

## A Founder Mutation in the GK1 Gene Is Responsible for Galactokinase Deficiency in Roma (Gypsies)

Luba Kalaydjieva,<sup>1</sup> Anna Perez-Lezaun,<sup>1,\*</sup> Dora Angelicheva,<sup>1</sup> Suna Onengut,<sup>1,†</sup> Danielle Dye,<sup>1</sup> Nils U. Bosshard,<sup>3</sup> Albena Jordanova,<sup>4</sup> Alexei Savov,<sup>4</sup> Peter Yanakiev,<sup>4</sup> Ivo Kremensky,<sup>4</sup> Brigitta Radeva,<sup>4</sup> Joachim Hallmayer,<sup>2</sup> Arseni Markov,<sup>5</sup> Vanya Nedkova,<sup>6</sup> Ivailo Tournev,<sup>4</sup> Lidia Aneva,<sup>7</sup> and Richard Gitzelmann<sup>3</sup>

<sup>1</sup>Centre for Human Genetics, Edith Cowan University, Joondalup Campus and <sup>2</sup>Centre for Clinical Research in Neuropsychiatry, Graylands Hospital/University of Western Australia, Perth, Australia; <sup>3</sup>Division of Metabolic and Molecular Diseases, University Children's Hospital, Zürich; <sup>4</sup>Medical University, Sofia, Bulgaria; <sup>5</sup>Institute of Medical Biochemistry, University of Münster, Germany; <sup>6</sup>Department of Pediatrics, Medical School, Pleven, Bulgaria; and <sup>7</sup>Department of Clinical Chemistry, Regional Hospital, Blagoevgrad, Bulgaria

### Summary

Galactokinase deficiency is an inborn error in the first step of galactose metabolism. Its major clinical manifestation is the development of cataracts in the first weeks of life. It has also been suggested that carriers of the deficiency are predisposed to presenile cataracts developing at age 20–50 years. Newborn screening data suggest that the gene frequency is very low worldwide but is higher among the Roma in Europe. Since the cloning of the galactokinase gene (GK1) in 1995, only two disease-causing mutations, both confined to single families, have been identified. Here we present the results of a study of six affected Romani families from Bulgaria, where index patients with galactokinase deficiency have been detected by the mass screening. Genetic linkage mapping placed the disease locus on 17q, and haplotype analysis revealed a small conserved region of homozygosity. Using radiation hybrid mapping, we have shown that GK1 is located in this region. The founder Romani mutation identified in this study is a single nucleotide substitution in GK1 resulting in the replacement of the conserved proline residue at amino acid position 28 with threonine (P28T). The P28T carrier rate in this endogenous population is ~5%, suggesting that the mutation may be an important cause of early childhood blindness in countries with a sizeable Roma minority.

### Introduction

Galactokinase (EC 2.7.1.6) catalyzes the first step of galactose phosphorylation in the Leloir pathway of intermediate metabolism. Galactokinase deficiency (MIM 230200), one of the three inborn errors of metabolism that lead to hypergalactosemia, was first described by Gitzelmann (1965, 1967). The disorder is inherited as an autosomal recessive trait. Unlike classic galactosemia, which is caused by deficiency of galactose-1-phosphate uridylyltransferase (MIM 230400), galactokinase deficiency does not present with severe manifestations in early infancy. Its major clinical symptom is the development of cataracts during the first weeks or months of life, as a result of the accumulation, in the lens, of galactitol, a product of an alternative route of galactose utilization. The development of early cataracts in homozygous affected infants is fully preventable through early diagnosis and treatment with a galactose-restricted diet. Some studies have suggested that, depending on milk consumption later in life, heterozygous carriers of galactokinase deficiency may be prone to presenile cataracts at 20–50 years of age (Stambolian et al. 1986).

Galactokinase deficiency is generally considered to be rare, compared to classic galactosemia; however, the estimates of gene frequency vary widely, and an east-to-west gradient appears to be present across Europe (table 1). Observations coming from independent sources suggest interpopulation variation and point to a high incidence of the disorder among the Roma. Galactokinase deficiency has been reported in Roma patients from Switzerland, Austria, and Italy (Gitzelmann 1967; Thalhammer et al. 1968; Bolgiani et al. 1984). All patients detected by the neonatal screening in Bulgaria (Kalaydjieva and Kremensky 1992) and 10 of the 21 galactokinase-deficient subjects studied by Gitzelmann (1987) are of Romani descent. Our previous studies of genetically isolated Romani populations (Kalaydjieva et al. 1996; Angelicheva et al. 1999) lead us to predict that galactoki-

Received February 22, 1999; accepted for publication August 18, 1999; electronically published October 11, 1999.

Address for correspondence and reprints: Dr. L. Kalaydjieva, Centre for Human Genetics, Edith Cowan University, Joondalup Campus, Perth WA 6027, Australia. E-mail: L.Kalaydjieva@cowan.edu.au

\*Current affiliation: Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut I de la Vida, Universitat Pompeu Fabra, Barcelona.

†Current affiliation: Department of Biology, Bogazici University, Istanbul.

© 1999 by The American Society of Human Genetics. All rights reserved. 0002-9297/1999/6505-0012\$02.00

**Table 1****Newborn Screening Data on the Birth Incidence of Galactokinase Deficiency**

Source	Tested	Detected	Incidence	Reference
“World” (Europe, USA, Asia) (1980)	6,000,000	6	1:1,000,000	Levy (1980)
Switzerland (1997)	2,200,000	1	1:2,200,000	Gitzelmann and Steinmann, unpublished data
Germany (1980)	1,100,000	7	1:157,000	Gitzelmann and Hansen (1980)
Austria (1980)	920,000	6	1:153,000	Gitzelmann and Hansen (1980)
Bulgaria (1988)	726,678	14	1:52,000	Kalaydjieva and Kremensky, unpublished data

nase deficiency in the Roma is genetically homogeneous and is caused by a private founder mutation.

In view of its low efficiency in detecting classic galactosemia, many mass-screening programs have discontinued testing for blood galactose. Instead, such testing has become part of the battery of selective-screening tests performed under clinical indications. The lack of early dramatic signs of a metabolic disorder in galactokinase deficiency implies that affected children will not attract medical attention and that their disorder will only be diagnosed after the development of lens opacities and blindness. Therefore, the identification of the founder galactokinase mutation would make possible the incorporation of a simple diagnostic test in the mass-screening programs of countries with a sizeable Roma population and, thus, would allow the early diagnosis and dietary treatment of a common cause of disability.

Since the cloning of the galactokinase gene GK1 (Stambolian et al. 1995), only two disease-causing mutations, V32M and Q80X, have been described (Stambolian et al. 1995), and conclusive data on the role of the gene in galactokinase deficiency are still missing. Here we present the results of refined genetic and physical mapping showing that GK1 is the gene responsible for galactokinase deficiency in the Roma. Sequencing analysis of the gene has revealed one single-nucleotide variation that results in an amino-acid substitution in the galactokinase protein and that is likely to be the founder Romani mutation. In view of the ethnic distribution of galactokinase deficiency, this may also be the major defect leading to this disorder worldwide.

## Subjects and Methods

### *Patients and Families with Galactokinase Deficiency*

The study included eight galactokinase-deficient Romani children from Bulgaria. The clinical and biochemical information on the patients is shown in table 2. Six were identified as being hypergalactosemic by the neonatal screening program and subsequently were shown, by means of enzyme analysis, to have galactokinase deficiency. In addition, two affected relatives (4-7 and 4-1), born either before the initiation or after the discontinuation of the hypergalactosemia screening, were identified in the family shown in fig. 1. In the same kindred, individual M.K.N. (4-5 in fig. 1), who is the mother of

an affected male, reported a history of juvenile cataracts, for which she had undergone surgery. The galactokinase activity of this individual, determined as part of the present study, was 0.23  $\mu\text{mol/h/ml}$  RBC. Enzyme activity has also been determined in individuals 4-10 (the mother of patient 4-11) and 4-13 (the maternal grandmother of patient 4-3). It was reduced in individual 4-10 (0.26  $\mu\text{mol/h/ml}$  RBC) and was within the normal range in individual 4-13 (0.83  $\mu\text{mol/hr/ml}$  RBC).

The patients belong to six unrelated families that originate from three different Romani groups: Thracian Tinkers, Kalderas, and Rudari. Despite their current social divergence, these groups have a common history of migrations in the northern part of the Balkan peninsula and a strong linguistic influence from Rumanian. The overall number of family members included in the study is 41.

The study complies with the ethical guidelines of the institutions involved. Informed consent was obtained from all adult subjects included in the study and from the parents of the children investigated.

### *Population Screening*

The population frequency of the mutation was analyzed in a sample of 194 anonymized adult controls with no family history of galactokinase deficiency: 130 from the Kalderas and Rudari Romani groups and 64 Bulgarians.

### *Linkage Analysis*

The genetic linkage study of the affected families started with the long arm of chromosome 17. This choice was based on the fact that the candidate gene, GK1, where mutations had been identified previously in two patients with galactokinase deficiency (Stambolian et al. 1995), had been roughly mapped, by use of FISH, to 17q24.

DNA was extracted from venous blood as described by Miller (1988). The six families were genotyped for 20 microsatellite markers spanning  $\sim 47$  cM. Microsatellites D17S808, D17S949, D17S802, D17S794, and D17S929 from the ABI mapping panel were used in the initial scan, which aimed at the exclusion of linkage to this region or roughly positioning the gene. These markers were PCR amplified using the ABI fluorescently la-

**Table 2****Clinical Information on the Galactokinase-Deficient Patients Detected by the Newborn Screening**

Patient	Year of Birth	Blood Galactose Screening Value (mg/dl)	Initiation of Treatment	Jaundice, Vomiting, Infection	<sup>a</sup> Serum GOT (IU/L)	<sup>a</sup> Serum GPT (IU/L)	<sup>b</sup> Galactokinase ( $\mu\text{mol/h/ml RBC}$ )	<sup>c</sup> Gal-1-P Uridyltransferase ( $\mu\text{mol/h/g Hb}$ )	Cataracts
D.P.H.	1981	110	45 d	No	20	8	.06	26.3	No
S.M.I.	1983	115	30 d	No	16	6	.02	35.6	No
S.H.G.	1984	90	36 d	No	11	2	.15	43.3	No
T.V.I.	1984	125	3.5 mo	No	15	9	0	33.4	Yes
P.P.V.	1985	99	20 d	No	12	12	.11	36.9	No
I.L.S.	1986	143	34 d	No	10	8	0	29.7	No

<sup>a</sup> Normal range for SGOT and SGPT < 18 IU/L.

<sup>b</sup> Normal range for galactokinase: 0.69–2.57 (2 wk–9 mo); 0.42–0.99 (>9 mo).

<sup>c</sup> Normal range for galactose-1-phosphate uridylyltransferase: 18.9–37.9.

beled primers, in a 5- $\mu\text{l}$  final volume containing 5–10 ng genomic DNA, 0.2 mM each primer, 0.25 mM each dNTP, 1.5 mM  $\text{MgCl}_2$ , 1  $\times$  PCR buffer (PE Biosystems N808-0010), and 0.2 U of *AmpliTaq*<sup>TM</sup> DNA polymerase (PE Biosystems N808-0101). The PCR products were length separated on an ABI 373 XL DNA sequencer in a 24-cm well-to-read 6% 19:1 acrylamide/bisacrylamide gel at 1400 V for 7 h. Polymorphic alleles were analyzed using ABI Genotyper<sup>TM</sup> (Version 2) software.

Additional markers used subsequently to create a denser map of the relevant interval were selected from the genetic maps in the public-domain databases and PCR amplified using the primer sequences indicated in the Genome Database. Radioactive labeling was done through incorporation of  $\alpha$ [<sup>32</sup>P]-dCTP in the product during PCR amplification. The PCR products were electrophoresed on 6% 19:1 acrylamide/bisacrylamide gels at 1400 V for 2–3 h. The gels were dried and autoradiographed for 12–48 h. Alleles were scored manually.

Linkage analysis was conducted under an autosomal recessive model with 100% penetrance with two different gene frequencies (0.01 and 0.05). Two-point analysis was performed using FASTLINK (Lathrop and Lalouel 1984) and multipoint analysis was done with GENE-HUNTER (Kruglyak et al. 1995). In the linkage analysis, the affection status of M.K.N. (individual 4-5 in the pedigree in fig. 1) was considered unknown. Haplotypes were constructed manually and were inspected for overlapping regions of homozygosity and for historical recombinations.

#### Radiation-Hybrid Mapping

The physical position of the GK1 gene, relative to the region of linkage, was established through radiation-hybrid mapping using the Stanford TNG panel. The analyses included a 293-bp sequence representing the GK1 gene (nucleotide positions 2071–2373, GenBank accession number L76927) and microsatellites D17S1839, D17S1603, and D17S801. PCR amplification of the microsatellite markers was performed using the primer sequences provided in the Genome Database.

The primers used for the GK1 fragment were 5'-CCTA-GAAACAGTTGCTAGGC-3' and 5'-ATAGTAGAAGC-TGGGACCAC-3'; cycling was in a touch-down regime with denaturing at 94°C for 30 s, elongation at 72°C for 30 s, and annealing from 63°C to 55°C, decreasing at 0.5° increments for the 16 initial cycles, followed by 20 cycles at 55°C for 30 s.

The results were analyzed using the software package "RH2PT," version 3.0 (September 1996), written by Elizabeth Hauser and Michael Boehnke.

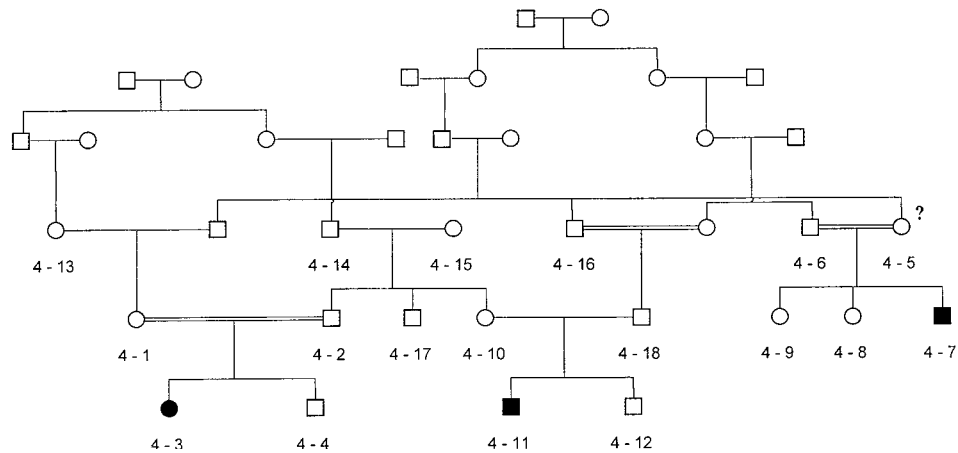
#### Sequence Analysis of the GK1 Gene

The analysis included the entire coding sequence of GK1, along with 100–150 bp of the flanking introns and the 5' UTR. Direct sequencing of both strands was performed after PCR amplification of genomic DNA from galactokinase-deficient individuals and control samples from a different population. The primer sequences are shown in table 3. PCR reactions were done with 30 ng total genomic DNA, 7 pmol each primer, 0.2 mM each dNTP, 1.5 mM  $\text{MgCl}_2$ , 5%–15% DMSO, 1  $\times$  PCR buffer (PE Biosystems N808-0010), and 0.5 U *AmpliTaq* (PE Biosystems N808-0101) in a final volume of 25  $\mu\text{l}$ . The PCR products were purified using Qiagen<sup>TM</sup> columns, according to the instructions of the manufacturer.

Sequencing reactions were performed using the ABI Prism<sup>TM</sup> Dye Terminator cycle sequencing kit and were run on an ABI 373 XL DNA sequencer. Alignment to the published genomic or cDNA sequence of GK1 was done using Sequence Navigator<sup>TM</sup> (PE Biosystems).

#### Restriction Assay for the Detection of the Mutation

A 272-bp DNA fragment spanning the region between nucleotides 478 and 749 in the GK1 sequence (GenBank accession number L76927) was amplified from genomic DNA by use of primers 5'-CGTCATGGCTGCTTT-GAGAC-3' and 5'-TAGAGGGCGCCCTTCTTGAG-3'. The PCR mix contained 30 ng total genomic DNA, 7 pmol each primer, 0.2 mM of each dNTP, 1.5 mM



**Figure 1** Romani family with galactokinase deficiency. The proband P.P.V. (individual 4-3) was detected by the neonatal screening program, whereas the two affected cousins, K.N.S (individual 4-7) and K.P.T. (individual 4-11), were born before the initiation of the screening and after the discontinuation of its hypergalactosemia component, respectively. Individual 4-5 reported a history of infantile cataracts and was found to be homozygous for the P28T mutation; her erythrocyte galactokinase activity was in the intermediate range.

MgCl<sub>2</sub>, 5% DMSO, 1 × PCR buffer (Perkin Elmer N808-0010), and 0.5 U AmpliTaq (PE Biosystems N808-0101) in a 20-ml total volume. PCR cycling was performed at initial denaturing at 94°C for 5 min; 15 cycles at a touch-down regime of denaturing at 94°C for 20 s, annealing from 60°C to 55°C (0.5°C increments) for 60 s, and extension at 72°C for 30 s; 15 cycles of 94°C for 20 s, 55°C for 60 s, and 72°C for 30 s; and a final extension step at 72°C for 7 min. The PCR product was ethanol precipitated, resuspended in 20 μl of dH<sub>2</sub>O, and then cut with *Ava* I. The restriction fragments were separated by electrophoresis on a 2.5% agarose gel and were visualized by ethidium-bromide staining.

#### Secondary Structure–Prediction Analysis of the Wild-Type and Mutant Galactokinase Proteins

In order to relate P28T and a known mutation, V32M (Stambolian et al. 1995), to conformational changes of functional significance, a secondary structure–prediction analysis was performed for the complete amino acid sequence (392 amino acids) of the wild-type and mutant proteins. The prediction analysis was performed by means of the PHDsec algorithm (Rost et al. 1994) on an 8-processor SGI Origin 2000 platform (dodo.cpmc.columbia.edu).

## Results

### Linkage of the Disease Gene to 17q24

Linkage analysis (table 4) excluded most of the long arm of chromosome 17. Positive LOD scores were obtained for the region 17q24-q25. The maximum two-point LOD score of 4.382 was obtained for D17S801

at recombination fraction ( $\theta$ ) 0. The results were not substantially affected by changing the gene frequency. Multipoint results peaked in the interval D17S1839/ D17S1603/ D17S801.

Haplotype analysis revealed a conserved haplotype, spanning a distance of ~6 cM, that was found in most disease chromosomes (fig. 2). The highly conserved region of homozygosity shared by all mutant chromosomes was defined by markers D17S1839 and D17S1603, which, according to the Genethon map, are located <1 cM apart. Allele 2 of marker D17S1839 is in strong linkage disequilibrium with the disease, whereas allele 1 of D17S1603, which is found in all disease chromosomes, is also the predominant normal allele—hence the low LOD score and TDT *P* value obtained for this marker (fig. 2).

These results point to 17q24 as the region containing the disease gene in the Romani families. The data coincide with the cytogenetic location of the GK1 gene, as determined by FISH (Stambolian et al. 1995). The conserved polymorphic haplotypes found in disease chromosomes from unrelated families suggest that affected individuals are homozygous for the same founder mutation, which, on the basis of linkage disequilibrium analysis, should be located in close proximity to marker D17S1839.

### Physical Mapping of GK1 Relative to the Linkage Region

Radiation-hybrid mapping was used to determine the physical position of GK1 relative to D17S1839 and D17S1603, the two markers defining the conserved homozygosity region, and to D17S801, where the highest

**Table 3**  
**PCR Primer Sequences Used in the Sequence Analysis of the GK1 Gene**

Sequence Name	Primer Sequence	Nucleotide Positions
5' UTR (F)	5'-CAGCTCCATTGCTCTGG-3'	74-93
5' UTR (R)	5'-CAGCAGCTGGATTCCCACG-3'	299-317
1 (F)	5'-CCCAGCATCCCCGCCGAC-3'	419-438
1 (R)	5'-GACAGGCTGTCCCCACGT-3'	800-818
2 (F)	5'-ACTGTGGAGGCATCAGAACC-3'	1436-1455
2 (R)	5'-CACAGAGCCCATTCAATTTGTCTGA-3'	1781-1805
3 (F)	5'-CCTAGAAACAGTTGCTAGGC-3'	2071-2090
3 (R)	5'-GTGGTCCCAGCTTCTACTAT-3'	2354-2373
4 (F)	5'-AGTGCATTGAAGCCACTGC-3'	2408-2427
4 (R)	5'-CAAGCACACGCTTGGCCTCGT-3'	2686-2706
5 (F)	5'-AGCAGCTCCTGGGTGGAGTGT-3'	2647-2667
5 (R)	5'-CTCAGTGTGGCCTTGACCT-3'	2971-2990
6 (F)	5'-ATCACCGCCTGCTGGTCCTC-3'	6945-6964
7 (F)	5'-CTCATGGTGGAGAGCCACCG-3'	7146-7165
7 (R)	5'-CTCAAGCAGCCGATGGAGCC-3'	7530-7549
8 (F)	5'-ACGCCATGCGGCACATCCAG-3'	7394-7413
8 (R)	5'-CTCTGTGCCCGGTGCCATCTT-3'	7624-7644

two-point LOD score had been obtained. The high resolution of the TNG panel allowed the separation of these loci and the positioning of GK1, despite the close proximity between galactokinase and thymidine kinase. Evidence of linkage was obtained between GK1 and D17S1839, with a LOD score of 12.86 and breakage probability (PBR) of .1139. For the other two markers, D17S1603 and D17S801, the LOD scores obtained were, respectively, 1.45 and 0, and the PBR values were .699 and 1. These findings indicate that GK1 is <300 kb from marker D17S1839.

#### *The GK1 Mutation in Galactokinase-Deficient Patients*

The analysis of the entire coding sequence of GK1, the flanking intronic sequences, and the 5' UTR has identified one single mutation. This is a C→A transversion in exon 1 that occurs either at nucleotide position 563, according to the GenBank numbering (accession number L76927), or at nucleotide position 82, according to the Bergsma et al. (1996) nucleotide numbering. The nucleotide substitution gives rise to a missense mutation—namely, the replacement of proline by threonine at amino acid position 28 (P28T). P28T was present in the homozygous state in affected individuals. The control subjects displayed the published wild-type sequence. No other differences were found between the galactokinase-deficient individuals, the controls, and the published GK1 sequence, as regards the remaining part of the coding region, the intronic sequences flanking the exons, and the 5' UTR.

The P28T mutation abolishes an Ava I-restriction site and thus can be detected through PCR-based restriction analysis. The Ava I-restriction assay was performed on all family members, including grandparents and unaf-

ected siblings. The restriction pattern coincided 100% with the carrier status of unaffected family members predicted on the basis of haplotype analysis. It also coincided with the measured levels of galactokinase activity in individuals 4-10 (carrier, decreased activity) and 4-13 (noncarrier, normal activity). Individual 4-5, who is the mother of an affected patient and who developed cataracts in early childhood, was found to be homozygous for both the disease haplotype and the mutation.

#### *Modeling of the Secondary Structure of the Mutant P28T and V32M Proteins*

Secondary-structure prediction analysis was performed in comparison with the wild-type protein and with the protein containing the known amino acid substitution V32M (Stambolian et al. 1995). The results indicate that both mutations induce topology changes in the loop-beta turn motif formed by the wild-type sequence AEP ELAVSA, immediately preceding the galactokinase signature domain GRVNLIGEHTDY (fig. 3). These differences are clearly visible on the detailed prediction maps (SUBsec). V32M causes an extension of the loop with one amino acid (G25) and disturbs the succeeding beta turn (LAV). P28T acts in a similar way, by extending and shifting the loop to F24 and by disturbing the succeeding beta turn.

#### *Population Screening for P28T*

We have used the Ava I assay to estimate the frequency of the mutation among a total of 194 unrelated adult control individuals. In the control group of 130 individuals originating from the Kalderas and Rudari Romani groups, we detected six heterozygotes for P28T.

**Table 4**  
Two-Point LOD Scores for Markers on Chromosome 17q

MARKERS	LOD SCORE AT $\theta =$									
	0	.05	.1	.15	.2	.25	.3	.35	.4	
D17S808	-1.27	-.38	-.14	-.04	-.006	.006	.009	.009	.008	
D17S807	-4.44	-.59	-.01	.22	.30	.30	.25	.18	.10	
D17S840	-1.09	.16	.25	.21	.17	.13	.09	.007	0	
D17S949	∞	-1.50	-.7	-.33	-.15	-.06	-.01	.004	.009	
D17S1797	2.43	2.08	1.73	1.39	1.06	.76	.50	.29	.04	
D17S1862	-1.7	.05	.32	.39	.39	.35	.30	.23	.16	
D17S1352	3.02	2.69	2.28	1.86	1.44	1.06	.73	.45	.24	
D17S1807	3.22	2.69	2.18	1.72	1.30	.95	.65	.41	.23	
D17S1602	2.32	2.15	1.86	1.52	1.18	.86	.59	.37	.21	
D17S1864	2.87	2.51	2.08	1.63	1.22	.86	.56	.34	.18	
D17S929	3.29	2.94	2.50	2.05	1.61	1.19	.83	.52	.28	
D17S1839	3.02	2.53	2.06	1.62	1.23	.90	.63	.41	.24	
D17S1603	1.34	1.16	.97	.79	.61	.45	.32	.20	.11	
D17S801	<b>4.38</b>	3.70	3.04	2.41	1.83	1.32	.88	.54	.27	
D17S1817	3.62	3.05	2.52	2.00	1.54	1.13	.78	.49	.28	
D17S802	.59	1.47	1.40	1.22	1.01	.79	.59	.41	.25	
D17S937	.28	.63	.58	.47	.34	.23	.14	.07	.03	
D17S1847	∞	-.3	-.1	0	.04	.05	.04	.03	.02	
D17S1806	-1.70	-1.19	-.85	-.59	-.40	-.26	-.17	-.098	-.04	
D17S784	-1.48	-.85	-.55	-.38	-.27	-.197	-.137	-.085	-.039	
D17S928	-.96	-.70	-.49	-.31	-.18	-.097	-.046	-.018	-.005	

No carriers were identified among the 64 Bulgarian controls.

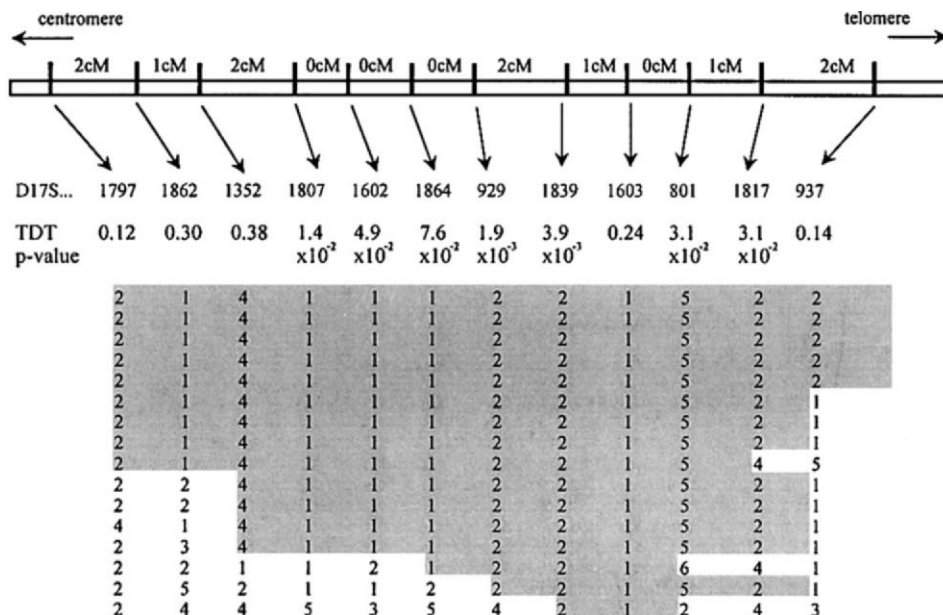
These results suggest that, in the Romani groups screened, the carrier rate for P28T is ~5%. By contrast, the mutation appears to be absent in control individuals from the non-Romani Bulgarian population.

## Discussion

The galactokinase-deficient individuals included in this study have been diagnosed, on days 4–5 after birth, as having hypergalactosemia. None of the patients have manifested clinical or biochemical symptoms suggestive of classic galactosemia; however, three of the eight affected individuals have developed cataracts. T.V.I. was diagnosed by the screening, but treatment was delayed; K.N.S. and K.P.T. were born before the initiation and after the discontinuation, respectively, of the hypergalactosemia screening. Galactokinase activity in these patients, which was measured during the first 2 mo of life, ranged from 0 to .15  $\mu\text{mol/h/ml}$  RBC, with a mean of 0.056, whereas the normal range for the same age group is .69–2.57, with a mean value of 1.29  $\mu\text{mol/h/ml}$  RBC. Diagnostic measurements of galactokinase activity are routinely performed in red blood cells, where the activity of the enzyme is ~50 times lower than that in liver (Shin-Buehring et al. 1977). The assay relies on the use of radioactively labeled substrate, and minor fluctuations in the measured activity can be caused by the quality of substrate purification as well as by reticulocyte counts.

The test is, therefore, highly reliable for the diagnosis of galactokinase deficiency but is not sensitive enough to discriminate between complete deficiency and very low residual activity. Our biochemical and clinical observations are sufficient to conclude that the Romani galactokinase mutation leads to a drastic reduction in neonatal enzyme activity (with the mean value in affected individuals constituting <5% of the mean in controls of the same age) and that, in untreated homozygotes, the reduced activity results in the development of cataracts during the first 3 mo of life.

The size and structure of the families included in this study, and the genetic homogeneity and recent founder effect observed in endogamous Romani groups, have allowed us to map the galactokinase-deficiency locus to 17q24, the same region that has been shown, by means of FISH, to contain GK1 (Stambolian et al. 1995). Radiation-hybrid mapping of the GK1 sequence, relative to the conserved region of homozygosity, confirmed that GK1 is located within this region. Both sets of results pointed to GK1 as the gene responsible for galactokinase deficiency in Bulgarian Gypsies. Haplotype analysis has indicated that, like other single-gene disorders in this population (Kalaydjieva et al. 1996; Angelicheva et al. 1999), galactokinase deficiency is caused by a single founder mutation shared by all affected individuals. Sequence analysis of GK1 identified a nucleotide substitution in exon 1 of the gene, resulting in the replacement of the conserved proline residue at amino acid position 28 with threonine. The role of the P28T mutation in



**Figure 2** Marker haplotypes in the 17q24 region of disease chromosomes from Gypsy families with galactokinase deficiency. Radiation-hybrid mapping (high-resolution TNG panel) placed the GK1 gene in very close proximity to D17S1839.

causing the disease is supported by the following arguments (Cotton and Scriver 1998):

(1) The analysis of the entire coding region of GK1, the flanking intronic sequences, and the 5' UTR in affected individuals and controls has revealed a single change in the sequence of the patients' DNA—namely, the C→A transversion at nucleotide position 563 of the gene (accession number L76927).

(2) Family analysis revealed 100% cosegregation with the disease phenotype. All affected individuals were found to be homozygous for the P28T mutation. P28T also occurred in the heterozygous state among unaffected siblings and followed the predictions of carrier state, on the basis of the study of 17q24 polymorphic haplotypes.

(3) The mutation was not detected among 128 Bulgarian control chromosomes.

(4) The screening of 130 individuals originating from the Gypsy groups to which the affected families belong revealed a carrier rate of ~5%, which is in good agreement with the number of affected births detected by the mass newborn screening (see below).

(5) The nature of the amino acid substitution, threonine for proline, suggests that changes in the structure and catalytic properties of the protein are likely to occur. A proline residue at this position of the galactokinase protein is conserved in evolution from *Escherichia coli* (Debouck et al. 1985), to mice (Mm Unigene EST database, AA255078) and humans. The mutation site is in close proximity upstream to the so-called “galacto-

kinase signature sequence.” It is also located very close to one of the two mutations described in the GK1 gene so far, a methionine-for-valine substitution at position 32 (Stambolian et al. 1995). Computer models of protein conformation show that V32M and P28T cause similar changes in the secondary structure of the protein domain immediately preceding the galactokinase signature. The role of the galactokinase signature domain and of the surrounding structures remains unknown; however, the fact that two of the three known mutations in GK1 are located within this region points to its functional significance.

The Romani population of Bulgaria—800,000 people in all—is a complex mosaic of a large number of endogamous groups. The families with galactokinase deficiency originate from three socially distinct but genetically related groups that, before migrating to Bulgaria during the last 100 years, were confined to slavery in the Wallachian Principalities, in the northern part of the Balkans. These groups, generally referred to as Vlax Roma, are believed to account for as much as 40%–50% of the Romani population of Bulgaria (Marushiakova and Popov 1997). Our preliminary estimates, on the basis of the analysis of 130 unrelated individuals from these groups, suggest a carrier rate of 4%–5% and an expected incidence of affected births of ~1:1,600 to ~1:2,500 in the high-risk population. Assuming that ~10% of the newborns in Bulgaria are of Romani origin and that 40%–50% of these belong to the Vlax groups where P28T is common would mean that, of the total of

**Galactokinase, normal sequence**

```

      .....1.....2.....3.....4.....5.....6
AA      |MAALRQPQVAELLAEARRAFREEFGAEP ELAVSAPGRVNLIGEHTDYNQGLVLPMALELM|
PHD sec |      HHHHHHHHHHHHHHHH      EEE  EEEEE      HHHH      |
subset: SUB sec |LLLLL.HHHHHHHHHHHHHH.LLLL.E.LLL.EEE.....LLLLLL.....|

```

**Galactokinase, V32M**

```

      .....1.....2.....3.....4.....5.....6
AA      |MAALRQPQVAELLAEARRAFREEFGAEP ELAMSAPGRVNLIGEHTDYNQGLVLPMALELM|
PHD sec |      HHHHHHHHHHHHHHHH      EE  EEEEE      HHHH      |
subset: SUB sec |LLLLLL.HHHHHHHHHHHHHH.LLLL...LL.EEE.....LLLLLL.....|

```

**Galactokinase, P28T**

```

      .....1.....2.....3.....4.....5.....6
AA      |MAALRQPQVAELLAEARRAFREEFGAETELAVSAPGRVNLIGEHTDYNQGLVLPMALELM|
PHD sec |      HHHHHHHHHHHHHHHH      EEE  EEEEE      HHHH      |
subset: SUB sec |LLLLL.HHHHHHHHHHHHHH.LLLL..EE.LLL.EEE.....LLLLLL.....|

```

**Figure 3** Secondary-structure prediction for the wild-type galactokinase protein and the V32M and P28T mutants, using the PHDsec algorithm. A selection of the first 60 amino acids is shown. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation >.69.

726,678 newborns screened for galactosemia, 30,000–40,000 belong to the high-risk groups. Based on the aforementioned carrier rate, one would predict between 12 and 25 affected births, which is in agreement with the actual number of 14 detected children with galactokinase deficiency. The Vlax Roma gave rise to the great Gypsy diaspora in the second half of the 19th century, after the abolition of slavery in Romania, and, currently, they account for a large proportion of the Romani minority in most countries. Thus, P28T may be an important cause of childhood blindness in the Roma worldwide and should be borne in mind in the diagnostic testing of any child with early development of cataracts.

Early investigations of the kinetic properties of rat-liver galactokinase (Cuatrecasas and Segal 1965) have shown that activity is high after birth and gradually declines with age and that the newborn and adult forms differ in terms of Michaelis constant, maximum velocity, and inhibition by galactose and galactose-1-phosphate. In human red blood cells, galactokinase activity is higher in small infants than in older children and adults (Sitzmann et al. 1973; also see table 2) and age differences in the kinetic properties of the enzyme have also been demonstrated (Mathai and Beutler 1967). The existence of two forms of galactokinase may help to explain why an adult female in our study (individual 4-5; see fig. 1), who has a history of juvenile cataracts and who is homozygous for P28T, had intermediate galactokinase activity (0.23  $\mu\text{mol/hr/ml}$  RBC). Perhaps related to the same issue is the discrepancy between the frequency of affected births observed in newborn screening programs (table 1) and that predicted on the basis of erythrocyte galactokinase activity in adult controls. Two studies (Gitzelmann 1967; Mayes and Guthrie 1968) have

found enzyme activity in the heterozygote range in 1:50–1:100 individuals. The predicted frequency of homozygotes, 1:10,000–1:40,000, is thus 50–100 times higher than that actually observed.

These observations may be due to age-related differences in the expression of the GK1-encoded protein, in terms of regulation and levels of expression, alternative splicing, or differences in posttranslational modification. They may also point to a possible sequential expression of two forms of galactokinase encoded by different genes. Whatever the mechanism of generating a second (adult) form of galactokinase, its existence would have important implications for the duration of dietary treatment in patients detected by newborn screening and, possibly, for understanding the genetic mechanisms of susceptibility to presenile cataracts.

## Acknowledgments

This study was supported by a grant from Edith Cowan University. We thank Drs. A. Akkari, A. Murch, and L. Heather for helpful discussions and D. Chandler, I. Hassanova, and O. Kamenov for technical assistance. Support from Bogazici University (to S.O.) is gratefully acknowledged.

## Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim> (for galactokinase deficiency [MIM 230200] and galactosemia due to galactose-1-phosphate uridylyltransferase deficiency [MIM 230400])



Généthon, <http://www.genethon.fr> (for the genetic map of chromosome 17q).  
 Genome Database, <http://gdbwww.gdb.org> (for the sequence of PCR primers for polymorphic microsatellites on 17q).  
 GenBank, <http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html> (for the complete sequence of the GK1 gene [accession number L76927])

## References

- Angelicheva D, Turnev I, Dye D, Chandler D, Thomas PK, Kalaydjieva L (1999) Congenital cataracts facial dysmorphism neuropathy syndrome: a novel developmental disorder in Gypsies maps to 18q. *Eur J Hum Genet* 7:560–566
- Bergsma DJ, Ai Y, Skach WR, Nesburn K, Anoaia E, Van Horn S, Stambolian D (1996) Fine structure of the human galactokinase GALK1 gene. *Genome Res* 6:980–985
- Bolgiani MP, Gallenca M, Barocelli PC (1984) Su un caso di galattosemia da deficit di galattochinasi. *Pediatr Med Chir* 6:333–336
- Cotton RGH, Sriver CR (1998) Proof of “disease causing” mutation. *Hum Mutat* 12:1–3
- Cuatrecasas P, Segal S (1965) Mammalian galactokinase, developmental and adaptive characteristics in the rat liver. *J Biol Chem* 240:2382–2388
- Debouck C, Riccio A, Schumperli D, McKenney K, Jeffers J, Hughes C, Rosenberg M (1985) Structure of the galactokinase gene of *E. coli*, the last (?) gene of the *gal* operon. *Nucleic Acids Res* 13:1841–1853
- Gitzelmann R (1965) Deficiency of erythrocyte galactokinase in a patient with galactose diabetes. *Lancet* 2:670–671
- (1967) Hereditary galactokinase deficiency, a newly recognized cause of juvenile cataracts. *Pediatr Res* 1:14–23
- (1987) Hereditary galactokinase deficiency: this week’s citation classic. *Curr Contents* 30:14
- Gitzelmann R, Hansen RG (1980) Galactose metabolism, hereditary defects and their clinical significance. In: Burman D, Holton JB, Pennock CA (eds) *Inherited disorders of carbohydrate metabolism*. Lancaster, UK, MTP Press Limited, pp 61–101
- Kalaydjieva L, Hallmayer J, Chandler D, Savov A, Nikolova A, Angelicheva D, King RHM, et al (1996) Gene mapping in Gypsies identifies a novel demyelinating neuropathy on chromosome 8q. *Nat Genet* 14:214–218
- Kalaydjieva L, Kremensky I (1992) Screening for phenylketonuria in a totalitarian state. *J Med Genet* 29:656–658
- Kruglyak L, Daly MJ, Lander E (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. *Am J Hum Genet* 56:519–527
- Lathrop GM, Lalouel J-M (1984) Easy calculations of LOD scores and genetic risks on small computers. *Am J Hum Genet* 36:460–465
- Levy HL (1980) Screening for galactosemia. In: Burman D, Holton JB, Pennock CA (eds) *Inherited disorders of carbohydrate metabolism*. Lancaster, UK, MTP Press Limited, pp 133
- Marushiakova E, Popov V (1997) *Gypsies (Roma) in Bulgaria*. Peter Lang, Frankfurt am Main
- Mathai CK, Beutler E (1967) Biochemical characteristics of galactokinase from adult and fetal human red cells. *Enzymologia* 33:224–230
- Mayes JS, Guthrie R (1968) Detection of heterozygotes for galactokinase deficiency in a human population. *Biochem Genet* 2:219–230
- Miller SA (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
- Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19:55–72
- Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ (1995) Cloning the galactokinase cDNA and identification of mutations in two families with cataracts. *Nat Genet* 10:307–317
- Stambolian D, Scarpino-Myers V, Hodes B, Harris H (1986) Cataracts in patients heterozygous for galactokinase deficiency. *Invest Ophthal Vis Sci* 27:429–433
- Shin-Buehring YS, Beier T, Tan A, Osang M, Schaub J (1977) The activity of galactose-1-phosphate uridylyltransferase and galactokinase in human fetal organs. *Pediatr Res* 11:1045–1051
- Sitzmann FC, Kaloud H, Teubl I, Müller HE (1973) Die aktivität des enzymes galaktokinase in den erythrozyten gesunder kinder und erwachsener. *Klin Padiatr* 185:444–448
- Thalhammer O, Gitzelmann R, Pantlitschko M (1968) Hypergalactosemia and galactosuria due to galactokinase deficiency in a newborn. *Pediatrics* 42:441–445