Increased Probability of Co-Occurrence of Two Rare Diseases in Consanguineous Families and Resolution of a Complex Phenotype by Next Generation Sequencing

Dennis Lal[1,2,3]*, Bernd A. Neubauer[2], Mohammad R. Toliat[1], Janine Altmüller[1], Holger Thiele[1], Peter Nürnberg[1,3,4], Clemens Kamrath[5], Anne Schänzer[6], Thomas Sander[1], Andreas Hahn[2‡], Michael Nothnagel[1‡]*

1 Cologne Center for Genomics, University of Cologne, 50931, Cologne, Germany, 2 Department of Neuropediatrics, University Medical Faculty Giessen and Marburg, 35392, Giessen, Germany, 3 Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50931, Cologne, Germany, 4 Center for Molecular Medicine Cologne (CMMC), University of Cologne, 50931, Cologne, Germany, 5 Department of Pediatrics, University Medical Faculty Giessen, 35392, Giessen, Germany, 6 Institute of Neuropathology University Medical Faculty Giessen and Marburg, 35392, Giessen, Germany

‡ These authors are shared senior authors on this work.
* Dennis.Lal@uni-koeln.de (DL); michael.nothnagel@uni-koeln.de (MN)

Abstract

Massively parallel sequencing of whole genomes and exomes has facilitated a direct assessment of causative genetic variation, now enabling the identification of genetic factors involved in rare diseases (RD) with Mendelian inheritance patterns on an almost routine basis. Here, we describe the illustrative case of a single consanguineous family where this strategy suffered from the difficulty to distinguish between two etiologically distinct disorders, namely the co-occurrence of hereditary hypophosphatemic rickets (HRR) and congenital myopathies (CM), by their phenotypic manifestation alone. We used parametric linkage analysis, homozygosity mapping and whole exome-sequencing to identify mutations underlying HRR and CM. We also present an approximate approach for assessing the probability of co-occurrence of two unlinked recessive RD in a single family as a function of the degree of consanguinity and the frequency of the disease-causing alleles. Linkage analysis and homozygosity mapping yielded elusive results when assuming a single RD, but whole-exome sequencing helped to identify two mutations in two genes, namely *SLC34A3* and *SEPN1*, that segregated independently in this family and that have previously been linked to two etiologically different diseases. We assess the increase in chance co-occurrence of rare diseases due to consanguinity, i.e. under circumstances that generally favor linkage mapping of recessive disease, and show that this probability can increase by several orders of magnitudes. We conclude that such potential co-occurrence represents an underestimated risk when analyzing rare or undefined diseases in consanguineous families and should be given more consideration in the clinical and genetic evaluation.

Introduction

Rare diseases (RD) are defined as disorders that affect only a minor proportion of the population. The European Union considers a disease to be rare if fewer than 5 people per 10,000 are affected (http://ec.europa.eu/health/rare_diseases/), corresponding to a prevalence < 0.05%. Similarly, any disease with less than 200,000 cases is considered rare in the United States (US) (http://rarediseases.info.nih.gov/), corresponding to a prevalence of < 0.063% in a population of approximately 316 million (http://www.census.gov/). Although being rare on their own, the large number of RD (5–8000; http://ec.europa.eu/health/rare_diseases/) leads to substantial proportions of the population being affected with any one of them. It is estimated that about 6–8% of the total population of approximately 500 million (http://epp.eurostat.ec.europa.eu/) suffer from such disorders throughout the 27 member states of the European Union, and a similar percentage (7.9–9.5%, http://rarediseases.info.nih.gov/) has been reported for the US. Thus, RD pose a serious challenge to population health and warrant investigation of their etiology. Despite their low prevalence, there have been a number of reports on the co-occurrence of two independent recessive diseases in consanguineous and non- consanguineous families in the past (e.g. [1–7]).

Diseases that follow a recessive Mendelian inheritance pattern have been an early target for indirect statistical mapping approaches, using parametric and non-parametric linkage mapping as well as homozygosity mapping (e.g. [8–10] for cystic fibrosis). However, factors such as locus and allelic heterogeneity as well as unspecific symptomatology can blur a clear relationship between causative factor and disease and thereby decrease the statistical power to detect such factors. Failures to repeat early mapping successes with other disorders and ensuing frustration led to a re-orientation of statistical mapping approaches towards association studies [11]. Next-generation sequencing (NGS) technologies now facilitate a direct assessment of causative variation, thereby obliterating the need for indirect mapping approaches. This has led to a recent renaissance of linkage-based approaches. Linkage analysis of variants obtained from whole-exome sequencing (WES) is currently the method of choice for deciphering recessive RD. In particular, large families living in or stemming from rural regions where marriages between relatives are common provide favorable conditions for statistical mapping approaches, because increased consanguinity levels lead to increased prevalence of some recessive diseases and to higher statistical power in these populations [12].

Still, some issues remain. The elucidation of genetic causes involved in RD can also be hampered by factors such as variability in the phenotypic manifestation with respect to age of onset, clinical features, and severity of symptoms and from difficulties to distinguish two etiologically distinct disorders with partially overlapping symptoms [13]. Successful identification of disease-associated genetic loci critically depends on precise phenotype information as much as on correctly specified inheritance models. Indeed, problems in delineating the exact phenotype are generally correlated with failures in discriminating between affected and unaffected family members. Furthermore, a situation where two phenotypically similar recessive diseases segregate independently in a single family is usually not taken into consideration given the low prevalence of any of these two diseases, despite previous reports of a co-occurrence of rare Mendelian diseases.

An example of a rare Mendelian disease is congenital myopathy (CM), with a prevalence of 1/26000 in the US (www.orpha.net; Orpha number: ORPHA97242, [14]). CM includes a group of neuromuscular disorders with variable clinical and histopathological features, in which muscular weakness and skeletal deformities are prominent findings. Recent technical advance has revealed marked genetic heterogeneity, and has shown that specific genetic defects are associated with variable clinical and histopathological phenotypes [15].

Hereditary hypophosphatemic rickets (HHR) is a further RD with an incidence of 1/21 000 [16], encompassing several genetically distinct disorders, all characterized by hypophosphatemia secondary to renal phosphate wasting, rickets, short stature, and skeletal deformities. Muscular weakness is also frequently reported, and studies in mice and humans with HHR have shown altered muscle composition and function [17].

In this study, we present a straightforward approximate approach for estimating the probability for the co-occurrence of two unlinked recessive RD in a single family as a function of the degree of consanguinity in the population and the frequency of the disease-causing alleles. We show that consanguinity can increase this probability by several orders of magnitudes, thereby increasing the chances of a manifestation of two etiologically different recessive Mendelian diseases in a single family. We illustrate our approach with a consanguineous family where several members had muscular weakness of variable degree and abnormal muscle biopsy findings; along with skeletal deformities, early ventilatory failure, and renal phosphate wasting occurring in some of them. Linkage analysis of this family yielded elusive results when assuming a single segregating RD, but WES helped to identify two mutations in two genes, namely *SLC34A3* and *SEPN1*, that segregated independently in this family and that have previously been linked to two etiologically different diseases.

Materials and Methods

Statistical basis for calculation of the joint occurrence of two recessive diseases in consanguineous populations

Assume that the occurrence of a recessive disease can be, at least in parts, parameterized by the homozygous carrier-ship of a deleterious allele A at a particular locus. Let q_A denote the population frequency of this allele and let the genotype AA be fully penetrant, i.e. carriers of genotype AA will always become affected. With little or no selection before the reproductive phase, the probability of becoming affected for an individual randomly drawn from an outbred population, $P(D)$, will equal the Hardy-Weinberg proportion of the genotype in the population:

$$P(D) = q_A^2 \qquad (1)$$

However, if consanguineous marriages are common in a population or subpopulation, the occurrence of the disease due to homozygous AA carrier-ship will increase. Its probability can then be parameterized by the inbreeding coefficient F_I [18], namely

$$P_I(D) = F_I \cdot q_A + (1 - F_I) \cdot q_A^2, \qquad (2)$$

which reduces to Eq 1 for the case of no inbreeding ($F_I = 0$).

A lower limit for the probability that two or more diseases occur within the same family is the co-occurrence of these diseases in a single individual. More specific, if some recessive disease, D_1, is caused by homozygous carrier-ship of allele A at one locus and another recessive disease, D_2, by homozygous carrier-ship of allele B at another unlinked locus, then the probability of their joint occurrence is simply the product $P_I(D_1)P_I(D_2)$. The increase in probability of a joint occurrence of both diseases in a single individual due to inbreeding can then expressed as a simple function of the allele frequencies, q_A and q_B, and the inbreeding

coefficient

$$f(F_I, q_A, q_B) = \frac{P_I(D_1) \cdot P_I(D_2)}{P(D_1) \cdot P(D_2)}$$
$$= \frac{F_I \cdot q_A + (1 - F_I) \cdot q_A^2}{q_A^2} \cdot \frac{F_I \cdot q_B + (1 - F_I) \cdot q_B^2}{q_B^2},$$
$$= \frac{F_I + (1 - F_I) \cdot q_A}{q_A} \cdot \frac{F_I + (1 - F_I) \cdot q_B}{q_B} \tag{3}$$

which reduces to

$$f(F_I, q) = \frac{(F_I + (1 - F_I) \cdot q)^2}{q^2}, \tag{4}$$

if one assumes, for the sake of simplicity, equal population frequencies of both alleles ($q = q_A = q_B$). Note that the consideration of less than fully penetrant disease loci would not change expressions ([3]) and ([4]).

Furthermore, we may also inquire about the probability that at least two out of a set of k diseases co-occur in an individual. Again assuming unlinked and fully penetrant loci and allele frequencies $q_1, \ldots, q_k$ for the different loci, respectively, the probability for this event can be approximated by

$$P(c \geq 2 | F_I, q_1, \ldots, q_k) = 1 - \prod_{i=1}^{k}(1 - q_i^I) - \sum_{i=1}^{k}\left(q_i^I \prod_{j=1, j \neq i}^{k}(1 - q_j^I)\right), \tag{5}$$

where q_i^I denotes the frequency of the i-th homozygous genotype under consanguinity:

$$q_i^I = F_I q_i + (1 - F_I) \cdot q_i^2. \tag{6}$$

Phenotypic characterization of family members suspected to suffer from a not yet described congenital myopathy with hypophospathemic rickets

We identified a consanguineous Turkish family with several members affected to a variable degree with muscular weakness and hypotonia in conjunction with early ventilatory failure and/or hypophosphatemic rickets in some of them ([Fig 1]). The institutional review board (IRB) of the Department of Medicine, University of Giessen, Germany, approved this study. Written informed consent for genetic analysis was obtained from all subjects or their legal guardians according to study protocols approved by this IRB. Clinical and paraclinical findings of the affected individuals are summarized in [Table 1], and muscle biopsy findings are shown in [Table 2], [Fig 2], and [S1 Fig].

Patient III-1 was referred at age 3 ½ years because of delayed motor development and muscular hypotonia. His height was between the third and 10th percentile. He was able to stand up from the floor and to climb stairs, but had marked difficulties to sit up from the supine position. Mobility of the spine was normal, and his creatine kinase (CK) level was within the normal range. Electromyography (EMG) showed no unequivocal abnormalities, but a muscle biopsy demonstrated myopathic changes. Determination of calcium, phosphate, 1,25-OH-dihydrocholecalciferol, and tubular phosphate re-absorption revealed normal values. A diagnosis of CM was made and regular follow-up was recommended, but appointments were not kept by the family. At age 8 years, the boy was admitted to our hospital as an emergency case with
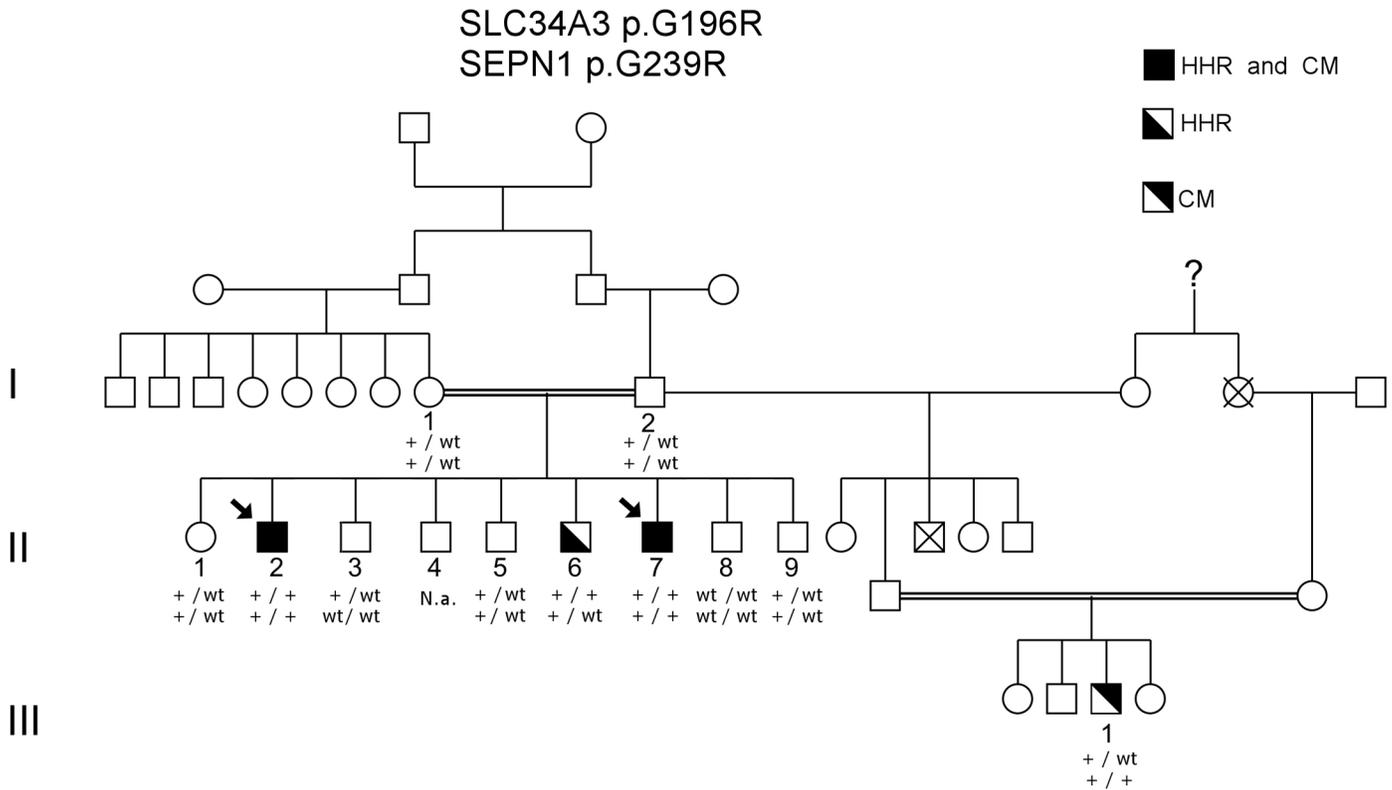
Fig 1. Pedigree and mutation segregation. All patients exhibiting symptoms of HHR are homozygous for the mutation *SLC34A3* p.G196R (c.586G>A NM_080877.2), whereas all patients expressing symptoms of CM are homozygous for the mutation *SEPN1* p.G239R (c.715G>A NM_206926.1). Therefore, patients II-2 and II-7 present features of both diseases. Arrows indicate index patients. Abbreviations: HHR = Hereditary hypophosphatemic rickets; CM = congenital myopathy.

doi:10.1371/journal.pone.0146040.g001

pneumonia and acute ventilatory failure, necessitating implementation of permanent non-invasive nocturnal ventilation.

In **patient II-2**, mild muscular weakness and reduced endurance were noticed since early childhood. The diagnosis of HHR was made at age 8 years due to dysproportionately short statue, and skeletal deformities such as rickets, flat occiput, and x-legs; in conjunction with hypophosphatemia and reduced tubular phosphate re-absorption. Mutations in *PHEX1*, and during the following years also in *FGF23*, *DMP1*, *and CLCN5* were ruled out. Muscular weakness initially improved during treatment with calcium and rocaltrol, and therefore was attributed to HHR. But after the age of 20 years, he became increasingly short of breath, and non-invasive ventilation had to be started at age 22 years. A muscle biopsy taken at this time showed mild to moderate myopathic changes. At age 26 years, his height was 5 cm below the third percentile. He was able to rise up from the floor, walked with a waddling gait, and was still climbing stairs. He had a rigid spine and bending of the head is nearly completely abolished.

Patient II-6's early motor development was delayed from the beginning. The diagnosis of HHR was established at age 6½ years, after the disease had been diagnosed in his older brother. Muscular hypotonia and weakness were documented from this age on. At age 13 years, his CK and EMG were normal, while a muscle biopsy taken during an osteotomy to correct his x-legs revealed single muscle fiber atrophy. At last follow-up, at age 15 years, he showed a mild proximal muscle weakness, and his height was on the 25th percentile.

Table 1. Synopsis of clinical and paraclinical findings in patients II-2, II-6, II-7 and III-1.

	II-2	II-6	II-7	III-1
SEPN1 mutation	+	-	+	+
SLC34A3 mutation	+	+	+	-
Delayed motor milestones	+	-	-	+
Muscular weakness	+	+	+	+
Muscular hypotonia	+	+	+	+
Rigid spine	+	-	-	-
Short stature	+	-	-	-
X-legs	+	+	+	-
Ventilatory failure (years)	22	-	-	8
Creatine kinase (12–174 U/l)	51	80	78	116
Abnormal EMG	+	-	+	-
Abnormal muscle biopsy	+	+	+	+
Loss of ambulance	-	-	-	-
Rickets	+	+	+	-
Serum calcium (2.0–2.6 mmol/l)	2.4	2.4	2.6	2.1
Serum phosphate(1.2–1.8 mmol/l)	**0.7**	**0.9**	**1.0**	1.4
Alkaline phosphatase (92–390 U/l)	**960**	**562**	271	251
Urine calcium/creatinine ratio (<0.3)	**0.76**	**1.13**	**0.59**	0.18
Phosphate re-absorption (82–90%)	**15**	**26**	**43**	82
1,25- dihydroxy-cholecalciferol (18–62 pg/ml)	21	60	111	45

Numbers in brackets correspond to normal and bold values to abnormal values

doi:10.1371/journal.pone.0146040.t001

In **patient II-7**, HHR was diagnosed at age 3 ½ years together with his older brothers. He also suffered from mild proximal muscle weakness. At age 13 years, his CK was normal, while the EMG showed a myopathic pattern. A muscle biopsy performed during a metal removal secondary to a femoral fracture showed myopathic changes. At last follow-up, at age 17 years, his height was on the 50th percentile. He walked with a waddling gait and was unable to run fast, but stood up from the floor without Gower's sign, and climbed stairs. Bending of his spine was not limited. He had difficulties to sit up, while his vital capacity is within the normal range.

Genotyping

DNA samples from the core family members I-1, I-2, II-1–II-3, and II-5–II-9 (Fig 1) were genotyped for linkage analysis using Illumina HumanLinkage-12 BeadChip, consisting of 6,090 single nucleotide polymorphism (SNP) probes according to the manufacturer's instructions. The genotype assignments were determined with the GenomeStudio version 1.9.4 genotyping module (Illumina Inc., San Diego, CA). All samples passed quality control (Array SNP call rate > 99%). Downstream quality control was automatically performed by the Superlink tool (for details see [19]) including marker filtering, linkage disequilibrium (LD) pruning as well as Mendelian error checking.

Initial mutational screening strategy

Given the family structure (Fig 1) and phenotype manifestations, we initially assumed an autosomal recessive mode of inheritance with incomplete penetrance for a novel rare disease in the family.

Table 2. Muscle morphology in patients II-2, II-6, II-7 and III-1.

|	II-2	II-6	II-7	III-1
Age of biopsy (years)	24	17	15	4
Biopsied muscle	vastus lat.	vastus lat.	vastus lat.	vastus lat.
Genetic phenotype	CM+HHR	HHR	CM+HHR	CM
Variability of fiber size	++	(+)	++	+
Internal nuclei	(+)	-	(+)	(+)
Core like lesions (NADH)	++	(+)	+	(+)
Endomysial Fibrosis	(+)	-	(+)	(+)
Fatty replacement	++	-	++	+
Inclusions (Desmin)	+	-	++	(+)
Inclusions (αβ Crystallin)	-	-	(+)	-
Typ II fiber predominance	+	+	+	(+)
Fiber necrosis	-	-	-	-
Fiber regeneration	-	-	-	-
CD56 positive fibers	-	-	+	+
Endomysial lymphozytes	+	-	-	-
Minicores (EM)	+	-	+	-

EM = electron microscopy

Initial mapping strategies

We pursued two mapping strategies based on the obtained SNP data. First, we performed genome-wide parametric linkage analysis under a recessive model of inheritance. To this end, we assumed a single locus, first with complete penetrance, and determined all non-healthy family members as affected, thereby assuming all affected members to suffer from the same rare disease. Two- and multi-point LOD scores were calculated with 99% penetrance using Superlink [19]. Second, we performed genome-wide homozygosity mapping, using Homozy-gosityMapper2012 [20].

Sequence capture and next generation sequencing

Since we expected at the beginning the segregation of a single disease, we generated exome sequencing data for patients II-2 and II-7 (Fig 1), both expressing the full range of symptoms observed in this family. We fragmented 1 μg of DNA from each of the two patients and their healthy parents with sonification technology (Covaris, Woburn, MA, USA). The fragments were end-repaired and adaptor-ligated, including incorporation of sample index barcodes. After size selection, we subjected the library to an enrichment process with the SeqCap EZ Human Exome Library version 2.0 kit (Roche NimbleGen, Madison, WI, USA). The enriched pools were then sequenced on one lane of an Illumina HiSeq 2000 sequencing instrument (Illumina, San Diego, CA, USA) with a paired-end 2×100 bp protocol. We filtered primary data according to signal purity with the Illumina Realtime Analysis software version 1.8. Subsequently, we mapped the reads to the human genome reference build hg19 with the ELANDv2 alignment algorithm on a multinode compute cluster. By use of CASAVA version 1.8 (Illumina processing software, Illumina), PCR duplicates were filtered out and the output was converted into BAM format.

An average coverage of >30x was achieved for 77% of the target sequence. Given the consanguinity of the family, variants were filtered for high-quality rare homozygous variants by
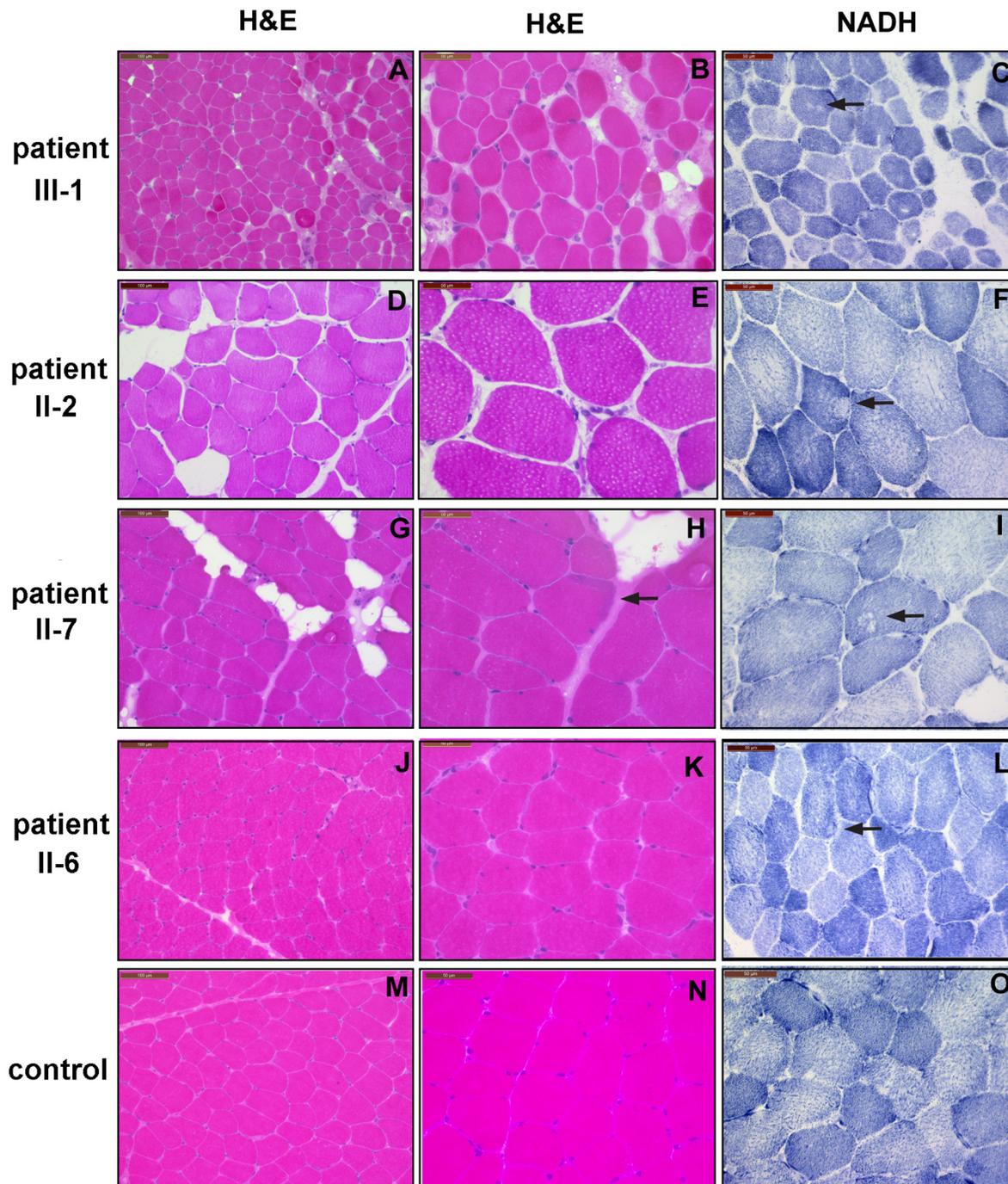
Fig 2. Histochemical and immunohistochemial biopsy analyses from affected family members and a normal control. H&E staining of transverse cryostat sections displays findings well compatible with a slowly progressive myopathy with some replacement of muscle by fat in patients III-1 (A), II-2 (D), and II-7 (G); and is also abnormal in patient II-6 due to increased fiber variability (J) compared to a control (M). There is increased variability in fiber size in all family members (B,E,H,K) Notice also additional small atrophic fibers in patient II-2 (E), and some subsarcolemnal muscle fiber disorganization in patient II-7 (H arrow). NADH enzymehistochemistry depicts core like lesions in patients II-2 (F) and II-6 (I), and minor myofibrillar disorganization in patients III-1 (C) and II-6 (L).

doi:10.1371/journal.pone.0146040.g002

comparison with external reference sources, including an in-house variation database (N> 1300 exomes) and more than 60,000 individuals without Mendelian disorders from the Exome Aggregation Consortium (ExAC; www.exac.broadinstitute.org; accessed Oct. 2015). We excluded those variants that had an allele frequency exceeding 1% in any of these sources from further investigation. The segregation was tracked *in silico* by haplotypes identified in the mapping analysis and then validated by Sanger sequencing. The Sanger sequencing was performed following standard protocols.

Results

We identified a consanguineous Turkish family with several members being affected to a variable degree with muscular weakness and hypotonia in conjunction with early ventilatory failure and/or hypophosphatemic rickets in some of them (Fig 1). Clinical and paraclinical findings of the affected individuals are summarized in Table 1, and muscle biopsy findings are shown in Table 2, Fig 2, and S1 Fig (see also Materials and Methods).

Whole-genome linkage analysis

An initial parametric linkage analysis was performed under the assumption of a single recessive disease with 99% penetrance and failed to identify a shared homozygous haplotype in all patients. Thereafter, linkage analysis was performed including all family members with symptoms of HHR (II-2, II-6 and II-7). This yielded a maximum genome-wide two-point and multi-point LOD score of 1.82 on 9q34 (Fig 3A). At this point, the muscle biopsies were re-assessed. Re-evaluation confirmed unequivocal myopathic features in patients II-2, II-7, and III-1, whereas the muscle histology from patient II-6 was interpreted as showing abnormal, but unspecific changes. In line with this observation we identified a homozygous p.G196R mutation (c.586G>A, NM_080877.2; p.G196R) in the *SLC34A3* gene in the three siblings affected by HHR.

Sequencing of a shared homozygous region

In addition, we performed homozygosity mapping in patients II-2 and II-7, who displayed signs of HHR and unequivocal myopathic features in the muscle biopsy, hypothesizing an independent molecular mechanism of this specific phenotype. The analysis revealed five large, continuous segments of homozygosity present in both patients and heterozygous in the parents. The second largest segment of 33Mb (chr1: 22Mb-55Mb; Fig 3B) contained the *SEPN1* gene, known to be associated with different forms of CM [21]. Consecutively, WES revealed the presence of a homozygous p.G239R mutation in *SEPN1* (c.715G>A, NM_206926.1, p.G239R) in both affected siblings. The p.G239R exchange had been previously reported by Moghadaszadeh et al. [22] in a patient with rigid spine muscular dystrophy. However, the variant had been falsely reported in their paper and has been corrected and annotated in HGMD (www.hgmd.cf.ac.uk) recently. The correct annotated variant exchange is (GRCh37: Chr1:26135586, NM_206926,NP_996809.1: p.Gly239Arg) and therefore identical with ours.

Although a total of 42 homozygous variants were identified in the whole coding region shared by both sequenced individuals with HHR and CM, no other homozygous variants were detected in genes that are known to relate to CM or HHR (S1 Table). Indeed, both detected mutations, p.G196R in *SLC19A3* and p.G239R in *SEPN1*, had been previously reported in patients with either HHRH [23] or rigid spine muscular dystrophy-1 (RSD-1) [22], respectively.

Finally, Sanger sequencing confirmed heterozygous carrier-ship of the *SLC19A3* and *SEPN1* mutations in the clinically healthy parents (I1 and I2), while patients II-3 and II-7 were found to be homozygous for both pathogenic variants. Furthermore, patient II-6 was found to be
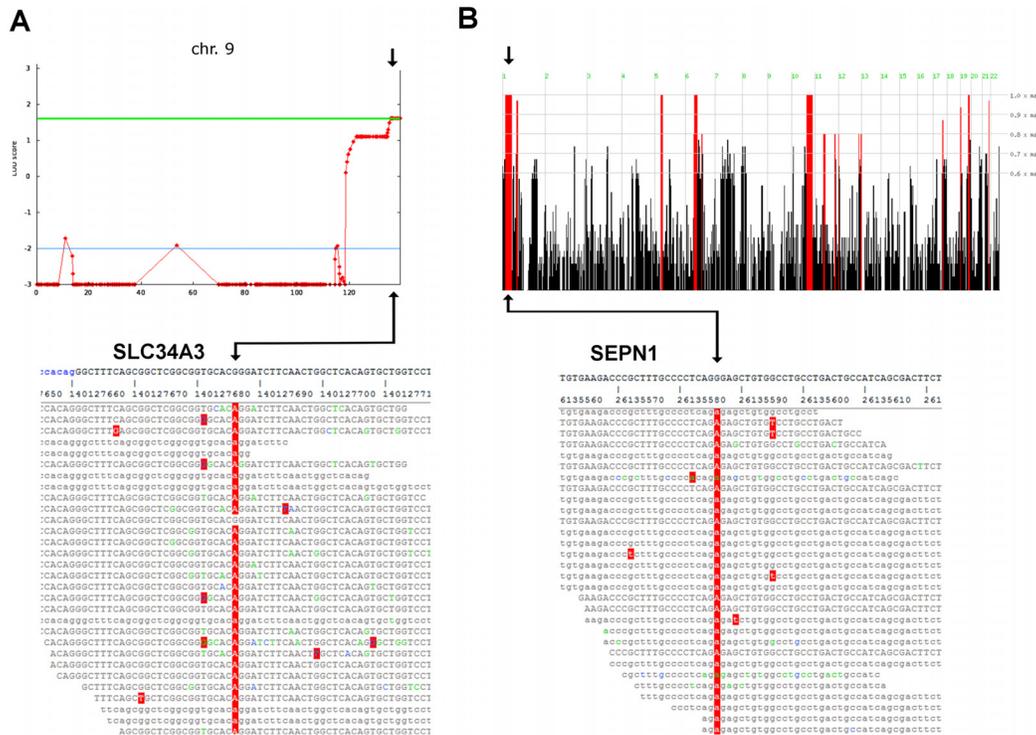
Fig 3. Variant discovery. (A) Top: The *SEPN1* p.G239R (c.715G>A NM_206926.1) mutation is located in a large, continuous segment of homozygosity identified in patients II-2 and II-7 of 33Mb (chr1:22Mb-55Mb). Bottom: Schematic representation of the mapped sequencing reads of the p.G239R mutation. **(B)** Top: Linkage analysis of all patients with symptoms of HHR (II-2, II-6 and II-7) yielded maximum genome-wide two-point and multi-point LOD score of 1.62 on 9q34, indicated by green line, harboring the *SLC34A3* p.G196R (c.586G>A NM_080877.2) mutation identified in II-2 and II-7. Bottom: Schematic representation of the mapped sequencing reads of the *SLC34A3* c.586G>A (NM_080877.2) variant.

doi:10.1371/journal.pone.0146040.g003

homozygous for the *SLC34A3* only, while patient III-1 carried merely the homozygous *SEPN1* mutation (Fig 3).

Co-occurrence of two rare recessive diseases

Given the refined phenotype assessment in our sample family, we theoretically explored the probability of a joint occurrence of two unlinked, etiologically different diseases in a single family and assessed how consanguinity impacts on this probability. To this end, we simplified the task to the question of how likely it is that two particular independent diseases co-occur in a single individual, thereby approximating the probability for more complex cases, such as the co-occurrence in two related family members (see Materials and Methods).

Assuming the same frequency, q, for the causative alleles at both disease-related loci and applying Eq 4, the probability of the joint occurrence of both diseases (Fig 4A) and the increase in probability due to consanguinity (Fig 4B) can be plotted as a function of q and the inbreeding coefficient F_I. For relatively common alleles of frequencies > 1%, the co-occurrence probability is largely determined by q and, correspondingly, increased consanguinity levels barely lead to an increase of this probability. With increasingly rare alleles, the impact of consanguinity grows steeply, making the co-occurrence of two diseases, although very rare in absolute terms, more likely by several orders of magnitude. For example, values of 0.01 for F_I would lead to an ~120-fold probability increase for $q = 0.001$ and even ~10,200-fold increase for $q = 0.0001$. With more pronounced consanguinity ($F_I = 0.05$), the corresponding numbers would equal ~2,600 and ~251,000, respectively.

A   B
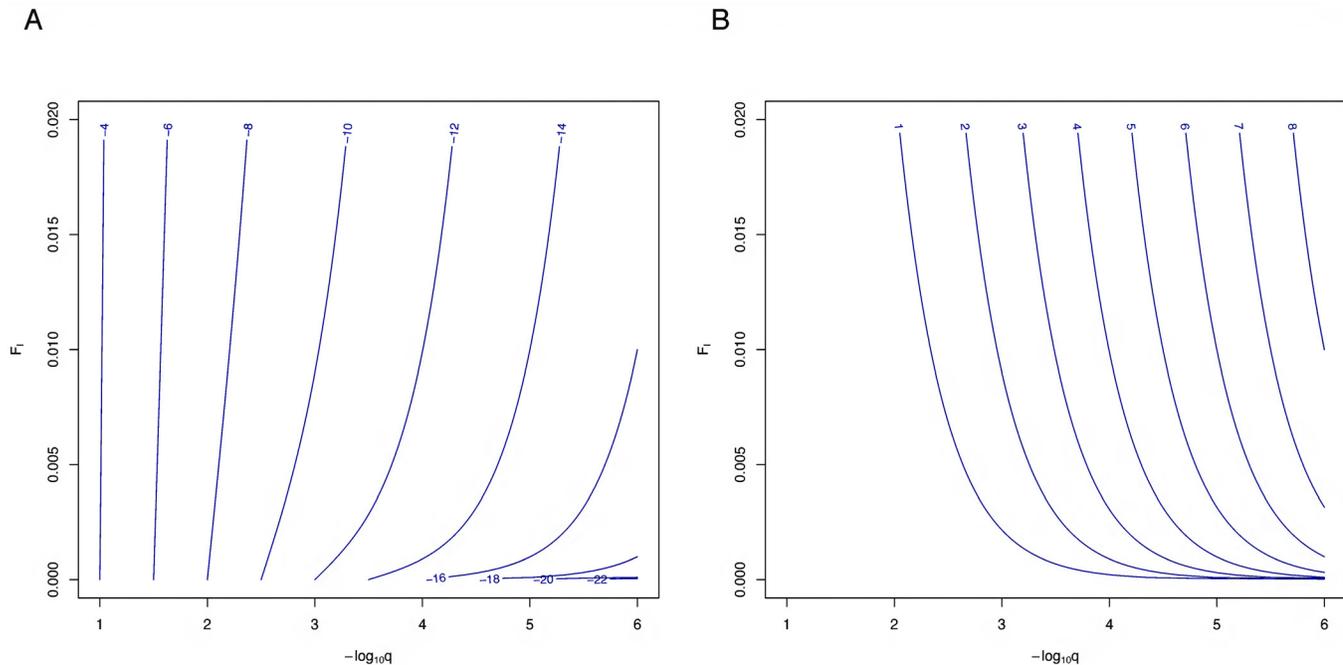
Fig 4. Co-occurrence of two recessive diseases under consanguinity. Probabilities are calculated for a model of two recessive diseases, each being caused by a single fully penetrant locus with identical susceptibility allele frequency, q, and both loci being unlinked. **(A)** $\log_{10}$ value of the probability to observe two recessive diseases in a single family by chance as a function of consanguinity level F_I in the general population and the same risk allele frequency, q, at both unlinked loci. **(B)** $\log_{10}$ value of the increase in the co-occurrence probability due to consanguinity compared to the non-consanguineous case ($F_I = 0$). Probabilities and allele frequencies are depicted using their decadic logarithm ($\log_{10}$). For example, when assuming values of $F_I = 0.01$ and q = 0.001 = 10^{-3}, the two diseases will jointly occur in a single family with a probability of roughly 10^{-10}, which equals approximately an $10^2 =$ 100-fold increase in probability in comparison to a non-consanguineous population.

doi:10.1371/journal.pone.0146040.g004

Considering the particular case of HHR und CM, we may also roughly assess the increase in probability to observe both diseases together in a single individual, assuming unlinked causal loci for these two diseases. To this end, we postulated that each of the diseases is exclusively caused by a single, fully penetrant mutation under a recessive disease model and assumed an average inbreeding coefficient of $F_I = 0.01$ for eastern Turkey (http://consang.net/), causing an only modest deviation of the genotype frequencies from Hardy-Weinberg equilibrium in absolute terms. Incidence values of 3.9/100,000 and 5.0/100,000 for HHR [16] and CM [14], respectively, then lead to frequency alleles estimates of 0.0030 and 0.0037, respectively, using the solution to the quadratic Eq 2. Applying Eq 3, the probability of observing both diseases together is then increased by factor ~16, compared to the case of a non-inbred population. We caution that this calculation represents only a rough approximation of the probability of disease co-occurrence, given that a number of other factors may also play a role in the etiology of both diseases.

We also considered the case where a set of k unlinked diseases show overlap of clinical symptoms that are difficult to distinguish and inquired how likely the co-occurrence of any two of them is (see Materials and Methods). Again, we assumed the same frequency, q, for the causative alleles at all disease-related loci, applied Eq 6 and considered in turn a set of 3, 5 and 10 diseases. Neither the probability of co-occurrence of at least two of these diseases, nor the increase in probability due to inbreeding did substantially differ from the case of two particular diseases (data not shown). This observation is likely due to the extremely small probability for more than two diseases to co-occur, being magnitudes smaller than the probability of co-occurrence of just two diseases and therefore contributing only little probability mass to the calculations.

Discussion

Recent advances in sequencing technology have raised high hopes for identifying genetic causes of many RD that had not been amenable to classic indirect linkage mapping or population-based association studies in the past. Indeed, many causative *de-novo* mutations and rare variants implicated in familial disease forms have been discovered in the past few years, sometimes based on family data that were sampled many years ago (e.g. [24]). However, a seemingly straightforward analysis can be hampered by several blurring factors. This includes incomplete penetrance of causative factors, locus and allelic heterogeneity as well as phenotypic heterogeneity, but also overlap of clinical symptomatology, and unspecific paraclinical features. Consanguineous pedigrees are a favorable sampling design for studying rare recessive diseases due to increased prevalence of some diseases and a higher statistical power for identifying causative genetic factors, and, thus, populations with increased levels of consanguinity have primarily been targeted for studying such diseases. However, these populations are precisely those where one would expect to see a much higher co-occurrence of rare recessive diseases by chance than in a completely non- consanguineous population. If these diseases show overlapping symptomatology, it is quite likely that they will be analyzed as a single disease in the first place and that the induced "locus heterogeneity" will hamper any successful linkage analysis. As we have shown here, the probability of co-occurrence can increase by several orders of magnitude, depending on the frequency of the risk allele and the degree of consanguinity. We note that the increase in co-occurrence probability is particularly severe for such rare and very rare variants. We therefore suggest that studies on rare recessive diseases should put strong emphasis on precise phenotyping and should consider the possibility of co-occurrence of two rare and phenotypically similar diseases within a single family especially when sampling from consanguineous populations.

As an instructive illustration, we presented the genetic analysis results of a consanguineous family with several members erroneously assumed to suffer from a yet not described recessive form of CM with renal phosphate wasting as an additional feature. Instead, linkage analysis combined with WES revealed independent segregation of two rare genetically distinct disorders, *SEPN1*-related myopathy and *HHRH*, both causing muscular weakness and abnormal muscle histology.

In this family, three affected members harbored a homozygous, previously reported mutation in *SEPN1*, coding selenoprotein N, a glycoprotein-localized within the endoplasmic reticulum [25]. As in our family, defects in *SEPN1* have been associated with a variable clinico-histopathological picture that includes rigid spine muscular dystrophy (RSMD-1), multimini-core disease, desmin-related myopathy with Mallory body-like inclusions (MR-DRM) [26, 27], and also non-specific myopathological features [28]. Retrospectively, some findings like rigid spine in patient II-2, early ventilatory failure despite preserved ambulance in subjects II-2 and III-1, core-like lesions in patients II-2 and II-7, and fibers showing minicores or expressing αβ-Crystallin in subject II-7 are consistent with a myopathy caused by *SEPN1* defects [26, 28–30]. However, the clinical symptoms and histopathological features observed in affected family members may also well be found in other neuromuscular disorders, and, thus, seemed us not sufficient to identify *SEPN1* as a candidate gene of high priority.

Mutations in *SLC34A3*, coding the renal sodium-phosphate co-transporter NaPi-IIc have been recently identified as a further rare cause of HHR [31]. As in other forms of HHR muscular weakness has been noted [31, 32]. The exact reasons why muscle function is impaired in patients with HHR are unclear. But assessment of lower extremity muscle function and determination of calf muscle cross-sectional area and density by peripheral quantitative computed tomography has revealed lower muscle force and reduced muscle density in patients with and

without limb deformities compared to controls [32]. Of note, the subject with the earliest onset of ventilatory failure was not affected by HHR, indicating that co-segregation of both genetic defects did not result in a more severe phenotype.

Conclusions

Taken together, the findings in this family demonstrate the difficulties in defining the affection status in patients with CM, show the problems in differentiating specific types of CM, and illustrate that an inheritance model assuming the occurrence of two genetically distinct disorders should not be neglected in larger consanguineous families.

Targeted massive parallel sequencing and candidate gene screening combined with haplotype tracking was performed to identify the underlying genetic cause for the disease spectra in this family. Our analysis unmasked that the phenotype in the family did not reflect a complex novel disease, but instead was due to two homozygous mutations in *SEPN1* and *SLC34A3*, respectively. [22, 23].

In retrospect, the event of exhibiting a combined recessive *SLC34A3* and *SEPN1* disease is more likely than expected initially, since in consanguineous unions autosomal recessive disorders are increased [33, 34], and because the p.G196R (*SLC34A3*) and the p.G239R (*SEPN1*) alleles were identified in Turkish and Iranian families, suggesting a higher frequency of the variants in middle eastern populations. Thus, in the clinical diagnostic procedure the presence of two rare diseases in one family should be taken into account, in particular if the disease risk is increased. Therefore, analysis of all known disease symptom associated genes should be considered, especially in clinically heterogeneous disorders like CM, where diagnosis is challenging due to phenotypic variability [35–38].

In the family reported, the final diagnosis could only be established by genetic analysis. The use of WES for clinical diagnostics offers a rapid and inexpensive opportunity and will improve genetic counseling in Mendelian diseases, even if masked as in the family presented in this report.

Supporting Information

S1 Fig. Histochemical, immunohistochemical, and ultrastructural muscle biopsy analyses of affected family members and a normal control. The muscle biopsies from all family members show type II fiber predominance (A,F,K,P) compared to a control (U). Immunohistochemical analyses display few myofibers expressing CD56 in patients III-1 (B, arrow) and II-7 (L, arrow). Small Desmin positive inclusions are present in patients III-1 (C, arrow), II-2 (H, arrow), and II-7 (M, arrow). αβCrystallin inclusion was only occasionally detected in patient II-7 (N, arrow). Ultrastructural analysis demonstrates unspecific myofibrillar disorganization with empty vacuoles in patients III-1 (E), II-2 (J) and II-6 (T) and Z-band streaming in patient II-7 (O, arrow).
(TIF)

S1 Table. Exome sequencing results. List of rare (MAF<0.1) recessive mutations present in both patients II-2 and II-7 (Fig 1) who presented the full range of symptoms observed in this family.
(XLSX)

Acknowledgments

We are indebted to the family who participated in this study.

Author Contributions

Conceived and designed the experiments: DL BAN AH MN. Performed the experiments: MRT JA PN. Analyzed the data: DL BAN AH MN. Contributed reagents/materials/analysis tools: BAN CK AS TS MN. Wrote the paper: DL BAN JA HT PN CK AS TS AH MN.

References

1. Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, et al. Pitfalls in homozygosity mapping. American journal of human genetics. 2000; 67(5):1348–51. doi: 10.1016/S0002-9297(07)62966-8 PMID: 11007652; PubMed Central PMCID: PMC1288578.

2. Lezirovitz K, Pardono E, de Mello Auricchio MT, de Carvalho ESFL, Lopes JJ, Abreu-Silva RS, et al. Unexpected genetic heterogeneity in a large consanguineous Brazilian pedigree presenting deafness. Eur J Hum Genet. 2008; 16(1):89–96. doi: 10.1038/sj.ejhg.5201917 PMID: 17851452.

3. Benayoun L, Spiegel R, Auslender N, Abbasi AH, Rizel L, Hujeirat Y, et al. Genetic heterogeneity in two consanguineous families segregating early onset retinal degeneration: the pitfalls of homozygosity mapping. Am J Med Genet A. 2009; 149A(4):650–6. doi: 10.1002/ajmg.a.32634 PMID: 19140180.

4. Frishberg Y, Ben-Neriah Z, Suvanto M, Rinat C, Mannikko M, Feinstein S, et al. Misleading findings of homozygosity mapping resulting from three novel mutations in NPHS1 encoding nephrin in a highly inbred community. Genetics in medicine: official journal of the American College of Medical Genetics. 2007; 9(3):180–4. PMID: 17413422.

5. Marin S, Casano-Sancho P, Villarreal-Pena N, Sebastiani G, Pinillos S, Perez-Duenas B, et al. Triple A syndrome in a patient with genetic growth hormone insensitivity: phenotypic effects of two genetic disorders. Hormone research in paediatrics. 2012; 77(1):63–8. doi: 10.1159/000335235 PMID: 22269647.

6. Adams DR, Yuan H, Holyoak T, Arajs KH, Hakimi P, Markello TC, et al. Three rare diseases in one Sib pair: RAI1, PCK1, GRIN2B mutations associated with Smith-Magenis Syndrome, cytosolic PEPCK deficiency and NMDA receptor glutamate insensitivity. Molecular genetics and metabolism. 2014; 113 (3):161–70. PMID: 24863970; PubMed Central PMCID: PMC4219933.

7. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369(16):1502–11. doi: 10.1056/NEJMoa1306555 PMID: 24088041; PubMed Central PMCID: PMC4211433.

8. Watkins PC, Schwartz R, Hoffman N, Stanislovitis P, Doherty R, Klinger K. A linkage study of cystic fibrosis in extended multigenerational pedigrees. American journal of human genetics. 1986; 39 (6):735–43. PMID: 2879440; PubMed Central PMCID: PMC1684122.

9. Ober C, Bombard A, Dhaliwal R, Elias S, Fagan J, Laffler TG, et al. Studies of cystic fibrosis in Hutterite families by using linked DNA probes. American journal of human genetics. 1987; 41(6):1145–51. PMID: 3479902; PubMed Central PMCID: PMC1684350.

10. Schmidtke J, Krawczak M, Schwartz M, Alkan M, Bonduelle M, Buhler E, et al. Linkage relationships and allelic associations of the cystic fibrosis locus and four marker loci. Human genetics. 1987; 76 (4):337–43. PMID: 2886421.

11. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273 (5281):1516–7. PMID: 8801636.

12. Weeks DE, Sinsheimer JS. Consanguinity and relative-pair methods for linkage analysis. American journal of human genetics. 1998; 62(3):728–36. Epub 1998/04/29. PMID: 9497239; PubMed Central PMCID: PMC1376938.

13. Hoefsloot LH, Feenstra I, Kunst HP, Kremer H. Genotype phenotype correlations for hearing impairment: approaches to management. Clinical genetics. 2014; 85(6):514–23. Epub 2014/02/20. doi: 10.1111/cge.12339 PMID: 24547994.

14. Jungbluth H. Multi-minicore Disease. Orphanet journal of rare diseases. 2007; 2:31. doi: 10.1186/1750-1172-2-31 PMID: 17631035; PubMed Central PMCID: PMC1947955.

15. Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. American journal of human genetics. 2014; 94(6):809–17. Epub 2014/06/07. doi: 10.1016/j.ajhg.2014.05.003 PMID: 24906018; PubMed Central PMCID: PMC4121481.

16. Beck-Nielsen SS, Brock-Jacobsen B, Gram J, Brixen K, Jensen TK. Incidence and prevalence of nutritional and hereditary rickets in southern Denmark. European journal of endocrinology / European Federation of Endocrine Societies. 2009; 160(3):491–7. doi: 10.1530/EJE-08-0818 PMID: 19095780.

17. Ward LM, Rauch F, White KE, Filler G, Matzinger MA, Letts M, et al. Resolution of severe, adolescent-onset hypophosphatemic rickets following resection of an FGF-23-producing tumour of the distal ulna. Bone. 2004; 34(5):905–11. Epub 2004/05/04. doi: 10.1016/j.bone.2003.12.025 PMID: 15121023.

18. Wright S. Coefficients of inbreeding and relationship. Amer Natur. 1922; 56:330–8.

19. Silberstein M, Weissbrod O, Otten L, Tzemach A, Anisenia A, Shtark O, et al. A system for exact and approximate genetic linkage analysis of SNP data in large pedigrees. Bioinformatics. 2013; 29(2):197–205. doi: 10.1093/bioinformatics/bts658 PMID: 23162081; PubMed Central PMCID: PMC3546794.

20. Seelow D, Schuelke M. HomozygosityMapper2012—bridging the gap between homozygosity mapping and deep sequencing. Nucleic acids research. 2012; 40(Web Server issue):W516–20. doi: 10.1093/nar/gks487 PMID: 22669902; PubMed Central PMCID: PMC3394249.

21. Clarke NF, Kidson W, Quijano-Roy S, Estournet B, Ferreiro A, Guicheney P, et al. SEPN1: associated with congenital fiber-type disproportion and insulin resistance. Annals of neurology. 2006; 59(3):546–52. Epub 2005/12/21. doi: 10.1002/ana.20761 PMID: 16365872.

22. Moghadaszadeh B, Petit N, Jaillard C, Brockington M, Quijano Roy S, Merlini L, et al. Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. Nature genetics. 2001; 29(1):17–8. doi: 10.1038/ng713 PMID: 11528383.

23. Kremke B, Bergwitz C, Ahrens W, Schutt S, Schumacher M, Wagner V, et al. Hypophosphatemic rickets with hypercalciuria due to mutation in SLC34A3/NaPi-IIc can be masked by vitamin D deficiency and can be associated with renal calcifications. Experimental and clinical endocrinology & diabetes: official journal, German Society of Endocrinology [and] German Diabetes Association. 2009; 117 (2):49–56. doi: 10.1055/s-2008-1076716 PMID: 18523928.

24. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013; 339(6122):959–61. Epub 2013/01/26. doi: 10.1126/science.1230062 PMID: 23348503.

25. Petit N, Lescure A, Rederstorff M, Krol A, Moghadaszadeh B, Wewer UM, et al. Selenoprotein N: an endoplasmic reticulum glycoprotein with an early developmental expression pattern. Human molecular genetics. 2003; 12(9):1045–53. PMID: 12700173.

26. Tajsharghi H, Darin N, Tulinius M, Oldfors A. Early onset myopathy with a novel mutation in the Selenoprotein N gene (SEPN1). Neuromuscular disorders: NMD. 2005; 15(4):299–302. PMID: 15792869.

27. Scoto M, Cirak S, Mein R, Feng L, Manzur AY, Robb S, et al. SEPN1-related myopathies: clinical course in a large cohort of patients. Neurology. 2011; 76(24):2073–8. Epub 2011/06/15. doi: 10.1212/WNL.0b013e31821f467c PMID: 21670436.

28. Schara U, Kress W, Bonnemann CG, Breitbach-Faller N, Korenke CG, Schreiber G, et al. The phenotype and long-term follow-up in 11 patients with juvenile selenoprotein N1-related myopathy. European journal of paediatric neurology: EJPN: official journal of the European Paediatric Neurology Society. 2008; 12(3):224–30. doi: 10.1016/j.ejpn.2007.08.011 PMID: 17951086.

29. Ferreiro A, Ceuterick-de Groote C, Marks JJ, Goemans N, Schreiber G, Hanefeld F, et al. Desmin-related myopathy with Mallory body-like inclusions is caused by mutations of the selenoprotein N gene. Annals of neurology. 2004; 55(5):676–86. doi: 10.1002/ana.20077 PMID: 15122708.

30. Ferreiro A, Quijano-Roy S, Pichereau C, Moghadaszadeh B, Goemans N, Bonnemann C, et al. Mutations of the selenoprotein N gene, which is implicated in rigid spine muscular dystrophy, cause the classical phenotype of multiminicore disease: reassessing the nosology of early-onset myopathies. American journal of human genetics. 2002; 71(4):739–49. doi: 10.1086/342719 PMID: 12192640; PubMed Central PMCID: PMC378532.

31. Bergwitz C, Roslin NM, Tieder M, Loredo-Osti JC, Bastepe M, Abu-Zahra H, et al. SLC34A3 mutations in patients with hereditary hypophosphatemic rickets with hypercalciuria predict a key role for the sodium-phosphate cotransporter NaPi-IIc in maintaining phosphate homeostasis. American journal of human genetics. 2006; 78(2):179–92. doi: 10.1086/499409 PMID: 16358214; PubMed Central PMCID: PMC1380228.

32. Veilleux LN, Cheung M, Ben Amor M, Rauch F. Abnormalities in muscle density and muscle function in hypophosphatemic rickets. The Journal of clinical endocrinology and metabolism. 2012; 97(8):E1492–8. doi: 10.1210/jc.2012-1336 PMID: 22639288.

33. Bittles A. Consanguinity and its relevance to clinical genetics. Clinical genetics. 2001; 60(2):89–98. PMID: 11553039.

34. Hamamy H, Antonarakis SE, Cavalli-Sforza LL, Temtamy S, Romeo G, Kate LP, et al. Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. Genetics in medicine: official journal of the American College of Medical Genetics. 2011; 13(9):841–7. PMID: 21555946.

35. Mercuri E, Muntoni F. The ever-expanding spectrum of congenital muscular dystrophies. Annals of neurology. 2012; 72(1):9–17. Epub 2012/07/26. doi: 10.1002/ana.23548 PMID: 22829265.

36. Nigro V, Aurino S, Piluso G. Limb girdle muscular dystrophies: update on genetic diagnosis and therapeutic approaches. Current opinion in neurology. 2011; 24(5):429–36. Epub 2011/08/10. PMID: 21825984.

37. Laing NG. Genetics of neuromuscular disorders. Critical reviews in clinical laboratory sciences. 2012; 49(2):33–48. Epub 2012/04/04. doi: 10.3109/10408363.2012.658906 PMID: 22468856.

38. Nance JR, Dowling JJ, Gibbs EM, Bonnemann CG. Congenital myopathies: an update. Current neurology and neuroscience reports. 2012; 12(2):165–74. Epub 2012/03/07. doi: 10.1007/s11910-012-0255-x PMID: 22392505.