Shin-ichi Kuno · Atsuo Taniguchi · Akira Saito
Sanae Tsuchida-Otsuka · Naoyuki Kamatani

Comparison between various strategies for the disease-gene mapping using linkage disequilibrium analyses: studies on adenine phosphoribosyltransferase deficiency used as an example

Received: 24 March 2004 / Accepted: 19 May 2004 / Published online: 28 July 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract Recently, linkage disequilibrium analyses have been used to detect disease-causing loci based on the common disease-common variant hypothesis. To see what methods can effectively identify the genes, we have to apply them to the practical data obtained from the human population. We extensively performed linkage disequilibrium and haplotype analyses on adenine phosphoribosyltransferase (*APRT*) genes in both control and deficient subjects. To examine the power to detect disease-causing loci, we analyzed SNPs, STRPs, and VNTR within and around the *APRT* gene. When only SNPs were used, *P* values did not necessarily show significant difference, even at loci close to the mutation site for *APRT***J* that is exclusively observed among Japanese. However, the examination of the same samples with haplotypes based on the haplotype block data gave sufficient significance. In the case of STRP and VNTR, some single-marker loci showed significant difference. Our study suggested that the use of haplotype analysis based on the haplotype-block structure is more

powerful than single-marker locus analysis for the detection of disease-related loci.

Keywords Adenine phosphoribosyltransferase deficiency · Linkage disequilibrium · Case-control analysis · SNP · Microsatellite · VNTR · Haplotype

Introduction

Common disease-common variant hypothesis predicts that causative mutations responsible for a common disease are likely to be attributed to common variants. The common variants presumably derived from common ancestor mutations. In such a case, the causative mutations can be identified by searching for the polymorphisms that are not the real causes of the disease but are in linkage disequilibrium (LD) with the causative mutations.

Recent genomic studies have been focused on the selection of single nucleotide polymorphisms (SNPs), short tandem repeat polymorphisms (STRPs; microsatellites), and variable numbers of tandem repeat (VNTR) that are in LD with common diseases such as diabetes mellitus (Mohlke et al. 2001) and hypertension (Angius et al. 2002). Such a selection has been attempted after narrowing the candidate regions by the linkage analyses or by directly searching the entire chromosomes. However, very few such attempts have so far been successful.

In the present study, we have tested the hypothesis as to whether attempts to identify the disease mutations can be successful by looking for polymorphisms that are in LD with the disease mutation. If it can be successful, we intended to ask what strategies would be optimal to solve this problem. Some researchers claimed that SNPs are superior to STRPs for that purpose while others argued against that. We also intended to test whether the use of haplotypes in addition to SNPs and STRPs is useful for finding the disease locus.

S. Kuno · A. Taniguchi · A. Saito · S. Tsuchida-Otsuka
N. Kamatani
Institute of Rheumatology,
Tokyo Women's Medical University, Tokyo, Japan

S. Kuno (✉)
Laboratory for Clinical Genome Informatics,
Translational Research Informatics Center,
Foundation for Biomedical Research and Innovation,
1-5-4 Minatojima-minamimachi, Chuo-ku,
Kobe 650-0047, Japan
E-mail: omoikane@tri-kobe.org
Tel.: +81-78-3032760
Fax: +81-78-3032761

A. Saito
Advanced Research Laboratory, Hitachi Ltd.,
Tokyo, Japan

N. Kamatani
Division of Genomic Medicine,
Department of Advanced Biomedical Engineering and Science,
Tokyo Women's Medical University,
Tokyo, Japan

To do this, we selected a genetic disease that is rather common and whose disease mutations have already been identified. We concealed the causative mutations in the samples and examined whether the search for the disease mutations by the method of LD in a case-control study would be successful and, if so, what strategies would be optimal. Adenine phosphoribosyltransferase (*APRT*) deficiency is an autosomal recessive genetic disease (Kelley et al. 1968; Henderson et al. 1969; Cartier et al. 1974) that is common, as a genetic disease, among Japanese, although it is far less common than common diseases such as diabetes, hypertension, rheumatoid arthritis, and chronic glomerulonephritis. This *APRT* deficiency includes partial deficiency called Type II (*APRT*J*), which is observed only in Japanese (Kamatani et al. 1985), and complete deficiency called as Type I (*APRT*Q0*), which is observed in other ethnic groups as well (Sahota et al. 1995). The incidence of the defective Type II gene has been estimated to be 3.65×10^{-3} (Kamatani et al. 1996), and the incidence of the homozygotes has been estimated to be 1.33×10^{-5} . Previous studies have shown that the majority of disease chromosomes among the Japanese had a few common variants (Kamatani et al. 1992). By analysis of a few flanking RFLPs, it was suggested that the majority of variants derived from a few common ancestors (Kamatani et al. 1990).

APRT is coded for by a gene 2.6 kbp in length and located in 16q24.3 (Fratini et al. 1986). The defective mutations of this gene cause an autosomal recessive disease *APRT* deficiency. The clinical features of this disease include 2,8-dihydroxyadenine crystalluria (Laxdal and Jonasson 1988) and urolithiasis (Debray et al. 1976). The penetrance of this disease was estimated to be 85% in Type I homozygotes (Sahota et al. 1995), and the severity of the disease fluctuates very much from without symptoms to acute renal failure leading to hemodialysis or renal implantation. Such a fluctuation probably depends on environmental factors.

The analysis of mutations for *APRT* deficiency among Japanese patients has elucidated that about 78% of the defective chromosomes contain a single mutation designated *APRT*J* (Kamatani et al. 1989). Including the most frequent mutation, three different changes, *APRT*J* (Hidaka et al. 1988), *APRT*Q0c* (Mimori et al. 1991), and *APRT*Q04* (Kamatani et al. 1992), in the nucleotide sequence have been found to account for 96% of all the defective mutations (Kamatani et al. 1992). The above characteristics in the region of the defective gene are what we expected to find in the responsible genes for the diseases compatible with the common disease-common variant hypothesis. Therefore, the precise analyses of LD around the *APRT* gene on defective and control chromosomes may give us valuable information as to how we can find the responsible mutations using the LD.

Subjects and methods

Subjects DNA samples were obtained from 174 Japanese including 42 *APRT*J* homogenous subjects, 14 *APRT*J/Q0* heterogeneous subjects (containing 9 *J/Q0c*, 2 *J/Q04* and 3 *J/Q0u*), 24 *APRT*Q0* homogenous subjects (containing 13 *Q0c/Q0c*, 3 *Q04/Q04*, 2 *Q0s/Q0s*, 5 *Q0u/Q0u* and 1 *Q0c/Q0u*), and 94 control subjects.

Laboratory analysis PCR was performed in 50 μ l reaction mixture containing 200 μ M deoxyribonucleoside triphosphates (dNTPs), 1 nM PCR forward and reverse primers, 2.5 U AmpliTaq Gold (Applied Biosystems), 100 ng DNA sample, and PCR buffer (Applied Biosystems). The forward and reverse PCR primer sequences for the products including marker locus are No. 1 (GCAATGCTCCGCGAGGTAT and GAT-AGAATGCGGCTAACCCA), No. 2 (TGGCAGGC-TATGGTTCCC and GGCTTCTCCTCTGTTGAG-CA), No. 3 (GGGCCTTGGTGGTTCTC and CTTC-TGTGGTAGCGGTCCT), No. 4 (AGGCATCCTGA-CCAGAGTCT and CAAGTGCCCTATGACAG), No. 5 (ACTGCCCAGGAGCTGAGGAT and GACT-GCGGTGGTCCCAT), No. 6 (GGGTCGCCACGCA-GAG and GAGGCAGGAGAATCGCTTGAA), No. 7 (ACCGCCCGCAGCCAGAGACC and mismatch primer GCTGGACAAGGCCGCGGAGC), No. 8 (CTT-GTCTGCCCAGCTTCTC and GGCGAACATGG-TGAAACCC), No. 9, No. 10, No. 11, and No. 12 (AGAGGGTGGTCGTCGTGGAT and GAACAGG-AGGACAGGAGACG), No. 13, No. 14, No. 15, and No. 16 (CTGCACTCTGACCTGGAAGC and GGG-AGACCCCTTACCACCAGT), No. 17 and No. 18 (GGAGCCACAACACTGCCAGA and CTTCCCGT-ACTCCAGGGAAT), No. 19 (CCCACCCCAGGCG-TGGTATT and GCAGCTCTGCACCAGGGCTT), No. 20 (CATCTCGCCCGTCTCTGAA and CCTGGG-TGGGCATTCCGTGA), No. 21 (GAGGCTGAGGC-GGGAGAATG and GAGGGAGGCCGGAAGG-TGT), No. 22 (CTGGACGTGGACGGCTGAAGTG-TAG and TTACAATTCAGCTTTGGCCTGGCAGT-TACT), No. 23 (GTGAAACGCCACACGCACAG and CCCACGAGCAGCTTCCCTC), No. 24 (CAC-CATGCCCAGCTAATTCT and GCCTGTAATCC-CAGCACCGT), No. 25 (CAGTTGTACGGGCA-TCTG and GGTCGCGGAGGGTCTCT), No. 26 (GGCATGTTCTAGATTCACGC and TTTCCCCAC-CTGTTAATCA), No. 27 (GACCCTCCGCTCCAAG-CTG and GATGGGCTCAAGCATGGCTG), No. 28 (CGCTGCCGAGTTCCTGT and AGCTGCAGG-TTGGGGTAATG), No. 29 (GTGGGTGGGGTA-GGTG and CCTTCTGCTTCTTACCTT), No. 30 (AGCCAAGATCACGCCAGTA and GCAGGAG-AATGGCACGAACC), No. 31 (TGTCACCAGGCG-AAAC and GAATTCCCACATACACCTGC), No. 32 (CGCTGCCACCCACTCTAGT and CTGCTTCT-CCAGGCGACTGT) and No. 33 (TCCTCTGGGCT-

CCCTACTCC and CTGGCATTACAGGCGTGAG), respectively. In these loci, No. 9, 10, 13, and 14 are disease-causing loci of *Q0s* (Taniguchi et al. 1998), *J*, *Q0c* and *Q04*, respectively. Cycling conditions included an initial denaturation at 95°C for 9 min 30 s; 35 cycles at 95°C for 30 s, 55–65°C for 30 s, and 72°C for 1 min., 72°C for 6 min at the end of PCR. In some products, typing of SNP and insertion markers were executed by sequencing. The rest of the products were treated with the following restriction enzymes: AgeI (No. 25), ApaI (No. 20), ApaLI (No. 5), Asp700 (No. 21), AvaI (No. 16), BglII (No. 6), BpmI (No. 23), BsiEI (No. 14), BsrI (No. 12), EarI (No. 1), Eco0109I (No. 8), HinfI (No. 9 and No. 19), NgoIV (No. 24), NlaIII (No. 10 and No. 11), PfiMI (No. 13 and No. 27), PshAI (No. 32), PvuII (No. 28), SacI (No. 7 and No. 33), Sau3A (No. 17), SphI (No. 2, No. 15 and No. 22), TaqI (No. 18), Tsp45I (No. 3). These marker loci were typed with the length of the fragments by restriction enzymes. The allele type of VNTR marker loci were decided by the length of PCR products with 2% agarose gel electrophoresis in TBE. For the typing of microsatellite marker loci, PCR products were analyzed by DNA analyzer 3100 with POP-4 capillary gel matrix (Applied Biosystems). The numbers of repeats were estimated by the length of the PCR products.

Statistical analysis Marker loci information was referred to the NCBI SNP database based on NCBI Build 34 (<http://www.ncbi.nlm.nih.gov>). The inferences of the frequencies of haplotypes and the individual diplotype configurations were executed by expectation-maximization (EM) algorithm (Laird 1993; Lange 2002). The tests of significances of the tables of case-control analysis were executed with Fisher's exact test, Pearson's χ^2 test for test of independence, Δ TAC (Total Allele Content Difference; Collins et al. 2000), and MCMC (Markov Chain Monte Carlo; Metropolis et al. 1953; Hastings 1970; Gilks et al. 1995) methods. Because individual diplotype configurations were inferred with the EM algorithm, frequencies of haplotypes include decimal fractions. Therefore, frequencies of haplotypes were normalized to integer values for conservative direction in the tests with Fisher's exact test, Δ TAC, and MCMC methods. Δ TAC method gave us *P* values by approximation to χ^2_{n-1} distribution (*n* is the numbers of alleles or haplotypes). MCMC method gave us *P* values by the sum of the probabilities of contingency tables less than given table in generating probability with the Metropolis-Hastings method. LD block was estimated by the four-gamete test (Hudson and Kaplan 1985). If the relative frequency of the fourth two-marker loci haplotype was lower than 0.05, historical recombination was regarded not to have occurred between two-marker loci. Bayesian analysis of haplotypes for LD mapping was by BLADE (Liu et al. 2001).

Results

Identification of polymorphisms and determination of genotypes By searching for the sequence differences between samples from 94 controls and 80 *APRT* deficient subjects, we identified 33 sequence differences within and near the *APRT* gene, of which 15 SNPs were already known. The locations of all such polymorphic positions, including disease-related mutations, are shown in Table 1. They were numbered from centromeric to telomeric direction, but this direction was reverse to the *APRT* gene because the *APRT* gene was coded on reverse strand. Positions No. 9, 10, 13, and 14 are those for disease-related mutations. Among the disease-related loci, No. 14 was an insertion, while all of the others, No. 9, 10, and 13 were SNPs. As described before, the mutation at No. 10 was designated *APRT*J* that accounted for about 68% of all the disease-related mutations among Japanese (Kamatani et al. 1992). Most of the marker loci were SNPs, while 2 and 3 of them were VNTR (No. 4 and 26) and microsatellite loci (No. 29, 30, and 31), respectively.

Based on the above knowledge, we determined genotypes at 29 marker and four disease-related loci within and near the *APRT* gene for 80 *APRT* patients and 94 control subjects. Table 2 shows the relative frequencies of the minor alleles for control subjects, the heterozygosities, and the relative frequencies of the genotypes at each polymorphic site (only for SNP loci). The proportions of the genotypes for each polymorphic site were in accord with the Hardy-Weinberg equilibrium in all cases (result not indicated).

Comparison of allele frequencies between *APRT* deficient and control subjects When the allele frequencies were compared between *APRT*-deficient and control subjects, there were significant differences in most of the SNP markers (Table 3). *P* values were under 0.05 for markers Nos. 1, 2, 5, 6, 7, 8, 11, 18, 19, 20, 22, 23, 24, 25, 27, 28, and 32. The minor allele frequencies were over 0.1 for markers with *P* values under 0.05, except for Nos. 11, 23, and 25. The significance of the difference as judged by the *P* values fluctuated between markers. For markers Nos. 12, 15, and 16, significance of differences was not observed, while the minor allele frequencies of these loci were zero in the control population. Although the cases contained minor alleles for these loci, the frequencies were too low (<0.02) even in the cases. This fact explains the failure of detecting significant differences between the cases and controls. For most of the loci in which relative frequencies of minor alleles were high in the cases, *P* values tended to be low. However, *P* value for locus No. 25 was quite low in spite of the low relative frequency of the minor allele. The evident significance is not merely a function of the location. As shown in the changes of Pearson's χ^2 statistics (Fig. 1a) and *P* values by MCMC (Fig. 1b) against the recombi-

Table 1 Marker loci around and inside the *APRT* gene

Locus no.	Marker	Chr. position ^a	Nucleotide position for <i>APRT</i> *J ^a	Gene	dbSNP ^b	Polymorphism ^c
1	SNP	87561943	-1058445	DKFZp434G0522	rs888649	AAGASAGTG
2	SNP	88371699	-248689		rs2334243	GCATRCCAG
3	SNP	88450078	-170310	IL17C	rs869410	GCCCKTCAC
4	VNTR	88582715	-37673			40 bp
5	SNP	88591827	-28561		rs558436	GTGCRCCCC
6	SNP	88611127	-9261		rs1563671	AGGGRGCAG
7	SNP	88618778	-1610	CDT1	rs572275	ACCTSGCCC
8	SNP	88619624	-764	CDT1		GGGCYTTCA
9	SNP	88620253	-135	APRT		TGGTSACTC
10	SNP	88620388	0	APRT		GTTTCRTGGT
11	SNP	88620547	159	APRT		TCCCRTGTC
12	SNP	88620626	238	APRT		ACCAYCAGT
13	SNP	88621004	616	APRT		AGGCYCACA
14	insertion	88621036	648	APRT		TT(TCGG)TC
15	SNP	88621048	660	APRT		AGCAYGCAG
16	SNP	88621098	710	APRT		GCCTYGGGA
17	SNP	88621270	882	APRT	rs3214057	CCAGRTCAG
18	SNP	88621459	1071	APRT		TCTCSAGGT
19	SNP	88621575	1187	APRT		TGTGRCTCC
20	SNP	88622075	1687	APRT	rs2242173	GAAGSGCCC
21	SNP	88623098	2710		rs537641	CCTTYTGTC
22	SNP	88623697	3309			CATGMGGCT
23	SNP	88634752	14364	GALNS	rs507540	GCTGRAGCG
24	SNP	88655808	35420	GALNS	rs506914	AGCCRGCG
25	SNP	88689210	68822	CBFA2T3		ACACSGGTT
26	VNTR	88700534	80146	CBFA2T3		49 bp
27	SNP	88708002	87614	CBFA2T3	rs522145	ATAAMCACC
28	SNP	88766727	146339	CBFA2T3	rs2341879	CAGCYGCCT
29	STRP	88775464	155076	CBFA2T3		ATCC
30	STRP	89019320	398932			ATTT
31	STRP	89032139	411751			TG
32	SNP	89102568	482180	LZ16	rs731135	CATCRACCT
33	SNP	89571046	950658	FANCA	rs2239357	CCAGRGCTC

^aThe location of SNPs, an insertion, STRPs, and VNTRs are shown as chromosome position numbers based on NCBI Build 34. Locations of STRPs and VNTRs are shown as the nucleotide position at the edges of the repeats closer to *APRT**J

^brs numbers for the SNP loci are shown as registered in dbSNP
^cSNPs are expressed as m, r, s, y, and k for the polymorphism of AC, AG, CG, CT, and GT, respectively. Polymorphisms of STRPs and VNTRs are shown by motifs and motif length, respectively

nation fraction for *APRT**J designated locus, the relation between the significance of the difference in allele frequencies and the distance to *APRT**J mutation locus was not linear.

D' and r^2 between marker loci and *APRT**J-causing locus in the general population LD were expected to decrease with increasing physical distance between marker loci and *APRT**J-causing locus. We estimated *D'* and r^2 in the virtual population using data from control and *APRT**J subjects. *D'* and r^2 were estimated using *APRT**J subjects under the assumption that the population frequency of the *APRT**J mutation was 0.00365 (Fig. 2). The decrease of *D'* was not merely proportional to physical distance. In the virtual population, which consisted of *APRT**J and control subjects, LD between a marker locus and *APRT**J mutation locus was complete ($D'=1$) for the region between 882 bp upstream of *APRT**J mutation and 1,610 bp downstream of *APRT**J mutation locus, whereas r^2 was very low. This region almost accurately corresponded to the haplotype block detected as in Table 4. The graph of r^2 was similar

to the graph of the χ^2 statistic (Fig. 1a), whereas the values of r^2 were very low because of the extremely low frequency of the *APRT**J mutation allele.

Estimation of haplotype blocks in control subjects Haplotype block, including *APRT**J mutation site, was estimated by four-gamete test (Hudson and Kaplan 1985). Among marker Nos. 5, 6, 7, 8, 17, and 18 (minor allele frequency > 10%), pairwise haplotypes and their relative frequencies were estimated by EM algorithm (Table 4). If the relative frequency of the fourth haplotype for two-marker loci was lower than 0.05, historical recombination was regarded not to have occurred between the two-marker loci. Among marker Nos. 6, 7, 8, and 17, the relative frequencies of the fourth haplotypes were lower than 0.05, whereas they were higher than 0.05 between No. 5 and others or between No. 18 and others. According to these results, it was thought that historical recombination had occurred in the sites between Nos. 5 and 6, between Nos. 17 and 18. Therefore, haplotype block, including the *APRT**J mutation site, occupied the region between Nos. 6 and 17.

Table 2 Relative frequencies of the minor alleles, heterozygosities, and relative frequencies of genotypes for SNP marker loci

Locus no.	Relative frequency	Heterozygosity	Genotype frequency ^a		
			11	12	22
1	0.385	0.473	0.418	0.396	0.187
2	0.176	0.289	0.660	0.330	0.011
3	0.027	0.052	0.947	0.052	0.000
5	0.457	0.496	0.340	0.404	0.255
6	0.420	0.487	0.372	0.415	0.213
7	0.250	0.375	0.564	0.372	0.064
8	0.430	0.490	0.355	0.430	0.215
11 ^b	0.000	0.000	1.000	0.000	0.000
12 ^b	0.000	0.000	1.000	0.000	0.000
15 ^b	0.000	0.000	1.000	0.000	0.000
16 ^b	0.000	0.000	1.000	0.000	0.000
17	0.186	0.303	0.681	0.266	0.053
18	0.479	0.499	0.287	0.468	0.245
19	0.473	0.499	0.298	0.457	0.245
20	0.117	0.207	0.777	0.213	0.011
21	0.160	0.268	0.745	0.191	0.064
22	0.433	0.491	0.356	0.422	0.222
23	0.097	0.175	0.849	0.108	0.043
24	0.234	0.359	0.596	0.340	0.064
25	0.011	0.021	0.978	0.022	0.000
27	0.430	0.490	0.323	0.495	0.183
28	0.101	0.182	0.809	0.181	0.011
32	0.309	0.427	0.447	0.489	0.064
33	0.074	0.138	0.851	0.149	0.000

Each value is obtained from the control population

^aAlleles 1 and 2 denote the major allele and minor alleles, respectively, at each locus

^bPolymorphisms were not observed in the control subjects at Nos. 11, 12, 15 and 16 loci

Table 3 Significance of the difference in allele frequencies between *APRT*-deficient and control subjects at SNP marker loci

Locus no.	Pearson's χ^2 test ^a <i>P</i> value	MCMC ^a <i>P</i> value
1	2.46×10 ⁻⁵	5.60×10 ⁻⁵
2	1.70×10 ⁻⁵	1.00×10 ⁻⁵
3	5.62×10 ⁻¹	7.60×10 ⁻¹
5	2.30×10 ⁻²	2.85×10 ⁻²
6	0.00 ^b	< 10 ⁻⁶
7	7.77×10 ⁻¹⁶	< 10 ⁻⁶
8	0.00 ^b	< 10 ⁻⁶
11	7.40×10 ⁻³	9.32×10 ⁻³
12	1.24×10 ⁻¹	2.12×10 ⁻¹
15	2.78×10 ⁻¹	4.60×10 ⁻¹
16	1.24×10 ⁻¹	2.12×10 ⁻¹
17	1.15×10 ⁻¹	1.22×10 ⁻¹
18	2.45×10 ⁻³	2.87×10 ⁻³
19	2.08×10 ⁻¹¹	< 10 ⁻⁶
20	7.80×10 ⁻⁶	2.00×10 ⁻⁶
21	9.10×10 ⁻²	1.03×10 ⁻¹
22	1.29×10 ⁻¹⁴	< 10 ⁻⁶
23	1.91×10 ⁻⁴	2.29×10 ⁻⁴
24	1.51×10 ⁻¹³	< 10 ⁻⁶
25	0.00 ^b	< 10 ⁻⁶
27	2.11×10 ⁻¹⁵	< 10 ⁻⁶
28	3.54×10 ⁻⁵	2.00×10 ⁻⁶
32	3.17×10 ⁻³	4.19×10 ⁻³
33	8.20×10 ⁻²	9.95×10 ⁻²

^a*P* values by Pearson's χ^2 test for independence and MCMC method are shown

^b*P* values could not be calculated in Nos. 6, 8, and 25 because the χ^2 statistics were too large

Haplotype inference We then inferred haplotypes for the region extending from Nos. 6–17 that was defined as haplotype block for the control subjects using common SNPs (minor allele frequency > 10%). Table 5 shows that the majority (95%) of haplotypes were accounted for by five haplotypes in the control chromosomes. Haplotype heterozygosity was 0.562. Among the chromosomes from *APRT*-deficient subjects, the majority of haplotypes were accounted for by only four different haplotypes, and the haplotype heterozygosity was 0.555. When the frequency of each haplotype was compared between *APRT*-deficient and control subjects, the frequency of haplotype No. 2 was much higher among the *APRT*-deficient subjects than the control subjects. When the significance of the difference in haplotype frequencies between *APRT*-deficient and control subjects was compared, *P* values by Fisher's exact test, Δ TAC, and MCMC methods were 9.74×10^{-23} , 1.29×10^{-9} (Δ TAC = 47.34), and $< 10^{-6}$, respectively. When the haplotype inference was performed using only the *APRT*J* chromosomes, the number of major haplotypes was two, while the haplotype heterozygosity was 0.133. In fact, the haplotype No. 2 that was present at a much higher frequency among the *APRT*-deficient chromosomes than the control chromosomes was the haplotype to which almost all the chromosomes containing *APRT*J* belonged. When the significance of the

difference in haplotype frequencies between *APRT*J* and control subjects was compared, *P* values by Fisher's exact test, Δ TAC, and MCMC methods were 1.92×10^{-34} , 3.55×10^{-15} (Δ TAC = 73.83), and $< 10^{-6}$, respectively.

When the number of haplotypes was estimated using disease chromosomes, it was fewer for disease chromosomes than control chromosomes. The haplotype heterozygosity was lower for disease chromosomes than for control chromosomes. Although the above characteristics are the most prominent, when all the disease chromosomes were of a single origin, they were still detectable even when the disease chromosomes were derived from a few origins. The test of the difference in frequency distributions of haplotypes between case and control chromosomes may be possible by the Fisher's exact test, Δ TAC, and MCMC methods.

Comparison of allele frequencies for STRP and VNTR markers The comparison of allele frequencies at multiallelic loci is generally more complicated than SNP loci. We tested the significances for independence of allele distributions between *APRT*-deficient subjects and control subjects with Fisher's exact test, Δ TAC values method, and MCMC method testing for the independence of $2 \times m$ contingency tables. We found three STRP and two VNTR marker loci, one locating downstream of *APRT* gene (No. 4) and the other four upstream of the

Fig. 1a,b Pearson's χ^2 statistics and logarithm of P values from MCMC to test differences in allele frequencies between cases and controls. **a** Pearson's χ^2 statistics. **b** Logarithm of P values by MCMC method. Horizontal axis shows the locations of the marker loci as shown by the logarithm of physical distances (bp) from the $APRT^*J$ mutation locus (J)

gene (Nos. 26, 29, 30, and 31). At No. 4 VNTR locus, 97.4% (152/156) of all $APRT$ -deficient chromosomes had the 11 repeat alleles, while the 11 repeat alleles were observed in only 69.4% (129/186) of control chromosomes (Table 6). P values by three tests indicated that 11 repeat alleles at No. 4 locus were significantly linked to $APRT$ -deficient mutations. As will be stated later in this manuscript, all chromosomes containing the $APRT^*J$ mutation had the same 11 repeat polymorphism at the No. 4 locus (Table 8). This result indicated that the $APRT^*J$ mutation was completely linked to a single 11-repeat polymorphism at the No. 4 VNTR locus locating downstream of the $APRT$ gene.

For the upstream STRP and VNTR loci, P values for the differences in allele distributions between $APRT$ -deficient (case) and control subjects were less than the general significance level (0.05), except for the No. 26 marker locus (Table 7). P values for the difference between $APRT^*J$ and control subjects were lower than between case and control subjects for these marker loci. Here, $APRT$ -deficient subjects include those with $APRT^*J$ and other mutant alleles. Therefore, the tests for the subjects with multiple mutant alleles resulted in

Fig. 2a,b D' and r^2 between the $APRT^*J$ mutation locus and a marker locus. $APRT^*J$ allele frequency was assumed to be 0.00365. **a** D' . **b** r^2 . Horizontal axis shows the locations of the loci as shown by the logarithm of physical distances (bp) from the $APRT^*J$ mutation locus (J)

higher P values than the test for the subjects with a single mutation. For upstream STRP and VNTR loci, P values were higher for loci (Nos. 26 and 29) closer to the $APRT$ gene than father loci (Nos. 30 and 31). However, the higher P values in closer loci do not indicate a weaker LD but seem to be caused by low heterozygosities at those loci in control subjects. Those results indicate that the analysis of STRP or VNTR loci with high heterozygosity in control subjects is necessary to detect disease-related loci efficiently.

Linkage between $APRT$ deficiency mutations and repeat markers Haplotypes were inferred using one of the $APRT$ mutation loci and a repeat marker locus. The results are shown in Table 8. The downstream VNTR locus No. 4 was in complete LD with the $APRT^*J$ mutation (including J homo and $J/Q0$ hetero), and a single 11 repeat polymorphism was completely linked to $APRT^*J$ mutation. The $Q0$ mutation was nearly completely linked to the same 11 repeat. For the upstream marker locus No. 30, the most frequent allele linked to $APRT^*Q0$ mutation was different from the most frequent allele linked to nonmutation. On the other hand, the most frequent allele linked to $APRT^*J$ mutation was

Table 4 Haplotype blocks inferred by four-gamete test. Pairwise haplotypes were constructed by two-marker loci between Nos. 5, 6, 7, 8, 17 and 18 in control subjects. For all pair of loci, possible four haplotypes and their relative frequencies were estimated by EM algorithm. Each cell contains four haplotypes and their relative frequencies

Locus no.	5	6	7	8	17
6	GG	0.391			
	AA	0.269			
	AG	0.189			
	GA	0.152			
7	GC	0.453	GC	0.554	
	AC	0.297	AG	0.224	
	AG	0.161	AC	0.196	
	GG	0.089	GG	0.026 ^a	
8	GC	0.387	GC	0.547	CC 0.561
	AT	0.273	AT	0.398	GT 0.243
	AC	0.185	GT	0.033	CT 0.189
	GT	0.155	AC	0.022 ^a	GC 0.007 ^a
17	GA	0.472	GA	0.567	CA 0.564
	AA	0.342	AA	0.247	TA 0.253
	AG	0.115	AG	0.173	CG 0.186
	GG	0.071	GG	0.013 ^a	CG 0.000 ^a
18	GG	0.316	GG	0.344	CC 0.411
	AC	0.252	AC	0.243	CG 0.339
	GC	0.227	GC	0.236	GG 0.183
	AG	0.206	AG	0.177	GC 0.067

^aIf the relative frequency of the fourth haplotype was less than 0.05, it was regarded that historical recombination had not occurred between the pair of the marker loci. By this test, the region between Nos. 6 and 17 was recognized as a haplotype block

Table 5 Frequencies of haplotypes within the haplotype block in the *APRT* gene and haplotype heterozygosities in *APRT*J*, *APRT*-deficient, and control subjects. Haplotype block (Nos. 6, 7, 8 and 17) was constructed using the data from control subjects in Table 4,

Haplotypes with higher frequencies were collected until the accumulated frequencies became over 95%, and only such common haplotypes were shown

	Haplotype no.	Haplotype	Frequency ^a			Heterozygosity
			Inferred	Relative	Normalized ^b	
<i>APRT*J</i>	2	agta	78.00	0.929	78	0.133
	4	ggta	6.00	0.071	6	
Case	2	agta	99.59	0.622	99	0.555
	3	actg	36.00	0.225	36	
	1	gccca	11.37	0.071	12	
Control	4	ggta	6.02	0.038	6	0.562
	1	gccca	100.62	0.535	100	
	2	agta	39.95	0.213	40	
	3	actg	31.14	0.166	32	
	4	ggta	5.26	0.028	6	
	5	accca	3.07	0.016	3	

^aFrequencies were inferred by EM algorithm

^bNormalized frequencies indicate the integers normalized from inferred values

different from the most frequent allele linked to non-mutation at No. 31. If a mutation causative of disease had occurred on the chromosome with low frequency marker allele, “the low frequency allele” linked to the mutation has spread on the case population. The marker loci Nos. 30 and 31 should be examples. These results indicate that chromosomes containing a common disease-causing mutation tend to have a common polymorphism at a nearby locus.

Bayesian analysis of haplotypes for linkage disequilibrium mapping Since the *APRT*-deficient chromosomes have multiple ancestral haplotypes, it is of interest to analyze the data using BLADE (Liu et al. 2001), a recently released software for the Bayesian analysis of haplotypes for LD mapping. Figure 3 indicates the results obtained

when the data for both *APRT*-deficient and control chromosomes were applied to the software. For this study, all data for the disease-related loci were removed. When the data for the control chromosomes were applied to the software, the posterior probabilities of the mutation clusters to which each current haplotype belongs tended to be dispersed (Fig. 3d). In contrast, when BLADE was applied with *APRT*-deficient chromosomes, *APRT*J* and *APRT*Q0c* chromosomes were separated from the other mutation chromosomes if the number of clusters was up to four (Fig. 3a,b). *APRT*Q04* chromosomes were separated from ‘null’ when the number of clusters became five (Fig. 3c). However, *APRT*J* chromosomes were not concentrated in a single cluster. This may reflect the fact that the *APRT*J* mutation is older than *APRT*Q0* mutations, and *APRT*J* chromosomes were judged to have two

Table 6 Allele number of repeat marker loci and heterozygosity in *APRT*J*, *APRT*-deficient and control subjects

Locus no.	Contingency table ^a							Heterozygosity
	Repeat	11	13	8	6	7	10	
4	<i>APRT*J</i>	82	0	0	0	0	0	0.000
	Case	152	2	0	2	0	0	0.050
	Control	129	27	19	9	1	1	0.485
	Repeat	3	2	4				
26	<i>APRT*J</i>	79	5	0				0.113
	Case	153	7	0				0.084
	Control	168	18	2				0.192
	Repeat	12	11	13	10	14		
29	<i>APRT*J</i>	80	4	0	0	0		0.091
	Case	147	10	3	0	0		0.152
	Control	153	19	7	7	2		0.325
	Repeat	14	13	12	9	15	11	
30	<i>APRT*J</i>	57	6	8	12	1	0	0.504
	Case	70	51	17	16	2	4	0.685
	Control	69	49	49	12	6	1	0.718
	Repeat	17	18	20	21	16	19	
31	<i>APRT*J</i>	20	52	9	1	0	0	0.526
	Case	75	63	17	1	0	2	0.604
	Control	110	37	28	8	2	1	0.586

^aThe differences in allele frequencies between symptomatic (*APRT*J* or *APRT*-deficient) and control subjects were tested

Table 7 Significance of difference in allele frequencies between *APRT*J* or *APRT*-deficient and control subjects

Locus no.		Fisher's exact test ^a	ΔTAC^a	MCMC ^a
4	<i>APRT*J</i>	2.01×10^{-8}	1.10×10^{-5}	$< 10^{-6}$
	Case	5.59×10^{-12}	3.50×10^{-5}	$< 10^{-6}$
26	<i>APRT*J</i>	4.70×10^{-1}	9.60×10^{-2}	3.69×10^{-1}
	Case	5.58×10^{-2}	4.37×10^{-2}	4.46×10^{-2}
29	<i>APRT*J</i>	2.91×10^{-2}	7.77×10^{-3}	2.74×10^{-2}
	Case	1.46×10^{-2}	3.29×10^{-2}	1.40×10^{-2}
30	<i>APRT*J</i>	1.87×10^{-7}	2.87×10^{-7}	$< 10^{-6}$
	Case	1.80×10^{-3}	3.35×10^{-3}	1.56×10^{-3}
31	<i>APRT*J</i>	2.57×10^{-10}	2.90×10^{-8}	$< 10^{-6}$
	Case	1.95×10^{-4}	9.19×10^{-4}	2.20×10^{-4}

^aThe differences in allele frequencies between symptomatic (*APRT*J* or *APRT*-deficient) and control subjects were tested by Fisher's exact test, ΔTAC , and MCMC methods

origins. On the other hand, *APRT*Q0c* and *Q04* chromosomes were concentrated in single clusters. Therefore, BLADE software successfully clustered the *APRT*Q0c* and *Q04* chromosomes according to the disease-related mutations without the direct data of the disease-related mutations, whereas *APRT*J* chromosomes were not successfully clustered. When the maximum likelihood of the data was calculated assuming varying numbers of clusters (excluding the null cluster), the curves showing the relation between the number of clusters and the maximum likelihood were quite different for *APRT*-deficient and control chromosomes. Thus, when data from the *APRT*-deficient chromosomes were used, the maximum likelihood increased, with increasing assumed number of clusters as long as the number of clusters was not more than three, and it stayed at the same level at assumed cluster numbers over three (Fig. 4). In contrast,

Table 8 Linkage between *APRT* deficiency mutation and repeat markers. Each cell contains the motif repeats in alleles and their frequencies. Frequencies were inferred by EM algorithm

	No. 4		No. 26		No. 29		No. 30		No. 31	
<i>APRT*J</i>	11	96.00	3	93.00	12	94.00	14	66.96	18	63.78
			2	5.00	11	4.00	9	12.00	17	22.07
							13	10.51	20	11.13
							12	7.53	21	1.02
							15	1.00		
<i>APRT*Q0</i>	11	56.00	3	60.00	12	53.00	13	40.49	17	52.95
	13	2.00	2	2.00	11	6.00	12	9.47	20	6.05
	6	2.00			13	3.00	9	4.00	19	2.00
							11	4.00	18	1.00
							14	3.04		
Nondisease							15	1.00		
	11	129.00	3	168.00	12	153.00	14	69.00	17	110.00
	13	27.00	2	18.00	11	19.00	13	49.00	18	37.00
	8	19.00	4	2.00	13	7.00	12	49.00	20	28.00
	6	9.00			10	7.00	9	12.00	21	8.00
	10	1.00			14	2.00	15	6.00	16	2.00
	7	1.0					11	1.00	19	1.00

Fig. 3a–d Affiliation of *APRT*-deficient and control chromosomes derived from the analysis with BLADE.

Haplotypes of chromosomes were constructed with 16 SNP marker loci with minor allele frequencies over 0.1 (Nos. 1, 2, 5, 6, 7, 8, 17, 18, 19, 20, 21, 22, 24, 27, 28, and 32). *APRT*-deficient chromosomes were clustered to **a** three, **b** four, and **c** five clusters, while control chromosomes were clustered to **d** seven clusters. Each vertical bar indicates a chromosome. *J*, *c*, *s*, 4 and *u* in horizontal axis denote *APRT***J*, *APRT***Q0c*, *APRT***Q0s*, *APRT***Q04*, and *APRT***Q0u*, respectively. Some chromosomes were clustered to the “null” cluster

the maximum likelihood for the data from control chromosomes increased with the increasing number of clusters of over three. These results indicate that the suitable cluster number in *APRT*-deficient chromosomes was three, whereas the suitable cluster number in control chromosomes was not decided.

Discussion

Common disease-common variant hypothesis assumes that the majority of the disease-related mutations for a common disease are accounted for by a limited number of the common variants derived from common origins. If this hypothesis is right, LD is expected to be present

between the disease-related loci and flanking marker loci. The disease chromosomes are expected to be derived from a limited number of ancestral haplotypes concerning the disease-related region. Based on this hypothesis, genome-wide studies searching for disease-related genes for common diseases are now under way. Such studies are not intended to identify disease-related mutations directly but to encounter some marker loci that are in LD with the disease-related loci.

However, it is not clear whether this approach can successfully identify disease-related genes for common diseases. Even if it can, the optimal conditions of such investigations remain to be clarified. The questions include what types of markers are to be used, how large the distance between markers should be, and whether

Fig. 4 Maximum log likelihood for the numbers of clusters. Circle and triangle indicate *APRT*-deficient and control chromosomes, respectively. The log likelihoods are output results from BLADE

the marker loci should be concentrated in the genes. We could use SNPs, STRPs, and VNTR markers, each of which has different characteristics. SNPs are present at the highest frequency, while a single STRP locus has multiple alleles. The mutation rates for SNPs are likely to be very low, while those for STRPs may be rather high.

Our results showed that the comparison of the allele frequencies at marker loci between *APRT*-deficient and control subjects was an effective method to identify loci that are in LD with the true disease-related loci. Although SNPs were present at the highest frequency, both STRP and VNTR may serve as markers by which the true disease-related loci can be identified, because the frequency distributions of the alleles at both STRP and VNTR loci were significantly different between *APRT*-deficient and control subjects when the statistic analysis was performed by the Fisher's exact test, Δ TAC, and MCMC method.

In the analyses with STRP and VNTR loci, *P* values for the tests of the differences in the frequencies at single loci were not sufficiently low when the heterozygosities were low. However, the marker loci with high heterozygosities generally gave significant difference. Especially at No. 4 VNTR locus that is 37.7 kbp apart from the *APRT***J* mutation locus, 11 repeat polymorphism was completely linked to *APRT***J* mutation (Table 8). At Nos. 30 and 31 loci, very low *P* values were also obtained. The most frequent alleles were different at Nos. 30 and 31 loci between *APRT***Q0* and *J* mutation chromosomes. The above evidence suggests that LD analysis to search for disease-causing loci is feasible in the population of common origin.

Comparison of frequencies of marker alleles of SNPs between affected and control subjects are also likely to be a sensitive method to identify candidate markers associated with the disease-related loci. *P* values tended to be low for the markers with high heterozygosities (Tables 2, 3). Even if the marker loci are close to the disease-related loci, the evident significance was not

shown when the frequencies of the minor alleles were low (Fig. 1, Table 3). Even though some marker loci were in the complete LD ($D' = 1$) with the disease loci, significant *P* values were not observed for those loci (Figs. 1, 2a). However, loci with low heterozygosities (No. 25, for example) occasionally gave significant *P* values. These results suggest that *P* values in single-marker loci do not always reflect D' between disease-related and marker loci. On the other hand, the r^2 values were extremely low as though the values denied the LD. Both D' and r^2 are insufficient as measure of LD when the haplotype frequencies are in extreme imbalance.

When we performed a haplotype analysis using SNP, haplotypes had to be constructed within haplotype blocks for the inferred haplotypes to be reliable. A haplotype block with a size of 10.1 kbp was observed using the data from the control chromosomes by four-gamete test. This haplotype block did not cover the entire *APRT* gene. When the haplotypes were inferred by the EM algorithm-based method, the frequencies of haplotype as well as the most frequent haplotypes were different between the *APRT*-deficient and control chromosomes (Table 5). Therefore, comparison of frequencies of haplotypes between case and control chromosomes is likely to serve as an effective method to identify disease-related loci.

Recently-released BLADE algorithm for the Bayesian analysis of haplotypes for LD mapping is of interest because of its flexibility. The analysis of our data using the algorithm indicated that it may serve as a powerful method to identify disease-related chromosomes from marker genotypes. Thus, when the data from the *APRT*-deficient chromosomes were applied, *APRT***Q0c* and *Q04* chromosomes were concentrated in single clusters, whereas *APRT***J* chromosomes were not concentrated in a single cluster (Fig. 3). This split of *APRT***J* chromosomes' affiliation may reflect the fact that *APRT***J* mutation is older than *APRT***Q0* mutations and original single haplotype may have been split into two by recombination in *APRT***J* chromosomes. Therefore, *APRT***J* chromosomes had been concentrated in a single cluster when the data from a more narrow region (Nos. 2–32) were applied to BLADE (data not shown). When BLADE software is applied, data from suitable-length regions should be used. If suitable data are used, this algorithm could successfully estimate disease-related chromosomes without the data of disease-related loci as shown in *APRT***Q0c* and *Q04*.

Although *APRT* deficiency is not a common disease, the combined frequency of the disease-related mutations is among the highest of all the genetic diseases in the Japanese. Data obtained from the analysis of LD and haplotypes related to this genetic deficiency may lead to a better understanding of the mode of LD and haplotypes and provide valuable information about the relationship between disease-related mutations and LD in common diseases in general. Our results suggest that the comparison of marker allele frequencies between case

and control chromosomes is a sensitive method to list up candidates while the analysis of haplotypes using EM-based haplotype inference and the Bayesian method may serve as tests to examine whether the candidate regions are real disease-related loci.

Acknowledgements This study was supported by a grant from Research for the Future Program from the Japan Society for the Promotion of Science (JSPS).

References

- Angius A, Petretto E, Maestrale GB, Forabosco P, Casu G, Piras D, Fanciulli M, Falchi M, Melis PM, Palermo M, Pirastu M (2002) A new essential hypertension susceptibility locus on chromosome 2p24–2p25, detected by genomewide search. *Am J Hum Genet* 71:893–905
- Cartier P, Hamet M, Hamburger J (1974) Une nouvelle maladie métabolique: le déficit en adénine phosphoribosyltransferase avec lithiase de 2,8-dihydroxyadénine. *CR Seances Acad Sci* 279:883–886
- Collins HE, Li H, Inda SE, Anderson J, Laiho K, Tuomilehto J, Seldin MF (2000) A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. *Hum Genet* 106:218–226
- Debray H, Cartier P, Temstet A, Cendron J (1976) Child's urinary lithiasis revealing a complete deficit in adénine phosphoribosyltransferase. *Pediatr Res* 10:762–766
- Fratini A, Simmers RN, Callen DF, Hyland VJ, Tischfield JA, Stambrook PJ, Sutherland GR (1986) A new location for the human adénine phosphoribosyltransferase gene (APRT) distal to the haptoglobin (HP) and fra(16)(q23) (FRA16D) loci. *Cytogenet Cell Genet* 43:10–13
- Gilks WR, Richardson S, Spiegelhalter DJ (1995) Markov Chain Monte Carlo in practice. Chapman & Hall, London
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Henderson JF, Kelley WN, Rosenbloom FM, Seegmiller JE (1969) Inheritance of purine phosphoribosyltransferases in man. *Am J Hum Genet* 21:61–70
- Hidaka Y, Tarle SA, Fujimori S, Kamatani N, Kelley WN, Pallela TD (1988) Human adénine phosphoribosyltransferase deficiency: demonstration of a single mutant allele common to the Japanese. *J Clin Invest* 81:945–950
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164
- Kamatani N, Takeuchi F, Nishida Y, Yamanaka H, Hishioka K, Tataru K, Fujimori S, Kaneko K, Akaoka I, Tofuku Y (1985) Severe impairment in adénine metabolism with a partial deficiency of adénine phosphoribosyltransferase. *Metabolism* 34:164–168
- Kamatani N, Kuroshima S, Terai C, Hidaka Y, Pallela TD, Nishioka K (1989) Detection of an amino acid substitution in the mutant enzyme for a special type of adénine phosphoribosyltransferase (APRT) deficiency by sequence-specific protein cleavage. *Am J Hum Genet* 45:325–331
- Kamatani N, Kuroshima S, Hakoda M, Pallela TD, Hidaka Y (1990) Crossovers within a short DNA sequence indicate a long evolutionary history of the APRT*J mutation. *Hum Genet* 85:600–604
- Kamatani N, Hakoda M, Otsuka S, Yoshikawa H, Kashiwazaki S (1992) Only three mutations account for almost all defective alleles causing adénine phosphoribosyltransferase deficiency in Japanese patients. *J Clin Invest* 90:130–135
- Kamatani N, Terai C, Kim SY, Chen CL, Yamanaka H, Hakoda M, Totokawa S, Kashiwazaki S (1996) The origin of the most common mutation of adénine phosphoribosyltransferase among Japanese goes back to a prehistoric era. *Hum Genet* 98:596–600
- Kelley WN, Levy RI, Rosenbloom FM, Henderson JF, Seegmiller JE (1968) Adénine phosphoribosyltransferase deficiency: a previously undescribed genetic defect in man. *J Clin Invest* 47:2281–2289
- Laird N (1993) The EM algorithm. In: Rao CR (ed) Computational statistics. Elsevier, New York, pp 509–520
- Lange K (2002) Mathematical and statistical methods for genetic analysis, 2nd edn. Springer, New York Berlin Heidelberg
- Laxdal T, Jonasson TA (1988) Adénine phosphoribosyltransferase deficiency in Iceland. *Acta Med Scand* 224:621–626
- Liu JS, Sabatti C, Teng J, Keats BJB, Risch N (2001) Bayesian analysis of haplotypes for linkage disequilibrium mapping. *Genome Res* 11:1716–1724
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equations of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Mimori A, Hidaka Y, Wu VC, Tarle SA, Kamatani N, Kelley WN, Pallela TD (1991) A mutant allele common to the type I adénine phosphoribosyltransferase deficiency in Japanese subjects. *Am J Hum Genet* 48:103–107
- Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson VL, Silander K, Watanabe RM, Chines PS, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2001) Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. *Genome Res* 11:1221–1226
- Sahota AS, Tischfield JA, Kamatani N, Simmonds HA (1995) Adénine phosphoribosyltransferase deficiency and 2,8-dihydroxyadénine lithiasis. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW, Vogelstein B (eds) The metabolic & molecular bases of inherited disease. McGraw-hill, New York, pp 2571–2584
- Taniguchi A, Hakoda M, Yamanaka H, Terai C, Hikiji K, Kawaguchi R, Konishi N, Kashiwazaki S, Kamatani N (1998) A germline mutation abolishing the original stop codon of the human adénine phosphoribosyltransferase (APRT) gene leads to complete loss of the enzyme protein. *Hum Genet* 102:197–202