Positional Cloning and Characterization of a Paired Box– and Homeobox-Containing Gene from the Aniridia Region

Carl C. T. Ton,* Harri Hirvonen,† Hiroshi Miwa,*
Michael M. Weil,* Paula Monaghan,‡ Tim Jordan,‡
Veronica van Heyningen,‡ Nicholas D. Hastie,‡
Hanne Meijers-Heijboer,§ Matthias Drechsler,§
Brigitte Royer-Pokora,§ Francis Collins,‖
Anand Swaroop,# Louise C. Strong,+
and Grady F. Saunders*
*Department of Biochemistry and Molecular Biology
+Department of Experimental Pediatrics
University of Texas M. D. Anderson Cancer Center
Houston, Texas 77030
†Department of Medical Biochemistry
University of Turku
Turku SF-20520
Finland
‡MRC Human Genetics Unit
Western General Hospital
Edinburgh EH4 2XU
Scotland
§Institut für Humangenetik und Anthropologie
Ruprecht-Karls Universität
6900 Heidelberg 1
Federal Republic of Germany
‖Departments of Internal Medicine and Human Genetics
Howard Hughes Medical Institute
#Departments of Ophthalmology and Human Genetics
University of Michigan
Ann Arbor, Michigan 48109

Summary

Based on the map location of the aniridia (AN) locus in human chromosomal band 11p13, we have cloned a candidate AN cDNA (D11S812E) that is completely or partially deleted in two patients with AN. The <70 kb smallest region of overlap between the two deletions encompasses the 3' coding region of the cDNA. This cDNA, which spans over 50 kb of genomic DNA, detects a 2.7 kb message specifically within all tissues affected in AN. The predicted polypeptide product possesses a paired domain, a homeodomain, and a serine/threonine-rich carboxy-terminal domain, structural motifs characteristic of certain transcription factors. The concordance between expression and pathology, map location, structure, and predicted function argues that the cDNA corresponds to the AN gene.

Introduction

Aniridia (AN) is a congenital, bilateral panocular disorder characterized by the complete or partial absence of the iris. Vision is often impaired by other associated abnormalities such as foveal and optic nerve hypoplasia and may progressively deteriorate, owing to cataracts, corneal opacification, and early onset of glaucoma (Nelson et al., 1984). Population studies in Michigan and Denmark show the incidence of AN to range between 1:64,000 and 1:96,000. About two-thirds of the children with AN have affected parents, with the disorder being inherited as an autosomal dominant trait with variable expressivity (Shaw et al., 1960). In the one reported case of a mating between two unrelated people with familial AN, a stillborn fetus with severe craniofacial abnormalities, marked by a total absence of eyes, nose, and adrenal glands, was produced (Hodgson and Saunders, 1980). These observations suggest that the AN trait might be lethal in the homozygous state.

Mannens et al. (1989) have shown evidence of linkage between AN and the 11p13 markers catalase, follicle cell–stimulating hormone β subunit, and two random DNA sequences (D11S16 and D11S151) in a large kindred transmitting AN with normal karyotypes. Although Ferrell et al. (1980) had previously reported genetic linkage between autosomal dominant AN ("AN1") and the chromosome 2 marker acid phosphatase-1 (ACP1), a reexamination of the same pedigree with additional 2p markers revealed no further indication of linkage to chromosome 2. Analysis with 11p13 markers, however, showed evidence of linkage to D11S323 and suggestion of linkage to D11S325 (L. A. Lyons et al., submitted). These findings indicate that the ACP1 linkage (and thus the chromosome 2 AN1 locus) could be spurious and that AN in this family might be determined by the AN locus in 11p13.

About a third of individuals with the sporadic form of AN also develop Wilms' tumor. Those with AN–Wilms' tumor are usually mentally retarded and often suffer from genitourinary abnormalities (Fraumeni and Glass, 1968). Patients with this constellation of systemic disorders, typical of the WAGR (Wilms' tumor, AN, genitourinary abnormalities, mental retardation) contiguous gene syndrome (Miller et al., 1964), are usually heterozygous for constitutional deletions of chromosomal band 11p13 (Riccardi et al., 1978; Francke et al., 1979). Molecular analyses of different WAGR-related deletions and translocations with 11p13 probes have made possible the localization of the WAGR complex to an area of several Mb between catalase and follicle cell–stimulating hormone β subunit (Glaser et al., 1986; Porteous et al., 1987) and resolution of two of its genetic components, AN and Wilms' tumor, into separate loci (Davis et al., 1988). Subsequent long-range physical mapping studies of the WAGR region revealed a series of NotI restriction fragments of sizes 500, 375, 325, and 1400 kb running in the telomeric direction. Analysis of deletion and translocation breakpoints in AN patients served to map the AN locus to the proximal end of the 1400 kb fragment near its junction with the 325 kb fragment (Compton et al., 1988; Gessler and Bruns, 1989).

Two families in which there was concordant transmission of AN with translocations involving 11p13 have been described, suggesting that these rearrangements might either physically disrupt the AN gene or impair its function through positional effects (Simola et al., 1983; Moore et al., 1986). The breakpoint of one of these translocations
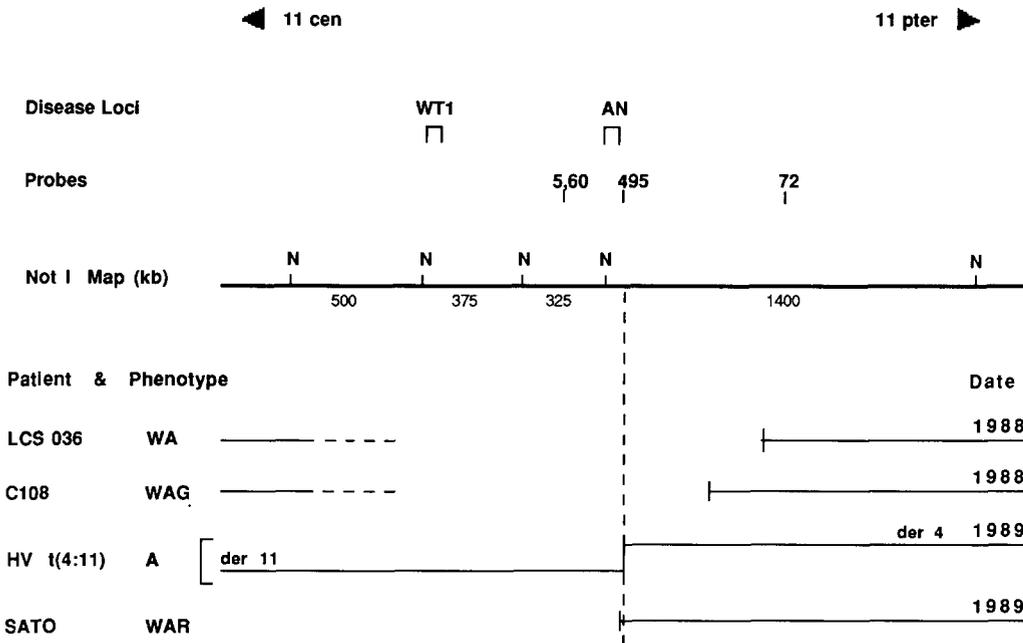
Figure 1. Map Location of the AN Locus within the WAGR Region

A NotI long-range physical map of the 11p13 WAGR region depicting the position of the AN cDNA and the WT1 Wilms' tumor susceptibility locus. Included are examples of some DNA probes (5, 60, and 72 [Compton et al., 1988]; 495 [Gessler et al., 1989]) within the region and the AN-associated chromosomal aberrations that were used in definition of the AN locus. The AN locus is denoted here as the AN locus since there is evidence that "AN1" does not exist (see text). The extent of the large 11p13 deletions in LCS 036, C108 (Compton et al., 1988), and SATO (Junien and McBride, 1989) and the detailed mapping of the HV translocation breakpoint (Davis et al., 1989) have been previously reported. The patient phenotypes are listed as W (Wilms' tumor), A (AN), G (genitourinary abnormalities), or R (mental retardation).

(t(4;11)(q23;p13)) from the patient HV has been cloned recently (Gessler et al., 1989), with its proximal terminus on derivative chromosome 11 defined by the probe 495 (D11S310). Initial sequence analysis of the interrupted region showed that the breakpoint was situated within an open reading frame (ORF) flanked by donor and acceptor splice sites, suggesting that the latter may represent an exon. However, the message corresponding to this putative exon has not been reported.

Using the DNA probes 1104 (D11S93) and 706 (D11S95), which flank the HV breakpoint, Davis et al. (1989) were able to deduce by long-range restriction mapping that the translocation occurred within about 100 kb of the proximal terminus of the 1400 kb NotI restriction fragment. However, in the human–mouse hybrid SAX, derived from the patient SATO (van Heyningen et al., 1985) with Wilms' tumor, complete AN, mental retardation, and a constitutional deletion of 11p13, the probe 495 was mapped outside of the distal deletion breakpoint (Junien and McBride, 1989). This situation therefore seems to suggest a more proximal location for the AN locus relative to the HV breakpoint, or that there may be more than one 11p13 gene involved in AN.

Small eye (Sey) is an autosomal dominant mutation of the mouse that affects craniofacial development, in particular, the embryonic development of the eyes and nose. Heterozygous Sey/+ mice and individuals with AN have similar eye defects, suffering chiefly from a complete or partial absence of iris tissue, though the phenotype in the mouse with respect to the entire eye can be much more severe. Homozygous Sey embryos are born without eyes and nose, and die soon after (Hogan et al., 1986, 1988), a phenotype comparable with that in the single reported case of probable homozygous AN in humans. In addition to these phenotypic parallels, there is both linkage (van der Meer-de Jong et al., 1990) and physical (Glaser et al., 1990) mapping data suggesting that the Sey gene is the murine homolog of the human AN gene.

We describe here the molecular cloning and characterization of a complementary DNA that spans over 50 kb of genomic DNA at the junction between the 1400 kb and 325 kb NotI restriction fragments. We report elsewhere (C. C. T. T. et al., submitted) on the cloning and characterization of a homologous murine cDNA as a candidate gene for the Sey locus. The human cDNA is eliminated by a deletion in the patient SATO whose distal deletion breakpoint lies about 30 kb telomeric of the 3' end of the cDNA, and is truncated at its 3' end by a deletion in another AN patient HO. The predicted polypeptide product of this cDNA possesses two specific DNA-binding domains, the paired domain and the homeodomain, features that suggest that it may be a transcription factor. Its map location, involvement in AN-associated chromosomal aberrations, specific expression in apparently all tissues affected in the disease, and its predicted function argue that this cDNA is a strong candidate for the AN gene.

Results

Approach to the AN Locus: Chromosome Jumping and Walking

We chose to use a positional cloning strategy in our search for the AN gene, directing our movement toward the estimated map location of the AN locus as defined by AN-associated chromosomal translocations and deletions (Figure 1). Based on such map information, we attempted to approach the AN locus from the nearest markers available to us (e.g., p5 [D11S323]) by means of the chromosome jumping technique (Collins et al., 1987). Pulsed field gel electrophoresis (PFGE) showed that the probe p5 BE0.8 maps to a 180–200 kb SfiI fragment located within the 325 kb NotI fragment (see Figure 3B). Direction-specific chromosome jumping entailed, first, the expansion of the D11S323 locus by chromosome walking to permit the selection of appropriate starting probes (p5 BE1.3 and p5 S1.6). This then enabled us to exploit the topological properties of the jumping library to execute two contradirectional jumps from the D11S323 locus, yielding the jump clones J5A1 and J5A6, respectively. The endpoint of the J5A1 jump gave a single-copy probe J5A1 EH0.8, which detected the same 325 kb NotI and 180 kb SfiI fragments as the starting probe (see Figure 3B). J5A1 was then mapped by a PaeR71–NotI double digest to a position within 9.5 kb of the NotI sites at the terminus of the 325 kb NotI fragment (Figure 2B). Similarly, by means of a BamHI–SfiI double digest, the J5A6 jump was found to have landed within 3.5 kb of the SfiI site within the 325 kb NotI fragment (Figure 2C). PFGE analysis indicates that the bidirectional jumps covered a total of 180–200 kb, as summarized in Figure 2A.

Using J5A1 as the starting point, we initiated chromosome walking in the telomeric direction. In the process, we generated a phage and cosmid contig of approximately 42 kb (Figure 3A), spanning two genomic NotI sites (separated by 16 kb) and extending from the 325 kb fragment into the 1400 kb fragment, as indicated by PFGE using the probes J5A1, pH2, and pH1-14 (Figures 3B and 3C). These probes detected, respectively, the 325 kb NotI fragment, a less than 23 kb sub-PFGE NotI fragment, and the 1400 kb NotI and BssHII fragments previously reported (Gessler and Bruns, 1989), thus indicating transit through the 325/1400 junction region. Four subcloned single-copy DNA sequences (Figure 3A) were found to meet two criteria that suggested that they might harbor candidate DNA sequences: the ability to detect cross-hybridizing sequences in other species (since DNA coding for genes tends to show evolutionary conservation) and the possession of CpG islands, which frequently mark the 5' end of mammalian genes (Bird, 1986). The presence of such islands in the subclones was revealed by a clustering in genomic DNA of recognition sites for rare-cutting restriction enzymes (e.g., HpaII, SalI, EagI, and PaeR71) that cleave only at unmethylated CpG motifs (Lindsay and Bird, 1987). However, we could proceed no further beyond this contig since the region immediately telomeric appeared not to be readily clonable from the genomic and jumping libraries

(both amplified and unamplified) that we used. It is possible that this area could be part of a genomic region containing DNA with unusual secondary structure or repetitive elements that are unstable in bacterial hosts (Wyman et al., 1986).

Complementary DNA Clones

An 8.5 day mouse whole embryo cDNA library was screened with the subcloned genomic DNA sequences. Of the four sequences tested, only the most distal one on the 1400 kb fragment (pH1-14 NB7.0) yielded positive clones. The insert sizes of the cDNAs ranged from 0.6 to 0.85 kb. The largest 0.85 kb insert was used to rescreen the cDNA library, producing 18 positive clones of which the largest was 2.1 kb in size. A subcloned 1.6 EcoRI fragment (pm1) from this murine cDNA was able to detect homologous sequences in human chromosome band 11p13 (EcoRI fragment sizes: 4.5 kb, 5.6 kb, and 7.5 kb) (Figure 4A). Furthermore, it was found to be evolutionarily conserved in all vertebrate and invertebrate species tested, from hamster to fruit fly and sea urchin (Figure 4B). The properties of the mouse cDNA (pm1) and its status as a candidate gene for the murine Sey genetic disorder are discussed elsewhere (C. C. T. T. et al., submitted). The pm1 clone was subsequently used to screen a human fetal eye cDNA library, yielding a total of 21 positive clones corresponding to five different overlapping cDNAs ranging in size from 0.7 to 1.0 kb (ph12) and 1.4 kb (ph1) (D11S182E). The genomic clone pH1-14 was also used to screen an adult retinal cDNA library to produce the clone pTP1, which detects the same genomic restriction fragments as pm1. The total span of the human cDNA clones is 1.6 kb. The partial mouse cDNA, while lacking 131 bp of 5' coding sequence, is longer than the human composite sequence by 473 bp in the 3' untranslated region. Southern blot analysis with fragments from ph1 and ph12 indicates that ph1 spans the two NotI sites marking the 325/1400 junction region such that 5' sequences of the cDNA are located on the 325 kb fragment while a major portion of the cDNA resides on the 1400 kb fragment. The sum of BamHI restriction fragments detected and spanned by the composite human and mouse cDNAs exceeds 50 kb (data not shown).

Analysis of AN Patient DNA

The human cDNAs were used to probe Southern blots of DNA from the 11p13 deletion patient SATO (with AN, Wilms' tumor, and mental retardation), the t(4;11) patient HV, also with familial AN (cell line name SIMO), two other patients with familial AN, and four patients with sporadic AN. Figure 5A shows the results obtained with DNA derived from the patients SATO and HV; no gross alterations were seen with the other patients (data not shown). The intensity of pTP1 hybridization in SATO for all three human bands (3 kb, 2.4 kb, and 1.8 kb) is half that in FATO and MATO (father and mother of SATO, respectively), reflecting a hemizygous state at this locus. Also, in the human–mouse hybrid SAX, which contains only the deleted 11 homolog, the human bands were all absent with no
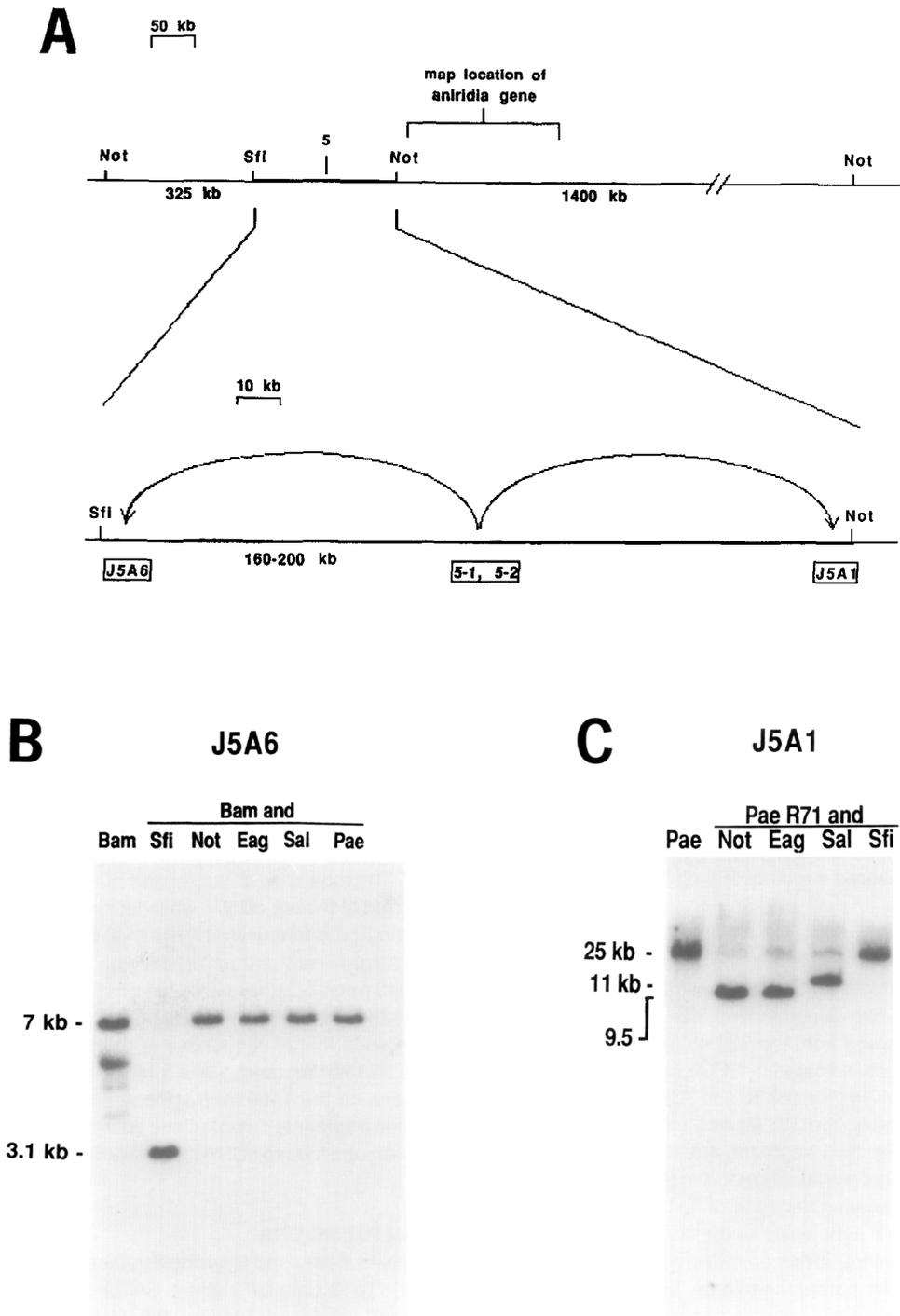
A

50 kb

map location of
aniridia gene

Not   Sfi   5   Not   Not

325 kb   1400 kb

10 kb

Sfi   Not

J5A6   160-200 kb   5-1, 5-2   J5A1

B   **J5A6**

Bam and

Bam Sfi Not Eag Sal Pae

7 kb -

3.1 kb -

C   **J5A1**

Pae R71 and

Pae Not Eag Sal Sfi

25 kb -
11 kb -

9.5

Figure 2. Approach to the AN Locus by Chromosome Jumping

(A) Expanded view of the junction between the 325 kb and 1400 kb Notl restriction fragments and schematic diagram of bidirectional jumps from the extended 5 (D11S323) locus toward the termini of the 160–200 kb Notl–Sfil fragment.

(B) Southern blot analysis with the jump probe J5A1 against genomic DNA doubly digested with PaeR71 and a range of rare-cutting enzymes (from left to right): PaeR71 (Pae), PaeR71 and Notl (Not), Eagl (Eag), Sall (Sal), Sfil (Sfi). Sizes of bands are indicated in kb (left).

(C) Southern blot with J5A6 against genomic DNA doubly digested with BamHI and a range of rare cutters (from left to right): BamHI (Bam), BamHI and Sfil (Sfi), Notl (Not), Eagl (Eag), Sall (Sal), and PaeR71 (Pae). The probes J5A1 and J5A6 detect the same 325 kb Notl and 180 kb Sfil restriction fragments (Figure 3B).
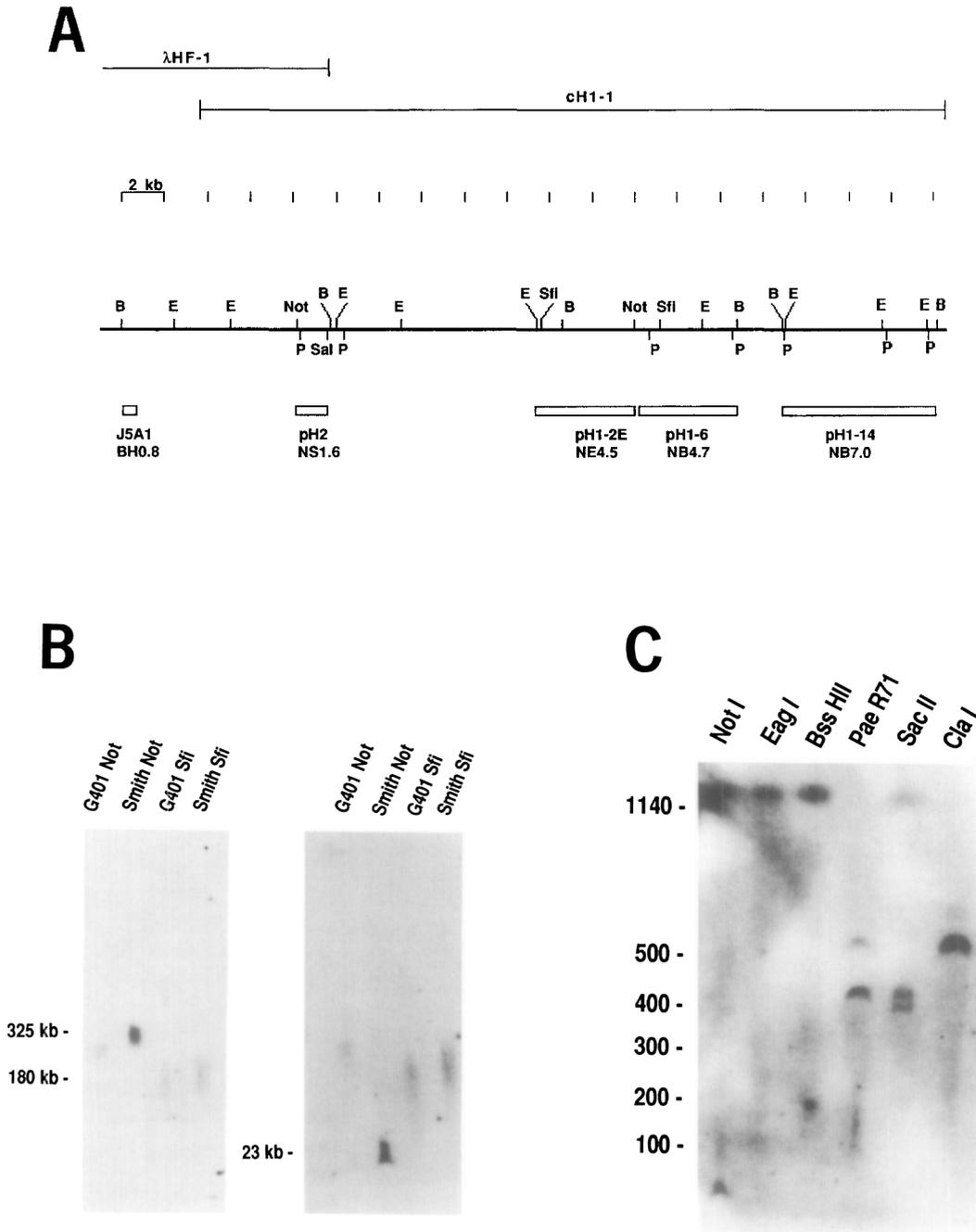
A

B

C

Figure 3. Chromosome Walking across the 325/1400 kb NotI Junction

(A) Restriction map of part of the 42 kb phage and cosmid contig spanning the junction between the 325 and 1400 kb NotI restriction fragments (the telomeric direction is left to right). The junction is defined by two NotI sites about 16 kb apart that exhibit different methylation patterns (unpublished data) in different individuals and tissues. Other restriction sites are: BamHI (B), EcoRI (E), PstI (P), SalI (Sal), and SfiI (Sfi). Open boxes at bottom represent conserved single-copy fragments subcloned from this contig.

(B) PFGE analysis of genomic DNA from two sources (G401, a Wilms' tumor cell line with an intact chromosome 11 [Weissman et al., 1987], and Smith, a Wilms' tumor patient with normal karyotype [peripheral blood lymphocytes]) digested with NotI and SfiI and hybridized the probes J5A1, p5, or J5A6 (left) and pH2 (right). DNA from G401 and Smith were used in order to control for the known inhibitory effects of methylated CpG dinucleotides on the activity of rare cutter enzymes.

(C) PFGE Southern blot with the probe pH1-14 detects the 1400 kb NotI and BssHII fragments reported previously (Gessler and Bruns, 1989).

indication of an altered fragment. This confirms that the DNA encompassed by the partial candidate AN cDNA ("AN cDNA") is completely lost through the 11p13 deletion in SATO. This deletion, however, spares the 495 probe lo-

cated at the t(4;11) breakpoint in SIMO (Junien and McBride, 1989; Gessler and Bruns, personal communication), which is estimated to be less than 100 kb from the 325/1400 kb NotI junction. Based on the sum of the restric-
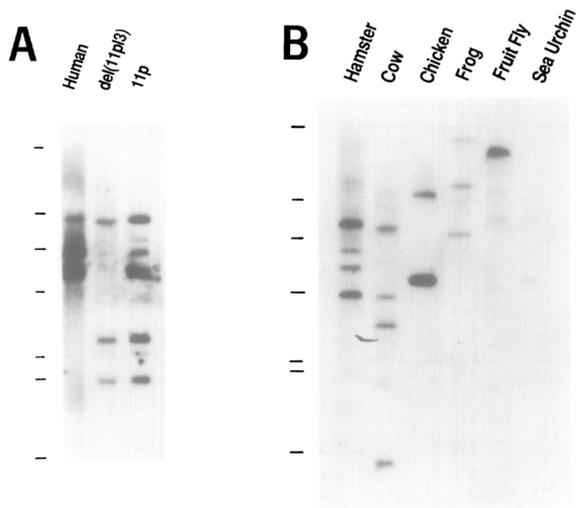
Figure 4. Southern Blot Analysis with the AN cDNA

(A) Southern blot of EcoRI-digested genomic DNA from normal human control (Human), the C2-1 human hamster hybrid containing a single chromosome 11 homolog deleted for 11p13 (del11(p13)), and the 901-110 human–hamster hybrid that contains only the short arm of chromosome 11 as its human complement (11p), probed with the ph12 AN cDNA. Size markers (lambda DNA cut with HindIII, kb) are shown on the left.

(B) Zoo blot of EcoRI-digested genomic DNA from a range of animal species hybridized with the ph12 AN cDNA. HindIII-digested lambda size (kb) markers are displayed on the left. Hybridization was conducted at reduced stringency.

tion fragments spanned by the cDNA, it is estimated that the distal deletion breakpoint in SATO lies about 30 kb from the 3' end of the gene.

In the case of the SIMO translocation cell line, however, Southern blot analyses with the cDNA clone pTP1, which extends as far as ph12 in the telomeric direction, revealed no altered bands. The presence of all the human bands in SMX B2, a somatic cell hybrid containing the chromosome 11 derivative, indicates that the cDNA lies wholly proximal of the translocation breakpoint. Since the cDNA is probably not full length at the 3' end (there is no definitive polyadenylation signal and poly(A)+ tail), it is unclear at this point whether the translocation physically disrupts the gene.

In one AN family (HO) with affected mother and son, the probes D11S16 (Feder et al., 1985) and G12 (Rose et al., 1990), which map to the distal and central thirds, respectively, of the 1400 kb NotI restriction fragment, hybridized to both the normal 1400 kb fragment and an altered 1100 kb fragment (Figure 5B) on a PFGE blot. This observation indicates the presence of an approximately 300 kb interstitial deletion of the proximal third of the 1400 kb NotI fragment. Southern blot analysis with a 1.4 kb HindIII fragment (probe 1) subcloned from the distal end of clone pH1-14 revealed rearranged fragments in the HO patients with the enzymes AccI, BamHI, EcoRI, and EcoRV (Figure 5C), while another genomic probe corresponding to the 3' end of the pm1 cDNA detected only the normal bands at reduced intensity (data not shown). These data show that

the proximal deletion breakpoint in HO truncates the AN gene within pH1-14 and eliminates the 3' end of the cDNA. The positional interrelationships between the AN gene, the SATO and HO deletions, their smallest region of overlap (SRO), and the SIMO translocation are summarized in Figure 5D.

Tissue-Specific Expression Pattern

The pm1 (1.6 kb) mouse cDNA was used to probe Northern blots of total RNA from a range of human fetal tissues from 17–20 wk therapeutic abortuses. The cDNA probe detected high levels of a 2.7 kb transcript specifically in the eyes and brain (Figure 6A), although a faint signal was also detected in the pancreas. Within the eye (Figure 6B), the strongest signal was seen in the neuroretina, while the pigment retina gave a much weaker one (<10%) that, given the proximity of the two tissues, could be due to contamination. No detectable levels of the transcript were seen in the sclera. In the brain, the cerebellum and pons had the highest amounts of message, while the temporal lobe, midregions, medulla oblongata, and choroid plexus all had less than 10% in comparison. Significantly, the transcript was also clearly present in the olfactory bulb, which was identified as one of the affected tissues in cases of homozygous AN and Sey. There was no discernable message in the meninges. It is interesting that a distinct signal can be seen in the intermediate brain layers, which consist of migratory, undifferentiated neuroepithelial cells, but is absent from the cortical plate that is abundant in differentiating neuroblasts. This suggests that within the brain the gene might be down-regulated during differentiation.

RNA In Situ Hybridization

Antisense riboprobe generated from the pm1 and pTP1 cDNAs showed strong hybridization signal in the human eye at 49 days of gestation; specifically, in the neural retina, the rim portion of the optic cup, i.e., the pars iridica retinae and pars ciliaris retinae (which constitute the neuroectodermal component of the presumptive iris/ciliary body) overlapping the lens (Last, 1968; Naumann and Apple, 1986), the lens itself, and the surface ectoderm that will form the cornea and/or conjunctiva (Figure 7). The apparent signal in the pigment epithelial layer of the optic cup is an artifact, since pigment granules are refractile in the same manner as silver grains under dark field imaging. In control albino mice where there is no pigment in the outer epithelial layer, no signal is seen at this site. Other parts of the eye show a similar pattern of gene expression in mouse and man. In carefully staged mouse embryos at 10 days of gestation, signal can also be seen in the prosencephalon and in parts of the neural tube (data not shown). The prosencephalon is the precursor of the diencephalon, from which the optic and olfactory rudiments are derived.

Sequence Analysis and Predicted Structure

Figure 8A shows the composite sequence of the transcript obtained from the overlapping fetal eye cDNA clones ph1 and ph12, with a total length of 1643 bp. There is an ATG
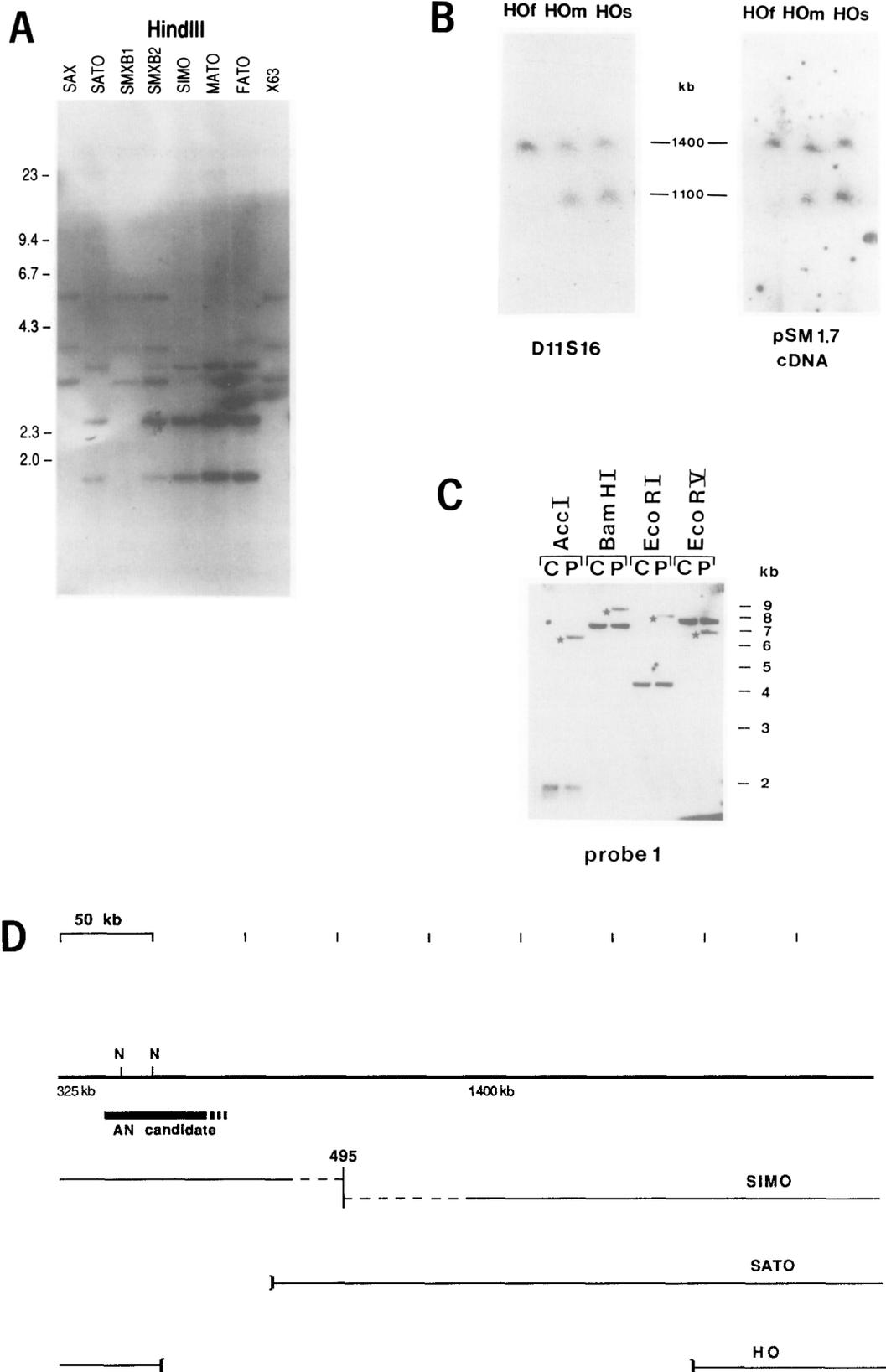
Figure 5. Southern Blot Analysis of Patient DNA

(A) Southern blot of patient and control DNA digested with HindIII and hybridized with the pTP1 AN cDNA. The samples are (left to right): SAX (human–mouse hybrid containing the 11p13 deleted homolog from SATO), SATO (EBV-transformed cell line), SMXB1 and SMXB2 (human–mouse somatic cell hybrids containing the der4 and der11 of the HV translocation), SIMO (EBV-transformed cell line from patient HV), MATO and FATO (mother and father of SATO, respectively), and X63 (mouse parental cell line). Molecular weight markers (kb) are on the left.

(B) PFG blot of DNA from the AN family HO probed with D11S16 or G12 (left panel) and with the murine cDNA clone pm1 (pSM1.7). HOf, normal father; HOm, mother with complete AN only; and HOs, affected son.

(C) Southern blot of control (C) and patient HO (P) DNA digested with AccI, BamHI, EcoRI, and EcoRV probed with a 1.4 kb HindIII fragment (probe 1) subcloned from the distal part of the genomic fragment pH1-14.

(D) Diagram showing the interrelationship between the AN cDNA, the SATO and HO deletions, and the SIMO translocation.
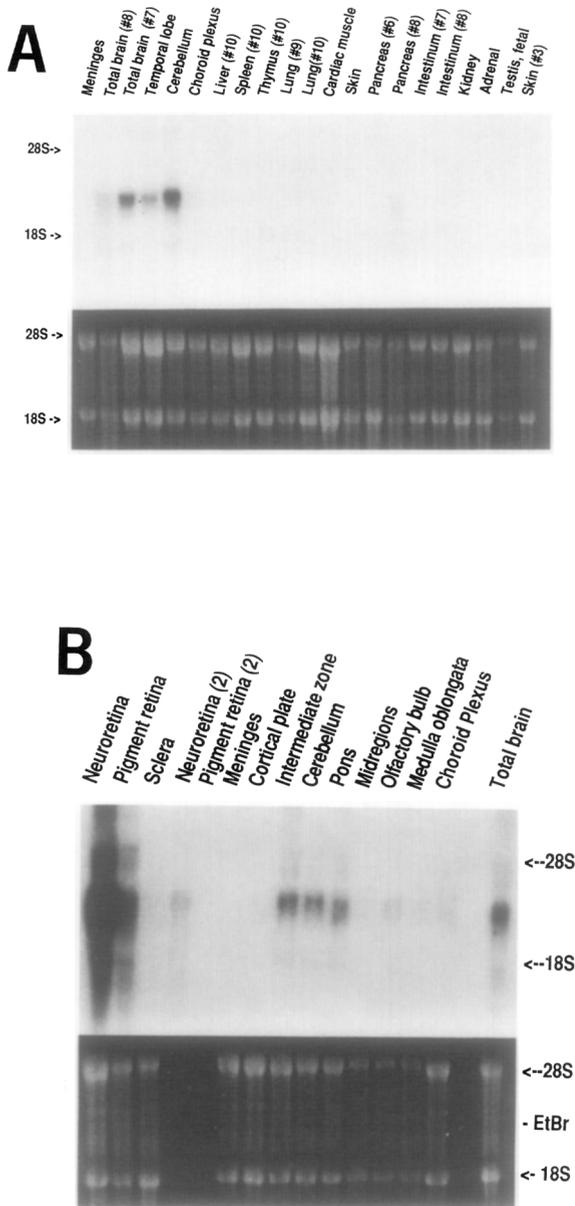
A

B

Figure 6. Expression Pattern of AN mRNA

(A) Northern blot of 17–20 week human fetal RNA from a range of tissues hybridized with the 1.6 kb pm1 mouse cDNA (left to right): meninges, two samples of total brain (specimen #8 is degraded, see ethidium-stained gel directly below), temporal lobe, cerebellum, choroid plexus, liver, spleen, thymus, two samples of lung, cardiac muscle, skin, two samples of pancreas and intestine, kidneys, whole adrenals, testis, and skin. The pm1 cDNA starts with the EcoRI site at the equivalent of nucleotide position 493 in the composite human sequence. Strong signals are detected in total brain, particularly in the cerebellum, and to a lesser degree in temporal lobes (<10%). A faint band is present in pancreas sample #8.

(B) Northern blot of RNA from tissues from the developing eye and brain hybridized with the pm1 cDNA (left to right): dissected neuroretina and pigment retina, sclera, a second (partially degraded) sample of neuroretina and pigment retina, meninges, cortical plate (differentiating neuroblasts), intermediate zone (mitotically active migratory neuroepithelial cells), cerebellum, pons, midregions, olfactory bulbs, medulla oblongata, choroid plexus, and total brain (control). The strongest signal by far is in the neuroretina. It is unclear if the band in pigment retina is due to contamination from the neuroretina. The positions of

codon at nucleotide position 363 with good context in comparison with the GCCA/GCCATGG Kozak consensus (Kozak, 1987). It initiates an extended ORF of 1266 nucleotides with a coding capacity of 422 amino acids. However, a search of the sequence for inverted repeats capable of forming secondary structures revealed two potential stem-loops: one 17 bp stem starting at position 389 (22 bp downstream of the ATG) and another one of 10 bp at position 301, 10 bp down from a nonoptimal in-frame GTGAGAGTGG initiation site. If the 301 stem-loop hinders ribosome scanning appropriately to promote initiation at the in-frame GTG site (Kozak, 1990), an extra 25 amino acids starting with a valine would be included in the polypeptide sequence, bringing the length to 447 residues. Some heterogeneity in length and sequence in the 5' untranslated leader upstream of the putative initiation codons was seen among the cDNA clones examined (unpublished data). There was no apparent canonical AATAAA polyadenylation signal, although the 3'-most cDNA (ph12) ends in a run of A's. It is uncertain whether this polyadenosine tract is internal to the message or a true poly(A) tail; a comparison with the mouse sequence (C. C. T. T. et al., submitted), which extends at least 473 bp further 3' without showing a poly(A) tract, suggests that it is not. The overall structural organization of the cDNA is shown in Figure 8B.

The conceptually translated polypeptide sequence possesses two interesting structural features. Nucleotides 436 to 685 were found to encode an 82 amino acid paired domain, a triple α helical motif recently shown to possess specific DNA-binding properties (Treisman et al., 1991). The paired box was first identified in the Drosophila segmentation genes paired (prd) (Kilcherr et al., 1986; Frigerio et al., 1986), gooseberry-distal (gsb), and gooseberry-proximal (Baumgartner et al., 1987; Côté et al., 1987), suggesting a functional role for this domain in Drosophila morphogenesis. Subsequently, homologs of the paired box have been discovered in the Pax family of developmental genes in mouse (reviewed by Kessel and Gruss, 1990) and in the HuP genes in humans (Burri et al., 1989). The predicted secondary structure of the paired domain consists of three α helices in which the first helix is separated from the others by a highly flexible region of 49 residues and the two remaining ones joined via a poorly defined α helical or β-pleated sheet structure indicative of a turn (Burri et al., 1989). It was shown by Treisman et al. (1991) that the sequence-specific DNA-binding activity of the paired domain in the prd protein resided wholly within the amino-terminal region of the structure, which includes the first of the three α helices (Figure 9A). The amino acid sequence of the helical core of the AN paired domain was aligned and compared with those of the prototypic prd protein and the murine Pax polypeptides (Figure 9A), generating a 45 amino acid consensus over the 82-residue domain. It is evident from inspection that AN is a member of the Pax gene family.

18S and 28S RNA are indicated on the right; ethidium bromide-stained gel before transfer is shown directly below.

Figure 7. RNA In Situ Hybridization of AN cDNA to Human Fetal Eye

Autoradiograph of antisense RNA probe generated from the pTP1 AN cDNA hybridized in situ to section of developing human eye at 49 days of gestation. Tissue section was stained with methyl green and riboprobe was labeled with [³⁵S]UTP. Silver grains visualized under dark field illumination were electronically recorded and superimposed in an arbitrary color (yellow) on the bright field image (Monaghan et al., 1991). Signal is seen over the neuroretina (interior layer of double-layered optic cup), the presumptive iris/ciliary body (corresponding to the lip of the optic cup overlapping the lens), the lens, and the surface ectoderm destined to form the transparent cornea and conjunctiva. Grains seen over the pigmented retina (outer layer of optic cup) are artifactual (see Results).

The second remarkable structural feature of the AN polypeptide is a paired-type homeodomain, a helix-turn-helix motif of 61 amino acid residues encoded by nucleotides 1017 to 1165. The 148 bp homeobox that codes for the homeodomain is a conserved DNA sequence first discovered in the *Antennapedia* (*Antp*) homeotic gene and other developmental control genes in Drosophila (McGinnis et al., 1984; Scott and Weiner, 1984). On the basis of this homology, other homeobox-containing developmental control genes have been isolated in diverse species ranging from yeast to humans (reviewed by Gehring, 1987; Scott et al., 1989). The homeodomain has been shown to adopt a helix-turn-helix conformation (Otting et al., 1988; Qian et al., 1989) in which helix 1 plays a stabilizing role, permitting the interaction of helix 2 with the sugar phosphate backbone of DNA and sequence-specific contacts between helix 3 and bases in the major groove (Hayashi and Scott, 1990). The primary function of homeodomain proteins appears to be the repression or activation of transcription at specific promoters (including those of the other homeobox genes as well as their own) to define

the spatial and temporal expression pattern of networks of developmental genes (reviewed by Hayashi and Scott, 1990). A comparison of the AN homeodomain with the *Antp* and *prd* (Figure 9B) prototypes reveals a greater similarity of AN to *prd* than to *Antp* and its murine *Hox* family homologs. AN and *prd* share with the *Hox* family and other Drosophila homeobox genes 11 invariant amino acids, but uniquely differ from all others in having a serine (S) residue in place of glutamine (Q) in the highly conserved WFQNRR motif of the DNA recognition helix. The AN polypeptide appears to be the first example of a human protein possessing both the paired domain and the homeodomain, although such an arrangement has been previously encountered in the Drosophila genes *prd* and *gsb*, and in the murine gene *Pax7* (Jostes et al., 1991). It has been shown that the two binding domains of *prd* can specifically bind to a bipartite recognition site in the promoter of *even-skipped* (*eve*). In addition, the C-terminal 140 amino acids possess a very high proportion (42/140; 30%) of serine and threonine residues, a feature reminiscent of the serine/threonine-rich domain in the transcription factor Oct-1

A

```
   1  TATCGATAAGTTTTTTTTTTATTGTCAATCTCTGTCTCCTTCCCAGGAATCTGAGGATTG
  61  CTCTTACACACCAACCCAGCAACATCCGTGGAGAAAACTCTCACCAGCAACTCCTTTAAA
 121  ACACCGTCATTTCAAACCATTGTGGTCTTCAAGCAACAACAGCAGCACAAAAAACCCCAA
 181  CCAAACAAAACTCTTGACAGAAGCTGTGACAACCAGAAAGGATGCCTCATAAAGGGGGAA
 241  GACTTTAACTAGGGGCGCGCAGATGTGTGAGGCCTTTTATTGTGAGAGTGGACAGACATC
 -25                                                    V  D  R  H  P
```

```
 301  CGAGATTTCAGAGCCCCATATTCGAGCCCCGTGGAATCCCGCGGCCCCCAGCCAGAGCCA
 -20   R  F  Q  S  P  I  F  E  P  R  G  I  P  R  P  P  A  R  A  S
```

```
 361  GCATGCAGAACAGTCACAGCGGAGTGAATCAGCTCGGTGGTGTCTTTGTCAACGGCGGC
   1   M  Q  N  S  H  S  G  V  N  Q  L  G  G  V  F  V  N  G  R  P
```

```
 421  CACTGCCGGACTCCACCCGGCAGAAGATTGTAGAGCTAGCTCACAGCGGGCCCGGCCGT
  21   L  P  D  S  T  R  Q  K  I  V  E  L  A  H  S  G  A  R  P  C
```

```
 481  GCGACATTTCCCGAATTCTGCAGGTGTCCAACGGATGTGTGAGTAAAATTCTGGGCAGGT
  41   D  I  S  R  I  L  Q  V  S  N  G  C  V  S  K  I  L  G  R  Y
```

```
 541  ATTACGAGACTGGCTCCATCAGACCCAGGGCAATCGGTGGTAGTAAACCGAGAGTAGCGA
  61   Y  E  T  G  S  I  R  P  R  A  I  G  G  S  K  P  R  V  A  T
```

```
 601  CTCCAGAAGTTGTAAGCAAAATAGCCCAGTATAAGCGGGAGTGCCCGTCCATCTTTGCTT
  81   P  E  V  V  S  K  I  A  Q  Y  K  R  E  C  P  S  I  F  A  W
```

```
 661  GGGAAATCCGAGACAGATTACTGTCCGAGGGGGTCTGTACCAACGATAACATACCAAGCG
  81   E  I  R  D  R  L  L  S  E  G  V  C  T  N  D  N  I  P  S  V
```

```
 721  TGTCATCAATAAACAGAGTTCTTCGCAACCTGGCTAGCGAAAAGCAACAGATGGGCGCAG
 101   S  S  I  N  R  V  L  R  N  L  A  S  E  K  Q  Q  M  G  A  D
```

```
 781  ACGGCATGTATGATAAACTAAGGATGTTGAACGGGCAGACCGGAAGCTGGGGCACCCGCC
 121   G  M  Y  D  K  L  R  M  L  N  G  Q  T  G  S  W  G  T  R  P
```

```
 841  CTGGTTGGTATCCGGGGACTTCGGTGCCAGGGCAACCTACGCAAGATGGCTGCCAGCAAC
 161   G  W  Y  P  G  T  S  V  P  G  Q  P  T  Q  D  G  C  Q  Q  Q
```

```
 901  AGGAAGGAGGGGGAGAGAATACCAACTCCATCAGTTCCAACGGAGAAGATTCAGATGAGG
 181   E  G  G  E  N  T  N  S  I  S  S  N  G  E  D  S  D  E  A
```

```
 961  CTCAAATGCGACTTCAGCTGAAGCGGAAGCTGCAAAGAAATAGAACATCCTTTACCCAAG
 201   Q  M  R  L  Q  L  K  R  K  L  Q  R  N  R  T  S  F  T  Q  E
```

```
1021  AGCAAATTGAGGCCCTGGAGAAAGAGTTTGAGAGAACCCATTATCCAGATGTGTTTGCCC
 221   Q  I  E  A  L  E  K  E  F  E  R  T  H  Y  P  D  V  F  A  R
```

```
1081  GAGAAAGACTAGCAGCCAAAATAGATCTACCTGAAGCAAGAATACAGGTATGGTTTTCTA
 241   E  R  L  A  A  K  I  D  L  P  E  A  R  I  Q  V  W  F  S  N
```

```
1141  ATCGAAGGGCCAAATGGAGAAGAGAAGAAAAACTGAGGAATCAGAGAAGACAGGCCAGCA
 261   R  R  A  K  W  R  R  E  E  K  L  R  N  Q  R  R  Q  A  S  N
```

```
1201  ACACACCTAGTCATATTCCTATCAGCAGTAGTTTCAGCACCAGTGTCTACCAACCAATTC
 281   T  P  S  H  I  P  I  S  S  S  F  S  T  S  V  Y  Q  P  I  P
```

```
1261  CACAACCCACCACACCGGTTTCCTCCTTCACATCTGGCTCCATGTTGGGCCTAACAGACA
 301   Q  P  T  T  P  V  S  S  F  T  S  G  S  M  L  G  L  T  D  T
```

```
1321  CAGCCCTCACAAACACCTACAGCGCTCTGCCGCCTATGCCCAGCTTCACCATGGCAAATA
 321   A  L  T  N  T  Y  S  A  L  P  P  M  P  S  F  T  M  A  N  N
```

```
1381  ACCTGCCTATGCAACCCCCAGTCCCCAGCCAGACCTCCTCATACTCCTGCATGCTGCCCA
 341   L  P  M  Q  P  P  V  P  S  Q  T  S  S  Y  S  C  M  L  P  T
```

```
1441  CCAGCCCTTCGGTGAATGGGCGGAGTTATGATACCTACACCCCCCACATATGCAGACAC
 361   S  P  S  V  N  G  R  S  Y  D  T  Y  T  P  P  H  M  Q  T  H
```

```
1501  ACATGAACAGTCAGCCAATGGGCACCTCGGGCACCACTTCAACAGGACTCATTTCCCCTG
 381   M  N  S  Q  P  M  G  T  S  G  T  T  S  T  G  L  I  S  P  G
```

```
1561  GTGTGTCAGTTCCAGTTCAAGTTCCCGGAAGTGAACCTGATATGTCTCAATACTGGCCAA
 401   V  S  V  P  V  Q  V  P  G  S  E  P  D  M  S  Q  Y  W  P  R
```

```
1621  GATTACAGTAAAAAAAAAAAAAAA
 421   L  Q  *
```
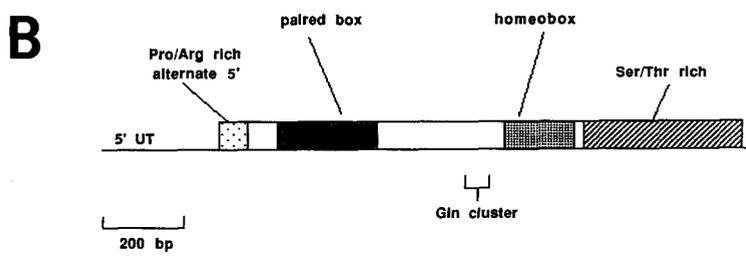
B

Figure 8. Sequence and Structural Organization of the AN cDNA

(A) The composite nucleotide sequence of the AN cDNA (1.62 kb) is shown, together with the predicted amino acid sequence of the ORF extending from position 288 to 1628. This ORF assumes that translation starts at the poor Kozak consensus at position 288, owing to the presence of a potential stem-loop structure arising from the inverted repeats (dotted nucleotides) starting 10 nucleotides downstream. An ATG with good consensus is located at position 363. The N-terminal paired domain (residues 26–88) and the downstream paired-type homeodomain (residues 211–260) are underlined. The asterisk marks the termination codon.

(B) Schematic diagram of the structural organization of the AN cDNA showing the 5' untranslated region (5' UT) and the long ORF (boxed region), subdivided into the paired box (dark stippled), homeobox (light stippled), the proline/arginine-rich (both 20%) alternative N-terminal peptide (box with dots), and the C-terminal serine/threonine-rich (30%) domain (hatched box). Scale in base pairs is indicated below.

A

COMPARISON OF PAIRED DOMAINS

Differences in Amino Acid Sequence from <u>Paired</u> Prototype

```
          α-helix
          --------
Prd    RLKIVEMA ADGIRPCVISRQHRVSHGCVSKILNRYQETGSIRPGVIGGSKPR*IATPE ...

AN     -Q----L- HS-A---D---ILQ--N-------G--Y-------RA------I*V---- ...

Pax7   -H------ HH----------L-----------C-----------A-------QV---D ...

Pax8   -QR--DL- HQ-V---D----L-----------G--Y---------------K*V---K ...

Pax2   -QR---L- HQ-V---D----L-----------G--Y-----K---------K*V---K ...

Pax1   -LR---L- QL-----D----L---------M-A--N-----L--A-------*VT--N ...

(Cons.) R..IV..A ..G.RPC.ISR...VS.GCVSK.L.RY.ETGSI.P..IGGSKP.*..TP. ...
```


```
          α-helix      turn      α-helix
          ---------    ----      ---------
Prd    ... IENRIEEYK  RSSPGMFS  WEIREKLIR

AN     ... VVSK-AQ--  -EC-SI-A  ----DR-LS

Pax7   ... V-KK-----  -EN-----  ----DR-LK

Pax8   ... VVEK-GD--  -QN-T--A  ----DR-LA

Pax2   ... VVDK-A---  -QN-T--A  ----AQ-LR

Pax1   ... VVKH-RD--  QGD--I-A  ----DR-LA

(Cons.) ... ....I..YK ...P..F.  WEIR..L..
```


B

COMPARISON OF HOMEODOMAINS

Differences in Amino Acid Sequence from <u>Antennapedia</u> Prototype

```
            Helix I                     Helix II              HelixIII/IV
            -------------                -------------         --------------------
Antp    KRERQTYT RYQTLELEKEFH FNRYLT RRRRIEIAHAL CLT ERQIKIWFQNR RMKWKKEN

AN      Q-N-TSF- QE-IEA-----E RTH-PD VFA-ERL-AKI D-P -AR-QV--S-- -A--RR-E

Prd     R-S-T-FS ND-IDA--RI-A RTQ-PD VYT-E-L-QST G-- -ARVQV--S-- -ARLR-QL

Hox1.4  --S-TA-- -Q-V-------- ------ ---------T- --S ---V------- --------

Dfd     --Q-TA-- -H-I-------- Y----- ---------T- V-S ----------- ------D-

Hox1.3  --A-TA-- -------R---- ------ ----------- --S ----------- ------D-

Hox1.5  --G-TA-- -P-LV------- -----M -P--V-M-NL- N-- ----------- ---Y--DQ

Iab-7   RKK-KP-S KF---------L --A-VS KQK-W-L-RN- Q-- ---V------- ---NKK--

Bcd     R-T-T-F- SS-IA---QH-L QG---- AP-LADLSAK- A-G TA-V------- -RRH-IQS

Eve     R-Y-TAF- -D-LGR-----Y KEN-VS -P--C-L-AQ- N-P -ST--V----- ---D-RQR

(Cons.) ...R.... ..Q..L...F. ...Y.. ............ .L. ......WF.NR R.......
```


Similarities in Amino Acid Sequence Between AN and <u>Paired</u> Prototype

```
AN/Prd  K..RT.F. ..QI.A'ᵣ.... RT.YFʋ V..RE.LA... .L. EAR.QVWFSNR RA..R...
```

Figure 9. Comparison of Amino Acid Sequence of the Paired Domain and Homeodomain with Corresponding Regions in Selected Paired-Type and Homeobox-Type Proteins

(A) Comparison of the amino acid sequence of the paired domain in AN and Pax family polypeptides with that of the Drosophila *paired* (Prd) prototype. The *paired* sequence is given in full (top line) and the predicted α helical domains are marked with dashed lines. The sequences of *Pax7* (Jostes et al., 1991), *Pax8* (Plachov et al., 1990), *Pax2* (Dressler et al., 1990), and *Pax1* (Deutsch et al., 1988) are listed below for comparison. Identity with the *paired* prototype is denoted with a dash, and letters show differing residues. The consensus sequence (Cons.) between all the polypeptides is given on the bottom line; an asterisk indicates the absence of a particular residue.

(B) Comparison of the amino acid sequence of the homeodomain in AN, paired (Prd), and selected representatives of homeobox subfamilies: *Hox-1.4* and *Dfd, Hox-1.3, Hox-1.5, lab-7* (reviewed by Duboule et al., 1989), *Bcd* (Frigerio et al., 1986), and *Eve* (Macdonald et al., 1986) against the *Antp* prototype (top line). The constituent α helices of the helix-turn-helix motif are marked with dashed lines. Similarities to the *Antp* sequence are denoted with dashes, and letters indicate the identity of the differing amino acid. The consensus sequence (Cons.) is given directly below. Bottom line (AN/Prd) shows the identical amino acids between the AN and *prd* homeodomains.

(Sturm et al., 1988). The occurrence of these two specific DNA interaction domains within the AN polypeptide together with the Ser/Thr-rich region strongly suggests that the AN gene encodes a transcription factor.

Discussion

Pursuing a positional cloning strategy, we have approached the map location of the human AN gene by means of chromosome jumping and cloned from that region a candidate AN cDNA that is, first, completely deleted in a patient (SATO) diagnosed with AN and cataracts, Wilms' tumor, and mental retardation, and, second, truncated at its 3' coding region by a 300 kb interstitial deletion in a patient (HO) with familial AN. The gene is clearly deleted in all WAGR patients previously documented, including SATO, whose distal deletion breakpoint is the most centromeric so far described. The deletion in HO is the first one characterized for a familial case of complete AN without other associated features of the WAGR syndrome. The distal breakpoint in SATO is estimated to be less than 30 kb telomeric of the gene and excludes the SIMO translocation breakpoint. Together with the HO deletion, it defines an SRO of about 70 kb that encompasses the 3' coding region of the AN cDNA. There is no indication that additional transcription units exist in this highly circumscribed area. The location of the AN cDNA within this narrowly defined SRO therefore strongly suggests that it is the genetic target in AN.

These data from deletion mapping, together with clear expression of the candidate gene specifically in all tissues affected in AN and Sey, i.e., developing ocular and olfactory structures, coupled with its predicted regulatory function in development, argue strongly that the AN cDNA corresponds to the AN gene. If this is the case, then detailed analysis of this gene in AN patients might be expected to reveal point mutations and/or other aberrations in a significant proportion of cases. Indeed, based on nucleotide sequence data from the Sey candidate gene that we have isolated (C. C. T. T. et al., submitted), evidence has been obtained (R. E. Hill et al., submitted) for the presence of functionally significant point mutations in three independent Sey mutant strains. The eye phenotype in these mutants is the same as that in the two Sey deletion mutants Dickie and Harwell (Theiler et al., 1978; Hogan et al., 1986), in which the Sey candidate is hemizygously deleted (C. C. T. T., unpublished data; R. E. Hill et al., submitted), indicating that point mutations in the murine homolog of AN are sufficient to account for the ocular defects caused by more extensive genomic aberrations.

Nevertheless, there remains the need to reconcile this body of data with the observation that the AN cDNA is not apparently disrupted by the t(4;11) translocation in the SIMO cell line. This could be accounted for by the fact that the longest AN cDNA available for Southern blot analysis does not include the full 3' untranslated sequence, as suggested by comparison with the Sey candidate cDNA and the known 2.7 kb size of the message. It is also possible that the AN locus might have been silenced by positional effects through its proximity to the translocation breakpoint. Such a situation is distinctly possible in view of the observation by Kennedy et al. (1991) that in at least four out of five cases of T cell ALL, specific disease-associated translocations occurred close to the target HOX11 locus but failed to disrupt the gene. A third possibility is that there might be genetic heterogeneity for AN involving another locus in 11p13. These are some issues that would need to be addressed in future studies.

Structurally, the predicted protein product of the AN gene is characterized by the presence of the paired and homeodomains, peptide motifs that are associated with developmental regulatory proteins and known to possess specific DNA-binding properties. In addition, the AN polypeptide possesses a serine/threonine-rich C-terminal domain that resembles a similar structure in the Oct-1 transcription factor. Based on these structural properties, it would be reasonable to conclude that the product of the AN gene is probably a transcription factor with important functions in eye and nasal development.

What might be the normal biological role of the AN gene, and how would perturbations in its function contribute to the etiology of AN? In view of the autosomal dominant nature of mutant AN and Sey alleles even in the hemizygously deleted state, it is likely that inadequate gene dosage (haploinsufficiency) would be a major determinant in the disease process. This suggests that the concentration of the polypeptide product needs to be at or above a critical threshold level in order for eye development to proceed normally. Thus, experiments presently underway to genetically complement some of the murine Sey mutations might be expected to yield interesting data on this issue.

In terms of their biological roles, it is probable that the AN and Sey genes would share with the Pax family certain mechanistic and functional characteristics. Pax genes are expressed in many tissues undergoing differentiation and morphogenesis (reviewed by Kessel and Gruss, 1990). For example, Pax1 (Deutsch et al., 1988) and Pax7 (Jostes et al., 1991) expression was found to occur in developing intervertebral disks and skeletal muscles, respectively. Pax2 (Dressler et al., 1990) and Pax8 (Plachov et al., 1990) appear to play a role in ureteric induction of metanephric mesenchyme condensation in nephrogenesis. Furthermore, there is evidence to show that Pax2 and Pax7 have yet another regulatory function in neurogenesis. Thus, our detection of AN transcript expression in the human fetal eye, olfactory bulbs, and in the cerebellum appears to be consistent with the bifunctional properties of some members of the Pax family.

We would therefore speculate on this basis that the AN and Sey genes might be involved in the complex inductive interaction(s) between the optic cup, surface ectoderm, and cephalic mesenchyme during formation of the iris and other ocular structures. Indeed, our in situ hybridization data show AN transcript expression in some of the participants and products of these interactions, e.g., the neuroretina and presumptive iris, lens, and the ectodermal precursors of the conjunctiva and/or cornea. It is also remarkable that AN has been described as a panocular disorder (Nelson et al., 1984), because of the wide range of eye structures that can sometimes be affected, particularly

so since they correspond to the sites of expression of the AN gene: the foveal area of the neuroretina (loss of macular reflex), the iris (iris hypoplasia), the lens (cataracts and ectopia lentis), and the cornea (corneal pannus or opacification). While some aspects of the pathology, e.g., the defect of the fovea and optic nerve, could be secondary to an excessive influx of light, others are not so readily attributable to this factor. It might be more reasonable to assume that different developmental processes, particularly those sensitive to morphogenetic gradient fields, e.g., the *bicoid* gradient morphogen in Drosophila (Struhl et al., 1989) and fields specified by homeobox genes (reviewed by de Robertis et al., 1991), could have stringent requirements for certain gene products at specific times and locations. Thus, inadequate gene dosage in such instances might lead to a global impairment of morphogenesis, as seen in the case of AN/Sey. Such a mechanism would also account for the phenotype observed in cases of homozygous AN (Hodgson and Saunders, 1980) and in homozygous Sey, where in the absence of both functional alleles, nasal structures fail to form as a result of abortive induction of the nasal placodes by the embryonic olfactory vesicles (Hogan et al., 1986, 1988).

It is well documented that the primary function of homeodomain proteins is the control of transcription at specific promoters. Some homeobox genes autoregulate their own expression as well as that of other homeobox genes, interacting in a combinatorial manner with each other and with other transcriptional factors in hierarchical cascades to execute a given developmental program (Hayashi and Scott, 1990). From this perspective, we could assume that there might be upstream agents that regulate the AN gene, cofactors that potentiate or antagonize the function of its gene product, and perhaps arrays of downstream target genes that are the ultimate mediators of eye development. The identification and cloning of this gene would be a first step in the dissection of this network of interactions.

Experimental Procedures

Cell Culture

SIMO is an Epstein–Barr virus (EBV)–transformed cell line derived from a patient with familial AN and has a balanced translocation t(4;11)(q22;p13). This patient's daughter and grandson also had the translocation and AN but no other anomalies (Simola et al., 1983). SMXB1 and SMXB2 are human–mouse hybrids into which the der14 and der11 chromosomes from SIMO have segregated, respectively. SATO is an EBV-immortalized cell line from a female patient diagnosed with AN soon after birth, who also showed developmental delay and mental retardation (van Heyningen et al., 1985) and has a del11(p13) that spares the probe 495 derived from the proximal end of the translocation breakpoint in SIMO (Junien and McBride, 1989). The cell line SAX is a human–mouse B cell somatic hybrid derived from SATO and contains the deleted chromosome 11. FATO and MATO are lymphoblastoid cell lines from the father and mother of SATO, respectively. X63 is a mouse parental cell line used as a negative control. G401 is a cell line with normal karyotype derived from a 3-month-old Wilms' tumor patient (Weissman et al., 1987) and was obtained from the American Type Culture Association. The human–hamster somatic cell hybrid 901-110 contains 11p as its sole human genetic component (Jones and Kao, 1978). The complement of human chromosomes in the C2-1 hybrid includes only one chromosome 11, which is deleted for 11p13 (Micholopoulos et al., 1985). Human EBV-transformed lymphoblastoid cell lines were maintained in RPMI 1640 medium (GIBCO) supplemented with 10% fetal calf serum (Hazleton). The 901-110 and C2-1

hybrids were maintained in Ham's F-12 (GIBCO) supplemented with 10% fetal calf serum and Dulbecco's modified Eagle's medium with 10% fetal calf serum, respectively.

Chromosome Jumping and Walking

The chromosome jumping library used has been previously described (Collins et al., 1987) and was prepared from 100 kb size-selected restriction fragments produced by partial MboI digestion of high molecular weight human genomic DNA. Human genomic libraries used for chromosome walking included an EMBL3 phage library prepared from human leukocyte DNA, commercially available from Clontech, and a pWE15 cosmid library prepared from human leukocyte DNA according to the method of Evans and Wahl (1987). For primary screens, about 6×10^5 to 8×10^5 clones of the jumping or genomic phage libraries were plated with LE392 host on NZCYM plates at a density of 1×10^4 pfu per 150 mm plate. Approximately 3×10^5 cosmid clones were spread on LB/Kanamycin (50 μg/ml) plates at a density of 5000–6000 per 150 mm plate. Plaques or colonies were replicated onto 137 mm, 0.2 μ Biotrans (ICN) nylon membranes and processed according to Maniatis et al. (1982). The master plates were regenerated by incubation at 37°C, then stored at 4°C.

Origin of Clones and DNA Probes

The isolation and physical mapping of the random DNA sequence p5 BE0.5 (D11S323) has previously been described (Compton et al., 1988). The λ phage clones λ5-1 and λ5-2 were obtained by screening the EMBL3 genomic library with p5 BE0.5. The single-copy DNA inserts of p5-1 S1.6 and p5-2 BE1.3 were subcloned from the corresponding phages and used as starting probes to screen the jumping library, yielding the jump clones J5A6 BE0.5 and J5A1 EH0.8, respectively. The probe J5-A1 EH0.8 was used to initiate chromosome walking, generating a 42 kb contig of phage λH1, λH2, and λH3) and cosmid (cH1 and cH7) clones. The genomic probes pH2 NotI-SalI 1.6 kb, pH1-2E NotI-EcoRI 4.5 kb, pH1-6 NotI-BamHI 4.7 kb, and pH1-14 NotI-BamHI 7.0 kb are single-copy DNA sequences subcloned from this contig. The pm1 (1.6 kb EcoRI) mouse cDNA fragment (referred to elsewhere as pSM1.7) was subcloned from the phage λm14-4, which carries a 2.1 kb insert with an internal EcoRI site that can be cleaved to give the 1.6 kb and 0.5 kb fragments. The human cDNA clones ph1 and ph12 have EcoRI inserts of 1.4 kb and 1.0 kb in size, respectively, and were obtained by screening the human fetal eye library with pm1. The cDNA clone pTP1 was obtained by screening an adult retinal library in λgt10 (Nathans and Hogness, 1984) with the genomic probe pH1-14. The pTP1 insert was subsequently used to rescreen the library to give pTP2, which hybridizes to the same pattern of bands detected by ph1 and ph12 together.

Southern Blot Analysis

For conventional Southern blot analysis, genomic DNAs were digested with restriction enzymes according to the manufacturer's instructions at 5–10 U/μg DNA for at least 5 hr. Samples were separated by electrophoresis through 0.7% agarose in TAE buffer. Gels were treated with 0.25 M HCl for 10 min and the DNA denatured with 0.5 M NaOH and 1.5 NaCl. After neutralization in 1.0 M Tris–HCl (pH 8.0) and 1.5 M NaCl, the DNA was transferred overnight in 10× SSC to Nytran (Schleicher & Schuell) nylon membranes (Southern, 1975). Prehybridization was performed at 42°C for 6 hr in 50% deionized formamide, 5× SSC, 1% SDS, and 100 μg/ml denatured autoclaved salmon sperm DNA. Hybridization was subsequently performed for 16–20 hr in fresh medium after the addition of 1×10^6 to 2×10^6 cpm of randomly primed [α-^{32}P]dCTP-labeled probe (Feinberg and Vogelstein, 1983). "Zoo blots" were prepared with EcoRI-digested DNA from hamster (CHO cells), mouse (liver), rat (liver), cow (pancreas), chicken (kidneys), frog (egg), fruit fly (whole animal), and sea urchin. Filters were hybridized at 37°C in 45% formamide, 6× SSC, 5% Denhardt's solution, 10% dextran sulfate, and 200 μg/ml salmon sperm DNA. Washes were at room temperature in 2× SSC, with final stringency washes at 37°C and 42°C for 30 min each.

Southern blots of contour-clamped homogeneous electric field (CHEF) gels were done as described above with the following modifications. High molecular weight DNAs within the gel were fragmented by UV irradiation (300 nm) for 5 min prior to transfer in alkaline denaturation medium (0.4 M NaOH, 0.6 M NaCl) for 48 hr onto Genescreen

Plus (NEN) nylon membranes. The membranes were neutralized with 0.5 M Tris–HCl (pH 8.0) and 1 M NaCl, air dried, and prehybridized as above, except for the addition of 10× Denhardt's solution, 10% dextran sulfate, and an increase in probe concentration to 3 × 10⁶ to 4 × 10⁶ cpm/ml.

CHEF Gel Electrophoresis

DNA from tissue culture cells was prepared as follows. After two washes in phosphate-buffered saline, cell concentration was adjusted to 3 × 10⁷ cells/ml. The cell suspension was mixed with an equal volume of 1% molten agarose (Seaplaque, FMC) in 0.125 M EDTA (pH 8.0) before being cast into 100 μl blocks. The blocks were incubated at 50°C for 48 hr in ESP solution (0.5 M EDTA, 1.0% Sarkosyl, 1.0 mg/ml proteinase K [pH 9.0]) and washed twice at room temperature with TE containing 170 μg/ml phenylmethylsulphonyl fluoride and three times with TE, for 1 hr each wash. Blocks were stored at 4°C in 0.5 M EDTA (pH 8.0). Each restriction enzyme digestion used half a block (50 μl). Blocks were washed four times in 1 ml of TE at room temperature for 15 min each wash, followed by a 30 min equilibration in 1 ml of restriction buffer. The DNA was then digested for at least 4 hr with 40 U of enzyme. After restriction digestion, the blocks were incubated for at least 2 hr in 200 μl of ESP at 50°C before loading into the CHEF gel.

Digested DNA samples were separated by CHEF gel electrophoresis (Chu et al., 1986). Electrophoresis was carried out in 0.5× TBE (0.0445 M Tris–borate, 0.045 M boric acid, 0.001 M EDTA) at a constant temperature of 15°C. Size markers were yeast chromosomes from Saccharomyces cerevisiae strain AB972 or Schizosaccharomyces pombe strain 972H⁻, or concatemers of lambda DNA. DNAs from the HO family (HOf, normal father; HOm, affected mother; and HOs, affected son, with complete AN) were processed in plugs as described above and digested with NotI enzyme. Digested samples were separated on a 0.8% gel at 150 V, 14°C, with consecutive switching times of 60, 120, and 180 s for 11, 24, and 20 hr, respectively.

Screening of cDNA Libraries

The human fetal eye cDNA library was constructed by the directional cloning method of Gubler and Hoffman (1983) and Dorssers and Postmes (1987) using the lambda vector Charon BS (−) developed by Swaroop and Weissman (1988). Eyes from four human fetuses at 11 weeks of gestation were used to prepare the poly(A)⁺ RNA used in the library. The 8.5 day λgt10 embryonic C57BL mouse cDNA library was a generous gift from Dr. B. Hogan. The library was constructed from poly(A)⁺ RNA extracted from whole mouse embryos with extraembryonic membranes removed. For primary screens of each library, 5 × 10⁵ to 6 × 10⁵ pfu was plated on NZCYM plates at a density of 1 × 10⁴ to 2 × 10⁴ per 150 mm plate and replicated onto Biotran's (ICN) nylon filters. The filters were treated with denaturing solution (0.5 M NaOH, 1.5 M NaCl), neutralizing solution (3.0 M NaOAc [pH 6.0]), and 2× SSC for 5 min each according to the manufacturer's instructions. Filters were prehybridized and hybridized as with conventional Southern blots. The libraries were screened with the conserved genomic DNA probes (pH2 NS1.6 kb, pH1-2E NE4.5 kb, pH1-6 NB4.7 kb, and pH1-14 NB7.0 kb) or with fragments of AN or Sey cDNAs.

Northern Blots

Total RNA was prepared by the guanidine isothiocyanate (GITC) method (Davis et al., 1986). About 0.5–1.0 g of tissue was pulverized in liquid nitrogen in a mortar and pestle and homogenized in 5 ml of GITC buffer (4 M GITC, 3 M NaOAc [pH 6.0] and 1.67 ml of β-mercaptoethanol/200 ml). The mixture was cleared by centrifugation and layered onto 4 ml of CsCl buffer (5.7 M CsCl, 3 M NaOAc [pH 6.0]) in an SW41 centrifuge tube. Centrifugation at 32 K, 20°C, for 48 hr in an SW41 swinging bucket rotor yielded the RNA pellet. The pellet was rinsed in 80% ethanol and redissolved in 0.3 M NaOAc, phenol extracted, ethanol precipitated, resuspended, quantitated, and stored at −20°C. Approximately 20 μg of total RNA was run on a 0.8% agarose–2.2 M formaldehyde gel in 1× MOPS buffer (0.02 M MOPS, 0.005 M NaOAc, 0.001 M EDTA) and blotted overnight with 10× SSC onto Nytran nylon membranes. The filters were prehybridized and hybridized at 42°C for at least 16 hr in 50% formamide, 5× Denhardt's solution, 0.5% SDS, 5× SSPE, and 200 μg/ml fragmented salmon sperm DNA. Blots were hybridized with the mouse 8.5 day embryo

cDNA pm1 (1.6 kb EcoRI insert) or human fetal eye cDNA ph1 (1.3 kb EcoRI fragment). After hybridization, filters were washed twice in 2× SSC for 30 min at room temperature and one or two times in the same medium for 30 min at 55°C.

In Situ Hybridization

Human embryonic tissue of about 42–49 days of gestation was collected under conditions stipulated by the Ethical Committee, Western General Hospital, Edinburgh (Scotland). Sense and antisense RNA transcripts were made from the insert of pTP1 human cDNA clone using the T7 and T3 promoters of the Bluescript vector. [³⁵S]UTP was added as one of the nucleotides in order to label the riboprobe for use in in situ hybridization against dewaxed, paraformaldehyde-fixed, wax-embedded tissue sections. Hybridization was carried out as described in Wilkinson et al. (1987) under coverslips in the presence of 50% formamide and 20% dextran sulfate for 16–18 hr at 55°C. Following thorough posthybridization washing, the slides were coated with Ilford K5 emulsion for autoradiography; exposure times were between 2 to 4 weeks. After photographic development and fixation, sections were stained with methyl green to allow visualization of tissue architecture. Silver grains were most easily seen under dark field illumination. To produce the image shown here, the bright field and dark field images were electronically recorded, and the dark field image subjected to threshold selection, in a procedure analogous to that for determining optimal photographic exposure, and superimposed in false color (yellow) on the black and white bright field image (Monaghan et al., 1991).

DNA Sequencing

Double-stranded DNA sequencing was performed by the the chain termination method of Sanger et al. (1977). Mouse cDNAs were subcloned from phage into the Bluescript SK vector. Human cDNAs, originally constructed within the NotI-flanked KS M13 plasmid "module" of Charon BS, were released by digestion with NotI, circularized at high dilution, and transformed into DH5α competent cells. Initial sequencing primers for Bluescript and KS M13 were obtained from New England Biolabs. Subsequent oligonucleotide primers (24- to 30-mers, custom synthesized by Genosys, Woodlands, TX) were designed from generated sequence data and used to analyze downstream segments of the cDNAs. Both strands were sequenced, and sequences in ambiguous areas were confirmed by data derived from other overlapping cDNA clones. Sequencing reactions were performed with [³⁵S]dATP and Sequenase, electrophoresed on 7% polyacrylamide gels, fixed in methanol-acetic acid, vacuum dried, and autoradiographed. The GenBank data base was searched by the Lipman and Pearson (1985) method, using the Genetics Computer Group program package developed by Devereux et al. (1984).

References

Baumgartner, S., Bopp, D., Burri, M., and Noll, M. (1987). Structure of

two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. Genes Dev. *1*, 1247–1267.

Bird, A. (1986). CpG islands and the function of DNA methylation. Nature *321*, 209–213.

Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G., and Noll, M. (1989). Conservation of the paired domain in metazoans and its structure in three isolated human genes. EMBO J. *8*, 1183–1190.

Chu, G., Vollrath, D., and Davis, R. W. (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science *234*, 1582–1585.

Collins, F. S., Drumm, M. L., Cole, J. L., Lockwood, W., Vande Woude, G. F., and Iannuzzi, M. C. (1987). Construction of a general human chromosome jumping library, with application cystic fibrosis. Science *235*, 1046–1049.

Compton, D. A., Weil, M. M., Jones, C., Riccardi, V. M., Strong, L. C., and Saunders, G. F. (1988). Long range physical map of the Wilms' tumor–aniridia region on human chromosome 11. Cell *55*, 827–836.

Côté, S., Preiss, A., Haller, J., Schuh, R., Keinlin, A., Seifert, E., and Jackle, H. (1987). The gooseberry-zipper region of *Drosophila*: five genes encode different spatially restricted transcripts in the embryo. EMBO J. *6*, 2793–2801.

Davis, L. G., Dibner, M. D., and Battey, J. F. (1986). Basic Methods in Molecular Biology (New York: Elsevier Science Publishing Co.), pp. 129–146.

Davis, L. M., Byers, M. G., Fukushima, Y., Qin, S., Nowak, N. J., Scoggin, C., and Shows, T. B. (1988). Two anonymous DNA segments distinguish the Wilms tumor and aniridia loci. Science *241*, 840–842.

Davis, L. M., Everest, A. M., Simola, K. O. J., and Shows, T. B. (1989). Long-range restriction map around 11p13 aniridia locus. Som. Cell Mol. Genet. *15*, 605–615.

de Robertis, E. M., Morita, E., and Cho, K. W. Y. (1991). Gradient fields and homeobox genes. Development *112*, 669–678.

Deutsch, U., Dressler, G. R., and Gruss, P. (1988). *Pax 1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. Cell *53*, 617–625.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. *12*, 387–395.

Dorssers, L., and Postmes, A. M. E. A. (1987). A simplified, orientation-specific cDNA cloning strategy. Nucl. Acids Res. *15*, 3629.

Dressler, G. R., Deutsch, U., Chowdhury, K., Nornes, H. O., and Gruss, P. (1990). *Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system. Development *109*, 787–795.

Duboule, D., Galliot, B., Baron, A., and Featherstone, M. S. (1989). Murine homeo-genes: some aspects of their organization and structure. In Cell to Cell Signals in Mammalian Development, NATO ASI Series, Vol. H26, S. W. de Laat, J. G. Bluemink, and C. L. Mummery, eds. (Berlin: Springer-Verlag), pp. 97–107.

Evans, G. A., and Wahl, G. M. (1987). Cosmid vectors for genomic walking and restriction mapping. Meth. Enzymol. *152*, 604–610.

Feder, J., Yen, L., Wijsman, E., Wang, L., Wilkins, L., Schroder, J., Spurr, N., Cann, H., Blumenberg, M., and Cavalli-Sforza, L. L. (1985). A systematic approach for detecting high frequency restriction fragment length polymorphisms using large genomic probes. Am. J. Hum. Genet. *37*, 635–649.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Ferrell, R. E., Chakarvarti, A., Hittner, H. M., and Riccardi, V. (1980). Autosomal dominant aniridia: probable linkage to acid phosphatase-1 locus on chromosome 2. Proc. Natl. Acad. Sci. USA *77*, 1580–1583.

Francke, U., Holmes, L. B., Atkins, L., and Riccardi, V. M. (1979). Aniridia-Wilms' tumor association: evidence for specific deletions of 11p13. Cytogenet. Cell Genet. *24*, 185–192.

Fraumeni, J. F., and Glass, A. G. (1968). Wilms' tumor and congenital aniridia. J. Am. Med. Assoc. *206*, 825–828.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene *paired* and the Drosophila PRD gene set as part of a gene network. Cell *47*, 735–746.

Gehring, W. J. (1987). Homeo boxes in the study of development. Science *236*, 1245–1252.

Gessler, M., and Bruns, G. A. P. (1989). A physical map around the WAGR complex on the short arm of chromosome 11. Genomics *5*, 43–55.

Gessler, M., Simola, K. O. J., and Bruns, G. A. P. (1989). Cloning of breakpoints of a chromosome translocation identifies the AN2 locus. Science *244*, 1575–1578.

Glaser, T., Lewis, W. H., Bruns, G. A. P., Watkins, P. C., Rogler, C. E., Shows, T. B., Powers, V. E., Willard, H. F., Goguen, J. M., Simola, K. O. J., and Housman, D. E. (1986). The β-subunit of follicle-stimulating hormone is deleted in patients with aniridia and Wilms' tumor, allowing a further definition of the WAGR locus. Nature *321*, 882–887.

Glaser, T., Lane, J., and Housman, D. E. (1990). A mouse model of the aniridia-Wilms tumor deletion syndrome. Science *250*, 823–827.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. Gene *25*, 263–269.

Hayashi, S., and Scott, M. P. (1990). What determines the specificity of action of Drosophila homeodomain proteins? Cell *63*, 883–894.

Hodgson, S. V., and Saunders, K. E. (1980). A probable case of the homozygous condition of the aniridia gene. J. Med. Genet. *6*, 478–480.

Hogan, B. L. M., Horsburgh, G., Cohen, J., Hetherington, C. M., Fisher, G., and Lyon, M. F. (1986). Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. J. Embryol. Exp. Morphol. *97*, 95–110.

Hogan, B. L. M., Hirst, E. A., Horsburgh, G., and Hetherington, C. M. (1988). Small eye (Sey): a mouse model for the genetic analysis of craniofacial abnormalities. Development *103*, 115–119.

Jones, C., and Kao, F. T. (1978). Regional mapping of the gene for human lysosomal acid phosphatase (ACP2) using a hybrid clone panel containing segments of human chromosome 11. Hum. Genet. *44*, 1–10.

Jostes, B., Walther, C., and Gruss, P. (1991). The murine paired box gene, *Pax7*, is expressed specifically during development of the nervous and muscular system. Mech. Dev. *33*, 27–38.

Junien, C., and McBride, O. W. (1989). Report of the committee on the genetic constitution of chromosome 11. Cytogenet. Cell Genet. *51*, 226–256.

Kennedy, M. A., Gonzales-Sarmiento, R., Kees, U. R., Lampert, F., Dear, N., Boehm, T., and Rabbitts, T. H. (1991). HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. Proc. Natl. Acad. Sci. USA *88*, 8900–8904.

Kessel, M., and Gruss, P. (1990). Murine developmental control genes. Science *249*, 374–379.

Kilcherr, F., Baumgartner, S., Bopp, D., Frei, E., and Noll, M. (1986). Isolation of the *paired* gene of *Drosophila* and its spatial expression during early embryogenesis. Nature *321*, 493–499.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15*, 8125–8148.

Kozak, M. (1990). Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA *87*, 8301–8305.

Last, R. J. (1968). Eugene Wolf's Anatomy of the Eye and Orbit, 6th Edition (Philadelphia: W. B. Saunders Co.), pp. 419–463.

Lindsay, S., and Bird, A. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature *327*, 336–338.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. Science *227*, 1435–1441.

Macdonald, P. M., Ingham, P., and Struhl, G. (1986). Isolation, structure, and expression of *even-skipped*: a second pair-rule gene of Drosophila containing a homeo box. Cell *47*, 721–734.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Clon-

ing: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 309–326.

Mannens, M., Bleeker-Wagemakers, M., Blick, J., Hoovers, J., Mandjes, S., van Tol, S., Frantz, R. R., Heyting, C., Westerveld, A., and Slater, R. M. (1989). Autosomal dominant aniridia linked to the chromosome 11p13 markers catalase and D11S151 in a large Dutch family. Cytogenet. Cell Genet. 52, 32–36.

McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984). A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature 308, 428–433.

Micholopoulos, E. E., Bevilacqua, P. J., Stokoe, N., Powers, V. E., Willard, H. F., and Lewis, W. H. (1985). Molecular analysis of gene deletion in aniridia- Wilms' tumor association. Hum. Genet. 70, 157–162.

Miller, R. W., Fraumeni, J. F., and Manning, M. D. (1964). Association of Wilms' tumor with aniridia, hemihypertrophy and other congenital anomalies. N. Engl. J. Med. 270, 922–927.

Monaghan, P., Davidson, D. R., Sime, C., Graham, E., Baldock, R., Battacharya, S. S., and Hill, R. E. (1991). The Msh-like homeobox genes define domains in the developing vertebrate eye. Development 112, 1053–1061.

Moore, J. W., Hyman, S., Antonarakis, S. E., Mules, E. H., and Thomas, G. H. (1986). Familial isolated aniridia associated with a translocation involving chromosomes 11 and 22 [t(11;22)(p13;q12.2)]. Hum. Genet. 72, 297–302.

Nathans, J., and Hogness, D. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. Proc. Natl. Acad. Sci. USA 81, 4851–4855.

Naumann, G. O. H., and Apple, D. J. (1986). Pathology of the Eye (New York: Springer-Verlag, Inc.), pp. 1–38.

Nelson, L. B., Spaeth, G. L., Nowinski, T. S., Margo, C. E., and Jackson, L. (1984). Aniridia. A review. Surv. Ophthalmol. 28, 621–642.

Otting, G., Qian, Y.-R., Müller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988). Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence of the helix-turn-helix motif. EMBO J. 7, 4305–4309.

Plachov, D., Chowdhury, K., Waslther, C., Simon, D., Guenet, J.-L., and Gruss, P. (1990). Pax8, a murine paired box gene expressed in the developing excretory system and thyroid gland. Development 110, 643–651.

Porteous, D. J., Bickmore, W., Christie, S., Boyd, P. A., Cranston, G., Fletcher, J. M., Gosden, J. R., Rout, D., Seawright, A., Simola, K. O. J., van Heyningen, V., and Hastie, N. D. (1987). HRAS-selected chromosome transfer generates markers that colocalize aniridia- and genitourinary dysplasia-associated breakpoints and the Wilms' tumor gene within 11p13. Proc. Natl. Acad. Sci. USA 84, 5355–5359.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., and Wüthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with procaryotic repressors. Cell 59, 573–580.

Riccardi, V. M., Sujansky, E., Smith, A. C., and Francke, U. (1978). Chromosomal imbalance in the aniridia-Wilms' tumor association: 11p interstitial deletions. Pediatrics 61, 604–610.

Rose, E. A., Glaser, T., Jones, C., Smith, C. L., Lewis, W. H., Call, K. M., Minden, M., Champagne, E., Bonetta, L., Yeger, H., and Housman, D. E. (1990). Complete physical map of the WAGR region of 11p13 localizes a candidate Wilms' tumor gene. Cell 60, 495–508.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Scott, M. P., and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA 81, 4115–4119.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W., III (1989). The structure and function of the homeodomain. Biochim. Biophys. Acta Rev. Cancer 989, 25–48.

Shaw, M. W., Falls, H. F., and Neel, J. V. (1960). Congenital aniridia. Am. J. Hum. Genet. 12, 389–415.

Simola, K. O. J., Knuutila, S., Kaitila, I., Pirkola, A., and Pohja, P. (1983). Familial aniridia and translocation t(4;11)(q22;p13) without Wilms' tumor. Hum. Genet. 63, 158–161.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Struhl, G., Struhl, K., and Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell 57, 1259–1273.

Sturm, R. A., Das, G., and Herr, W. (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev. 2, 1582–1599.

Swaroop, A., and Weissman, S. M. (1988). Charon BS (+) and (−), versatile λ phage vectors for constructing directional cDNA libraries and their efficient transfer to plasmids. Nucl. Acids Res. 16, 8739.

Theiler, K., Varnum, D. S., and Stevens, L. C. (1978). Development of Dickie's small eye, a mutation in the house mouse. Anat. Embryol. 155, 81–86.

Treisman, J., Harris, E., and Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the Paired homeo domain protein. Genes Dev. 5, 594–604.

van Heyningen, V., Boyd, P. A., Seawright, A., Fletcher, J. M., Fantes, J. A., Buckton, K. E., Spowart, G., Porteous, D., Hill, R. E., Newton, M. S., and Hastie, N. D. (1985). Molecular analysis of chromosome 11 deletions in aniridia-Wilms' tumor syndrome. Proc. Natl. Acad. Sci. USA 82, 8592–8596.

van der Meer-de Jong, R., Dickinson, M. E., Woychik, R. P., Stubbs, L., Hetherington, C., and Hogan, B. L. M. (1990). Location of the gene involving the small eye mutation on mouse chromosome 2 suggests homology with human aniridia 2 (AN2). Genomics 7, 270–275.

Weissman, B. E., Saxon, P. J., Pasquale, S. R., Jones, G. R., Geiser, A. G., and Stanbridge, E. J. (1987). Introduction of a normal chromosome 11 into a Wilms' tumor cell line controls its tumorigenic expression. Science 236, 175–180.

Wilkinson, D. G., Bailes, J. A., Champion, J. E., and McMahon, A. P. (1987). Molecular analysis of mouse development from 8–10 days post coitum detects changes only in embryonic globin expression. Development 99, 493–500.

Wyman, A. R., Wolfe, L. B., and Botstein, D. (1986). Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant Escherichia coli hosts. Proc. Natl. Acad. Sci. USA 82, 2880–2884.

GenBank Accession Number

The accession number for the sequence reported in this paper is M77844.