be significant, because the *Aqr* ORF also contains the four-segment RRP motif corresponding to the polymerase active site (Fig. 2B; Poch et al., 1989; Sousa, 1996). The RRP motifs within the viral RRPs are in a different region of their sequence from those that are shown in Figure 2A. With the active site motif, *Aqr* may possess RNA polymerase activity. Alternatively, it may have descended from a common ancestor shared with RRP genes but, later, lost its original function. The 2.4-Kb cDNA that we have cloned is similar in size to the 2.5-Kb isoform of the *Aqr* transcript on Northern blot. However, our cDNA is not complete, because it does not contain the initiation codon and the 5′-untranslated region. *Aqr* ORF also contains two putative nuclear localization signals similar to those present in known nuclear proteins, such as p53 and DIM (Fig. 2C), suggesting nuclear localization of the *Aqr* gene product (Robbins et al., 1991). The nucleotide sequence (GenBank accession no. U90333) includes a 5′ palindromic region within the ORF that is capable of forming a 27-bp stem and a five-base-long loop structure. There is also an RNA instability motif (AUUUA) in the 3′-untranslated region (Ross, 1996), suggesting a short half-life for the *Aqr* transcript.

We searched several on-line sequence databases to identify other genes with significant homology to *Aqr*. No significant homology was detected in the yeast or *C. elegans* databases. However, there were four Expressed Sequence Tag (EST) sequences, two from human and two from the mouse EST databases, that were highly homologous to the *Aqr* ORF. The four ESTs show 88–100% amino acid sequence identity with *Aqr* in overlapping regions. Human EST 564231 (GenBank accession no. AA121582) was cloned from NT2 neuronal precursor cells and is 94% identical to *Aqr* over 167 amino acid residues; human EST 626472 (GenBank accession no. AA188145) was cloned from HeLa cells and is 88% identical to *Aqr* over 92 amino acid residues; mouse EST 765493 (GenBank accession no. AA274735) was cloned from mouse lymph node and is 100% identical to *Aqr* sequence over its full length of 157 amino acids; and mouse EST 535171 (GenBank accession no. AA073476) was cloned from RA-induced P19 embryonic carcinoma cells and is 89% identical to *Aqr* over 123 amino acid residues. This last EST was cloned from an RA-induced cell line, suggesting that it may also be responsive to RA.

### Germ-Line Transmission of the Gene Trapped *Aquarius* Locus

The trapped *Aqr* allele was transmitted through the germ line of mouse chimeras to produce heterozygous I193 mice. Heterozygotes were identified and bred to obtain F2 progeny in three different out-bred backgrounds (see Experimental Procedures). All genotyping was performed by using a quantitative Southern blot analysis (Fig. 3A,B). The F2 progeny results generally indicate a 1:2:1 Mendelian distribution of +/+, +/−, and −/− animals (Fig. 3C). The distribution in the 75% 129 background was slightly skewed, suggesting that a few homozygous mutant embryos may not have survived to birth. However, when embryos were examined from midgestation to late gestation, no abnormalities were observed in −/− embryos. At 10 months of age, +/− and −/− animals were the same weight and size as their wild-type (+/+) litter mates, they bred normally, and they did not manifest any discernable abnormalities. Newborn skeletons were also prepared and stained with alizarin red and alcian blue stains for skeletal and cartilage structures. No abnormalities were detected in craniofacial bone or cartilage structures where *Aqr* is expressed.

To confirm genomic disruption of the *Aqr* locus by the gene trap vector, genomic DNA from +/+, +/−, and −/− animals was amplified by PCR using primers from the 5′ *Aqr* exon and the *lac*Z sequences (Fig. 3D). We obtained amplified recombinant DNA products from +/− and −/− animals but not from +/+ animals, confirming the integration of the *lac*Z vector in the genomic *Aqr* locus (Fig. 3E).

To determine whether *Aqr* transcripts are present in the homozygous animals, we performed Northern blot analysis on total RNA from genotyped midgestation embryos. *Aqr* transcripts were present in homozygous as well as heterozygous and wild-type embryos (data not shown), indicating that the insertion of the gene-trap vector in the *Aqr* locus does not lead to a null mutation. The gene trap vectors employed in these experiments were designed so that transcription is terminated after the *lac*Z polyadenylation site, leading

**A**

```
          1                                                                 80
Aqr       PRFSWAAQEPPSQPKKIVAP .TVSQINAE.FVTQLACKYW APHIKKKSPFDIKVIE.EIY EKEIVKSRFA.I.RKIMLLE
Vfihjh    FRFQRV.D.EEGN.K.LDKF FVVKRTNLEVYNKEKEC.Y. .E.LTKDC..GV.VAEHEFF TFDVEGSRVPHIVRK.DLSK
S15760    CRFQRV.D.EDGN.K.LDKF FVVKRTNLEVYNKEKEC.Y. .E.LTKEC..GV.VAEHEFF TFDVEGSRVPHIVRK.DLSK
Vfihb2    ARFQELRDTEDGNLEYLDSY FVVKQTTPSNYEHEKSC.Y. .EDLKSE....V.TADHDFF VFN...KNIYNISRQ.RLTK

          81                                                               160
Aqr       FSQYLE.NYLWMNYSP.EVS S.KAYLMSIC.CM..VNEK. FRENVPAWETFKKKPDHFPF FFKC..ILKAALAETDGEFS
Vfihjh    FTM.LDLCYALRHFDRNDCS TLKEILLTYAECDESY.... FQK..KDWYDFVENPDIINV YKKLGPIFNRALLNT.ANFA
S15760    FTM.LDLCYALRHFDRNDCS TLKEILLTYAECEESY.... FQK..KDWYDFVENPDIINV YKKLGPIFNRALLNT.AKFA
Vfihb2    YTM.MDFCYALRHFDPKDCE VLKEILVTYG.CIEDYHPKW FEEN.KDWYDPIENSKYYVM LAKMGPIVRRALLNA.IEFG

          161                                                              240
Aqr       ..LHEQTLL.LLFLDH..CF NS.LEV.DLIRS......QV .QQLIS..LPM.WM..GLQP ARL.ELELKKTPKLRKFWNL
Vfihjh    DTLVEAGLVGVLTLDNQDLY GQWYDFGDFVKTVPCCGVAV ADSYYSYMMPMLTMCHALDS E.LF...VN..GTYREF.DL
S15760    DALVEAGLVGVLTLDNQDLY GQWYDFGDFVKTVPGCGVAV ADSYYSYMMPMLTMCHALDS E.LF...VN..GTYREF.DL
Vfihb2    NLMVEKGYVGVITLDNQDLN GKFYDFGDFQKTAPGAGVPV FDTYYSYMMPIIAMTDALAP ERYFEYDVHK.G.YKSY.DL

          241                                                              320
Aqr       IKKN..DEKMDPEAREQAYQ ERRFLSRLIQKFISVLKSIP .LSEPVTMDK..VHYCERFI ELMIDLEALLPTRRWFNTI.
Vfihjh    VQYDFTDFKL..ELFNK.Y. ...F......KHWS.MTYHP NTSE.CEDDRCIIH.CANF. NILFSM..VLP.KTCFGPLV
S15760    VQYDFTDFKL..ELFTK.Y. ...F......KHWS.MTYHP NTCE.CEDDRCIIH.CANF. NILFSM..VLP.KTCFGPLV
Vfihb2    LKYDYTEEKQ..ELFQK.Y. ...F......KYWD.QEYHP NCRD.CSDDRCLIH.CANF. NILFST..LIP.QTSFGNLC

          321                                                              400
Aqr       ....LDDSHLLVHC.Y.... LSSLVHREEDGHLFSQL.L. DMLKFYTGFEINDQTGNALT ENEMTTIHYD..RITS.L..
Vfihjh    RQIFVDGVPFVVSIGYHYKE LGVVMNMDVDTHRY.RLSLK DLLLYAADPALHVASASALL ..DLRTCCFSVAAITSGVKF
S15760    RQIFVDGVPFVVSIGYHYKE LGVVMNMDVDTHRY.RLSLK DLLLYAADPALHVASASALL ..DLRTCCFSVAAITSGVKF
Vfihb2    RKVFVDGVPFIATCGYHSKE LGVIMNQD.NTMSFSKMGLS QLMQFVGDPALLVGTSNNLV ..DLRTSCFSVCALTSGITH

          401                                                              480
Aqr       QRAAFAHFS.ELYDFALSNV AEVDARDSL.VK.FFGPLSS NTLHQVASYLCLLPTLPKNE DTTFD.KE..FLLELLVSR.
Vfihjh    QTVKPGNFNQDFYEFILSK. GLLKEGSSVDLKHFFFTQDG NA..AITDYNYYKYNLP... .TMVDIKQLLFVVE.VVNKY
S15760    QTVKPGNFNQDFYEFILSK. GLLKEGSSVDLKHFFFTQDG NA..AITDYNYYKYNLP... .TMVDIKQLLFVLE.VVNKY
Vfihb2    QTVKPGHFNKDFYDF.AEKA GMFKEGSSIPLKHFFYPQTG NA..AINDYDYYRYNRP... .TMFDICQLLFCLE.VTSKY

          481                                                              560
Aqr       ...HERR.I..SQ..IQQLN QMPLYPTEKI....IWDENI VPTE....Y.YSGEGCL.AL PKLNLQF.LTLHDYLLRNF.
Vfihjh    FEIYEGGCIPATQVIVNNYD KSAGYPFNKFGKARLYYEAL SFEEQDEIYAYTKRNVLPTL TQMNLKYAISAKN.RARTVA
S15760    FEIYEGGCIPATQVIVNNYD KSAGYPFNKFGKARLYYEAL SFEEQDEIYAYTKRNVLPTL TQMNLKYAISAKN.RARTVA
Vfihb2    FECYEGGCIPASQVVVNNLD KSAGYPFNKFGKARLYYE.M SLEEQDQLFEITKKNVLPTI TQMNLKYAISAKN.RARTVA

          561                                                              640
Aqr       NLFRLESTYEIRQDIEDSVS RMKPWQSEYG.GVVV.VFRW LGHGCPAHCGFHCSRGCQTP TSVKTG..QPEFVQMSPSI*
Vfihjh    GVSIL.STMTGRM.FHQKC. .LKSIAATRGVPVVIGTTKF YG.GWDDMLR.RLIKDVDSP ..VLMGWDYPKCDRAMPNIL
S15760    GVSIL.STMTGRM.FHQKC. .LKSIAATRGVPVVIGTTKF YG.GWDDMLR.RLIKDVDSP ..VLMGWDYPKCDRAMPNIL
Vfihb2    GVSIL.STMTNRQ.FHQKI. .LKSIVNTRNASVVIGTTKF YG.GWDNMLR.NLIQGVEDP ..ILMGWDYPKCDRAMPNLL
```

**B**

```
                A                    B                    C          D
Aqr      41 HIKKKSPFDIKVIEEIY 75 ILKAALAETDGEFS.LHEQTLLLLFLDH 116 FNTILDDSHLLV 162 YSGEGCLALPKLN
Vfihjh  717 CNGHK.IEDLS.IRELQ 89 SQHTMLVKMDGDEVYLPYPDPSRI.LGA 203 KWTEVDDY.VLA  44 EIGKVRPPLNKNY
S15760  717 CNGHK.IEDLS.IRELQ 89 SQHTMLVKMDGDEVYLPYPDPSRI.LGA 203 KWTEVDDY.VLA  44 EIGKVRPPLNKNY
Vfihb2  729 LAKQGLVADISGFREVL 27 SQHTMLVEVDGEPKYLPYPDPSRI.LGA 189 GSENVDDFNQLA  92 GEGKDVV.YYKAT
                 * *  *         **        *     * *         *    ***        *
```

**C**

```
Aqr     43  KKKspfdiKvieeiyeKeivKsR
Aqr    195  KKtpKlRKfwnliKKndeKmdpeaR
p53    316  KKKpldgeyftlKiRgR
DIM    519  RKKyRaigtfmsvyyKsKKgRK
TGA1A   74  KvlRRlaqnReaaKRsRlRkk
TGA1B  185  KKRaRlvRnResaqlsRqRKK
```

Fig. 2. *Aqr* is similar to viral RNA-dependent RNA polymerases (RRPs). **A:** Alignment of the *Aqr*-deduced amino acid sequence with the murine hepatitis virus RRPs (Vfihjh and S15760) and the avian infectious bronchitis virus RRP (Vfihb2). Gaps (dots) are introduced for the best alignment and are counted in numbering of residues. Numbering is based on the *Aqr* amino acid residues. Amino acid identities are shaded. **B:** Alignment of the RRP conserved motifs for the *Aqr* and the three viral RRP sequences shown in A. The strictly conserved residues are in boldface, and the conserved hydrophobic residues are indicated by asterisks. Numbers indicate the amino acid residues between the fragments or from the beginning of the sequence. **C:** *Aqr* contains two nuclear localization signals (NLSs) aligned here with known NLSs from four other genes. Numbers indicate the residues from the beginning of the sequence.

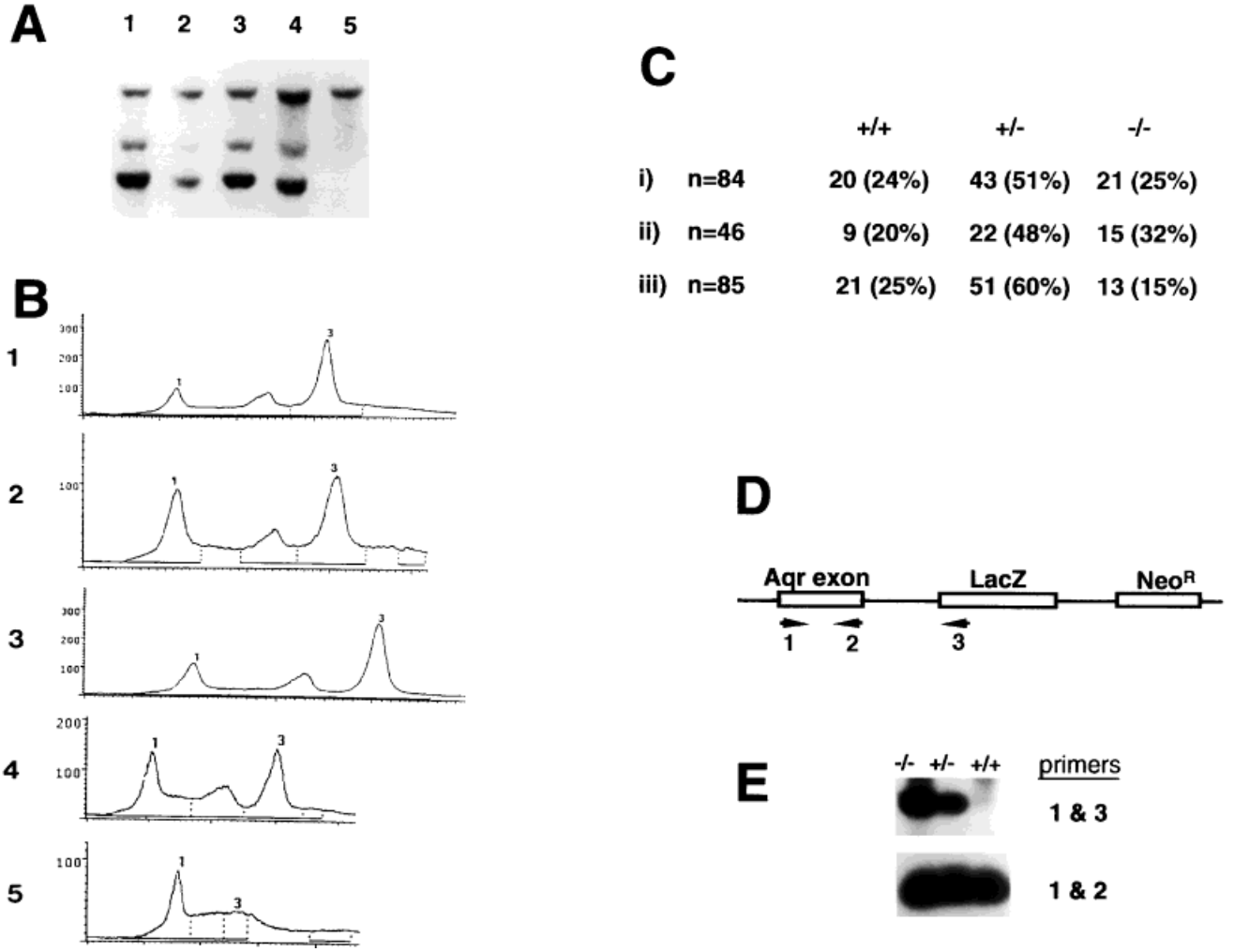


Fig. 3. Genotyping of F2 progeny and the genomic disruption of the *Aqr* locus. **A:** Southern blot. The upper band is endogenous en-2 gene used as internal control for DNA loading. The lower two bands are *lac*Z transgene. Ratio of the intensity of the faster *lac*Z band (peak 3 in B) to the en-2 band (peak 1 in B) indicates the genotype of the animal: +/+ (ratio = 0; lane 5), +/− (ratio = 1; lanes 2 and 4), and -/- (ratio = 2; lanes 1 and 3). **B:** The densitometric measurements of the band intensities. **C:** Genotypes of F2 progeny in three mixed genetic backgrounds: i) 50% 129sv-cp/50% C57-Bl6, ii) 25% 129sv-cp/25% C57-Bl6/50% CD-1, and iii) 75% 129sv-cp/25% C57-Bl6. **D:** *lac*Z disrupts *Aqr* genomic locus. *Aqr* exon was cloned by 5′ RACE-PCR. The primers 1–3 were used for PCR on genomic DNA to confirm disruption of the *Aqr* locus by the gene trap vector. **E:** PCR on genomic DNA from +/+, +/−, and −/− animals and Southern blot probed with *Aqr* exon.

to a transcription disruption of the endogenous trapped gene. It is possible, however, that RNA polymerases proceed with transcription downstream of a polyadenylation site. Polyadenylation site selection by RNA polymerase II and RNA transcript cleavage are not well understood (Darnell et al., 1990; Alberts et al., 1994), and it remains a concern in gene-trapping experiments. In the case of a run-through transcription, splicing across the gene trap vector sequences can lead to a wild-type mRNA. The presence of the wild-type *Aqr* transcript indicates why we did not detect any abnormalities in the homozygous mice.

### Embryonic Expression in Neural Crest and Mesodermal Tissues

The developmental pattern of *Aqr* gene expression was examined during embryogenesis by staining for β-galactosidase activity (Fig. 4) and by using RNA and wholemount in situ hybridization techniques employing *Aqr* cDNA as a probe (Figs. 5, 6). At day 8.5 of gestation, expression was detected in the cephalic neural folds and mesoderm, paraxial mesoderm, and somites and throughout the lateral ridges of neural folds, including the fused regions in between the caudal and rostral extremities of neural tube closures (Fig. 4A). The lateral ridges of neural folds are where premigratory neural crest (NC) cells emerge and subsequently migrate in a rostral-to-caudal sequence at the cranial, trunk, and tail levels (Serbedzija et al., 1990, 1992). Trunk NC cells migrate through the somites along well-characterized pathways in a segmental fashion and give rise to numerous derivatives, including the sensory and sympathetic ganglia (Krull et al., 1995).

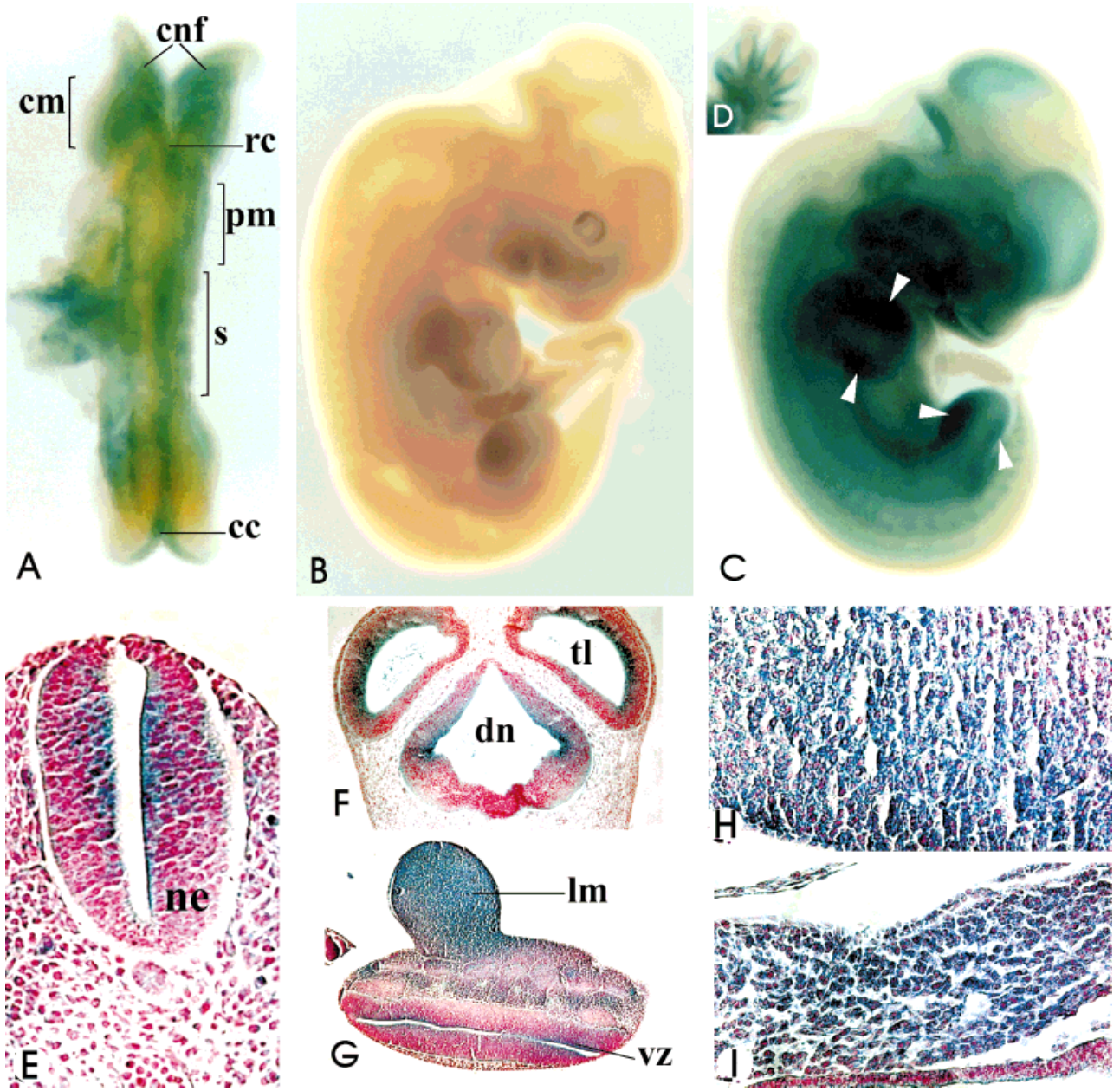


Fig. 4. *lac*Z expression in wholemount embryos and sections. **A:** Embryonic day 8.5 (E8.5) *lac*Z staining in the cephalic neural folds (cnf), in the cephalic mesoderm (cm), at the lateral ridges of neural folds, including the fused regions in between the caudal (cc) and rostral (rc) extremities of neural tube closures in which premigratory neural crest cells originate, in somites (s), and in the paraxial mesoderm (pm). **B,C:** E10.5 nonstaining control (B) and *lac*Z-expressing (C) litter mate embryos showing expression in the forebrain, posterior midbrain, hindbrain, craniofacial structures, limb buds, and internally along the length of the body. Note the expression in both anterior and posterior regions of the limbs (arrowheads). **D:** E13.0 forefoot plate expression in the circumference of the cartilaginous anlagen of digits. E–I: Expression in E11 embryonic sections. **E:** Ventricular zone in the dorsal region of the neuroepithelium (ne) in the neural tube. **F:** Ventricular cells of neuroepithelium in the telencephalon (tl) and diencephalon (dn). **G:** Widespread expression in the limb mesoderm (lm) and in the ventricular zone (vz) of the neuroepithelium. **H:** Widespread expression in the mesenchyme within the first branchial arch. **I:** Expression in the paraxial mesoderm at the lower lumbar region.

At day 10.5, expression is detected in forebrain, posterior midbrain, hind brain, craniofacial structures, limb buds, and internally along the length of the body (Fig. 4C), indicating that *Aqr* expression is maintained in NC target tissues after their migration is terminated. To examine the expression pattern in more detail, histological sections of the *lac*Z-stained embryos at day 11 were prepared (Fig. 4E–I). The ventricular zone of neuroepithelium throughout the nervous system was positive for *lac*Z. Ventricular expression in

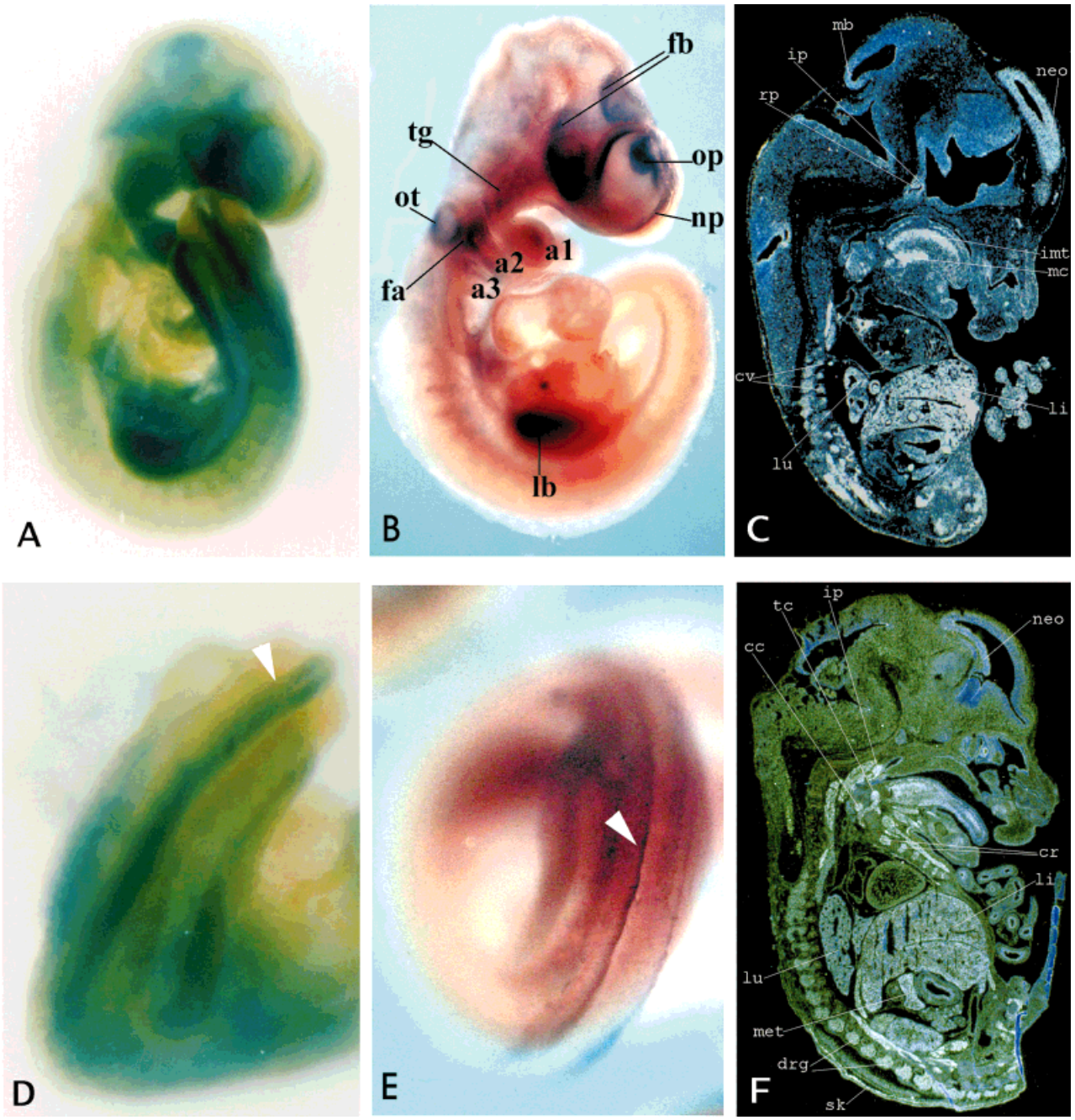


Fig. 5. RNA in situ hybridization recapitulates *lacZ* expression. RNA in situ hybridization of wholemount embryos (B,E) recapitulates *lacZ* expression in E9.5 embryos (A,D). **A,B:** Expression in the forebrain (fb), trigeminal neural crest tissue (tg), the region surrounding the otic vesicle (ot), facioacoustic neural crest tissue (fa), the olfactory placode (op), the nasal process (np), branchial arches 1–3 (a1-a3), and the limb bud (lb). **C:** E12.5 sagittal section showing expression in the neopallial cortex (neo), the midbrain (mb), the infundibulum of the pituitary (ip), Rathke's pouch (rp), the cartilage primordium of the lumbar vertebral body (cv), and neural-crest derived structures: the intrinsic muscle of the tongue (imt) and Meckel's cartilage (mc). Expression is also seen in the liver (li) and the lung (lg). **D,E:** Expression in the lateral ridges of neural folds, where premigratory neural crest cells originate (arrows). RNA in situ of embryonic sections. **F:** E14.5 sagittal section showing expression in the neopallial cortex (neo), infundibulum of the pituitary (ip), the lung (lg), the liver (li), cartilage primordium of the ribs (cr), thyroid and cricoid cartilages (tc and cc), the dorsal root ganglion (drg), the metanephros (met), and the skin (sk).

neuroepithelium was observed in all stained embryos and was also confirmed by wholemount in situs (Fig. 6B). Figure 4E–G presents the expression in the neural tube, telencephalon, and diencephalon. Widespread expression was also detected in the limb mesoderm (Fig. 4G), in the mesenchyme within the branchial arches (Fig. 4H), and in the paraxial mesoderm at the lower lumbar region (Fig. 4I).

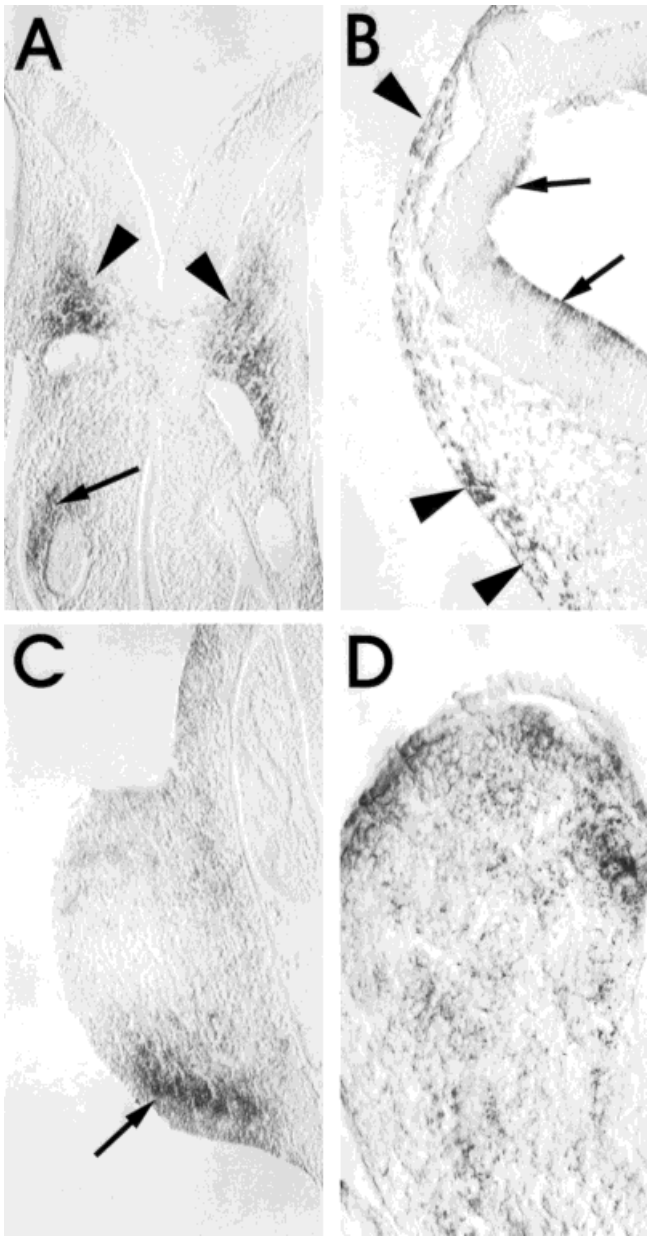


Fig. 6. Expression in sections from wholemounts in situ. *Aqr* expression in E9.0–9.5 embryonic sections. **A:** Expression in trigeminal neural crest tissue (arrowheads) just above the rostral extension of dorsal aorta and in the mesenchyme within the first branchial arch (arrow) around the first branchial arch artery. **B:** Expression in the neural crest cells migrating laterally next to surface ectoderm (arrowheads) and in the ventricular zone of neuroepithelium (arrows). **C:** Expression in the posterior end of limb bud (arrow). **D:** Expression in the mesenchyme within the first branchial arch and stronger expression in the rostral regions.

To determine whether the gene trap vector insertion deregulated *Aqr* expression, wild-type embryonic day 9.5 (E9.5) embryos were stained by wholemount RNA in situ hybridization with *Aqr* cDNA and compared with *lac*Z staining of embryos (Fig. 5A,B,D,E). Both *lac*Z and in situ hybridization recapitulated the same expression pattern for *Aqr* in forebrain tissues, trigeminal tissues, facioacoustic NC tissues, olfactory placode, nasal process, branchial arches 1–3, limb buds, and the lateral ridges of neural folds. These results indicate that the pattern of *lac*Z expression faithfully reproduced the endogenous *Aqr* gene expression pattern. The expression surrounding the otic vesicle and in the olfactory placode was more pronounced in the wholemounts in situ than in the *lac*Z-stained embryos, perhaps due to an unspecific staining rather than a deregulation in gene expression.

*Aqr* expression was further delineated by examining E9 and E9.5 sections from wholemount in situs (Fig. 6). *Aqr* RNA was expressed in the trigeminal NC, the branchial arch mesenchyme (Fig. 6A,D), laterally migrating NC cells, and the ventricular region of the neuroepithelium (Fig. 6B). Furthermore, *Aqr* was expressed in a dynamic and spatially regulated manner in the limbs in the zone of polarizing activity (ZPA) at the posterior end of limb bud at day 9.5 (Fig. 6C). The ZPA constitutes a group of mesenchyme cells that establish the anteroposterior patterning of the limb. Cells in the ZPA express *Sonic hedgehog*, a key signal involved in limb patterning (Cohn and Tickle, 1996). At day 10.5, *Aqr* was expressed in both anterior and posterior ends of the limb bud (Fig. 4C), and, at day 13, expression of *Aqr* in the foot plate was restricted to the circumference of the cartilaginous anlagen of digits (Fig. 4D). This distinct and dynamic limb expression suggests a role for *Aqr* in the determination or patterning of limb structures.

To observe *Aqr* expression at later stages of embryonic development, we performed RNA in situ of embryonic sections at E12.5 (Fig. 5C) and E14.5 (Fig. 5F). *Aqr* was expressed in the neopallial cortex, midbrain, infundibulum of the pituitary, Rathke's pouch, intrinsic muscle of the tongue, Meckel's cartilage, thyroid and cricoid cartilages, cartilage primordium of the ribs and lumbar vertebral body, liver, lung, metanephros, dorsal root ganglion, and skin. Some of these structures, such as Meckel's cartilage and ganglia, are derived from NC cells, indicating the persistence of *Aqr* expression during the differentiation of these cells. *Aqr* may potentially be a useful marker in following the development of these lineages.

### Chromosome Localization by Fluorescence In Situ Hybridization Analysis

To localize *Aqr* in the mouse genome, we isolated a genomic phage clone and used it as a probe in fluorescence in situ hybridization (FISH) analysis (Fig. 7). The assignment between signals from the probe and mouse chromosome 2 was obtained by using 4,6-diamidino-2-phenylindole (DAPI) banding and superimposing FISH signals with DAPI-banded chromosomes. The detailed position on chromosome 2, region E5 to F2, was based on the analysis of ten mitotic chromosome spreads. There are two semidominant mutations, Strong's luxoid and tight skin, in the E5 to F2 region of mouse chromosome 2 that are potentially interesting, because

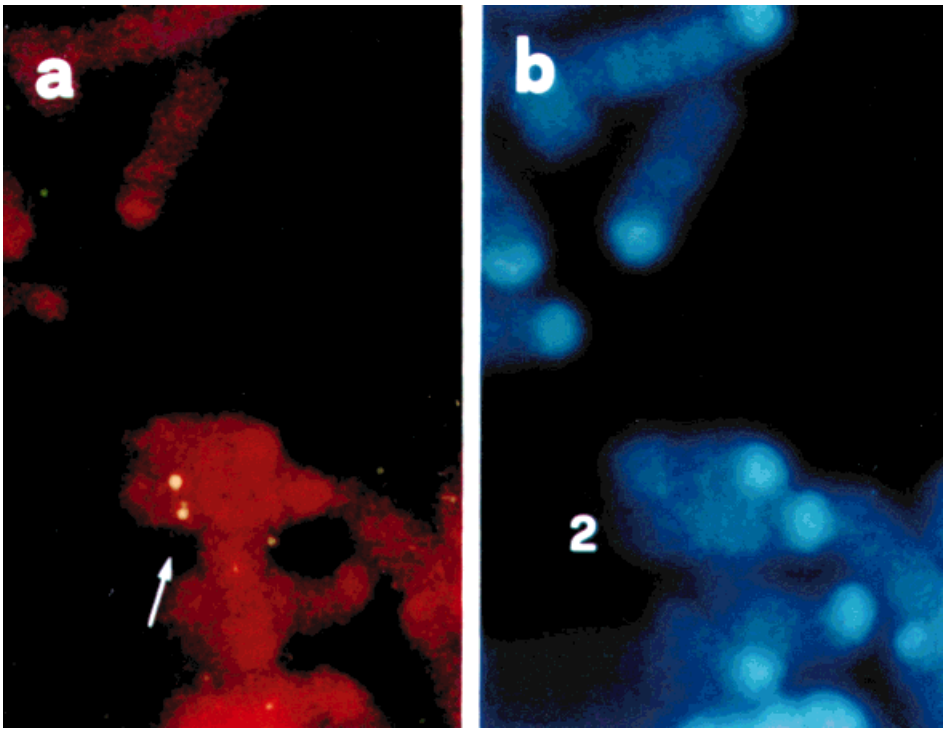


Fig. 7. By using fluorescence in situ hybridization (FISH), the *Aqr* gene was localized in a set of mouse metaphase chromosome spreads. **a:** Phosphatidyl inositol-stained chromosomes with fluorescein isothiocyanate signals. **b:** Chromosomes were stained with 4,6-diamidino-2-phenylindole to assign signal from the probe with the mouse chromosome 2 bands. Under the conditions used, the hybridization efficiency was 93% for the probe (93 of 100 mitotic figures showed signals on one pair of the chromosomes).

they exhibit phenotypes in *Aqr*-expressing tissues. Strong's luxoid (*lst*) mice present limb phenotypes, including polydactyly, and reductions or duplications of the radius (Lyon et al., 1996). Tight-skin (*Tsk*) mice have increased growth of cartilage and bone, tight skin, and hyperplasia of the subcutaneous connective tissue (Lyon et al., 1996).

## DISCUSSION

We have identified a novel mammalian gene, termed *Aquarius* (*Aqr*), that may be distantly related to RRPs. *Aqr* shares 25% amino acid identity over the length of its ORF with murine hepatitis virus RRPs, and it has the conserved active site RRP motif. Very little is known about the structural and functional domains of RRPs. Following entry into the host cell and uncoating in the cytoplasm, positive-stranded RNA viruses replicate their genetic material by using their (+) RNA genome as a template to synthesize a complementary (−) RNA molecule, which, in turn, serves as a template for the synthesis of progeny genomic (+)-strand RNA. There have been a few reports suggesting the presence of RNA-dependent RNA synthesis activity in eukaryotic cells (Volloch et al., 1991; Schiebel et al., 1993a,b). In addition, although hepatitis delta virus does not encode any polymerase, it can still replicate autonomously through an RNA-dependent RNA-synthesis mechanism in animal cells, raising the possibility that eukaryotic cells have the enzymatic machinery necessary to replicate RNA molecules (Lai, 1995).

The polymerase subunit of the Tetrahymena telomerase, p95, also yields an alignment with the polymerase active site of a family of viral RNA-dependent RNA polymerases (Collins et al., 1995). Telomerase is a specialized polymerase that synthesizes long, repetitive, telomeric DNA (Shay, 1996). The mouse homolog of the p80 subunit of Tetrahymena telomerase has been identified, but the mouse p95 homolog, which is thought to carry the catalytic polymerase domain, has not been identified (Harrington et al., 1997). Clearly, it will be of interest to determine whether the *Aqr* gene product possess RNA-dependent RNA polymerase activity. *Aqr* ORF also carries two nuclear localization signals, suggesting that *Aqr* protein is localized to the cell nucleus.

During embryogenesis, *Aqr* is expressed predominantly in NC tissues from early on in their ontogeny to their differentiated derivatives in cartilaginous head structures, such as the hyoid bone and Meckel's cartilage. Expression was first detected in the dorsal edges along the whole length of the neural groove and in its adjacent mesenchyme in the head fold, which contains the earliest migrating NC cells. NC cells constitute a transient embryonic cell population that arises from the dorsal region of the neural tube, extensively migrates, and gives rise to a number of tissues and structures, including ectomesenchyme that forms most

of the skeleton, dermis, and connective tissue of the vertebrate head (Couly et al., 1993; Schilling, 1997). Cranial mesoderm and NC cells are codistributed in the craniofacial mesenchyme but are distinctly segregated in the branchial arches, although the interactions of these cell populations are not yet fully understood (Trainor and Tam, 1995). *Aqr* expression in NC cells and in craniofacial and branchial mesoderm suggests a possible role for its gene product in craniofacial development. *Aqr* is expressed in paraxial and branchial mesoderm and in other mesodermal tissues, such as the ZPA in the limb buds, where *Sonic hedgehog* signaling is thought to establish the anteroposterior limb pattern (Perrimon, 1995). There is also evidence that *Sonic hedgehog* cooperates with an RA-inducible cofactor to establish ZPA-like activity (Ogura et al., 1996). The distinct and dynamic expression in the limbs and its RA inducibility suggest a possible role for *Aqr* in the development of limb structures or their patterning.

Through germ-line transmission of the trapped locus, we created animals homozygous for the disrupted *Aqr* gene. Homozygous mutant mice did not present any obvious abnormalities. At 10 months of age, they were the same size and weight as their heterozygous and wild-type siblings, and they bred normally. Newborn skeletons from homozygotes also did not present any abnormalities in craniofacial structures. The absence of any discernable phenotype in mice homozygous for the gene trap allele does not rule out the possibility that *Aqr* plays an essential role in the cells that express it. We found that the gene trap allele described here is not a null mutation, because *Aqr* RNA is present in homozygous animals. The lack of a discernible phenotype is likely due to read-through transcription and splicing across the trapping vector. *Aqr* was identified in an RA gene trap screen of mouse ES cells. In vitro, *Aqr* expression is induced ninefold in response to RA (Forrester et al., 1996), and its transcript includes the RA-induced Dr repeat element in its 5′ end (Sam et al., 1996). We did not detect RA responsivity of the *Aqr* locus when we examined embryonic *lac*Z expression in vivo (E9.5–E11.0). This lack of in vivo RA responsiveness might be due to the inability to expose embryos to a concentration of RA in vivo that corresponds to the optimal dose in vitro (Morriss-Kay, 1993).

RA is known to induce both alterations in NC cell migration as well as craniofacial defects (Lee et al., 1995; Gale et al., 1996). RA receptor α (RARα) is expressed in migrating crest cells and in facial mesenchyme (Ruberte et al., 1991). Also, mutant mice with targeted mutations in both RARα and other RARs have severe defects in the craniofacial complex (Lohnes et al., 1994). The transcription factor AP-2, which is induced in response to RA, is expressed in NC cell lineages, and its disruption in mice also leads to anencephaly and craniofacial defects (Zhang et al., 1996). *Aqr* induction in response to RA and its expression in ectomesenchyme presents *Aqr* as a candidate gene in an RA-signaling pathway that may also play a role in craniofacial development.

In a search of EST databases, we identified *Aqr* itself and three other genes that were similar to *Aqr*, two from humans and one from mouse. These genes are highly similar in sequence to *Aqr* (88–94% amino acid identity). The high degree of similarity suggests that they may be homologous gene pairs in the two species. We did not detect any other similar sequences in the GenBank, yeast, or *C. elegans* databases, suggesting that *Aqr* may be the founding member of a mammalian-specific gene family.

The availability of significant numbers of sequence databases for different organisms, the utility of analytical software products, and the ready access to these tools through the internet has led to an increase in gene identification by computer (Fickett, 1996). Approximately 50% of all human genes are represented already in the EST databases (http://www.ncbi.nlm.nih.gov/dbEST; Gerhold and Caskey, 1996). Up to half of ESTs are sequences from the 5' end of cDNAs, and gene-trapping experiments can also be designed to trap genes in their 5' exons, which was the case in the present study. Thus, one can envision obtaining rapidly sequence information about the trapped gene by using RACE-PCR cloning of 5' exons, searching the EST databases for the 5' exon sequences, and obtaining the commercially available EST clones. This combination of gene trap experimentation and EST databases provides ready access to the sequence of trapped genes and has the potential to increase several-fold the pace of gene discovery and functional characterization. At the same time, the gene-trapped mice allow for expression and functional analyses of the trapped gene.

In summary, in an RA gene trap screen of mouse ES cells, we have identified and characterized the RA-responsive *Aqr* gene. *Aqr* is expressed developmentally and is a novel mammalian gene with some homology to RNA-dependent RNA polymerases. Our results demonstrate the feasibility of in vitro manipulation and screening of ES cells to rapidly identify and characterize genes that lie along a cellular-response pathway. Modification of this gene trap strategy in ES cells should make it possible to trap and study genes that either are induced by other biological, chemical, or physical agents or are subject to developmental regulation.

## EXPERIMENTAL PROCEDURES

### Gene-Trapped ES Cell Lines, *lac*Z Staining, and Germ-Line Transmission

Construction, handling, and characterization of the I193 gene-trapped ES cell line, *lac*Z staining, and the transmission of the trapped locus through the mouse germ line have been described previously (Forrester et al., 1996).

### Mouse Genotyping and F2 Breeding in Different Backgrounds

Genomic DNA was isolated from tails at the time of weaning, digested with EcoR1 restriction enzyme, and separated by electrophoresis on a 1% agarose gel. Southern blots were double probed with both *engrailed*-2 (en-2) and *lac*Z probes. En-2 was used as an internal control for quantitative Southern, and *lac*Z detected the transgene in the *Aqr* locus. Densitometric analyses were used to measure the intensity of the *lac*Z signal (4.0-Kb band) relative to the en-2 signal (11-Kb band). Three F1 males and six F1 females, all of which were heterozygous for the disrupted locus, were bred to obtain F2 progeny in a 50% 129sv-cp/50% C57BL6 out-bred background. Because the chimeric animals generated by blastocyst injection died shortly after germ-line transmission, out-bred F1 males were also bred to CD-1 and 129 females to obtain 50% CD-1/25% 129/25% C57BL6 and 75% 129/25% C57BL6 backgrounds. Three males and six females from these backgrounds were also bred to obtain F2 progeny. Heterozygous animals (+/−) were expected to have a *lac*Z/en-2 signal ratio of 1, whereas homozygotes (−/−) were expected to yield a *lac*Z/en-2 signal ratio of 2 on Southern blots. To confirm the genotyping, two heterozygotes and two homozygotes were back-crossed to wild-type animals, and their progeny were genotyped for the presence or absence of the transgene. In all four crosses, half of the progeny from the heterozygotes and all of the progeny from the homozygotes carried the transgene, confirming the original genotypes.

### RACE, PCR Amplification, Cloning of the 5′ *Aqr* Exon, and Genomic PCR

5′-RACE-PCR on fusion transcripts from the gene-trapped ES cell lines was performed by using a 5′-RACE kit (Life Technologies, Burlington, Ontario, Canada) essentially according to the manufacturer's instructions, with the following modifications: Reverse transcription was carried out at 42°C for 30 min in the presence of [α-32P]dCTP to monitor the synthesis of the first strand cDNA. SuperScript II (Life Technologies) and GT-*lac*Z-1 [5′-GCAAGGCGATTAAGTTGGGT-3′] primers were used for reverse transcription.

DNA amplification of the RACE products was performed in two rounds in 50-μl volumes containing 1 / PCR buffer II (Perkin Elmer, Foster City, CA), 1.25 mM MgCl2, 200 mM deoxynucleoside triphosphates, 200 nM of each primer, and 2.0 U of AmpliTaq DNA Polymerase (Perkin Elmer) in a DNA thermal cycler (Perkin Elmer; one cycle at 94°C for 5 min; one cycle at 80°C for 8 min during which Taq polymerase was added to the reaction mix; 35 cycles at 94°C for 1 min, at 60°C for 2 min, and at 72°C for 3 min; one cycle at 72°C for 5 min). The second round of amplification was performed under the same conditions, using nested primers (see below) and 5 ml of the reaction product from the first round. The primers used for DNA amplification were as follows: for the first round, GT-*lac*Z-2A [5′-CCGTCGACTCTGGCGCCGCTGCTCTGTCAG-3′] and Anchor [5′-GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3′] (Life Technologies); for the second round, Nested 2AU [5′-CAUCAUCAUCAUTTGTCGACCTGT TGGTC-TGAAACTCAGCCT-3′] and Universal Amplification Primer [5′-CUACUACUACUAGGCCACGCGTCGAC-TAGTAC-3′] (Life Technologies).

The RACE-PCR products were digested with *Sal*I and cloned into pBlueScript II vector (Stratagene, La Jolla, CA). Alternatively, they were cloned into the pAMP1 plasmid by using CloneAmp System (Life Technologies). The cDNA inserts were sequenced on both strands by using the AutoRead sequencing kit (Pharmacia, Baie d'Urfe, Quebec, Canada) on an A.L.F. DNA Sequencer (Pharmacia).

PCR amplification on genomic DNA from +/+, +/−, and −/− animals was performed with primers flanking the 5′ *Aqr* exon, 1 [5′-GCATGTAAATACTGGGCT-3′] and 2 [5′-CTAAATTCCAGCAGCATT-3′], and the GT-*lac*Z-1 primer (above). PCR conditions were the same as those described above.

### cDNA Cloning, Sequencing, GenBank, and Internet Utilities

A 12.5-dpc mouse embryonic cDNA library was screened with standard procedures (Sambrook et al., 1989) by using the nonrepeat region of the RACE product (*Aqr* exon) as a DNA probe. Hybridization was carried out in 50% formamide, 5 / sodium saline phosphate EDTA buffer (SSPE; 1 / SSPE is 150 mM NaCl, 10 mM $NaH_2PO_4$, and 1 mM EDTA), 5 / Denhardt's solution, 100 mg/ml boiled and chilled salmon sperm DNA, and 0.5% sodium dodecyl sulfate (SDS) at 42°C for 24 hr. High-stringency washes were carried out twice in 0.2 / standard saline citrate (SSC), 0.2% SDS at 60°C for 15 min each. DNA was prepared from the only positive plaque, and its 2.4-Kb cDNA insert was sequenced as described above. GenBank search and sequence analyses were performed by using Fasta, Translate, Bestfit, and Pileup software from the Wisconsin Package (version 8.0, Open VMS; Genetics Computer Group, Inc., Madison, WI). The *Aqr* ORF was submitted to the National Center for Biotechnology Information XREF internet service (http://www.ncbi.nlm.nih.gov/XREFdb/), where it was searched against the most recent update of the EST database. Positive matches were examined for homology with *Aqr*. On-line sequence databases for yeast (100% of the genome complete; http://speedy.mips.biochem.mpg.de/mips/yeast/) and *C. elegans* (60% of the genome complete; http://www.sanger.ac.uk/~sjj/C.elegans_Home.html) were also searched for homology with *Aqr*.

### RA Induction, RNA Preparation, Northern Blot Analyses, and RT-PCR

In the RA-induction experiments, ES cells were maintained in ES cell medium (without LIF) containing 5% fetal calf serum (FCS) and $10^{-6}$ M all-trans RA (Sigma, St. Louis, MO). The medium was changed

every 12 hr on all plates, so that cells were in the continuous presence of RA for varying time periods (0, 12, 24, or 48 hr). Control samples (0 hr) were in ES cell medium containing 5% FCS in the absence of LIF and RA throughout the experiment. In in vivo RA experiments, pregnant mice at E8.5–E9.5 were fed 2 mg of RA per 100 gram of body weight; then, embryos were removed surgically 24 hr or 48 hr after RA treatment and stained for *lac*Z histochemical staining.

Total RNA was prepared from ES cell lines by using a modified version of the method described by Chomczynski and Sacchi (1987). Confluent plates (60 mm) of cells were lysed with 5 ml guanidinium thiocyanate (4.4 M) solution containing 25 mM sodium citrate, pH 7.0, 0.6% sarcosyl, and 100 mM β-mercaptoethanol. The lysate was mixed with 5 ml Tris EDTA-saturated phenol, 1 ml chloroform, and 0.5 ml 2 M sodium acetate. After centrifugation, RNA was precipitated from the aqueous phase with an equal volume of isopropanol. Northern blot analyses were carried out according to standard procedures (Sambrook et al., 1989). The blots were hybridized overnight with $^{32}$P-labelled probes and were washed twice under stringent conditions in 0.2 / SSC (1 / SSC is 150 mM NaCl plus 15 mM sodium citrate)/0.2% SDS at 60°C for 15 min each.

RT-PCR was carried out with the GeneAmp RNA PCR kit (Perkin Elmer, Foster City, CA) according to the manufacturer's instructions. The primers used for PCR amplification were *Aqr* exon-flanking primers A and B and Dr-1a repeat-flanking primers C and D: A (5′-GAGCTCTTGCTATTGCTC-3′), B (5′-TAGAGTCG-GCAGCCCAAT-3′), C (5′-GCATGTAAATACTGGGCT-3′), and D (5?-CTAAATTCCAGCAGCATT-3′). The PCR products were run on a 1% gel, blotted, and hybridized with $^{32}$P-labelled Dr1a-*Aqr* probe.

### RNA and Wholemount In Situ Hybridization

RNA in situ hybridization was performed as described previously (Hogan et al., 1994). Briefly, tissues were cryostat sectioned at 10 μm, mounted on glass slides, and refixed in 4% paraformaldehyde. Prehybridization treatments were performed as described previously (Hogan et al., 1994). $^{35}$S-labeled, single-stranded RNA probes were prepared by using T3 and T7 RNA polymerases (Boehringer-Mannheim, Laval, Quebec, Canada). Adjacent sections were hybridized with *Aqr* sense and antisense probes. Posthybridization washings included treatment with 50 μg/ml RNase A (Sigma, St. Louis, MO) at 37°C for 30 min. Following dehydration, the slides were dipped into NTB-2 film emulsion (Eastman-Kodak, Rochester, NY), exposed at 4°C for 6–10 days, developed, and stained with toluidine blue. Wholemount in situ hybridization was performed as described previously (Conlon and Herrmann, 1993) with *Aqr* riboprobes.

### FISH

A genomic clone for *Aqr* was obtained and used as a probe for FISH analysis. FISH was performed according to published procedures (Heng et al., 1992; Heng and Tsui, 1993).

## NOTE ADDED IN PROOF

We have carried out a further statistical analysis of the sequence alignment shown in Figure 2A using the GAP and Bestfit programs in GCG package with 100 randomizations. This analysis gave z score values between 0.15 to 4.06 within the significance range ($z < 3.0$).

## ACKNOWLEDGMENTS

We thank Sandra Tondat and Ken Harpal for technical assistance.

## REFERENCES

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Molecular Biology of the Cell. New York: Garland Publishing, 1994:456.

Chen J, Nachabah A, Scherer C, Ganju P, Reith A, Bronson R, Ruley HE. Germ-line inactivation of the murine Eck receptor kinase by gene trap retroviral insertion. Oncogene 1996;12:979–988.

Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 1987;162:156–159.

Cohn MJ, Tickle C. Limbs: A model for pattern formation within the vertebrate body plan. Trends Genet. 1996;12:253–257.

Collins K, Kobayashi R, Greider CW. Purification of Tetrahymena telomerase and cloning of genes encoding the two protein components of the enzyme. Cell 1995;81:677–686.

Conlon RA, Herrmann BG. Detection of messenger RNA by in situ hybridization to postimplantation embryo whole mounts. In: Wassarman PM, DePamphilis ML, eds. Methods in Enzymology, Guides to Techniques in Mouse Development, Vol. 225. New York: Academic Press, 1993:373–383.

Couly GF, Coltey PM, Le Douarin NM. The triple origin of skull in higher vertebrates—A study in quail-chick chimeras. Development 1993;117:409–429.

Darnell J, Lodish H, Baltimore D. Molecular Cell Biology. New York: Scientific American Books, 1990:296.

Fickett JW. Finding genes by computer: The state of the art. TIG 1996;12:316–320.

Forrester L, Nagy A, Sam M, Watt A, Stevenson L, Bernstein A, Joyner AL, Wurst W. An induction gene trap screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro. Proc. Natl. Acad. Sci. USA 1996;93;1677–1682.

Gale E, Prince V, Lumsden A, Clarke J, Holder N, Maden M. Late effects of retinoic acid on neural crest and aspects of rhombomere identity. Development 1996;12:783–793.

Gerhold D, Caskey CT. It's the genes! EST access to human genome content. Bioessays 1996;18:973–981.

Harrington L, McPhail T, Mar V, Zhou W, Oulton R, Amgen EST, Bass MB, Arruda I, Robinson MO. A mammalian telomerase-associated protein. Science 1997;275:973–977.

Heng HHQ, Tsui L-C. Modes of DAPI banding and simultaneous in situ hybridization. Chromosoma 1993;102:325–332.

Heng HHQ, Squire J, Tsui L-C. High resolution mapping of mammalian genes by in situ hybridization to free chromatin. Proc. Natl. Acad. Sci. USA 1992;89:9509–9513.

Hogan B, Beddington R, Costantini F, Lacy E. Manipulating the Mouse Embryo: A Laboratory Manual, 2nd Ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1994.

Kerr WG, Heller M, Herzenberg LA. Analysis of lipopolysaccharide-response genes in B-lineage cells demonstrates that they can have differentiation stage-restricted expression and contain SH2 domains. Proc. Natl. Acad. Sci. USA 1996;93:3947–3952.

Krull CE, Collazo A, Fraser SE, Bronner-Fraser M. Segmental migration of trunk neural crest: Time-lapse analysis reveals a role for PNA-binding molecules. Development 1995;121:3733–3743.

Lai MMC. The molecular biology of hepatitis delta virus. Annu. Rev. Biochem. 1995;64:259–286.

Lee YM, Osumi-Yamashita N, Ninomiya Y, Moon CK, Eriksson U, Eto K. Retinoic acid stage-dependently alters the migration pattern and identity of hindbrain neural crest cells. Development 1995;121:825–837.

Lohnes D, Mark M, Mendelsohn C, Dolli P, Dierich A, Gorry P, Gansmuller A, Chambon P. Function of the retinoic acid receptors (RAR) during development. I. Craniofacial and skeletal abnormalities in RAR double mutants. Development 1994;120:2723–2748.

Lyon MF, Rastan S, Brown SDM. Genetic Variants and Strains of the Laboratory Mouse, 3rd Ed. Oxford: Oxford University Press, 1996.

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM. The nuclear receptor superfamily: The second decade. Cell 1995;83:835–839.

Morriss-Kay G. Retinoic acid and craniofacial development: Molecules and morphogenesis. Bioessays 1993;15:9–15.

Ogura T, Alvarez IS, Vogel A, Rodriguez C, Evans RM, Belmonte JCI. Evidence that *Shh* cooperates with a retinoic acid inducible co-factor to establish ZPA-like activity. Development 1996;122:537–542.

Perrimon N. Hedgehog and beyond. Cell 1995;80:517–520.

Poch O, Sauvager I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 19898:3867–3874.

Robbins J, Dilworth SM, Laskey RA, Dingwall C. Two interdependent basic domains in nucleoplasm in nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell 1991;64:615–623.

Ross J. Control of messenger RNA stability in higher eukaryotes. TIG 1996;12:171–175.

Ruberte E, Dolle P, Chambon P, Morriss KG. Retinoic acid receptors and cellular retinoid binding proteins. II. Their differential pattern of transcription during early morphogenesis in mouse embryos. Development 1991;111:45–60.

Sam M, Wurst W, Forrester L, Vauti F, Heng H, Bernstein A. A novel family of repeat sequences in the mouse genome responsive to retinoic acid. Mammalian Genome 1996;7:741–748.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.

Schiebel W, Haas B, Marinkovic S, Klanner A, Sanger HL. RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties. J. Biol. Chem. 1993a;268:11851–11857.

Schiebel W, Haas B, Marinkovic S, Klanner A, Sanger HL. RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. J. Biol. Chem. 1993b;268:11858–11867.

Schilling TF. Genetic analysis of craniofacial development in the vertebrate embryo. Bioessays 1997;19:459–468.

Serbedzija GN, Fraser SE, Bronner-Fraser M. Pathways of trunk neural crest cell migration in the mouse embryo as revealed by vital dye labelling. Development 1990;108:605–612.

Serbedzija GN, Bronner-Fraser M, Fraser SE. Vital dye analysis of cranial neural crest cell migration in the mouse embryo. Development 1992;116:297–307.

Shay JW. Telomerase and cancer. TIG 1996;12:486–487.

Simeone A, Acampora D, Arcioni L, Andrews PW, Boncinelli E, Mavilio F. Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. Nature 1990;346:763–766.

Skarnes WC, Moss JE, Hurtley SM, Beddington RSP. Capturing genes encoding membrane and secreted proteins important for mouse development. Proc. Natl. Acad. Sci. USA 1995;92:6592–6596.

Sousa R. Structural and mechanistic relationships between nucleic acid polymerases. TIBS 1996;21:186–190.

Takeuchi T, Yamazaki Y, Katoh FY, Tsuchiya R, Kondo S, Motoyama J, Higashinakagawa T. Gene trap capture of a novel mouse gene, *jumonji*, required for neural tube formation. Genes Dev. 1995;9:1211–1222.

Tini M, Otulakowski G, Breitman ML, Tsui LC, Giguere V. An everted repeat mediates retinoic acid induction of the γF-crystallin gene: Evidence of a direct role for retinoids in lens development. Genes Dev. 1993;7:295–307.

Torres M, Stoykova A, Huber O, Chowdhury K, Bonaldo P, Mansouri A, Butz S, Kemler R, Gruss P. An α-E-catenin gene trap mutation defines its function in preimplantation development. Proc. Natl. Acad. Sci. USA 1997;94:901–906.

Trainor PA, Tam PPL. Cranial mesoderm and neural crest cells of the mouse embryo: Co-distribution in the craniofacial mesenchyme but distinct segregation in branchial arches. Development 1995;121: 2569–2582.

Volloch V, Schweitzer B, Zhang X, Rits S. Identification of negative-strand complements to cytochrome oxidase subunit III RNA in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA 1991;88:10671–10675.

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

Zhang J, Donaldson SH, Serbedzija G, Elsemore J, Plehn-Dujowich D, McMahon AP, Flavell RA, Williams T. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. Nature 1996;381:238–241.