SMPD1 Expression Profile and Mutation Landscape Help Decipher Genotype–Phenotype Association and Precision Diagnosis for Niemann–Pick Disease Types A and B

Ruisong Wang (ruisong.wang@huas.edu.cn)
Hunan University of Arts and Science https://orcid.org/0000-0002-7523-9644
Ziyi Qin
Hunan University of Arts and Science
Long Huang
Hunan University of Arts and Science
Huiling Luo
Hunan University of Arts and Science
Han Peng
HUAS: Hunan University of Arts and Science
Xinyu Zhou
HUAS: Hunan University of Arts and Science
Zhixiang Zhao
HUAS: Hunan University of Arts and Science
Mingyao Liu
HUAS: Hunan University of Arts and Science
Pinhong Yang
HUAS: Hunan University of Arts and Science
Tieliu Shi
HUAS: Hunan University of Arts and Science https://orcid.org/0000-0001-5592-2836

Research Article

Keywords: Niemann-pick Disease, Genotype, Phenotype, Type A and B, novel target for the subtypes.

Posted Date: August 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1899568/v1

Abstract

Background: Types A and B of the rare genetic disease Niemann–Pick disease (NPD) are caused by mutations in the *SMPD1* gene, which encodes sphingomyelin phosphodiesterase (ASM). Except for the liver and spleen enlargement and lung disease, the two subtypes have different onset times, survival times, ASM activities, and neurological abnormalities. To comprehensively explore the genotype-phenotype association and pathophysiological characteristics of NPD, we collected 144 NPD cases with strict quality control through literature mining.

Results: The difference in ASM activity can differentiate NPD type A from other subtypes, with the ratio of ASM activity to the reference values being lower in type A (threshold 0.045 (4.45%)). Severe variations, such as deletion and insertion, can cause complete loss of ASM function, leading to type A, whereas relatively mild missense mutations generally result in type B. Among reported mutations, p.Arg3AlafsX76 mutation is highly prevalent in the Chinese population, and p.R608del mutation is common in Mediterranean countries. The expression profiles of *SMPD1* from GTEx and single-cell RNA sequencing data of multiple fetal tissues showed that high expressions of *SMPD1* can be observed in the liver, spleen, and brain tissues of adults and in hepatoblasts, hematopoietic stem cells, STC2_TLX1-positive cells, mesothelial cells of the spleen, vascular endothelial cells of the cerebellum and the cerebrum of fetuses, indicating that *SMPD1* dysfunction is highly likely to have a significant effect on the function of those cell types during development and the clinicians need pay attention to these organs or tissues as well during diagnosis. In addition, we also predicted 21 new pathogenic mutations in the *SMPD1* gene that potentially cause the NPD, signifying that more rare cases will be detected with those mutations in *SMPD1*.

Conclusions: Our study is the first one to elucidate the effects of *SMPD1* mutation on cell types and at the tissue level, which provides new insights into the genotype-phenotype association and can help in the precise diagnosis of NPD.

1 Background

Lysosome storage disease (LSD) is a collection of inherited metabolic illnesses. LSD occurs when lysosomes are unable to degrade macromolecules, and the deposition of macromolecules in organelles forms cell inclusions, which causes various signs and symptoms. Niemann–Pick disease (NPD), as a type of LSD, is a group of rare autosomal recessive inherited diseases in which deposition of different lipids are caused by congenital abnormal lipid metabolism. In patients with hepatosplenomegaly, lipid-filled foam-like cells can be seen in the bone marrow, brain, and organs often called Niemann–Pick cells. Recently, the cause for NPD type A (NPA, MIM257200) and NPD type B (NPB, MIM607616) have been clarified with the mutations in the *SMPD1* gene encoding sphingomyelin phosphodiesterase-1 (1, 2). Thus, patients with NPA and NPB can be diagnosed with acid sphingomyelinase (ASM) deficiency. NPD type C (NPD-C1, MIM257220; or NPD-C2, MIM601015) results from a defect in the transport of low-

density lipoprotein cholesterol in cells (3, 4). Type D is also known as Nova Scotia NPD; some cases are an allelic variant of NPD-C1 (5).

The *SMPD1* gene comprises six exons and spans 5 kb on chromosome 11p15.4-p15.1 region. It encodes the human ASM protein (UniProt ID P17405) with 631-amino acids composed of a saposin domain, a proline-rich linker, a metallophosphatase catalytic domain, and a C-terminal domain. Six potential N-linked glycosylation sites, eight disulfides, and two zinc ions in ASM play key roles in protein folding and stability (Callahan et al., 1983; Zhou et al., 2016).

Currently, NPA and NPB have no efficient treatment. Patients with NPA often die before age three years from diagnosis (early onset) (6). Symptoms of NPA include central nervous system (CNS) deterioration, cherry-red macula, and massive hepatosplenomegaly, leading to death at an early age, whereas patients with NPB have a better prognosis (7), which symptoms are non-neuropathic. Most of them could survive into adulthood and even until the 70s.

Since the pathogenic factor for the two NPD types are mutations in *SMPD1* gene, systematically exploring the underlying mechanism that causes these two subtypes is necessary. Because the pathogenic mutations of *SMPD1* gene are primarily found in compound heterozygotes and clinical case reports, investigating the phenotype-genotype association is the key to distinguishing the two subtypes (6–9). Genotype-phenotype relationships for pathogenicity could be associated with the molecular basis, such as gene mutation and expression (10). Therefore, it is hypothesised that various mutation sites of *SMPD1* gene and its expressions in various cells are associated with early-onset and late-onset phenotypes. Therefore, in this study, we collected clinical cases reported in the literature and illustrated the landscape of mutations in *SMPD1* to explore how those mutations affect the protein function, including the physical and chemical properties of the protein and ASM activity. In addition, we also applied several algorithms to predict variants that potentially cause dysfunction of SMPD1 gene and result in NPD.

2 Results

2.1 Case Collection

To thoroughly investigate the genotype–phenotype association in NPA and NPB, reported cases were extracted from the PubMed database through a manual curation process. A total of 144 cases with mutation information and additional 23 cases reporting ASM levels but lack of detailed mutation information of the patients were collected (the detailed information can be found in the Supplementary Table S1), and the following data were extracted: onset age, sex, race, mutation site of *SMPD1* gene, clinical phenotypes and symptoms, and ASM activity level for each case based on the standardised format and terms.

2.2 ASM activity level in the collected clinical cases

Currently, there are no conclusive criteria to identify NPA and NPB at the physiological and biochemical levels. A previous study reported that the ASM level in plasma could be used as an index to differentiate NPA from NPB in Chinese populations (11). To test whether the ASM level in plasma as an index can be extended to other populations, we analysed the ASM levels in collected cases to find the threshold value to differentiate NPA from NPB based on the ratio of the ASM activity in patients to the reference constructed with healthy people (or normal ASM enzymes). When excluding the Chinese cases for this analysis, we then collected additional papers consisting of 23 non-Chinese cases (most of them are clinical reports) that reported ASM levels (in the Supplementary Table S2). These additional papers are not used for phenotype-genotype correlation due to a lack of mutation information. Thus, 108 cases were included. As a result, NPA and NPB demonstrated a significant difference in ASM activities by t-test ($p = 0.0005$), but no significant difference was found between the intermediate NPD and NPB. NPA can be differentiated from other subtypes (NPB and the intermediate group) at a threshold of 4.45% of ASM activity with an AUC value of 0.740, the Sensitivity of 0.800, Specificity of 0.705, and Youden's index of 0.505 (Figure 1B) indicating that patients without neurological involvement normally have the ASM residual activity-to-control ratio over 0.045 (4.45%).

2.3 Mutations in the clinical cases collected

Previous studies have demonstrated that patients with diverse NPD subtypes have significantly different severities (11, 12). Normally, patients with NPA have high pathogenicity mutations in *SMPD1*, with early onset of disease. In the present study, we collected 144 cases and showed that the onset age for most patients with NPA was between 0 and 10 months (less than one year), while the onset age of patients with NPB was aggregated between 0 and 200 months (16.7 years) (Figure 2A). Among those patients, sixty patients with NPD are living in Mediterranean countries, including Italy, Algeria, Spain, Turkey, Maghreb, Jordan, and North Africa. Thirty-five patients with NPD have Asian backgrounds (China and Japan) (Figure 2B). All 22 patients from Europe were Caucasians, Polish, Gypsy, and Dutch. In the cases collected, there were more patients with NPD from the Mediterranean area, although the results might be biased during the case collection from PubMed. From these cases, the most common mutation is p.Arg608del, followed by p.Arg3AlafsX76 and p.Arg610del (Figure 2C).

It has been noticed that different mutation types would lead to distinct phenotypes of NPD; therefore, we annotated the different types of mutations were annotated (Figure 2E). Duplications, nonsense mutation, deletions and insertions are considered severe mutations, while missenses are the mild to severe ones. About 50% of the mutations belong to severe type in NPA in the collected cases, a higher prevalence than that in NPB (42% roughly). The mild mutations counted for 57% of all mutations in NPB, higher than the frequency in NPA. Of all the annotated missense mutations, over 70% of mutations occurred in the metallophosphatase domain, while only a small portion of missense mutation happened in Saposin (B) Domains (Figure 2F).

2.4 Pathogenicity at different sites in *SMPD1*

ANNOVAR is a software tool that utilises genetic and evolutionary information to annotate genetic variants detected from diverse genomes functionally. A total of 1203 variants of the *SMPD1* gene have been annotated in ANNOVAR. We retrieved their corresponding SIFT_score, Polyphen2_HVAR _score, Polyphen2_HDIV_score, MutationTaster, M-CAP score and CADD_Phred. We also collected the annotation information for *SMPD1* variants from the ClinVar database; 591 variants in the ClinVar and 203 the pathogenic/likely pathogenic variants in exon regions were filtered and mapped in the diagrammatic sketch (Figure 2G). There are 44 variants annotated in the ANNOVAR but not ClinVar database. Following the pathogenicity threshold defined by their corresponding authors (SIFT_score< 0.05, Polyphen2_HDIV_score >=0.957, MetaLR_score> 0.5, CADD_phred> 20, and M-CAP_score> 0.025 and disease_causing labeled by MutationTaster_pred), 38 of the 44 variants are annotated as pathogenicity (Supplementary Table S3), but none of them have been reported in literature. In addition, 8 variants, including c.G491T, c.G1026T, c.C1279A, c.C1288G, c.C1288T, c.T1309C, c.A1351C and c.A1382C were not found in the gnomAD and other variants frequency were quite low, only 3 (c.G394T, c.C995A, and c.C1598A) were annotated in Han Chinese people (Huabiao), which indicated the population-specific pattern for some mutations.

Next, we adapted a deep learning algorithm, the EVE model, to further predict the clinical significance for those mutations in the *SMPD1* gene. EVE is a new method to predict the clinical significance of human variants based on sequences of diverse organisms across evolution (13). With the EVE model, 23 of the above 44 variants were predicted to be pathogenic. Among the 44 variants, 21 were predicted to be pathogenic with both methods (their locations were shown in Figure 2G roughly).

2.5 Differences in the distribution of variant sites among different races

To survey whether the mutation profile in the *SMPD1* gene has any ethnic prevalence, we first plotted the top 20 variants that have been annotated as pathogenic/likely pathogenic ones according to their allele frequency (Figure 3A). p.Arg610del has been found in nearly all the races. The variants at the 498th amino acid (missense and changed to His or Leu), namely p.Arg498His or p.Arg498Leu, share the second place. Among 36 pathogenic/likely pathogenic mutations detected in gnomAD (Figure 3B). p.His461ArgfsTer3 was found specific to African/African American. The Jewish people carry a unique variant p.Leu304Pro, while p.Ala195SerfsTer14, p.Arg364Gly and p.Pro155Arg are the three variants specific reported in East Asians. There are two unique mutations in Finnish populations. South Asians carry three protein changes induced by single nucleotide variants.

We further explored whether any significant difference could be observed between the Han Chinese group and other East Asians. Comparing gnomAD and Huabiao databases, we have found that the allele frequency of 24 variants was differently recorded (Figure 3C). For example, the variants occurring at sites

of 6411878 (p.Arg17Gln or p.Arg17Pro), 6411954 (p.Ala44_Leu49del) and 6415243 (p.Ser486Arg) in gnomAD are nearly 0, but in Huabiao, their prevalence is pretty high. In contrast, the frequency at the mutation sites of 6412081 and 6412854 was much higher than that of the variants in the Huabiao. Additionally, in the East Asian people (the subtype of the gnomAD), the 24 mutation sites have not been included.

2.6 Phenotype-genotype correlation

In this analysis, 144 cases with correspondingly detailed documentation of the complete mutation information were utilised.

2.6.1 Different variations are associated with different phenotypes

The amino acid glycine at site 247, located in the conserved metallophosphatase domain, is a highly prevalent mutation site affected by nucleotide changes of c.740delG, c.741delG, and c.739G>A in 7 patients. Variations at this site have been associated with NPA (12, 14, 15). Patients hosting the deletions (c.740delG or c.741delG) and c.739G>A show disease onset age less than 6 months and died by age 3 years, which result in a global developmental delay, seizure, psychosis, and other nervous system-related diseases, together with the liver- and spleen-related symptoms. In contrast, patients with most missense mutations at glycine site 247 were not diagnosed with NPA but NPB (mean onset age: 45.6 10.9 years)., except c.739G>A and c.1159T>C (p.[Cys387Arg]). Similarly, patients with homozygous or heterozygous variants c.1828_1830delCGC (p.Arg610del) were all associated with NPB without nervous system involvement.

2.6.2 Recurrent variants of c.4delC (p.Arg3AlafsX76) and c.842-849dup8 (p.His284SerfsX17) in Chinese origin correlates with NPA/B or the intermediate

Among all 28 reported variants in our collection, c.4delC (p.Arg3AlafsX76) and c.842-849dup8 (p.His284SerfsX17) in Chinese patients have high prevalence (16). The duplication variant occurs in six patients (mean onset age: 2.5 1.6 months) with seven alleles. The mean of ASM activities was significantly low (5.5 0.67% to the reference). Psychomotor regression and hypotonia were the main phenotypes related to the nervous system of patients with the mutation. For deletion of c.4delC (p.Arg3AlafsX76), nine alleles were detected in six patients. Although the deletion variants can cause severe ASM dysfunction, Chinese patients with one variant on the alleles (heterozygous) were diagnosed with the intermediate type, indicating haploinsufficiency, while patients with homozygous alleles (two c.4delC on the alleles) were diagnosed with NPB. This gene mutation appeared non-neurotoxic as the detected ASM activity was relatively high ( ). Thus, the two variants mentioned above are associated with the discrimination of the NPD.

2.6.3 Variants of c.1823_1825delCCG (p.R608del) correlate with the NPB phenotype

Among our collected cases, 14 patients had at least one *SMPD1* p.R608del allele variant (homozygous, n = 9; heterozygous, n = 5) associated with NPB clinical phenotypes (17-19). Interestingly, 13 of these patients were reported to live in Mediterranean countries (Italy, Algeria, Spain, Turkey, *etc*.). This variant has not been reported from people living in other areas except America. Nearly all patients with homozygous or heterozygous mutations survive to adulthood with the oldest patient 60 years old. None of them was reported to have neurological diseases and most of them had active ASM. Thus, c.1823_1825delCCG (p.R608del) variants are mostly associated with NPB in Mediterranean patients and the deletion of amino acid proline has minor impact on the patients since it happens near the C-terminal of the protein.

2.7 The expression pattern of *SMPD1* gene

Gene performs its functions only in the cells/tissues it expresses. To comprehensively explore the tissues affected by *SMPD1* gene mutations, we collected the expression pattern of the *SMPD1* gene from the GTEx portal and Descartes database. GTEx is a data resource and tissue bank that currently includes approximately 11,688 RNA-seq samples across 53 tissue sites, and the Descartes database contains the gene expressions of over 4 million cells of 121 human tissues during development. Considering that hepatosplenomegaly and splenomegaly are the most commonly observed clinical syndrome of NPB and higher incidence of neuronopathy with rapid progressive psychomotor deterioration are reported in NPA, we mainly focused on the expression patterns of *SMPD1* in the brain, liver, and spleen.

According to the expression patterns of the *SMPD1* gene in various tissues from GTEx, thyroid tissue was found to have the highest expression of *SMPD1*, followed by the pituitary, aorta, cerebellum, lungs, skin exposed to the sun, cerebellar hemisphere, kidney cortex, tibial nerve, and coronary artery. Liver is ranked as 13th among 53 tissues, while spleen is ranked 20th.

Considering that NPA is the most severe clinical form with early-onset CNS involvement, we believed that dysfunctions caused by pathogenic mutations of the *SMPD1* gene could significantly affect fetal development and functions of related organs. Therefore, we evaluated the expression patterns of the *SMPD1* gene in various fetal tissues from Descartes database. The single-cell dataset of fetuses in Descartes database is generated from 121 human fetal samples of 72–129 days in estimated postconceptual age and represents 15 organs (20). As a result, we observed that the *SMPD1* gene nearly expressed in all studied fetal organs (Figure 4A), and the highest expression level is found in CLC_IL5RA-positive cells of heart, followed by endocrine cells in pancreatic islets, endocardial cells, retinal microglia, and megakaryocytes in kidney. The expression pattern of *SMPD1* in these organs indicates that *SMPD1* dysfunction should significantly affect the functions and development of heart, pancreatic islets, eyes, and kidneys. The highest expression level of the *SMPD1* gene in liver is detected in hepatoblast cells,

followed by hematopoietic stem cells. In spleen, the highest expression level was observed in STC2_TLX1-positive cells, followed by mesothelial cells. In the brain, the highest expression level was found in the vascular endothelial cells of cerebellum, followed by the vascular endothelial cells of cerebrum and astrocytes of cerebellum. In cerebellum, *SMPD1* presented a relatively low expression profile in astrocytes, Purkinje neurons, SLC24A4_PEX5L-positive cells, oligodendrocytes, and microglia. A similar expression profile was observed for *SMPD1* in inhibitory neurons, astrocytes, limbic system neurons, oligodendrocytes and microglia of cerebrum. Interestingly, *SMPD1* was highly expressed in placenta; the highest expression level was found in IGFBP1_DKK1-positive cells, followed by PAEP_MECOM-positive cells and myeloid cells.

3 Discussion

In the present study, we tried to explore the association of phenotype-genotype in NPD comprehensively and lay the foundation for understanding the mechanisms of this rare disease. With the strict quality-controlled literature search, we collected 144 cases with comprehensive pathophysiological characteristics of NPD. We also found the connection between some variants and the phenotypes (NPD type A or B). Type A is correlated with more severe mutations, while patients with the non-neuro-related type NPB normally have mutations in *SMPD1* with a mild effect. In addition, following the model, we have found a threshold of 4.45% and that it can be taken into account to discriminate the majority of cases with NPA from clinical phenotypes less severe of NPD (intermediate and NPB). Furthermore, we also explored the expression landscape of the *SMPD1* in different cell types of fetal development and adult tissues, which offered us the opportunities to better understand pathogenic mechanisms underlying NPA and NPB at a single cell type level. At the same time, the difference in *SMPD1* expression levels on different cell types provides an important resource for the precise diagnosis of the disease in clinical application.

The residual ASM activity has been regarded as one of the clinical features to distinguish NPA from NPB in Chinese people (11), while some bibliographic data that supports ASM residual activity threshold is not definitive for discriminating between type A and B (21–23). Normally, < 5% of effective residual ASM activity *in situ* is observed in NPA, whereas 5–20% is detected in NPB (6, 24–26). However, in the literature we have mined, > 5% of the cases were still diagnosed as NPA (Ding et al., 2016; Ceron-Rodriguez et al., 2019; Al-Eitan et al., 2020). Several intermediate cases indicated that the residual activities of the ASM enzyme were broad (21, 27). Therefore, the ASM activity is not always associated with the so-called well-defined subtypes. Our model indicated that 4.45% of ASM activity could be the threshold to distinguish NPA from other subtypes. However, Hu, Maegawa (11) suggested that the cutoff value for differentiating the two clinical forms was 1.685 nmol/17 h/mg protein (approximately 12.2% to the reference, 13.7 nmol/17 h/mg protein as the reference value in Chinese patients only, and all ASM activities were measured within single laboratory using the same method. Although the *SMPD1* gene sequencing appears to be a golden standard for NPD diagnosis, the method might be less available in the less-developed regions. ASM activity is still a standard for deficiency diagnosis (28). The threshold proposed in the present study was derived from multinational samples (i.e., leukocytes, skin fibroblasts,

or dried blood spots) and different ethnic groups (29) with different measurement approaches to ASM activity, which further indicates that ASM activity is a common feature used to differentiate NPA from other subtypes for counselling, prognostication, and the interpretation.

SMPD1 gene mutations have been reported in many countries and ethnic groups, and its mutation prevalence varies from one ethnic group to another (7, 14–16, 18, 30). In this study, we observed frequency differences in the same sites of *SMPD1* protein between the East Asians in gnomAD and Han Chinese in the "Huabiao" project (Fig. 3C). In the gnomAD, the most common variant in East Asians is p.Lys189GlnfsTer4 (c.564dup), while p.Glu508Lys (c.1522G > A) is the most common one in Han Chinese. Besides, the high-frequency mutation sites of the *SMPD1* gene highly vary in different populations reported in the literature, such as Ashkenazi Jews, Italians, Spanish, Turks, Chinese, and Dutch. Moreover, p.F333SfsX52, p.L304P, and p.R498L are the most common *SMPD1* gene mutations among Ashkenazi Jews, which morally cause NPA (31, 32). This racial difference in the mutation frequency may also contribute to a massive difference in the prevalence of NPA and NPB in various races and cause different phenotypes. Globally, the most common mutation is p.Arg610del, which has been associated with NPB (16, 33). Similarly, p.Arg610del is also dominant in our collected cases. In contrast, the most common mutations among Chinese patients are p.Arg3AlafsX76 and p.H284SfsX7 (16).

Furthermore, with the two databases, ClinVar and ANNOVAR, and a deep learning algorithm to improve the reliability of the EVE model, we predict unreported 21 variants that could be pathogenic (34), which can provide new information to interpret the related variants in SMPD1 gene testing for NPD. The comparison of different databases shows the frequency of variant sites of the *SMPD1* gene in the Chinese Han group from Huabiao to the east Asian in the gnomAD. It is believed that these high-risk mutations might lead to spontaneous abortion as the *SMPD1* gene expression is high in the CNS system during development. These novel variants aggregate in the domains of the protein. Fourteen variants (66.67%) are involved in sphingolipid metabolism reactions in the Calcineurin-like phosphodiesterase domain (from 255th to 462nd amino acid). Sphingomyelin was included in this type of phosphodiesterase superfamily, demonstrating the amino acid changes due to variants would impact the function of the ASM and finally lead to the severe phenotype - NPA or the milder one - NPB. Metallo-dependent phosphatase-like domain (202nd to 497th amino acid) found 80.95% of the variants. This domain is associated with metabolite damage-control (35). Hence, if the variants occur in these domains, it is highly likely to lead to LSD, even NPD (36).

Studies that comprehensively expounded pathways related to NPD are barely found. In the present results, within the data above, we infer that the following scenario could be the mechanisms underlying NPD. The expression profiles of *SMPD1* in cells and tissues in healthy people help explain the complex symptoms of NPD. We can further connect the clinical phenotypes to the mutation pattern based on the SMPD1 expression profiles in fetal and adult tissues. ASM is an enzyme essential for neurodevelopment. Normally, the mutations in the catalytic domain of *SMPD1* have severe pathogenic effects because the lost catalytic function of the enzyme can significantly decrease ASM activity, which causes the accumulation of sphingomyelin and other sphingolipids that are toxic at elevated and nonphysiological

levels. The clinical manifestations include rapid progressive psychomotor deterioration, liver and spleen enlargement, respiratory disease, jaundice, and death within three years (11, 15, 37). Liver and spleen enlargement could be a compensation mechanism of the body to sustain ASM activities. Considering that *SMPD1* is universally expressed in many different cell types and tissues, it is expected that the dysfunction of *SMPD1* should have a significant impact on many tissues, indicating that symptoms of NPD should present in the whole body without abundant specificity. Consistently, the phenotypes reported in NPD are all consistent with short stature, osteoporosis, sea-blue histiocytosis, microcytic anaemia, and bone-marrow foam cells. The expression profile of SMPD1 in various cell types during development and various adult tissues can help us comprehensively decipher the potentially affected cell types and tissues of SMPD1 mutation, which might have been ignored clinically. In addition, according to its expression profile, SMPD1 is highly expressed in the heart, pancreas, eyes, and kidneys during fetal development, which indicates that the dysfunction of the SMPD1 gene should have possibly influenced these organs and the related phenotypes such as renal involvement in NPD is rarely reported (38, 39). Therefore, in clinical application, clinicians are recommended to conduct comprehensive examination during diagnosing patients with potential NPD, paying attention to the pathological abnormalities of these organs, such as the heart, pancreas, and kidneys, during fetal development hepatosplenomegaly, splenomegaly, and neurological abnormalities.

We further deduced that potential pathogeny at a gene level could correlate with the types of the mutations. Severe mutations resulting from deletion or insertion and stop gain led to the premature termination of the synthesis of the polypeptide chain of the *SMPD1* gene, or mutated polypeptide chains produce enzymes without biological activity or barely active domains. In patients with NPB, a single missense mutation only changes an amino acid, resulting in defective ASM with partial catalytic activity. Therefore, the ASM activity of NPB is higher than that of NPA, which explains the perspective that patients with NPA/NPB have the same pathogenic mutated genes, but the clinical manifestations are quite different. The pathogenic mutations of the SMPD1 gene are primarily found in compound heterozygotes; the phenotype-genotype association study is particularly complicated. Thereupon, the expression profiles of the *SMPD1* in different cell types of fetal development and adult tissues would be a tool for understanding the pathogenic mechanisms underlying NPA and NPB. The period from 2 weeks post-conception to early childhood is crucial for developing the brain and other CNS organs (40). Large amounts of sphingomyelin are needed to be converted to develop non-CNS cells (40). In the present study, the expression pattern of *SMPD1* (Fig. 4A) in various cell types of fetal demonstrated that *SMPD1* dysfunction should significantly affect the functions and development of the circulatory system (heart and kidneys) and nervous system, which are essential for the survival of the fetal and infants. The placenta, vital to support fetal growth, also presents a high *SMPD1* gene expression. The individual clinical symptoms strongly correlate with the severity of *SMPD1* mutations, as the mutations would result in the functional decrease or even loss of the ASM activity in those cell types. Once those mutations cause decreased or forfeit ability of ASM results in the ASM substrates and sphingomyelin accumulation, which would negatively affect individual fetuses. Finally, the excessive amount of accumulated sphingomyelin might lead to NPD phenotypes at an early age (namely, Type A) or the NPB (the late-onset

phenotype). Among 53 adult tissues, the expression of *SMPD1* is relatively high in the liver and the spleen (liver ranked 13, and the spleen ranked 20, Fig. 4C), which also suggests high levels of sphingomyelin in both tissues. Individuals with low ASM activities might not be able to convert sphingomyelin timely; thus, patients with NPA and NPB are featured with progressive hepatosplenomegaly and other organ dysfunction (1, 15, 41).

Our study is the first to comprehensively elucidate the effects of *SMPD1* mutation on cell types and at the tissue level, which provides new insights into the genotype-phenotype association and can help in the precise diagnosis of NPD. Admittedly, our study has certain limitations; the number of cases included in this study is relatively small, which could influence the AUC results; more cases should improve the model's performance. In this study, we fail to comprehensively detect the relationship between phenotypes and genotypes because of incomplete phenotype data from some reported cases. In addition, we found the area or race specificity to the frequency of the variants, but it should be noted that some mutations with population or area-specific prevalence could also result from the bias of case study and collection. However, we compared the frequency of the variants collected to the public databases, gnomAD and the results are consistent. For example, p.Arg610del are the most domain variant in both the documented NPD patients (gnomAD databases) and the cases we collected.

4 Conclusions

As a rare disease, symptoms of NPD are scattered in the whole body without abundant specificity. It is often misdiagnosed in different specialities. Many researchers have reported that patients with the same mutation site in a candidate pathogenetic gene always have different phenotypes; we also observe similar events in patients with NPD, which indicates that other unidentified factors can also contribute to the clinical manifestation. Therefore, caution should be taken when interpreting the effect of gene mutation in inherited diseases. It can be anticipated that with the whole genome sequence technology being gradually applied to the clinical diagnosis field, comprehensively deciphering the underlying mechanism for inherited disease will be a routine procedure with which the genetic factors and their interaction with diseases will be fully illustrated. It is also suggested that noninvasive prenatal testing with the whole genome sequence technology could be incorporated into the national healthcare program that has reduced the prevalence of inherited diseases in China (42); thus, the prediction for pathogenicity will also be improved with the application of deep learning (43).

5 Materials And Methods

5.1 Case collection

We searched PubMed using 'Niemann-Pick disease', and 'mutation' as keywords. We selected the data with a clearly statement of the patient's information (age, gender, nationality or race background etc.) and the corresponding types of NPD, mutation sites, clinical symptoms and/or ASM level. Then, the published result was evaluated by the

5.2 Cuttoff value to distinguish NPA and NPB based on ASM activity

The t-test was used to test whether ASM activity has a difference between the two groups, with $p < 0.05$ considered as significant. The "pROC" package (version 0.2.3) in R (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria) was used to analyse ASM ratio data (the ratio of ASM activity in patients to the reference value of healthy people that were stated by the authors of the publications) to classify response groups and visualised by ggplot2 R package. The predictive performance of each model was evaluated using the receiver operating characteristic curve (ROC) and the area under the curve (AUC).

5.3 Pathogenicity annotation for mutations in *SMPD1* gene

SMPD1 has been confirmed to cause NPA and NPB. To date, many mutation sites have been detected in patients with NPD. To comprehensively identify pathogenic mutations in *SMPD1* that cause NPD, a pathogenicity analysis was conducted for all potential mutation sites in *SMPD1* genes with ANNOVAR (https://annovar.openbioinformatics.org/en) (Wang, Li (44). The annotated pathogenetic effect of each mutation in *SMPD1* was retrieved from ClinVar. PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) was used to predict the pathogenic effect of missense on the protein. EVE scores for the *SMPD1* variants in proteins sequence were retrieved from the EVE model (https://evemodel.org/proteins/ASM_HUMAN) (13).

To comprehensively detect the mutation profile of *SMPD1*, we extracted the mutation frequency of *SMPD1* from the gnomAD (https://gnomad.broadinstitute.org/). The Han Chinese populations' allele frequency of *SMPD1* was obtained from Huabiao (https://www.biosino.org/wepd).

5.4 The expression pattern of *SMPD1* gene in various cell types and tissues

The expression patterns of *SMPD1* in adult tissues were extracted from the GTEx portal (https://www.gtexportal.org/), and the expression patterns of *SMPD1* at the cell type level were downloaded from Descartes (https://descartes.brotmanbaty.org/). GTEx is a data resource and tissue bank and is used in investigating the relationship between genetic variation and gene expression in human tissues. The currently released platform includes genotype data from approximately 714 donors and approximately 11,688 RNA-seq samples across 53 tissue sites. The Descartes database hosted the human gene expressions of over 4 million cells of 121 human tissues during fetal development. Fetal gene expressions of *SMPD1* in different cells were downloaded from the Gene Expression Omnibus platform (GSE156793).

5.5 Statistical analysis

R software (version 3.6.1) was used to conduct all statistical analyses. is seen as the significant difference.

Declarations

Ethics approval and consent to participate

Patient consent was not required because this study used public data.

Consent for publication

Patient consent was not required because this study used public data.

Availability of data and materials

All data are submitted within the paper.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The China Hunan Changde ChangCaiQiZhi (Grant No. 2020-59) and Hunan University of Arts and Science PhD Boosting Project 20BSQD08 supported this work.

Author Contributions

Ruisong Wang finished the manuscript. Mingyao Liu provided the idea of the research. Ziyi Qin, Long Huang, Huiling Luo, Han Peng, and Xinyu Zhou collected the data. Ruisong Wang analysed the clinical findings and genetic assay. Pinhong Yang and Tieliu Shi designed and supervised the study.

Acknowledgements

We thank all the authors of the case reports in the collection and the GTEx project, NCBI GEO, and Xena database and their contributors for these valuable public data sets.

References

1. Al-Eitan L, Alqa'qa K, Amayreh W, Aljamal H, Khasawneh R, Al-Zoubi B, et al. Novel mutations in the SMPD1 gene in Jordanian children with Acid sphingomyelinase deficiency (Niemann-Pick types A and B). Gene. 2020;747:144683. doi:10.1016/j.gene.2020.144683.
2. Aneja A, Sharma A, Dalal A, Sondhi V. R542X mutation in SMPD1 gene: genetically novel mutation with phenotypic features intermediate between type A and type B Niemann-Pick disease. BMJ Case Rep. 2012;2012. doi: 10.1136/bcr-2012-006959.
3. Vanier MT. Niemann–pick diseases. Handb Clin Neurol. 2013;113:1717–21. doi.
4. de Frutos LL, Cebolla JJ, Irún P, Köhler R, Giraldo P. The erythrocyte osmotic resistance test as screening tool for cholesterol-related lysosomal storage diseases. Clin Chim Acta. 2018;480:161–5. doi.
5. Greer WL, Riddell DC, Gillan TL, Girouard GS, Sparrow SM, Byers DM, 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. 1998;63(1):52–4. doi:https://doi.org/10.1086/301931.
6. Ota S, Noguchi A, Kondo D, Nakajima Y, Ito T, Arai H, et al. An Early-Onset Neuronopathic Form of Acid Sphingomyelinase Deficiency: A SMPD1 p.C133Y Mutation in the Saposin Domain of Acid Sphingomyelinase. Tohoku J Exp Med. 2020;250(1):5–11. doi:10.1620/tjem.250.5.
7. Nasereddin A, Ereqat S. Deep sequencing of SMPD1 gene revealed a heterozygous frameshift mutation (p.Ser192Alafs) in a Palestinian infant with Niemann–Pick disease type A: a case report. J Med Case Rep. 2018;12(1):272. doi:10.1186/s13256-018-1805-x.
8. Cheema HA, Rasool IG, Anjum MN, Zahoor MY. Mutational spectrum of SMPD1 gene in Pakistani Niemann-Pick disease patients. Pak J Med Sci. 2020;36(3):479–84. doi:10.12669/pjms.36.3.467.
9. Ordieres-Ortega L, Galeano-Valle F, Mallén-Pérez M, Muñoz-Delgado C, Apaza-Chavez JE, Menárguez-Palanca FJ, et al. Niemann-Pick disease type-B: a unique case report with compound heterozygosity and complicated lipid management. BMC Med Genet. 2020;21(1):94. doi:10.1186/s12881-020-01027-9.
10. Li G, Eriani G, Wang ED, Zhou XL. Distinct pathogenic mechanisms of various RARS1 mutations in Pelizaeus-Merzbacher-like disease. Science China Life sciences. 2021. doi.
11. Hu J, Maegawa GHB, Zhan X, Gao X, Wang Y, Xu F, et al. Clinical, biochemical, and genotype-phenotype correlations of 118 patients with Niemann-Pick disease Types A/B. Hum Mutat. 2021;42(5):614–25. doi:10.1002/humu.24192.
12. Irun P, Mallen M, Dominguez C, Rodriguez-Sureda V, Alvarez-Sala LA, Arslan N, et al. Identification of seven novel SMPD1 mutations causing Niemann-Pick disease types A and B. Clin Genet. 2013;84(4):356–61. doi:10.1111/cge.12076.
13. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021;599(7883):91–5. doi:10.1038/s41586-021-04043-8.

14. Galehdari H, Tangestani R, Ghasemian S. New Single Nucleotide Deletion In the SMPD1 Gene Causes Niemann Pick Disease Type A in a Child from Southwest Iran: A Case Report. Iran J Pediatr. 2013;23(2):233–6. doi.
15. Hashemian S, Eshraghi P, Dilaver N, Galehdari H, Shalbafan B, Vakili R, et al. Niemann-Pick Diseases: The Largest Iranian Cohort with Genetic Analysis. Iran J Child Neurol. 2019;13(2):155–62. doi.
16. Zhang H, Wang Y, Gong Z, Li X, Qiu W, Han L, et al. Identification of a distinct mutation spectrum in the SMPD1 gene of Chinese patients with acid sphingomyelinase-deficient Niemann-Pick disease. Orphanet J Rare Dis. 2013;8:15. doi:10.1186/1750-1172-8-15.
17. Rodriguez-Pascau L, Gort L, Schuchman EH, Vilageliu L, Grinberg D, Chabas A. Identification and characterization of SMPD1 mutations causing Niemann-Pick types A and B in Spanish patients. Hum Mutat. 2009;30(7):1117–22. doi:10.1002/humu.21018.
18. Pittis MG, Ricci V, Guerci VI, Marcais C, Ciana G, Dardis A, et al. Acid sphingomyelinase: identification of nine novel mutations among Italian Niemann Pick type B patients and characterization of in vivo functional in-frame start codon. Hum Mutat. 2004;24(2):186–7. doi:10.1002/humu.9263.
19. Sikora J, Pavlu-Pereira H, Elleder M, Roelofs H, Wevers RA. Seven novel acid sphingomyelinase gene mutations in Niemann-Pick type A and B patients. Ann Hum Genet. 2003;67(Pt 1):63–70. doi:10.1046/j.1469-1809.2003.00009.x.
20. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, et al. A human cell atlas of fetal gene expression. Science. 2020;370(6518). doi.
21. Mihaylova V, Hantke J, Sinigerska I, Cherninkova S, Raicheva M, Bouwer S, et al. Highly variable neural involvement in sphingomyelinase-deficient Niemann–Pick disease caused by an ancestral Gypsy mutation. Brain. 2007;130(4):1050–61. doi:10.1093/brain/awm026.
22. Dardis A, Zampieri S, Filocamo M, Burlina A, Bembi B, Gabriela Pittis M. Functional in vitro characterization of 14 SMPD1 mutations identified in Italian patients affected by Niemann Pick Type B disease. Hum Mutat. 2005;26(2):164-. doi:https://doi.org/10.1002/humu.9353.
23. Boustany R-M, Al-Shareef I, El-Haddad S. Chapter 104 - Sphingolipid Disorders and the Neuronal Ceroid Lipofuscinoses or Batten Disease (Wolman Disease, Cholesteryl Ester Storage Disease, and Cerebrotendinous Xanthomatosis). In: Rimoin D, Pyeritz R, Korf B, editors. Emery and Rimoin's Principles and Practice of Medical Genetics (Sixth Edition). Oxford: Academic Press; 2013. pp. 1–85.
24. Desnick JP, Kim J, He X, Wasserstein MP, Simonaro CM, Schuchman EH. Identification and characterization of eight novel SMPD1 mutations causing types A and B Niemann-Pick disease. Mol Med. 2010;16(7–8):316–21. doi:10.2119/molmed.2010.00017.
25. Mukherjee SB, Pandey M, Kapoor S, Priya TP. Infant with type A Niemann Pick disease and undetectable Niemann Pick cells in bone marrow. Indian Pediatr. 2012;49(6):490–2. doi:10.1007/s13312-012-0095-4.
26. Graber D, Salvayre R, Levade T. Accurate differentiation of neuronopathic and nonneuronopathic forms of Niemann-Pick disease by evaluation of the effective residual lysosomal sphingomyelinase activity in intact cells. J Neurochem. 1994;63(3):1060–8. doi:10.1046/j.1471-4159.1994.63031060.x.

27. Pavlů-Pereira H, Asfaw B, Poupctová H, Ledvinová J, Sikora J, Vanier MT, et al. Acid sphingomyelinase deficiency. Phenotype variability with prevalence of intermediate phenotype in a series of twenty-five Czech and Slovak patients. A multi-approach study. J Inherit Metab Dis. 2005;28(2):203–27. doi:10.1007/s10545-005-5671-5.
28. McGovern MM, Dionisi-Vici C, Giugliani R, Hwu P, Lidove O, Lukacs Z, et al. Consensus recommendation for a diagnostic guideline for acid sphingomyelinase deficiency. Genet Med. 2017;19(9):967–74. doi:10.1038/gim.2017.7.
29. Lipinski P, Kuchar L, Zakharova EY, Baydakova GV, Lugowska A, Tylki-Szymanska A. Chronic visceral acid sphingomyelinase deficiency (Niemann-Pick disease type B) in 16 Polish patients: long-term follow-up. Orphanet J Rare Dis. 2019;14(1):55. doi:10.1186/s13023-019-1029-1.
30. Ding Y, Li X, Liu Y, Hua Y, Song J, Wang L, et al. Seven novel mutations of the SMPD1 gene in four Chinese patients with Niemann-Pick disease type A and prenatal diagnosis for four fetuses. Eur J Med Genet. 2016;59(4):263–8. doi:10.1016/j.ejmg.2015.11.012.
31. Ricci V, Stroppiano M, Corsolini F, Di Rocco M, Parenti G, Regis S, et al. Screening of 25 Italian patients with Niemann-Pick A reveals fourteen new mutations, one common and thirteen private. Hum Mutat. 2004;in SMPD1(1):105-. doi:. 24(.
32. Zampieri S, Filocamo M, Pianta A, Lualdi S, Gort L, Coll MJ, et al. SMPD1 mutation update: database and comprehensive analysis of published and novel variants. Hum Mutat. 2016;37(2):139–47. doi.
33. Fernández-Burriel M, Peña L, Ramos JC, Cabrera JC, Marti M, Rodríguez-Quiñones F, et al. The R608del mutation in the acid sphingomyelinase gene (SMPD1) is the most prevalent among patients from Gran Canaria Island with Niemann-Pick disease type B. Clin Genet. 2003;63(3):235–6. doi:10.1034/j.1399-0004.2003.00025.x.
34. Liu J, Li J, Wang H, Yan J. Application of deep learning in genomics. Sci China(Life Sciences). 2020;63(12):92–110. doi.
35. Huang L, Khusnutdinova A, Nocek B, Brown G, Xu X, Cui H, et al. A family of metal-dependent phosphatases implicated in metabolite damage-control. Nat Chem Biol. 2016;12(8):621–7. doi:10.1038/nchembio.2108.
36. Munford RS, Sheppard PO, O'Hara PJ. Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. J Lipid Res. 1995;36(8):1653–63. doi.
37. Thurberg BL. Autopsy pathology of infantile neurovisceral ASMD (Niemann-Pick Disease type A): Clinicopathologic correlations of a case report. Mol Genet Metab Rep. 2020;24:100626. doi:10.1016/j.ymgmr.2020.100626.
38. Grafft CA, Fervenza FC, Semret MH, Orloff S, Sethi S. Renal involvement in Neimann-Pick Disease. NDT Plus. 2009;2(6):448–51. doi:10.1093/ndtplus/sfp101.
39. Jerbi M, Sayhi M, Gaied H, Hedri H, Aoudia R, Goucha R, et al. Renal Thrombotique microangiopathy: An unusual renal involvement in Niemann-Pick disease type B. Clin Case Rep. 2020;8(12):3315–20. doi:https://doi.org/10.1002/ccr3.3408.

40. Tierney AL, Nelson CA 3. Brain Development and the Role of Experience in the Early Years. Zero Three. 2009;30(2):9–13. doi:. rd. .
41. Tirelli C, Arbustini E, Meloni F. Bilateral Cystic Bronchiectasis as Novel Phenotype of Niemann-Pick Disease Type B Successfully Treated With Double Lung Transplantation. Chest. 2021;159(5):e293-e7. doi:10.1016/j.chest.2020.11.074.
42. Tian C, Deng T, Zhu X, Gong C, Zhao Y, Wei Y, et al. Evidence of compliance with and effectiveness of guidelines for noninvasive prenatal testing in China: a retrospective study of 189,809 cases. Sci China Life Sci. 2020;63(3):319–28. doi:10.1007/s11427-019-9600-0.
43. Zhang T, Tian X, Xu KF. Cystic fibrosis: a rare disease emerging in China. Sci China Life Sci. 2020;63(7):1082–4. doi.
44. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164-e. doi.

Figures

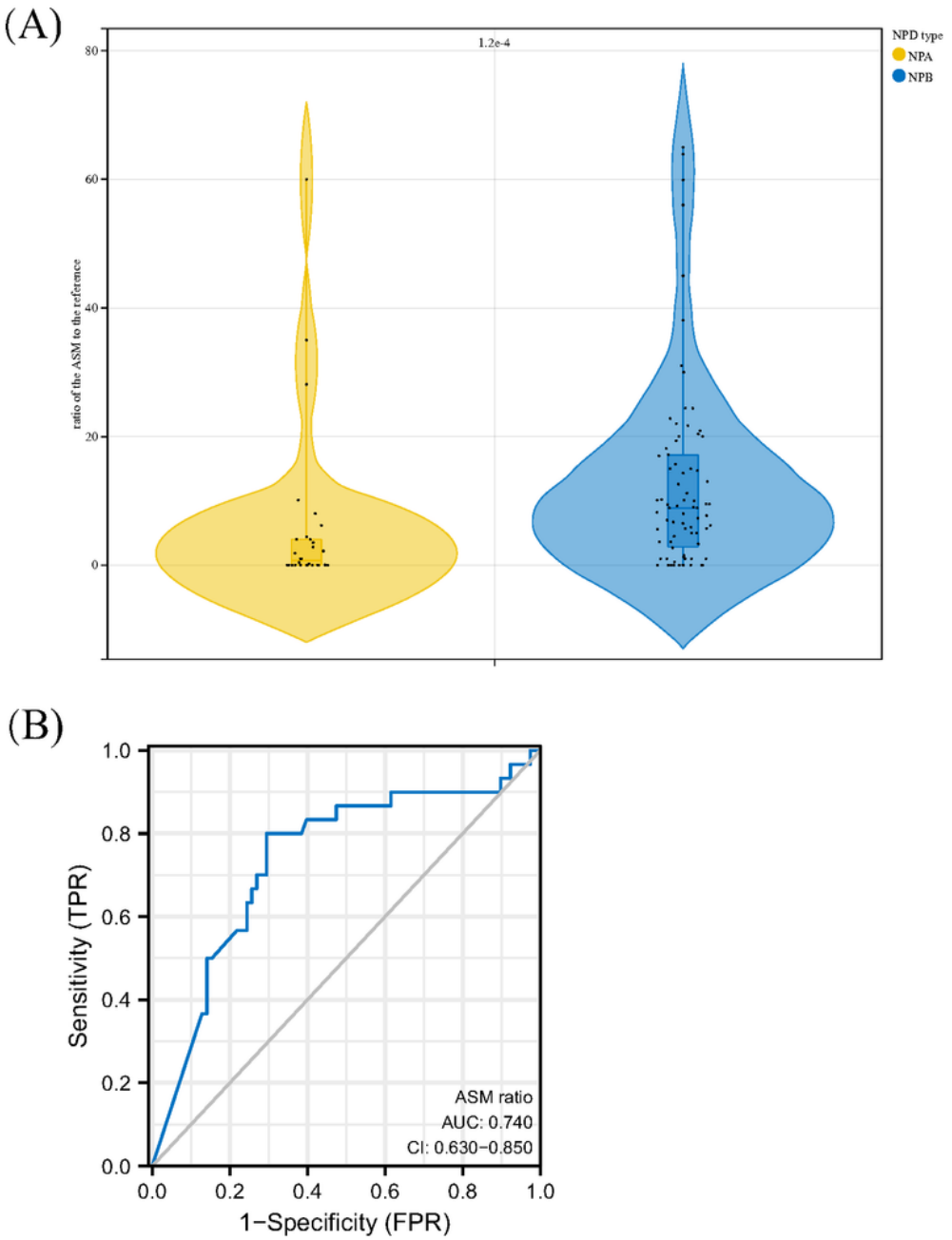

Figure 1

A novel threshold of determining NPD subtypes. (A) ASM levels in the clinical cases are collected. (B) ROC curve for the predictability of the threshold. ASM ratio, the ratio of the activity of acid sphingomyelin phosphodiesterase of the patients to the reference value; NPD, Niemann–Pick disease; NPB, Niemann–Pick disease type B; ROC, receiver operating characteristic curve; AUC, area under the curve.

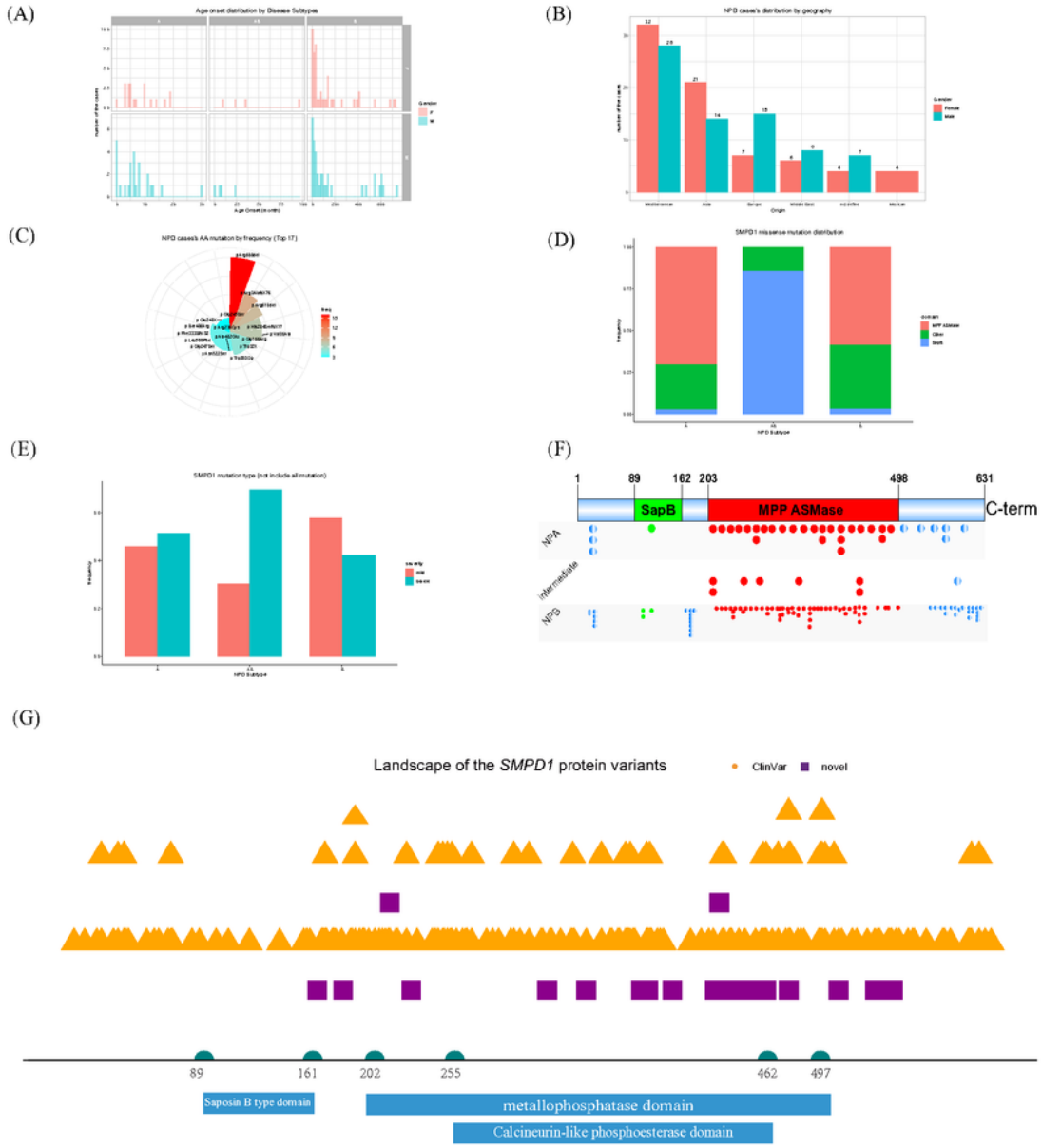

Figure 2

Statistics of the mutation sites on the *SMPD1* gene. (A) Distribution of the patients by onset age. (B) Country of origin of the patients: Mediterranean countries including Italy, Algeria, Spain, Turkey, Maghreb, Jordan, and North Africa; Asian countries such as China and Japan; European countries, such as Caucasian (documented by the research), Poland, Gypsy (documented by the research), and the Netherlands; Middle East countries including Iran and Palestine. (C) Distribution of amino acids (AA) with

altered mutations. (D) *SMPD1* mutation types (do not include all mutations). Severe mutations include deletions, insertions, and nonsense. (E) Distribution of missense mutations on the conserved domain of the human ASM protein. Domains were retrieved from the NCBI (NP_000534.3), namely, saposin (B) (Accession no. 00741), and acid sphingomyelinase and related proteins (MPP ASMase). Each point represents one reported mutation in the collected cases. Points are coloured according to the domains. (G) The landscape of *SMPD1* mutations and 21 novel pathogenic variants prediction based on the databases of ClinVar, ANNOVAR and the EVE model. Purple squares, novel predicted pathogenic variants. Orange triangles, pathogenic/Likely pathogenic variants from the ClinVar. Domains were annotated by ANNOVAR.

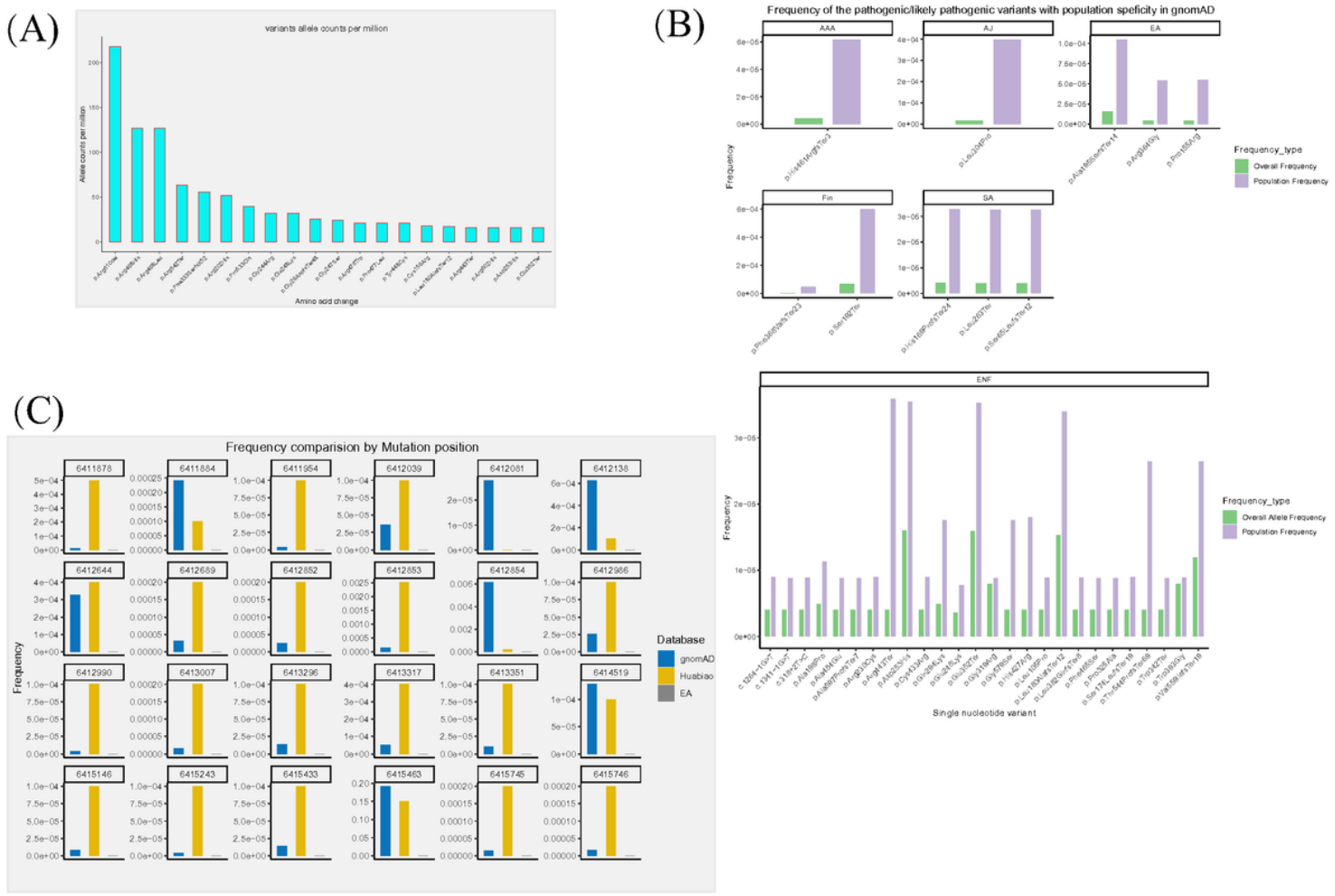

Figure 3

Pathogenic variants allele distribution. (A) Top 20 pathogenic variatns in gnomAD. (B) Frequency of the pathogenic/likely pathogenic variants with population speficity in gnomAD. AAA, African/African American. AJ, Ashkenazi Jewish. EA, East Asian. ENF, European(non-Finnish). Fin, Finnish. LAA, Latino/Admixed American. SA, South Asian. (C) *SMPD1* mutation allele frequency in two data sources. Huabiao, the public project database of whole-exomes of the Chinese Han. gnomAD_EA, data from east Asian of the gnomAD.

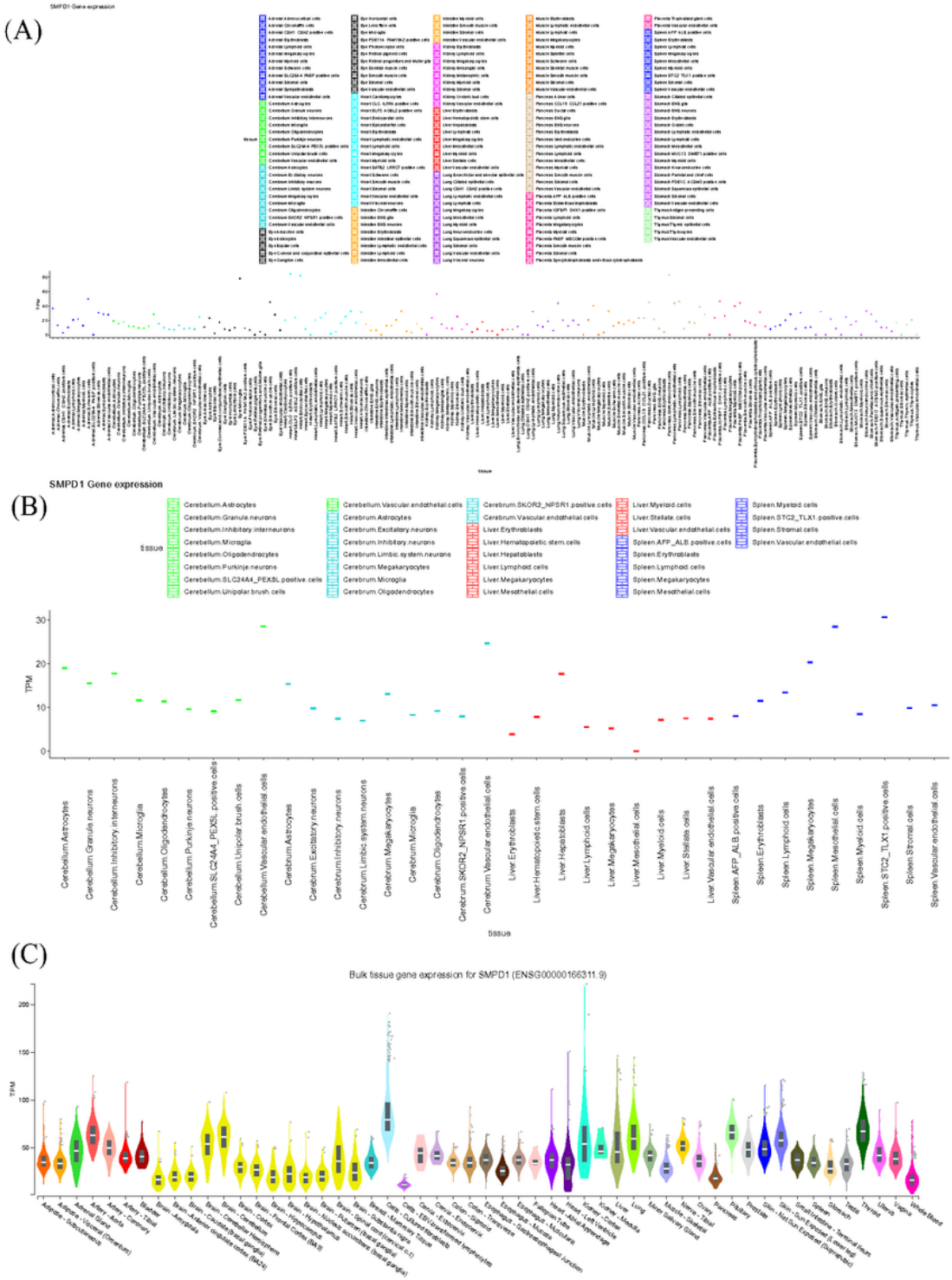

Figure 4

SMPD1 gene expression in all tissues. (A) Expression profiles of the *SMPD1* gene in different organs based on the GSE156793. (B) Expression profiles of the *SMPD1* gene in different cell lines based on the GSE156793. (C) Expression profiles of the *SMPD1* gene in the liver, spleen, and brain based on GTEx data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryTableS3.xlsx