



OPEN

Estimated prevalence of Niemann–Pick type C disease in Quebec

Marjorie Labrecque^{1,2,7}, Lahoud Touma^{2,3,7}, Claude Bhérier⁴, Antoine Duquette^{2,3,5,6} & Martine Tétreault^{2,3}✉

Niemann–Pick type C (NP-C) disease is an autosomal recessive disease caused by variants in the *NPC1* or *NPC2* genes. It has a large range of symptoms depending on age of onset, thus making it difficult to diagnose. In adults, symptoms appear mainly in the form of psychiatric problems. The prevalence varies from 0.35 to 2.2 per 100,000 births depending on the country. The aim of this study is to calculate the estimated prevalence of NP-C in Quebec to determine if it is underdiagnosed in this population. The CARTaGENE database is a unique database that regroups individuals between 40 and 69 years old from metropolitan regions of Quebec. RNA-sequencing data was available for 911 individuals and exome sequencing for 198 individuals. We used a bioinformatic pipeline on those individuals to extract the variants in the *NPC1/2* genes. The prevalence in Quebec was estimated assuming Hardy–Weinberg Equilibrium. Two pathogenic variants were used. The variant p.Pro543Leu was found in three heterozygous individuals that share a common haplotype, which suggests a founder French-Canadian pathogenic variant. The variant p.Ile1061Thr was found in two heterozygous individuals. Both variants have previously been reported and are usually associated with infantile onset. The estimated prevalence calculated using those two variants is 0.61:100,000 births. This study represents the first estimate of NP-C in Quebec. The estimated prevalence for NP-C is likely underestimated due to misdiagnosis or missed cases. It is therefore important to diagnose all NP-C patients to initiate early treatment.

Lysosomal storage disorders, like Gaucher's disease, Tay-Sachs disease or Niemann–Pick type C (NP-C) disease are a group of diseases characterized by cholesterol trafficking problems¹. The collective prevalence of lysosomal storage disorders is 1:5,000 births². NP-C (MIM 257220 and MIM 607625) are neurodegenerative autosomal recessive and pan-ethnic diseases with a prevalence varying between 0.35 and 2.2 per 100,000 births^{3,4}. NP-C has different symptoms depending on the age of onset, which can be infantile, juvenile or in adolescence and adult-onset. The classical phenotype for NP-C is found in the infantile and juvenile populations¹. The clinical presentation is neuro-visceral with hepatosplenomegaly and neurological signs like delay in motor skills, clumsiness, hypotonia and ataxia⁵. However, adult-onset subtypes have been reported and often present an atypical phenotype with psychiatric symptoms. In some cases, it can mimic other neurodegenerative diseases such as Alzheimer, Wilson or Parkinson^{6–8}. The clinical heterogeneity observed in NP-C makes the diagnosis challenging and it is often delayed, especially in atypical subtypes. As a result, the number of misdiagnoses could be high, and this may affect the estimated prevalence.

NP-C is caused by pathogenic variants in the *NPC1* gene (NM_000271) in 95% of cases and the rest is due to pathogenic variants in the *NPC2* gene (NM_006432)⁵. In the gnomAD database, there are 2,227 variants in the *NPC1* gene and 467 in the *NPC2* gene (accessed on 7 April 2021) including intronic and UTR variants. An o/e (LOEUF) score above 0.35 for missense variants in both *NPC1* and *NPC2* suggests that these genes are somewhat tolerant to change and may be unlikely to have heterozygote pathogenic variants for the severe pediatric typical NP-C⁹. However, these values need to be interpreted with caution in the context of atypical or adult-onset NPC.

¹Bioinformatics Program, Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, QC, Canada. ²CHUM Research Center, Tour Viger, 900 rue Saint-Denis, R, Montréal, QC H2X 0A9, Canada. ³Department of Neurosciences, Université de Montréal, Montréal, QC, Canada. ⁴Department of Human Genetics, McGill University, Montréal, Canada. ⁵Neurology Service, Department of Medicine, André-Barbeau Movement Disorders Unit, CHUM, Montréal, Canada. ⁶Genetic Medicine Service, Department of Medicine, CHUM, 1000 rue Saint-Denis, Montréal, QC H2X 0C1, Canada. ⁷These authors contributed equally: Marjorie Labrecque and Lahoud Touma. ✉email: antoine.duquette@umontreal.ca; martine.tetreault@umontreal.ca

The NPC-db2 database (see “Patients and methods” section) is also often used to assess the pathogenicity of variants. It regroups 692 variants in the *NPC1* gene and their classification. Amongst all those variants, 200 in *NPC1* and 5 in *NPC2* have been reported as pathogenic¹⁰. Furthermore, there is growing evidence that heterozygote carriers of recessive mutations can increase the risk of developing a disease and even lead to milder and adult-onset phenotypes^{11,12} including for neurodegenerative disorders such as Parkinson and NP-C¹³. Regarding NP-C specifically, one study described hepatosplenomegaly in 71% of heterozygotes carriers and several individuals had impaired cognitive functions¹⁴. The same study also suggests that NP-C heterozygosity may lead to late-onset neurodegeneration. Thus, the late-onset form of NP-C, due to heterozygous mutations or simply the high heterogeneity of symptoms, can be misdiagnosed and result in NP-C being underdiagnosed.

The prevalence of the disease is variable worldwide due to population differences, diagnostic awareness and diagnostic methodology^{5,15}. Some populations may also have higher disease prevalence because of a founder effect, leading some rare variants to become more frequent. French-Canadians from Quebec are a well-known example of such founder population, with more than 30 monogenic conditions showing an increased prevalence and/or particular variants/phenotypes^{16,17}. Among the French-Canadian population, we also observe regional founder effects such as in the Saguenay and Gaspesia regions¹⁸. A founder effect was previously reported in Canada for Niemann–Pick disease. Originally, Niemann–Pick type D (NP-D), had been reported clinically in Nova Scotia in four individuals of Acadian ancestry¹⁹. The founder pathogenic variant, p.Gly992Trp in *NPC1*, has a carrier frequency between 10 and 26%²⁰. The high frequency of this variant is the result of a homogeneous genetic pool in the community and is unknown outside of Nova Scotia¹⁹. Since the identification of the underlying genetic abnormality, NP-D is no longer recognized as a different disease and it is now included in the NP-C spectrum⁵. Considering the founder effect in the French-Canadian population, we want to explore how it may affect the prevalence of rare variants in *NPC1* and *NPC2* and, consequently, of the NP-C disease. In Quebec, CARTaGENE (CaG) is a unique prospective cohort of healthy individuals²¹. It was developed to help study diseases and the genetics of the Quebec population.

Therefore, our goal is to measure the prevalence of NP-C in the Quebec population using population genetics data from CaG. Our results will inform on genetic testing approaches to be applied in a clinical setting in cases presenting clinical features reminiscent of NP-C and, more importantly, when an atypical form is encountered.

Patients and methods

CaG contains genomic data and health information for 20,000 residents of Quebec between the age of 40 and 69²¹. The data used in our analysis comes from the RNA-sequencing (RNA-seq) of 911 people and from the exome-sequencing (exome-seq) of 198 people in the CaG cohort²¹. Among the 1,109 data sets, 93 people are in both the RNA-seq and exome-seq groups, which gives us sequencing information on 1,016 distinct individuals. The cohort used in this study is representative of Quebec’s population based on sex, age and ethnicity²². Individuals in our cohort had no known neurological diseases. The population structure was previously verified with a principal component analysis to confirm the French Canadians ancestry of the individuals of CaG²³.

The use of RNA-sequencing to identify variants has been proven to be reliable and enable the identification of high-quality variants^{24–27}. The pipeline for identification and classification of NP-C variants was used as described before²². Briefly, the sequences from the 1016 individuals were aligned to the human reference genome version GRCh37. The RNA-seq was aligned with HISAT2²⁸ and the variants were identified with VarDict²⁹. The exome-seq data was aligned with BWA³⁰ and the variants were identified using GATK³¹. Both datasets were annotated with ANNOVAR³² as well as custom scripts. Annotations include different pathogenicity scores (CADD³³, SIFT³⁴, Polyphen2³⁵), conservation scores (phastCons³⁶ and GERP³⁷) and allele frequencies of different databases (GnomAD⁹, NPC-db2 (<https://medgen.medizin.uni-tuebingen.de/NPC-db2/> accessed on 18 July 2019)). We then kept the rare missense and indel variants with the following filters: allelic frequency (AF) < 1%, a pathogenicity score CADD > 15 and inverted SIFT or Polyphen2 > 0.75 and for conservation a GERP > 5 or phastCons > 500. The pathogenicity of variants identified in *NPC1* and *NPC2* was assessed using the ACMG guidelines³⁸. Using this approach, two pathogenic variants were selected for further analysis. No evidence of allele degradation through nonsense mediated decay was observed in our cohort.

For each variant, we estimated the allele frequency in our cohort as the count of alternate allele divided by 2,032, the total number of alleles. The allele frequencies observed in our cohort were compared to that observed in exome and genome sequencing data from 55,852 Non-Finnish Europeans from the gnomAD v2.1 database⁹. To estimate which variants are found at significantly higher frequencies in the CaG cohort compared to Non-Finnish Europeans, we applied Fisher’s Exact test one-sided alternative as implemented in R version 4.1.0 to obtain the p-value and the odds ratio.

Summing allele frequencies over all pathogenic variants we obtained the cumulative frequency of pathogenic variants in *NPC1/2* genes. The prevalence was estimated using Hardy–Weinberg equation $p^2 + 2pq + q^2 = 1$, where q is the cumulative frequency of pathogenic variants. We can estimate q^2 , the frequency of homozygotes, as the expected birth incidence assuming that it is at Hardy–Weinberg equilibrium and that the penetrance is complete. We were then able to calculate the number of births that could be affected by NP-C every year in Quebec by multiplying q^2 by the number of births in Quebec in 2019. By dividing the number of births in Quebec by the number of births that could be affected by NP-C we can measure the prevalence and report it per 100,000 births.

Ethics approval. The Sample and Data Access Committee (SDAC) of CaG approved the use of the genetic and baseline characteristics for our study. All genetic and bioinformatic analysis were carried out in accordance with relevant guidelines and regulations. Our protocol was approved by our institution’s research ethics board (CR-CHUM REB, Project 18.116).

cDNA	Protein	Exon	Variant type	rsID	Allele number detected
c.1628C>T	p.P543L	10	missense	rs369368181	3
c.3182T>C	p.I1061T	21	missense	rs80358259	2

Table 1. NP-C variants classified as pathogenic.

Position	21109 250	21111 680	21120 444	21131 617	21140 432	21148 863	21166 545	21713 456	21714 228	21714 791	21714 934	21715 084	21715 162	21715 264	21742 586	21743 039	21891 975	22020 543	22032 913	22033 261	22056 970	22057 738
P1	T	G	C	A	C	A	G	T	T	A	T	C	T	CT	A	AT	G	C	G	G	A	C
	T	G	T	G	T	G	G	T	T	A	T	C	T	C	A	A	A	G	GT	G	A	C
P2	T	G	C	A	C	A	G	T	T	A	T	C	T	CT	A	AT	G	C	G	G	A	C
	T	G	C	G	C	A	G	T	T	A	T	C	T	C	A	A	A	G	GT	A	A	C
P3	C	G	C	A	C	A	G	T	T	A	T	C	T	CT	A	AT	G	C	G	G	A	C
	C	G	C	G	C	A	G	T	T	A	T	C	T	C	AT	A	A	G	G	G	A	T

Table 2. Haplotypes of individuals with the P543L variant (in red).

Consent to participate. Informed consent was obtained from all individual participants included in the study.

Results

Identification of two *NPC1/2* pathogenic variants in the CaG cohort. In the CaG cohort, we found two variants classified as pathogenic in the *NPC1* gene (Table 1). The first one is in exon 10, c.1628C>T:p.Pro543Leu (P543L). This variant was found once before in a homozygote individual of French Canadian origin³⁹. In our cohort, all three individuals with the P543L variant were found in the RNA-seq data and were heterozygote carriers of the variant. The first person identified with the P543L variant is a 46-year-old Caucasian woman. She has migraine occurrences but no known neurodegenerative disease. The second individual is a 51-year-old Caucasian male, also without known neurodegenerative disease. The third individual carrying the P543L variant, is a 62-year-old Caucasian woman also without neurodegenerative disease. All three individuals were sampled in the region of Quebec City, a region previously associated with founder mutations or diseases^{16,40}. Thus, considering that all carriers of the P543L variant came from the same region in Quebec, we wanted to explore if they shared a haplotype. We found a haplotype of 9.4 Mb surrounding the variant that was shared among all carriers, suggesting a recent common ancestor, putatively of French-Canadian origin (Table 2).

The second pathogenic variant found is in exon 21, c.3182T>C:p.Ile1061Thr (I1061T). This variant is the most common associated with NP-C⁴¹. The first I1061T variant in CaG was found in the RNA-seq data, in a 62-year-old Caucasian female with no known neurological condition. The second I1061T variant comes from the exome-seq data and was found in a 54-year-old Caucasian, who did not mention any neurodegenerative problems. The two individuals with the I1061T variant were also heterozygous. Next, we wanted to see if both the variants identified as pathogenic in our cohort were more frequent in the Quebec population compared to the European population.

A higher frequency of pathogenic variants in Quebec compared to Europeans. In total, we had a cohort of 1,016 individuals or 2032 alleles. The P543L variant found in three heterozygotes individuals has an allele frequency (AF) of 1.48e−03 in the CaG cohort and in Non-Finnish Europeans (NFE) from the gnomAD database, the AF is 1.79e−05. The I1061T is observed in two heterozygotes which means an AF of 9.84e−04 and 3.94e−04 in CaG and gnomAD NFE, respectively. Both variants show enrichment in allele frequency in our cohort compared to gnomAD NFE as shown in Fig. 1. The P543L variant is a lot more common in our population (one-sided Fisher exact test p-value = 5.577e−5; Odds Ratio (OR) = 82.3) and the I1061T variant is also enriched although not significantly (p-value = 0.1985; OR = 2.5).

Estimated prevalence of NP-C in the Quebec population. Using the allele frequency from both pathogenic variants, P543L and I1061T, we estimated the prevalence of NP-C in the Quebec population. The frequency of allele mutant (q) was 0.0025 (5/2032) and the frequency of the disease (q²) is 6.055e−6. Assuming Hardy–Weinberg equilibrium, there should be 0.51 new case of NP-C every year in Quebec based on the 84,200 births in 2019⁴². We estimate a prevalence of 0.61 per 100,000 births in Quebec.

Discussion

In this study, we defined the prevalence of NP-C in the Quebec population. Analysis of RNA- and exome-sequencing from the CaG cohort led to the identification of two pathogenic variants P543L and I1061T. The allele frequencies of these variants were used to calculate the prevalence as well as establish if the variants were enriched in our population.

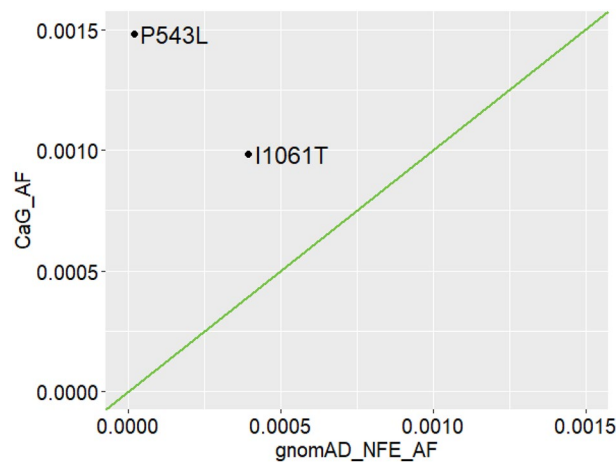


Figure 1. Comparison of variant allele frequencies (AF) in the CARTaGENE cohort compared to gnomAD NFE. The AF of the two identified pathogenic variants, P543L and I1061T is plotted in the CaG cohort (CaG_AF) as a function of the allele frequency in gnomAD NFE (gnomAD_NFE_AF). Made with the ggplot package in R (version 4.1.0).

Pathogenic variants. The P543L variant was previously seen in one infantile homozygote French-Canadian with severe neurological symptoms reminiscent of classical infantile NPC phenotype and passed away at age 6³⁹. Although the study described genetic and clinical findings in 35 patients of different ethnicities, this variant was only found in the French-Canadian patient, suggesting that the P543L variant could be more prevalent in Quebec. In our cohort, P543L was found in three heterozygote individuals all originating from the same region in Quebec. Those individuals shared a large haplotype that suggests a common French-Canadian ancestor. Since the genotype of the parents of these individuals is not available, phasing of the haplotype is not possible, and thus a direct link cannot be established. However, the French-Canadian population, and specifically the Quebec-Charlevoix region, is well known for founder effects and several diseases or variants have been described more frequently in this region^{16,40}. In addition, the P543L variant was significantly enriched in our cohort when compared to Non-Finnish Europeans. Thus, the suggestive haplotype, the enrichment in our population, and the report of this variant in a French-Canadian patient support P543L as a founder mutation. The association of this variant with a regional founder effect is of great importance as it will inform future genetic testing, where this variant should be prioritized in patients presenting a NP-C phenotype in Quebec.

The second pathogenic variant identified, I1061T, is the most prevalent variant associated with NP-C to date. Its prevalence varies depending on the population and the highest, 20%, is found in Western Europe (mainly in France and the UK), followed by Spain (10%), Portugal (6%) and Italy and Germany (5%)^{4,41,43}. This variant was also very prevalent in the Hispanic population of the Upper Rio Grande in the United States. This high prevalence is due to a founder effect from the Spanish settlers in Mexico at the beginning of the eighteenth century⁴¹. Since Quebec was founded by French immigrants in 1600–1700 and English settlers later on, it would explain why this variant is also prevalent in our cohort. On the other hand, the prevalence of I1061T does not significantly differ from that of Non-Finnish Europeans.

Prevalence of NP-C. In the world, we observe a heterogeneous estimation of NP-C prevalence, depending on the population. In Canada, and more specifically in Quebec, no prevalence had been calculated or estimated for NP-C. Elsewhere in the world, the prevalence of NP-C for 100,000 live births is 0.47 in Australia⁴⁴, 0.91 in Czech Republic¹⁵, 2.2 in Northern Portugal⁴, 0.35 in the Netherlands³, 0.82 for France⁵, 0.78 for the United Kingdom⁴⁵ and 1.12 for the United States of America (USA)⁴⁶. By combining these different prevalences, we obtain an average of 0.95 cases per 100,000 births. In all studies except for the USA, prevalence was measured using the number of patients diagnosed with NP-C divided by the number of births in the same period, reported on 100,000 births. For the USA, the prevalence is derived from an estimate based on four databases. The low prevalence observed in Australia and the Netherlands (0.47 and 0.35 per 100,000 respectively) could be explained by an underdiagnosis in the 1990s, when the phenotype spectrum was not clearly defined. More recent prevalence estimates were not found for those countries. Several factors, such as the inclusion of prenatal cases or of heterozygotes can influence prevalence. For example, in France, adding prenatal cases increases prevalence from 0.82 per 100,000 births to 0.96 per 100,000 births⁵. In the UK, many heterozygote individuals for the I1061T variant had neurological symptoms in all age groups⁴⁵. NP-C in the heterozygous form may predispose patients to a late-onset form of the disease with symptoms of dementia, tremors similar to Parkinson's disease or psychosis¹³. Individuals who are heterozygous for pathogenic variants and who present with atypical symptoms are likely to be misdiagnosed and this could underestimate the true prevalence of NP-C.

As there is no registry of NP-C cases in Quebec and it is impossible to measure the prevalence using the classical method, we used the Hardy–Weinberg equation. We also calculated the prevalence if one NP-C case was born in Quebec every year, every two years, every three years, or every four years. Using this method, the prevalence

at birth would be, respectively, 1.19:100,000, 0.59:100,000, 0.39:100,000 or 0.30:100,000. Our prevalence estimate based on the two pathogenic variants found in the CaG cohort is 0.51 case every year or 0.61 case per 100,000 births. Based on clinical data, less than one case appears to be diagnosed in Quebec each year. If one NP-C case was diagnosed every two years, the disease would probably not be underdiagnosed in the province. Otherwise, it would suggest that NP-C could be underdiagnosed clinically in the Quebec population. Since our estimation is only based on heterozygote asymptomatic carriers of two pathogenic variants, it could explain the low estimated prevalence in Quebec. The death of early infantile cases may also play a role in an underestimation. Considering the founder effect, it would be interesting to measure the prevalence of NP-C using data only from individuals from French-Canadian origin in the CaG cohort. It could help us reinforce the role of the P543L variant as a founder mutation in French-Canadians. In the future, establishing a registry of NP-C cases in Quebec is of great importance since it would allow a more precise estimation of the prevalence. It is also important to ensure that all NP-C cases are properly diagnosed in order to start treatment as soon as possible. It is particularly true for adult-onset cases who are more difficult to diagnose but for whom therapy has proven helpful to reduce neurological symptoms⁴⁷. Identifying atypical cases is also a priority to ensure a better understanding of the causes and mechanisms underlying this rare disease.

Genetic testing is readily available for NP-C, for both *NPC1* and *NPC2* genes. In clinical practice, it is generally used as a confirmatory test after initial screening by measuring levels of oxysterols⁴⁸. The level of oxysterols is only done when clinical features suggest the diagnosis of NP-C. Some clinicians use a Suspicion Index tool based on visceral, neurologic and psychiatric symptoms to evaluate the suspicion of NP-C⁴⁹. In Quebec, newborn screening includes cystic fibrosis, congenital hypothyroidism, hemoglobinopathies and several metabolic disorders⁵⁰. The disease prevalence of NP-C remains relatively low but is more prevalent than some disorders that are systematically screened. Given the founder effect in Quebec, the difficult diagnosis and the potential treatment strategies that can improve outcomes, it would be interesting to consider adding oxysterols to blood screening in newborns. One important pitfall to screening with oxysterols is the risk of false positive in patients with neonatal cholestasis⁵¹. Thus, the optimal screening method for NP-C in this population requires further investigation. There is currently a pilot study of newborn screening assays in New York, including bile acids for NP-C, which will inform us on the clinical validity of screening newborns for complex disorders⁵².

Conclusion

Certain diseases can be misdiagnosed for a variety of reasons including lack of awareness and heterogeneous clinical presentation. Being able to measure disease prevalence is very important to increase awareness among clinicians. Unfortunately, this can be difficult for rare diseases, where one misdiagnosis can have an important effect on prevalence. For hereditary diseases, estimating prevalence genetically can help compensate for these difficulties and NP-C is an excellent example of a condition for which symptom heterogeneity can make the diagnosis challenging. By identifying a founder variant in the Quebec City region and obtaining data suggesting that the disease is probably underdiagnosed, we are able to encourage clinicians of that area to consider the NP-C diagnosis more readily. This will hopefully lead to earlier detection and treatment of the disease.

Data availability

Data was obtained from the CARTaGENE database.

Code availability

All software used are listed in patients and methods. Custom scripts are available on demand.

Received: 11 August 2021; Accepted: 8 November 2021

Published online: 19 November 2021

References

- Evans, W. R. & Hendriks, C. J. Niemann–Pick type C disease—the tip of the iceberg? A review of neuropsychiatric presentation, diagnosis and treatment. *BJPsych Bull.* **41**(2), 109–114 (2017).
- Wraith, J. E. Lysosomal disorders. *Semin Neonatol.* **7**(1), 75–83 (2002).
- Poorthuis, B. J. *et al.* The frequency of lysosomal storage diseases in The Netherlands. *Hum. Genet.* **105**(1–2), 151–156 (1999).
- Pinto, R. *et al.* Prevalence of lysosomal storage diseases in Portugal. *Eur. J. Hum. Genet.* **12**(2), 87–92 (2004).
- Vanier, M. T. Niemann–Pick disease type C. *Orphanet J. Rare Dis.* **5**(1), 16 (2010).
- Malnar, M. *et al.* Bidirectional links between Alzheimer’s disease and Niemann–Pick type C disease. *Neurobiol. Dis.* **72**, 37–47 (2014).
- Sakiyama, Y. *et al.* Abnormal copper metabolism in Niemann–Pick disease type C mimicking Wilson’s disease. *Neurol. Clin. Neurosci.* **2**(6), 193–200 (2014).
- Zavala, L., *et al.* Niemann Pick type C as presentation of Huntington-like syndrome (P4. 043). AAN Enterprises (2018).
- Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**(7809), 434–443 (2020).
- Patterson, M., Niemann–Pick disease type C. GeneReviews®[Internet] (2019).
- Tétrault, M. *et al.* Adult-onset painful axonal polyneuropathy caused by a dominant NAGLU mutation. *Brain J. Neurol.* **138**(Pt 6), 1477–1483 (2015).
- Bras, J., Guerreiro, R. & Hardy, J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat. Rev. Neurosci.* **13**(7), 453–464 (2012).
- Schneider, S.A., *et al.* Do heterozygous mutations of Niemann–Pick type C predispose to late-onset neurodegeneration: A review of the literature. *J. Neurol.*, 2019: 1–10.
- Bremova-Ertl, T. *et al.* Clinical, ocular motor, and imaging profile of Niemann–Pick type C heterozygosity. *Neurology* **94**(16), e1702 (2020).
- Poupětová, H. *et al.* The birth prevalence of lysosomal storage disorders in the Czech Republic: Comparison with data in different populations. *J. Inherit. Metab. Dis.* **33**(4), 387–396 (2010).

16. Laberge, A. M. *et al.* Population history and its impact on medical genetics in Quebec. *Clin. Genet.* **68**(4), 287–301 (2005).
17. Bchetnia, M., *et al.*, Genetic burden linked to founder effects in Saguenay-Lac-Saint-Jean illustrates the importance of genetic screening test availability. *J. Med. Genetics*, 2021.
18. Roy-Gagnon, M.-H. *et al.* Genomic and genealogical investigation of the French Canadian founder population structure. *Hum. Genet.* **129**(5), 521–531 (2011).
19. Winsor, E. & Welch, J. Genetic and demographic aspects of Nova Scotia Niemann–Pick disease (type D). *Am. J. Hum. Genet.* **30**(5), 530 (1978).
20. Greer, W. L. *et al.* The Nova Scotia (type D) form of Niemann–Pick disease is caused by a G3097→T transversion in NPC1. *Am. J. Hum. Genet.* **63**(1), 52–54 (1998).
21. Awadalla, P. *et al.* Cohort profile of the CARTaGENE study: Quebec’s population-based biobank for public health and personalized genomics. *Int. J. Epidemiol.* **42**(5), 1285–1299 (2012).
22. Touma, L. *et al.* Identification and classification of rare variants in NPC1 and NPC2 in Quebec. *Sci. Rep.* **11**(1), 10344 (2021).
23. Hussin, J. G. *et al.* Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nat. Genet.* **47**(4), 400–404 (2015).
24. Frésard, L. *et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* **25**(6), 911–919 (2019).
25. Gonorazky, H. D. *et al.* Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* **104**(3), 466–483 (2019).
26. Nicolau, S. *et al.* A molecular diagnosis of LGMDR1 established by RNA sequencing. *Can. J. Neurol. Sci.* **48**(2), 293–296 (2021).
27. Vasli, N. *et al.* Recessive mutations in the kinase ZAK cause a congenital myopathy with fibre type disproportion. *Brain J. Neurol.* **140**(1), 37–48 (2017).
28. Kim, D. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**(8), 907–915 (2019).
29. Lai, Z. *et al.* VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **44**(11), e108–e108 (2016).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14), 1754–1760 (2009).
31. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**(9), 1297–1303 (2010).
32. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**(16), e164–e164 (2010).
33. Rentszsch, P. *et al.* CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**(D1), D886–D894 (2019).
34. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**(13), 3812–3814 (2003).
35. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**(1), 7–20 (2013).
36. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**(8), 1034–1050 (2005).
37. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**(12), e1001025 (2010).
38. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**(5), 405–423 (2015).
39. Millat, G. *et al.* 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.* **86**(1–2), 220–232 (2005).
40. Thiffault, I. *et al.* Diversity of ARSACS mutations in French-Canadians. *Can. J. Neurol. Sci.* **40**(1), 61–66 (2013).
41. Millat, G. *et al.* Niemann–Pick C1 disease: The I1061T substitution is a frequent mutant allele in patients of Western European descent and correlates with a classic juvenile phenotype. *Am. J. Hum. Genet.* **65**(5), 1321–1329 (1999).
42. Québec, I.D.L.S.D., *Le bilan démographique du Québec. Édition 2020* (2020) p. 183.
43. Fernandez-Valero, E. *et al.* Identification of 25 new mutations in 40 unrelated Spanish Niemann–Pick type C patients: Genotype-phenotype correlations. *Clin. Genet.* **68**(3), 245–254 (2005).
44. Meikle, P. J. *et al.* Prevalence of lysosomal storage disorders. *JAMA* **281**(3), 249–254 (1999).
45. Imrie, J. *et al.* Observational cohort study of the natural history of Niemann–Pick disease type C in the UK: A 5-year update from the UK clinical database. *BMC Neurol.* **15**, 257–257 (2015).
46. Wassif, C. A. *et al.* High incidence of unrecognized visceral/neurological late-onset Niemann–Pick disease, type C1, predicted by analysis of massively parallel sequencing data sets. *Genet. Med.* **18**(1), 41 (2016).
47. Wraith, J. E. *et al.* Miglustat in adult and juvenile patients with Niemann–Pick disease type C: Long-term data from a clinical trial. *Mol. Genet. Metab.* **99**(4), 351–357 (2010).
48. Jiang, X. *et al.* A sensitive and specific LC-MS/MS method for rapid diagnosis of Niemann–Pick C1 disease from human plasma. *J. Lipid Res.* **52**(7), 1435–1445 (2011).
49. Wijburg, F. A. *et al.* Development of a suspicion index to aid diagnosis of Niemann–Pick disease type C. *Neurology* **78**(20), 1560–1567 (2012).
50. Québec, G.d. *Blood and Urine Screening in Newborns. 2020*; <https://www.quebec.ca/en/health/advice-and-prevention/screening-and-carrier-testing-offer/blood-and-urine-screening-in-newborns/diseases-screened>.
51. Polo, G. *et al.* High level of oxysterols in neonatal cholestasis: A pitfall in analysis of biochemical markers for Niemann–Pick type C disease. *CCLM* **54**(7), 1221–1229 (2016).
52. Wasserstein, M. P. *et al.* The New York pilot newborn screening program for lysosomal storage diseases: Report of the First 65,000 Infants. *Genet. Med.* **21**(3), 631–640 (2019).

Acknowledgements

The authors would like to highlight the collaboration with the CARTaGENE database for the genetic and survey data used in this study and Compute Canada for computational resources. The authors would also like to express their appreciation to Dr Eric Bareke and Dr Alina Levtova for their great support in designing the methodology of this study.

Author contributions

M.L. performed the bioinformatic analysis and prevalence calculation and wrote the first draft of the manuscript. L.T. performed the characterization and classification of the variants and revised the manuscript. C.B.

participated in the prevalence calculation and revised the manuscript. A.D. participated in study design, interpretation of data and revised the manuscript. M.T. participated in study design, interpretation of data, supervised M.L. in bioinformatic analysis and revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by an unrestricted educational grant from Actelion Pharmaceuticals Canada, RMGA. M.T. received a Junior 1 salary award from the Fond de recherche du Québec—Santé. M.L. received an Excellence bursary from the Bioinformatic program, Université de Montréal.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.D. or M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021