Published in final edited form as:
Genet Med. 2016 January ; 18(1): 41–48. doi:10.1038/gim.2015.25.

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[#1,2], **Joanna L. Cross**[#1,2], **James Iben**[1], **Luis Sanchez-Pulido**[3], **Antony Cougnoux**[1], **Frances M. Platt**[2], **Daniel S. Ory**[4], **Chris P. Ponting**[3], **Joan E. Bailey-Wilson**[5], **Leslie G. Biesecker**[6], and **Forbes D. Porter**[1]

[1]Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, US

[2]Department of Pharmacology, University of Oxford, Oxford OX1 3QT, United Kingdom

[3]MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, OX1 3PT, UK

[4]Diabetic Cardiovascular Disease Center, Washington University School of Medicine, St. Louis, MO, USA

[5]Statistical Genetics Section, Computational and Statistical Genomics Branch, Department of Health and Human Services, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

[6]Clinical Genomics Section, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA

[#] These authors contributed equally to this work.

Abstract

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 due to both these factors and that the full phenotypic spectrum may not have been fully delineated. Given the recent development of a blood-based diagnostic test and development of potential therapies, it is important to understand the incidence of NPC and to define at risk patient populations.

Corresponding Author: Christopher A. Wassif M.S., Senior Technical Specialist/Research Associate, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Candidate for the degree of Doctorate in Philosophy, University of Oxford, 9000 Rockville Pike Building CRC Room 1E-3288, Bethesda, MD 20892-1830, (301) 496-6664, (301) 480-5791 (fax).

L. G. B. is an uncompensated advisor to the Illumina Corporation and receives royalties from the Genentech Corporation; all other authors report no conflict of interest.

Europe PMC Funders Author Manuscripts

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

Results—Our data suggests an incidence rate for NPC1 and NPC2 of 1/92,104 and 1/2,858,998, respectively. However, evaluation of common *NPC1* variants, suggests that there may be a late-onset NPC1phenotype with a markedly higher incidence on the order of 1/20,000-39,000.

Conclusions—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.

Keywords

Next Generation sequence study; Niemann-Pick disease; type C; NPC; Allele frequency

Introduction

Niemann-Pick Type C (NPC) is an autosomal recessive, neurodegenerative lethal disorder with a clinical incidence of 1:104,000 [1-3]. This was considered to be a minimal estimate due to incomplete ascertainment of atypical phenotypes or limitations of current diagnostic testing. NPC is caused by disruption of either *NPC1*or *NPC2* with mutations of *NPC1* accounting for 95% of patients [1; 2]. 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 of neurological onset. These categories are early-infantile, late-infantile, juvenile 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 NPC patients die of systemic liver disease usually during the neonatal period [1]. However in the majority of NPC patients the liver disease frequently resolves, but neurological signs and symptoms follow [1; 2]. 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 [4; 5] and ultimately result in death.

The current diagnosis of NPC is based upon 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, lack of study of allele segregation, existence of one (possibly 2) unidentified mutant allele. Combined with the frequently

Europe PMC Funders Author Manuscripts

nonspecific and insidious nature of the neurological disease onset, the difficulty of diagnosis contributes to a diagnostic delay on the order of 4-5 years [2] for the late infantile and juvenile forms of the disease. The diagnostic delay in the adolescent/adult-onset is likely greater but the full extent of that delay cannot be determined due to a limited number of reported cases. Recently a sensitive blood-based diagnostic test, which detects elevated oxysterols, has been developed and this blood-based test could economically and rapidly be used to screen potential patients[6].

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-β-cyclodextrin (HP-β-CD) has shown significant promise in both mouse and feline (Charles Vite personal communication) models of NPC1 and is currently in a phase 1/2 trial (NCT01747135) at the NIH. The development of HP-β-CD for NPC1 has been reviewed by Ottinger *et al.*[7] Other potential therapies under development include HDAC inhibitors [8-10], HSP70 (F. Platt), and delta-tocopherol[11]. 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 adult-onset NPC patients are being reported [1; 12-14]. Psychiatric symptoms can be prominent [12-15] although affected adults without neurological manifestations have also been reported [16-18]. 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 due to 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 utilizing data from four independent massively parallel exome sequencing projects, or next generation sequencing projects. Our data indicates that the classical incidence of NPC likely occurs at the clinically predicted rate of approximately 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.

Material and Methods

We have recently reported the determination of the pathogenic allele frequency of the 7-dehydrocholesterol reductase gene (DHCR7)[19]. We utilized 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 utilized. These data sets are the NHLBI GO Exome Sequencing Project (ESP) [20], V3 release of the 1000 Genomes Project [21], ClinSeq® [22], and a database from a NIH inter-institute collaboration on Autism (PIs: FD Porter, J Bailey-Wilson, E Tierney, A. Thurm). ESP contributed a maximum number of 13,006 chromosomes, 1000 Genome Project contributed 2,184 chromosomes, ClinSeq® contributed 1,902 chromosomes and the NIH inter-institute collaboration on Autism project contributed 662 chromosomes.

Thus, a maximum total of 17,754 chromosomes were analyzed and this number was utilized as the denominator in total frequency calculations. None of these datasets 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/minus 5 base pairs of exon sequence when available. Mutations were annotated using SNPnexus [23] using Refseq annotations [24], pathogenicity predictions were performed using Polyphen-2, [25] SIFT,[26] Mutation assessor [27]. Intronic variations detected within 5 bases of intron exon boundaries were analyzed by MaxEntScan [28]. Untranslated regions 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 utilized 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®) [29] and the existing database of 78 patients with NPC1 in the NIH Cohort (PI: FD Porter) to determine which variants had been previously identified in patients known to have NPC. Since inclusion in HGMD does not require identification in a patient, 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). Modeling of the variant NPC2 protein was performed using I-TASSER[30].

Single coding nucleotide variants were interrogated *in silico* by three different predictive software packages Polyphen-2, [25] SIFT, [26] Mutation assessor [27]. Polyphen-2 provides a predicted assignment of “Benign”, “Possibly Damaging”, or “Probably Damaging” as well as a false discovery rate (FDR) for each single coding nucleotide variant call. For the determination of the pathogenesis of a single coding nucleotide variants 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 non-functional low, non-functional 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 was available, we accepted the prediction of two of the three programs. Potential splice variants were processed in MaxEntScan [28] to provide a predictive determination of the variants 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.

Europe PMC Funders Author Manuscripts

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 (HWE). For this estimate we assumed that all pathogenic variants were fully penetrant. The HWE model also assumes that allelic variation is at equilibrium and thus 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 dataset. The potential error making this assumption is greatest for the ESP cohort given that it includes large number of individuals of either European or African descent. We evaluated our data for reduction of heterozygosity due to ethnic difference (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 inter-institute collaboration on Autism. These 2 lines and one control line were grown under standard growth conditions15% fetal bovine serum in RPMI (Life Technologies) for 3 days. Cell pellets were isolated and mRNA isolated as per manufactures protocol (Qiagen). Forward primer NPC2-F3 5′-GGTGGAGTGGCAACTTCAGG-3′ and Reverse primer NPC2-R2 5′-CACTGGATACCATTGGAGAGC-3′ were used to reverse transcribe the mRNA using Superscript III One-step RT-PCR System (Life Technologies). The cDNA was visualized on a 1.5% agarose gel. One band was observed for WT 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 LB-ampicillin and plasmid DNA was isolated (Qiagen). Sequencing was performed on a 3500×L Genetic Analyzer (Life Technologies) using BigDye sequencing kit as per manufacturer’s protocol.

Results

Analysis of exomic sequence data from 17,754 chromosomes as compared to 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 (Table 1). 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 2).

The Human Gene Mutation Database HGMD® [29] was queried to establishing what observed variants in this data set might be pathogenic. For *NPC1* (Table 1) and *NPC2* (Table 2), 33 (32 pathogenic and one benign variant) of 147 (22.4%) and 5 out of 14 (35.7%) variants, respectively, had previously been reported in HGMD®. One additional novel *NPC1* variant, c.2524T>C (p.F842L), was present in the NIH cohort (Table 1). The combination of Polyphen-2, SIFT, and Mutation assessor classified 53 *NPC1* and 8 *NPC2* coding nucleotide variants as pathogenic by 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 (FDR). For *NPC1* and *NPC2* predicted variants, the average FDR for a prediction of "Probably" or "Possibly" damaging were 0.04% and 0.03% respectively. These low mean FDRs 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*, 2 out of 9 and 1 of 2, potential splice mutations were predicted to be pathogenic. Of the nine insertion/deletions (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 insertion/deletions). 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 out of the 371 (68%) predicted pathogenic alleles. Allelic frequencies for these four alleles were 0.400, 0.287, 0.389 and 0.355 percent respectively. Given that their individual allelic frequencies exceed the allelic frequency of p.I1061T (0.028%), the most commonly reported mutant allele in patients with mutations in *NPC1*, by more than a factor of 10 (Table 1), 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 high frequency 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 non-pathogenic 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 adult onset (35 yr) patient with variant filipin staining [32]. This patient initially presented with visceral disease (hepatosplenomegaly) and later manifested ataxia at 44 years of age. We have 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 sib 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 on 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 hepatosplenamegaly and lymphadenopathy noted on autopsy following death due to acute pulmonary embolism and myocardial infarction [16]. 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 1/19,077-1/36,420.

Although predicted to be probably damaging by Polyphen-2, neither p.T511M nor p.S1200G have been reported in NPC1 patients. Millat et al [33] reported p.T511M as a novel nonpathological coding single nucleotide variant. The p.S1200G variant was reported in an "NPC uncertain" case in the recent ZOOM study [14]. This subject, patient 5 in the ZOOM study, was a compound heterozygote for p.V664M, a known *NPC1* mutation, but plasma cholestane-3β,5α,6β-triol testing[6] was negative. Current data do not support classification of either p.T511M or p.S1200G variants as pathogenic alleles.

Sequence analysis of *NPC2* (Table 2) 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 appear to be realistic unless one proposed an extreme degree of under-ascertainment. 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 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 that leads to the alteration of the terminal 4 amino acids and the addition of 86 additional amino acids to the protein (supplemental figure1). 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[30] found no alterations to the cholesterol binding pocket or stability of the protein (data not shown). Finally, Huang et al. have 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*$^{-/-}$ mouse embryonic fibroblasts[34]. As such we have 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 considered non-pathogenic by mutation assessor. The one reported NPC subject with the p.V30M variant was classified as a phenotypic NPC variant, a second mutation was not identified and near normal levels of cholesterol esterification was 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 account for only 2-5% of all patients with NPC [1; 2]. 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 is found at a higher frequency than any known pathogenic *NPC2* allele, 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 of 1/2,858,998 conceptions for NPC2.

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[2] when she accounted for prenatal cases and the fraction of NPC2 cases is consistent with prior clinical observation of 2-5%[1-3].

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[36]. 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 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, as 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 percent 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 the 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 late onset variant of NPC1 disease would increase to 1/19,077. Another possible to explanation of 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 only manifest subclinical signs.

Europe PMC Funders Author Manuscripts

Leveraging existing "whole exome sequence" data we have estimated the disease incidence of NPC utilizing both bioinformatic tools and manual curation. With respect to classical NPC 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 approximately 1/89,000. These estimates are in agreement with previous clinical estimates. Thus, our data does not support significant under-ascertainment 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 "whole exome sequencing." However, our data suggests that there may be significant under-ascertainment of a late-onset NPC1 phenotype. This late-onset phenotype may present as visceral-only or neurological mild NPC1, and with a potential incidence of 1/19,000-1/39,000. Further work is necessary to fully delineate this late-onset NPC1 phenotype, but the current study suggests that NPC should be considered in individuals with visceral lipidosis or unexplained neurological and psychiatric symptoms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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. supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development or the National Human Genome Research Institute intramural research programs, DHHS. J.L.C. C.A.W. is a NIH/Oxford Scholar and J.L.C is a Wellcome Trust/NIH Scholar. F.M.P. is a Royal Society Wolfson Research Merit Award holder.

References

1. Patterson MC, Hendriksz CJ, Walterfang M, Sedel F, Vanier MT, Wijburg F, Group N-CGW. Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. Molecular genetics and metabolism. 2012; 106:330–344. [PubMed: 22572546]
2. Vanier MT. Niemann-Pick disease type C. Orphanet journal of rare diseases. 2010; 5:16. [PubMed: 20525256]
3. Jahnova H, Dvorakova L, Vlaskova H, Hulkova H, Poupetova H, Hrebicek M, Jesina P. 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 journal of rare diseases. 2014; 9:140. [PubMed: 25236789]
4. Yanjanin NM, Velez JI, Gropman A, King K, Bianconi SE, Conley SK, Brewer CC, Solomon B, Pavan WJ, Arcos-Burgos M, 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. [PubMed: 19415691]
5. te Vruchte D, Speak AO, Wallom KL, Al Eisa N, Smith DA, Hendriksz CJ, Simmons L, Lachmann RH, Cousins A, Hartung R, et al. Relative acidic compartment volume as a lysosomal storage disorder-associated biomarker. Journal of Clinical Investigation. 2014; 124:1320–1328. [PubMed: 24487591]
6. Porter FD, Scherrer DE, Lanier MH, Langmade SJ, Molugu V, Gale SE, Olzeski D, Sidhu R, Dietzen DJ, Fu R, et al. Cholesterol oxidation products are sensitive and specific blood-based biomarkers for Niemann-Pick C1 disease. Science translational medicine. 2010; 2:56ra81.
7. Ottinger EA, Kao ML, Carrillo-Carrasco N, Yanjanin N, Shankar RK, Janssen M, Brewster M, Scott I, Xu X, Cradock J, et al. Collaborative development of 2-hydroxypropyl-beta-cyclodextrin for the treatment of Niemann-Pick type C1 disease. Current topics in medicinal chemistry. 2014; 14:330–339. [PubMed: 24283970]

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 by valproic acid, a histone deacetylase inhibitor. Biochem Biophys Res Commun. 2007; 360:593–599. [PubMed: 17624314]
9. Munkacsi AB, Chen FW, Brinkman MA, Higaki K, Gutierrez GD, Chaudhari J, Layer JV, Tong A, Bard M, Boone C, 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. [PubMed: 21489983]
10. Pipalia NH, Cosner CC, Huang A, Chatterjee A, Bourbon P, Farley N, Helquist P, Wiest O, Maxfield FR. Histone deacetylase inhibitor treatment dramatically reduces cholesterol accumulation in Niemann-Pick type C1 mutant human fibroblasts. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:5620–5625. [PubMed: 21436030]
11. Xu M, Liu K, Swaroop M, Porter FD, Sidhu R, Firnkes S, Ory DS, Marugan JJ, Xiao J, Southall N, et al. delta-Tocopherol reduces lipid accumulation in Niemann-Pick type C1 and Wolman cholesterol storage disorders. J Biol Chem. 2012; 287:39349–39360. [PubMed: 23035117]
12. Zech M, Nubling G, Castrop F, Jochim A, Schulte EC, Mollenhauer B, Lichtner P, Peters A, Gieger C, Marquardt T, et al. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PloS one. 2013; 8:e82879. [PubMed: 24386122]
13. Schicks J, Muller Vom Hagen J, Bauer P, Beck-Wodl S, Biskup S, Krageloh-Mann I, Schols L, Synofzik M. Niemann-Pick type C is frequent in adult ataxia with cognitive decline and vertical gaze palsy. Neurology. 2013; 80:1169–1170. [PubMed: 23427322]
14. Bauer P, Balding DJ, Klunemann HH, Linden DE, Ory DS, Pineda M, Priller J, Sedel F, Muller A, Chadha-Boreham H, et al. Genetic screening for Niemann-Pick disease type C in adults with neurological and psychiatric symptoms: findings from the ZOOM study. Human molecular genetics. 2013; 22:4349–4356. [PubMed: 23773996]
15. Bonnot O, Klunemann 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 journal of rare diseases. 2014; 9:65. [PubMed: 24775716]
16. Dvorakova L, Sikora J, Hrebicek M, Hulkova H, Bouckova M, Stolnaja L, Elleder M. Subclinical course of adult visceral Niemann-Pick type C1 disease. A rare or underdiagnosed disorder? Journal of inherited metabolic disease. 2006; 29:591. [PubMed: 16802107]
17. Fensom AH, Grant AR, Steinberg SJ, Ward CP, Lake BD, Logan EC, Hulman G. An adult with a non-neuronopathic form of Niemann-Pick C disease. Journal of inherited metabolic disease. 1999; 22:84–86. [PubMed: 10070623]
18. Frohlich E, Harzer K, Heller T, Ruhl U. Ultrasound echogenic splenic tumors: nodular manifestation of type C Niemann-Pick disease. Ultraschall in der Medizin. 1990; 11:119–122. [PubMed: 2200110]
19. Cross JL, Iben J, Simpson C, Thurm A, Swedo S, Tierney E, Bailey-Wilson J, Biesecker LG, Porter FD, Wassif CA. Determination of the Allelic Frequency in Smith-Lemli-Opitz Syndrome by Analysis of Massively Parallel Sequencing Data Sets. Clinical genetics. 2014
20. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP). Seattle, WA:
21. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
22. Biesecker LG, Mullikin JC, Facio FM, Turner C, Cherukuri PF, Blakesley RW, Bouffard GG, Chines PS, Cruz P, Hansen NF, et al. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. Genome research. 2009; 19:1665–1674. [PubMed: 19602640]
23. 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. [PubMed: 19098027]
24. Haas D, Garbade SF, Vohwinkel C, Muschol N, Trefz FK, Penzien JM, Zschocke J, Hoffmann GF, Burgard P. Effects of cholesterol and simvastatin treatment in patients with Smith-Lemli-Opitz syndrome (SLOS). Journal of inherited metabolic disease. 2007; 30:375–387. [PubMed: 17497248]

25. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7:248–249. [PubMed: 20354512]

26. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009; 4:1073–1081. [PubMed: 19561590]

27. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic acids research. 2011; 39:e118. [PubMed: 21727090]

28. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Journal of computational biology: a journal of computational molecular cell biology. 2004; 11:377–394. [PubMed: 15285897]

29. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. The Human Gene Mutation Database: 2008 update. Genome medicine. 2009; 1:13. [PubMed: 19348700]

30. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010; 5:725–738. [PubMed: 20360767]

31. Al-Daghri NM, Cagliani R, Forni D, Alokail MS, Pozzoli U, Alkharfy KM, Sabico S, Clerici M, Sironi M. Mammalian NPC1 genes may undergo positive selection and human polymorphisms associate with type 2 diabetes. BMC medicine. 2012; 10:140. [PubMed: 23153210]

32. Tangemo C, Weber D, Theiss S, Mengel E, Runz H. Niemann-Pick Type C disease: characterizing lipid levels in patients with variant lysosomal cholesterol storage. Journal of lipid research. 2011; 52:813–825. [PubMed: 21245028]

33. Millat G, Bailo 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. Molecular genetics and metabolism. 2005; 86:220–232. [PubMed: 16126423]

34. 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.

35. Park WD, O'Brien JF, Lundquist PA, Kraft DL, Vockley CW, Karnes PS, Patterson MC, Snow K. Identification of 58 novel mutations in Niemann-Pick disease type C: correlation with biochemical phenotype and importance of PTC1-like domains in NPC1. Human mutation. 2003; 22:313–325. [PubMed: 12955717]

36. Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S, Durand E, Vatin V, Degraeve F, Proenca C, 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. [PubMed: 19151714]

37. 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. [PubMed: 17573352]

38. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic acids research. 2005; 33:W299–W302. [PubMed: 15980475]

39. 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. [PubMed: 15614759]

40. Kwon HJ, Abi-Mosleh L, Wang ML, Deisenhofer J, Goldstein JL, Brown MS, Infante RE. Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. Cell. 2009; 137:1213–1224. [PubMed: 19563754]

Europe PMC Funders Author Manuscripts

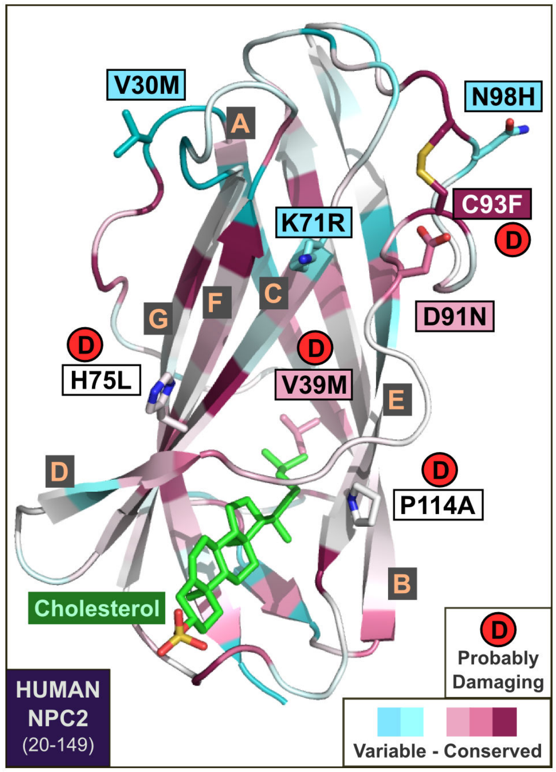

Figure 1.

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) [37]. Human NPC2 ribbon is colored according to evolutionary conservation using ConSurf server[38; 39]. Cholesterol sulfate (from PDB:2HKA)[37] is shown in sticks. Beta strands are labeled (A to G).

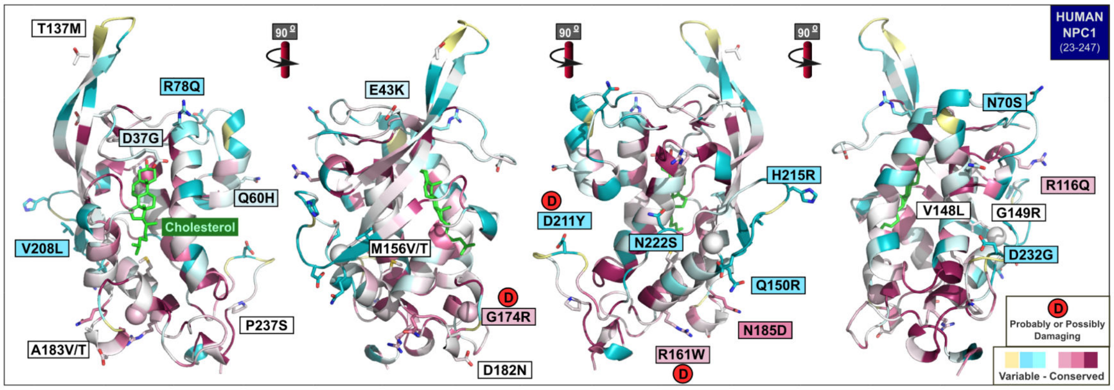

Figure 2.

Mapping human N-terminal domain (NTD)-NPC1 mutants. Probably and possibly damaging mutations are labeled with red circles. The human NTD-NPC1 (PDB:3GKI)[40] ribbon was colored according to evolutionary conservation using the ConSurf server [39; 40]. Cholesterol is shown in sticks. None of the NTD-NPC1 mutants is located at cholesterol interacting residues.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

This table summarizes the 16,455 distinct variants detected in *NPC1*. Each variant has a corresponding cDNA nucleotide number, protein change, and reference SNP “RS” number when available. The majority have been assigned either a Polyphen-2, SIFT, Mutation assessor, or MaxEntScan scores, as well variants that have been previously published are noted. Variants considered non-pathogenic are shaded grey. 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. The asterisk indicates the one novel variant detected in the NIH patient.

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.110A>G	p.D37G		BENIGN	BENIGN	Predicted non-functional (low)	17754	1	0	0	1	0	0.006%
c.127G>A	p.E43K	rs138277307	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.180G>T	p.Q60H	rs145566943	BENIGN	POSSIBLY DAMAGING	Predicted functional (medium)	17754	5	4	1	0	0	0.028%
c.181-4A>C	intronic	rs374571310	Neutral			17754	2	2	0	0	0	0.011%
c.181-3A>G	intronic	rs371126954	Strong Negative Effect			17754	1	1	0	0	0	0.006%
c.209A>G	p.N70S	rs200291759	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	0	1	0	0	0.006%
c.233G>A	p.R78Q	rs373274825	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.346C>T	p.R116X	rs144973225	PROBABLY DAMAGING/Published			17754	1	1	0	0	0	0.006%
c.347G>A	p.R116Q	rs140952850	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	2	2	0	0	0	0.011%
c.410C>T	p.T137M	rs372947142	BENIGN/Published	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.424_425insGA	p.K142Rfs					17266	1	1	0	0	0	0.006%
c.442G>C	p.V148L	rs200323346	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	2	0	2	0	0	0.011%
c.445G>A	p.G149R	rs143205855	BENIGN	BENIGN	Predicted non-functional (low)	17754	2	1	1	0	0	0.011%
c.449A>G	p.Q150R	rs37594D577	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.466A>G	p.M156V	rs149074243	BENIGN	BENIGN	Predicted non-functional (low)	17754	6	6	0	0	0	0.034%
c.467T>C	p.M156T	rs147615070	BENIGN	BENIGN	Predicted non-functional (low)	17734	2	1	0	1	0	0.011%
c.481C>T	p.R161W	rs141243713	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	2	0	0	0	0.011%
c.520G>C	P.G174R	rs37009B528	PROBABLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.544G>A	p.D182H	rs201021988	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	2	1	1	0	0	0.011%
c.547G>A	p.A183T	rs111256741	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	7	4	3	0	0	0.039%
c.548C>T	P.A183V	rs192963719	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	0	1	0	0	0.006%
c.553A>G	p.N185D	rs139485263	BENIGN	POSSIBLY DAMAGING	Predicted functional (medium)	17752	4	3	0	1	0	0.023%
c.622G>C	p.V208L	rs372416248	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.631G>T	p.D211Y	rs367851289	PROBABLY DAMAGING	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.644A>G	p.H215R	rs1805081	BENIGN/Known polymorphism	BENIGN	Predicted non-functional (low)	17696	5250	3849	535	646	220	29.668%
c.665A>G	p.N222S	rs55680026	BENIGN/Published	BENIGN	Predicted non-functional (low)	17748	71	59	2	9	1	0.400%
c.688_693del	p.S230_V231del		Published			17226	3	3	0	0	0	0.017%
c.695A>G	p.D232G	rs201956601	BENIGN	BENIGN	Predicted functional (medium)	17754	1	0	1	0	0	0.006%
c.709C>T	p.P237S	rs80358251	BENIGN/Published benign	BENIGN	Predicted non-functional (low)	17730	183	150	14	15	4	1.032%
c.749A>C	p.K250T		BENIGN	BENIGN	Predicted non-functional (low)	17754	1	0	0	0	1	0.006%
cn.763C>T	p.P255S	rs373815982	POSSIBLY DAMAGING	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.764C>G	p.P255R	rs3710239B3	PROBABLY DAMAGING	BENIGN	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.769C>T	p.P257S	rs368776731	BENIGN	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.782C>T	p.T261M	rs374169117	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.797A>G	p.D266G	rs370188327	POSSIBLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.806A>G	p.Y269C		POSSIBLY DAMAGING	BENIGN	Predicted non-functional (neutral)	17746	1	0	0	1	0	0.006%
c.811A>G	p.1271V	rs370810779	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.841C>T	p.L281F	rs377132020	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.873G>T	p.W291C	rs138151007	BENIGN			17754	10	7	2	1	0	0.056%
c.901G>A	p.E301K	rs150154006	POSSIBLY DAMAGING	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.962C>T	p.A321V	rs138079168	BENIGN	BENIGN	Predicted non-functional (low)	17750	2	1	1	0	0	0.011%
c.979G>A	p.V3271	rs141361998	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.1001G>C	p.C334S	rs199693280	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	0	0	1	0	0.006%
c.1010G>A	p.R337Q	rs373390781	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
C.1022G>C	p.R341P	rs370181667	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.1039G>A	p.V3471	rs376741451	BENIGN	BENIGN	Predicted non-functional (low)	17754	2	2	0	0	0	0.011%
c.1055G>T	p.C352F	rs149020783	BENIGN	BENIGN	Predicted non-functional (low)	17754	2	2	0	0	0	0.011%
c.1094C>T	p.S365L	rs200243024	POSSIBLY DAMAGING			17754	2	0	0	2	0	0.011%
c.1115G>A	p.R372Q	rs150053420	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.1166G>A	p.R389H	rs373751051	POSSIBLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.1208T>C	p.F403S	rs371234970	PROBABLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.1211G>A	p.R404Q	rs139751448	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (high)	17754	1	1	0	0	0	0.006%
c.1232G>A	p.R411Q	rs77080672	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	34	26	5	1	2	0.192%

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.1270C>G	p.P424A	rs143797098	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.1274C>T	p.S425L	rs140149624	BENIGN			17754	1	1	0	0	0	0.006%
c.1300c>T	p.P434S	rs61731962	BENIGN/Known polymorphism	BENIGN	Predicted non-functional (low)	17752	212	189	18	3	2	1.194%
c.1346C>T	p.A449V	rs372289265	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.1348A>G	p.1450V	rs141892620	BENIGN	BENIGN	Predicted non-functional (low)	17752	5	3	0	2	0	0.028%
c.1367C>T	p.5456F	rs374159264	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.1412C>T	p.P471L	rs201226297	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	0	1	0	0	0.006%
c.1421C>T	p.P4741	rs372445155	PROBABLY DAMAGING/Published	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.1472G>A	p.S491N	rs37075B521	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.1480G>A	p.V494M	rs199B12609	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.1506C>G	p.D502E	rs191537721	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	0	1	0	0	0.006%
c.1532C>T	p.T511M	rs13381670	PROBABLY DAMAGING			17754	51	42	8	0	1	0.287%
c.1549G>A	p.V517l	rs201791992	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	0	1	0	0	0.006%
c.1552C>T	p.R518W	rs377515417	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	2	0	0	0	0.011%
c.1561G>T	p.A521S	rs138184115	BENIGN/Published	BENIGN	Predicted non-functional (neutral)	17754	9	9	0	0	0	0.051%
c.1628C>T	p.P543L	rs369368181	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.1672G>T	p.A558S	rs201156397	POSSIBLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17704	1	0	0	1	0	0.006%
c.1756G>A	p.E586K	rs369753548	BENIGN	BENIGN	Predicted non-functional (low)	17754	2	2	0	0	0	0.011%
c.1766A>G	p.N589S	rs147021046	BENIGN	BENIGN	Predicted non-functional (low)	17732	8	7	0	1	0	0.045%
c.1780_1781insT	p.Y594Lfs					17266	1	1	0	0	0	0.006%
c.1793A>G	p.N598S	rs201236716	BENIGN	PROBABLY DAMAGING	Predicted non-functional (low)	17754	1	0	1	0	0	0.006%
c.1870G>A	p.V6241	rs76615690	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	2	0	2	0	0	0.011%
c.1901A>G	p.Y634C	rs202140203	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17752	1	0	0	1	0	0.006%
c.1936C>T	p.R646C	rs368129141	POSSIBLY DAMAGING	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.1937G>A	p.R646H	rs112387550	BENIGN	BENIGN	Predicted non-functional (low)	17754	7	4	3	0	0	0.039%
c.1976C>T	p.A659V	rs140786703	POSSIBLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.1990G>A	p.V664M	rs376213990	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.2020_2021del	p.V674Lfs					17266	319	319	0	0	0	1.848%
c.2027G>C	p.5676T		PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	0	0	0	1	0.006%

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.2083C>G	p.L695V	rs370323921	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (high)	17754	1	1	0	0	0	0.006%
c.2141G>A	p.R714H	rs375047023	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted non-functional (low)	17754	2	2	0	0	0	0.011%
c.2209C>G	p.L737V	rs201100763	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	0	0	1	0	0.006%
c.2257G>A	p.V753M	rs146874573	BENIGN	POSSIBLY DAMAGING	Predicted non-functional (low)	17754	2	2	0	0	0	0.011%
c.2338G>A	p.V780M	rs193182840	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	1	0	0	0.006%
c.2428G>T	p.V810L	rs145362908	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	8	5	3	0	0	0.045%
c.2428G>C	p.V810L	rs145362908	BENIGN	BENIGN	Predicted non-functional (neutral)	17752	1	0	0	1	0	0.006%
c.2501T>C	p.M834T	rs373435883	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
*c.2524T>C	p.F842L	rs19029B665	PROBABLY DAMAGING			17754	1	0	1	0	0	0.006%
c.2525T>C	p.F842S	rs374068891	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.2551G>A	p.A851T	rs139297968	POSSIBLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	2	0	0	0	0.011%
c.2572A>G	p.1858V	rs1805082	BENIGN/Known polymorphism	BENIGN	Predicted non-functional (neutral)	17752	8005	5758	1100	864	283	45.094%
c.2605-6_2605-3del	intronic		Neutral			17266	1	1	0	0	0	0.006%
c.2621A>T	p.D874V	rs372030650	POSSIBLY DAMAGING/Published	BENIGN	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.2705C>G	p.S902C	rs374656358	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.2731G>A	p.G911S	rs34302553	BENIGN	POSSIBLY DAMAGING	Predicted non-functional (low)	17754	69	69	3	0	3	0.389%
c.2796-4C>T	intronic	rs374406578	Neutral			17754	1	1	0	0	0	0.006%
c.2800C>T	p.R934X	rs370721218	PROBABLY DAMAGING/Published			17754	1	1	0	0	0	0.006%
c.2819C>T	p.S940L	rs143124972	PROBABLY DAMAGING/Published			17754	1	1	0	0	0	0.006%
c.2873G>A	p.R958Q	rs120074132	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.2882A>G	p.N961S	rs34084984	BENIGN/Published	BENIGN	Predicted non-functional (low)	17754	69	56	11	1	1	0.389%
c.2908_2909insTT	p.S970Ffs					17266	1	1	0	0	0	0.006%
C.2911+4C>T	intronic	rs186588103	Neutral			17754	3	3	0	0	0	0.017%
c.2929+4C>T	intronic	rs186588103	Neutral			17754	1	0	1	0	0	0.006%
c.2972_2973del	p.991_fs		Published	POSSIBLY DAMAGING	Predicted functional (medium)	17724	2	0	0	1	1	0.011%
c.2974G>T	p.G992W	rs80358254	PROBABLY DAMAGING/Published	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3011C>T	p.S1004L	rs150334966	PROBABLY DAMAGING/Published			17750	11	8	0	2	1	0.062%
c.3019C>G	p.P1007A	rs80358257	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	3	2	1	0	0	0.017%
c.3028A>C	p.K1010Q	rs191876836	BENIGN	BENIGN	Predicted functional (medium)	17754	1	0	1	0	0	0.006%

Europe PMC Funders Author Manuscripts

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.3047A>G	p.H1016R	rs140211089	POSSIBLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3052G>A	p.A1018T	rs146666146	PROBABLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (high)	17754	2	2	0	0	0	0.011%
c.3059G>C	p.S1020T	rs374719153	BENIGN	BENIGN	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3182T>C	p.11061T	rs80358259	BENIGN/Published	POSSIBLY DAMAGING	Predicted functional (medium)	17754	5	5	0	0	0	0.028%
c.3184G>A	p.A1062T	rs369960141	POSSIBLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3217G>A	p.G1073S	rs141440861	BENIGN	BENIGN	Predicted non-functional (neutral)	17750	23	18	4	1	0	0.130%
c.3265G>A	p.E1089K	rs374526072	PROBABLY DAMAGING/Published	POSSIBLY DAMAGING	Predicted functional (high)	17754	1	1	0	0	0	0.006%
c.3343G>T	p.V1115F	rs34226296	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	43	34	7	0	2	0.242%
c.3364T>C	p.W1122R	rs148571882	BENIGN	BENIGN	Predicted non-functional (low)	17754	2	1	0	0	1	0.011%
c.3422T>G	p.V1141G	rs144725473	PROBABLY DAMAGING/Published	POSSIBLY DAMAGING	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.3477+4A>G	intronic	rs114073738	Neutral			17754	28	28	0	0	0	0.158%
c.3498+4A>G	intronic		Negative			17754	4	0	4	0	0	0.023%
c.3506G>T	p.S11691	rs139612110	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	2	0	0	0	0.011%
c.3535A>G	p.M1179V	rs61731969	BENIGN	BENIGN	Predicted non-functional (neutral)	17694	61	54	5	2	0	0.345%
c.3548G>A	p.R1183H	rs148035987	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	8	7	0	0	1	0.045%
c.3550G>A	p.V1184M		POSSIBLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17676	1	1	0	1	0	0.006%
c.3556C>T	p.R1186C	rs145297180	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	2	0	0	0	0.011%
c.3557G>A	p.R1183H	rs200444084	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	2	1	1	0	0	0.011%
c.3560C>T	p.A1187V	rs113371321	POSSIBLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (high)	17754	1	0	1	0	0	0.006%
c.3566A>G	p.E1189G	rs369098773	POSSIBLY DAMAGING/Published	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.3577C>T	p.H1193Y	rs375309094	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	1	1	0	0	0	0.006%
c.3598A>G	p.S1200G	rs35248744	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	63	61	1	0	1	0.355%
c.3611_3G14del	p.L1204Qfs		Published			17264	1	1	0	0	0	0.006%
c.3619T>C	p.F1207L	rs140827681	BENIGN	POSSIBLY DAMAGING	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.3667A>G	p.11223V	rs368658600	BENIGN	POSSIBLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3689T>C	p.L12305	rs374150662	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3741_3745A	0					17748	1	0	0	1	0	0.006%
c.3742_3745del	p.L1248Vfs		Published			17266	1	1	0	0	0	0.006%
c.3755-5_3755-4insTC	intronic		Neutral			17266	4	4	0	0	0	0.023%

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.3796C>T	p.R1266X	rs376164368	PROBABLY DAMAGING			17754	1	1	0	0	0	0.006%
c.3797G>A	p.R1266Q	rs1805084	BENIGN/Known polymorphism	BENIGN	Predicted non-functional (low)	17754	1724	1275	307	101	41	9.710%
c.3799T>G	p.Y1267D	rs373435628	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.3811G>C	p.E1271Q	rs140527006	POSSIBLY DAMAGING	BENIGN	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.3814C>T	p.R1272C	rs200264267	PROBABLY DAMAGING	BENIGN	Predicted non-functional (neutral)	17754	1	0	1	0	0	0.006%
c.3818A>G	p.E1273G	rs374032318	BENIGN	BENIGN	Predicted non-functional (low)	17754	1	1	0	0	0	0.006%
c.3821G>A	p.R1274Q	rs151305963	BENIGN	BENIGN	Predicted non-functional (neutral)	17754	5	4	0	0	1	0.028%

Europe PMC Funders Author Manuscripts

Table 2

This table summarizes the 271 distinct variants detected in *NPC2*. Each variant has a corresponding cDNA number; protein change and reference SNP "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 non-pathogenic are shaded grey. The number of alleles analyzed for each variant and the total number of times the variant was detected is noted in conjunction with the frequency of each variant in each of the four data sets and the carrier rate for each variant.

cDNA	Protein	rs#	Polyphen-2/MaxEntScan/Published	SIFT	MutationAssessor Pred	Total Alleles	Total Variants	NHLBI	1000 Genome	Clinseq	Autism	Rate
c.38T>C	p.L13P	rs147602717	PROBABLY DAMAGING	POSSIBLY DAMAGING	Predicted functional (medium)	17360	5	4	0	1	0	0.029%
c.49G>A	p.A17T	rs145302203	BENIGN	BENIGN	Predicted non-functional (low)	17748	1	1	0	0	0	0.006%
c.56C>A	p.A19D	rs369392502	PROBABLY DAMAGING	PROBABLY DAMAGING	Predicted functional (medium)	17748	1	1	0	0	0	0.006%
c.58G>T	p.E20X	rs80358260	Published	PROBABLY DAMAGING	Predicted functional (medium)	17750	1	1	0	0	0	0.006%
c.88G>A	p.V30M	rs151220873	POSSIBLY DAMAGING/Published	POSSIBLY DAMAGING	Predicted non-functional (low)	17252	34	25	1	4	4	0.197%
c.115G>A	p.V39M	rs80358261	PROBABLY DAMAGING/Published	PROBABLY DAMAGING	Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.212A>G	p.K71R	rs142075589	POSSIBLY DAMAGING	BENIGN	Predicted non-functional (neutral)	17754	3	2	1	0	0	0.017%
c.224A>T	p.H75L	rs369221608	PROBABLY DAMAGING		Predicted functional (medium)	17754	1	1	0	0	0	0.006%
c.271G>A	p.D91N	rs148607507	POSSIBLY DAMAGING	POSSIBLY DAMAGING	Predicted non-functional (low)	17754	10	8	0	2	0	0.056%
c.278G>T	p.C93F	rs143960270	PROBABLY DAMAGING/Published		Predicted functional (high)	17754	1	1	0	0	0	0.006%
c.292A>C	p.N98H	rs142858704	BENIGN	BENIGN	Predicted non-functional (low)	17754	12	11	1	0	0	0.068%
c.340C>G	p.P114A	rs371363324	PROBABLY DAMAGING		Predicted functional (high)	17754	1	1	0	0	0	0.006%
c.441+1G>A	intronic	rs140130028	Strong Negative/Published			17752	96	83	0	6	7	0.541%
c.442-4A>C	intronic	rs114950106	Neutral			17754	104	104	0	0	0	0.586%