The Mouse *Aire* Gene: Comparative Genomic Sequencing, Gene Organization, and Expression

Karin Blechschmidt,[1] Michal Schweiger,[2] Karin Wertz,[3] Richard Poulson,[4] Hoang-My Christensen,[2] Andre Rosenthal,[1] Hans Lehrach,[2] and Marie-Laure Yaspo[2,5]

[1]*Institute of Molecular Biotechnology, Department of Genome Analysis, D-07745 Jena, Germany;* [2]*Max-Planck Institute for Molecular Genetics, D-14195 Berlin-Dahlem, Germany;* [3]*Max-Planck Institute for Immunobiology, D-79108 Freiburg, Germany;* [4]*Imperial Cancer Research Fund; Histopathology Unit, London, UK*

Mutations in the human *AIRE* gene (h*AIRE*) result in the development of an autoimmune disease named APECED (autoimmune polyendocrinopathy candidiasis ectodermal dystrophy; OMIM 240300). Previously, we have cloned h*AIRE* and shown that it codes for a putative transcription-associated factor. Here we report the cloning and characterization of *Aire*, the murine ortholog of h*AIRE*. Comparative genomic sequencing revealed that the structure of the *AIRE* gene is highly conserved between human and mouse. The conceptual proteins share 73% homology and feature the same typical functional domains in both species. RT–PCR analysis detected three splice variant isoforms in various mouse tissues, and interestingly one isoform was conserved in human, suggesting potential biological relevance of this product. In situ hybridization on mouse and human histological sections showed that AIRE expression pattern was mainly restricted to a few cells in the thymus, calling for a tissue-specific function of the gene product.

There is a wide range of human autoimmune diseases, but the molecular background of autoimmunity remains poorly understood (Ollier 1992). Despite the identification of a number of genetic susceptibility factors, the etiology of most autoimmune diseases remains elusive. In this context, the study of autoimmune conditions with Mendelian inheritance could provide a boost for unraveling pathogenic pathways involved in human autoimmunity.

Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy [APECED, Online Mendelian Inheritance in Man (OMIM) 240300] is an autosomal recessive disease resulting in a variable combination of failure of the parathyroid glands, adrenal cortex, gonads, and pancreatic β cells (Ahonen 1985). Ectodermal dystrophies, vitiligo, and chronic mucocutaneous candidiasis are also frequently observed among APECED patients (Ahonen et al. 1990; Perheentupa 1996). APECED is a rare disease particularly enriched in genetic isolates, such as the Finnish population, Iranian Jews, and Sardinians (Ahonen 1985; Zlotogora and Shapiro 1992; Clemente et al. 1997). After demonstration of genetic linkage and locus homogeneity on chromosome 21q22.3 (Aaltonen et al. 1994; Bjorses et al. 1996), the APECED gene was cloned and called *AIRE* for autoimmune regulator (The Finnish–German APECED Consortium 1997; Nagamine et al. 1997).

The *AIRE* gene encodes a nuclear factor of unknown function that harbors two PHD zinc fingers—a modular domain found in many proteins involved in chromatin-mediated regulation of transcription (Schindler et al. 1993; Aasland et al. 1995). Moreover, striking structural similarities were observed between *AIRE* and human Sp100/Sp140 proteins, suggesting that they derive from a common ancestor. In addition to the zinc fingers, AIRE and Sp140 also share a putative DNA-binding domain called SAND domain and a stretch of 90 amino acids in their amino-terminal region (Gibson 1998). Sp100 proteins localize to specific nuclear structures called nuclear bodies and represent a target of autoantibodies in patients with primary biliary cirrhosis (Szostecki et al. 1990).

As a first step toward investigating AIRE biochemical properties and for engineering a mouse model for APECED, we have cloned and characterized the murine *AIRE* homolog. Here we report comparative genomic sequence analysis of the *AIRE* loci and *AIRE* expression pattern on mouse and human histological sections.

RESULTS

Identification of the Mouse *Aire* Gene

We have isolated the mouse homolog of the human *AIRE* gene by cross-species screening of mouse genomic libraries with a human cDNA containing the complete *AIRE* coding sequence [B1-1pA (The Finnish–German APECED Consortium 1997), referred to here as h*AIRE*]. Six positive clones were isolated: one PAC (RPCIP711H2150), four P1s (ICRFP703A23152,

[5]Corresponding author.
E-MAIL yaspo@impimg-berlin-dahlem.mpg.de; FAX 49-30-8413-1380.

9:158–166 ©1999 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/99 $5.00; www.genome.org

A10129, G23152, and J2183), and one cosmid (MPMGc121L12287). After restriction digest with *Eco*RI and hybridization with h*AIRE*, all clones showed a similar pattern of four *Eco*RI fragments totaling a size of 20.6 kb, except for P1 A10129, which showed an AIRE pattern of only 13.54 kb (data not shown). Hybridization with the most 5′ end and 3′ ends of the human cDNA indicated that A10129 was missing at least the first exon, whereas the five other genomic clones contained the whole *Aire* gene (data not shown).

Comparative Genomic Sequencing and *AIRE* Gene Organization

We have sequenced cosmid MPMGc121L12287 (GenBank accession no. AF073797) and deduced the mouse *Aire* gene structure by comparative analysis with the previously published h*AIRE* locus (cosmid LLNCO22G11; EMBL accession no. HSAJ9610). L12287 contained the 14 Aire exons spanning 13,276 bp from the proposed initiation codon to the termination codon. This compares with a length of 11,714 bp for the human gene (Fig. 1). The mouse Aire intron/exon boundaries were confirmed experimentally after alignment of L12287 genomic sequence and mouse cDNA sequence (see below) using the EST:GENOME program (Mott 1997). In both species, splice acceptor and donor sequences were found to conform to the GT-AG rule, and the intron phase is completely conserved (Table 1). The GC content of *AIRE* coding sequence is 61% in mouse versus 68% in human. Genomic information was analyzed by first-pass automatic annotation using the Rummage package (http:genome.imb-jena.de/rummage.html). Features conserved between the two loci include a CpG island overlapping with *AIRE* first exon and a potential promoter associated with a TATA box located 200 bp upstream of the proposed translation initiation site (Fig. 1). Two other genes were identified in the mouse cosmid, the *PFKL* promoter and a novel C_2H_2 zinc finger gene predicted in silico 6 kb proximal to *AIRE* on the opposite DNA strand (Fig. 1). This gene model is incomplete but shows significant EST matches (GenBank accession no. AA413561) and strong homology (78%) with a human trapped exon previously located 60 kb proximal from the *PFKL* promoter on 21q22.3 (HC21EXc32; D86111) (Kudoh et al. 1997). *PFKL* is 3 kb telomeric to *AIRE* in human (Fig. 1), and data indicate that the linkage group HC21EXc32, *AIRE*, *PFKL* is conserved in human and mouse. To detect potentially conserved elements, the murine and human sequences were plotted on a dot matrix using the DOTTER program (Fig. 2a; Sonnhammer and

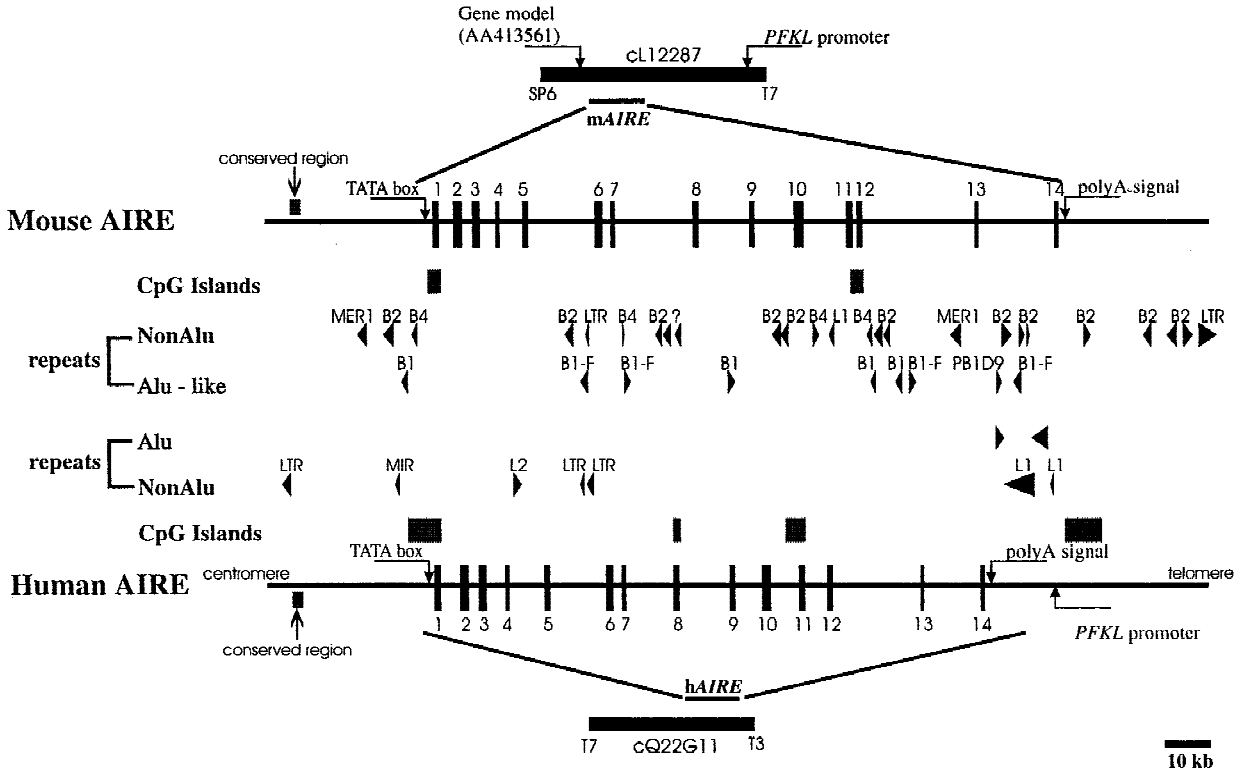

Figure 1 Comparative genomic organization of the *AIRE* locus. Exons are represented by solid boxes numbered from 1 to 14. Repetitive elements are depicted by arrowheads. CpG islands are represented by solid boxes. Putative TATA box promoter and conserved region are indicated by arrows.

Table 1. Human and Mouse Gene Structure Information

Exon No.	Exon Size (bp)	Position in cDNA	Position in genomic DNA	Intron size (bp)	Splice Acceptor	Splice Donor	Intron phase
1	132	121-252	22648-22779	418	5' UTR	CAGgtggg	0
	135	1-135	9555-9689	312	5' UTR	CAGgtggg	
2	175	253-427	23198-23372	246	tgcagGAG	AAGgtggg	1
	175	136-310	10002-10176	229	tgcagGAG	AAGgtggg	
3	156	428-583	23619-23774	383	tgcagATG	CAGgtacc	1
	156	311-466	10406-10561	381	tgcagATG	CAGgtaca	
4	75	584-658	24158-24232	753	ttcagGCT	ACGgtgag	1
	75	467-541	10943-11017	447	cgcagGCT	ACGgtgag	
5	114	659-772	24986-25099	1198	cccagGGA	CAGgtaga	1
	114	542-655	11465-11578	1420	tccagGAA	CAGgtaaa	
6	146	773-918	26298-26443	185	cccagGCG	CCCgtaag	0
	149	656-804	12999-13147	188	cccagGAA	CCTgtaag	
7	81	919-999	26629-26709	1026	tgcagGGT	CAGgtaat	0
	81	805-885	13336-13416	1674	catagGGT	CAGgtaag	
8	116	1000-1115	27736-27851	1091	gccagAAG	CAGgtgag	2
	116	886-1001	15091-15206	1088	gtcagAAG	CAGgtaag	
9	100	1116-1215	28943-29042	590	agcagTGG	CCGgtatg	0
	100	1002-1101	16295-16394	851	cacagTGG	CCGgtagt	
10	183	1216-1398	29633-29815	612	tccagCTC	CAGgtgag	0
	177	1102-1278	17246-17422	949	tccagATC	CCAgtgag	
11	122	1399-1520	30428-30549	490	cacagAAC	CGGgtgag	2
	122	1279-1400	18372-18493	96	tgcagGGT	GGGgtgag	
12	103	1521-1623	31040-31142	1879	tgcagGAC	AAGgtcag	0
	109	1401-1509	18590-18698	2491	gacagGAC	AAGgtcag	
13	63	1624-1686	33022-33084	1206	tccagGAT	GAGgtaac	0
	69	1510-1578	21190-21258	1492	tccagGTA	GAGgtaat	
14	69	1687-1755	34291-34359		cgcagCAC	3' UTR after stop	
	78	1579-1656	22751-22828		ctcagCAC	3' UTR after stop	

Numbering of exon 1 begins from the translation start site (A of ATG start codon is position 1). Numbering of exon 14 ends at the stop codon. Nucleotide residues in the cDNA and genomic sequences refer to GenBank accession nos. Z97990 and HSAJ9610 for h*AIRE* and accession nos. AJ132243 and AF073307 for mouse *Aire*. The *top* and *bottom* lines for each exon refer to the human and mouse gene, respectively.

Durbin 1995). Most of the exons were identified, albeit exons 4, 7, and 10 are barely distinguishable because of to their shorter size (exons 4 and 7) or sequence divergence. Interestingly, a highly conserved region of 90 nucleotides was identified 3 kb upstream of the AIRE first exon, displaying a stretch of 40 nucleotides with 80% identity (Fig. 2A,B). This region sharing no homology with other known regulatory elements may have a role in modulating *AIRE* expression, but this hyppothesis has to be formally demonstrated.

Localization of *Aire* to Chromosome 10

Comparative mapping between mouse and human has shown that human chromosome 21q22.3 shares conserved synteny with mouse Chromosomes 10 and 17 (Irving et al. 1994). The chromosomal localization of *Aire* was determined by PCR analysis of monochromosomal hybrids containing mouse Chromosome 10 or 17. The primer set Mforw2/Mrev32 amplified a specific product of the expected size in total mouse genome and Chromosome 10 DNAs (Fig. 3), in agreement with the expected conserved synteny in this region around the *Pfkl* locus (Irving et al. 1994).

The Predicted Mouse Aire Protein

Genomic information allowed in silico characterization of the murine cDNA sequence and corresponding conceptual protein. Nucleotide sequence identity between mouse and human *AIRE* coding sequences is 77%. h*AIRE* encodes a 545-amino-acid protein. The predicted mouse Aire protein is 552 residues with a calculated pI of 8.43 and a theoretical molecular mass of 59 kD. The overall identity between the mouse and human AIRE proteins is 73% and similarity is 76% (Fig. 4). The two proteins appear remarkably conserved and harbor the same modular domains: a SAND domain, two PHD zinc fingers, a LXXLL motif, which is a signature for nuclear receptor binding site (Heery et al. 1997), and a nuclear targeting signal (Fig. 4).

AIRE Gene Expression

Using primers designed from genomic sequence information, mouse *Aire* cDNA fragments were isolated by PCR amplification of a cDNA source prepared from ES cells. A cDNA sequence of 2015 bp deduced from overlapping PCR products contained an open reading frame (ORF) of 1656 bp (GenBank accession no. AJ132243). Northern blot analysis using a PCR product spanning exons 1–7 failed to detect any transcripts in the panel of mouse tissues analyzed containing heart, brain, spleen, lung, liver, skeletal muscle, kidney, and testis, indicating that *Aire* is seldom expressed in these tissues (data not shown). The screening of EST databases using BLAST (Altschul et al. 1990) identified only one partially processed cDNA from a 4-week mouse thymus (GenBank accession no. AA866822).

RT–PCR amplification was performed on a panel of mouse normalized first-strand cDNAs. Sequencing of cloned PCR products indicated the presence of *Aire* transcripts at 11 dpc and in adult heart, spleen, lung, skeletal muscle, and testis. Three potentially functional alternatively spliced transcripts (type I, II, and III) were seen in some tissues (Table 2). Type I isoform corresponds to the skipping of exon 10 (Fig. 5A). Type II splice variant shows a 3-bp deletion at the splice ac-

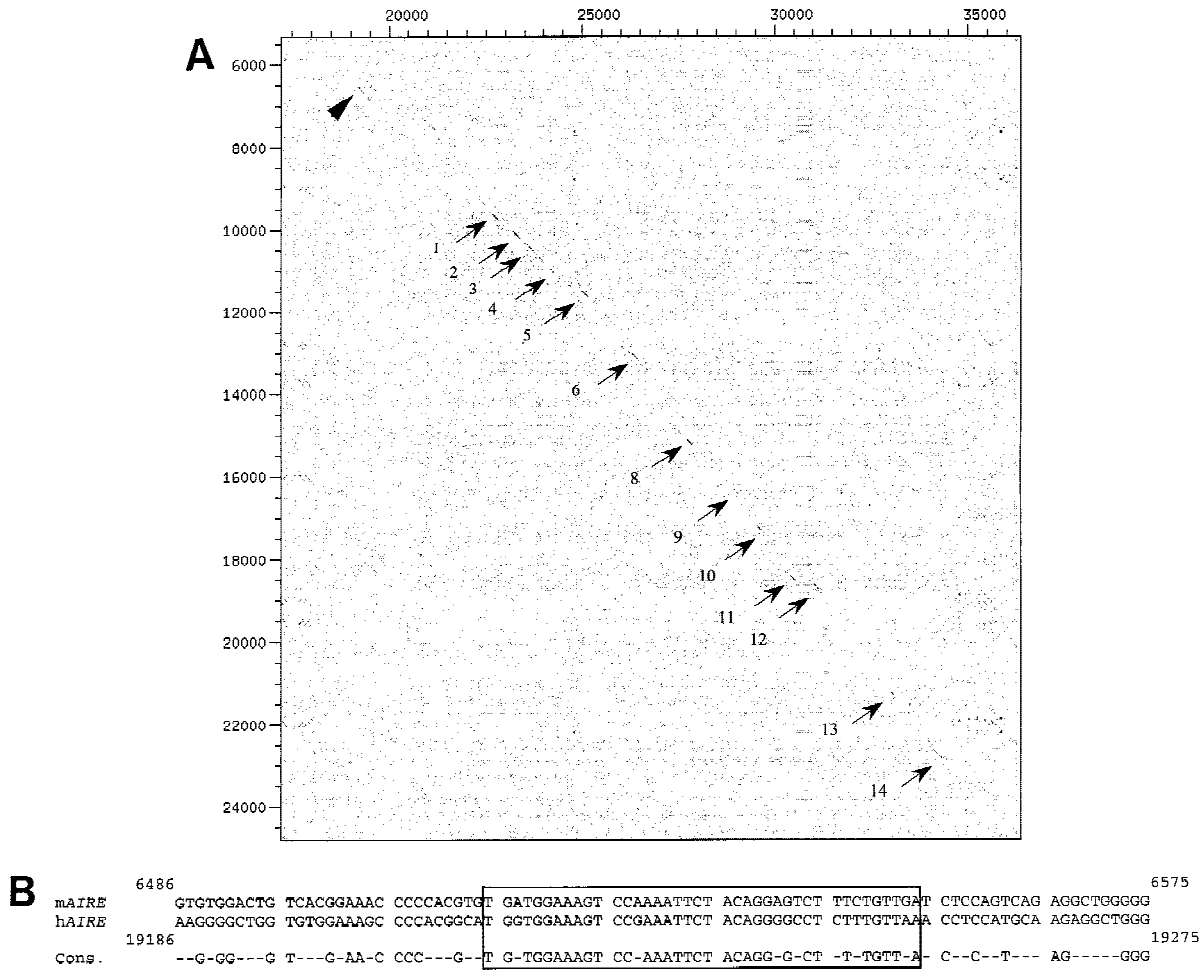

Figure 2 (*A*) Dot-matrix comparative analysis of hAIRE and mouse *Aire* genomic sequences: HSAJ9610 is represented on the *x*-axis and AF073797 on the *y*-axis. Arrows denote exons; the arrowhead points to a conserved region located ~3 kb upstream of the gene. (*B*) Alignment of the conserved nucleotide sequence identified in *A*. Numbers at the end of lines indicate nucleotide positions. Consensus sequence is drawn below the alignment. Box corresponds to the core conserved sequence.

ceptor site in exon 8 (Fig. 5B), leading to a predicted protein lacking Lys-296. Type III isoform has an in-frame 12-bp deletion at exon 6 splice donor site, and the putative peptide is lacking Val-265, Thr-266, Ile-267, and Pro-268 (Fig. 5C). No *Aire* transcripts could be obtained from 7 dpc, 17 dpc, or from adult brain or kidney. Control *Hprt* gene PCR amplification led to a single product of comparable intensity in all tissues from the panel (not shown). In human, direct sequencing of uncloned RT–PCR products generated from a panel of tissues (see Methods) identified the type II transcript in spleen and bone marrow (data not shown). However, our data did not address whether type I and type III isoforms were conserved.

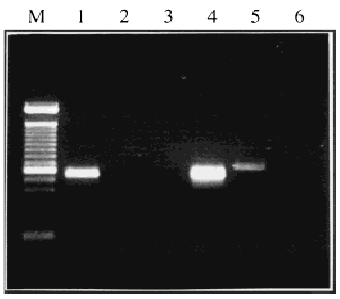

Figure 3 Mapping of *Aire* to mouse Chromosome 10. Electrophoresis of PCR amplification of mouse DNAs with primers Mforw2 and Mrev32.(Lane *1*) Hybrid SN17C3 (Chr. 10), (Lane *2*) SN11CS3 (chr. 3); (Lane *3*) EJ167 (chr. 3+17); (Lane *4*) mouse genomic DNA; (Lane *5*) human genomic DNA; (Lane *6*) no DNA. (M) 100-bp ladder (Life Technologies).

Spatial *AIRE* distribution was investigated by in situ hybridization on histological sections. In mouse embryo, *Aire* could be detected from 14.5 dpc, in which a peculiar pattern of expression was confined to a few cells in the developing thymus (Fig. 6). The cells expressing *Aire* are located in the medulla of the organ anlage but cannot be correlated with a particular cell type. This restricted staining pattern could either re-

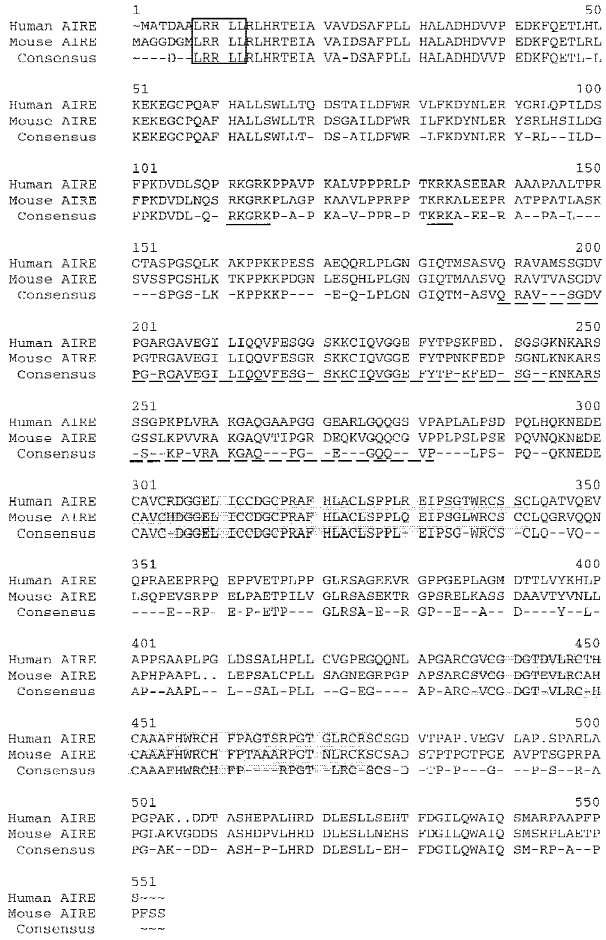

Figure 4 Amino acid aligment of hAIRE and mouse AIRE proteins. The LXXLL motif is shown by an open box; the nuclear localization signal is underlined; the SAND domain is shown by a broken line. Shaded boxes indicate the PHD zinc fingers.

flect *Aire* expression in only a very limited subset of cells or for a very short period of time in a larger cell population. In human, the spatial expression profile was found comparable with a signal restricted to foci of cells in the lobule of juvenile thymus medulla (Fig. 7).

DISCUSSION

We present here the cloning and characterization of the mouse ortholog of human *AIRE*, the gene causative for APECED disease. Comparative genomic sequencing indicated that the gene organization was highly conserved in human and mouse featuring 14 exons spanning 13 kb of genomic DNA, a TATA box promoter associated with a CpG island, and a potential controller element located 3 kb upstream of the first exon. The mouse and human *AIRE* genes are highly homologous at both the nucleotide and amino acid levels, and the two proteins contain similar structural hallmarks. By virtue of two PHD zinc fingers shared by a number of chromatin-associated transcriptional regulators, it was postulated that AIRE may have a role in gene regulation. PHD fingers are often found together with other functional domains, such as a RING zinc finger in KRIP-1 (Kim et al. 1996) or a helicase domain in the Mi2 autoantigen identified in some dermatomyositis patients (Ge et al. 1995; Seelig et al. 1995). AIRE's closest structural homolog is Sp100, which localizes to discrete nuclear dots (Grotzinger et al. 1996; Zuchner et al. 1997). hAIRE protein localizes to speckled domains in the cell nucleus (Rinderle et al. 1999), and its murine counterpart probably exhibits a similar subcellular localization. However, AIRE function is as yet elusive even if it provides the third example of a PHD finger protein involved in autoimmunity.

It is of paramount importance to determine the temporal and spatial distribution of *AIRE* for gaining insights into its primary function. Controversy revolved around the expression of human *AIRE* assessed previously by Northern blot. We reported *AIRE* expression as a 2-kb cDNA in a range of tissues with most prevalent expression in thymus, pancreas, and adrenal cortex, using a probe spanning exons 2–5 or with the

Table 2. Summary of the Sequenced RT–PCR Products

Tissue	No. of clones sequenced	No. of AIRE sequences	Canonical	Type I	Type II	Type III	Other forms
11 dpc	6	6/6	3			1	2
17 dpc	5	0	—	—	—	—	—
Heart	10	1/10				1	
Spleen	6	5/5	1			1	3
Lung	9	6/9	5				1
Testis	7	2/7			1	1	
Sk. muscle	7	1/7	1				
Brain	4	0	—	—	—	—	—
ES cells	20	19/20	5	3	2		9

Type I, II, and III isoforms are described in the left. Other forms refer to transcripts processed incompletely.

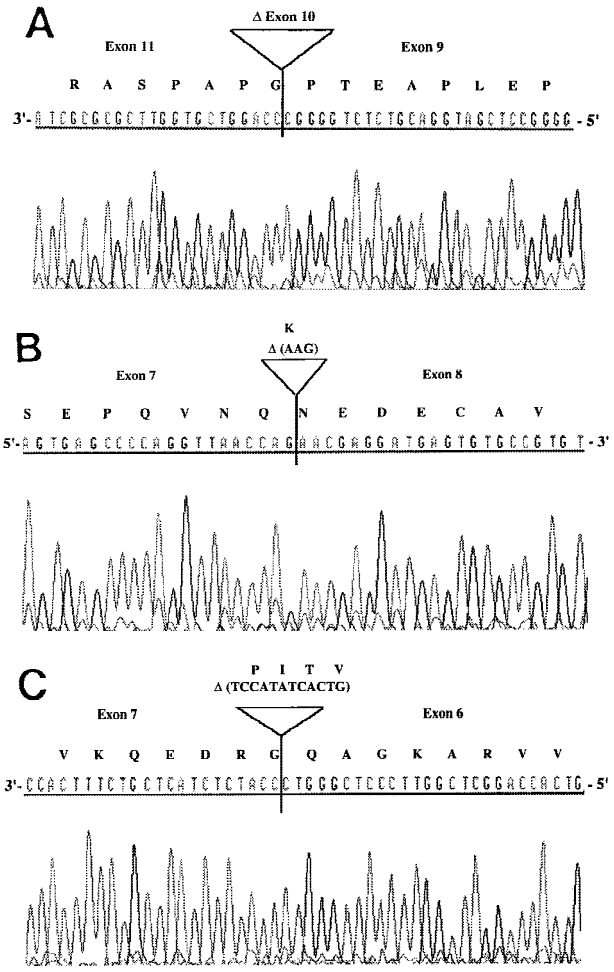

Figure 5 Differential splicing of *Aire* transcripts. (*A*) Deletion of exon 10. Sequence is reversed as indicated 3′ → 5′; (*B*) deletion of 3 nucleotides at the start of exon 8; (*C*) deletion of 12 nucleotides at the end of exon 6. Sequence is reversed as indicated 3′ → 5′.

whole cDNA (The Finnish–German APECED Consortium 1997). However, we identified a strong 2.4-kb signal in fetal liver but not in other tissues, using a probe spanning exons 11–13 (M.L. Yaspo, unpubl.). Nagamine et al. (1997) reported a cDNA of 2.4 kb in fetal liver, and several transcripts of 2, 3, and 4 kb in lymph node and thymus, using a probe spanning exons 12–14. The difference between these results can be explained in part by GC-rich regions found in the 5′ end of the probe, which may hybridize with nonlegitimate transcripts. In mouse, Northern blot analysis failed to detect *Aire* in the tissues analyzed, indicating a rare or restricted expression profile. In human, AIRE protein expression investigated by Western blot analysis failed to detect the gene product in a range of fetal and adult tissues that included organs affected in APECED, such as thyroid and parathyroid (M.L. Yaspo et al. unpubl.) Although we (M.L. Yaspo et al. unpubl.) and others (Nagamine et al. 1997) detected *AIRE* mRNA expression in human fetal liver, this could not be confirmed by Western blot. Correlation may be difficult to draw from different samples if the gene product is expressed during a short period of time and/or at a particular developmental stage. Interestingly, in situ hybridization performed at 14.5 dpc in the mouse indicated that *Aire* is expressed in only a few cells of the thymus, which are probably located in the medulla. Analysis of histological sections originating from human juvenile thymus corroborate this observation. Taken together, data confirm that AIRE is seldom expressed in most tissues. In situ hybridization data would explain the RT–PCR negative results at 17 dpc on whole mouse embryo, considering the very low proportion of cells expressing the gene. RT–PCR analysis detected three potentially functional isoforms. These variants occur with a relatively high frequency in independent PCR reactions and are unlikely to represent artifacts. Isoform type I would lead to an in-frame deletion of 59 residues between the two PHD fingers. Examples of such splice variants in zinc finger proteins have been reported previously. For instance, alternate splicing isoforms of WT1 occurring in the hinge region spacing two Krüppel zinc fingers are associated with differential subnuclear localization (Larsson et al. 1995). However, the significance of *Aire* isoforms remains to be addressed formally, sensitivity of RT–PCR may reflect residual activity of a "leaky" promoter rather than true physiological expression. In situ analysis of the expression pattern on histological sections appears to be the most informative approach for tackling the temporal and spatial *AIRE* expression pattern. Identification of those cells expressing *AIRE* in the thymus will be of fundamental relevance for shedding light onto some of the pathological mechanisms leading to autoimmunity.

Mutations in AIRE represent the primary genetic defect leading to APECED, presumably because of a defective AIRE protein. *AIRE* expression profile in embryo and adult tissues suggest that if proven to act as a mediator of transcription, AIRE is not a global transcription factor but rather is involved in modulating the expression of tissue-specific genes, for example, in the thymus. Highly conserved protein structure and similar spatial expression profile in the thymus argue for a comparable function of AIRE in human and mouse. Characterization of the *Aire* murine ortholog will thus provide a tool for exploring AIRE function and for engineering a murine model of APECED

METHODS

Isolation of Mouse *Aire* Genomic Clones

Mouse clones were screened by hybridization of mouse genomic libraries with a human cDNA probe containing the complete *AIRE* coding sequence. Six positive mouse clones were isolated: PAC RPCIP711H2150 (129/SvevTACfBr); P1

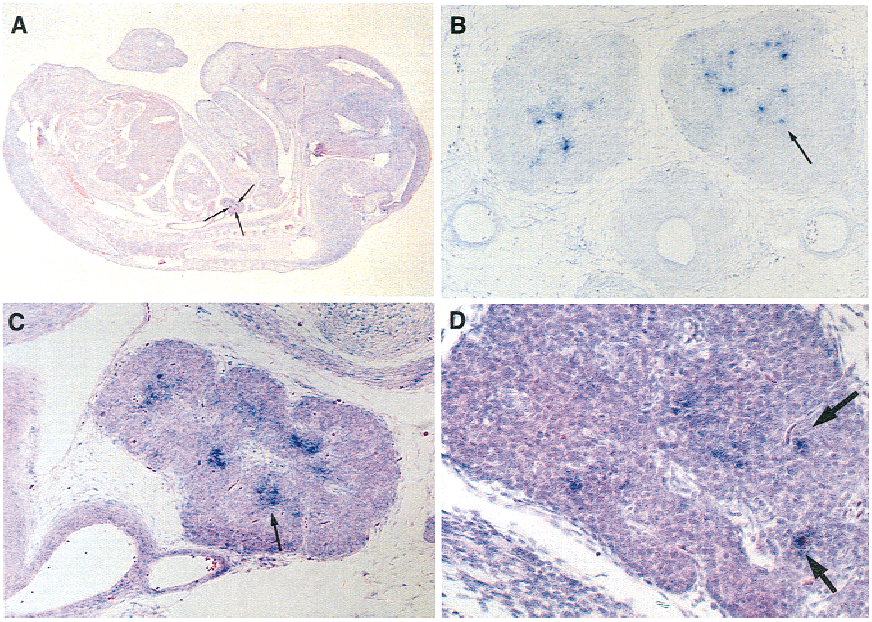

Figure 6 Expression of *Aire* at 14.5 dpc is restricted to few cells in the thymus. RNA in situ hybridization with *Aire* antisense riboprobe recognizing exons 1–7. No signal was detected upon hybridization with a sense probe. (*A*) Sagittal section through 14.5 dpc. mouse embryo, counterstained with eosin. (*B*) Transverse section of thymic lobes of a 14.5 dpc. embryo. (*C*) Sagittal section of 14.5 dpc. thymus, counterstained with eosin. (*D*) Sagittal section of 14.5 dpc. thymus, counterstained with hematoxylin and eosin, at higher magnification. Arrows point to single cells or cell groups expressing *Aire*.

clones ICRFP703A23152, A10129, G23152, and J2183 (C57/Black6); and cosmid MPMGc121L12287 (129/Ola).

Genomic Sequencing

Cosmid DNAs were isolated using a standard lysis method and purified on a CsCl gradient. DNA was sonicated, size fractionated, and ligated into M13 vector for shotgun sequencing using Thermo Sequenase (Amersham) and dye-terminator chemistry (Perkin Elmer). Data were collected using ABI 377 automated sequencers and assembled with Gap4 (Staden 1996). Gaps were closed by resequencing the M13 templates with ET (Energy Transfer) dye primers (Amersham).

Computer Analysis

Repeats were identified with the Repeat masker program http://ftp.genome.washington.edu/RM/RepeatMasker.html); (A.F.A Smit and P. Green). Homology searches were performed using BLAST version 1.4 (Altschul et al. 1990) and FASTA version 2.0 (Pearson and Lipman 1988). Programs GRAIL2 (Uberbacher and Mural 1991), XPOUND (Thomas and Skolnick 1994), MZEF (Zhang 1997), and GENSCAN (Burge and Karlin 1997) were used for exon prediction. Promoter predictions were done with Promoter Scan II (Prestridge 1995) and Transcription Start Site using both Ghosh/Prestridge (TSSG) and Wigender (TSSW) motif databases (http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html); (V.V. Solovyev, A.A. Salamov, and C.B. Lawrence). Dot matrix comparison was performed on a DEC-α station using the DOTTER program (Sonnhammer and Durbin 1995), and analysis was done using set default parameters.

RNA and RT–PCR analysis

Analysis of Mouse Aire

A Northern blot containing 2 mg of poly(A)$^+$ RNA was purchased from Clontech. RT–PCR analysis was performed on ES cells cDNA and on a normalized first-strand cDNA panel from mouse multiple tissues (Clontech). Primers Mforw4 (5′-TGGCAGGTGGGGATGGAA-3′) and Mrev15 (5′-GGAGGGATGGAAGGGGAGGA-3′) amplified a product spanning exons 1–7. PCR reactions were performed in a Biometra UNO II thermocycler. An initial denaturation at 94°C for 2 min was followed by 35 cycles at 94°C for 45 sec, 56°C for 40 sec, 72°C for 1 min, and a final extension at 72°C for 5 min. Primers Mforw6 (5′-AAAGCCAGTGGTCCGAGCCAA-3′) and Mrev34. (5′-GGAAGTGGCAGCGCCAGT-3′) amplified exons 6–11. Primers Mforw7 (5′-TGGTCCGAGCCAAGGGAG-3′) and MR4 (5′-GCCACCTGTCATCAGGAAGAG-3′) were used to amplify a cDNA fragment spanning exon 7–14 and extending in the 3′ direction outside of the translated region. Conditions for all PCRs were basically identical with the exception of the annealing temperature specific for each primer pair.

*Analysis of h*AIRE

RT–PCR analysis was performed on a normalized first-strand cDNA panel from human immune system tissue (Clontech). Primers B127FR4-21 (5′-GGCTTCTGAGGCTGCACC-3′) and B127FR4-29 (5′-GCTCTGGATGGCCTACTG C-3′) were used to amplify a 1.6-kb fragment. Nested PCRs were performed using primers B127FR4-17 (5′-AGAAGTGCATCCAGGTTGGC-3′) and B127FR4-33 (5′-GTGTGCTCGCTCAGAAGGG-3′) and products were sequenced directly.

Sequencing of PCR Products

Products from PCR amplifications were purified using the QIAquick PCR Purification Kit (Qiagen). Purified products were sequenced using the dye-terminator chemistry on an ABI 377 automated sequencer (Perkin Elmer).

Chromosomal Localization of *Aire*

PCR amplifications were performed using mouse-specific primers Mforw2 (5′-TCCCACCTGAAGACTAAGC-3′) and Mrev32 (5′-TCACAGCTC TCTGGACAGAA-3′) on hybrids SN11CS3 and SN17C3 containing mouse chromosome 3 and 10, respectively (Sabile et al. 1997) and hybrid EJ167 containing mouse chromosomes 17 and 3 (Cox et al. 1991). PCR reactions were performed in a Biometra UNO II thermocycler. Initial denaturation at 94°C for 2 min was followed by 35 cycles at 94°C for 45 sec, 51°C for 40 sec, 72°C for 2 min, and a final extension at 72°C for 5 min.

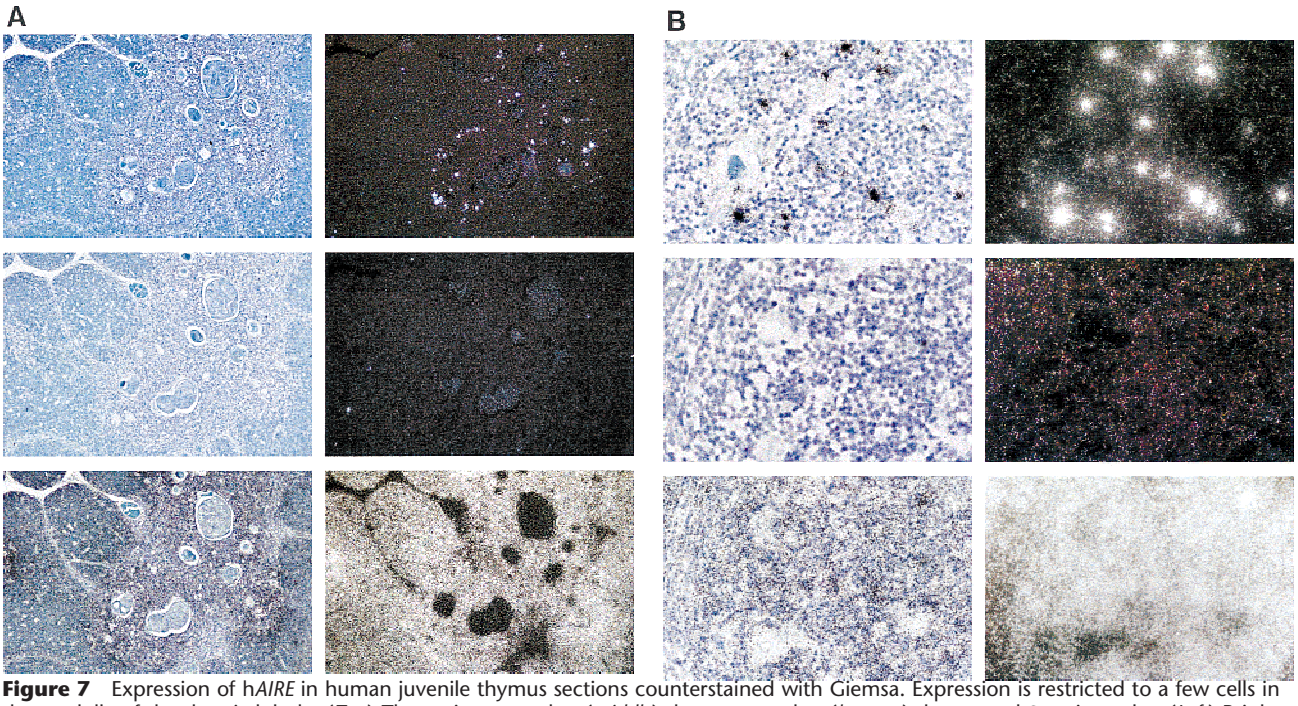

Figure 7 Expression of h*AIRE* in human juvenile thymus sections counterstained with Giemsa. Expression is restricted to a few cells in the medulla of the thymic lobule. (*Top*) The antisense probe; (*middle*) the sense probe; (*bottom*) the control β-actin probe. (*Left*) Bright field; (*right*) dark field. (*A*) Magnification, 100×; (*B*) another section with magnification, 500×.

In Situ Hybridization

In situ hybridization on mouse sections was performed essentially according to Henry et al. (1996). The cloned RT–PCR product used for riboprobe synthesis recognizes all splice variants and spans exons 1–7. Hybridization stringency was 5× SSC, 50% formamide, at 65°C. Final washing stringency was 1× SSPE, 50% formamide, at 50°C. The sections were stained for 2–3 days with BM-Purple (Boehringer). Counterstaining with eosin or with hematoxylin and eosin was performed according to standard procedures. In situ hybridization on human sections was performed with ^{35}S radiolabeled riboprobe spanning human exons 7–14 (sense and antisense), using a standard protocol described previously (Poulsom et al. 1988; Senior et al. 1988;). Sections were 4 µm thick from formaline, paraffin-embedded samples. Final washing stringency was 0.5× SSC at 65°C. Counterstaining with Giemsa was performed according to standard procedures. Exposure was for 31 days for *AIRE* and 10 days for β-actin control.

ACKNOWLEDGMENTS

We thank the Resource Center Team (RZPD) for providing mouse library filters and genomic clones. We thank Dr. Tilman Vogel (UniKlinik, Duesseldorf) for providing tissue samples. We thank Dr. Heinz Himmelbauer for the mouse monochromosomal hybrid DNAs and Dr. Michael Wiles for providing ES cell first-strand cDNA. Margit Teuchtler and Bärbel Ukena are appreciated for excellent technical assistance. This work was supported by the Deutsche Human Genome Projekt (grants BMBF 01KW9608 and 01KW9617).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aaltonen, J., P. Bjorses, L. Sandkuijl, J. Perheentupa, and L. Peltonen. 1994. An autosomal locus causing autoimmune disease: Autoimmune polyglandular disease type I assigned to chromosome 21. *Nat. Genet.* **8:** 83–87.

Aasland, R., T.J. Gibson, and A.F. Stewart. 1995. The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20:** 56–59.

Ahonen, P. 1985. Autoimmune polyendocrinopathy–candidosis–ectodermal dystrophy (APECED): Autosomal recessive inheritance. *Clin. Gen.* **27:** 535–542.

Ahonen, P., S. Myllarniemi, I. Sipila, and J. Perheentupa. 1990. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N. Engl. J. Med.* **322:** 1829–1836.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Bjorses, P., J. Aaltonen, A. Vikman, J. Perheentupa, G. Benzion, G. Chiumello, P. Heideman, J.J.G. Hoorwegnijman, L. Mathivon, P.E. Mullis et al. 1996. Genetic homogeneity of autoimmune polyglandular disease type I. *Am. J. Hum. Genet.* **59:** 879–886.

Burge, C. and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268:** 78–94.

Clemente, M.G., P. Obermayer-Straub, A. Meloni, C.P. Strassburg, V. Arangino, R.H. Tukey, S. De Virgiliis, and M.P. Manns. 1997. Cytochrome P450 1A2 is a hepatic autoantigen in autoimmune polyglandular syndrome type 1. *J. Clin. Endocrinol. Metab.* **82:** 1353–1361.

Cox, R.D., L. Stubbs, T. Evans, and H. Lehrach. 1991. A mouse specific polymerase chain reaction (PCR) primer: Probe generation from somatic cell hybrids. *Nucleic Acids Res.* **19:** 2503.

The Finnish–German APECED Consortium. 1997. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. Autoimmune

Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *Nat. Genet.* **17:** 399–403.

Ge, Q., D.S. Nilasena, C.A. O'Brien, M.B. Frank, and I.N. Targoff. 1995. Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen. *J. Clin. Invest.* **96:** 1730–1737.

Gibson, T.J., C. Ramu, C. Gemünd, and R. Aasland. 1998. The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. *Trends Biol. Sci.* **23:** 242–244.

Grotzinger, T., T. Sternsdorf, K. Jensen, and H. Will. 1996. Interferon-modulated expression of genes encoding the nuclear-dot-associated proteins Sp100 and promyelocytic leukemia protein (PML). *Eur. J. Biochem.* **238:** 554–560.

Heery, D.M., E. Kalkhoven, S. Hoare, and M.G. Parker. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387:** 733–736.

Henry, G.L., I.H. Brivanlou, D.S. Kessler, A. Hemmati-Brivanlou, and D.A. Melton. 1996. TGF-beta signals and a pattern in Xenopus laevis endodermal development. *Development* **122:** 1007–1015.

Irving, N.G., D.E. Cabin, D.A. Swanson, and R.H. Reeves. 1994. Gene order is conserved within the human chromosome 21 linkage group on mouse chromosome 10. *Genomics* **21:** 144–149.

Kim, S.S., Y.M. Chen, E. O'Leary, R. Witzgall, M. Vidal, and J.V. Bonventre 1996. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc. Natl. Acad. Sci.* **93:** 15299–15304.

Korenberg, J.R., J. Aaltonen, C. Brahe, D. Cabin, N. Creau, J.M. Delabar, J. Doering, K. Gardiner, R.S. Hubert, J. Ives et al. 1997. Report of the sixth international workshop on human chromosome 21 mapping 1996. *Cytogenet. Cell. Genet.* **79:** 22–34.

Kudoh, J., K. Nagamine, S. Asakawa, I. Abe, K. Kawasaki, H. Maeda, S. Tsujimoto, S. Minoshima, F. Ito, and N. Shimizu. 1997. Localization of 16 exons to a 450-kb region involved in the autoimmune polyglandular disease type I (APECED) on human chromosome 21q22.3. *DNA Res.* **4:** 45–52.

Larsson, S.H., J.P. Charlieu, K. Miyagawa, D. Engelkamp, M. Rassoulzadegan, A. Ross, F. Cuzin, V. van Heyningen, and N.D. Hastie. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* **81:** 391–401.

Mott, R. 1997. EST:GENOME: A program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* **13:** 477–478.

Nagamine, K., P. Peterson, H.S. Scott, J. Kudoh, S. Minoshima, M. Heino, K.J. Krohn, M.D. Lalioti, P.E. Mullis, S.E. Antonarakis et al. 1997. Positional cloning of the APECED gene. *Nat. Genet.* **17:** 393–398.

Ollier, W.A.S., and D.P.M. 1992. *Autoimmunity*. Bioscientific Publishers, Oxford, UK.

Pearson, W.R., and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85:** 2444–2448.

Perheentupa, J. 1996. Autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED). *Hormone Metab. Res.* **28:** 353–356.

Poulsom, R., J.M. Longcroft, R.E. Jefferey, L.A. Rogers, and A. Steel. 1988. A robust method for isotopic in situ hybridization to localize mRNAs in routine pathology specimens. *Eur. J. Histochem.* **42:** 121–132.

Prestridge, D.S. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. *J. Mol. Biol.* **249:** 923–932.

Rinderle, C., H-M. Christensen, S. Schweiger, H. Lehrach, and M. L. Yaspo. 1999. AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: Altered sub-cellular distribution of mutants lacking the PHD zinc fingers. *Hum. Mol. Genet.* (in press).

Sabile, A., I. Poras, D. Cherif, P. Goodfellow, and P. Avner. 1997. Isolation of monochromosomal hybrids for mouse chromosomes 3, 6, 10, 12, 14, and 18. *Mamm. Genome* **8:** 81–85.

Schindler, U., H. Beckmann, and A.R. Cashmore. 1993. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J.* **4:** 137–150.

Seelig, H.P., I. Moosbrugger, H. Ehrfeld, T. Fink, M. Renz, and E. Genth. 1995. The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis Rheum.* **38:** 1389–1399.

Senior, P.V., D.R. Critchley, F. Beck, R.A. Walker, and J.M.Varley. 1988. The localization of laminin mRNA and protein in the postimplantation embryo and placenta in the mouse: An in situ immunohistological study. *Development* **104:** 431–446.

Sonnhammer, E.L. and R. Durbin. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167:** GC1–10.

Staden, R. 1996. The Staden sequence analysis package. *Mol. Biotechnol.* **5:** 233–241.

Szostecki, C., H.H. Guldner, H.J. Netter, and H. Will. 1990. Isolation and characterization of cDNA encoding a human nuclear antigen predominantly recognized by autoantibodies from patients with primary biliary cirrhosis. *J. Immunol.* **145:** 4338–4347.

Thomas, A., and M.H. Skolnick. 1994. A probabilistic model for detecting coding regions in DNA sequences. IMA *J. Math. Appl. Med. Biol.* **11:** 149–160.

Uberbacher, E.C., and R.J. Mural. 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci.* **88:** 11261–11265.

Zhang, M.Q. 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis [published erratum appears in *Proc. Natl. Acad. Sci.* 1997 **94:** 5495]. *Proc. Natl. Acad. Sci.* **94:** 565–568.

Zlotogora, J. and M.S. Shapiro. 1992. Polyglandular autoimmune syndrome type I among Iranian Jews. *J. Med. Genet.* **29:** 824–826.

Zuchner, D., T. Sternsdorf, C. Szostecki, E.J. Heathcote, K. Cauch-Dudek, and H. Will. 1997. Prevalence, kinetics, and therapeutic modulation of autoantibodies against Sp100 and promyelocytic leukemia protein in a large cohort of patients with primary biliary cirrhosis. *Hepatology* **26:** 1123–1130.

Received October 26, 1998; accepted in revised form January 11, 1999.