

ARTICLE

Genetic analysis of autosomal recessive osteopetrosis in Chuvashiya: the unique splice site mutation in *TCIRG1* gene spread by the founder effect

Elena A Bliznetz*,¹, Svetlana M Tverskaya¹, Rena A Zinchenko¹, Anna V Abrukova², Ekaterina N Savaskina², Maxim V Nikulin¹, Alexander G Kirillov², Evgeny K Ginter¹ and Alexander V Polyakov¹

¹Russian Academy of Medical Sciences, Research Center for Medical Genetic, Moscow, Russia; ²Ministry of Health and Social Development of Chuvash Republic, Presidential Perinatal Center, Chuvash Republic, Cheboksari, Moscow, Russia

The rare malignant disorder autosomal recessive osteopetrosis (OPTB) is one of the most prevalent autosomal recessive diseases in the Chuvash Republic of Russia. The purpose of this study was to determine the underlying molecular cause of osteopetrosis in Chuvashiya and to reveal the factors causing the unusual high frequency of the disease in this region. Having assumed a founder effect, we performed linkage disequilibrium (LD) mapping of the OPTB locus at the *TCIRG1* region and found a unique splice site mutation c.807 + 5G > A in all Chuvashian OPTB patients studied. We then analyzed the mutational change in mRNA and detected an intron insertion within the mutant transcript, resulting in a frameshift and premature stop-codon formation (*p.Leu271AspfsX231*). A decreased expression of the mutant transcript was also detected, which may have been the result of nonsense-mediated decay. Real-time qPCR and MLPA[®] melting curve analysis-based systems were designed and used for c.807 + 5G > A mutation screening. In addition to analyzing the gene frequency in Chuvashiya, we also estimated three other populations in the Volga-Ural region (Mari, Udmurt and Bashkir). We found a 1.68% prevalence in Chuvashiya (calculated disease frequency, 1/3500 newborns) and a 0.84% in the Mari population (1/14 000 newborns). The haplotype analysis revealed that all OPTB cases in Chuvashians and Marians originated from a single mutational event and the age of the mutation in Chuvashians was estimated to be approximately 890 years.

European Journal of Human Genetics (2009) 17, 664–672; doi:10.1038/ejhg.2008.234; published online 28 January 2009

Keywords: osteopetrosis; *TCIRG1*; founder effect

Introduction

Osteopetrosis is included in a heterogeneous group of inherited osteosclerosis and is characterized by a general

increase in bone density due to osteoclast deficiency (ie, the inability to resorb bone matrix). The most known forms of osteopetrosis are the infantile autosomal recessive form (OPTB), which is comprised of six subtypes, and the milder autosomal dominant form (OPTA), which is comprised of two subtypes (OMIM no. 259700). For each subtype, there is a certain causative gene. All forms of the disease are infrequent and found all over the world, the average disease frequency of the most rare malignant OPTB being 1/200 000–1/300 000.¹ The high frequency of

*Correspondence: Dr EA Bliznetz, Russian Academy of Medical Sciences, Research Center for Medical Genetic, Moskvorechye str. 1, Moscow 115478, Russia.

Tel: +7 495 3248110; Fax: +7 495 3248110;

E-mail: bliznetzelena@rambler.ru

Received 2 May 2008; revised 24 September 2008; accepted 29 October 2008; published online 28 January 2009

OPTB1, found in Costa-Rica, was 3.4/100 000^{2,3} and was caused by the founder effect.⁴ The Chuvash population of Russia is of great interest, as it has been recorded with the highest frequency of malignant OPTB (1/3879 newborns).⁵ It should be noted that apart from OPTB, two other diseases endemic to the Chuvash Republic are known, autosomal recessive erythrocytosis⁶ and autosomal recessive hypotrichosis.⁷ The presence of such a genetic differentiation in the Chuvash population presumably results from the long reproductive isolation of the population and the genetic drift.^{8,9}

Malignant OPTB is the most severe form of osteopetrosis and is characterized by an early manifestation (severe anemia), a progressing clinical course (hydrocephalus, hepatosplenomegaly, blindness, deafness and so on) and early lethality. Three genes responsible for infantile malignant OPTB are known: *TCIRG1* (OPTB1), *CLCN7* (OPTB4) and *OSTM1* (OPTB5). These genes encode the proteins for the osteoclast-specific enzymatic system. These proteins are highly expressed in the ruffled border of mature osteoclasts and directly participate in the bone dissolution process.^{10,11} Most cases of OPTB are due to various mutations in the *TCIRG1* gene,^{4,12} which is mapped in 11q13.4–13.5 and encoding the osteoclast-specific α_3 subunit of the vacuolar H^+ -ATPase proton pump. The *CLCN7* gene, which is associated with 10–15% of OPTB cases,^{13–16} encodes the monomer for the dimer molecule of chloric channel 7.¹¹ *CLCN7* functions in the complex with *OSTM1* protein, which is suggested to be a β -subunit of *CLCN7*.¹⁷

In this study, a causal molecular defect for Chuvash OPTB is shown and possible mechanisms for the endemic nature of the disorder in Chuvashiya are considered.

Materials and methods

Study subjects

A total of 11 OPTB families, including 11 affected and 25 relatives from Chuvashiya, and one OPTB family with one affected and three relatives from Mari were analyzed. A population sample comprising 327 Chuvashians, 299 Marians, 396 Udmurts and 271 Bashkirs was also analyzed. All samples consisted of unrelated aborigines (or unrelated indigenous families) from the Chuvashiya, Mari, Udmurtiya and Bashkortostan Republics of Volga-Ural region of Russia, respectively.

DNA sample preparation

Genomic DNA was extracted from whole blood using a *DIAtom™ DNA Prep100* kit (Isogene Lab. Ltd., Russia) according to the manufacturer's instruction.

RNA and cDNA sample preparation

Total RNA was extracted from the skin fibroblast culture using *RNAgens™* Total RNA Isolation System kit (Promega,

USA) according to the manufacturer's instruction. To obtain cDNA, reverse transcriptase PCR was obtained using a *Promega™ Reverse Transcription System* kit, Technical Bulletin no. 099 (Promega) according to the manufacturer's instruction. To obtain the skin fibroblast culture, the skin biopsies were incubated in DMEM medium containing 10% embryonic bovine serum with a periodic change of medium.

Genetic markers

Ten microsatellite markers from chromosome 11q13.4–q13.5 covering a 12.25 Mb region around the *TCIRG1* gene as well as six intragenic SNP markers were studied. All markers with the exception of the CA repeat in AC034259 were selected using Sequence and Marshfield maps in the database on the NCBI Web site. The CA repeat in AC034259 was absent in the database and was found by analyzing the gene sequence. The allele designations correspond to the Human Genome Database nomenclature (http://www.genoscope.cns.fr/externe/gmap/Nature-1995/alleles/all_chrom11). Supplementary Tables S1 and S2 summarize the markers used, the primer sequences and reaction conditions.

The studies of microsatellites and SNPs were carried out using ALFP and RLFP analyses, respectively. An amplification of the DNA/cDNA fragments was performed by PCR using a mix composed of 0.1–1.0 μ g DNA/cDNA, 0.5 μ M primer pair (Litech or 'SYNTOL', Russia), 200 μ M each dNTP, 1–10 mM $MgCl_2$, 1 U Taq-DNA polymerase (BioMaster, Russia), PCR buffer (67 mM Tris-HCl, 16.6 mM $(NH_4)_2SO_4$, 0.01% Tween-20, pH 8.8) and mineral oil. The restriction reaction was carried out according to the manufacturer's instruction in 20 μ l total volume of the reaction mix composed of 50–200 ng/ μ l PCR product, 5 U endonuclease and 1 \times buffer. Restriction enzymes from SibEnzym (Russia), Fermentas (Lithuania) and Promega were used. The PCR/restriction fragments were assessed by polyacrylamide gel electrophoresis with subsequent ethidium bromide stain and ultraviolet visualization.

SSCP analysis and sequencing

The PCR was performed as described above and in Supplementary Table S3. To analyze SSCP, alkaline denaturation and the non-denaturing polyacrylamide gel electrophoresis with subsequent 1 \times SYBR[®] Gold nucleic acid gel stain (Invitrogen™ Molecular Probes™, USA) for ultraviolet visualization were carried out on the PCR fragments. The denaturation was performed for 30 min at 45°C in 15 μ l total volume of the reaction mixture, which was composed of 5 μ l PCR product, 1 μ l NaOH (5 M) and 1 μ l EDTA (0.5 M, pH 8.8). The vertical electrophoresis lasted 24 h at 25°C; the length of the gel was 20 sm. For sequencing, the PCR products were purified by isopropanol-ethanol precipitation. Sequencing was performed by using

the ABI Dye Terminator, version 1 (Applied Biosystems), in a final reaction volume of 10 μ l and run on a 3130 ABI genetic analyzer (Applied Biosystems). Sequences were aligned using Chromas version 2 (Technelysium). The reference cDNA sequence was obtained from GeneBank (NM_006019; the numbering used starts with nucleotide +1 for the A of the ATG translation-initiation codon).

Mutation identification of c.807 + 5G > A

To detect the c.807 + 5G > A mutation, RLFP analysis was used in conjunction with OPTB family analysis and estimating the frequency of heterozygous carriers of the mutation in the Chuvashian, Marian and Udmurt populations. Restriction reaction was performed as described above and in Supplementary Table S2.

To reveal the heterozygous mutation carriers in the Bashkir population, we applied the real-time qPCR method using the Real-Time PCR detecting system iCycler[®] ThermalCycler (Bio-Rad). The system simultaneously revealed two fragments, one containing the c.807 + 5G > A mutation and the other containing exon 6 of the *TBP* gene as a reference fragment. The PCR was performed in the mixture described above, adding 0.25 μ M of each the TaqMan probe ('SYNTOL', Russia) to PCR mixture under elongation with a fluorescent registration step at 65°C. The real-time qPCR-based system was used only as a qualitative test for the presence of the mutation in a sample because Pfaffl ratio¹⁸ was not different at the comparison of heterozygous with homozygous mutation carriers.

For the mutation screening in health Chuvashians, MLPA[®] (offered by MRC-Holland) with subsequent product registration using a melting curve analysis was applied. As a result of MLPA, two DNA fragments were obtained – mutation-bearing fragment and a shorter normal allele-bearing another fragment. Using melting curve analysis, distinctions of melting temperatures between these fragments were successfully discovered (Supplementary Figure S1). The oligonucleotide sequences for real-time qPCR and MLPA melting curve analysis-based systems are given in Supplementary Table S4. Ligase reaction was performed for 1–2 h at 57°C in 5 μ l total volume of the reaction mixture, which was composed of 0.1–1.0 μ g DNA, 0.16–10 fmol/ μ l of each of the ligation probes ('SYNTOL', Russia), 0.4 U Pfu-DNA-ligase, ligase buffer (20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM MgCl₂, 0.1% Igepal, 0.01 mM rATP and 1 mM DTT) and mineral oil. The PCR was performed in a 20 μ l total volume mixture composed of 5 μ l of the ligation mix, 0.5 μ M primer pair ('SYNTOL', Russia), 200 μ M of each dNTP, 1 U Taq-DNA polymerase (BioMaster), 0.2 \times SYBR GREEN I (Sigma, USA), PCR buffer (67 mM Tris-HCl, 16.6 mM (NH₄)₂SO₄, and 0.01% Tween-20, pH 8.8) and mineral oil. For melting curve analysis, an iCycler ThermalCycler (Bio-Rad) was used.

Linkage disequilibrium analysis and mutation dating

The statistical comparison of allele frequencies between disease and control chromosomes was based on a χ^2 test or Fisher's exact probability test for a 2 \times 2 contingency table with alleles classified into two groups, one for the associated allele and all others combined into a single group. VassarStats Statistical Computation Web Site was used for the calculations. A confidence interval (CI) for a proportion was calculated according to Wilson procedure without a correction for continuity (VassarStats Statistical Computation Web Site). To assess the degree of linkage disequilibrium (LD), we employed the formula described by Bengtsson and Thomson¹⁹: $\delta = (P_D - P_N)/(1 - P_N)$, where δ equals the measure of LD, P_D is the frequency of the associated alleles on the diseased chromosome and P_N is the frequency of the same allele on the control chromosome. A CI for δ was calculated as described earlier by Diaz *et al.*²⁰ To date the mutation, we used the algorithm first applied by Risch *et al.*²¹: $g = \log \delta / \log (1 - \theta)$, where g equals the time (in generations) back to the initial spreading of mutant chromosomes in population and θ is the recombination fraction between the disease locus and the marker. We also included a correction for growing population (Labuda *et al.*²²): $g + g_0$, where g_0 is defined as $-1/d \ln(\theta f_d)$, $f_d = e^d / (e^d - 1)$. Here, d is equal to the population growth rate.

Results

Clinical follow-up

Symptoms of OPTB in Chuvashiya were typical of malignant OPTB. At the age of 3 months, the 11 patients from Chuvashiya and the one patient from Mari had severe hypoplastic anemia, thrombocytopenia, hepatosplenomegaly and a model phenotype (macrocephaly, frontal bossing, swelling abdomen, failure to thrive and disem-bryogenetic stigmas). All patients presented with hydrocephalus, myatony, neuropsychic and psychomotor retardation, compromised vision and heavy nasal breathing. Pathologic fractures were absent in all patients. Radiological examination revealed a generalized sclerosis of all bones of the skeleton. Proximal metaphyses of femurs and tibias were irregularly shaped and, together with forearms, showed an absence of the medullar cavities. Femurs, tibias, forearms, phalanges and iliac crests showed a 'bone-within-bone' appearance and vertebral bodies showed a 'sandwich' appearance. All the patients were subjected to regular packed red cell transfusion and symptomatic therapy. According to the older data,⁵ 94% of the Chuvashian OPTB cases resulted in death at the age of 1–3 years. Only 6% of the patients exhibited longer survival. In our cohort of patients, four patients died within 2–3 years of life. One girl was 9 years old, and at the age of 8 years, she had a model phenotype, mild anemia, hepatosplenomegaly, optic atrophy, nystagmus, growth

and mental retardation, dental caries and milder sclerosis of bones in contrast with remaining patients (medullar cavities were partially showing). The life of the remaining patients studied is unknown.

Fine-scale mapping and mutation identification

As the founder effect was expected to cause the high frequency of OPTB in Chuvashiya, we performed an LD mapping of OPTB locus in the Chuvashian population. For this purpose, four polymorphic markers closely linked to the *TCIRG1* gene (as it was the most likely candidate gene) in eight OPTB families including eight affected and 16 relatives from Chuvashiya were analyzed. Two studied polymorphic markers were intragenic SNPs (dbSNP884826:A>G and dbSNP2075609:A>G) and two other markers were microsatellites (CA repeat in AC034259 and D11S987) flanking 0.27 Mb of the *TCIRG1* gene region. A complete LD between the disease locus and all studied markers was revealed, and all patients were found to be homozygous for the same haplotype (CA repeat in AC034259-dbSNP884826-dbSNP2075609-D11S987: 4-A-A-5). This was not revealed in the homozygous state in any relatives (Figure 1). The estimated association between the patients' haplotype alleles and the disease locus is shown in Table 1.

During the subsequent research on the coding regions and exon-intron junctions of *TCIRG1* gene in one OPTB patient, using SSCP analysis and sequencing, only sequence change was found. A unique substitution in intron 8, c.807 + 5G>A, was identified and later revealed in all eight patients at the homozygous state, whereas it was found at heterozygous state in all of the parents. We observed the same in three more OPTB Chuvashian

families, among whom three were affected in a total of nine relatives.

Mutation effect analysis

We preliminarily studied the influence of the c.807 + 5G>A mutation on splicing by using the NetGene2 program to predict splice sites. We found that the mutation results in a decrease in confidence of the normal donor splice site in intron 8 from 0.95 to 0.42 and an increase in confidence of an additional donor splice site 37 nt downstream from the normal one (from 0.93 to 0.95). RT-PCR analysis carried out on skin fibroblasts showed the presence of two bands in the parent heterozygous for the c.807 + 5G>A mutation, whereas individuals not bearing the mutation showed a smaller band (Figure 2a). During the direct sequencing of each of the two purified bands, apart from the normal transcript (Figure 2b), a longer splice variant bearing a 37 nt intron insertion was detected (Figure 2c). The identified insertion was the same as that predicted by the NetGene2 program. This intron insertion resulted in a frameshift and a premature termination at exon 13 (*p.Leu271AspfsX231*). Investigation of fibroblasts of patients failed as affected cells had not produced the colonies and had been lost.

We also analyzed the relative level of the mutant transcript in the cells. For this purpose, we performed an RLFP analysis of the intragenic SNP in an individual who was heterozygous for both the SNP and for the mutation *versus* that without the mutation for comparison. Here, the SNP allele linked to the mutation can be regarded as a quantitative marker of the mutant transcript. All intragenic SNPs (dbSNP3808973:C>T, dbSNP1129633:C>G, dbSNP2471829:G>A and dbSNP1047817:C>T) were analyzed in 38 heterozygous mutation carriers (23 members of 11 studied OPTB families, 11 revealed during the estimation of the OPTB gene frequency and four members of additional OPTB families for whom diagnostics were carried out during this study, see below) and in three Chuvashians not bearing the mutation. In total, seven (of 38 mutation carriers) and one (of three non-carriers) heterozygotes were revealed at the SNP in exon 4 dbSNP3808973:C>T. All revealed genotypes in the mutation carriers included allele C of the SNP, both in the heterozygous and in the homozygous states, that is the mutation was linked to allele C of the SNP. There were accessible fibroblasts from only one person who was heterozygous both for this SNP

Table 1 Estimation of the difference in frequencies of alleles in OPTB haplotype in patients' chromosomes (D) and control relatives' chromosomes not transmitted to the patients (C)

Marker locus	Allele	D, % n = 16	C, % n = 15	P-value ^a
CA repeat in AC034259	4	100	26	<0.0001
dbSNP884826:A>G	A	100	40	0.0002
dbSNP2075609:A>G	A	100	80	0.1012
D11S987	5	100	33	<0.0001

^aFisher's exact probability test.

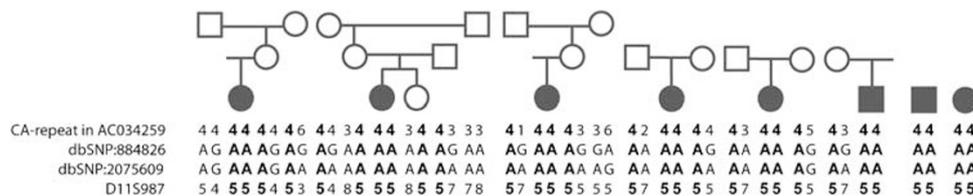


Figure 1 Haplotypes of chromosomes at polymorphic markers tightly linked to the *TCIRG1* gene in OPTB Chuvashian families.

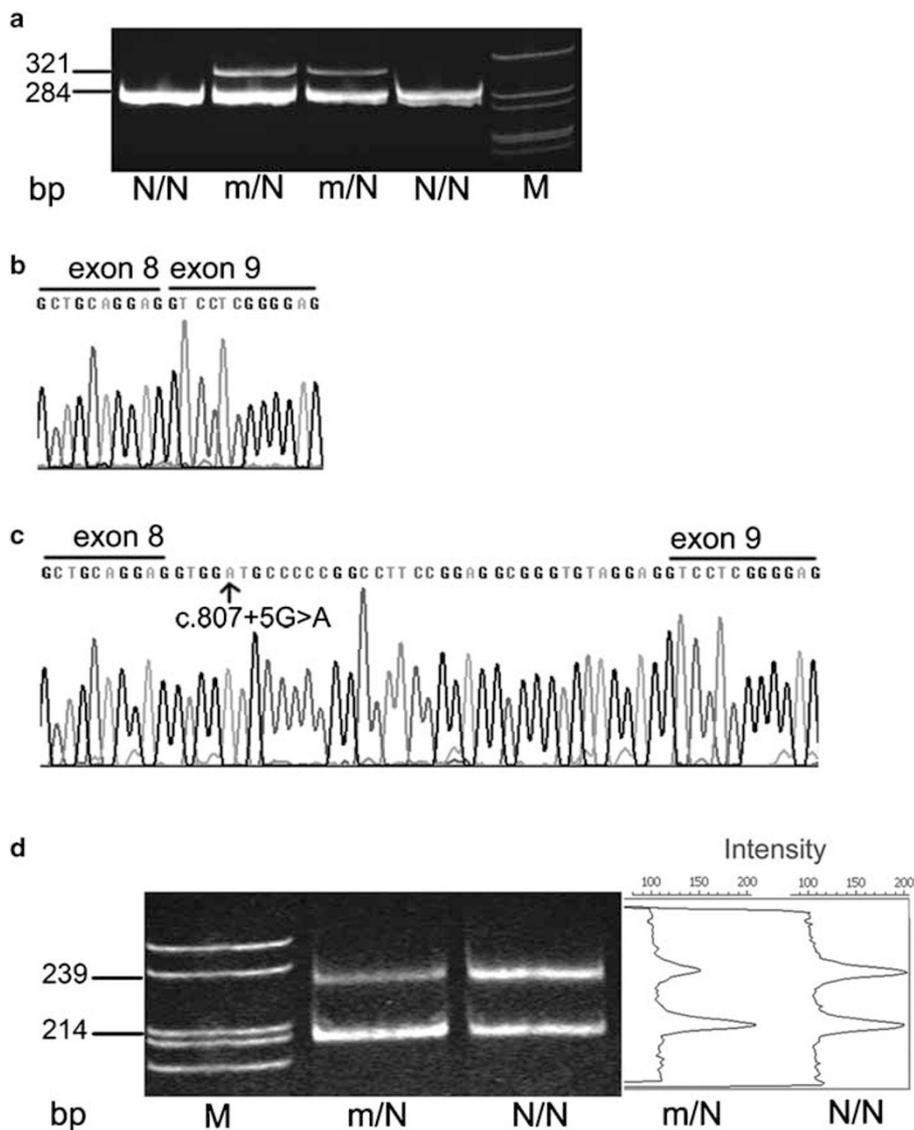


Figure 2 c.807 + 5G > A mutation in the *TCIRG1* gene effect analysis. (a) RT-PCR analysis of heterozygous parents, m/N; individuals not bearing the mutation, N/N; the fragments of chromatograms obtained by sequencing of purified smaller (b) and larger (c) bands observed during (a). (d) RFLP analysis of dbSNP3808973:C>T in individuals double heterozygous for SNP and the mutation, m/N, and those heterozygous for SNP and not bearing the mutation, N/N; on the right a fluorescence intensity of bands is shown in the graph (it is obtained by *Quantity One*[®] version 4.2 program (Bio-Rad)).

and the mutation, and from one person heterozygous for the SNP and not bearing the mutation. As a result of RFLP analysis, we detected a decreased fluorescence intensity of the band corresponding to the allele C at the SNP in the heterozygous mutation carrier (Figure 2d).

Decay of LD between the OPTB locus and the *TCIRG1* markers: dating the mutation in Chuvashiya

As described above, the complete LD between the OPTB locus in Chuvashiya and the four markers tightly linked to *TCIRG1* was revealed. We carried out an analysis for eight additional microsatellites in the gene region and have studied eight OPTB families described above and three new

families that included three OPTB patients and nine relatives. Allele frequencies in the mutation-bearing (*D*) and control population (*N*) chromosomes for the nine microsatellite loci are represented on a Marshfield map (D11S4076, D11S1883, D11S913 and D11S1889) upstream and (D11S987, D11S4113, D11S4136, D11S4139 and D11S4207) downstream to the *TCIRG1* gene and are illustrated in Supplementary Table S5 (available online at European Journal of Human Genetics web site). Results of χ^2 tests (or Fisher's test) for alleles showing the highest disequilibrium parameter δ are given in Table 2.

As expected, the farther a marker was from *TCIRG1*, the lower it showed LD with the OPTB locus. A significant

Table 2 Linkage disequilibrium analysis between OPTB locus in Chuvashians and microsatellites closed to the *TCIRG1* gene

Marker	Allele	P-value	χ^2	$\delta \pm 95\% CI$	<i>g</i>	<i>g</i> ₀	<i>g</i> + <i>g</i> ₀
D11S4076	1	0.2435	01.36	0.242 ± 0.385			
D11S1883	2	<0.0001	27.03	0.778 ± 0.200	10.18	9.35	19.53
D11S913	2	<0.0001	23.90	0.943 ± 0.111			
D11S1889	2	<0.0001	36.35	0.954 ± 0.088			
<i>TCIRG1</i>	mut						
D11S987	5	<0.0001	21.35	0.939 ± 0.119			
D11S4113	12	<0.0001	46.14	0.850 ± 0.159	30.45	16.05	46.50
D11S4136	4	0.0353	04.43	0.282 ± 0.272	30.08	7.02	37.10
D11S4139	6	0.0089	06.84	0.320 ± 0.254	20.76	5.88	26.64
D11S4207	15	0.1202 ^a		0.117 ± 0.174			
				32.4 ± 11.8 ^b			

^aP-value, Fisher's exact probability one-tailed test.

^bMean age in generations and standard deviation.

δ , the degree of linkage disequilibrium by Bengtsson and Thomson¹⁹; *g*, by Risch *et al*²¹; *g*₀, by Labuda *et al*.²² Estimating the age of the 807+5G>A mutation in *TCIRG1* gene in Chuvashians.

excess in the frequency of the '2' alleles at D11S1883, D11S913 and D11S1889, the '5' allele at D11S987, the '12' allele at D11S4113, the '4' allele at D11S4136 and the '6' allele at D11S4139 on OPTB chromosomes *versus* controls was observed. The remaining two flanking markers, D11S4076 and D11S4207, did not show significant associations.

We were able to phase all mutation-bearing chromosomes into haplotypes using other family members. As a result, full haplotypes for 23 mutation-bearing chromosomes (22 patients' chromosomes and one chromosome not transmitted to a patient from an unaffected member of OPTB family) for nine microsatellites are presented in Table 3. Of these, 23/23 = 100% carry either a '2-2-2' haplotype (or its decay derivative) at D11S1883-D11S913-D11S1889, or a '5-12' haplotype (or its decay derivative) at D11S987-D11S4113, or both. Moreover, 8/23 = 34.8% of all the disease chromosomes carried a '5-12-4-6' haplotype (or its decay derivative) at D11S987-D11S4113-D11S4136-D11S4139 and 7/23 = 30.4% carried a '5-12-9-6' haplotype (or its decay derivative) at one. A significant association was found with only the '4' allele at D11S4136. Given the strong association with seven of the markers, we were able to determine the probable 'founder haplotype' as '2-2-2-5-12-4-6' at D11S1883-D11S913-D11S1889-D11S987-D11S4113-D11S4136-D11S4139.

The frequencies of various 'founder haplotype' decay derivatives were used to estimate the original time of the spreading of the c.807 + 5G>A mutation in Chuvashians. To accomplish this, the algorithm offered by Rish²¹ with *g*₀ correction for growing population²² was applied. To define the Chuvashian population growth rate, we used historical data from a census of the Chuvashian population of 1723–1989.²³ An average of 27.4 years was used as the generation length, corresponding to current data for the rural population of the Republic.²⁴ The expansion dynamics of the Chuvashian population is approximated by an exponential curve with the approximation reliability

$R^2 = 0.9952$, where growth rate $d = 0.2272$ (Supplementary Figure S2).

The obtained estimates of the age of the c.807 + 5G>A mutation in Chuvashians are shown in Table 2. We did not apply the data from D11S913, D11S1889 and D11S987 to estimate the time of the mutation because these markers are located in the same linked cluster on Marshfield map as the *TCIRG1* gene. Thus, the mean age is equal to 32.4 generations (SD = ± 11.8). Assuming a mean generation length of 27.4 years, the time since origin would be about 890 ± 320 years. Assuming an average year of birth of 2000 for all studied patients, we can then place the approximate year of the beginning of the spreading of the c.807 + 5G>A mutation in Chuvashians as 1110, with a range of 790–1430.

Historical data regarding the Chuvashian population suggest that it is constantly growing. Extrapolating the growing exponential curve for eight centuries, we attempted to estimate a population size at the moment of an origin for the mutation in the population. According to the graph in Supplementary Figure S2, at the beginning of the twelfth century, the population size reached approximately 1000 (0.1–10 000) persons, and the drift influence can be assumed to be accruing OPTB in Chuvashiya.

Estimating the OPTB gene frequency

Among 327 unaffected unrelated Chuvashians, we identified 11 heterozygous c.807 + 5G>A mutation carriers (the gene frequency is equal to 1.68% (95% CI 0.95 ÷ 2.96%)). Thus, the frequency of the occurrence of homozygotes is 0.000282, and the disease frequency is 1/3500 newborns. In 299 Marians, the gene frequency is equal 0.84% (95% CI 0.36 ÷ 1.93%) and the calculated disease frequency is 1/14 000 newborns. In 396 Udmurts and 271 Bashkirs, the c.807 + 5G>A mutation was not found. Furthermore, the mutation in the homozygous state was not found in either of the studied OPTB-free individuals.

Table 3 Haplotypes of mutation-bearing chromosomes at nine microsatellite loci

Marker	D11S4076	D11S1883	D11S913	D11S1889	TCIRG1	D11S987	D11S4113	D11S4136	D11S4139	D11S4207
cM	62.62	65.05	67.48	67.48	67.48	67.48	68.01	71.60	72.82	76.13
Mb	61.10	63.13	65.70	67.07	67.59	67.80	68.52	69.32	70.18	73.35
<i>Chuvashians' chromosome</i>										
1	8	2	2	2	m	5	12	4	6	15
2	8	2	2	2	m	5	12	4	6	15
3	2	2	2	2	m	5	12	4	6	8
4	1	2	2	2	m	5	12	4	6	14
5	1	2	2	2	m	5	12	4	1	2
6	1	2	2	2	m	5	12	4	3	5
7	1	2	2	2	m	5	12	4	17	14
8	3	2	2	2	m	5	12	4	13	11
9	1	2	2	2	m	5	12	9	6	5
10	1	2	2	2	m	5	12	9	6	5
11	1	2	2	2	m	5	12	9	6	14
12	6	2	2	2	m	5	12	9	6	16
13	2	2	2	2	m	5	12	9	4	15
14	1	2	2	2	m	5	12	2	1	15
15	1	2	2	2	m	5	12	2	1	7
16	1	2	2	2	m	5	12	11	9	14
17	7	2	2	2	m	5	12	8	9	2
18	2	2	2	2	m	5	5	7	13	2
19	9	2	2	2	m	5	5	4	9	11
20	1	3	2	2	m	5	12	9	6	5
21	8	4	2	2	m	3	5	4	9	16
22	1	9	3	2	m	5	12	9	6	10
23	1	3	2	3	m	5	12	8	11	5
<i>Marians' chromosome</i>										
1	1	2	2	2	m	5	12	9	1	5
2	5	1	1	2	m	5	12	2	6	6

The saved fragments of Chuvashian 'founder haplotype' are in bold.

OPTB in the Mari population

We studied one family with one OPTB case from the Mari population. The same c.807+5G>A mutation in the *TCIRG1* gene in the homozygous state was detected in the patient. In both disease chromosomes, the persisting fragments of the Chuvashian 'founder haplotype' were revealed (Table 3). Thus, OPTB cases in Chuvashians and Marians are the result of a single mutation event.

Diagnostics of OPTB in Chuvashiya and Mari

To diagnose OPTB cases in Chuvashian and Marian families, we used the RLFP analysis system described in 'Material and methods'. Diagnostics of OPTB was carried out for 22 Chuvashian and 2 Marian families. All patients carried the c.807+5G>A mutation in a homozygous state and all parents of affected children carried the mutation in heterozygous state. Five prenatal diagnostics were also carried out, and the prognosis proved favorable in three cases. For the mutation screening in health Chuvashians and Marians, MLPA melting curve analysis-based systems were designed and were used to detect the mutation in 225 related/unrelated Chuvashians (we identified 14 heterozygous mutation carriers).

Discussion

In this study, a molecular defect causing OPTB in Chuvashiya was discovered. As a result of LD mapping of the disease locus in the *TCIRG1* gene region and gene sequence analysis, the unique c.807+5G>A mutation was identified in all Chuvashian OPTB patients studied. Proceeding from the observed frequency of heterozygous carriers of the c.807+5G>A mutation in indigenous Chuvashians, a calculated disease frequency was found to be 1/3500 newborns. These data agree with the data from epidemiological research, according to which the disease frequency is equal to 1/3879 newborns.⁵ According to our results, it is possible that all OPTB cases in Chuvashiya are caused by a single mutation c.807+5G>A in the *TCIRG1* gene.

The mutation does not affect invariant 5'-GU nucleotides in intron 8 and results in a G>A substitution in the +5 position at a partially variant region within a canonical sequence of a donor splice site. In this position, nucleotide G occurs in 75–86% of exon–intron junction sequences and nucleotide A occurs in 5–11% of these.²⁵ The predicting splice site program has shown that the mutation results in decrease at twice the confidence of the normal donor splice site in intron 8. We were not able to show any

effect of failure of the normal donor splice site in intron 8 due to the lack of material from the patient. By using RT-PCR analysis in the heterozygous c.807 + 5G>A carrier, however, we have revealed, apart from the normal transcript, a longer splice variant bearing a 37 nt intron insertion. The insertion occurs as a result of using an additional donor splice site downstream of the normal one. The program predicted this additional donor splice site with high confidence in sequences bearing the mutation and those that were normal. The additional site, however, was used only in the presence of the mutation. It is therefore possible to assume that we deal with a cryptic donor splice site used in the failure of the normal one.

It is interesting that a fluorescence intensity of the band corresponding to the mutant transcript decreased in comparison with the normal band on the electrophoregram (Figure 2a). By using RLFP analysis of the intragenic SNP upstream of the mutation, we have confirmed a partial loss of SNP heterozygosity at the allele linked to the mutation. The fact that the identified intron insertion within the mutant transcript results in a frameshift and premature stop-codon formation can explain the partial loss of the mutant transcript quantity by the presence of the nonsense-mediated decay (NMD) process. With this connection, it was interesting to analyze other published mutations of the *TCIRG1* gene, as it is known that approximately 40% of the mutations are splice site mutations. We have analyzed the described mutations in Sobacchi *et al*⁴ and Susani *et al*¹² by using the predicting splice site NetGene2 and/or EditSeq programs. It is apparent that 17/21 (81%) of the splice site mutations result in a frameshift and a premature stop-codon formation. These mutations occur both in donor and in acceptor splice sites and cause an insertion/deletion of a full intron/exon. Five mutations result in the complex rearrangements in a mutation transcript caused by the use of the cryptic splice sites, as a rule. Only 4/21 mutations cause an in-frame long insertion/deletion. It is also clear that 21/29 (72%) of the intraexon mutations are nonsense mutations or other sequence modifications resulting in a premature stop-codon formation. Thus, the effects of 76% of the mutations in the *TCIRG1* gene may have been implemented by a common mRNA degradation mechanism, that is, an NMD.

During LD mapping of the OPTB locus, we found that all c.807 + 5G>A-bearing chromosomes had the same haplotype at SNP and microsatellite markers tightly linked to the *TCIRG1* gene. These data confirm the assumption of the founder effect for OPTB in Chuvashians, suggesting that we could estimate the age of the mutation. The obtained date of the beginning of the spreading of the mutation in the Chuvashian population corresponds to the twelfth century (with a range of 8–15 centuries). This result does not contradict to the theory traditionally shared by the majority of historians that Chuvash ethnoses had been

formed around the thirteenth to sixteenth centuries AD from the non-indigenous Turkic-speaking population of an ancient Volga Bulgar empire partially assimilating with the indigenous Finno-Ugric population in the north territory of contemporary Chuvashiya. The mutation most likely was fixed in the tribe of Bulgars-Suvars, which populated the North of current Chuvashiya during the period from thirteenth to fourteenth centuries or in old Marians (also known as Cheremis) who before this period occupied the territory. In fact, we also have found the c.807 + 5G>A mutation in the Marian population, and the frequency of heterozygous carriers appears to be insignificantly different from that found in Chuvashians. We then investigated one Mari OPTB family and revealed the same haplotype described in Chuvashian patients in both chromosomes bearing the c.807 + 5G>A mutation. This mutation is unique to Chuvashians and Marians and did not arise in the Udmurt and Bashkir populations geographically close to Chuvashiya. Furthermore, the c.807 + 5G>A mutation did not arise in Central Russian OPTB patients (data are not shown). These data confirm the hypothesis regarding the long reproductive isolation of the Chuvashian and Marian populations. These results also correspond to the data showing that the Volga region Chuvashians are genetically closest to Marians whereas Udmurts are genetically distant from both.

References

- 1 Balemans W, Van Wesenbeeck L, Van Hul W: A clinical and molecular overview of the human osteopetroses. *Calcif Tissue Int* 2005; **77**: 263–274.
- 2 Phadke SR, Gupta A, Pahi J, Pandey A, Gautam P, Agarwal SS: Malignant recessive osteopetrosis. *Indian Pediatr* 1999; **36**: 69–74.
- 3 Fasth A, Porras O: Human malignant osteopetrosis: pathophysiology, management and the role of bone marrow transplantation. *Pediatr Transplant* 1999; **3** (Suppl 1): 102–107.
- 4 Sobacchi C, Frattini A, Orchard P *et al*: The mutational spectrum of human malignant autosomal recessive osteopetrosis. *Hum Mol Genet* 2001; **10**: 1767–1773.
- 5 Ginter EK, Kirillov AG, Rogaev EI: [Autosomal-recessive osteopetrosis in Chuvashiya]. *Genetika* 2001; **37**: 1152–1155.
- 6 Vasserman NN, Karzakova LM, Tverskaya SM *et al*: Localization of the gene responsible for familial benign polycythemia to chromosome 11q23. *Hum Hered* 1999; **49**: 129–132.
- 7 Kazantseva A, Goltsov A, Zinchenko R *et al*: Human hair growth deficiency is linked to a genetic defect in the phospholipase gene LIPH. *Science* 2006; **314**: 982–985.
- 8 El'chinova GI, Ginter EK: [Malecot's parameters of isolation by distance and the degrees of endogamy in three regions of the Chuvash Republic]. *Genetika* 2001; **37**: 684–689.
- 9 Ginter EK, Zinchenko RA, El'chinova GI *et al*: [Genetic epidemiological study of populations in three regions of Chuvashia Republic]. *Genetika* 2001; **37**: 840–847.
- 10 Chalhoub N, Benachenhou N, Rajapurohitam V *et al*: Grey-lethal mutation induces severe malignant autosomal recessive osteopetrosis in mouse and human. *Nat Med* 2003; **9**: 399–406.
- 11 Janssens K, Van Hul W: Molecular genetics of too much bone. *Hum Mol Genet* 2002; **11**: 2385–2393.
- 12 Susani L, Pangrazio A, Sobacchi C *et al*: *TCIRG1*-dependent recessive osteopetrosis: mutation analysis, functional identification

- of the splicing defects, and *in vitro* rescue by U1 snRNA. *Hum Mutat* 2004; **24**: 225–235.
- 13 Kornak U, Kasper D, Bosl MR *et al*: Loss of the CIC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* 2001; **104**: 205–215.
 - 14 Cleiren E, Benichou O, Van Hul E *et al*: Albers-Schonberg disease (autosomal dominant osteopetrosis, type II) results from mutations in the CLCN7 chloride channel gene. *Hum Mol Genet* 2001; **10**: 2861–2867.
 - 15 Frattini A, Pangrazio A, Susani L *et al*: Chloride channel CLCN7 mutations are responsible for severe recessive, dominant, and intermediate osteopetrosis. *J Bone Miner Res* 2003; **18**: 1740–1747.
 - 16 Shin YJ: Chloride channel CLCN7 mutations in a Korean patient with infantile malignant osteopetrosis initially presenting with neonatal thrombocytopenia. *J Perinatol* 2004; **24**: 312–314.
 - 17 Lange PF, Wartosch L, Jentsch TJ, Fuhrmann JC: CIC-7 requires Ostml as a beta-subunit to support bone resorption and lysosomal function. *Nature* 2006; **440**: 220–223.
 - 18 Pfaffl MW: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001; **29**: e45.
 - 19 Bengtsson BO, Thomson G: Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* 1981; **18**: 356–363.
 - 20 Diaz GA, Gelb BD, Risch N *et al*: Gaucher disease: the origins of the Ashkenazi Jewish N370S and 84GG acid beta-glucosidase mutations. *Am J Hum Genet* 2000; **66**: 1821–1832.
 - 21 Risch N, de Leon D, Ozelius L *et al*: Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 1995; **9**: 152–159.
 - 22 Labuda D, Zietkiewicz E, Labuda M: The genetic clock and the age of the founder effect in growing populations: a lesson from French Canadians and Ashkenazim. *Am J Hum Genet* 1997; **61**: 768–771.
 - 23 Ivanov VP, Nikolayev VV, Dmitriev VD: [*Chuvashians. An ethnic history and the traditional culture*]. Moscow: DIK, 2000.
 - 24 Ginter EK, Zinchenko RA: [*Genetic Structure and Hereditary Diseases of Chuvash Population*]. Cheboksari: Publishing house 'Pegas', 2006.
 - 25 Zhang MQ: Statistical features of human exons and their flanking regions. *Hum Mol Genet* 1998; **7**: 919–932.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)