Journal of Pathology
J Pathol 2011; **225:** 12–18
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.2941

ORIGINAL PAPER

Using next-generation sequencing for the diagnosis of rare disorders: a family with retinitis pigmentosa and skeletal abnormalities

Kasmintan A Schrader,[1,2#] Alireza Heravi-Moussavi,[3#] Paula J Waters,[4] Janine Senz,[1] James Whelan,[5] Gavin Ha,[3] Patrice Eydoux,[4] Torsten Nielsen,[6] Barry Gallagher,[7] Arusha Oloumi,[3] Niki Boyd,[3] Bridget A Fernandez,[8] Terry-Lynn Young,[8] Steven JM Jones,[3] Martin Hirst,[3] Sohrab P Shah,[3] Marco A Marra,[3] Jane Green[5,8] and David G Huntsman[1]*

[1] *Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada*
[2] *Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada*
[3] *BC Cancer Agency, Vancouver, BC, Canada*
[4] *Department of Pathology and Laboratory Medicine, BC Children's Hospital, Vancouver, BC, Canada*
[5] *Department of Ophthalmology, Memorial University of Newfoundland, St. John's, NL, Canada*
[6] *Department of Anatomical Pathology, Vancouver Hospital and Health Sciences Centre, Vancouver, BC, Canada*
[7] *Department of Pathology, James Paton Memorial Regional Health Centre, Gander, NL Canada*
[8] *Discipline of Genetics, Memorial University of Newfoundland, St. John's, NL Canada*

**Correspondence to: David G Huntsman, Centre for Translational and Applied Genomics, Room 3427, BC Cancer Agency, 600 W 10th Avenue, Vancouver, BC, Canada. e-mail: dhuntsma@bccancer.bc.ca*

[#]These authors contributed equally to this study.

Abstract

Linkage analysis with subsequent candidate gene sequencing is typically used to diagnose novel inherited syndromes. It is now possible to expedite diagnosis through the sequencing of all coding regions of the genome (the exome) or full genomes. We sequenced the exomes of four members of a family presenting with spondyloepiphyseal dysplasia and retinitis pigmentosa and identified a six-base-pair (6-bp) deletion in *GNPTG*, the gene implicated in mucolipidosis type IIIγ. The diagnosis was confirmed by biochemical studies and both broadens the mucolipidosis type III phenotype and demonstrates the clinical utility of next-generation sequencing to diagnose rare genetic diseases.

Keywords: spondyloepiphyseal dysplasia; retinitis pigmentosa; mucolipidosis type III; exome; *GNPTG*; next-generation sequencing; familial

Received 25 April 2011; Revised 17 May 2011; Accepted 18 May 2011

No conflicts of interest were declared.

Introduction

Targeted and whole-genome sequencing has revolutionized the discovery of novel disease genes [1]. Eventually the low cost of sequencing and growing bioinformatic capacity will permit the routine clinical use of these tools in the rapid diagnosis of Mendelian disorders with heterogeneous phenotypes. Diagnostic algorithms used in clinical practice will evolve to reflect this, resulting in faster diagnosis, rapid expansion of disease phenotypes and better clinical management.

We describe the successful use of a next-generation sequencing approach to diagnose the genetic basis of the occurrence in an extended family of autosomal recessive spondyloepiphyseal dysplasia (SED) and retinitis pigmentosa (RP). Individual family members initially presented with spondyloepiphyseal dysplasia leading to multiple surgeries in the third to sixth decade. It was later determined that a retinitis pigmentosa phenotype, presenting in the third and fourth decades with decreased night vision and leading to significant peripheral and central vision loss or blindness by the fifth to seventh decade, co-segregated with the SED phenotype. Additionally, affected individuals had a high incidence of corneal abnormalities.

Materials and methods

Study design

The family originates from an isolated fishing community in Newfoundland, Canada (Figure 1A). Due to geographical isolation, known consanguinity and the autosomal recessive inheritance pattern, it was predicted that the causative mutation would most likely

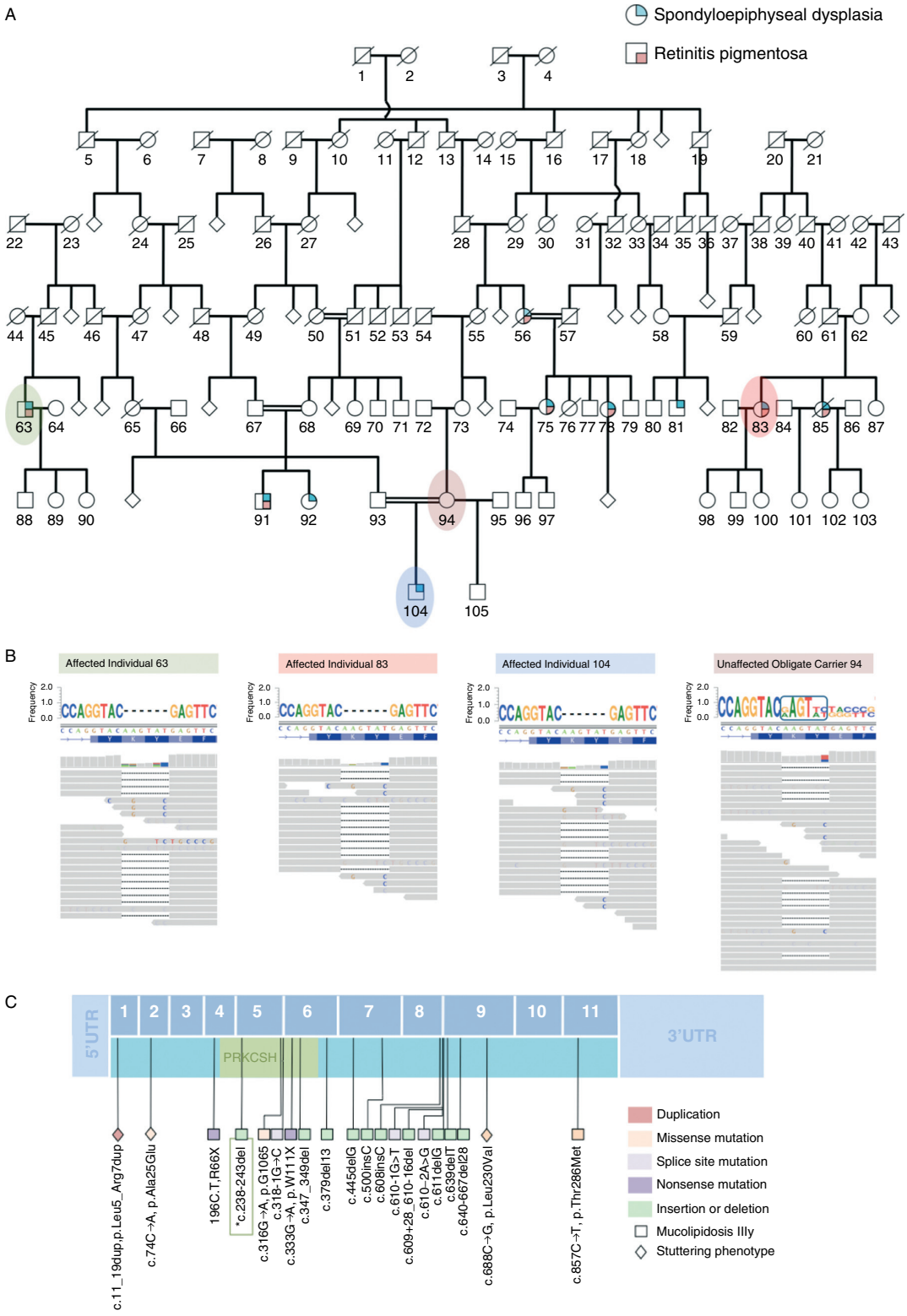

Figure 1. Identification of the family's candidate mutation by exome sequencing. (A) Pedigree of the family, indicating disease status. Individuals 63, 83, 104 and 94, shaded in green, red, blue and brown, respectively, were exome sequenced. (B) Entropy figures depicting sequence data post-filtering, using a base-calling threshold of 10 and allowing for up to three mismatches per paired-end read. Below the entropy figures are corresponding screen shots of the BAM files viewed with IGV 1.4.2, showing the reference DNA and amino acid sequence above the raw paired-end reads overlying the 6 bp deletion in *GNPTG*. (C) Mutations found in *GNPTG* and its encoded protein *N*-acetylglucosamine-1-phosphotransferase subunit-γ. The 11 exons of *GNPTG* are represented (as numbered blue boxes) above a schematic of the *N*-acetylglucosamine-1-phosphotransferase subunit γ-protein (the teal segment, with the conserved PRKCSH domain in green). Below the schematic are all of the reported nucleotide mutations, including the novel mutation reported in this manuscript (boxed and denoted by an asterisk). Nucleotide position numbering is based on starting with the A in the ATG start codon for *GNPTG* in position 1 (based on the sequence given in record number NM_032520.4 in Entrez Gene; UTR denotes untranslated region).

be novel, and would lie within an extended block of linkage that was homozygous in the affected individuals and heterozygous in the unaffected obligate carrier. We performed exome sequencing of three affected individuals (nos 63, 83 and 104) and one obligate carrier (no. 94) to look for novel variants that were consistent with this pattern of inheritance in all individuals. The research protocol was approved by the BC Cancer Agency research ethics board and the Memorial University Human Investigation Committee. All participants gave written informed consent.

Ruling out homozygous microdeletions

Affymetrix 6.0 SNP chips were also performed on the same individuals to rule out homozygous microdeletions and to confirm blocks of linkage surrounding candidate novel variants.

Exome capture, sequencing and bioinformatic analysis

Exome capture was achieved through solution hybrid selection with the Human All Exon kit SureSelect Target Enrichment System (Agilent) for Illumina Genome Analyser paired-end sequencing [2]. Two lanes of Illumina paired-end sequencing were prepared for each exome capture library. Short paired-read (75 bp) sequences obtained from the Illumina Genome Analyser were mapped to the reference human genome (NBCI Build 36.1, hg18) using MAQ (version 0.7.1) in paired-end mode [3]. The Sequence Alignment/Map (SAM 0.1.7) format was used for downstream processing. Insertion and deletion (indel) information was extracted from the alignment data using the Samtools package [4]. SNVMix, capable of distinguishing homozygous variants, was also used for the inference of SNVs [5] and was modelled on the expectation of normal Hardy–Weinberg frequencies, consistent with a diploid genome. Variants were enriched for novel non-synonymous, indel and splice-site variants, by filtering with those already present in dbSNP130, The 1000 Genomes Project, and in in-house exomes.

Biochemical confirmation of the pathogenic variant

The pathogenicity of the candidate variant in *GNPTG* was confirmed by measuring the serum enzyme activities of hexosaminidase, α-*N*-acetylglucosaminidase, α-mannosidase and β-glucuronidase, which are lysosomal enzymes known to be dependent on mannose-6-phosphate targeting. Each of the four lysosomal enzymes was assayed fluorometrically, by incubating serum together with the appropriate 4-methylumbelliferyl(4MU)–glycoside conjugate, at 37 °C and in acid pH. Reaction was stopped after a specified time period by the addition of alkaline buffer. Free 4MU, released from the substrate by enzymatic hydrolysis, was quantitated by measuring its fluorescence (excitation wavelength 350 nm, emission 440 nm). Biochemical analysis was performed on samples from four affected and three unaffected-carrier family members in parallel with seven unrelated healthy controls. For each enzyme, all serum samples were assayed simultaneously within a single batch.

Results

Filtered exome sequence data revealed only two novel variants fulfilling criteria

On average, coverage of targeted exons for > 10 reads was 87%. Filtering revealed an average of 412 novel non-synonymous variants, 242 indels and 60 splice site variants per case.

Based on our enrichment, only two novel variants appeared as homozygous in all three affected individuals and heterozygous in the obligate carrier, a non-synonymous variant in *RPL3L* and a 6-bp deletion in *GNPTG*. These were validated and found to segregate with disease in 14 further family members, using Sanger sequencing.

Both variants were located within a 3.5 Mb region of linkage containing 202 UCSC genes on chromosome (chr) 16: 482 030–3 938 263 (hg18) (Supplementary Table 1). Within the 3.5 MB interval, 134 genes out of the 202 UCSC genes had been targeted by exome capture. Coverage of targeted exons in the 3.5 MB region of linkage was in the range 46–59% and there were, on average, five novel non-synonymous variants, one indel and no splice-site variants per affected individual. Manual review of novel variants within the region revealed one further non-synonymous variant in *MSLNL* that followed the expected pattern of inheritance. This was called in more than one individual, although had not met our full criteria by filtering strategies. The remaining novel variants in the region were not shared. This was confirmed by manual review.

Confirmation of pathogenicity of the 6-bp deletion in *GNPTG*

The most promising candidate variant was the 6-bp deletion c.238-243del, p.80K_81Ydel in exon 5 of *GNPTG* (Figure 1b), an OMIM disease gene associated with mucolipidosis type IIIγ (MLIIIγ) [OMIM No. 252605]. The 6-bp deletion removes a highly conserved lysine and tyrosine residue. The deletion of each of these amino acids is predicted by Polyphen to be 'probably damaging'. Further support for pathogenicity was based on the absence of the deletion in 368 chromosomes from an ethnically matched control population.

MLIIIγ is caused by biallelic germline loss-of-function mutations in *GNPTG* (Figure 1c), which encodes the γ-subunit of GlcNAc-1-phosphotransferase [6–13], while the related disorder MLIII alpha/beta (MLIIIα/β) is caused by mutations in *GNPTAB*, which encodes the α- and β-subunits of the same phosphotransferase. The phosphotransferase complex plays a

Table 1. Lysosomal enzyme activities in serum from affected patients, carriers and unrelated normal controls

	Hexosaminidase	α-*N*-Acetylglucosaminidase	α-Mannosidase	β-Glucuronidase
Activities* in normal controls ($n = 7$)	Mean 1267 Range 743–1718	Mean 33 Range 27–40	Mean 51 Range 36–73	Mean 243 Range 129–380
Activities* in carrier individuals ($n = 3$)	Mean 1166 Range 1036–1374	Mean 36 Range 32–44	Mean 55 Range 52–63	Mean 209 Range 165–286
Activities* in affected patients ($n = 4$)	Mean 19 139 Range 16 469–21 156	Mean 415 Range 318–470	Mean 7888 Range 5850–10 016	Mean 16 454 Range 13 055–19 017
Range of fold-elevations in affected patients versus mean of normal controls	13–17-fold	10–14-fold	114–196-fold	54–78-fold

Elevations of serum lysosomal enzyme activities by at least 10-fold have been regarded as biochemical diagnostic criteria for MLIII. Comparison of serum enzyme activities by least square means (using Tukey HSD) showed statistically significant ($p < 0.0001$) elevations for all four enzymes in affected subjects versus normal control subjects and versus carriers. Serum enzyme activities in carriers did not differ significantly from normal controls. *All enzyme activities are expressed in nmol/h/ml serum.

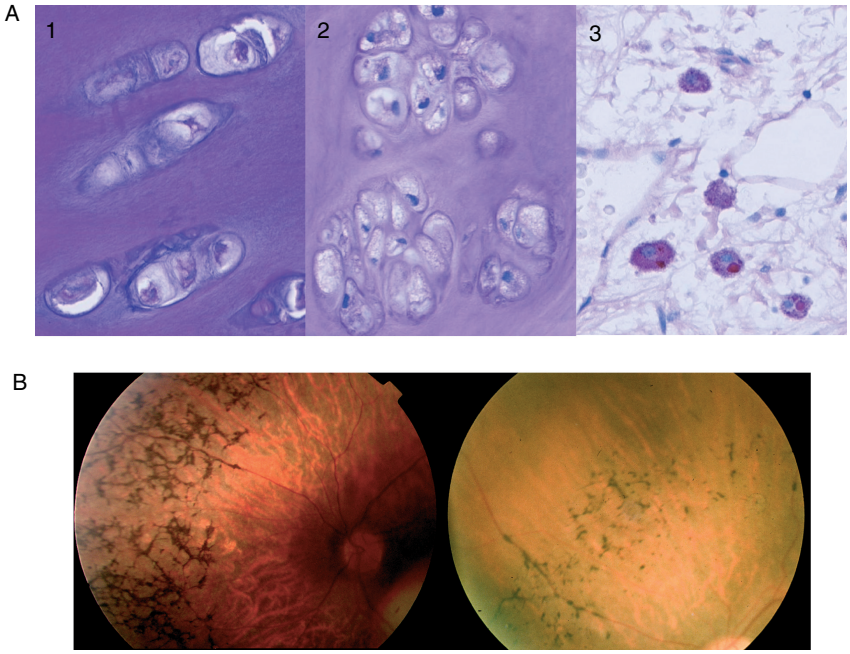

Figure 2. Demonstration of the associated bone and retinal abnormalities. (A) Panel 1, section of normal cartilage; panel 2, section of cartilage from the femoral head of individual 104. Note the pronounced microvesicular change in the chondrocytes. Of note, the appearance of the chondrocytes are in keeping with those seen in *GNPTAB*-null mice, compared with *GNPTG*-null mice, which do not display an abnormal chondrocyte morphology [15]. Panel 3, section of the femoral head of individual 104, showing macrophages in the subchondral marrow space with accumulated glycosolated intracellular material and inclusions. All sections were stained with periodic acid–Schiff with diastase digestion (PAS-D) and photographs were taken at ×400 original magnification. (B) Retinal photographs of bone spicules in individual 91 at age 28 years (left) and individual 85 at age 35 years (right).

crucial role in targeting of many enzymes to the lysosome. In both subtypes of MLIII, loss of phosphotransferase function causes mistargeting of lysosomal enzymes, resulting in abnormal accumulation of various lysosomal substrates in tissues [7,14].

Lysosomal enzyme activities were found to be markedly elevated in the serum of affected individuals compared to levels in carriers and healthy controls, indicating enzyme mistargeting and thus confirming the diagnosis of MLIIIγ (Table 1). Carriers did not demonstrate a phenotype.

Histochemical analysis of a section of the femoral head of individual 104, who had undergone joint replacement at age 23 years, revealed microvesicular change in the chondrocytes consistent with the abnormal chondrocyte morphology observed in the mucolipidosis mouse model [15] (Figure 2a).

Discussion

Mucolipidosis type III is an autosomal recessive, lysosomal storage disorder typically defined by progressive joint stiffness, scoliosis, coarse facies, mild intellectual disability, dysostosis multiplex with progressive destruction of the hip joint, increased lysosomal enzyme levels in serum, reduced lysosomal enzyme levels in cultured fibroblasts, corneal clouding and opacities [6,16]. Most published clinical descriptions of MLIII predate the genetic distinction between MLIIIγ and MLIIIα/β. As the latter is apparently more common, the specific manifestations

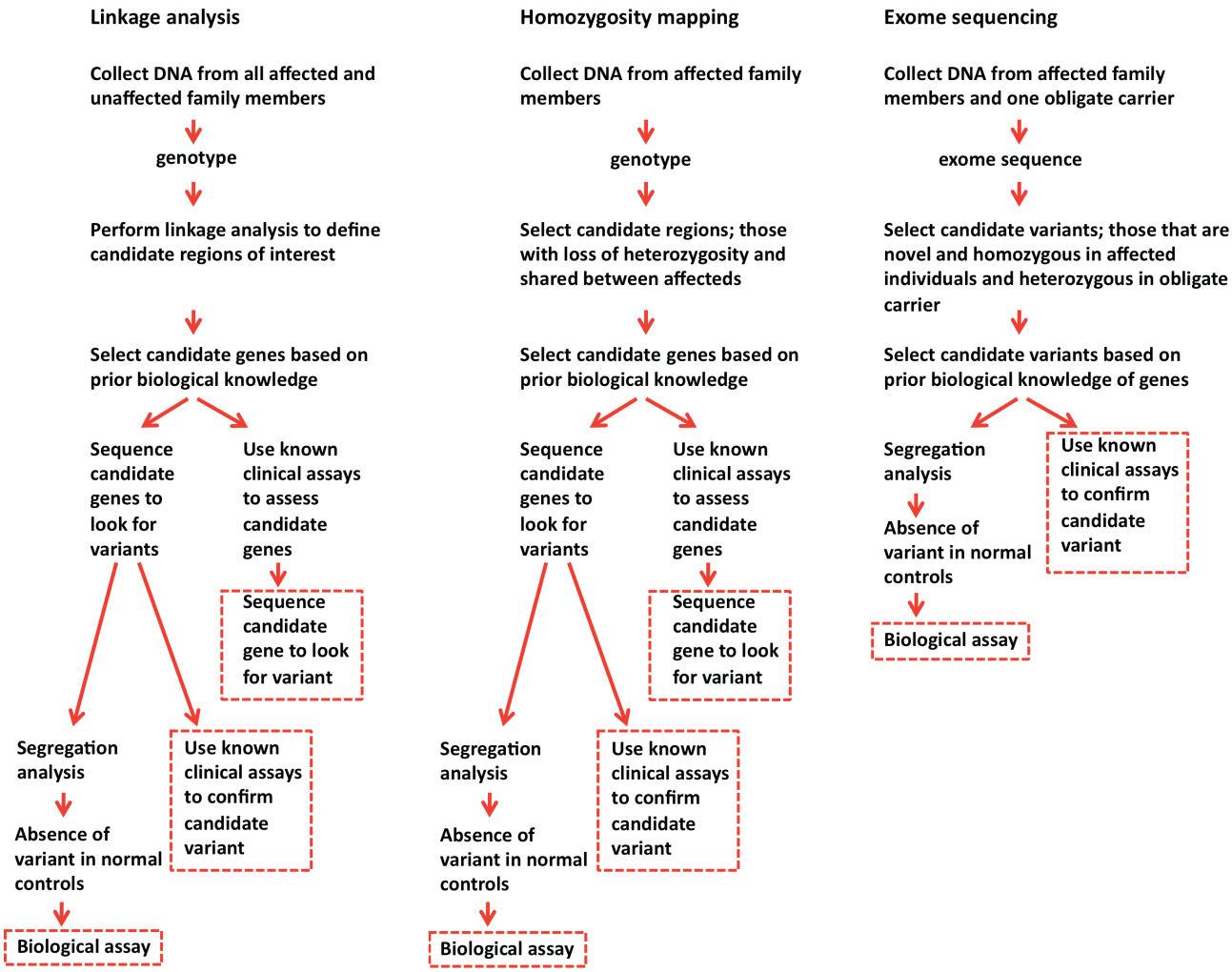

Figure 3. Expedited diagnosis of an autosomal recessive condition in a consanguinous family by next-generation sequencing. Linkage analysis, homozygosity mapping and exome sequencing are three strategies that can be used to search for disease genes. Potential end points are indicated in the red hatched boxes. Homozygosity mapping in consanguineous families can direct the investigator to candidate genes within regions of identity-by-descent (IBD). However, depending on the degree of consanguinity, there can be multiple regions of IBD and further sequencing to determine the pathogenic variant is required. The exome-sequencing approach can, in one step, provide the investigator with a list of candidate variants. Although separate genome-wide analysis of polymorphisms is still necessary to rule out deletions or duplications missed by exome sequencing, these would be detected by whole-genome sequencing. A combined homozygosity-mapping and exome-sequencing approach is also discussed in the text.

and natural history of MLIIIγ have yet to be fully defined. However, the spectrum of disease in MLIII overall is notoriously broad. Furthermore, rare variants in *GNPTAB* and *GNPTG* have recently been implicated in non-syndromic familial stuttering [17].

The degree of retinal degeneration seen in this family has not yet been reported in association with MLIII (Figure 2b), nor even in MLII, a severe infantile-onset disorder which is allelic to MLIIIα/β and involves complete ablation of phosphotransferase activity [18]. Severe retinal degeneration is seen in *GNPTAB*-null mice, implying a causal link between loss of phosphotransferase function and retinal pathology [19]. However, *GNPTG*-null mice have not been shown to develop retinal disease [15]. This family's mutation in *GNPTG* causes the deletion of a lysine and tyrosine residue in the phosphotransferase γ-subunit. This subunit is thought to have a role in regulation of phosphotransferase complex structure and function, rather than a direct role in catalysis [10]. It is possible that the presentation with severe retinal degeneration reflects specific effects of the particular mutation on phosphotransferase function.

Variable presentations of MLIII now need to be considered as a possible aetiology for other reported syndromes of spondyloepiphyseal dysplasia or spondylometaphyseal dysplasia associated with retinal manifestations [20–22]. In addition, reported findings of low neuraminidase activity in fibroblasts, but normal activity in leukocytes, in patients with spondylometaphyseal dysplasia with cone–rod dystrophy [23], suggest MLIII as the cause [24]. This broadens the phenotypic spectrum of MLIII even further.

This next-generation sequencing approach allowed a rapid molecular diagnosis of this family's complex phenotype of MLIIIγ, with a newly reported severity

of rod–cone dystrophy. This enabled a molecular diagnosis for the family and furthermore has significant clinical implications for management with regard to cardiac surveillance, anaesthetic considerations and management of potential atlanto–axial instability.

The use of exome sequencing to study four individuals from a single family allowed us to rapidly narrow the list of candidate mutations to two, and to correctly diagnose the disease gene using a standard clinical assay. This mutation could have been found by a combined approach of homozygosity mapping and exome sequencing of one affected individual, followed by segregation analysis on each novel variant in the region. However, as we had expected more false positives, our strategy was to collect multiple exome sequences and identify rapidly, without bias, novel non-synonymous, indel and splice-site variants. We could then select the true candidate variants based on their expected homozygous or heterozygous state in the affected and carrier individuals.

Traditional linkage and homozygosity mapping approaches identify regions of linkage or identity-by-descent which contain multiple candidate genes. In all approaches, ranking of candidate genes or variants usually requires some knowledge of the gene functions and possible disease mechanisms. This process can be time-consuming; however, it is expedited by next-generation sequencing approaches such as the one outlined above (Figure 3). Furthermore, as the costs of whole-genome sequencing come down to the costs of SNP arrays, sequencing of multiple individuals will become the most direct approach to the diagnosis of Mendelian disorders that are phenotypically and genetically heterogeneous.

Acknowledgment

We thank the family members who participated in this research. This work was funded in part by the Canadian Cancer Society Research Institute, the Michael Smith Foundation for Health Research, the Canadian Institutes for Health Research and Genome Canada/Genome Atlantic.

Author contributions

DH, JG and KS designed the study; KS, PW, AHM, JG and DH wrote the manuscript; JG identified the family and consented the patients; BF examined several affected family members and collected phenotypic data; AHM conducted the bioinformatic analysis; MM, SJ, MH performed the next-generation sequencing; SS and AHM designed the bioinformatic algorithms; KS, AHM, DH, PW, GH and PE analysed the data; JS, KS, MM performed molecular analysis; PW performed biochemical analysis; JG and JW diagnosed and managed the family's retinitis pigmentosa; TN and BG performed the histopathology and analysis; AO and TLY coordinated the experiment and NB helped prepare the manuscript.

References

1. Ng SB, Buckingham KJ, Lee C, *et al*. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010; **42**: 30–35.
2. Gnirke A, Melnikov A, Maguire J, *et al*. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009; **27**: 182–189.
3. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008; **18**: 1851–1858.
4. Li H, Handsaker B, Wysoker A, *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.
5. Goya R, Sun MG, Morin RD, *et al*. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* 2010; **26**: 730–736.
6. Raas-Rothschild A, Cormier-Daire V, Bao M, *et al*. Molecular basis of variant pseudo-hurler polydystrophy (mucolipidosis IIIC). *J Clin Invest* 2000; **105**: 673–681.
7. Varki AP, Reitman ML, Kornfeld S. Identification of a variant of mucolipidosis III (pseudo-Hurler polydystrophy): a catalytically active *N*-acetylglucosaminylphosphotransferase that fails to phosphorylate lysosomal enzymes. *Proc Natl Acad Sci USA* 1981; **78**: 7773–7777.
8. Tiede S, Cantz M, Raas-Rothschild A, *et al*. A novel mutation in UDP-*N*-acetylglucosamine-1-phosphotransferase γ-subunit (GNPTAG) in two siblings with mucolipidosis type III alters a used glycosylation site. *Hum Mutat* 2004; **24**: 535.
9. Raas-Rothschild A, Bargal R, Goldman O, *et al*. Genomic organisation of the UDP-*N*-acetylglucosamine-1-phosphotransferase γ-subunit (GNPTAG) and its mutations in mucolipidosis III. *J Med Genet* 2004; **41**: e52.
10. Pohl S, Tiede S, Castrichini M, *et al*. Compensatory expression of human *N*-acetylglucosaminyl-1-phosphotransferase subunits in mucolipidosis type IIIγ. *Biochim Biophys Acta* 2009; **1792**: 221–225.
11. Encarnacao M, Lacerda L, Costa R, *et al*. Molecular analysis of the *GNPTAB* and *GNPTG* genes in 13 patients with mucolipidosis type II or type III—identification of eight novel mutations. *Clin Genet* 2009; **76**: 76–84.
12. Persichetti E, Chuzhanova NA, Dardis A, *et al*. Identification and molecular characterization of six novel mutations in the UDP-*N*-acetylglucosamine-1-phosphotransferase γ-subunit (*GNPTG*) gene in patients with mucolipidosis IIIγ. *Hum Mutat* 2009; **30**: 978–984.
13. Pohl S, Encarnacao M, Castrichini M, *et al*. Loss of *N*-acetylglucosamine-1-phosphotransferase γ-subunit due to intronic mutation in *GNPTG* causes mucolipidosis type IIIγ: Implications for molecular and cellular diagnostics. *Am J Med Genet A* 2010; **152A**: 124–132.
14. Reitman ML, Varki A, Kornfeld S. Fibroblasts from patients with I-cell disease and pseudo-Hurler polydystrophy are deficient in uridine 5′-diphosphate-*N*-acetylglucosamine: glycoprotein *N*-acetylglucosaminylphosphotransferase activity. *J Clin Invest* 1981; **67**: 1574–1579.
15. Vogel P, Payne BJ, Read R, *et al*. Comparative pathology of murine mucolipidosis types II and IIIC. *Vet Pathol* 2009; **46**: 313–324.
16. Traboulsi EI, Maumenee IH. Ophthalmologic findings in mucolipidosis III (pseudo-Hurler polydystrophy). *Am J Ophthalmol* 1986; **102**: 592–597.

17. Kang C, Riazuddin S, Mundorff J, *et al*. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *N Engl J Med* 2010; **362**: 677–685.
18. Leroy J, Cathey S. Friez M. Mucolipidosis II. In: GeneReviews at GeneTests: Medical Genetics Information Resource (database online), © University of Washington, Seattle, 1997–2010 [last updated 7 July 2009]: http://www.genetests.org [accessed 12 December 2010].
19. Gelfman CM, Vogel P, Issa TM, *et al*. Mice lacking α/β-subunits of GlcNAc-1-phosphotransferase exhibit growth retardation, retinal degeneration, and secretory cell lesions. *Invest Ophthalmol Vis Sci* 2007; **48**: 5221–5228.
20. Isidor B, Le Merrer M, Ramos E, *et al*. Cone–rod dystrophy, growth hormone deficiency and spondyloepiphyseal dysplasia: report of a new case without nephronophtisis. *Am J Med Genet A* 2009; **149A**: 788–792.
21. Walters BA, Raff ML, Hoeve JV, *et al*. Spondylometaphyseal dysplasia with cone–rod dystrophy. *Am J Med Genet A* 2004; **129A**: 265–276.
22. Isidor B, Baron S, Khau van Kien P, *et al*. Axial spondylometaphyseal dysplasia: confirmation and further delineation of a new SMD with retinal dystrophy. *Am J Med Genet A* 2010; **152A**: 1550–1554.
23. Sousa SB, Russell-Eggitt I, Hall C, *et al*. Further delineation of spondylometaphyseal dysplasia with cone–rod dystrophy. *Am J Med Genet A* 2008; **146A**: 3186–3194.
24. Kuriyama M, Miyatake T, Owada M, *et al*. Neuraminidase activities in sialidosis and mucolipidosis. *J Neurol Sci* 1982; **54**: 181–187.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Table S1. Affymetrix 6.0 SNP data indicating copy number (CN) state and regions of loss of heterozygosity (LOH) in each individual.

Analysis ruled out homozygous microdeletions shared between all affected individuals and confirmed an extended region of LOH (chr16:482,030-3,938,263), shared between the affected individuals (nos. 63, 83, 104), surrounding the candidate variants.