

J. Aaltonen^a
J. Komulainen^b
A. Vikman^a
A. Palotie^c
C. Wadelius^d
J. Perheentupa^b
L. Peltonen^a

- ^a Department of Human Molecular Genetics, National Public Health Institute, Helsinki;
^b Children's Hospital;
^c Department of Clinical Chemistry, Helsinki University Central Hospital, Helsinki, Finland, and
^d Department of Clinical Genetics, University Hospital, Uppsala, Sweden

.....
Key Words

Autoimmune disease
PGD type I
APECED
Microsatellites
Linkage analysis
Gene mapping

Introduction

The autoimmune reactions associated with many human diseases are still only partially understood. Unravelling the molecular pathogenesis of inherited diseases with a strong autoimmune component in their clinical ex-

Autoimmune Polyglandular Disease Type I

Exclusion Map Using Amplifiable Multiallelic Markers in a Microtiter Well Format

.....
Abstract

The pathogenetic background of human autoimmunity is only partially understood. By discovering the defective gene causing the autosomal recessive polyglandular autoimmune disease type I (PGDI, APECED) we hope to provide new insights into autoimmune responses in general. Here we have taken advantage of newly developed amplifiable multiallelic microsatellite markers and performed the analyses using the microtiter well format of the polymerase chain reaction. This rapid semiautomated protocol was applied to analyze 62 assigned highly polymorphic loci. The linkage analyses coupled with the EXCLUDE analysis resulted in an exclusion map of this polyglandular autoimmune disease and in the preliminary assignment of the APECED locus to chromosome 22. The method proved to be an effective and economical tool for gene mapping compared with standard blotting and hybridization.

.....
pression could help to dissect the molecular background of abnormal immune responses. One such genetic disorder is autosomal recessive polyglandular autoimmune disease type I, also known as autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED, MIM number 240300) [1, 2]. The

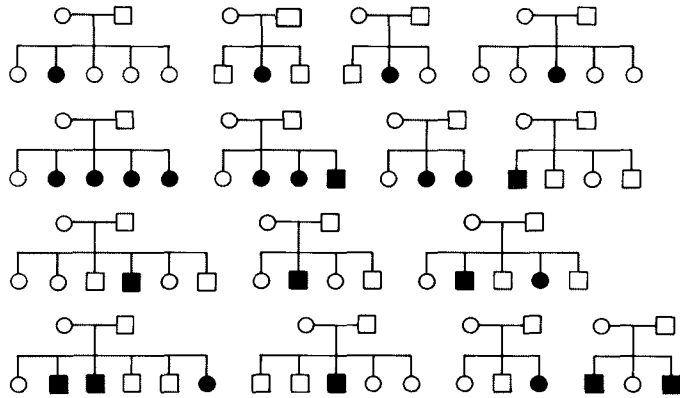


Fig. 1. The collection of APECED pedigrees included in this study.

disease is especially enriched in the genetically isolated population of Finland [3]: to date 70 patients have been identified in Finland and less than 100 elsewhere, the incidence of APECED in Finland being at least 1:25,000. The clinical phenotype of APECED includes variable combinations of the failure of parathyroid glands, the adrenal cortex, gonads, pancreatic beta cells, gastric parietal cells and the thyroid gland. Additional features include hepatitis, chronic mucocutaneous candidiasis, dystrophy of dental enamel and nails, alopecia, vitiligo and keratopathy. The diagnosis is typically made in childhood, but symptoms may appear as late as the fifth decade of life [4].

The pathogenesis of APECED is unknown and there are no reports of abnormal biochemical findings or chromosomal rearrangements in the affected individuals. Consequently, a random search of the human genome directly at the DNA level is a relevant approach for revealing the defective gene locus. Here we have taken advantage of recently identified amplifiable polymorphisms assigned to human chromosomes [5]. The microtiter well format of the PCR [6] was

adapted for screening these markers, and high informativeness of the markers guarantees relatively wide regions of exclusion in linkage analysis. The map constructed from multiallelic markers was complemented with biallelic markers also detected in the microtiter well format using the solid-phase minisequencing technique [7]. The screening for the APECED locus using 62 polymorphisms suggests that the most probable location for this autoimmune-disease gene is chromosome 22.

Materials and Methods

A total of 15 APECED families were included in this study (fig. 1). In total, 94 blood samples were collected, 25 from affected siblings. Genomic DNA was extracted by standard procedures from 10-ml blood samples [8].

In this study 62 polymorphic markers were analyzed, most of them highly polymorphic microsatellite markers. Alleles were detected by PCR amplification [6] followed by polyacrylamide gel electrophoresis. PCR was performed in a volume of 15 μ l containing 20 ng of template DNA, 2 pmol of primers, 0.2 mM each of dATP, dCTP, dGTP and dTTP, 20 mM Tris-HCl (pH 8.8), 15 mM $(\text{NH}_2)\text{SO}_4$, 1.5 mM MgCl_2 , 0.1% Tween 20, 0.01% gelatin and 0.8 U of *Thermus aquati-*

cus DNA polymerase (Promega). One of the primers was labeled at the 5' end using ^{32}P . The PCR reactions were performed in multiwell microtitration plates for 28 cycles, the first cycle consisting of 95 °C for 3 min, 55 °C for 30 s and 72 °C for 30 s followed by 26 cycles consisting of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s and the final cycle of 95 °C for 30 s and 55 °C for 30 s followed by a 4-min extension at 72 °C. The amplified fragments were separated by electrophoresis in 6% denaturing polyacrylamide sequencing gels. Twelve of the analyzed polymorphisms were single-nucleotide variations. These biallelic markers were analyzed by PCR and the solid-phase minisequencing technique was performed as described in detail elsewhere [7, 9].

To test if the family material available was sufficient for conclusive results in the linkage analyses we first carried out data simulation analysis assuming a single marker locus and double heterozygosity for the parents. We used the SLINK program, specifically its MSIM option, with 1,000 replicates (one replicate equals one round of generating marker genotype for each individual) to obtain the ELOD (expected lod score) values [10, 11].

The MLINK option of the LINKAGE package program (version 5.10) was used to calculate the linkage between the disease and the individual markers [12]. Because the age at onset of the disease varies, the following age-dependent categories of penetrance were adapted for the healthy individuals for the linkage analyses: for age categories of <2, 2–5, 6–10, 11–15, 16–20, 21–30 and >30 years a penetrance of 5, 20, 45, 60, 70, 80 and 95%, respectively, was assigned. The original MLINK data from the pairwise linkage analysis was subsequently analyzed with the EXCLUDE program to construct an exclusion map of the disease [13]. The data input into the program consisted of the chromosomal position of the locus, locus name, the linkage data in the form of recombination fraction (we used five θ values: 0.00, 0.10, 0.20, 0.30 and 0.40) and the corresponding lod score calculated by the MLINK program. In all calculations the Haldane map function was used to convert recombination fractions to map distances. The locations of the markers analyzed were based on HGM11 and the map (Marshfield markers 8) released by Dr. J.L. Weber (Marshfield Medical Research Foundation, Marshfield, Wisc., USA) and expressed as a percentage from pter to qter. When the precise chromosomal location was unknown or expanded over a region in a chromosome, the marker was placed in the middle of that region.

Results and Discussion

The family material analyzed is summarized in figure 1. Although the clinical phenotype of APECED is highly variable, the disease is enriched in the genetically isolated Finnish population. Consequently the assumption of locus homogeneity is valid within this population. To test if the available family material was sufficient to obtain conclusive data in linkage analysis of this disease with heterogeneity in phenotype and age-dependent penetrance, we carried out simulations in our family material with 1,000 replicates using the SLINK program [10, 11]. Under the one-locus model for this family material the mean ELOD is 7.110 with a standard deviation of 1.424. Thus, the expected lod score at a recombination fraction of 0.01 ranges from 4.262 to 9.958 (95% confidence interval). The lod score still remained significant in over 50% of the replicates at the recombination fraction of 0.10. This simulation analysis suggests that an evenly spread set of informative markers should result in a conclusive lod score in our family material.

We analyzed the linkage data from a total of 62 polymorphic markers on all 22 autosomes. Fifty markers detected multiallelic polymorphisms. The informativeness of these multiallelic markers is very high and we applied the PCR in the microtiter well format and multi-channel pipetting which provides a rapid semiautomated method for detecting these multiallelic DNA polymorphisms. Two subsequent loadings of the samples on large sequencing-type polyacrylamide gels made it possible to obtain information from two markers of 60 individuals per gel. In the multiwell format, 96 samples can conveniently be simultaneously analyzed and, although loading of the sequencing gels was carried out manually, the system significantly improved

the efficiency in detecting allelic polymorphisms.

The problems encountered with multiallelic polymorphisms were to a large extent associated with the identification of individual alleles in the case of each polymorphism. Frequently this required simultaneous analyses of representative samples of individual families on the same sequencing gel. For linkage analyses per se this would not have been necessary since within a family the allelic pattern could be identified without any problem but for determining allelic frequencies the precise typing of alleles is a necessity.

Twelve biallelic markers were analyzed to bridge the gaps still existing in the chromosomal maps of highly polymorphic amplifiable markers. The alleles of these markers were also identified after PCR amplification, and the specific solid-phase minisequencing technique was applied to detect the alleles [7]. Using this method the results are obtained in numeric values which unequivocally identify heterozygotes and homozygotes for a given polymorphism [9].

In our family material originating from the genetically isolated Finnish population the informativeness of the markers analyzed was close to the published heterozygosity values obtained in other populations (table 1). Pairwise linkage data to individual markers resulted in some low positive lod scores but no definitive proof of linkage to any of the markers analyzed was found. The total length of the chromosomal regions revealing a relative exclusion (lod score < -1.0) was 1,597.7 cM (table 1) compared to total autosomal length of 2,885 cM (table 2).

In addition to performing the traditional pairwise linkage analyses using the MLINK program we combined the MLINK data and analyzed them with the EXCLUDE program [13]. The result gave evidence of some chromosomal locations with a clearly higher possi-

bility of representing the site of the APECED locus than analyzed loci on the other chromosomes (table 2, figure 2). The result obtained suggested a fairly high probability for the disease being located on chromosome 22, corresponding to a lod score of 3.1 (table 2). This value cannot be directly interpreted as proof of linkage, however. The corresponding maximal lod score achieved in multipoint (LINK-MAP) analysis with the disease against the three markers of chromosome 22 is 1.4. This apparent numeric discrepancy is due to the nature of the EXCLUDE program in which the obtained lod scores are not comparable with the values achieved in the classical linkage analyses. The program takes advantage of the fact that a gene excluded from one region must have a higher probability of being in another region of the genome and this assumption increases the numeric values of the likelihood compared to the values achieved with the more classical linkage approaches adopted in the LINKAGE package programs [12].

Although it is problematic to specify candidate genes for the multicomponent APECED disease with multiple phenotypic features and clinical symptoms from several tissues, the genes encoding proteins that control normal immune responses could serve as tentative candidate loci. Such proteins include, besides histocompatibility antigens and T cell receptors for antigen, immunoglobulins, as well as genes affecting antigen processing and presentation, lymphocyte proliferation and differentiation and the effector molecules that mediate immune destruction in normal immune responses.

Association with the HLA region on chromosome 6 was excluded in an earlier study based on serologically determined HLA haplotypes in APECED families [14]. So far we have analyzed genes coding for the T cell antigen receptors (TCR) on chromosomes 7

Table 1. Polymorphic markers analyzed

Marker	Chromosome	Heterozygosity		Region revealing lod score < -1.0, cM
		published	observed	
D1S80	1p36-p35	0.80	0.79	32.8
D1S7	1p35-p33	0.99	0.92	27.4
HSD3B ^a	1p13.1	0.30	0.29	2.0
APOA2	1q21-q23	0.74	0.79	15.0
D1S103	1q32-q44	0.88	0.68	4.0
APOB	2p24-p23	0.79	1.00	66.0
CD8A	2p12	0.71	0.68	24.7
CTLA4	2q33	0.92	0.82	47.8
GCG	2q36-q37	0.79	0.81	32.8
PROS1 ^a	3p11.1-q11.2	0.48	0.29	8.0
GLUT2	3q26.2-q26.3	0.84	0.89	51.0
D3S196	3q27-q28	0.67	0.71	44.6
D4S174	4p15-p11	0.92	0.96	58.0
ADH3 ^a	4q21-q23	0.50	0.50	12.8
FABP2	4q28-q31	0.64	0.38	51.0
D4S175	4q21-qter	0.82	0.84	10.0
D4S171	4q33-q35	0.75	0.93	15.0
F11	4q35	0.48	0.42	<1.0
D5S117	5p15.3-p15.1	0.55	0.64	12.8
ARSB ^a	5q11-q13	0.47	0.43	<1.0
D5S107	5q11.2-q13.3	0.82	0.82	4.0
FBN2	5q23-q31	n.d.	1.00	27.4
D5S207	5q31.3-q33.3	0.69	0.65	38.6
D5S211	5q33.3-qter	0.72	0.79	41.6
D6S105	6p	0.87	0.86	51.0
DMDL	6q24	n.d.	0.68	69.4
D7S435	7p15.1-p13	0.59	0.68	27.4
MET ^a	7q31	0.50	0.46	35.6
LPL ^a	8p22	0.50	0.43	22.4
D8S87	8p12	0.71	0.79	22.4
D9S54	9pter-p22	0.56	0.76	32.8
D9S43	9p21-qter	0.83	0.93	32.8
D10S89	10pter-p11.2	0.80	0.82	32.8
D10S169	10q11.2-qter	0.73	0.68	27.4
HBG1	11p15.5	0.49	0.50	17.4
IGF2 ^a	11p15.5	0.44	0.50	6.2
F2 ^a	11p11-q12	0.32	0.32	4.0
D11S527	11q13.5	0.88	0.86	32.8
CD3D	11q23	0.74	0.81	51.0
COL2A1	12q13	0.58	0.61	4.0
IGF1	12q23	0.54	0.61	17.4
D13S71	13	0.72	0.75	51.0
FLT1	13q12	0.51	0.19	12.8
TCRD	14q11.2	0.77	0.79	51.0
IGHJ	14q32.32-q32.33	0.73	0.61	22.4

Table 1 (continued)

Marker	Chromosome	Heterozygosity		Region revealing lod score <-1.0, cM
		published	observed	
FBN1	15q21.1	n.d.	0.39	4.0
FES	15q25-qter	0.75	0.82	12.8
D16S265	16q	0.77	0.89	51.0
D1755	17p13.3	0.85	0.86	35.6
GH	17q22-q24	0.83	0.96	22.4
BCL2 ^a	18q21	0.48	0.49	32.8
MBP	18q22-qter	0.76	0.69	17.4
LDLR ^a	19p13.3	0.49	0.50	8.0
APOC2	19q12-q13.2	0.81	0.93	38.6
PRNP ^a	20pter-p12	0.38	0.39	27.4
D20S27	20p12	0.70	0.75	27.4
D20S66	20	0.77	0.93	38.6
D21S13E ^a	21q11	0.46	0.38	8.0
PFKL	21q22.3	0.70	0.32	<1.0
D22S264	22q11.2	0.70	0.68	<1.0
D22S156	22q11.2	0.69	0.86	<1.0
D22S258	22	0.81	0.82	22.4
				1,597.7

^a Bi-allelic marker [23].

and 14 [15–18], as well as the immunoglobulin heavy chain (IgH) gene cluster segments on chromosome 14 [18], the gene encoding the CD3 δ chain on chromosome 11 [19], and the CD8 α -chain gene on chromosome 2 [20]. The markers D7S435 on chromosome 7, TCRD and IGHJ on chromosome 14, CD3D on chromosome 11 and CD8A on chromosome 2 are located in the immediate vicinity of these regions. The result of the linkage analyses was a definitive exclusion of all these areas. Interestingly, chromosome 22, which is the most likely chromosome to carry the APECED locus, contains a gene cluster coding for structural proteins of the immunoglobulin λ chain [21], as well as genes for pre-B lymphocyte

gene 1 and the interleukin-2 β receptor [22]. This is further justification for future analysis focusing on the chromosomal regions containing these genes.

Efficient chromosomal screening with the microsatellite markers represents a major improvement in the mapping of inherited human diseases [5]. Here we analysed both multiallelic and biallelic markers in the microtiter well format suitable for automatization. Even the semiautomated protocol applied here proved to be effective and economical. The highly polymorphic, multiallelic markers are efficient tools in the mapping of any inherited disease, even with limited family resources. Coupling of the microtiter well format ampli-

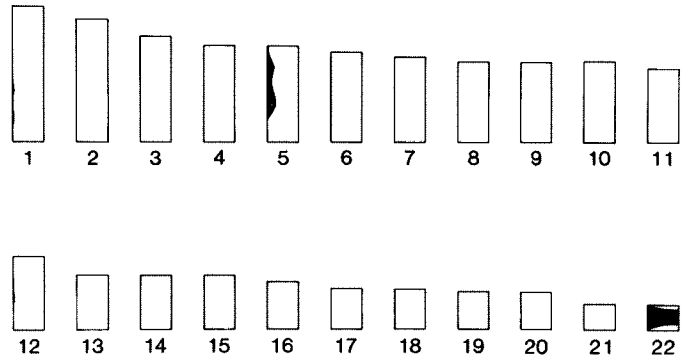


Fig. 2. Graphical presentation of the exclusion map for APECED disease. The most probable areas for the gene location are colored black while the remaining white areas represent the excluded areas.

Table 2. Exclude analysis

Chromosome	Loci	Length cM	Percentage of autosome length	Chromosome probability	Maximum likelihood
1	5	250	8.67	3.36	14.69
2	4	225	7.80	0.08	1.05
3	3	200	6.93	0.23	1.02
4	6	180	6.24	0.23	6.12
5	6	180	6.24	17.84	220.22
6	2	170	5.89	0.02	0.97
7	2	160	5.55	0.14	1.39
8	2	150	5.20	0.33	1.33
9	2	150	5.20	0.13	0.99
10	2	150	5.20	0.12	1.30
11	5	140	4.85	0.09	1.92
12	2	140	4.85	0.74	3.11
13	2	100	3.47	0.02	1.05
14	2	100	3.47	0.00	0.66
15	2	100	3.47	0.61	3.71
16	1	90	3.12	0.03	0.55
17	2	80	2.77	0.04	0.85
18	2	80	2.77	0.05	0.80
19	2	70	2.43	0.01	1.78
20	3	70	2.43	0.03	1.04
21	2	50	1.73	0.21	2.76
22	3	50	1.73	75.69	1,393.22

Maximum chromosomal probability = 75.69%. Maximum likelihood 1,393.22 = lod score 3.14.

fication of these markers to the nonradioactive labeling of the product and utilization of an automated sequencer in the identification and storage of allelic information will provide in the future the technical basis for mapping inherited human diseases.

Acknowledgements

The authors gratefully acknowledge the help of Dr. Jurg Ott and Dr. Lodewijk Sandkuijl in linkage analysis and Dr. Ann-Christine Syvänen with the minisequencing technique. The grants from the Foundation for Paediatric Research and the Academy of Finland are appreciated. The marker resource of Nordic Hugo is appreciated.

References

- McKusick VA: Mendelian Inheritance in Man, ed 9. Baltimore, Johns Hopkins University Press, 1990.
- Ahonen P: Autoimmune polyendocrinopathy-candidosis-ectodermal-dystrophy (APECED): Autosomal recessive inheritance. *Clin Genet* 1985;27:535-542.
- Norio R: Diseases of Finland and Scandinavia; in Rothschild HR (ed): *Biocultural Aspects of Diseases*. New York, Academic Press Inc., 1981, pp 359-415.
- Ahonen P, Myllärniemi S, Sipilä I, Perheentupa J: Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med* 1990;322:1829-1836.
- Weber JL, May PE: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989;44:388-396.
- Mullis KB, Faloona FA: Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 1987;155:335-350.
- Syvänen AC, Aalto-Setälä K, Harju L, Kontula K, Söderlund H: A primer-guided nucleotide assay in the genotyping of apolipoprotein E. *Genomics* 1990;8:684-692.
- Vandenplas S, Wiid I, Grobler-Rabie A, Brebner K, Ricketts M, Wallis G, Bester A, Boyd C, Mathew C: Blot hybridization analysis of genomic DNA. *J Med Genet* 1984;21:164-172.
- Syvänen AC, Ikonen E, Manninen T, Bengtström M, Söderlund H, Aula P, Peltonen L: Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: Application to aspartylglucosaminuria in Finland. *Genomics* 1992;12:1-6.
- Ott J: Computer-stimulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 1989;86:4175-4178.
- Weeks DE, Ott J, Lathrop GM: SLINK: A general simulation program for linkage analysis. *Am J Hum Genet* 1990;47(suppl):A204.
- Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;81:3443-3446.
- Edwards JH: Exclusion mapping. *J Med Genet* 1987;24:539-543.
- Ahonen P, Koskimies S, Lokki ML, Tiilikainen A, Perheentupa J: The expression of autoimmune polyglandular disease type I appears associated with several HLA-A antigens but not with HLA-DR. *J Clin Endocrinol Metab* 1988;66:1152-1157.
- Collins MKL, Goodfellow PN, Spurr NK, Solomon E, Tanigawa G, Tonegawa S, Owen MJ: The human T-cell receptor alpha chain maps to chromosome 14. *Nature* 1985;314:273-274.
- Bernard F, Chuchana P, Lefranc G, Lefranc MP: An STS in the human T cell receptor gamma locus (located at 7p14-15). *Nucleic Acids Res* 1991;18:6170.
- Jordan SA, McWilliam P, O'Brian DS, Humphries P: Dinucleotide repeat polymorphism at the T cell receptor delta locus (TCRD). *Nucleic Acids Res* 1990;19:1959.
- Rabbits TH, Lefranc MP, Stinson MA, Sims JE, Schroder J, Steinmetz M, Spurr NL, Solomon E, Goodfellow PN: The chromosomal location of T-cell receptor genes and a T cell rearranging gene: Possible correlation with specific translations in human T cell leukaemia. *EMBO J* 1985;4:1461-1465.
- Van Den Elsen P, Bruns G, Gerhard DS, Pravtcheva D, Jones C, Housman D, Ruddle FA, Orkin S, Terhorst C: Assignment of the gene coding for the T3-delta subunit of the T3-T-cell receptor complex to the long arm of human chromosome 11 and to mouse chromosome 9. *Proc Natl Acad Sci USA* 1985;82:2920-2924.
- Polymeropoulos MH, Xiao H, Rath DS, Merrill CR: Dinucleotide repeat polymorphism at the human MHC class I CD8 alpha-chain (Leu-2/T8) gene. *Nucleic Acids Res* 1991;19:1718.
- Chuchana P, Alexandre D, Lefranc G, Lefranc MP: The IGLVA gene as a STS in the human immunoglobulin lambda light chain variable gene locus (located at 22q11). *Nucleic Acids Res* 1991;19:4786.
- Human Gene Mapping 11. *Cytogenet Cell Genet* 1991;58:1-2200.
- Syvänen AC, Sajantila A, Lukka M: Identification of individuals by analysis of biallelic DNA markers using PCR and solid-phase minisequencing. *Am J Hum Gen*, in press.