# High incidence of unrecognized visceral/neurological late-onset Niemann-Pick disease, type C1, predicted by analysis of massively parallel sequencing data sets

Christopher A. Wassif<sup>1,2</sup>, Joanna L. Cross<sup>1,2</sup>, James Iben<sup>1</sup>, Luis Sanchez-Pulido<sup>3</sup>, Antony Cougnoux<sup>1</sup>, Frances M. Platt<sup>2</sup>, Daniel S. Ory<sup>4</sup>, Chris P. Ponting<sup>3</sup>, Joan E. Bailey-Wilson<sup>5</sup>, Leslie G. Biesecker<sup>6</sup> and Forbes D. Porter<sup>1</sup>

**Purpose:** Niemann-Pick disease type C (NPC) is a recessive, neurodegenerative, lysosomal storage disease caused by mutations in either *NPC1* or *NPC2*. The diagnosis is difficult and frequently delayed. Ascertainment is likely incomplete because of both these factors and because the full phenotypic spectrum may not have been fully delineated. Given the recent development of a bloodbased diagnostic test and the development of potential therapies, understanding the incidence of NPC and defining at-risk patient populations are important.

**Method:** We evaluated data from four large, massively parallel exome sequencing data sets. Variant sequences were identified and classified as pathogenic or nonpathogenic based on a combination of literature review and bioinformatic analysis. This methodology provided an unbiased approach to determining the allele frequency.

## **INTRODUCTION**

Niemann-Pick type C (NPC) is a lethal autosomal recessive, neurodegenerative disorder with a clinical incidence of 1:104,000.1-3 This was considered to be a minimal estimate because of incomplete ascertainment of atypical phenotypes or limitations of current diagnostic testing. NPC is caused by disruption of either NPC1 or NPC2; mutations of NPC1 account for 95% of patients.<sup>1,2</sup> Loss of function of either NPC1 or NPC2 results in the accumulation of unesterified cholesterol and glycosphingolipids within the late endosome/lysosome of all cells. Although the clinical presentation and progression of NPC is a continuous spectrum, patients can be classified into four general categories based on age at neurological onset. These categories are early infantile, late infantile, juvenile onset, and adolescent/adult onset.1 In the early infantile, late infantile, and juvenile forms of the disease patients may initially present with neonatal cholestasis or hepatosplenomegaly. A small subset of patients with NPC die of systemic liver disease, usually during the neonatal period.<sup>1</sup> However, in the majority of NPC patients **Results:** Our data suggest an incidence rate for NPC1 and NPC2 of 1/92,104 and 1/2,858,998, respectively. Evaluation of common *NPC1* variants, however, suggests that there may be a late-onset NPC1 phenotype with a markedly higher incidence, on the order of 1/19,000-1/36,000.

**Conclusion:** We determined a combined incidence of classical NPC of 1/89,229, or 1.12 affected patients per 100,000 conceptions, but predict incomplete ascertainment of a late-onset phenotype of NPC1. This finding strongly supports the need for increased screening of potential patients.

Genet Med advance online publication 12 March 2015

**Key Words:** allele frequency; next-generation sequence study; Niemann-Pick disease; NPC; type C

the liver disease frequently resolves, but neurological signs and symptoms follow.<sup>1,2</sup> Neurological symptoms are insidious and heterogeneous in nature, often initially manifesting in a non-specific manner (e.g., clumsiness or difficulty with school work) but commonly progress to include variable degrees of cerebellar ataxia, vertical supranuclear gaze palsy, gelastic cataplexy, seizures, and dementia. These neurological manifestations are invariably progressive<sup>4,5</sup> and ultimately result in death.

The current diagnosis of NPC is based on filipin staining of unesterified cholesterol in cultured fibroblasts or molecular testing. Filipin staining requires a skin biopsy, is performed in only a few specialized diagnostic laboratories worldwide, and is not always conclusive. Molecular testing of *NPC1* and *NPC2* is also available; however, molecular testing in practice also has weaknesses. It is currently still inconclusive in 12–15% of the cases because of unknown pathogenicity of the changes, the lack of study of allele segregation, and the existence of one (possibly two) unidentified mutant allele. Combined with the frequently nonspecific and insidious nature of the neurological

Submitted 27 September 2014; accepted 21 January 2015; advance online publication 12 March 2015. doi:10.1038/gim.2015.25

The first two authors contributed equally to this work.

<sup>&</sup>lt;sup>1</sup>Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA; <sup>2</sup>Department of Pharmacology, University of Oxford, Oxford, UK; <sup>3</sup>Department of Physiology, Anatomy and Genetics, MRC Functional Genomics Unit, University of Oxford, Oxford, UK; <sup>4</sup>Diabetic Cardiovascular Disease Center, Washington University School of Medicine, St Louis, Missouri, USA; <sup>8</sup>Statistical Genetics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, USA; <sup>6</sup>Clinical Genomics Section, Awasif (wassifc@mail.nih.goy)

disease onset, the difficulty of diagnosis contributes to a diagnostic delay on the order of 4–5 years<sup>2</sup> for the late infantile and juvenile forms of the disease. The diagnostic delay in the adolescent/adult-onset form is likely greater, but the full extent of that delay cannot be determined because of a limited number of reported cases. A sensitive blood-based diagnostic test, which detects elevated oxysterols, was recently developed; this test could be used to screen potential patients economically and rapidly.<sup>6</sup>

A number of therapies for NPC are actively being developed. Miglustat, a glycosphingolipid synthesis inhibitor, although not approved in the United States for treatment of NPC1, has been approved in the European Union and other countries for the treatment of NPC. 2-Hydroxypropyl- $\beta$ -cyclodextrin has shown significant promise in both mouse and feline models of NPC1 (C. Vite, personal communication) and is currently in a phase I/II trial (NCT01747135) at the National Institutes of Health (NIH). The development of 2-hydroxypropyl- $\beta$ -cyclodextrin for NPC1 has been reviewed by Ottinger et al.<sup>7</sup> Other potential therapies under development include histone deacetylase inhibitors,<sup>8-10</sup> heat shock protein 70 (unpublished data F.M.P.), and  $\delta$ -tocopherol.<sup>11</sup> Given the rapid development of potential therapeutic interventions, it is critical that the incidence of NPC and its full clinical spectrum be fully defined.

An increasing number of patients with adult-onset NPC are being reported.<sup>1,12-14</sup> Psychiatric symptoms can be prominent,<sup>12-15</sup> although affected adults without neurological manifestations have also been reported.<sup>16-18</sup> The full phenotypic spectrum of adult-onset NPC disease has yet to be delineated. This led us to question whether the incidence of NPC might be greater than previous clinical estimates because of incomplete ascertainment. To estimate the incidence of NPC in a manner that is independent of clinical recognition of cases, we sought to determine a pathogenic carrier frequency of NPC1 and NPC2 variants using data from four independent, massively parallel exome sequencing projects, or next-generation sequencing projects. Our data indicate that the classical incidence of NPC likely occurs at the clinically predicted rate of ~1:90,000 and suggest that there may be a late-onset phenotype or variant form with an incidence potentially as high as 1:19,000-36,000.

## MATERIALS AND METHODS

We recently reported the determination of the pathogenic allele frequency of the 7-dehydrocholesterol reductase gene.<sup>19</sup> We used a similar approach for the determination of the variant frequency in NPC.

### Data sets

Four large, independent, massively parallel exome sequencing projects, or next-generation sequencing projects, were used. These data sets are the National Heart, Lung, and Blood Institute GO Exome Sequencing Project (ESP),<sup>20</sup> version 3 release of the 1000 Genomes Project,<sup>21</sup> ClinSeq,<sup>22</sup> and a database from an NIH interinstitute collaboration on autism (primary investigators: F.D.P., J.E.B.-W., E Tierney, and A Thurm). ESP contributed a

maximum number of 13,006 chromosomes, the 1000 Genomes Project contributed 2,184 chromosomes, ClinSeq contributed 1,902 chromosomes, and the NIH interinstitute collaboration on autism project contributed 662 chromosomes. Thus, a maximum total of 17,754 chromosomes were analyzed, and this number was used as the denominator in total frequency calculations. None of these data sets included patients evaluated for NPC, nor did we identify any individuals with two pathogenic mutations, so we considered them to be unbiased with respect to variation in *NPC1* and *NPC2*.

### Determination of variant calls and annotation

Variant calls were downloaded for regions overlapping *NPC1* and *NPC2*, by Perl script, for every base of the coding exons plus or minus five base pairs of exon sequence, when available. Mutations were annotated using SNPnexus<sup>23</sup> with Refseq annotations<sup>24</sup>; pathogenicity predictions were performed using PolyPhen-2,<sup>25</sup> SIFT,<sup>26</sup> and Mutation assessor.<sup>27</sup> Intronic variations detected within five bases of intron–exon boundaries were analyzed by MaxEntScan.<sup>28</sup> Untranslated region variations were excluded from the analysis of these data sets.

### Determination of pathogenicity of the variant call

Determination of the pathogenicity of a variant allele was a multistep process that used both bioinformatic tools and manual curation. We began by comparing the variants found in the data sets against the professional version of the Human Gene Mutation Database (HGMD)<sup>29</sup> and the existing database of 78 patients with NPC1 in the NIH Cohort (primary investigator: F.D.P.) to determine which variants had been previously identified in patients known to have NPC. Because inclusion in HGMD does not require identification in a patient, the primary literature was reviewed to determine the nature and manner in which the variations were detected. Variants were mapped onto known protein tertiary structures as part of the bioinformatic approach, identifying variable to conserved residues and possible interactions (**Figures 1** and **2**). The variant NPC2 protein was modeled using I-TASSER.<sup>30</sup>

Single coding nucleotide variants were interrogated in silico by three different predictive software packages: PolyPhen-2,25 SIFT,<sup>26</sup> and Mutation assessor.<sup>27</sup> PolyPhen-2 provides a predicted assignment of "benign," "possibly damaging," or "probably damaging," as well as a false discovery rate for each single coding nucleotide variant call. For the determination of the pathogenesis of a single coding nucleotide variant, PolyPhen-2 calls of possibly damaging or probably damaging were considered pathogenic. SIFT uses the same terminology as PolyPhen-2, and the same approach was used. Mutation assessor has four predictive determinants: predictive nonfunctional low, nonfunctional neutral, functional (medium), and functional (high). Mutation assessor predictions of functional (medium) and functional (high) where considered pathogenic. When the three predictive algorithms were discrepant and no published data supporting pathogenicity were available, we accepted the prediction of two of the three programs. Potential Prediction of high incidence of visceral/neurological late-onset NPC1 | WASSIF et al

splice variants were processed in MaxEntScan<sup>28</sup> to provide a predictive determination of the variant's affect on splicing; these were reported as "strongly negative," "negative," or "neutral." Potential splice variants that were classified as negative or strongly negative were considered to be pathogenic. All pathogenic variants were assumed to be fully penetrant.



Figure 1 Mapping of the coding variants onto the known structure of NPC2. Probably damaging mutations are labeled with red circles. The human NPC2 structural model (from positions 20 to 149) was created using Modeller based on the bovine NPC2 structure (PDB:2HKA).<sup>37</sup> The human NPC2 ribbon is colored according to evolutionary conservation using ConSurf server.<sup>38,39</sup> Cholesterol sulfate (from PDB:2HKA)<sup>37</sup> is shown in sticks.  $\beta$  strands are labeled (A–G).

## **ORIGINAL RESEARCH ARTICLE**

## Determination of predicted disease incidence

Once potential pathogenic variants were identified and a carrier frequency determined, the predicted disease incidence was calculated assuming a Hardy-Weinberg equilibrium. For this estimate we assumed that all pathogenic variants were fully penetrant. The Hardy-Weinberg equilibrium model also assumes that allelic variation is at equilibrium and thus is not undergoing active selective pressure. Given that NPC1 is a receptor for filoviruses and its association with body mass, an assumption of neutral selection may not be correct. However, Al-Daghri et al.31 concluded selective pressure on NPC1 in humans is weak to neutral. We made the assumption that allelic frequencies were consistent across different ethnic groups represented in our data set. The potential error making this assumption is greatest for the ESP cohort, given that it includes a large number of individuals of either European or African descent. We evaluated our data for reduction of heterozygosity resulting from ethnic difference (the Wahlund effect) by determining a weighted frequency; however, only negligible changes were observed for any of the NPC1 or NPC2 pathogenic alleles (data not shown). Given the negligible effect the weighted frequencies were not applied to carrier frequency calculations.

# Cloning and sequence analysis of the c.441+1G>A variant discovered in NPC2

Two heterozygous Epstein-Barr virus-transformed lymphoblast cell lines for the c.441+1G>A variant, NIMH 42 and NIMH 77, were identified in the NIH interinstitute collaboration on autism. These two lines and one control line were grown under standard growth conditions in 15% fetal bovine serum in RPMI (Life Technologies, Foster City, CA) for 3 days. Cell pellets were isolated and messenger RNA isolated per the manufacturer's protocol (Qiagen, Gaithersburg, MD). Forward primer NPC2-F3 5'-GGTGGAGTGGCAACTTCAGG-3' and reverse primer NPC2-R2 5'-CACTGGATACCATTGGAGAGC-3' were used to reverse transcribe the messenger RNA using the Superscript III One-step RT-PCR System (Life Technologies).



**Figure 2 Mapping human N-terminal domain (NTD)** *NPC1* **mutants.** Probably and possibly damaging mutations are labeled with red circles. The human NTD *NPC1* (PDB:3GKI)<sup>40</sup> ribbon was colored according to evolutionary conservation using the ConSurf server.<sup>39,40</sup> Cholesterol is shown in sticks. None of the NTD *NPC1* mutants is located at residues that interact with cholesterol.

The complementary DNA was visualized on a 1.5% agarose gel. One band was observed for Control and two for NIMH 42 and NIMH 77. All bands were gel purified and cloned into the TOPO TA Cloning Kit for Sequencing (Life Technologies). Isolated colonies were grown overnight in Luria broth-ampicillin, and plasmid DNA was isolated (Qiagen). Sequencing was performed on a 3500xL Genetic Analyzer (Life Technologies) using a BigDye sequencing kit, per the manufacturer's protocol.

## RESULTS

Analysis of exomic sequence data from 17,754 chromosomes compared with the human reference sequence for *NPC1* and *NPC2* led to the identification of 16,455 and 271 nonsynonymous sequence variants in *NPC1* and *NPC2*, respectively. The 16,455 variants identified in *NPC1* were comprised of 147 distinct variants that included 129 coding single-nucleotide base variants, 9 splice-site changes, and 9 insertions/deletions (indels; **Supplementary Table S1** online). The 271 nonsynonymous variants identified in *NPC2* included 14 distinct changes consisting of 12 coding single-nucleotide base variants and 2 splice-site changes (**Table 1**).

The HGMD<sup>29</sup> was queried to establish what observed variants in this data set might be pathogenic. For NPC1 (Supplementary Table S1 online) and NPC2 (Table 1), 33 (32 pathogenic variants and 1 benign variant) of 147 variants (22.4%) and 5 of 14 variants (35.7%), respectively, had previously been reported in HGMD. One additional novel NPC1 variant, c.2524T>C (p.F842L), was present in the NIH cohort (Supplementary Table S1 online). The combination of PolyPhen-2, SIFT, and Mutation assessor classified 53 NPC1 and 8 NPC2 coding nucleotide variants as pathogenic based on our criteria. Of the predicted pathogenic variants, 27 (51%) and 6 (75%) have not been reported in HGMD for NPC1 and NPC2, respectively. PolyPhen-2 also calculates a false discovery rate. For predicted NPC1 and NPC2 variants, the average false discovery rate for a prediction of probably or possibly damaging were 0.04% and 0.03%, respectively. These low mean false discovery rates had a negligible effect on the carrier incidence estimate and thus were not applied to either NPC1 or NPC2 carrier frequency calculations.

For *NPC1* and *NPC2*, two of nine and one of two potential splice mutations, respectively, were predicted to be pathogenic. Of the nine indels identified in *NPC1*, a two-base pair deletion, c.2020\_2021del, was observed 319 times only in the ESP data set and thus was removed as a technical artifact unique to the ESP data set. The eight other *NPC1* indels result in a frameshift and thus were considered pathogenic. No indels were identified in *NPC2*.

Based on the above analysis, for *NPC1* we initially considered the 68 distinct variants meeting the criteria of pathogenic (54 identified by predictive software to be pathogenic, 4 indicated by the predictive software as benign but known to be pathogenic, the 2 splice variants, and the 8 indels). This accounted for 371 pathogenic alleles, with an estimated carrier rate of 2.09% (371/17,754) and a predicted NPC incidence of 1/9,160. Given the order of magnitude difference between this number and clinical estimates, this prediction is likely a significant overestimation. Thus we applied manual curation to the NPC1 data set. Four variants-c.665A>G (p.N222S), c.1532C>T (p.T511M), c.2882A>G (p.N961S), and c.3598A>G (p.S1200G)-accounted for 254 of the 371 predicted pathogenic alleles (68%). Allelic frequencies for these four alleles were 0.400, 0.287, 0.389, and 0.355%, respectively. Given that their individual allelic frequencies exceed by more than a factor of 10 the allelic frequency of p.I1061T (0.028%), the most commonly reported mutant allele in patients with mutations in NPC1 (Supplementary Table S1 online), it is not plausible that these alleles are associated with classical NPC disease. Excluding these four high-frequency variants based on this assertion left 117 pathogenic alleles, or a 0.659% (117/17,754) carrier rate. This carrier rate predicts an incidence of NPC attributable to NPC1 of 1/92,104.

We further evaluated the decision to exclude the four highfrequency alleles based on lack of an association with classical NPC disease. Although all three predictive packages indicate both p.N222S and p.N961S to be nonpathogenic, these two variants have been reported in "visceral-only" or adult-onset NPC1 cases. The p.N222S variant was reported in combination with a p.I1061T mutation in a single patient (35 years old) with adult-onset NPC with variant filipin staining.32 This patient initially presented with visceral disease (hepatosplenomegaly) and later manifested ataxia at 44 years of age. We identified a p.N222S variant in combination with c.1402T>G, (p.C468G) in teenage sisters diagnosed based on splenomegaly. The second allele in this sibling pair, p.C468G, is predicted by PolyPhen-2 to be probably damaging. Pathological analysis of the spleen in the older sibling was suggestive of Niemann-Pick disease, but filipin staining was inconclusive. Neurological symptoms were absent and signs were very minor, with deep tendon hyperreflexia and minor auditory brainstem response abnormalities noted upon evaluation at 15 and 13 years of age, respectively. NIH severity score for both was 1.4 Plasma oxysterol concentrations were consistent with a diagnosis of NPC in these two subjects. Mapping of p.N222S to the known tertiary structure provided no additional evidence for the pathogenicity of this residue (Figure 2). The p.N961S (c.2882A>C) variant has been reported in a compound heterozygous state with p.S666N (c.1997G>A, with a PolyPhen-2 prediction of probably damaging) in an adult case with subclinical hepatosplenomegaly and lymphadenopathy noted on autopsy following death caused by acute pulmonary embolism and myocardial infarction.<sup>16</sup> Although no neurological symptoms were reported, brain pathology was notable for distended neurons with increased lipofuscin granules. Assuming one or both of these variants are pathogenic, fully penetrant, and associated with late-onset NPC disease, the total disease incidence of NPC1 would range from 1/19,077 to 1/36,420.

Although predicted to be probably damaging by PolyPhen-2, neither p.T511M nor p.S1200G have been reported in NPC1 patients. Millat et al.<sup>33</sup> reported p.T511M as a novel nonpathological coding single-nucleotide variant.

Table 1 The 2	71 distinct	t variants dete	cted in NPC2								
			Polyphen-2/ MaxEntScan/		Mutation assessor	Total	Total		1000		
cDNA	Protein	rs no.	published	SIFT	prediction	alleles	variants	NHLBI	Genome	Clinseq	Autism
c.38T>C	p.L13P	rs147602717	Probably damaging	Possibly damaging	Predicted functional (medium)	17,360	Ŋ	4	0	<del>~ -</del>	0
c.49G>A	p.A17T	rs145302203	Benign	Benign	Predicted nonfunctional (low)	17,748	-	-	0	0	0
c.56C>A	p.A19D	rs369392502	Probably damaging	Probably damaging	Predicted functional (medium)	17,748	←	<del>~</del>	0	0	0
c.58G>T	p.E20X	rs80358260	Published	Probably damaging	Predicted functional (medium)	17,750	-	<del>~</del>	0	0	0
c.88G>A	p.V30M	rs151220873	Possibly damaging/ published	Possibly damaging	Predicted nonfunctional (low)	17,252	34	25	-	4	4
c.115G>A	p.V39M	rs80358261	Probably damaging/ published	Probably damaging	Predicted functional (medium)	17,754	-	-	0	0	0
c.212A>G	p.K71R	rs142075589	Possibly damaging	Benign	Predicted nonfunctional (neutral)	17,754	m	2	←	0	0
c.224A>T	p.H75L	rs369221608	Probably damaging		Predicted functional (medium)	17,754	-	-	0	0	0
c.271G>A	p.D91N	rs148607507	Possibly damaging	Possibly damaging	Predicted nonfunctional (low)	17,754	10	Ø	0	2	0
c.278G>T	p.C93F	rs143960270	Probably damaging/ published		Predicted functional (high)	17,754	-	-	0	0	0
c.292A>C	p.N98H	rs142858704	Benign	Benign	Predicted nonfunctional (low)	17,754	12	11	~	0	0
c.340C>G	p.P114A	rs371363324	Probably damaging		Predicted functional (high)	17,754	-	-	0	0	0
c.441+1G>A	Intronic	rs140130028	Strong negative/ published			17,752	96	83	0	9	2
c.442-4A>C	Intronic	rs114950106	Neutral			17,754	104	104	0	0	0

0.017%

0.006%

0.006%

0.056%

0.006%

0.068%

#### Prediction of high incidence of visceral/neurological late-onset NPC1 | WASSIF et al

0.006%

0.006%

0.006%

0.197%

0.029%

Rate

## **ORIGINAL RESEARCH ARTICLE**

0.586%

Each variant has a corresponding complementary DNA (cDNA) number; protein change, and reference single-nucleotide polymorphism "rs" number, when available. The majority have been assigned either a PolyPhen-2, SIFT, Mutation assessor, or MaxEntScan score, as well; variants that have been previously published are noted. Variants considered nonpathogenic are shaded gray. The number of alleles analyzed for each variant and the total number of times the variant was detected are noted in conjunction with the frequency of each variant in each of the four data sets and the carrier rate for each variant.

0.541%

0.006%

NHLBI, National Heart, Lung, and Blood Institute.

WASSIF et al | Prediction of high incidence of visceral/neurological late-onset NPC1

The p.S1200G variant was reported in an "NPC uncertain" case in the recent ZOOM study.<sup>14</sup> This subject, patient 5 in the study, was a compound heterozygote for p.V664M, a known *NPC1* mutation, but plasma cholestane- $3\beta$ , $5\alpha$ , $6\beta$ -triol testing<sup>6</sup> was negative. Current data do not support classification of either p.T511M or p.S1200G variants as pathogenic alleles.

Sequence analysis of NPC2 (Table 1) identified 151 potential pathogenic alleles and calculated a pathogenic carrier frequency of 0.85% (151/17,754). Again, the predicted disease incidence (1/55,297) did not seem to be realistic unless one proposed an extreme degree of underascertainment. Thus we similarly applied manual curation to the NPC2 data set. Review of the NPC2 data identified two high-frequency variants that dominated the frequency calculation: c.441+1G>A and c.88G>A (p.V30M); both variants are reported in the HMGD. The splice variant c.441+1G>C was predicted to be "strongly negative" by MaxEntScan. Molecular analysis of independent cell lines revealed multiple splicing events. The most prominent errant splicing event results in the insertion of 16 bases, which leads to the alteration of the terminal 4 amino acids and the addition of 86 additional amino acids to the protein (Supplementary Figure S1 online). Multiple lines of evidence strongly indicate that this errant splicing results in a functional protein. First, the variant has not been reported in association with a patient with NPC. Second, modeling of the variant protein using I-TASSER<sup>30</sup> found no alterations to the cholesterol binding pocket or stability of the protein (data not shown). Finally, Huang et al.<sup>34</sup> demonstrated that generation of an NPC2 fusion protein with mCherry fused to the carboxy-terminal end of the protein is fully functional and is able to correct the NPC cellular phenotype in Npc2<sup>-/-</sup> mouse embryonic fibroblasts. As such, we excluded c.441+1G>C as a pathogenic allele. The p.V30M variant, with a allelic frequency of 0.197%, is predicted to be possibly damaging by PolyPhen-2 and SIFT but is considered nonpathogenic by Mutation assessor. The one reported subject with NPC with the p.V30M variant was classified as having a phenotypic NPC variant, a second mutation was not identified, and near normal levels of cholesterol esterification were reported in skin fibroblasts.35 Inclusion of the p.V30M allele predicts a disease incidence of NPC attributable to NPC2 of 1/402,400 and that NPC2 should account for 18.6% of patients with NPC. This latter prediction conflicts with clinical data indicating that NPC2 accounts for only 2-5% of all patients with NPC.<sup>1,2</sup> Sequence alignment and structural analyses demonstrate that the p.V30 residue is not evolutionarily conserved and is present in a structurally variable region of the NPC2 protein well away from its binding pocket (Figure 1). Furthermore, p.V30M occurs a higher frequency than any known pathogenic NPC2 allele; this, coupled with the lack of evidence supporting functional importance and ultimately the lack of any clinical correlation, has lead us to exclude p.V30M as a pathogenic allele. We are, therefore, left with 21 pathogenic alleles (0.118% carrier frequency) and a predicted disease incidence for NPC2 of 1/2,858,998 conceptions.

Based on the above analysis of both *NPC1* and *NPC2*, the combined incidence is predicted to be 1/89,229, or 1.12 cases per 100,000 conceptions, and the fraction of NPC2 cases is predicted to be 3.1%. The predicted number of cases is slightly more than the 0.96 cases per 100,000 conceptions reported by Vanier<sup>2</sup> when she accounted for prenatal cases, and the fraction of NPC2 cases is consistent with prior clinical observation of 2-5%.<sup>1-3</sup>

## DISCUSSION

The impact of NPC1 variation on human health may be significant. Work by multiple groups has demonstrated that c.644A>G (p.H215R) is associated with obesity.<sup>36</sup> In this analysis of NPC1 variants, we identified the p.H215R variant in almost a third of the NPC1 alleles. Our work now demonstrates that two relatively common NPC1 variants, with a combined carrier frequency approaching 0.8%, may contribute, in a compound heterozygous state, to a late-onset NPC1 phenotype for which the phenotypic spectrum and clinical significance remains to be defined. This late-onset NPC1 phenotype may represent a milder manifestation of NPC1 deficiency with predominately visceral manifestations. The degree to which this late-onset NPC1 phenotype is associated with high-frequency NPC1 alleles and the adult-onset NPC1 phenotype that includes significant neurological and psychological symptoms also remains to be defined.

Failure to ascertain certain alleles in patients, such as the p.V30M in *NPC2* or the p.T511M and p.S1200G in *NPC1*, could be due to prenatal lethality; however, because NPC is an autosomal recessive disorder, it is difficult to hypothesize a plausible mechanism, such as a dominant inhibitory function, by which these alleles would uniquely result in prenatal lethality.

Based on clinical case reports, one needs to consider the possibility that p.N222S and p.N961S maybe pathogenic, with allelic frequencies of 0.400 and 0.389%, respectively. The evidence for clinical relevance is strongest for p.N222S, which has been observed in two independent cases with similar visceral and delayed neurological manifestations, variant filipin staining in fibroblasts, and positive plasma oxysterol testing in siblings. Assuming pathogenicity is related to a compound heterozygous state and full penetrance, the combined frequency of p.N222S with another pathogenic NPC1 mutation would be 1/35,667. Although only limited data are available, if one includes p.N961S based on a single report with no supporting diagnostic testing, the incidence of a lateonset variant of NPC1 disease would increase to 1/19,077. Another possible explanation for these high predicted incidences is that some individuals harboring these variants, either in combination with another pathogenic allele or in the homozygous state, may be asymptomatic or manifest only subclinical signs.

Leveraging existing "whole-exome sequence" data, we estimated the disease incidence of NPC using both bioinformatic tools and manual curation. With respect to classical NPC Prediction of high incidence of visceral/neurological late-onset NPC1 | WASSIF et al

disease, we estimate that the incidence of NPC1 and NPC2 are on the order of 1/92,000 and 1/2,900,000, respectively, with a combined incidence of ~1/89,000. These estimates are in agreement with previous clinical estimates. Thus, our data do not support significant underascertainment of classical NPC cases. Concurrence with clinical data also suggests that we are not missing a significant number of alleles, such as large indels or intronic mutations that are not detected by wholeexome sequencing. However, our data suggest that there may be significant underascertainment of a late-onset NPC1 phenotype. This late-onset phenotype may present as visceral-only or neurological mild NPC1 and have a potential incidence of 1/19,000–1/36,000. Further work is necessary to fully delineate this late-onset NPC1 phenotype, but this study suggests that NPC should be considered in individuals with visceral lipidosis or unexplained neurological and psychiatric symptoms.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

### DISCLOSURE

L. G. B. is an uncompensated adviser to the Illumina Corporation and receives royalties from the Genentech Corporation. The other authors declare no conflict of interest.

## ACKNOWLEDGMENTS

J.L.C., J.I., C.L., A.T., S.S., J.E.B.-W., L.G.B., F.D.P., and C.A.W. are supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Human Genome Research Institute intramural research programs, Department of Health and Human Services. C.A.W. is an NIH/ Oxford Scholar. J.L.C. is a Wellcome Trust/NIH Scholar. F.M.P. is a Royal Society Wolfson Research Merit Award holder. L.S.P. and C.P.P. are funded by the MRC.

### REFERENCES

- Patterson MC, Hendriksz CJ, Walterfang M, Sedel F, Vanier MT, Wijburg F; NP-C Guidelines Working Group. Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. *Mol Genet Metab* 2012;106:330–344.
- 2. Vanier MT. Niemann-Pick disease type C. Orphanet J Rare Dis 2010;5:16.
- Jahnova H, Dvorakova L, Vlaskova H, et al. Observational, retrospective study of a large cohort of patients with Niemann-Pick disease type C in the Czech Republic: a surprisingly stable diagnostic rate spanning almost 40 years. Orphanet J Rare Dis 2014;9:140.
- Yanjanin NM, Vélez JI, Gropman A, et al. Linear clinical progression, independent of age of onset, in Niemann-Pick disease, type C. Am J Med Genet B Neuropsychiatr Genet 2010;153B:132–140.
- te Vruchte D, Speak AO, Wallom KL, et al. Relative acidic compartment volume as a lysosomal storage disorder-associated biomarker. J Clin Invest 2014;124:1320–1328.
- Porter FD, Scherrer DE, Lanier MH, et al. Cholesterol oxidation products are sensitive and specific blood-based biomarkers for Niemann-Pick C1 disease. Sci Transl Med 2010;2:56ra81.
- Ottinger EA, Kao ML, Carrillo-Carrasco N, et al. Collaborative development of 2-hydroxypropyl-β-cyclodextrin for the treatment of Niemann-Pick type C1 disease. *Curr Top Med Chem* 2014;14:330–339.
- 8. Kim SJ, Lee BH, Lee YS, Kang KS. Defective cholesterol traffic and neuronal differentiation in neural stem cells of Niemann-Pick type C disease improved

## **ORIGINAL RESEARCH ARTICLE**

by valproic acid, a histone deacetylase inhibitor. *Biochem Biophys Res Commun* 2007;360:593–599.

- Munkacsi AB, Chen FW, Brinkman MA, et al. An "exacerbate-reverse" strategy in yeast identifies histone deacetylase inhibition as a correction for cholesterol and sphingolipid transport defects in human Niemann-Pick type C disease. *J Biol Chem* 2011;286:23842–23851.
- 10. Pipalia NH, Cosner CC, Huang A, et al. Histone deacetylase inhibitor treatment dramatically reduces cholesterol accumulation in Niemann-Pick type C1 mutant human fibroblasts. *Proc Natl Acad Sci USA* 2011;108:5620–5625.
- Xu M, Liu K, Swaroop M, et al. d-Tocopherol reduces lipid accumulation in Niemann-Pick type C1 and Wolman cholesterol storage disorders. *J Biol Chem* 2012;287:39349–39360.
- Zech M, Nübling G, Castrop F, et al. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. *PLoS ONE* 2013;8:e82879.
- Schicks J, Müller Vom Hagen J, Bauer P, et al. Niemann-Pick type C is frequent in adult ataxia with cognitive decline and vertical gaze palsy. *Neurology* 2013;80:1169–1170.
- Bauer P, Balding DJ, Klünemann HH, et al. Genetic screening for Niemann-Pick disease type C in adults with neurological and psychiatric symptoms: findings from the ZOOM study. *Hum Mol Genet* 2013;22:4349–4356.
- Bonnot O, Klünemann HH, Sedel F, Tordjman S, Cohen D, Walterfang M. Diagnostic and treatment implications of psychosis secondary to treatable metabolic disorders in adults: a systematic review. Orphanet J Rare Dis 2014;9:65.
- Dvorakova L, Sikora J, Hrebicek M, et al. Subclinical course of adult visceral Niemann-Pick type C1 disease. A rare or underdiagnosed disorder? J Inherit Metab Dis 2006;29:591.
- 17. Fensom AH, Grant AR, Steinberg SJ, et al. An adult with a non-neuronopathic form of Niemann-Pick C disease. *J Inherit Metab Dis* 1999;22:84–86.
- Fröhlich E, Harzer K, Heller T, Rühl U. [Ultrasound echogenic splenic tumors: nodular manifestation of type C Niemann-Pick disease]. Ultraschall Med 1990;11:119–122.
- Cross JL, Iben J, Simpson CL, et al. (2014) Determination of the allelic frequency in Smith-Lemli-Opitz syndrome by analysis of massively parallel sequencing data sets. *Clin Genet*. doi: 10.1111/cge.12425 (Epub ahead of print 9 May 2014).
- 20. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). Seattle, WA. http://evs.gs.washington.edu/EVS/
- Abecasis GR, Auton A, Brooks LD, et al.; 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
- Biesecker LG, Mullikin JC, Facio FM, et al.; NISC Comparative Sequencing Program. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res* 2009;19:1665–1674.
- Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics* 2009;25:655–661.
- 24. Haas D, Garbade SF, Vohwinkel C, et al. Effects of cholesterol and simvastatin treatment in patients with Smith-Lemli-Opitz syndrome (SLOS). J Inherit Metab Dis 2007;30:375–387.
- 25. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4: 1073–1081.
- Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011;39:e118.
- 28. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–394.
- 29. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: 2008 update. *Genome Med* 2009;1:13.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;5:725–738.
- Al-Daghri NM, Cagliani R, Forni D, et al. Mammalian NPC1 genes may undergo positive selection and human polymorphisms associate with type 2 diabetes. BMC Med 2012;10:140.
- Tängemo C, Weber D, Theiss S, Mengel E, Runz H. Niemann-Pick Type C disease: characterizing lipid levels in patients with variant lysosomal cholesterol storage. *J Lipid Res* 2011;52:813–825.

WASSIF et al | Prediction of high incidence of visceral/neurological late-onset NPC1

- Millat G, Baïlo N, Molinero S, Rodriguez C, Chikh K, Vanier MT. Niemann-Pick C disease: use of denaturing high performance liquid chromatography for the detection of NPC1 and NPC2 genetic variations and impact on management of patients and families. *Mol Genet Metab* 2005;86:220–232.
- Huang L, Pike D, Sleat DE, Nanda V, Lobel P. Potential pitfalls and solutions for use of fluorescent fusion proteins to study the lysosome. *PLoS ONE* 2014;9:e88893.
- Park WD, O'Brien JF, Lundquist PA, et al. Identification of 58 novel mutations in Niemann-Pick disease type C: correlation with biochemical phenotype and importance of PTC1-like domains in NPC1. *Hum Mutat* 2003;22:313–325.
- Meyre D, Delplanque J, Chèvre JC, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 2009;41:157–159.
- Xu S, Benoff B, Liou HL, Lobel P, Stock AM. Structural basis of sterol binding by NPC2, a lysosomal protein deficient in Niemann-Pick type C2 disease. J Biol Chem 2007;282:23525–23531.
- Landau M, Mayrose I, Rosenberg Y, et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 2005;33(Web Server issue):W299–W302.
- Glaser F, Rosenberg Y, Kessel A, Pupko T, Ben-Tal N. The ConSurf-HSSP database: the mapping of evolutionary conservation among homologs onto PDB structures. *Proteins* 2005;58:610–617.
- Kwon HJ, Abi-Mosleh L, Wang ML, et al. Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. *Cell* 2009;137:1213–1224.