Universität Rostock Traditio et Innovatio

Liquid Chromatography - Mass Spectrometry Untargeted Metabolomics Application Towards Biomarker Discovery for Neuronal Ceroid Lipofuscinosis Type 6 Disease

Kumulative Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Rostock

vorgelegt von

Corina-Marcela Rus

aus Hamburg, geboren am 13.12.1985 in Sânnicolau Mare

Rostock, Februar 2024

https://doi.org/10.18453/rosdok_id00004756

Gutachter:

Prof. Dr. Ralf Zimmermann
Institut für Chemie, Universität Rostock

Prof. Dr. Michael Lämmerhofer
Eberhard-Karls-Universität Tübingen

Jahr der Einreichung: 2024

Jahr der Verteidigung: 2024

Doktorandinnen/Doktoranden-Erklärung gemäß § 4 Absatz 1 Buchstaben g und h der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Rostock

Corina-Marcela Rus
Richardstraße 48
22081, Hamburg

Ich habe eine Dissertation zum Thema:

"Liquid Chromatography - Mass Spectrometry Untargeted Metabolomics Application Towards Biomarker Discovery for Neuronal Ceroid Lipofuscinosis Type 6 Disease"

an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Rostock angefertigt. Dabei wurde ich von Herrn **Prof. Dr. Ralf Zimmermann** betreut.

Ich gebe folgende Erklärung ab:

1. Die Gelegenheit zum vorliegenden Promotionsvorhaben ist mir nicht kommerziell vermittelt worden. Insbesondere habe ich keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen/Betreuer für die Anfertigung von Dissertationen sucht oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

2. Ich versichere hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig angefertigt und ohne fremde Hilfe verfasst habe. Dazu habe ich keine außer den von mir angegebenen Hilfsmitteln und Quellen verwendet und die den benutzten Werken inhaltlich und wörtlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Rostock, den ....

Corina-Marcela Rus

LIST OF PUBLICATIONS

First authorship

Title: Dried Blood Spot (DBS) Methodology Study for Biomarker Discovery in Lysosomal Storage Disease (LSD)

Author: **Corina-Marcela Rus**, Sebastiano Di Bucchianico, Claudia Cozma, Ralf Zimmermann, and Peter Bauer

Journal: *Metabolites.* 11(6):382

Year: 2021

Doi: 10.3390/metabo11060382

Status: Published

Contributions: *Corina-Marcela Rus organized and performed the experimental work, conducted the MS experiments, analyzed and interpreted the data, created the figure, and wrote the manuscript.*

Title: Clinical and genetic characterization of a cohort of 97 CLN6 patients tested at a single center

Author: **Corina-Marcela Rus**, Thomas Weissensteiner, Catarina Pereira, Iuliana Susnea, Bright D Danquah, Galina Morales Torres, Maria Eugenia Rocha, Claudia Cozma, Deepa Saravanakumar, Sumanth Mannepalli, Krishna K Kandaswamy, Sebastiano Di Bucchianico, Ralf Zimmermann, Arndt Rolfs, Peter Bauer, and Christian Beetz

Journal: *Orphanet J Rare Dis.* 17(1):179

Year: 2022

Doi: 10.1186/s13023-022-02288-8

Status: Published

Contributions: *Corina-Marcela Rus contributed to the study conceptualization and investigation, data collection and analysis, figure creation, and the writing of the original draft.*

Title: Neuronal progenitor cells-based metabolomics study reveals dysregulated lipid metabolism and identifies putative biomarkers for CLN6 disease

Author: **Corina-Marcela Rus**, Daniel L. Polla, Sebastiano Di Bucchianico, Steffen Fischer, Jörg Hartkamp, Guido Hartmann, Yunus Alpagu, Claudia Cozma, Ralf Zimmermann and Peter Bauer

Journal: *Sci Rep* **13**, 18550

Year: 2023

DOI: 10.1038/s41598-023-45789-7

Status: Published

Contributions: *Corina-Marcela Rus performed the cell cultivation, sample preparation, MS experiments, data processing, statistical analysis, result interpretation, figure design, and wrote the manuscript.*

Related publications

Cao X, Padoan S, Binder S, Bauer S, Orasche J, **Rus CM**, Mudan A, Huber A, Kuhn E, Oeder S, Lintelmann J, Adam T, Di Bucchianico S, Zimmermann R. **A comparative study of persistent DNA oxidation and chromosomal instability induced in vitro by oxidizers and reference airborne particles**. *Mutat Res Genet Toxicol Environ Mutagen*. 2022 Feb-Mar;874-875:503446. doi: 10.1016/j.mrgentox.2022.503446.

ACKNOWLEDGMENTS

I am grateful to have received support and guidance from various individuals and departments throughout my academic journey. Firstly, I acknowledge **Centogene GmbH**, the Rare Disease Company in Rostock, for supporting this study. During this time, I held the position of an industrial Ph.D. student working full-time, which allowed me to conduct this research.

I am deeply appreciative to **Prof. Dr. Ralf Zimmermann** for the opportunity to undertake my thesis as an external doctoral student at the University of Rostock. His support and acceptance were crucial in making this work possible.

Furthermore, I am thankful to **Prof. M.D. Peter Bauer** for his mentorship throughout the project. His feedback and expertise were instrumental in shaping my research and overcoming the numerous challenges that arose. I would also like to acknowledge **Prof. M.D. Arndt Rolfs** for the chance to pursue an industrial Ph.D. at Centogene GmbH. Without his vision and belief, this work would not have been initiated.

My gratitude extends to **Dr. Claudia Cozma** and **Dr. Yunus Alpagu** for their scientific support and guidance, which let me conduct my research in alignment with my scientific objectives. Thank you for the freedom to pursue my research as I deemed appropriate, for which I am grateful.

I thank **Dr. Sebastiano Di Bucchianico** for his unwavering guidance throughout the project, for addressing my queries with clarity and precision, and **Dr. Nikenza Viceconte** for consistently providing the necessary support and effort to ensure this project's smooth progression and completion. Your contributions ensured the project remained on track and reached its ultimate goal.

Lastly, I sincerely thank the entire Centogene team, especially the R&D Biochemistry & iPSC department, for providing essential resources, facilities and support throughout the process.

Egan, you have my deepest gratitude.

ZUSAMMENFASSUNG

Neuronale Ceroid-Lipofuszinose Typ 6 (CLN6) ist eine seltene, erbliche, neurodegenerative Krankheit im Kindesalter, die das Nervensystem betrifft. Diese Studie zielt darauf ab, mithilfe der Flüssigchromatographie – Massenspektrometrie (LC-MS) - basierte Metabolomik das globale Stoffwechselprofil und die Stoffwechselwege bei CLN6-betroffenen Patienten zu erforschen. Die Metabolomik ist ein leistungsstarkes Instrument zur Identifizierung und Quantifizierung kleinen Molekülen, die an Stoffwechselreaktionen beteiligt sind, und LC-MS ist eine hochsensitive Plattform zur Entdeckung von Biomarkern. Herausforderungen bei der Entwicklung von Biomarkern für seltene Krankheiten umfassen begrenzte Proben, genetische und Umweltvariabilität, strenge Studiendesigns und die Einhaltung etablierter Protokolle.

In dieser Studie wurde die Machbarkeit der Verwendung von Trockene Blutspots Proben zur Entdeckung von Biomarkern auf Metabolomikbasis untersucht. Blutproben wurden von nicht betroffenen Probanden entnommen, auf Filterkarten aufgetropft, getrocknet und bei -20°C gelagert. Um mögliche Einflüsse auf das Metabolom der Probanden zu untersuchen, wurde eine vielfältige Patientenkohorte einbezogen, die Personen mit unterschiedlichen demografischen und medizinischen Merkmalen umfasste, darunter Alter, Geschlecht, Tageszeit der Probenahme und Lagerungszeit. Es wurde die Auswirkung von Lagerungsbedingungen, Extraktionsprotokollen und der Schichtung der Proben auf die Genauigkeit der Metaboliten untersucht. In einer sechsjährigen Stabilitätsstudie wurden optimale Lagerungsbedingungen ermittelt. Die DBS-Proben wurden in verschiedenen Batch-Kombinationen analysiert, um metabolische Unterschiede zwischen den Kontrollen und den von der Diagnose betroffenen Patienten zu identifizieren.

Darüber hinaus wurden genotypisch-phenotypische Parameter bei 97 CLN6-Patienten analysiert, was zur Identifizierung von 24 neuen Varianten führte und das Verständnis der CLN6-Krankheit erweiterte. Des Weiteren erweiterte die Studie ihren Fokus auf die Metabolomik von Zelllinien, um tiefere Einblicke in den Stoffwechsel bei CLN6 zu gewinnen und einen Rahmen für zukünftige Forschung an neuronalen Zellen zu etablieren. Potenzielle Biomarker für die Frühdiagnose von CLN6, wie C16 GlcCer, C24 GlcCer, C24:1 GlcCer und Glycerophospholipide PG 40:6 und PG 40:7, wurden identifiziert.

Diese Forschung zeigte das Potenzial der LC-MS-basierten Metabolomik zur Erforschung der mit CLN6-Erkrankungen verbundenen Stoffwechselveränderungen und zur Entdeckung potenzieller

Biomarker mit diagnostischen und klinischen Implikationen auf. Sie betont die Bedeutung eines stabilen LC-MS-Ansatzes, einer gut definierte Patienten-Kohorte und geeigneter Probentypen für die Entdeckung von Biomarkern bei der CLN6-Krankheit.

ABSTRACT

Neuronal Ceroid Lipofuscinosis Type 6 (CLN6) is a rare, inherited, neurodegenerative childhood disease that affects the nervous system. This thesis aimed to explore the global metabolic profiling in CLN6-affected patients using LC-MS-based metabolomics. Metabolomics is a powerful tool for identifying and quantifying small molecules involved in metabolic reactions, and Liquid Chromatography - Mass Spectrometry (LC-MS) is a highly sensitive platform suitable for biomarker discovery. Challenges in developing biomarkers for rare diseases include limited samples, genetic and environmental variability, rigorous study design, and adherence to established protocols.

This thesis assessed the feasibility of dried blood spot (DBS) card samples for metabolomics-based biomarker discovery. Whole blood from unaffected donors was collected and dripped onto filter cards, dried, and stored at -20°C. A diverse patient cohort, comprising individuals with varying demographics and medical conditions as well as age, sex, time of day at sampling, and storage time, was sampled to address potential influences on the subjects' metabolome. The impact of storage conditions, extraction protocols, and sample stratification on metabolite accuracy was investigated. A six-year stability study determined optimal storage conditions. DBS samples were analyzed across different batch combinations to compare metabolic disparities between controls and disease-affected patients.

Additionally, genotypic-phenotypic features were analyzed in 97 CLN6 patients, leading to the identification of 24 new variants and enhancing the understanding of CLN6 disease. Furthermore, the study expanded its focus on cell lines' metabolomics to gain deeper insights into CLN6 metabolism and establish a framework for future research using neuronal cells. Potential biomarkers for early diagnosis of CLN6, such as C16 GlcCer, C24 GlcCer, C24:1 GlcCer, and glycerophospholipids PG 40:6 and PG 40:7, were identified.

This research demonstrated the potential of LC-MS-based metabolomics in understanding the metabolic changes associated with CLN6 diseases and uncovering potential biomarkers with diagnosis and clinical trial implications. It highlights the importance of a stable LC-MS approach, a well-defined patient cohort, and suitable sample types for biomarkers discovery in CLN6 disease.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

EDTA	ethylenediaminetetraacetic acid
BDNF	brain-derived neurotrophic factor
CLN6	neuronal ceroid lipofuscinosis type 6
CSF	cerebrospinal fluid
DBS	dried blood spot
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EI	electron ionization
EMA	European medicines agency
ER	endoplasmic reticulum
ESI	electrospray ionization
GC-MS	gas chromatography-mass spectrometry
GenBank	genetic sequence data bank
GFAP	glial fibrillary acidic protein
$HDMS^E$	high-definition MS^e
HMDB	human metabolome database
IPA	ingenuity pathway analysis
kDa	kilodalton
KEGG	Kyoto encyclopedia of genes and genomes
LC-MS	liquid chromatography - mass spectrometry
LSD	lysosomal storage disorder
LTQ	linear trap quadrupole
MAP2	Microtubule-associated protein 2
MRI	magnetic resonance imaging
MS	mass spectrometry
NCL	neuronal ceroid lipofuscinosis
NGS	next generation sequencing
NMR	nuclear magnetic resonance
NPC	neuronal progenitor cell
OMIM	online Mendelian inheritance in man

OPL- DA	orthogonal partial least squares discriminant analysis
PCA	principal component analysis
PG	phosphatidylglycerol
PLS-DA	partial least squares-discriminant analysis
QQQ	triple quadrupole
Q-ToF	quadrupole time of flight
RNA	ribonucleic acid
RT	retention time
TUJ1	neuron-specific class iii beta tubulin
VCAM-1	vascular cell adhesion protein 1
VIP	variable importance for the projection
WES	whole exome sequencing
WGS	whole-genome sequencing

LIST OF FIGURES

1 INTRODUCTION

1.1 RARE DISEASES

1.1.1 Understanding Rare Diseases: prevalence and impact

Rare diseases are a group of unique diseases numbering between 6000-8000. Genetic factors influence them in 80% of cases, of which 50 to 75% have pediatric onset and altogether, it impacts over 300 million individuals worldwide (Plaiasu et al., 2010, Orphanet, 2024). Although each condition is rare, affecting less than 50 per 100,000 individuals in the European Union or fewer than 200,000 individuals in the USA, their combined prevalence is significant (European Parliament, 2000, The Food and Drug Administration, 2024). Rare diseases affect approximately 3.5–5.9% of the global population worldwide, totaling 263 to 446 million people. In Europe and the USA alone, around 60 million individuals are affected (Nguengang Wakap et al., 2020). Regardless of how many individuals are affected, rare diseases deeply impact individuals and their families. Rare diseases can affect individuals of diverse ages, races, and ethnicities, surpassing demographic boundaries. Nevertheless, the overall statistic is particularly concerning as it underscores the significant impact on children, with a mortality rate of 30% among affected individuals even before they reach the age of 5 (Lunke et al., 2023). Typically chronic, progressive, and life-threatening, they impose significant challenges on patients, their families, and healthcare systems alike. While rare diseases can manifest in various organ systems, a notable proportion presents with neurological symptoms, including cognitive decline, motor dysfunction, and sensory impairments (Witt et al., 2023, Sirbu et al., 2023).

The occurrence of rare diseases varies widely depending on the specific condition. Some rare diseases, such as multiple sclerosis (90:100,000), narcolepsy (50:100,000), Fabry disease (30:100,000), and cystic fibrosis (25:100,000), are relatively well-known and affect more significant numbers of individuals, while others are exceedingly rare, with only a handful of cases reported worldwide. Despite their rarity, the impact of rare diseases on affected individuals and their families is profound. Patients often face significant challenges in obtaining an accurate diagnosis, accessing appropriate medical care, and receiving effective treatments. The rarity of these diseases can lead to delays in diagnosis, misdiagnosis, and limited treatment options, resulting in increased mortality rates (Delaye et al., 2022). Rare diseases also present significant economic challenges beyond the physical and emotional impact on patients and their families. The substantial

costs of managing rare diseases, limited treatment options, and the necessity for specialized medical care can impose a significant financial burden on individuals, families, and healthcare systems (Navarrete-Opazo et al., 2021).

Neurodegenerative rare diseases represent a significant subset of rare diseases and encompass a diverse range of conditions affecting primarily the neurons of the central nervous system. These disorders can result from genetic mutations, metabolic abnormalities, autoimmune reactions, infectious agents, or environmental factors, leading to various clinical presentations and disease courses (Perrone et al., 2021, Matilla-Dueñas et al., 2017). Among neurodegenerative rare diseases, neuronal ceroid lipofuscinoses (NCLs), also called Batten disease, are a group of inherited disorders characterized by the abnormal buildup within the cells of a lipofuscin pigment. There are several subtypes of NCLs, each caused by mutations in different genes and characterized by distinct clinical features and disease courses. CLN6 is one of the less common variants of NCL, with mutations in the *CLN6* gene leading to progressive neurological deterioration and premature death (Mole et al., 2020).

1.1.2 Challenges in diagnosis and therapeutic approaches

The diagnostic process is one of the most significant challenges in managing rare diseases. Due to the rarity and heterogeneity of these conditions, healthcare providers require familiarity with specific diseases, leading to delays or errors in diagnosis. Patients with rare diseases often undergo extensive diagnostic evaluations before final diagnosis, including genetic testing, imaging studies, and biochemical assays. Furthermore, the symptoms of rare diseases often overlap with those of more familiar conditions, complicating the diagnostic process further. Misdiagnosis and delayed diagnosis can have severe consequences for patients, delaying the initiation of appropriate treatments and potentially worsening disease progression and complications. In addition to diagnostic challenges, rare diseases also present significant therapeutic obstacles. Many rare diseases lack effective treatments, limiting patients' chances to address their conditions. The development of treatments for rare diseases is often hindered by factors such as small patient populations, limited or no research funding, and regulatory burdens. Moreover, the high cost of developing therapies for rare diseases can discourage pharmaceutical companies from investing in drug development efforts. The small market size for rare disease therapies may not justify the

substantial investments required to bring a new therapy to market, resulting in a lack of incentives for drug development (Uhlenbusch et al., 2019, Lopes et al., 2018, Groft et al., 2021).

1.1.3 Rare diseases diagnosis through analytical techniques

Despite the challenges in diagnosing and treating rare diseases, ongoing progress in biomedical research brings hope to those affected. Advanced omics technologies have transformed our understanding of these conditions, playing a crucial role in revealing disease mechanisms and potential treatment targets. These technologies help pinpoint disease-causing mutations accurately, guiding treatment decisions when options are available. As these tools become more accessible and cost-effective, healthcare professionals and researchers can delve into the genetic aspects of rare diseases in individual patients, uncover disease-specific protein patterns, and identify markers for diagnosis and prognosis. Furthermore, these advancements revealed disrupted metabolic pathways linked to rare diseases (Crowther et al., 2018, Kerr et al., 2020, Lunke et al., 2023).

Recent advancements in analytical chemistry offer a promising pathway toward enhancing the accuracy of diagnosing rare diseases. For instance, in CLN6 disease, characterized by progressive cognitive and motor function deterioration, implementing high-throughput omics technologies presents significant potential in revolutionizing diagnostic and therapeutic strategies. Furthermore, the discovery of omics-driven biomarkers facilitates non-invasive monitoring of disease progression and evaluation of treatment responses, thereby increasing clinical decision-making and optimizing patient care. The integration of high-throughput omics technologies marks a pivotal change in the diagnosis and treatment of rare diseases. Through comprehensive molecular profiling, these methodologies provide invaluable insights into the underlying mechanisms of diseases, enabling early detection and fast therapeutic interventions (Casetta et al., 2020, Chen et al., 2023, Jiang et al., 2011).

1.2 NEURONAL CEROID LIPOFUSCINOSIS

1.2.1 History of neuronal ceroid lipofuscinosis

Neuronal ceroid lipofuscinosis (NCL; Batten disease) are a group of fatal rare inherited disorders primarily affecting the nervous system. The worldwide incidence of NCL, as a group, is estimated at a rate of 2–4 per 100,000 live births, while the prevalence is approximately 2-4 cases per 1,000,000, mainly data from Western countries where the molecular diagnosis has become the

norm over the last 10-15 years (Mole and Haltia, 2015, Rider and Rider, 1988, Santavuori, 1988, Williams, 2011, Teixeira et al., 2003, Sleat et al., 2016); therefore global wide, the exact frequency and prevalence is difficult to determine due to under -and - misdiagnosed cases.
The first published NCL cases were in 1903 and 1826 by Batten and Stengel. Batten reported two siblings with brain and macular degeneration (Batten), and Stengel documented four siblings with vision impairment, cognitive decline, speech loss, and seizures (Stengel, 1982). Since then, 14 subtypes of NCL (**Figure 1**) have been discovered (Mole and Cotman, 2015) that exhibit broadly similar clinical and pathological features such as the buildup of ceroid lipofuscin storage material within lysosomes, accompanied by the deterioration of motor and cognitive functions, seizures, and vision impairment (Nelvagal et al., 2020, Zeman and Dyken, 1969) yet, each caused by a mutation unique to a specific gene (*CLN1-CLN8, CLN10-CLN14*) (Schulz et al., 2013, Kousi et al., 2012, Butz et al., 2020).

Nevertheless, the onset and disease progression exhibit diverse patterns , even among members of the same family (Warrier et al., 2013). The first gene identified was the PPT1 (Vesa et al., 1995), followed by CLN3 (1995). Subsequently, TPP1 (Sleat et al., 1997), CLN5 (Savukoski et al., 1994), CLN8 (Ranta et al., 1999), CLN6 (Wheeler et al., 2002, Gao et al., 2002), CTSD (Steinfeld et al., 2006, Siintola et al., 2006), MFSD8 (Siintola et al., 2007) and DNAJC5 (Nosková et al., 2011) were identified. The most recent discoveries include GRN (Smith et al., 2013), ATP13A2 (Bras et al., 2012), CTSF (Smith et al., 2013) and KCTD7 (Staropoli et al., 2012). However, no causative gene has been identified for CLN9 disease (Warrier et al., 2013).

There is still anticipation for additional NCL genes to be discovered as CLN15 has partially been associated with NCL (Beck-Wodl et al., 2018). Depending on the disease's genetic variant, the disease's onset and progression vary considerably (Warrier et al., 2013). The inheritance pattern of all NCL subtypes is autosomal recessive except for the adult-onset form (ANCL) which can be inherited in either an autosomal recessive or dominant manner (Nosková et al., 2011). Previously, NCLs were classified by the child's age at onset and their clinical features into congenital, infantile (INCL), late-infantile (LINCL), juvenile (JNCL), and adult (ANCL) neuronal ceroid lipofuscinoses. Their shared clinical spectrum encompassed a combination of cognitive and motor decline, visual impairment, seizures, and psychomotor regression. Furthermore, there is notable neuronal, primarily observed in the cerebral and cerebellar cortices (Mole et al., 2005). The

sequence in which these symptoms manifest varies depending on the subtype, affecting children during neonatal, early childhood, or young adulthood. Unfortunately, the prognosis is invariably fatal.

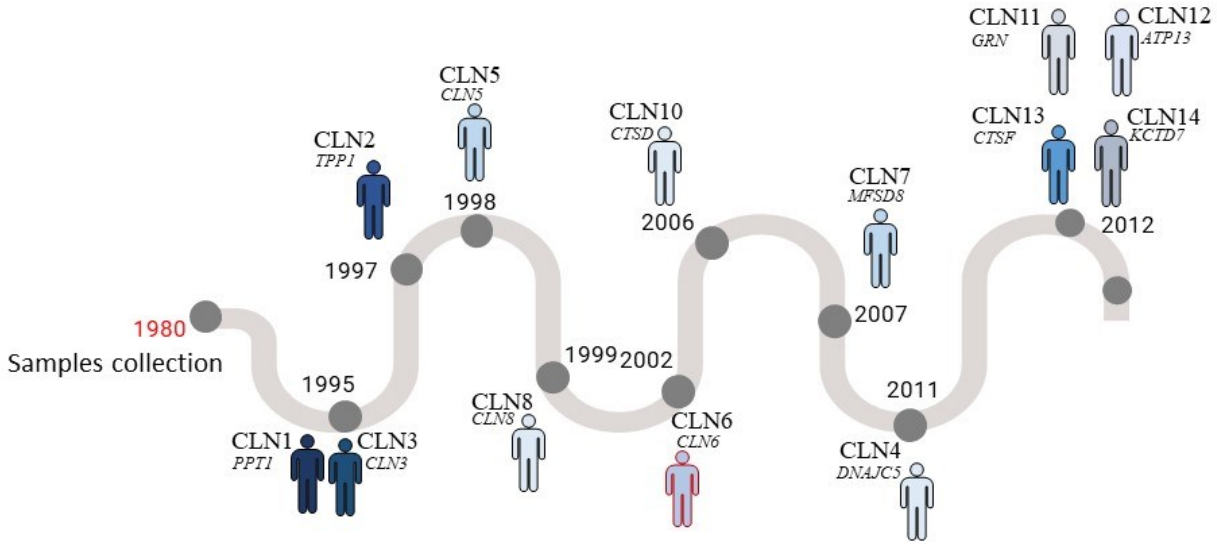


Figure 1. *Timeline for identifying genes associated with CLN diseases.* ***The*** *figure shows the chronological order of the identification of genes related to various types of CLNs, the corresponding disease names, and the years in which the genes were identified.*

1.2.2 Neuronal ceroid lipofuscinosis type 6 (CLN6)

Neuronal ceroid lipofuscinosis type 6 (CLN6) (OMIM 601780) is the late infantile form of NCL and, like other types of NCL, is an inherited rare autosomal recessive disorder caused by a deficiency in the *CLN6* gene (Mole et al., 2004). The *CLN6* gene is located on 15q23 and encodes a 311-amino acid transmembrane endoplasmic reticulum protein with a structure composed of an N-terminal cytoplasmic domain, seven putative transmembrane domains, and a luminal C-terminus (Wheeler et al., 2002, Gao et al., 2002). Mutations in the *CLN6* gene, which encodes the ceroid lipofuscinosis neuronal protein 6 (a transmembrane protein of unknown function), lead to the loss of function of the *CLN6* gene (GenBank accession no. NM_017882.3.) and trigger the onset of CLN6 disease (Teixeira et al., 2003, Gao et al., 2002).

Despite two decades of ongoing research since the CLN6 disease identification (Gao et al., 2002, Wheeler et al., 2002), the precise function of the CLN6 proteins has not yet been established. Although it is assumed to be involved in intracellular trafficking and lysosomal function, it is not

yet clear how dysfunction of CLN6 protein - leads to a buildup of lysosomal storage material (Heine et al., 2004, Benedict et al., 2009), posing a continuous obstacle to developing targeted therapies for CLN6 disease. Clinically, CLN6 disease is characterized by a broad spectrum of progressive features that can vary in severity and age of onset (Canafoglia et al., 2015). The first signs appear between the ages of 2 and 4 years (Mole and Cotman, 2015), affected individuals exhibit motor regression, visual impairment, tremors, and seizures (Cannelli et al., 2009, Nita et al., 2016), and as the disease progresses, patients experience ataxia, advanced cognitive decline, and motor impairments (Nafi et al., 2019).

1.2.3 Diagnosis

Considering that there is rapid disease progression, an early diagnosis is imperative for the effective management and future treatment of the CLN6 disease and also for other forms of NCL to improve the quality of life of affected individuals (Augustine et al., 2021, Fietz et al., 2016). The timely implementation of appropriate disease management may help slow the disease progression and facilitate genetic counseling and testing for family members to identify at-risk individuals (Kaminiów et al., 2022). Furthermore, early diagnosis is critical for screening and randomizing the patients for inclusion in clinical trials and research studies, which, if successful, will lead to effective treatments (Kliegman et al., 2017). The Clinical trials investigating potential disease-modifying therapies for NCL are more effective when initiated early in the disease course. Efforts are underway to enhance the diagnosis of NCL, focusing on developing novel biomarkers and imaging techniques (Johnson et al., 2020). Overall, early diagnosis of NCL is essential for optimal management and treatment of the disease, and further research is needed to improve diagnostic methods besides identifying effective therapies (Trivisano et al., 2022, Simonati and Williams, 2022).

Currently, the diagnosis of CLN6 is based on clinical symptoms, neurological examination, genetic tests, and imaging tests. Genetic testing involves analyzing an individual's DNA sample to identify mutations in the *CLN6* gene that cause the disease. Clinical examinations and imaging tests, such as brain magnetic resonance imaging (MRI), help identify characteristic features of CLN6 disease, such as progressive brain atrophy and white matter abnormalities (Johnson et al., 2019). As these testing methods are costly and invasive, there is a continuous need for more advanced, less invasive, cost-effective diagnostic methods. One suitable method would be using non-invasive

biomarkers, potentially reducing the need for invasive diagnostic procedures (Johnson et al., 2020, Best et al., 2021, Iwan et al., 2021).

1.2.4 Treatment options

In recent years, several clinical trials have been initiated to develop targeted treatments for various types of CLNs. Presently, targeted therapy is underway for CLN2, CLN3, CLN5, and CLN7 disease reflecting significant advancement in this area of research (Rosenberg et al., 2019, Johnson et al., 2019) (ClinicalTrials.gov). Among the promising developments is the possibility of small-molecule therapies, gene therapies, and enzyme replacement treatments (Specchio et al., 2020, Johnson et al., 2019). Upon their effectiveness and more extensive scale application, these clinical trials may have relevance to CLN6 disease, yet it is crucial to acknowledge that the likelihood of success is uncertain (Mole et al., 2019).

The only interventional trial focused on human CLN6 disease was the Phase I/II Gene Transfer Clinical Trial, an open-label, single-dose study of AT-GTX-501 administered by a single intrathecal injection started in 2016 by Amicus Therapeuticus (Clini calTrials.gov: NCT02725580). However, this clinical attempt has yet to be successful so far. Unfortunately, the current lack of knowledge regarding the functions and mechanisms of CLN6 has presented a significant obstacle in identifying effective therapeutic targets for treating CLN6 disease (Kollmann et al., 2013, Kyttala et al., 2006). Currently, clinical treatment is focused on symptom management and includes the management of seizures, sleep alterations, extrapyramidal symptoms, behavioral disturbances, anxiety, and psychosis (Nita et al., 2016). Thus, developing a targeted treatment for CLN6 is of utmost importance. However, until such a therapy becomes available, early disease diagnosis remains critical in assisting affected individuals and their families (Fietz et al., 2016).

1.2.5 Biomarker studies on NCL

Various technologies have been utilized in NCL biomarker research to identify and monitor disease progression. One notable contribution in this field comes from the study conducted by Hersrud et al. Their research employed a plasma-based approach and revealed elevated Clusterin, Adiponectin, Apolipoprotein E, and Apolipoprotein E levels in juvenile NCLs (Hersrud et al., 2016). Similarly, Sleat et al. utilized a CSF-based proteomics approach in CLN1, CLN2, and

CLN3, identifying 18 altered proteins across these diseases (Sleat et al., 2017). Although these biomarker candidates show potential for early diagnosis and treatment monitoring, their effectiveness is yet to be determined (Kline et al., 2020). Until then, they remain potential indicators of early-onset neurodegeneration. This demonstrates the difficulty in finding clinically valuable biomarkers despite substantial advancements and the application of cutting-edge technologies. To progress towards the clinical stage, it is crucial to analyze a suitable cohort of samples (Pepe et al., 2015). Therefore, in the context of rare diseases like NCL, integrating multiple platforms is essential to enhance our understanding of the disease and uncover a promising path for biomarker identification.

1.3 METABOLOMICS

Metabolomics has emerged as a powerful tool for identifying and quantifying small molecules (80-1200 kDa) known as metabolites, collectively forming the human metabolome. These metabolites provide valuable insights into biochemical pathways, their regulation, and their role in biological growth, maintenance, and functioning (Hollywood et al., 2006, Nicholson and Lindon, 2008, Wishart, 2007). However, the extensive datasets generated by metabolomics make it challenging to fully comprehend, particularly when studying the highly complex human metabolome. When adding to this the dynamic nature of the metabolome earned to the fact that the composition of the metabolome varies across different organisms and is constantly changing due to various chemical processes, it brings a significant challenge in fully defining its extent (Holmes et al., 2008a, Holmes et al., 2008b).

The human metabolome is influenced by internal factors like genetics, age, and health (Carayol et al., 2015, Shin et al., 2014, Xiao et al., 2014) and the experimental methodologies employed during research. These methodologies include sample collection, sample type, sample storage, stratification of patient samples, and technical processes used. Considering these parameters is crucial for obtaining reliable and accurate results during data analysis. The examination of metabolite levels in diverse sample types, such as DBS or cell lines, nonetheless offers insightful information about the pathophysiology of diseases (Čuperlović-Culf et al., 2010). Dried blood spots (DBS), particularly, have gained interest as a non-invasive and practical tool for blood-based research due to their processing, storage, and transportation simplicity. Analyzing metabolites in DBS has shown promising results in identifying biomarkers associated with various diseases and

metabolic disorders. Similarly, studying the metabolic profiles of cultured cells can reveal important insights into disease mechanisms and disease-specific biomarkers.

By incorporating biomarker analyses based on DBS and cultured cells, we can enhance our understanding of the human metabolome, paving the way for better disease monitoring, diagnostics, and advancements in biomedical research.

However, several variables must be considered when using dried blood spots (DBS) in biomarker discovery studies. For instance, the age of the DBS card can affect the stability of analytes, potentially impacting the reliability of biomarker studies. Storage conditions, including temperature, humidity, and light exposure, can influence metabolite stability. Therefore, proper storage protocols are essential to minimize sample quality degradation. Another critical consideration is the choice of extraction solvent for metabolomics analysis, which can impact metabolite recovery and detection. Optimization of the extraction protocol is necessary to target specific metabolites of interest. These factors and others must be carefully addressed when utilizing DBS in biomarker discovery studies. By understanding and effectively managing these elements, researchers can improve the reliability and reproducibility of metabolomic analyses using DBS samples.

1.3.1 Untargeted and targeted metabolomics

Two primary approaches are used in metabolomics: targeted and untargeted metabolomics (Patti et al., 2012a). Targeted metabolomics involves the analysis of a specific set of metabolites in a biological sample, whereas untargeted metabolomics is the identification and quantification of as many metabolites as possible without prior knowledge of their chemical identity. Both approaches have advantages and limitations, and the choice of methods depends on the specific research question and available technical resources (Schrimpe-Rutledge et al., 2016).

Targeted metabolomics is ideal for quantifying a specific set of metabolites accurately but may not be suitable for detecting novel metabolites or unexpected changes in the metabolome. In contrast, untargeted metabolomics can capture the entire metabolome and identify unknown metabolites but requires more accurate quantification of specific metabolites than targeted metabolomics (Ribbenstedt et al., 2018, Patti et al., 2012b). While untargeted metabolomics has great potential, further application of the unknown metabolites obtained from this method can take time and effort. One of the challenges is the need for peak identification in the metabolite profile. This requires

more than precise mass measurement to annotate and comprehensively determine the structure of these metabolites (Goodacre et al., 2004).

1.3.2 Analytical platforms

Recent years have brought significant advancements in metabolomics, resulting in the development of disease-targeted procedures and techniques precisely tailored for diverse clinical applications (Zhang et al., 2016). However, due to the chemical complexity of the metabolome, no single analytical methodology can capture all the metabolites present in a biological sample. Therefore, a combination of techniques is often used for metabolome analysis (Halket et al., 2005, Shulaev, 2006). Yet, the vast number of metabolites and their continuous fluctuations make a comprehensive analysis challenging. Further research and development of novel analytical tools are necessary to enhance our understanding of the metabolome. Nonetheless, with the evolution of metabolomics technology, other mass detectors have caught attention for investigating a more comprehensive range of metabolites (Liebal et al., 2020, Cao et al., 2022).

Mass spectrometry (MS) is one of the major detectors used in metabolomics due to its sensitivity and the ability to screen on a large scale. The MS-based technique offers a quantitative analysis of metabolites with high selectivity, sensitivity, and accuracy (Emwas, 2015), requiring separation techniques to distinguish metabolites based on their properties, such as mass, charge, and retention time (RT), and they have become the most widely used approach for analyzing the metabolome. Separation technologies most commonly coupled with MS include gas chromatography–mass spectrometry (GC-MS), gas chromatography–tandem mass spectrometry (GC-MS/MS), liquid chromatography–mass spectrometry (LC-MS), liquid chromatography–tandem mass spectrometry (LC-MS/MS), ultra-performance liquid chromatography–mass spectrometry (UPLC-MS), nuclear magnetic resonance – mass spectrometry (LC-NMR-MS), and capillary electrophoresis – mass spectrometry (CE-MS) (Kałużna-Czaplińska, 2011).

Among these, liquid chromatography MS (LC-MS), gas chromatography MS (GC-MS), and high-performance liquid chromatography (HPLC-MS) are widely used for metabolomic analysis due to their high precision and sensitivity (Lu et al., 2008, O'Hagan et al., 2007). GC-MS has the highest resolving power but limited selectivity (Wei et al., 2012, Lenz and Wilson, 2007), whereas LC-MS is often preferred due to its rapid sample preparation and small sample size, despite having lower

resolving power than GC-MS (Lu et al., 2008). LC is the most widely used separation technology in metabolomics, as it allows for a wide range of metabolites and avoids chemical derivatization (Lopes et al., 2017). These techniques generate large and complex datasets, which can be used to identify and quantify a wide range of metabolites. In recent years, high-resolution mass spectrometry (HRMS) has emerged as a promising tool for metabolite profiling due to its superior mass accuracy and resolving power, fostering better characterization of unknown metabolites and precise identification of adducts (Junot et al., 2014). Overall, MS-based techniques are commonly used to profile the metabolome of various diseases to identify disease-associated metabolites. Regarding the pros and cons, the three primary analytical methods used in metabolomics, namely NMR, GC-MS, and LC-MS, have been extensively reviewed (Wishart et al., 2022).

1.3.3 Untargeted metabolomics workflow

Untargeted metabolomics has emerged as a powerful technique for identifying disease biomarkers as it enables the simultaneous analysis of numerous metabolites without prior knowledge of the metabolites of interest, making it an attractive tool for biomarker discovery. The typical process involves sample collection and preparation, conducting LC-MS analysis, performing statistical analysis, and identifying biomarkers (Lee et al., 2010, Alseekh et al., 2021).

1.3.3.1 Sample source

Biological samples commonly analyzed in untargeted metabolomics include EDTA blood, plasma, serum, urine, and cerebrospinal fluid (Beckonert et al., 2007, Wang et al., 2016) and DBS samples that lastly gained popularity for clinical metabolomics applications (Drolet et al., 2017, Jung et al., 2013). However, it is essential to acknowledge that variations in health status may limit the accuracy of these biofluids in reflecting the metabolic state; consequently, although standardized procedures exist for metabolomics analysis, additional standardization is required (Hernandes et al., 2017, Wu and Li, 2016). Biofluids are favorable and preferred due to their ease of collection (Beckonert et al., 2007) and non-invasive nature (Holmes et al., 2008b) . However, less frequently used samples, such as cerebrospinal fluid and tissue samples, hold significant potential for metabolomics research. Tissue samples, in particular, can provide organ-specific metabolic profiling (Lin et al., 2007), allowing for investigations into the origins of metabolites.

1.3.3.2 Sample collection and processing

The collection and handling of samples during experiments are critical factors in guaranteeing reproducibility, particularly regarding the metabolite composition. Consistent sample handling and immediate storage are essential for minimizing variability in subsequent analyses and maintaining data quality. Storing samples at extremely low temperatures, such as −80°C or liquid nitrogen, is recommended to prevent metabolite degradation. Other factors that should be considered when collecting biofluids include the type of syringe for blood collection, anticoagulant, temperature, duration of centrifugation, and speed (Lehmann, 2021, Yin et al., 2015). In the case of tissue samples, various methods are available for harvesting and processing them, such as lyophilization, homogenization, and tissue cell lysis. However, the sample stability is sensitive to other factors, including the freezing and washing methods, storage conditions, and time storage (Dapic et al., 2019). Human metabolomics studies typically analyze biofluids or tissue samples, making biofluids more accessible for collection and analysis. It is essential to consider the choice of the biological sample and to ensure that the collection and handling procedures are consistent and adequately documented to obtain reproducible and reliable results (Yin et al., 2013, Kirwan et al., 2018, Kamlage et al., 2014, Hirayama et al., 2015).

1.3.4 Metabolomics open-source software and statistical analysis

The massive volume of data generated in metabolomics presents a significant challenge that requires sophisticated data analysis tools. Regardless of the technology used or the approach taken, capturing the entire spectrum of metabolites in a sample is impossible, leading to a tremendous amount of generated data that cannot be evaluated (Lu et al., 2008). Thus, developing and applying effective computational methods is crucial extracting the relevant information for biological interpretation from this vast amount of data. So far, established software tools such as XCMS online (Huan et al., 2017), MZmine (Katajamaa et al., 2006), and OpenMS (Pfeuffer et al., 2017) have been efficiently employed in metabolomics studies for biomarker discovery. Additionally, publicly available databases such as METLIN (Tautenhahn et al., 2012), KEGG (Kanehisa, 2016), HMDB (Wishart, 2007, Wishart et al., 2022), LIPID MAPS (Sud et al., 2007), ChemSpider (Little et al., 2012) and PubChem (Kim et al., 2021) have proven valuable resources in assisting in metabolite identification and data integration.

Metabolomics data analysis includes multiple critical steps and necessitates significant preprocessing of the raw data, followed by specialized mathematical, statistical, and bioinformatics tools for conducting multivariate statistical analysis. At an advanced stage, it also involves data mining, integration with other omics data, and mathematical network modeling (Shulaev, 2006). Translating metabolomic data into practical applications for clinical use poses a significant challenge, starting with acquiring data, preprocessing, and subsequent application into clinical trials. The raw data must undergo preprocessing, which involves various steps such as quality processing, normalization, or rescaling. In metabolomics, preprocessing of raw data entails noise subtraction, chromatographic peak alignment, data normalization, and scaling before statistical analysis. This preprocessing can be done using vendor-specific or open-source tools (Sugimoto et al., 2012).

The data obtained through mass spectrometry (MS) appears in the form of spectra, comprising a complex arrangement of peaks representing various metabolites. These peaks, alone or in combination with others, represent distinct compounds (Stricker et al., 2021). Although the peaks belonging to a specific molecule are generally located at the same position on the spectra, various factors may cause slight shifts (Beckonert et al., 2007, Lenz and Wilson, 2007). To minimize the risk of misidentifying metabolites, this issue can be addressed by integrating the spectra over small chemical shift windows (Keun et al., 2004) or applying peak alignment algorithms before analysis to reduce the likelihood of incorrect metabolite characterization. Peak alignment and deconvolution techniques are typically utilized for preprocessing MS datasets to account for variances caused by temperature, column fluctuations, instrument settings, and other sources of variability (Torgrip et al., 2003).

Biomarker identification is crucial for advancing our comprehension of biological systems, yet detecting unknown metabolites presents a significant challenge in untargeted metabolomics studies. Addressing this challenge requires extensive LC-MS-based metabolomic studies and the establishment of comprehensive databases, which hold immense potential for decreasing the number of unknown metabolites identified in untargeted metabolomics, thereby providing a promising approach for identifying previously unknown metabolites and furthering our knowledge of biomarkers and biological systems (Ma et al., 2006). Numerous software programs have been developed to facilitate the biomarker identification process, including free solutions such as

MetaboAnalyst (Xia et al., 2009), which offers valuable data preprocessing, metabolite identification, and various analysis methods such as PCA and PLS-DA for effectively identifying biomarkers in metabolomic data. In conclusion, the availability of efficient and user-friendly software programs is crucial for biomarker identification, and free programs provide a cost-effective solution for researchers.

1.3.5 Analytical challenges

The field of metabolomics has made significant progress, yet several challenges remain to be overcome. The foremost among these is the need for more annotated metabolites since, to date, a small fraction of the detected spectra have been identified and quantified, leaving many metabolites unannotated (Cui et al., 2018, Muthubharathi et al., 2021). To address this issue, the Human Metabolome Database (HMDB) (Wishart et al., 2022) and METLIN (Xue et al., 2020) databases are frequently employed in metabolomics research, and attempts are being made to expand these databases continuously. However, determining the specific pathways affected by metabolites is demanding, and ongoing efforts are being pursued to discover innovative approaches to these limitations (Cui et al., 2018).

Another obstacle encountered in untargeted metabolomics pertains to annotating and identifying metabolites, a persistent challenge despite the enhanced accessibility of mass spectral libraries (Vinaixa et al., 2016, Dunn et al., 2013). Although such libraries contain many small molecules, most are not derived from experimental data using pure standards. Moreover, even when standards are obtainable, they encompass only 40% of the compounds detected in the human genome-scale metabolic network (Frainay et al., 2018). As such, efforts to expand annotated metabolite databases and develop innovative methods for identifying metabolites continue to be key research areas in metabolomics.

1.4 GENOMICS FOR RARE DISEASE DIAGNOSIS

The application of genomic sequencing has reshaped conventional diagnostic methods, allowing for fast, precise, and cost-effective genetic screening. This transformative technology has particularly enhanced the detection of rare diseases, reshaping conventional diagnostic approaches (Stark and Scott, 2023). Furthermore, the integration of sequencing methods, such as whole-exome sequencing (WES) and whole-genome sequencing (WGS), into research and clinical areas has

considerably enhanced the diagnosis of rare diseases. This progress has provided a more thorough understanding of the molecular mechanisms underlying diseases, thereby driving significant advancements in treatment strategies (Tambuyzer et al., 2020).

Nevertheless, there is a pressing need for innovation because the detection rate of uncommon genetic disorders using genomic testing is much below 50%. Yet, improved sequencing techniques and the incorporation of different omics data are some examples of recent advancements in screening technology that may lead to increased accuracy. This advancement benefits patients by enabling earlier and more accurate diagnostics and furthering our understanding of genetic diseases (Smedley et al., 2021, Marwaha et al., 2022). For rare hereditary diseases, understanding the causative variant and mode of inheritance provides patients with crucial insights into the risk of transmitting the condition to offspring and facilitates facilitating informed decisions regarding family planning alternatives (Mitsuhashi and Matsumoto, 2020).

Despite the benefits of genomic sequencing, diagnostic delays for rare diseases can be extensive, often spanning over a decade (Molster et al., 2016, Heuyer et al., 2017), influenced by multifaceted factors encompassing patient phenotype, age, and financial resources. Consequently, patients often remain undiagnosed or even misdiagnosed, amplifying emotional distress. Challenges in identifying causal variants using exome or genome sequencing are often addressed through Next-Generation Sequencing (NGS) technologies. NGS-based genetic testing shows significant potential as the most reliable approach for detecting rare diseases and delivering the fastest diagnosis outcome (Liu et al., 2019, Clark et al., 2018). However, limited real-world evidence supporting NGS-based genetic testing poses challenges as its lack of evidence may impede physician decision-making and reduce patients' willingness to undergo testing (Liu et al., 2019). Overall, genomics is leading a significant change in diagnosing rare diseases, transitioning its focus from analyzing symptoms to assessing molecular causes. This molecular diagnostic approach forms the cornerstone of precision medicine offering, personalized treatment plans based on solid, evidence-based results.

1.5 BIOMARKER DISCOVERY

1.5.1 Biomarkers conceptualization

While the term "biomarker" was first introduced in 1973 to refer to the presence or absence of biological material (Rho et al., 1973), the concept can be traced back to 1949 (Mundkur, 1949) and 1957 (Porter, 1957), when it was known as a "biochemical marker" and a "biological marker," respectively. In time, our understanding of biomarkers has advanced, and they have become increasingly valuable for diagnosing and monitoring disease, guiding treatment decisions, and predicting patient outcomes (Aronson and Ferner, 2017). However, despite their potential, the reliability of the biomarkers has been hampered by limitations, inaccuracies, and gaps.

In 2001 (Group et al., 2001) and 2016 (FitzGerald, 2016), different teams aimed to define biomarkers in such a way that would confer sustained usefulness. Regardless of their best efforts, these definitions failed to capture biomarkers' true nature and potential fully. Subsequent investigations have shown that these definitions needed to be revised and aligned with the essential principles of biomarkers (Califf and Medicine, 2018). A group of experts collaborated to create a more resilient, simple, accurate, and adaptable definition, defining biomarkers as ''a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention" (Group, 2016). This definition is intended to withstand the test of time and accommodate future scientific advances.

Biomarkers have emerged as essential components in clinical trials, providing significant insights into the underlying pathophysiology of various neurodegenerative disorders (Koníčková et al., 2022). Specifically, the use of biomarkers within the framework of clinical trials involves measuring and assessing the biological samples before, during, and after treatment, enabling the monitoring of the efficacy of the therapeutic interventions and the refinement of the existing therapeutic strategies. Hence, incorporating biomarkers in clinical trials presents a promising approach to enhancing our understanding of the common neurodegenerative disease as well as the rare NCL group and improving clinical outcomes (Nickel and Schulz, 2022, Ganesalingam and Bowser, 2010, Hersrud et al., 2016),

In conclusion, biomarkers provide objective measurements and indicators of disease and offer valuable insight for identifying potential therapeutic targets, monitoring disease progression, and evaluating treatment efficacy (Ehrenberg et al., 2020). In clinical trials, biomarkers can aid in

developing personalized treatment plans, resulting in better patient outcome (Qiu et al., 2023, Fountzilas et al., 2022). However, it is essential to determine the differences in the sensitivity and specificity of biomarkers to ensure the appropriate application of biomarkers in clinical and research settings. Only then can biomarkers hold promises for advancing disease diagnosis, monitoring, and treatment (Rachakonda et al., 2004).

1.5.2 The quest for biomarkers

Biomarkers are measurable entities that provide valuable information about various aspects of a disease (Group et al., 2001, Mendrick and Tong, 2013). Their identification is possible at multiple disease course stages, from pathogenesis, initial clinical signs, diagnosis, and therapy to recovery (Jain, 2010, Khan et al., 2020). As a result, biomarkers may serve multiple purposes or possess unique features that suit a specific application (Cagney et al., 2018). With the emergence of precision medicine, it has become increasingly evident that diagnostic biomarkers serve multiple functions beyond their traditional applications (Hartl et al., 2021). One of these functions includes aiding in the stratification of patients (risk assessment) within a disease, which is particularly important given the diverse disease subtypes associated with a disorder (Trusheim et al., 2007). Such an example is neuronal ceroid lipofuscinosis (NCLs), which has distinct subtypes with various onset, clinical features, and therapeutic responses, rendering diagnostic biomarkers critical in enhancing the development of future therapies (Torres Moral et al., 2022, Hersrud et al., 2016, Schulz and Kohlschütter, 2013).

On the other hand, monitoring biomarkers are a type of biomarkers that are assessed at regular intervals to monitor disease progression and treatment efficacy (Group, 2016). It provides valuable insight into the emergence of novel disease symptoms, worsening of pre-existing symptoms, and variations in disease severity or clinical manifestations (Verberk et al., 2021, Buchhave et al., 2012). However, the conventional "one-size-fits-all" approach in clinical trials assumes a uniform response in the population studies, which may only benefit a subset of patients. Thus, biomarker-based clinical research can uncover response differences among patients. Nevertheless, finding an ideal biomarker effective for everyone in real-life situations is rare (Gromova et al., 2020, Frangogiannis, 2012).

Consequently, caution should be exercised when evaluating clinical trials based on this approach (Behl et al., 2022). Personalized medicine and discovering effective biomarkers for all patients require an extensive and complex process. Nonetheless, the benefits of these endeavors are substantial, benefiting many patients and guiding the development of additional studies to identify more biomarkers and treatments. The widely accepted paradigm in biomarker research recognizes the crucial role of practicality and cost-effectiveness in enabling routine clinical use of biomarkers (Davis et al., 2009, Sturgeon et al., 2010). This paradigm emphasizes the need for biomarkers to be easily quantifiable, readily available in samples, and cost-effective (Lescuyer et al., 2007). Access to brain samples is critical to investigating neurological disorders (Shepherd et al., 2019, Blennow and Zetterberg, 2018). However, the inability to monitor disease progression in living patients hinders systematic biochemical studies using these samples (Bujak et al., 2015, Farahany et al., 2018). As a result, biofluid samples such as blood, serum, and urine have been employed as an alternative, although they often fail to provide a complete representation of the brain's condition (Attard et al., 2019, Bujak et al., 2015, Huan et al., 2017). Alternatively, cell samples represent substitutes (Son et al., 2017). However, their use presents numerous challenges, including high costs, time consumption, and limited availability of patient samples, resulting in various compromises that complicate the biomarker discovery process yet are necessary for advancing research (Hosoya and Czysz, 2016).

1.5.3 Metabolomics for biomarker discovery

The need to identify biomarkers for rare and neurological disorders is growing, driving research in this field. Metabolomics technology is significant in biomarker discovery and analyzing small-molecule (metabolites) in human biofluid and cell line samples. However, the human metabolome is subject to rapid changes caused by various metrics such as genetics, lifestyle, environment, pathology, and development (Monteiro et al., 2012). Compared to genomics and proteomics approaches, metabolomics offers several advantages for biomarker discovery, including capturing a more comprehensive picture of cellular functions and physiological conditions, being more dynamic and time-sensitive, and identifying biomarkers that reflect complex interactions between proteins, signaling pathways, and cellular settings (Quinones and Kaddurah-Daouk, 2009, Martins-de-Souza, 2022). This approach presents new possibilities for personalized medicine, particularly diagnosing and treating rare and neurological disorders.

Numerous metabolomics-based studies have aimed to identify potential biomarkers for neurodegenerative diseases (Lanznaster et al., 2022). However, the practical usefulness and efficacy of these biomarkers have yet to be demonstrated in clinical applications (Parker et al., 2018). In particular, the translation of biomarker candidates from laboratory research to clinical practice, especially in rare diseases, has been limited, which underscores the pressing need for an effective methodology to measure biomarkers routinely (Seyhan, 2019, Perlis, 2011). Despite sophisticated and sensitive methods, multiple factors can affect the results, and no single factor alone can satisfy the requirements for developing a reliable and robust biomarker. These factors include small sample sizes, sample complexity, and human metabolism's genetic and environmental variability (Bax, 2021). Therefore, specific criteria must be met to develop biomarkers successfully that can be used in future research studies and implemented on a large scale (Califf and Medicine, 2018).

Over the last three decades, extensive research has been conducted on rare neuronal ceroid lipofuscinosis (NCL) disorders, with some focusing mainly on identifying potential biomarkers. Various omics-based approaches have been employed to achieve this goal. For instance, the blood-plasma-based proteomics approach to juvenile and late infantile NCL patients identified seven candidate biomarkers with potential therapeutic applications. Interestingly, some putative biomarkers, such as brain-derived neurotrophic factor (BDNF), neuronal cell adhesion molecule, adiponectin, and Apolipoprotein E (ApoE), due to their neuroprotective and lipid metabolism functions have already been associated with or proposed as potential biomarkers for other common neurodegenerative disorders. Among the identified biomarkers, myoglobin and vascular cell adhesion protein 1 (VCAM-1) were significantly elevated in NCL, with VCAM-1 being the only biomarker previously appointed as a potential biomarker in NCL (Hersrud et al., 2016).

The most recent study aimed at biomarker discovery was conducted in 2022 and highlighted the potential development of a quantitative marker to measure disease progression in NCL-affected patients. This biomarker study involved magnetic resonance imaging (MRI) analysis of 35 patients diagnosed with CLN3 disease. It was found that the supratentorial cortical grey matter (GM) volume exhibited a significant and consistent decline with age and was closely linked to the disease-specific clinical scoring. These findings suggest that supratentorial cortical GM volume can be a potential marker for monitoring disease progression (Hochstein et al., 2022). In 2020, a

study conducted by Kline *et al*. reviewed the potential of Omics technologies to understand NCL disease mechanisms and identify specific biomarkers. The authors examined previous applications of Omics technologies to various forms of NCL and identified common pathways affected by these NCL disorders. The study concludes that using more advanced study models and well-defined patient populations combined with *omics* techniques may identify clinically relevant biomarkers for disease diagnosis, progression tracking, and treatment response (Kline et al., 2020).

Even with extensive attempts, identifying clinical applicable biomarkers that pass the validation stage remains challenging. This is primarily due to the challenges associated with developing robust biomarker development protocols and the heterogeneity of studies. In addition, only a few metabolomics biomarkers for neurodegenerative diseases have met the regulatory standards for clinical use. These limitations arise from the limited sample size for rare disorders and the insufficient sensitivity of current mass spectroscopy technology, which hinder the identification of reliable metabolomics biomarkers. To overcome these obstacles, advanced AI algorithms (Filipp, 2019, Schneider et al., 2020, Mamoshina et al., 2016) and chromatographic separation methods may be necessary to mitigate potential interferences in human samples. For example, patients with different subtypes of NCL disorder may require specific biomarkers, and advanced technology can help identify these biomarkers and facilitate their translation into clinical use (Mayeux, 2004).

One study done in 2018 used a metabolomics-based approach to investigate potential biomarkers for CLN2 disease and the efficacy of untargeted metabolite profiling of cerebrospinal fluid (CSF). The study used LC-MS-based untargeted metabolite profiling and identified 29 metabolite features that showed a linear correlation with disease severity. Among the eight confirmed metabolites identified, seven acetylated species had decreased levels in CLN2 patients. These results may indicate a deficiency in acetylated species in CSF that causes cerebral energy deficiency and neurodegeneration. The study suggests that targeted analysis of these metabolites in CSF may help monitor CLN2 disease progression and evaluate the effectiveness of current and future therapies (Sindelar et al., 2018). The thorough evaluation of the sensitivity and specificity of potential biomarkers for rare diseases holds paramount significance (Lescuyer et al., 2007). The limited reproducibility can be addressed by utilizing modern workflows that enable high-throughput analyses with large cohorts of samples that have undergone comprehensive characterization and stratification. This approach provides a promising avenue to enhance previous methodologies (Comes et al., 2018).

A metabolomics-based approach utilizing blood, plasma, or serum samples is highly desirable for profiling biomarkers in neurodegenerative disorders. These biological samples are commonly used in clinical practice for routine diagnostic analyses, making their collection much more accessible. While the cerebrospinal fluid (CSF) is particularly interesting for identifying potential biomarkers due to its proximity to the brain, collecting it is a complex and invasive procedure. There are visible signs and symptoms of neurodegeneration in some cases, but there is no cost-effective method to identify them on time. Thus, biomarker studies based on accessible samples could provide a means of identifying early changes in neuronal degradation (Bujak et al., 2015).

In conclusion, evaluating the sensitivity, specificity, and predictive values of potential biomarkers for the disease of interest is crucial. Modern workflows that enable high-throughput studies with large cohorts of well-defined samples represent an opportunity to improve the reproducibility of past studies. As mentioned earlier, specific compounds, have been associated with certain NCL diseases. Nevertheless, none of these have been proven to be a robust and reproducible biomarker. Although identifying metabolomics biomarkers for NCL and neurodegenerative diseases is an ongoing process, three crucial criteria must be met to make progress in this area: 1) accurate selection and stratification of patients in the sample cohort, 2) shortened sampling time and suitable storage, and 3) standardizing procedures for sample processing targeted to the disease. These criteria are particularly important because of the broad spectrum of phenotypes in patients with the same NCL disorder. The overlap of these traits in different CLN diseases often makes diagnosis difficult and increases the heterogeneity of the clinical population being studied.

The current state of biomarker discovery has been affected by differences in sample processing procedures and the varying sensitivity and power of different MS technologies. Identifying reliable biomarkers could be possible by developing new combined technologies that have improved quantity, power, and sensitivity. Nonetheless, integrating the extensive data generated by MS platforms remains challenging (Misra et al., 2019). To address this issue, bioinformatics tools have to be created to aid in our comprehension of the potential of this technology (Liang and Kelemen, 2018). Until then, biomarker validation has remained a lengthy and complex process involving method and clinical validation to establish a direct connection between the biomarker and the targeted disease or clinical endpoint.

1.5.4 Biomarker discovery study design

The application of available technologies is instrumental for detecting variations in protein and metabolite levels in biological samples with clinical relevance for disease diagnosis and prognosis. Nevertheless, a biomarker must undergo extensive validation procedures that evaluate both the analytical methodology and the putative biomarker to be deemed suitable for clinical applications. This rigorous process ensures that only reliable and robust biomarkers are approved for clinical use. Only after the successful completion of the validation process can the discovery of a viable biomarker be confirmed (Issaq and Veenstra, 2020).

The ideal characteristics of a biomarker include non-invasive sample collection, such as biofluids, and a technology that requires minimal sample preparation and is cost-effective. It should also be easy to use, fast, reliable, and have user-friendly software and tools for data analysis and interpretation. In addition, the biomarker should be readily available, sensitive, accurate, and reproducible while remaining affordable (**Figure 2**) (Bodaghi et al., 2023). Regardless, a biomarker's most desired features are perfect accuracy, sensitivity, and specificity (Califf and Medicine, 2018). However, it is essential to note that this is an ideal scenario.

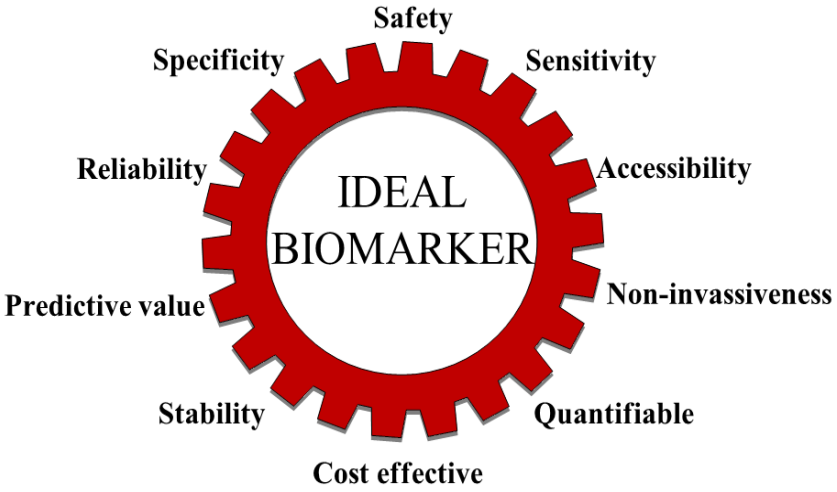


Figure 2. *General characteristics of an ideal biomarker. Adapted from Bodaghi et al. (2023).*

1.5.4.1 Study design and execution

Ensuring a well-designed biomarker discovery study depends on careful consideration of certain critical factors. These factors include, but are not limited to, the biological characteristics of the

targeted disease, the accessibility of patient samples, the appropriate selection of controls, the stratification of patients based on age, gender, clinical and genetic features, the type, quantity, and suitability of samples being analyzed, and the molecular nature of the compounds under investigation). Additionally, it is essential to ensure that a sufficient sample size is available and easily accessible for analysis (McDermott et al., 2013, Pepe et al., 2015).

Therefore, designing a study for biomarker development depends highly on several other laboratory variables. Primary elements that must be incorporated into the study design as they will impact the accuracy and reliability of the results include sample collection, which entails a range of factors such as the method and timing of collection and the tools employed. Storage conditions (room temperature, refrigeration, freezing, or liquid nitrogen), storage time (short-term, long-term, or no storage), sample type and preparation (fresh samples, frozen samples, extraction method, and pre-post processing), the analytical method, number of replicates, and data analysis approach must all be meticulously evaluated. Integrating these parameters into the study is of utmost importance as they serve as the foundational framework for initiating and guiding the proper execution of the study (Metcalf and Orloff, 2004, Zheng, 2018).

Despite challenges in achieving the mandatory sensitivity and specificity, searching for the ‘'easy-to-use and fit-for-purpose metabolomic biomarkers remains a top priority if aiming for early disease diagnosis and improved patient outcomes. Despite the observed slow progress, which can be attributed to small sample sizes, unstable standard operating procedures, limited coverage, and several other limitations, the current successful biomarker studies have shown promise in identifying biomarkers for other diseases (Minno et al., 2022). The potential for discovering qualitative biomarkers, particularly for rare and neurodegenerative disorders, is high if there are advancements in technology and analytical methods (Ehrenberg et al., 2020) since incorporating multiple approaches can speed up the process, enhance accuracy and ensure the long-term stability of biomarkers (Xiao et al., 2022).

Dried blood spots (DBS) for biomarker metabolomics studies

Applying dried blood spots (DBS) for biomarker discovery in metabolomics is a long process. Despite DBS's advantages over other sample types in terms of sampling and storage, apparent drawbacks are associated with its use in metabolomics studies. One of the initial challenges is related to the pre-analytical steps, namely the quality and consistency of the collected blood spots.

Variables like hematocrit, dried blood spot size, and sample homogeneity can significantly impact the accuracy and reproducibility of metabolomic measurements. Therefore, these variables must be uniform across different samples to ensure accurate results. Another issue concerns the analytical techniques employed for DBS-based metabolomics research. To mitigate the interference caused by the filter paper and the presence of contaminants in the blood spots, specific techniques are required for extracting and analyzing metabolites from DBS.

Consequently, there is continuing research into developing robust and effective analysis techniques for DBS metabolomics studies. Despite the advancement in metabolomics research employing DBS technology, the practical application of these findings in clinical practice remains limited. Although DBS has the potential to be used in real-world clinical settings for various diseases and clinical tasks, such as disease diagnosis, treatment effectiveness monitoring, and monitoring therapeutic medication levels, its application is still limited, and further research and validation are necessary to assess its practical value in clinical settings and address existing challenges. This thesis addresses the challenges of using DBS in biomarker discovery studies, particularly in DBS metabolomics studies. Several factors that can influence the reliability and accuracy of metabolomic analysis, including card age, storage conditions, and analyte extraction solvents, were comprehensively evaluated using samples from patients affected by several lysosomal storage disorders.

2 SCOPE OF THE THESIS

This study aimed to identify specific biomarkers and the disease-related pathways associated with CLN6 disease using LC-MS-based metabolomics techniques. To achieve this, a comprehensive methodology was developed that enabled the identification of potential biomarkers with a role in the early diagnosis of CLN6 and possible monitoring of future clinical trials. The findings of this study provide insights into the CLN6 disease pathophysiology, having significant clinical implications that could accelerate the development of new therapeutic targets. The three studies are aimed at achieving the following objectives:

• The **first study** focusing on ***Dried Blood Spot (DBS) Methodology Study for Biomarker Discovery in Lysosomal Storage Disease (LSD)*** aimed to evaluate the suitability of using dried blood spot (DBS) card samples for metabolomics-based biomarker discovery specifically targeting

CLN6 disease. It compared the efficacy of EDTA-blood and filter cards in metabolomics studies to identify the most suitable sample type for disease-specific compound detection. Additionally, it sought to optimize metabolite extraction to address the limitations of small sample sizes typical in rare disease research. This involved evaluating various solvents for their effectiveness in metabolite recovery and assessing the impact of storage conditions on metabolite stability. The study employed DBS-based metabolomics, a promising methodology for diagnosing and monitoring rare diseases.

The research addressed the challenges associated with using blood samples for biomarker discovery, aiming to advance biomarker development techniques and improve the reliability of compound detection. It employed three pivotal experimental strategies: Firstly, it involved a comparative analysis of EDTA-blood versus filter card samples in metabolomics research to identify the sample types most effective for detecting disease-specific compounds. Secondly, it focused on determining the optimal storage conditions for these samples, a critical factor, considering the challenges of small sample sizes in rare disease studies and the necessity of reanalyzing the same samples due to the difficulties in obtaining new samples from individuals with rare diseases. Lastly, it aimed to enhance metabolite extraction methods, significantly increasing the number of detected compounds.

In summary, by carefully designing experiments to improve the detection and reliability of compounds, the study aimed to establish a universally applicable testing framework, thereby enhancing the process of discovering biomarkers.

- Recognizing the importance of careful sample selection and patient stratification, the **second study** focusing on ***Clinical and genetic characterization of a cohort of 97 CLN6 patients tested at a single center*** – aimed to comprehensively analyze this cohort, representing the largest CLN6 group screened to date in a single center. Given the wide variability in symptoms and variants among rare diseases like CLN6, meticulous patient selection is crucial to assemble a homogeneous group. To achieve this, the study focused on conducting a thorough analysis of CLN6 genetic variants and investigating these patients clinical and genetic profiles.

Through comprehensive analyses of demographic, phenotypic, and genotypic data from CLN6 patients, the study sought to uncover correlations between genetic variants and various factors, including gender, ethnicity, and clinical presentations. This approach provided deeper insights into

the intricate nature of the disease, revealing genetic heterogeneity within the CLN6 population and establishing links between specific variants and aspects of patient demographics, clinical symptoms, and disease progression. The study underscores the critical role of assembling a well-defined cohort of CLN6 sample, laying the groundwork for identifying potential compounds specific to CLN6 disease.

• Furthermore, to refine the precision and applicability of biomarker discovery in CLN6 disease studies and to mitigate the impact of sample or patient variability on data accuracy, a robust methodology was implemented. This led to the **third study** ***"Neuronal progenitor cells-based metabolomics study reveals dysregulated lipid metabolism and identifies putative biomarkers for CLN6 disease"***. This study aimed to utilize untargeted LC-MS-based metabolomics to investigate the metabolomic profile of human CLN6 neuronal progenitor-like cells, aiming to gain insights into CLN6 disease metabolism. Despite the investigation being confined to only three available CLN6 cell lines - representing the entirety accessible in the market at that time - the study's efficacy remained notable. The lack of iPSC availability posed additional challenges, as their potential use could have significantly expedited the research process. This limitation necessitated modifying the research approach by directly differentiating these cell lines into specific neuronal cells through chemical conversion. This ensured consistent neuronal differentiation across samples, facilitating the metabolomics approach for identifying potential biomarkers for CLN6. Notably, the study revealed significant downregulation of several sphingolipids and glycerophospholipids CLN6-affected cells compared to controls, a finding consistent with patterns observed in other neurodegenerative disorders. These compounds serve as a foundation for further focused investigations and future studies on their potential as candidate biomarkers for CLN6 disease, thereby establishing a new framework for neuronal cell-based metabolomics studies.

3 MATERIALS AND METHODS

This section provides a brief overview of the techniques and analytical approaches utilized in each study. However, referring to the original publication is recommended for a comprehensive and detailed presentation of the methodology employed.

3.1.1 Study design and human samples

The **first study** involved a combined cohort of 66 subjects from diverse geographical backgrounds categorized into four distinct groups based on their genetic features. Group one comprised 16

subjects with CLN6 disease, while group two served as the control group with 27 individuals. Group three constituted a mixed group of 26 NCL subjects, including 10 subjects with eight different types of CLN disease and 16 CLN6 subjects. The fourth group comprised 16 subjects with 11 types of various lysosomal storage disorders (LSD). Each group underwent various experimental design settings, including a stability study, an extraction optimization study, and a cumulative study based on all the previous experimental data. The study was based on blood samples from disease-affected patients sent by physicians to Centogene GmbH in Rostock, Germany, for routine genetic diagnosis. The screening involved analyzing the patients' genetic material to identify potential genetic mutations or variations.

The **second study was** a retrospective cross-sectional study of 97 CLN6 patients screened over ten years at a single center. These genetically confirmed CLN6 sample, were sent by physicians to Centogene GmbH in Rostock, Germany, between 2010 and 2020 for genetic diagnosis. The samples provided included DNA, EDTA blood, dried blood spots, amniotic fluid, or saliva.

The **third study** utilized cell lines procured from a commercial source derived from patients diagnosed with CLN6 disease (n=3), who carried bi-allelic pathogenic variants, and control/wild-type cell lines (n=3) with no pathogenic genetic variants. Each cell line was cultured in triplicate, resulting in a total of 9 CLN6 and control cell lines, respectively.

3.1.2 Sample preparation and metabolite extraction

3.1.2.1 Dried blood spot card preparation

Before LC-MS analysis, the EDTA blood sample had to be transferred to a dried blood card format. For the preparation of dried blood spots (DBS), a 50 µL aliquot (EDTA) was dripped onto each spot of the CentoCard® and left to dry for at least two hours at room temperature (**Figure 3**). For analysis, five center punches with a diameter of 3.2 mm were taken from each card and collected into a 96-well microtitration plate using a PerkinElmer puncher.

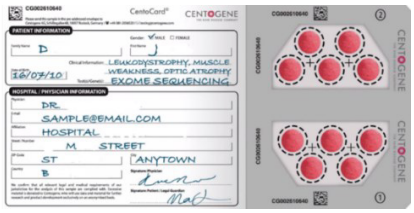


***Figure 3**. Standard CentoCard used to collect EDTA–treated blood. The blood is dripped onto the pre-printed circles, and the card is left to dry at room temperature for 2 hours.*

3.1.2.2 Metabolite extraction from DBS

The five punched blood spots were automatically transferred into a microtitration plate. To extract the analyte, 50 µL of extraction solution, consisting of a single solvent or mixture-based e.g., dimethyl sulphoxide, isopropanol, acetonitrile, methanol, ammonium acetate, and water in varying ratios based on the experimental design, along with 100 µL of internal standards solution containing 200 ng/mL lyso-Gb2, were added to each well of the microtitration plate. The plate was vortexed for 10 minutes, sonicated, and then incubated for 30 minutes at 37°C and 700 rpm. Following incubation, the liquid was transferred to an AcroPrep Filter Plate with a PTFE membrane, which was placed on a 96-well V-shape bottom plate and centrifuged for 5 minutes at 3500 rpm to remove any solid particles from the solution. This process effectively extracted the analytes from the blood spots for subsequent mass spectrometric analysis.

3.1.2.3 Cell culture

The CLN6 (CLF064, CLF121, CLF210) and wild-type (GM0839, GM0565, GM0203) fibroblast cell lines were grown in high glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin, and streptomycin until they reached 90% confluence. Next, these were directly converted into neuronal cells using a cocktail of six chemicals known to aid in NPC differentiation and then cultured in a neuronal medium consisting of a mixture of DMEM/F12 and neurobasal medium (Dai et al., 2015). The cells were cultured for 21 days with medium changes every third day. The cells were fixed, permeabilized, and blocked before being incubated with primary and secondary antibodies. Cell images were acquired using a Keyence fluorescence microscope BZ-X710E equipped with the BZ-X800 Analyzer software (Keyence, Osaka, Japan) with a 20X Plan-Apo Gamma NA 0.75 objective and fluorescence filter set for GFP, TRITC, and DAPI.

3.1.2.4 Metabolite extraction from cell lines

The cell culture plates were placed on dry ice, the medium was removed, and the cells were rinsed with cold 0.9% NaCl solution. A prechilled extraction solvent consisting of methanol and water was added to the cells, which were then detached using a scraper and transferred to a tube

containing an internal standard. After vortex and centrifuging, the resulting supernatant was transferred to a glass vial for LC-MS analysis (**Figure 4**). Quality control samples were generated by pooling a small amount from each sample, while blank samples comprised 100% LC-MS water. The standard, blank, and pooled samples were injected five times before the first sample injection to establish system equilibrium and ensure stability throughout the batch.

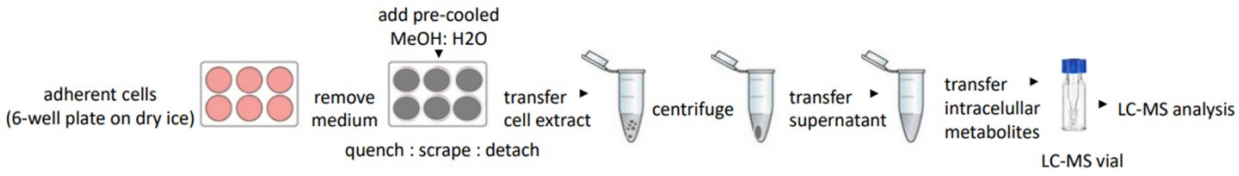


Figure 4. *Schematic overview of the cell metabolites extraction procedure.* ***Cells*** *were cultured in 6-well plates, scraped using methanol: water (3:1, v/v), quenched and washed with 0.9 % NaCl. The harvesting and scraping procedures were performed with the plates kept on dry ice.*

3.1.3 LC-MS analysis

Extract analysis was performed on a Waters® i-Class ACQUITY UPLC (Waters, Borehamwood, UK) coupled to a Vion™ IMS Q-Tof™ mass spectrometer (Waters, Borehamwood, UK) **(Figure 5)** equipped with an ESI ion source, system operating in negative (ESI−) and positive (ESI+) ionization modes. The LC-MS method used a C18 Kinetex EVO LC (Phenomenex, Aschaffenburg, Germany) column with a linear gradient of 1% to 100% B (50 mM formic acid in methanol: acetonitrile 1:1 *v/v*) and A (50 mM formic acid in water) at a flow rate of 0.5 mL/min. Mass spectrometry was acquired using High Definition MS^E ($HDMS^E$) with specific settings. Each signal was identified by three parameters: retention time in minutes (RT), ion mass (m/z), and collision cross-section (CCS) **(Figure 6).**

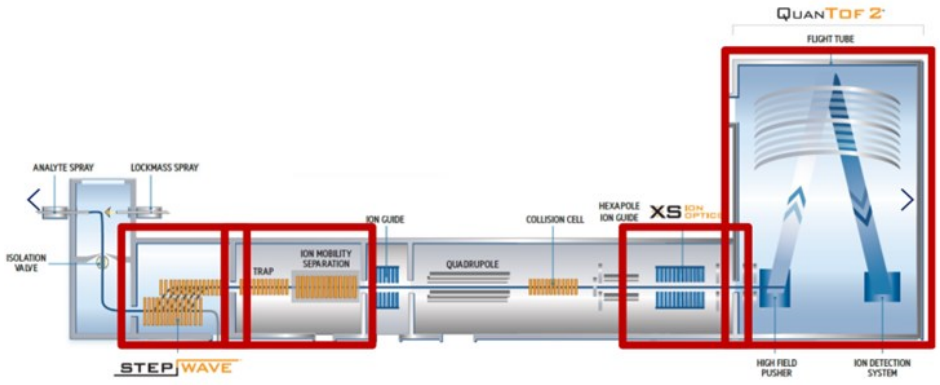


Figure 5. *Vion IMS QTof Mass Spectrometer depicting ion mobility separation device between the StepWave and a mass-resolving quadrupole (Note. Retrieved from https://www.waters.com, Waters Corporation (Wang, 2017)).*

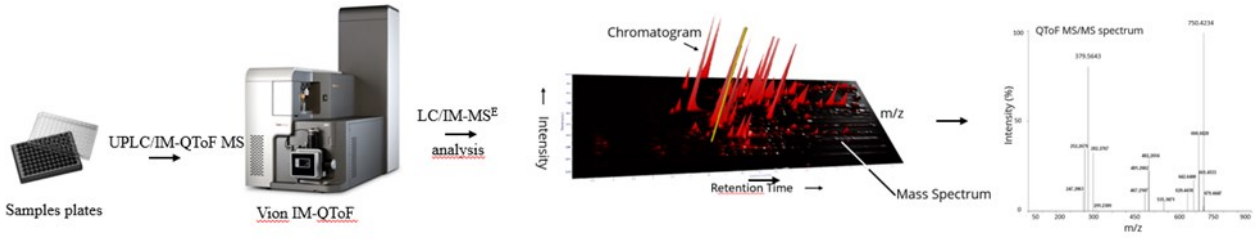


Figure 6. *From sample to spectrum: peak processing visualization for* HDMS^{E} *data and MS/MS spectrum of a random compound.*

3.1.3.1 Overview of LC-MS

LC-MS is a robust analytical technique that identifies target compounds based on their mass, collision cross section (CCS), and retention time (RT). Confined initially to specialized settings, LC-MS systems have transitioned from expensive and intricate research instruments into accessible and user-friendly detectors, reshaping analytical processes across scientific fields. Fundamentally, LC-MS integrates two powerful analytical techniques: liquid chromatography (LC) and mass spectrometry (MS). Liquid chromatography (LC) serves as the initial step where components within a sample undergo separation using various modes such as partition chromatography, ion-exchange chromatography, size-exclusion chromatography, or affinity chromatography. This separation process relies on the interaction between the analytes and the column's stationary phase. The retention time (RT) indicates their traversal duration. Retention time in chromatography denotes the duration required for a compound or analyte to traverse the column and reach the detector. This metric is essential in identifying and characterizing compounds within a given sample. Through comparison with established standards, the retention time facilitates precise determination of the compound's identity.

The separated components are then introduced into the mass spectrometer, where electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) is commonly used to generate charged ions from the sample. These ions are then analyzed using their mass-to-charge ratio (m/z)

to identify the desired analytes. Liquid chromatography-mass spectrometry (LC-MS) is now widely used in biochemistry due to the simplicity and reliability of electrospray ionization (ESI). As LC-MS instruments become more affordable and reliable, they are increasingly important in clinical biochemistry, challenging traditional techniques like liquid chromatography and immunoassay. Liquid chromatography involves injecting a precise volume of the sample solution into a continuous flow of solvent, known as the mobile phase. As the sample solution-mobile phase mixture traverses the column packed with a stationary phase, analytes interact differentially based on their properties. Partition chromatography separates analytes based on their distribution between the stationary and mobile liquid phases, while ion-exchange chromatography relies on ion charges. Size-exclusion chromatography separates analytes based on size, and affinity chromatography allows highly selective separation based on analytes' binding affinity to the stationary phase.

Following separation via liquid chromatography, the separated components are introduced into the mass spectrometer. Electrospray ionization (ESI) or atmospheric pressure chemical ionization produces charged ions from the sample. These ions are then analyzed based on their mass-to-charge ratio (m/z), enabling the identification of desired analytes with high sensitivity and accuracy.

The ion intensities detected during the LC-MS sample analysis are depicted as a total ion chromatogram (TIC). This visualization presents the intensity peaks of the analyte ions and their retention times (RT). Each data point on the chromatogram corresponds to a mass spectrum, which depicts the abundance of ions plotted versus their m/z (mass-to-charge ratio) values. The mass spectrometer operates in two modes: scan and selected ion monitoring (SIM). In scan mode, it detects ions across a range of m/z values within a set time, suitable for analyzing unknown samples. Specific m/z values are targeted in SIM mode, which is ideal for precisely quantifying known compounds.

By integrating two series-operating mass analyzers, enhanced sample identification and precise quantification are achieved. The predominant tandem mass spectrometers used are the triple quadrupole (QQQ) and quadrupole time-of-flight (QTOF) configurations.

In the QQQ setup, two quadrupole mass analyzers (Q1 and Q3) are interposed by a collision cell (q/Q2). Q1 and Q3 are mass analyzers for scanning a mass range or monitoring specific m/z values. Meanwhile, the collision cell facilitates the fragmentation of precursor ions isolated in Q1 through high-energy collisions with a neutral gas.

3.1.3.2 *Chromatography columns*

With different compositions and separation methods, chromatography columns are designed to meet different analytical demands and requirements for separation. Typical varieties include some of the following:

Reversed-phase columns are widely used in high-performance liquid chromatography (HPLC) to separate non-polar or slightly polar compounds. They are equipped with a non-polar stationary phase, which is usually hydrophobic.

In the normal phase column, the stationary phase is more polar than the mobile phase. Silica, a polar material, is commonly used as the stationary phase. The separation of sample components occurs based on their polarity, with more polar components interacting more strongly with the stationary phase and less polar components interacting more strongly with the mobile phase. This results in the separation of components based on their polarity.

Size Exclusion Columns utilize the stationary phase of porous beads, typically composed of polymers such as polysaccharides or silica. These beads feature varying-sized pores, allowing smaller molecules to penetrate and thus experience slower movement through the column. In contrast, larger molecules cannot access the pores and consequently elute earlier. This separation is based on the differential permeability of molecules through the porous stationary phase, enabling the separation of molecules based on their size.

Ion exchange columns consist of a charged stationary phase and a polar liquid mobile phase, such as a salt solution in water. With this configuration, molecules can be separated based on their affinity for the charged groups on the stationary phase. The method facilitates the exchange of ions between sample components and the stationary phase, leading to separation based on their ionic properties. Readily ionizable compounds are often analyzed using this method.

3.1.4 Metabolomic data processing

The raw MS data were acquired using Unifi software v1.9 (Waters, Borehamwood, UK) and exported as Unify export packages (.uep). Progenesis QI software v3 (Nonlinear Dynamics,

Newcastle upon Tyne, UK) was used for automated data processing, which included retention time correction, experimental design setup, peak picking, probabilistic quotient normalization (PQN) (Dieterle et al., 2006), deconvolution, and compound identification. The statistical relevance and robustness of the metabolites were evaluated based on predefined quality criteria, such as fold change of at least 2-fold, charge ≤ 3, the mass-to-charge ratio (m/z) ≥ 179, and median normalized abundance of ≥ 350 counts relative to the reference compound in at least one of the cohorts. The metabolites' features with significant differences were identified using Student's t-tests.

3.1.5 Data analysis

The following software and tools were used to perform the data analysis and generate the figures: MetaboAnalyst (Pang et al., 2021) statistical analysis toolbox, which was utilized for study I (v4.0) and III (v5.0), available at http://www.metaboanalyst.ca. Additionally, the "stats" package in R (Team, 2021) was used for study 1 (v3.6.2) and study 2 (v4.0.4) and can be found at www.r-project.org. GraphPad Prism (v9.5.0) was used for study 3, and this software was provided by GraphPad Software, Inc., (San Diego, CA), and accessed at http://www.graphpad.com. Canonical pathway analysis (study 3) was performed using Ingenuity Pathway Analysis (IPA) software from QIAGEN (Ingenuity Systems, QIAGEN, Redwood City, CA, USA). Overall, the results obtained from these analyses were critical to the conclusions and insights presented in all three studies.

3.1.6 Metabolites identification

Metabolites were identified by matching the obtained features against different metabolite databases, which included the in-house compound library, Human Metabolome Database (HMDB) (Wishart et al., 2022), PubChem (Sayers et al., 2022), ChemSpider (Pence and Williams, 2010), and LIPID MAPS® Structure Database (LMSD) (Sud et al., 2007). The metabolites' identities were determined through accurate mass, retention time, and MS spectra.

3.1.7 Genetic analysis

All procedures were undertaken according to the German Gene Diagnostic Act (Gendiagnostikgesetz) and the General Data Protection Act (Bundesdatenschutzgesetz) provisions. Samples were processed at Centogene GmbH (Rostock, Germany) in a facility certified under the Clinical Laboratory Improvement Amendments of 1988 and accredited by the College of American Pathologists.

Various sample types were utilized, including extracted DNA, EDTA blood, and dried blood spots (DBS) on filter cards (CentoCard®), amniotic fluid, or saliva. DNA extraction was performed using a QIAsymphony instrument with recommended reagents and kits from Qiagen (Hilden, Germany). Variant screening procedures were performed as described previously (Trujillano et al., 2017). Sequencing methods varied depending on the referring physician's specifications, encompassing whole-exome sequencing (WES), gene panel sequencing (CentoMetabolic® or Ceroid lipofuscinosis panels), or sequencing for CLN6 alone. A custom double-stranded DNA capture bait pool was used for the gene panel. A custom double-stranded DNA capture bait pool was employed for gene panel sequencing, targeting coding regions, flanking intronic sequences, and known relevant variants beyond coding regions. Libraries were prepared with Illumina-compatible adaptors and sequenced on an Illumina platform (Illumina, San Diego, CA) to achieve adequate coverage depth. Sanger sequencing was utilized to complete missing fragments when necessary. For whole-exome sequencing, human consensus coding sequences were enriched from fragmented genomic DNA using the Nextera Rapid Capture Exome kit (Illumina)/SureSelect Human All Exon V6 (Agilent)/ TWIST Human Core Exome (Twist Bioscience). The generated libraries were sequenced on an Illumina platform to an average coverage depth of 70-100x. Any relevant variants detected by WES were validated by Sanger sequencing in both directions.

4 RESULTS AND DISCUSSIONS

4.1 Study 1: Dried Blood Spot (DBS) Methodology Study for Biomarker Discovery in Lysosomal Storage Disease (LSD)

Liquid Chromatography-Mass Spectrometry (LC-MS) is a powerful tool in metabolomics research that enables the detection and quantification of metabolites in a biological sample. However, LC-MS sensitivity can be affected by several factors: sample type, storage conditions, sample age, and patient stratification (**Figure 7**). This instability in the workflow can pose challenges in untargeted metabolomics analysis (Li et al., 2020, Trifonova et al., 2019). Therefore, identifying and understanding the biological and environmental factors that impact metabolites during analysis is crucial to obtaining reliable data upon analyzing the targeted disease’s metabolomics profile.

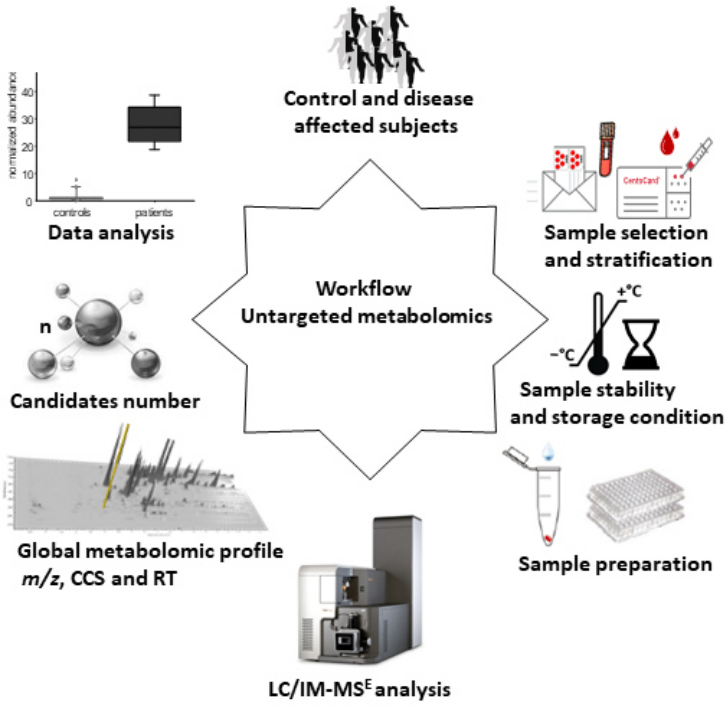


***Figure** 7. Standard workflow for DBS-based untargeted metabolomics. It entails study design, sample collection, preparation, and LC-MS analysis. Preprocessing of the raw data includes normalization, peak picking, filtering, and statistical analysis. Accurate mass and retention time are used for metabolite identification, and MS2 fragmentation spectra are used for structural elucidation. Adapted from Lee et al. (2010).*

The first study focused on identifying and managing the factors that impact LC-MS sensitivity, specifically when analyzing dried blood spot samples. The study comprised a series of experiments aimed at quantifying the impact of various triggers on metabolome variability. This helped identify critical factors that, when controlled, could enhance the biomarker discovery process. Recognizing and managing the factors influencing LC-MS sensitivity proved essential in obtaining reliable and precise results in metabolomics research. The findings from study 1 are further important in developing effective metabolomics strategies, ultimately leading to improved efficiency in biomarker discovery processes.

Thus, in the first study, the assessment of the variability of the human blood metabolome was based on two distinct approaches: (1) the stability of the metabolites was assessed by analyzing the impact of sample age (time since collection) and storage conditions; (2) the optimization of the extraction

solvent for metabolomics analysis was conducted to determine the solvent that generated the highest yield in metabolite extraction efficiency.

The stability study was further divided into two primary objectives: (a) to investigate the stability of filter card samples from preparation to analysis and (b) to compare the effectiveness of using old dripped DBS cards versus new dripped cards. For this, 36 blood samples were collected from six participants over a six-year period before analysis. The results showed high fluctuations in metabolite variations, implying that even if stored at -20 °C, the DBS cards' yearly stability cannot be guaranteed (**Figure 8a, b**).

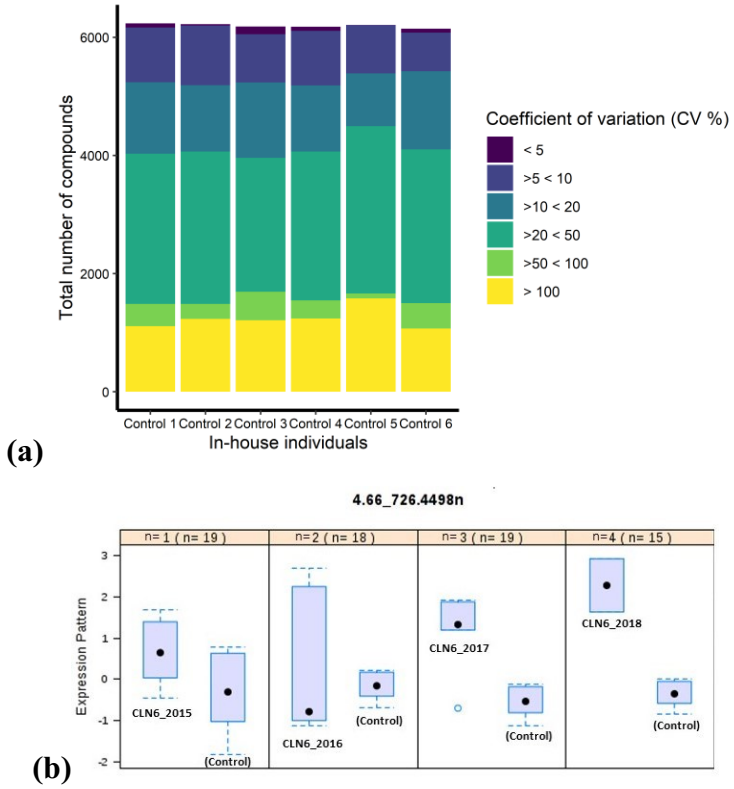


Figure 8. *The impact of long-term storage on metabolites stability. The effect of six years of storage time on the stability of the dried blood spot cards* ***(a).*** *Boxplots illustrate the variability in the abundance of a specific metabolite from 2015 to 2018* ***(b)****.*

After that, a follow-up experiment was carried out to investigate the impact of short-term storage on analyte stability in the DBS samples. For this, duplicate DBS cards were prepared using freshly collected blood and were then stored at various temperatures (RT, 8°C, -20°C) for three days before undergoing metabolomics analysis. The results showed that even the DBS short storage

significantly affects analyte stability, underscoring the importance of precise temperature control to guarantee accurate results. An additional experiment was conducted to evaluate the suitability of new dripped DBS cards stored at RT versus the new dripped DBS cards prepared the EDTA-whole blood stored at -20°C. The results showed fluctuations in metabolite measurements, underscoring the importance of preparing samples uniformly and storing them under consistent conditions to minimize variability. These findings have considerable implications for biomarker development studies, highlighting the need to adhere to established sample storage and handling protocols when conducting metabolomics research to ensure reliable and reproducible results.

The second objective aimed to optimize the extraction process for LC/MS-based metabolomics analysis and involved a comparison of four solvents based on the total number and abundance of metabolites screened. The results indicated that DMSO in water provided the best coverage of the metabolome. However, methanol-acetonitrile (3:1, *v/v*) yielded the highest number of high intensity metabolites. In contrast, ammonium acetate generated slightly more metabolites with an abundance of over one thousand than methanol-acetonitrile (3:1, *v/v*) but had lower global metabolome coverage. Isopropanol-acetonitrile-water ranked second in metabolome coverage but had fewer high-intensity compounds (**Figure 9**). It was ultimately determined that methanol was the most suitable extraction solvent for detecting metabolites with high abundance, which could be advantageous in the LC/MS-based metabolomics analysis extraction process.

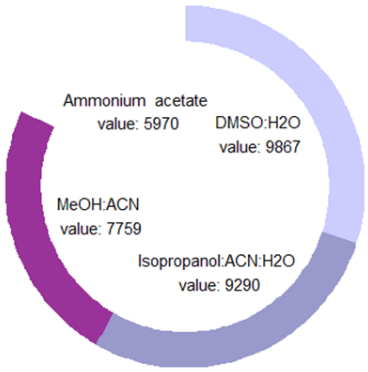


Figure 9. *The impact of four different extraction solvent mixtures on the metabolome coverage in DBS samples. The data shows the relative number of metabolites detected using each mixture.*

Finally, a cumulative comparative analysis was conducted to evaluate the efficacy of various methanol mixtures (e.g., methanol 100%, methanol-water (1:1, *v/v*), methanol-acetonitrile (1:1, *v/v*), and methanol-acetonitrile (3:1, *v/v*)) in extracting high-intensity metabolites. The results

demonstrated that the most efficient mixture was methanol-acetonitrile (3:1, *v/v*). Moreover, to ensure the reproducibility of the optimized extraction method, 95 samples were analyzed, including samples from CLN6 disease, control subjects, eight additional NCL, and nine types of LSD diseases. The primary objective of this study was to identify specific compounds associated with the diseases CLN6 disease, which was successfully achieved **(Figure 10).**

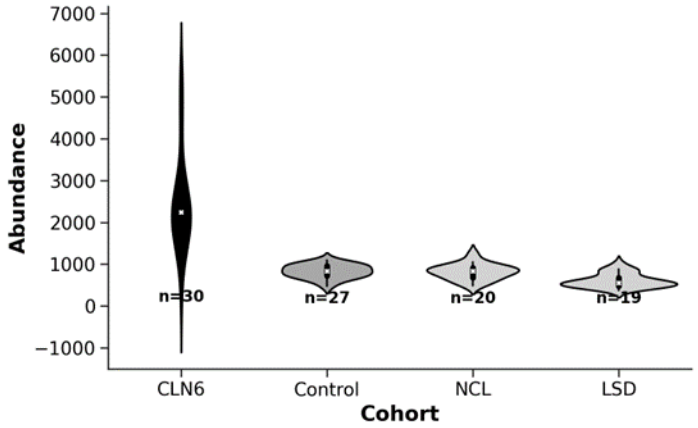


Figure 10. *The violin plot compares the abundance of a CLN6 compound (4.40_421.2811 m/z) across four cohorts - CLN6, Control, Other NCLs, Other LSDs - with outliers represented by rhombuses.*

4.2 Study 2: Clinical and genetic characterization of a cohort of 97 CLN6 patients tested at a single center

Neuronal ceroid lipofuscinosis (NCLs) is a group of rare genetically heterogeneous disorders with a diverse phenotype and global distribution. A thorough analysis of the NCL cohort (n=489) has uncovered significant trends using data gathered from NCL patients diagnosed at Centogene GmbH over a decade. The cohort comprised patients from around 40 countries, grouped into seven formal regions. Notably, the Middle East accounted for the majority of patients, followed by Africa and Europe (**Figure 11**). Specifically, the study has identified that certain regions, such as the Middle East and North Africa, display a higher incidence of patients with NCL disorder. The NCL cohort was composed of variants that were categorized as pathogenic (47%), likely pathogenic (24%), and variants of uncertain significance (29%). Out of the total cohort of 489 patients, 40% were classified as affected cases, 32% as carriers, 16% as probable carriers, and in 12% of cases the status could not be determined (**Figure 12**).

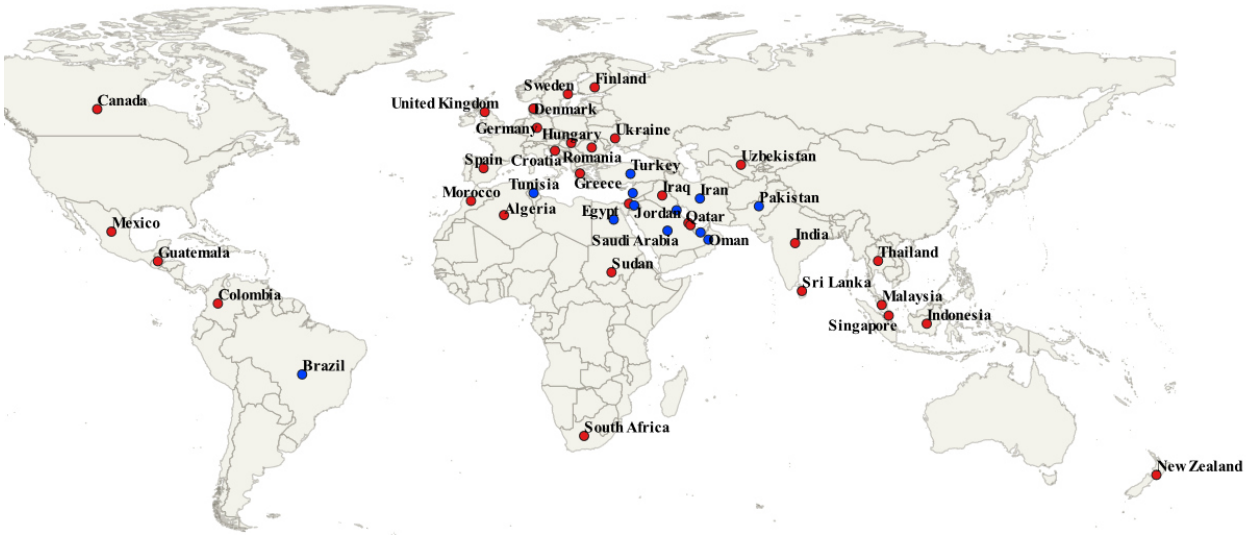


Figure 11. *Geographical distribution of the NCL pathological patients screened at Centogene, 2009-2019. Blue dots represent the CLN6 patients' distribution. (Map created with QGIS 3.10.1).*

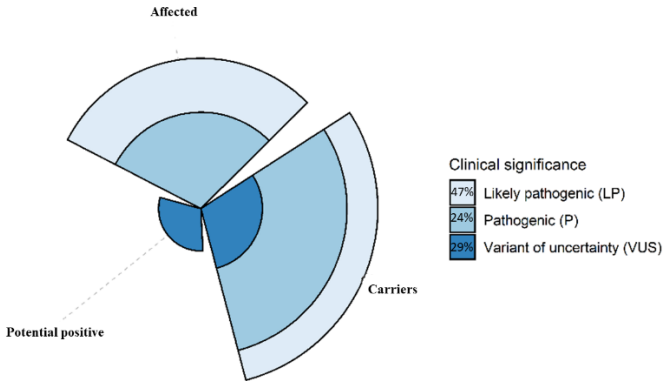


Figure 12. *Distribution of the genetic variants in genes responsible for NCL (Neuronal Ceroid Lipofuscinosis). Classified according to the clinical significance in the internal database CentoMD. Affected – compound heterozygous or homozygous for pathogenic (P) or likely pathogenic (LP), Carrier-heterozygous for pathogenic (P), likely pathogenic (LP) or variant of unknown significance (VUS), Potential positive – homozygous or compound heterozygous for the variant of unknown significance (VUS).*

Among the most prevalent clinical manifestations seen in the NCL cohort screened during the ten years were ataxia, seizures, and development regression (**Figure 13**), critical characteristics for the majority of CLNs, as mentioned in the literature.

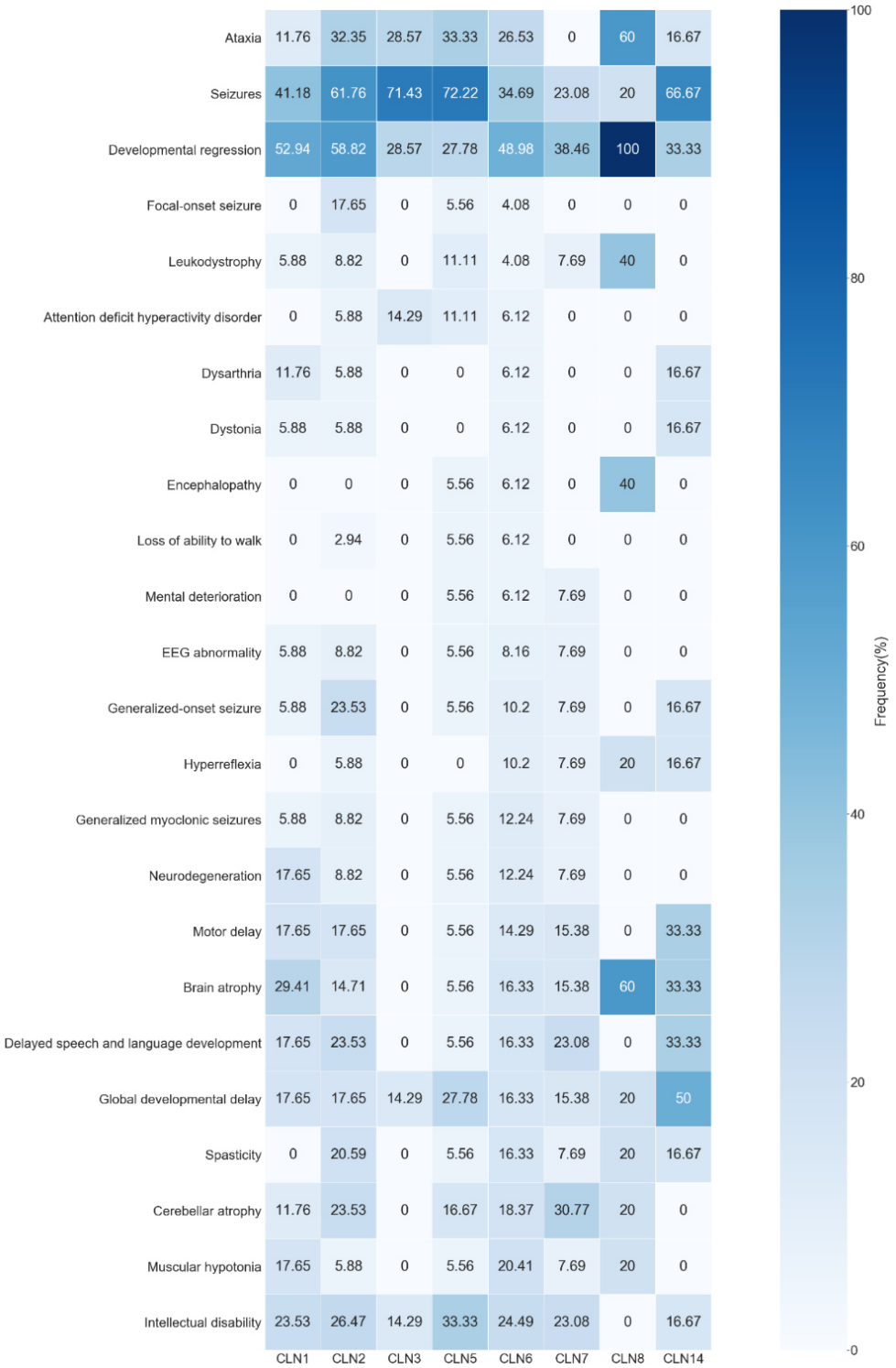


Figure 13. *The heatmap displays the prevalence of the most commonly encountered clinical symptoms in the NCL. It includes CLN diseases with more than five pathologic (P) or likely pathologic (LP) patients screened at Centogene GmbH.*

CLN6 disease is a subtype of NCL that results from mutations in the *CLN6* gene. Because of its rarity and diverse clinical manifestations, knowledge is scarce regarding the clinical and genetic characteristics of CLN6 disease, particularly among underrepresented populations. The second study reports the findings of a retrospective cross-sectional study that aimed to explore the clinical and genetic characteristics of CLN6 disease patients. The study analyzed data from 97 individuals who underwent routine genetic testing for the CLN gene at Centogene GmbH over ten years. The analysis identified 45 distinct variants, including 24 that had not been previously reported. Of the 45 unique variants identified, 16% were classified as pathogenic, 26% likely pathogenic, and 57% were categorized as variants of unknown significance (VUS).

This study has revealed that the most common mutation detected among the examined patients was in-frame deletion c.794_796del (p.265Sdel), with a detection rate of 31%. Furthermore, the novel discovered variant, c.257A>G (p.H86R), was identified in five unrelated cases, making it the most frequently reported new mutation (**Figure 14**).

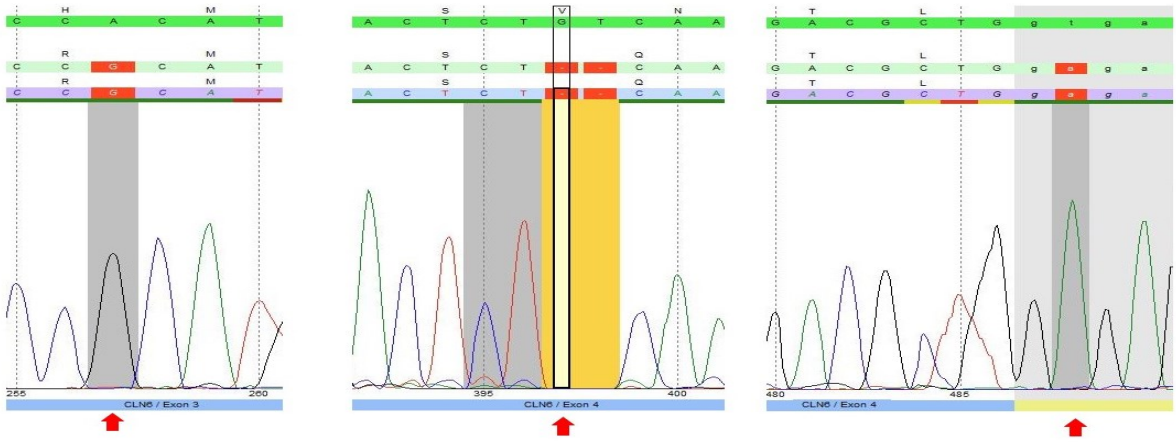


Figure 14. *Sequence chromatograms depict the novel CLN6 variants. The control © sequence is displayed above the mutation sequence. The affected nucleotide is underlined by a red line, indicating the variant's location. The electropherograms of CLN6 variants were identified by Sanger sequencing and included three novel variants: c.397_398del (**a**), c.257A>G (b), and c.486+2T>A (c).*

The most prevalent symptoms observed included "Developmental regression" (n= 46, 53%), "Seizure" 43%, "Ataxia" 33%, and "Intellectual disability" 30% (**Figure 15 a, b**). Other less common symptoms included speech impairment, muscle weakness, hypertonia, and microcephaly.

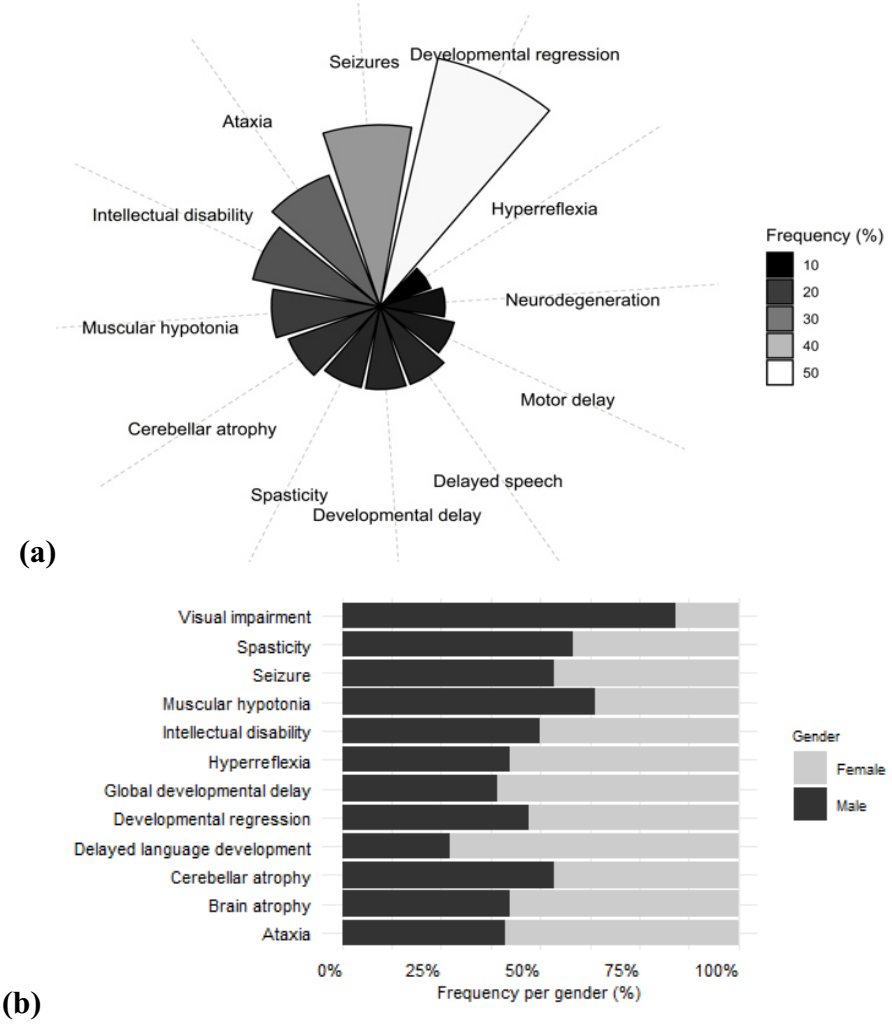


Figure 15. *Clinical characteristics representative of CLN6 disease. The prevalence of the most encountered clinical symptoms in the CLN6 disease* ***(a).*** *Gender differences in the clinical symptom's occurrence* ***(b).***

This study presents a comprehensive description of the clinical and genetic characteristics of CLN6 disease, with the aim of identifying potential correlations between causative variants. The study is particularly significant as it represents the largest cohort of CLN6 patients to undergo genetic testing at a single center, thereby expanding our understanding of the disease's mutational spectrum and highlighting its genetic and ethnic diversity.

As illustrated in **Figure 11,** the origins of the highest proportion of patients are Middle East and North Africa. This study emphasizes the importance of characterizing rare diseases in underrepresented populations. It recommends including CLN6 in genetic testing for individuals with developmental regression, seizures, ataxia, intellectual disability, and ocular symptoms. These

findings will raise awareness for CLN6 disease, shorten the time to diagnosis, and ultimately lead to more effective treatment.

4.3 Study 3: Neuronal progenitor cells-based metabolomics study reveals dysregulated lipid metabolism and identifies putative biomarkers for CLN6 disease

The primary objective of the third study was the comprehensive investigation into the metabolomics profile of individuals affected by CLN6 disease and to determine the corresponding metabolomics response in human CLN6 neuronal cell lines. This study expanded upon the initial CLN6 biomarker discovery (study 1), which involved the utilization of DBS cards, and delved further into investigating the potential effectiveness of employing cell lines in metabolomics studies.

Since CLN6 disease is a rare genetic disorder that affects the nervous system (Mole et al., 2004), it is reasonable to anticipate that the changes associated with the disease would be noticeable at the neuronal level. However, obtaining neuronal samples for research purposes can present significant challenges. Notably, commercially induced pluripotent stem cell (iPSC) models for CLN6 were unavailable, and the available cell lines were limited to three fibroblast lines. Thus, the reduced number of CLN6 cell samples, the lack of commercially available induced pluripotent stem cell (iPSC), and challenges associated with generating iPSCs from human dermal fibroblasts hinder cellular-level investigations. To address this, based on a methodology previously described by Dai *et al.* (Dai et al., 2015), neural progenitor cell-like cells were directly generated from human dermal fibroblasts (**Figure 16**).

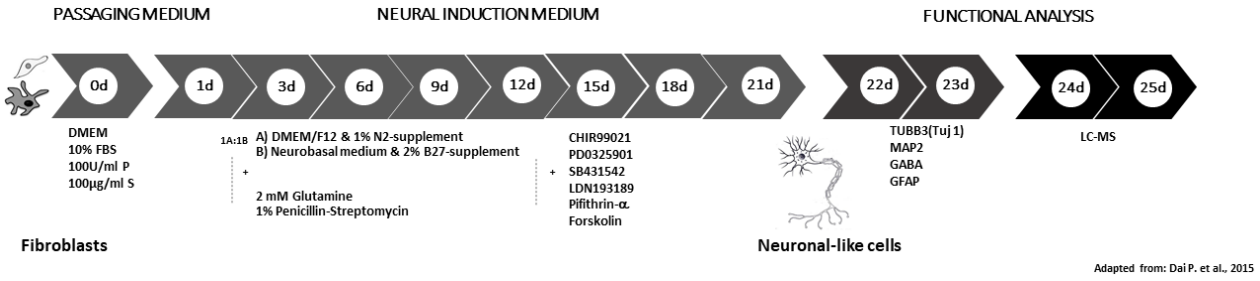


Figure 16. *Scheme of direct conversion of human dermal fibroblasts into neuronal-like cells. The experiment starts by plating the fibroblasts in the DMEM medium, designated "Day"0." After one day, the cells were transferred to an induction medium containing chemical compounds and supplementary chemicals to promote differentiation into neuronal progenitor cells. Adapted from Dai et al. (2015).*

The differentiation process resulted in the formation of neuronal cell-like clusters observed on day 12. Over time, from day 12 to day 20, the chemical-induced neuronal progenitor cells underwent visible transformations in their morphology. Notably, they exhibited the growth of axons and dendrites, indicating the development of essential neuronal structures. Moreover, these cells demonstrated distinct neuronal properties, including a well-defined central cell body housing a nucleus and multiple organelles. Furthermore, the formation of synapses was observed, highlighting their capacity for establishing functional connections (**Figure 17 a**). Following a three-week differentiation period, the expression of NPC markers, represented by GABA, GFAP, TUJ1, and MAP2, was confirmed through immunofluorescence staining, and comparable marker expression levels were observed between CLN6 and control (wild-type) NPCs (**Figure 17 b**).

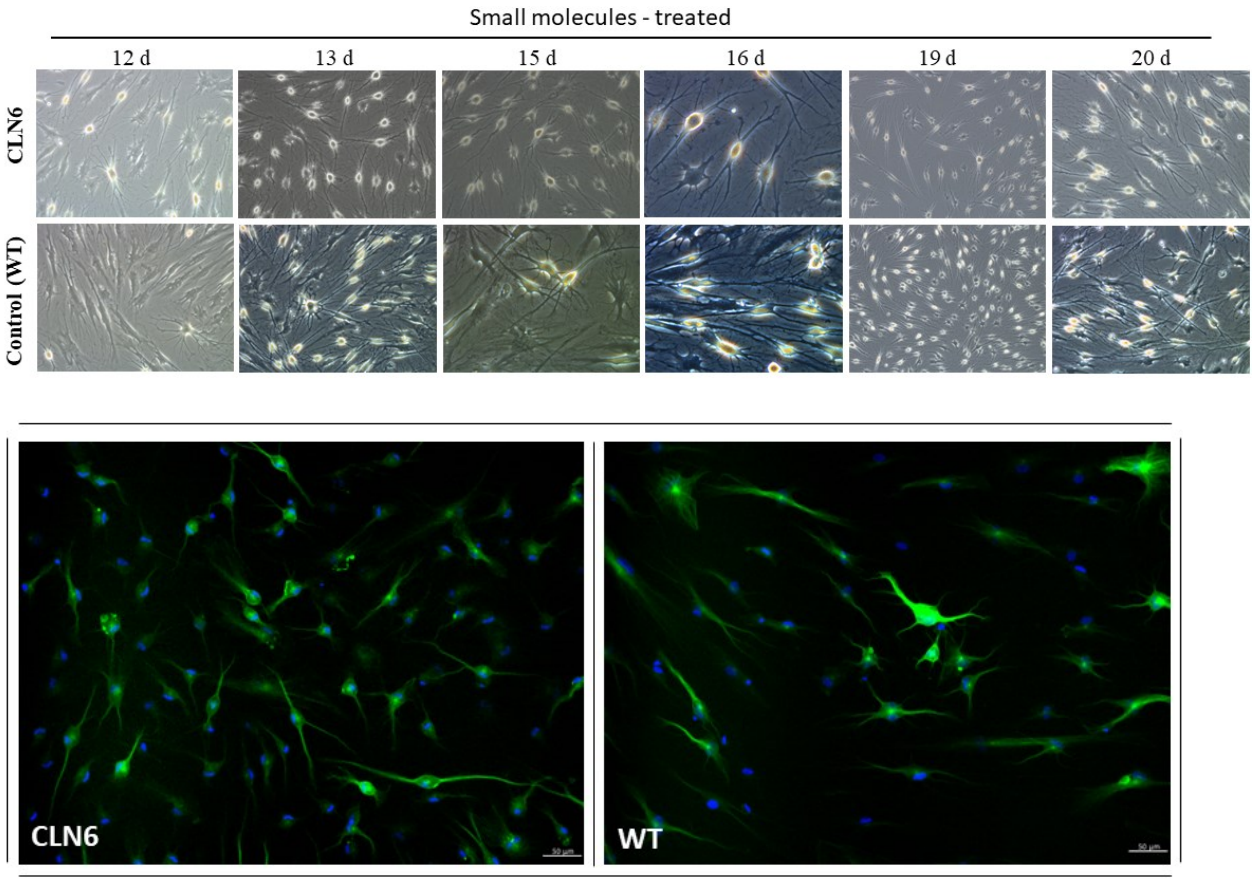


Figure 17. *Representative microscope images of chemical-induced neuronal cells at various developmental stages in the induction medium* ***(a).*** *Immunofluorescence staining of neuronal cultures with anti-MAP2 antibody (mature neurons). Nuclei were stained with DAPI (blue). Scale bars = 50 µm. MAP, microtubule-associated protein 2* ***(b).***

In this study, the cells were cultured in triplicates to enhance the statistical power, and an untargeted LC-MS metabolomics approach was employed. Upon automated data preprocessing, 2720 spectral features (ESI^-) were identified, revealing the significant influence of sample type and preparation method on the number and types of metabolites detected. It is worth noting that this number is four times smaller than the number detected using DBS cards.

The variation in metabolite quantity and diversity observed can be attributed to the influence of sample type and the variability in pre- and post-preparation processes. Dried blood spots could encompass a broader range of metabolites due to external factors, such as collection and storage, compared to cultured cells. Accordingly, distinct sample types may exhibit distinct metabolic profiles, affecting the detection of metabolites in terms of both quantity and variety. Additionally, disparate preparation techniques may result in the identification of different metabolites.

Metabolomics studies are complex and subject to various sources of variability, including but not limited to the sample type, storage temperature, sample processing, and pre-analytical methods. These factors require careful consideration and management to ensure accurate and reproducible results. The current lack of a standardized sample preparation protocol significantly contributes to the observed variability in the number of detected metabolites. Therefore, a rigorous and standardized protocol is required to minimize variability and ensure accurate and reproducible results in metabolomics studies (Ehrenberg et al., 2020, Kiseleva et al., 2021).

Furthermore, cell lines are helpful in metabolomics research as they provide consistency and reproducibility, allowing for easy manipulation of various factors, which are important to analyze and identify metabolites accurately. Cell-based samples can be grown in large quantities for high-throughput screening and are easy to maintain, increasing the reproducibility of results. Although cell lines may not fully capture the complexity of in vivo metabolic processes due to their limited nature compared to biofluids, they still play an essential role as a tool in metabolomics research (Čuperlović-Culf et al., 2010, Hayton et al., 2017). Therefore, while cell lines can be helpful for metabolomics research, it is essential to consider their limitations and potential impact on the number of metabolites detected.

After automatic data preprocessing, qualitative and quantitative filters were applied to enhance the accuracy and reliability of the metabolomics data while reducing the risk of false positives and negatives in downstream analysis. The filters included a signal-to-noise ratio, coefficient of

variation < 30%, a maximum abundance > 350, and maximum fold change > 2. Only the metabolites meeting the criteria above and passing the quality filters of m/z > 179 and retention time > 1 were retained for subsequent analysis to ensure data quality. Finally, the identified metabolites were analyzed using uni -and multivariate approaches. A student t-test was used for metabolite feature ranking in the univariate analysis, and the top 250 ($p < 0.005$) were selected for further multivariate analysis. The principal component analysis score plot indicated a notable separation between the CLN6 and control groups, with glycerophospholipids showing the most significant influence on variations.

In the final multivariate analysis, it was found through fold-change calculations that 15 lipids displayed regulation in opposing directions when comparing the two cohorts, with significant intensity differences noted. The top 15 metabolites were carefully selected based on their ability to differentiate between groups. Following the exclusion of compounds with low peak intensity, the analysis revealed that five of these metabolites, namely C16 GlcCer, C24 GlcCer, C24:1 GlcCer, and glycerophospholipids PG 40:6 and PG 40:7, were identified as potential biomarkers for early disease prognosis. The ROC curves revealed high precision and accuracy in distinguishing between the two groups (**Figure 18 a, b**). Despite the need for more thorough clinical investigations to evaluate the effectiveness of these metabolite markers, they show potential as diagnostic markers for individuals affected by CLN6 disease.

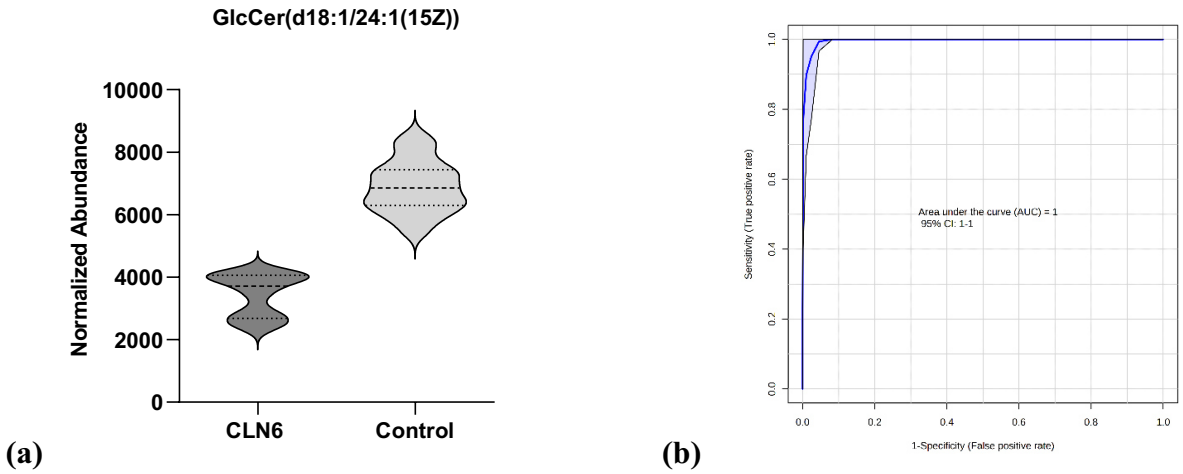


Figure 18*. Violin plot* ***(a)*** *and ROC curve* ***(b)*** *representation of one of the top selected biomarker candidates for CLN6 disease.*

Metabolomics data enrichment investigated the metabolic pathways and networks involved in CLN6 disease pathogenesis. The most significantly affected pathways were sphingolipid and glycerophospholipid metabolism, as the MetaboAnalyst pathway analysis module revealed. Further analysis using Ingenuity Pathway Analysis identified NAD signaling pathway and cyclic ADP-ribose deregulation significantly associated with differentially expressed metabolites in differentiated neurons from CLN6 patients. Moreover, a regulator effect network highlighted cell signaling, molecular transport, as well as vitamin and mineral metabolism functions.

5 Summary and outlook

This dissertation emphasizes the importance of in-depth research in addressing rare diseases like CLN6, highlighting LC-MS-based metabolomics as a valuable tool for biomarker discovery. It explores the potential of dried blood spot (DBS) samples in metabolomics-based biomarker research, investigating factors affecting metabolite stability over time and assessing various extraction solvents. Results indicate that while DBS samples may maintain short-term stability at room temperature, long-term storage at -20 °C is necessary for ensuring metabolite reliability. Sample type, card age, and storage conditions significantly impact metabolite yield, underscoring the need for consistent sample preparation and storage procedures. Methanol emerges as the most effective extraction solvent, enhancing metabolome coverage and peak intensity in LC-MS-based analyses.

Integrating DBS-based metabolomics into biomarker discovery holds promise for identifying disease-specific biomarkers. However, future studies should prioritize large cohort studies, carefully considering the influence of sample type, age, and storage conditions on metabolite stability. Standardizing these parameters is crucial for ensuring the reliability of metabolomics data. Furthermore, optimizing extraction protocols to maximize metabolome coverage is essential. These findings offer valuable insights into advancing biomarker research through DBS-based metabolomics, potentially enabling cost-effective and accessible biomarker screening methods.

The comprehensive analysis presented in this thesis highlights the intricate process of identifying biomarkers within rare diseases, particularly emphasizing the challenges posed by the heterogeneity of patients. Patient diversity, encompassing variations in age, sex, geographical

origin, and clinical presentation, complicates biomarker discovery, necessitating a new approach to address these complexities effectively.

Examining 97 CLN6 patients provided valuable insights into the disease's demographic, phenotypic, and genotypic characteristics. The prevalence of cases from the Middle East, particularly Egypt and Saudi Arabia, underscores the global impact of CLN6 disease, while the observed slight male predominance highlights the diverse nature of its affected population. The wide range of onset ages further emphasizes the significant implications of CLN6 disease across different stages of life. Key clinical features such as developmental regression, seizures, ataxia, and intellectual disability are prominently observed, illustrating the multifaceted nature of CLN6 disease. Identifying 42 unique CLN6 variants, including 22 novel ones, increases our understanding of the disease's genetic landscape.

Additionally, this study points out the importance of ongoing genetic testing and clinical research to establish genotype-phenotype correlations and develop personalized therapeutic interventions for CLN6 disease. It emphasizes the necessity of global awareness and diagnosis to enhance diagnostic capacities worldwide, supporting the inclusion of diverse populations in genetic investigations. Early and accurate genetic diagnosis remains crucial for effectively managing rare diseases like CLN6 and ultimately improving patient outcomes.

In summary, this thesis marks significant progress in understanding the metabolic changes linked with CLN6. Affected pathways and potential biomarker candidates were identified through a comprehensive analysis of the metabolome of human CLN6 ciNPCs using LC-MS-based metabolomics. Insights into alterations in sphingolipids and glycerophospholipid metabolism deepen our understanding of the fundamental pathophysiology of the condition. Investigating the global metabolic profiling and pathways in CLN6-affected patients constitutes a step toward understanding this disease. The insights obtained from LC-MS-based metabolomics shed light on the metabolic shifts associated with CLN6 disease, facilitating the identification of potential biomarker candidates for early disease detection. The promise of these biomarker candidates in aiding early diagnosis or treatment monitoring of CLN6 disease is considerable. Continued research and validation efforts are imperative to ascertain the clinical efficacy of these compounds, especially given the current absence of treatment options for CLN6 disease and the limitations of existing genetic tests in tracking disease progression and therapy response.

REFERENCES

1995. Isolation of a novel gene underlying Batten disease, CLN3. The International Batten Disease Consortium. *Cell,* 82**,** 949-57.

ALSEEKH, S., AHARONI, A., BROTMAN, Y., CONTREPOIS, K., D'AURIA, J., EWALD, J., J, C. E., FRASER, P. D., GIAVALISCO, P., HALL, R. D., HEINEMANN, M., LINK, H., LUO, J., NEUMANN, S., NIELSEN, J., PEREZ DE SOUZA, L., SAITO, K., SAUER, U., SCHROEDER, F. C., SCHUSTER, S., SIUZDAK, G., SKIRYCZ, A., SUMNER, L. W., SNYDER, M. P., TANG, H., TOHGE, T., WANG, Y., WEN, W., WU, S., XU, G., ZAMBONI, N. & FERNIE, A. R. 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods,* 18**,** 747-756.

ARONSON, J. K. & FERNER, R. E. 2017. Biomarkers-A General Review. *Curr Protoc Pharmacol,* 76**,** 9.23.1-9.23.17.

ATTARD, J. A., DUNN, W. B., MERGENTAL, H., MIRZA, D. F., AFFORD, S. C. & PERERA, M. T. P. J. O. A. J. O. I. B. 2019. Systematic review: clinical metabolomics to forecast outcomes in liver transplantation surgery. 23**,** 463-476.

AUGUSTINE, E. F., ADAMS, H. R., DE LOS REYES, E., DRAGO, K., FRAZIER, M., GUELBERT, N., LAINE, M., LEVIN, T., MINK, J. W., NICKEL, M., PEIFER, D., SCHULZ, A., SIMONATI, A., TOPCU, M., TURUNEN, J. A., WILLIAMS, R., WIRRELL, E. C. & KING, S. 2021. Management of CLN1 Disease: International Clinical Consensus. *Pediatr Neurol,* 120**,** 38-51.

BATTEN, F. E. Cerebral degeneration, with symmetrical changes in the maculae, in two members of a family.

BAX, B. E. 2021. Biomarkers in Rare Diseases. *Int J Mol Sci,* 22.

BECK-WODL, S., HARZER, K., STURM, M., BUCHERT, R., RIESS, O., MENNEL, H. D., LATTA, E., PAGENSTECHER, A. & KEBER, U. 2018. Homozygous TBC1 domain-containing kinase (TBCK) mutation causes a novel lysosomal storage disease - a new type of neuronal ceroid lipofuscinosis (CLN15)? *Acta Neuropathol Commun,* 6**,** 145.

BECKONERT, O., KEUN, H. C., EBBELS, T. M., BUNDY, J., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc,* 2**,** 2692-703.

BEHL, T., KAUR, I., SEHGAL, A., SINGH, S., ALBARRATI, A., ALBRATTY, M., NAJMI, A., MERAYA, A. M. & BUNGAU, S. 2022. The road to precision medicine: Eliminating the “One Size Fits All” approach in Alzheimer’s disease. *Biomedicine & Pharmacotherapy,* 153**,** 113337.

BENEDICT, J. W., GETTY, A. L., WISHART, T. M., GILLINGWATER, T. H. & PEARCE, D. A. 2009. Protein product of CLN6 gene responsible for variant late-onset infantile neuronal ceroid lipofuscinosis interacts with CRMP-2. *J Neurosci Res,* 87**,** 2157-66.

BEST, H. L., CLARE, A. J., MCDONALD, K. O., WICKY, H. E. & HUGHES, S. M. 2021. An altered secretome is an early marker of the pathogenesis of CLN6 Batten disease. *J Neurochem,* 157**,** 764-780.

BLENNOW, K. & ZETTERBERG, H. 2018. The Past and the Future of Alzheimer’s Disease Fluid Biomarkers. *Journal of Alzheimer's Disease,* 62**,** 1125-1140.

BODAGHI, A., FATTAHI, N. & RAMAZANI, A. 2023. Biomarkers: Promising and valuable tools towards diagnosis, prognosis and treatment of Covid-19 and other diseases. *Heliyon,* 9**,** e13323.

BRAS, J., VERLOES, A., SCHNEIDER, S. A., MOLE, S. E. & GUERREIRO, R. J. 2012. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Hum Mol Genet,* 21**,** 2646-50.

BUCHHAVE, P., MINTHON, L., ZETTERBERG, H., WALLIN, A. K., BLENNOW, K. & HANSSON, O. 2012. Cerebrospinal fluid levels of β-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Arch Gen Psychiatry,* 69**,** 98-106.

BUJAK, R., STRUCK-LEWICKA, W., MARKUSZEWSKI, M. J. & KALISZAN, R. 2015. Metabolomics for laboratory diagnostics. *Journal of Pharmaceutical and Biomedical Analysis,* 113**,** 108-120.

BUTZ, E. S., CHANDRACHUD, U., MOLE, S. E. & COTMAN, S. L. 2020. Moving towards a new era of genomics in the neuronal ceroid lipofuscinoses. *Biochim Biophys Acta Mol Basis Dis,* 1866**,** 165571.

CAGNEY, D. N., SUL, J., HUANG, R. Y., LIGON, K. L., WEN, P. Y. & ALEXANDER, B. M. J. N.-O. 2018. The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology. 20**,** 1162-1172.

CALIFF, R. M. J. E. B. & MEDICINE 2018. Biomarker definitions and their applications. 243**,** 213-221.

CANAFOGLIA, L., GILIOLI, I., INVERNIZZI, F., SOFIA, V., FUGNANESI, V., MORBIN, M., CHIAPPARINI, L., GRANATA, T., BINELLI, S., SCAIOLI, V., GARAVAGLIA, B., NARDOCCI, N., BERKOVIC, S. F. & FRANCESCHETTI, S. 2015. Electroclinical spectrum of the neuronal ceroid lipofuscinoses associated with CLN6 mutations. *Neurology,* 85**,** 316-24.

CANNELLI, N., GARAVAGLIA, B., SIMONATI, A., AIELLO, C., BARZAGHI, C., PEZZINI, F., CILIO, M. R., BIANCHERI, R., MORBIN, M., DALLA BERNARDINA, B., GRANATA, T., TESSA, A., INVERNIZZI, F., PESSAGNO, A., BOLDRINI, R., ZIBORDI, F., GRAZIAN, L., CLAPS, D., CARROZZO, R., MOLE, S. E., NARDOCCI, N. & SANTORELLI, F. M. 2009. Variant late infantile ceroid lipofuscinoses associated with novel mutations in CLN6. *Biochem Biophys Res Commun,* 379**,** 892-7.

CAO, J., WANG, Y., ZHANG, Y. & QIAN, K. 2022. Emerging Applications of Mass Spectrometry-Based Metabolic Fingerprinting in Clinics. 4**,** 2100191.

CARAYOL, M., LICAJ, I., ACHAINTRE, D., SACERDOTE, C., VINEIS, P., KEY, T. J., ONLAND MORET, N. C., SCALBERT, A., RINALDI, S. & FERRARI, P. 2015. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoS One,* 10**,** e0135437.

CASETTA, B., MALVAGIA, S., FUNGHINI, S., MARTINELLI, D., DIONISI-VICI, C., BARONE, R., FIUMARA, A., DONATI, M. A., GUERRINI, R. & LA MARCA, G. 2020. A new strategy implementing mass spectrometry in the diagnosis of congenital disorders of N-glycosylation (CDG). *Clin Chem Lab Med,* 59**,** 165-171.

CHEN, C. J., LEE, D. Y., YU, J., LIN, Y. N. & LIN, T. M. 2023. Recent advances in LC-MS-based metabolomics for clinical biomarker discovery. *Mass Spectrom Rev,* 42**,** 2349-2378.

CLARK, M. M., STARK, Z., FARNAES, L., TAN, T. Y., WHITE, S. M., DIMMOCK, D. & KINGSMORE, S. F. 2018. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med,* 3**,** 16.

CLINICALTRIALS.GOV. *Neuronal Ceroid Lipofuscinosis* [Online]. Available: https://www.clinicaltrials.gov/ct2/results?term=neuronal+ceroid+lipofuscinosis&Search=Search [Accessed 12 March 2023].

COMES, A. L., PAPIOL, S., MUELLER, T., GEYER, P. E., MANN, M. & SCHULZE, T. G. J. T. P. 2018. Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future. 8**,** 160.

CROWTHER, L. M., POMS, M. & PLECKO, B. 2018. Multiomics tools for the diagnosis and treatment of rare neurological disease. *J Inherit Metab Dis,* 41**,** 425-434.

CUI, L., LU, H. & LEE, Y. H. 2018. Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases. *Mass Spectrom Rev,* 37**,** 772-792.

ČUPERLOVIĆ-CULF, M., BARNETT, D. A., CULF, A. S. & CHUTE, I. 2010. Cell culture metabolomics: applications and future directions. *Drug Discovery Today,* 15**,** 610-621.

DAI, P., HARADA, Y. & TAKAMATSU, T. 2015. Highly efficient direct conversion of human fibroblasts to neuronal cells by chemical compounds. *J Clin Biochem Nutr,* 56**,** 166-70.

DAPIC, I., BALJEU-NEUMAN, L., UWUGIAREN, N., KERS, J., GOODLETT, D. R. & CORTHALS, G. L. 2019. Proteome analysis of tissues by mass spectrometry. *Mass Spectrom Rev,* 38**,** 403-441.

DAVIS, J. C., FURSTENTHAL, L., DESAI, A. A., NORRIS, T., SUTARIA, S., FLEMING, E. & MA, P. J. N. R. D. D. 2009. The microeconomics of personalized medicine: today's challenge and tomorrow's promise. 8**,** 279-286.

DELAYE, J., CACCIATORE, P. & KOLE, A. 2022. Valuing the "Burden" and Impact of Rare Diseases: A Scoping Review. *Front Pharmacol,* 13**,** 914338.

DIETERLE, F., ROSS, A., SCHLOTTERBECK, G. & SENN, H. 2006. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem,* 78**,** 4281-90.

DROLET, J., TOLSTIKOV, V., WILLIAMS, B. A., GREENWOOD, B. P., HILL, C., VISHNUDAS, V. K., SARANGARAJAN, R., NARAIN, N. R. & KIEBISH, M. A. 2017. Integrated Metabolomics Assessment of Human Dried Blood Spots and Urine Strips. *Metabolites,* 7.

DUNN, W. B., ERBAN, A., WEBER, R. J., CREEK, D. J., BROWN, M., BREITLING, R., HANKEMEIER, T., GOODACRE, R., NEUMANN, S. & KOPKA, J. J. M. 2013. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. 9**,** 44-66.

EHRENBERG, A. J., KHATUN, A., COOMANS, E., BETTS, M. J., CAPRARO, F., THIJSSEN, E. H., SENKEVICH, K., BHARUCHA, T., JAFARPOUR, M., YOUNG, P. N. E., JAGUST, W., CARTER, S. F., LASHLEY, T., GRINBERG, L. T., PEREIRA, J. B., MATTSSON-CARLGREN, N., ASHTON, N. J., HANRIEDER, J., ZETTERBERG, H., SCHÖLL, M. & PATERSON, R. W. 2020. Relevance of biomarkers across different neurodegenerative diseases. *Alzheimers Res Ther,* 12**,** 56.

EMWAS, A. H. 2015. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol Biol,* 1277**,** 161-93.

EUROPEAN PARLIAMENT, C. O. T. E. U. 2000. *Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on Orphan Medicinal Products* [Online]. Brussels: European Union. Available: https://eur-lex.europa.eu/EN/legal-content/summary/medicines-for-rare-diseases-orphan-drugs.html [Accessed February 11 2024].

FARAHANY, N. A., GREELY, H. T., HYMAN, S., KOCH, C., GRADY, C., PAȘCA, S. P., SESTAN, N., ARLOTTA, P., BERNAT, J. L., TING, J., LUNSHOF, J. E., IYER, E. P. R., HYUN, I., CAPESTANY, B. H., CHURCH, G. M., HUANG, H. & SONG, H. 2018. The ethics of experimenting with human brain tissue. *Nature,* 556**,** 429-432.

FIETZ, M., ALSAYED, M., BURKE, D., COHEN-PFEFFER, J., COOPER, J. D., DVOŘÁKOVÁ, L., GIUGLIANI, R., IZZO, E., JAHNOVÁ, H., LUKACS, Z., MOLE, S. E., NOHER DE HALAC, I., PEARCE, D. A., POUPETOVA, H., SCHULZ, A., SPECCHIO, N., XIN, W. & MILLER, N. 2016. Diagnosis of neuronal ceroid lipofuscinosis type 2 (CLN2 disease): Expert recommendations for early detection and laboratory diagnosis. *Mol Genet Metab,* 119**,** 160-7.

FILIPP, F. V. 2019. Opportunities for Artificial Intelligence in Advancing Precision Medicine. *Curr Genet Med Rep,* 7**,** 208-213.

FITZGERALD, G. A. J. S. T. M. 2016. Measure for measure: biomarker standards and transparency. 8**,** 343fs10-343fs10.

FOUNTZILAS, E., TSIMBERIDOU, A. M., VO, H. H. & KURZROCK, R. 2022. Clinical trial design in the era of precision medicine. *Genome Medicine,* 14**,** 101.

FRAINAY, C., SCHYMANSKI, E. L., NEUMANN, S., MERLET, B., SALEK, R. M., JOURDAN, F. & YANES, O. 2018. Mind the Gap: Mapping Mass Spectral Databases in Genome-Scale Metabolic Networks Reveals Poorly Covered Areas. *Metabolites,* 8.

FRANGOGIANNIS, N. G. 2012. Biomarkers: hopes and challenges in the path from discovery to clinical practice. *Transl Res,* 159**,** 197-204.

GANESALINGAM, J. & BOWSER, R. 2010. The application of biomarkers in clinical trials for motor neuron disease. *Biomark Med,* 4**,** 281-97.

GAO, H., BOUSTANY, R. M., ESPINOLA, J. A., COTMAN, S. L., SRINIDHI, L., ANTONELLIS, K. A., GILLIS, T., QIN, X., LIU, S., DONAHUE, L. R., BRONSON, R. T., FAUST, J. R., STOUT, D., HAINES, J. L., LERNER, T. J. & MACDONALD, M. E. 2002. Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. *Am J Hum Genet,* 70**,** 324-35.

GOODACRE, R., VAIDYANATHAN, S., DUNN, W. B., HARRIGAN, G. G. & KELL, D. B. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol,* 22**,** 245-52.

GROFT, S. C., POSADA, M. & TARUSCIO, D. 2021. Progress, challenges and global approaches to rare diseases. *Acta Paediatr,* 110**,** 2711-2716.

GROMOVA, M., VAGGELAS, A., DALLMANN, G. & SEIMETZ, D. 2020. Biomarkers: Opportunities and Challenges for Drug Development in the Current Regulatory Landscape. 15**,** 1177271920974652.

GROUP, B. D. W., ATKINSON JR, A. J., COLBURN, W. A., DEGRUTTOLA, V. G., DEMETS, D. L., DOWNING, G. J., HOTH, D. F., OATES, J. A., PECK, C. C., SCHOOLEY, R. T. J. C. P. & THERAPEUTICS 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. 69**,** 89-95.

GROUP, F. D. A. N. I. H. B. W. 2016. *BEST (Biomarkers, EndpointS, and Other Tools) Resource*, Food and Drug Administration (US).

HALKET, J. M., WATERMAN, D., PRZYBOROWSKA, A. M., PATEL, R. K., FRASER, P. D. & BRAMLEY, P. M. 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot,* 56**,** 219-43.

HARTL, D., DE LUCA, V., KOSTIKOVA, A., LARAMIE, J., KENNEDY, S., FERRERO, E., SIEGEL, R., FINK, M., AHMED, S., MILLHOLLAND, J., SCHUHMACHER, A., HINDER, M., PIALI, L. & ROTH, A. 2021. Translational precision medicine: an industry perspective. *Journal of Translational Medicine,* 19**,** 245.

HAYTON, S., MAKER, G. L., MULLANEY, I. & TRENGOVE, R. D. 2017. Experimental design and reporting standards for metabolomics studies of mammalian cell lines. *Cell Mol Life Sci,* 74**,** 4421-4441.

HEINE, C., KOCH, B., STORCH, S., KOHLSCHUTTER, A., PALMER, D. N. & BRAULKE, T. 2004. Defective endoplasmic reticulum-resident membrane protein CLN6 affects lysosomal degradation of endocytosed arylsulfatase A. *J Biol Chem,* 279**,** 22347-52.

HERNANDES, V. V., BARBAS, C. & DUDZIK, D. 2017. A review of blood sample handling and pre-processing for metabolomics studies. *Electrophoresis,* 38**,** 2232-2241.

HERSRUD, S. L., GERAETS, R. D., WEBER, K. L., CHAN, C. H. & PEARCE, D. A. 2016. Plasma biomarkers for neuronal ceroid lipofuscinosis. *Febs j,* 283**,** 459-71.

HEUYER, T., PAVAN, S. & VICARD, C. 2017. The health and life path of rare disease patients: results of the 2015 French barometer. *Patient Relat Outcome Meas,* 8**,** 97-110.

HIRAYAMA, A., SUGIMOTO, M., SUZUKI, A., HATAKEYAMA, Y., ENOMOTO, A., HARADA, S., SOGA, T., TOMITA, M. & TAKEBAYASHI, T. 2015. Effects of processing and storage conditions on charged metabolomic profiles in blood. *Electrophoresis,* 36**,** 2148-2155.

HOCHSTEIN, J. N., SCHULZ, A., NICKEL, M., LEZIUS, S., GROSSER, M., FIEHLER, J., SEDLACIK, J. & LÖBEL, U. 2022. Natural history of MRI brain volumes in patients with neuronal ceroid lipofuscinosis 3: a sensitive imaging biomarker. *Neuroradiology,* 64**,** 2059-2067.

HOLLYWOOD, K., BRISON, D. R. & GOODACRE, R. 2006. Metabolomics: current technologies and future trends. *Proteomics,* 6**,** 4716-23.

HOLMES, E., LOO, R. L., STAMLER, J., BICTASH, M., YAP, I. K., CHAN, Q., EBBELS, T., DE IORIO, M., BROWN, I. J., VESELKOV, K. A., DAVIGLUS, M. L., KESTELOOT, H., UESHIMA, H., ZHAO, L., NICHOLSON, J. K. & ELLIOTT, P. 2008a. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature,* 453**,** 396-400.

HOLMES, E., WILSON, I. D. & NICHOLSON, J. K. 2008b. Metabolic phenotyping in health and disease. *Cell,* 134**,** 714-7.

HOSOYA, M. & CZYSZ, K. 2016. Translational Prospects and Challenges in Human Induced Pluripotent Stem Cell Research in Drug Discovery. *Cells,* 5.

HUAN, T., FORSBERG, E. M., RINEHART, D., JOHNSON, C. H., IVANISEVIC, J., BENTON, H. P., FANG, M., AISPORNA, A., HILMERS, B., POOLE, F. L., THORGERSEN, M. P., ADAMS, M. W. W., KRANTZ, G., FIELDS, M. W., ROBBINS, P. D., NIEDERNHOFER, L. J., IDEKER, T., MAJUMDER, E. L., WALL, J. D.,

RATTRAY, N. J. W., GOODACRE, R., LAIRSON, L. L. & SIUZDAK, G. 2017. Systems biology guided by XCMS Online metabolomics. *Nat Methods,* 14**,** 461-462.

ISSAQ, H. J. & VEENSTRA, T. D. 2020. Chapter 1 - Biomarker discovery: Study design and execution. *In:* ISSAQ, H. J. & VEENSTRA, T. D. (eds.) *Proteomic and Metabolomic Approaches to Biomarker Discovery (Second Edition).* Boston: Academic Press.

IWAN, K., CLAYTON, R., MILLS, P., CSANYI, B., GISSEN, P., MOLE, S. E., PALMER, D. N., MILLS, K. & HEYWOOD, W. E. 2021. Urine proteomics analysis of patients with neuronal ceroid lipofuscinoses. *iScience,* 24**,** 102020.

JAIN, K. K. 2010. Role of Biomarkers in Health Care. *The Handbook of Biomarkers***,** 115-88.

JIANG, X., SIDHU, R., PORTER, F. D., YANJANIN, N. M., SPEAK, A. O., TE VRUCHTE, D. T., PLATT, F. M., FUJIWARA, H., SCHERRER, D. E., ZHANG, J., DIETZEN, D. J., SCHAFFER, J. E. & ORY, D. S. 2011. A sensitive and specific LC-MS/MS method for rapid diagnosis of Niemann-Pick C1 disease from human plasma. *J Lipid Res,* 52**,** 1435-45.

JOHNSON, T. B., BRUDVIG, J. J., LEHTIMÄKI, K. K., CAIN, J. T., WHITE, K. A., BRAGGE, T., RYTKÖNEN, J., HUHTALA, T., TIMM, D., VIHMA, M., PUOLIVÄLI, J. T., POUTIAINEN, P., NURMI, A. & WEIMER, J. M. 2020. A multimodal approach to identify clinically relevant biomarkers to comprehensively monitor disease progression in a mouse model of pediatric neurodegenerative disease. *Progress in Neurobiology,* 189**,** 101789.

JOHNSON, T. B., CAIN, J. T., WHITE, K. A., RAMIREZ-MONTEALEGRE, D., PEARCE, D. A. & WEIMER, J. M. 2019. Therapeutic landscape for Batten disease: current treatments and future prospects. *Nat Rev Neurol,* 15**,** 161-178.

JUNG, S., TRAN, N. T., GOSPE, S. M., JR. & HAHN, S. H. 2013. Preliminary investigation of the use of newborn dried blood spots for screening pyridoxine-dependent epilepsy by LC-MS/MS. *Mol Genet Metab,* 110**,** 237-40.

JUNOT, C., FENAILLE, F., COLSCH, B. & BÉCHER, F. 2014. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom Rev,* 33**,** 471-500.

KAŁUŻNA-CZAPLIŃSKA, J. 2011. Current medical research with the application of coupled techniques with mass spectrometry. *Med Sci Monit,* 17**,** Ra117-23.

KAMINIÓW, K., KOZAK, S. & PAPROCKA, J. 2022. Recent Insight into the Genetic Basis, Clinical Features, and Diagnostic Methods for Neuronal Ceroid Lipofuscinosis. 23**,** 5729.

KAMLAGE, B., MALDONADO, S. G., BETHAN, B., PETER, E., SCHMITZ, O., LIEBENBERG, V. & SCHATZ, P. 2014. Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling. *Clin Chem,* 60**,** 399-412.

KANEHISA, M. 2016. KEGG Bioinformatics Resource for Plant Genomics and Metabolomics. *Methods Mol Biol,* 1374**,** 55-70.

KATAJAMAA, M., MIETTINEN, J. & ORESIC, M. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics,* 22**,** 634-6.

KERR, K., MCANENEY, H., SMYTH, L. J., BAILIE, C., MCKEE, S. & MCKNIGHT, A. J. 2020. A scoping review and proposed workflow for multi-omic rare disease research. *Orphanet Journal of Rare Diseases,* 15**,** 107.

KEUN, H. C., EBBELS, T. M., BOLLARD, M. E., BECKONERT, O., ANTTI, H., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. 2004. Geometric trajectory analysis of metabolic responses to toxicity can define treatment specific profiles. *Chem Res Toxicol,* 17**,** 579-87.

KHAN, S., BARVE, K. H. & KUMAR, M. S. 2020. Recent Advancements in Pathogenesis, Diagnostics and Treatment of Alzheimer's Disease. *Curr Neuropharmacol,* 18**,** 1106-1125.

KIM, S., CHEN, J., CHENG, T., GINDULYTE, A., HE, J., HE, S., LI, Q., SHOEMAKER, B. A., THIESSEN, P. A., YU, B., ZASLAVSKY, L., ZHANG, J. & BOLTON, E. E. 2021. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res,* 49**,** D1388-D1395.

KIRWAN, J. A., BRENNAN, L., BROADHURST, D., FIEHN, O., CASCANTE, M., DUNN, W. B., SCHMIDT, M. A. & VELAGAPUDI, V. 2018. Preanalytical Processing and Biobanking Procedures of Biological Samples for Metabolomics Research: A White Paper, Community Perspective (for "Precision Medicine and Pharmacometabolomics Task Group"-The Metabolomics Society Initiative). *Clin Chem,* 64**,** 1158-1182.

KISELEVA, O., KURBATOV, I., ILGISONIS, E. & POVERENNAYA, E. 2021. Defining Blood Plasma and Serum Metabolome by GC-MS. *Metabolites,* 12.

KLIEGMAN, R. M., BORDINI, B. J., BASEL, D. & NOCTON, J. J. 2017. How Doctors Think: Common Diagnostic Errors in Clinical Judgment—Lessons from an Undiagnosed and Rare Disease Program. *Pediatric Clinics of North America,* 64**,** 1-15.

KLINE, R. A., WISHART, T. M., MILLS, K. & HEYWOOD, W. E. 2020. Applying modern Omic technologies to the Neuronal Ceroid Lipofuscinoses. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1866**,** 165498.

KOLLMANN, K., UUSI-RAUVA, K., SCIFO, E., TYYNELA, J., JALANKO, A. & BRAULKE, T. 2013. Cell biology and function of neuronal ceroid lipofuscinosis-related proteins. *Biochim Biophys Acta,* 1832**,** 1866-81.

KONÍČKOVÁ, D., MENŠÍKOVÁ, K., TUČKOVÁ, L., HÉNYKOVÁ, E., STRNAD, M., FRIEDECKÝ, D., STEJSKAL, D., MATĚJ, R. & KAŇOVSKÝ, P. 2022. Biomarkers of Neurodegenerative Diseases: Biology, Taxonomy, Clinical Relevance, and Current Research Status. *Biomedicines,* 10.

KOUSI, M., LEHESJOKI, A. E. & MOLE, S. E. 2012. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum Mutat,* 33**,** 42-63.

KYTTALA, A., LAHTINEN, U., BRAULKE, T. & HOFMANN, S. L. 2006. Functional biology of the neuronal ceroid lipofuscinoses (NCL) proteins. *Biochim Biophys Acta,* 1762**,** 920-33.

LANZNASTER, D., DINGEO, G., SAMEY, R. A., EMOND, P. & BLASCO, H. 2022. Metabolomics as a Crucial Tool to Develop New Therapeutic Strategies for Neurodegenerative Diseases. *Metabolites,* 12.

LEE, D., BOWEN, B. & NORTHEN, T. 2010. Mass spectrometry-based metabolomics, analysis of metabolite-protein interactions, and imaging. *BioTechniques,* 49**,** 557-65.

LEHMANN, R. 2021. From bedside to bench-practical considerations to avoid pre-analytical pitfalls and assess sample quality for high-resolution metabolomics and lipidomics analyses of body fluids. *Anal Bioanal Chem,* 413**,** 5567-5585.

LENZ, E. M. & WILSON, I. D. 2007. Analytical strategies in metabonomics. *J Proteome Res,* 6**,** 443-58.

LESCUYER, P., HOCHSTRASSER, D. & RABILLOUD, T. J. J. O. P. R. 2007. How shall we use the proteomics toolbox for biomarker discovery? 6**,** 3371-3376.

LI, K., NAVIAUX, J. C., MONK, J. M., WANG, L. & NAVIAUX, R. K. 2020. Improved Dried Blood Spot-Based Metabolomics: A Targeted, Broad-Spectrum, Single-Injection Method. *Metabolites,* 10.

LIANG, Y. & KELEMEN, A. J. B. I. B. 2018. Dynamic modeling and network approaches for omics time course data: overview of computational approaches and applications. 19**,** 1051-1068.

LIEBAL, U. W., PHAN, A. N. T., SUDHAKAR, M., RAMAN, K. & BLANK, L. M. 2020. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. 10**,** 243.

LIN, C. Y., WU, H., TJEERDEMA, R. S. & VIANT, M. R. J. M. 2007. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. 3**,** 55-67.

LITTLE, J. L., WILLIAMS, A. J., PSHENICHNOV, A. & TKACHENKO, V. 2012. Identification of "known unknowns" utilizing accurate mass data and ChemSpider. *J Am Soc Mass Spectrom,* 23**,** 179-85.

LIU, Z., ZHU, L., ROBERTS, R. & TONG, W. 2019. Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? *Trends in Genetics,* 35**,** 852-867.

LOPES, A. S., CRUZ, E. C., SUSSULINI, A. & KLASSEN, A. 2017. Metabolomic Strategies Involving Mass Spectrometry Combined with Liquid and Gas Chromatography. *Adv Exp Med Biol,* 965**,** 77-98.

LOPES, M. T., KOCH, V. H., SARRUBBI-JUNIOR, V., GALLO, P. R. & CARNEIRO-SAMPAIO, M. 2018. Difficulties in the diagnosis and treatment of rare diseases according to the perceptions of patients, relatives and health care professionals. *Clinics (Sao Paulo),* 73**,** e68.

LU, X., ZHAO, X., BAI, C., ZHAO, C., LU, G. & XU, G. 2008. LC-MS-based metabonomics analysis. *J Chromatogr B Analyt Technol Biomed Life Sci,* 866**,** 64-76.

LUNKE, S., BOUFFLER, S. E., PATEL, C. V., SANDARADURA, S. A., WILSON, M., PINNER, J., HUNTER, M. F., BARNETT, C. P., WALLIS, M., KAMIEN, B., TAN, T. Y., FRECKMANN, M.-L., CHONG, B., PHELAN, D., FRANCIS, D., KASSAHN, K. S., HA, T., GAO, S., ARTS, P., JACKSON, M. R., SCOTT, H. S., EGGERS, S., ROWLEY, S., BOGGS, K., RAKONJAC, A., BRETT, G. R., DE SILVA, M. G., SPRINGER, A., WARD, M., STALLARD, K., SIMONS, C., CONWAY, T., HALMAN, A., VAN BERGEN, N. J., SIKORA, T., SEMCESEN, L. N., STROUD, D. A., COMPTON, A. G., THORBURN, D. R., BELL, K. M., SADEDIN, S., NORTH, K. N., CHRISTODOULOU, J. & STARK, Z. 2023. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nature Medicine,* 29**,** 1681-1691.

MA, S., CHOWDHURY, S. K. & ALTON, K. B. 2006. Application of mass spectrometry for metabolite identification. *Curr Drug Metab,* 7**,** 503-23.

MAMOSHINA, P., VIEIRA, A., PUTIN, E. & ZHAVORONKOV, A. 2016. Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics,* 13**,** 1445-1454.

MARTINS-DE-SOUZA, D. J. D. I. C. N. 2022. Proteomics, metabolomics, and protein interactomics in the characterization of the molecular features of major depressive disorder.

MARWAHA, S., KNOWLES, J. W. & ASHLEY, E. A. 2022. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Med,* 14**,** 23.

MATILLA-DUEÑAS, A., CORRAL-JUAN, M., RODRÍGUEZ-PALMERO SEUMA, A., VILAS, D., ISPIERTO, L., MORAIS, S., SEQUEIROS, J., ALONSO, I., VOLPINI, V., SERRANO-MUNUERA, C., PINTOS-MORELL, G., ÁLVAREZ, R. & SÁNCHEZ, I. 2017. Rare Neurodegenerative Diseases: Clinical and Genetic Update. *Adv Exp Med Biol,* 1031**,** 443-496.

MAYEUX, R. 2004. Biomarkers: Potential uses and limitations. *NeuroRX,* 1**,** 182-188.

MCDERMOTT, J. E., WANG, J., MITCHELL, H., WEBB-ROBERTSON, B. J., HAFEN, R., RAMEY, J. & RODLAND, K. D. 2013. Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data. *Expert Opin Med Diagn,* 7**,** 37-51.

MENDRICK, D. L. & TONG, W. 2013. Biomarkers. *In:* DUBITZKY, W., WOLKENHAUER, O., CHO, K.-H. & YOKOTA, H. (eds.) *Encyclopedia of Systems Biology.* New York, NY: Springer New York.

METCALF, S. W. & ORLOFF, K. G. 2004. Biomarkers of exposure in community settings. *J Toxicol Environ Health A,* 67**,** 715-26.

MINNO, A., GELZO, M., CATERINO, M., COSTANZO, M., RUOPPOLO, M. & CASTALDO, G. 2022. Challenges in Metabolomics-Based Tests, Biomarkers Revealed by Metabolomic Analysis and the Promise of the Application of Metabolomics in Precision Medicine. *International Journal of Molecular Sciences,* 23**,** 5213.

MISRA, B. B., LANGEFELD, C., OLIVIER, M. & COX, L. A. J. J. O. M. E. 2019. Integrated omics: tools, advances and future approaches. 62**,** R21-R45.

MITSUHASHI, S. & MATSUMOTO, N. 2020. Long-read sequencing for rare human genetic diseases. *Journal of Human Genetics,* 65**,** 11-19.

MOLE, S. E., ANDERSON, G., BAND, H. A., BERKOVIC, S. F., COOPER, J. D., KLEINE HOLTHAUS, S. M., MCKAY, T. R., MEDINA, D. L., RAHIM, A. A., SCHULZ, A. & SMITH, A. J. 2019. Clinical challenges and future therapeutic approaches for neuronal ceroid lipofuscinosis. *Lancet Neurol,* 18**,** 107-116.

MOLE, S. E. & COTMAN, S. L. 2015. Genetics of the neuronal ceroid lipofuscinoses (Batten disease). *Biochim Biophys Acta,* 1852**,** 2237-41.

MOLE, S. E. & HALTIA, M. 2015. Chapter 70 - The Neuronal Ceroid-Lipofuscinoses (Batten Disease). *In:* ROSENBERG, R. N. & PASCUAL, J. M. (eds.) *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease (Fifth Edition).* Boston: Academic Press.

MOLE, S. E., MICHAUX, G., CODLIN, S., WHEELER, R. B., SHARP, J. D. & CUTLER, D. F. 2004. CLN6, which is associated with a lysosomal storage disease, is an endoplasmic reticulum protein. *Exp Cell Res,* 298**,** 399-406.

MOLE, S. E., SCHULZ, A. & HALTIA, M. 2020. Chapter 4 - The neuronal ceroid-lipofuscinoses (Batten disease). *In:* ROSENBERG, R. N. & PASCUAL, J. M. (eds.) *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease (Sixth Edition).* Academic Press.

MOLE, S. E., WILLIAMS, R. E. & GOEBEL, H. H. 2005. Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. *Neurogenetics,* 6**,** 107-26.

MOLSTER, C., URWIN, D., DI PIETRO, L., FOOKES, M., PETRIE, D., VAN DER LAAN, S. & DAWKINS, H. 2016. Survey of healthcare experiences of Australian adults living with rare diseases. *Orphanet J Rare Dis,* 11**,** 30.

MONTEIRO, M., CARVALHO, M., BASTOS, M. & GUEDES DE PINHO, P. 2012. Metabolomics Analysis for Biomarker Discovery: Advances and Challenges. *Current medicinal chemistry,* 20.

MUNDKUR, B. D. 1949. Evidence Excluding Mutations, Polysomy, and Polyploidy as Possible Causes of Non-Mendelian Segregations in Saccharomyces. *Annals of the Missouri Botanical Garden,* 36**,** 259-280.

MUTHUBHARATHI, B. C., GOWRIPRIYA, T. & BALAMURUGAN, K. 2021. Metabolomics: small molecules that matter more. *Mol Omics,* 17**,** 210-229.

NAFI, O., RAMADAN, B., RIESS, O., BUCHERT, R. & FROUKH, T. 2019. Two cases of variant late infantile ceroid lipofuscinosis in Jordan. *World J Clin Cases,* 7**,** 203-208.

NAVARRETE-OPAZO, A. A., SINGH, M., TISDALE, A., CUTILLO, C. M. & GARRISON, S. R. 2021. Can you hear us now? The impact of health-care utilization by rare disease patients in the United States. *Genetics in Medicine,* 23**,** 2194-2201.

NELVAGAL, H. R., LANGE, J., TAKAHASHI, K., TARCZYLUK-WELLS, M. A. & COOPER, J. D. 2020. Pathomechanisms in the neuronal ceroid lipofuscinoses. *Biochim Biophys Acta Mol Basis Dis,* 1866**,** 165570.

NGUENGANG WAKAP, S., LAMBERT, D. M., OLRY, A., RODWELL, C., GUEYDAN, C., LANNEAU, V., MURPHY, D., LE CAM, Y. & RATH, A. 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics,* 28**,** 165-173.

NICHOLSON, J. K. & LINDON, J. C. 2008. Systems biology: Metabonomics. *Nature,* 455**,** 1054-6.

NICKEL, M. & SCHULZ, A. 2022. Natural History Studies in NCL and Their Expanding Role in Drug Development: Experiences From CLN2 Disease and Relevance for Clinical Trials. *Front Neurol,* 13**,** 785841.

NITA, D. A., MOLE, S. E. & MINASSIAN, B. A. 2016. Neuronal ceroid lipofuscinoses. *Epileptic Disord,* 18**,** 73-88.

NOSKOVÁ, L., STRÁNECKÝ, V., HARTMANNOVÁ, H., PŘISTOUPILOVÁ, A., BAREŠOVÁ, V., IVÁNEK, R., HŮLKOVÁ, H., JAHNOVÁ, H., VAN DER ZEE, J., STAROPOLI, J. F., SIMS, K. B., TYYNELÄ, J., VAN BROECKHOVEN, C., NIJSSEN, P. C., MOLE, S. E., ELLEDER, M. & KMOCH, S. 2011. Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *Am J Hum Genet,* 89**,** 241-52.

O'HAGAN, S., DUNN, W. B., KNOWLES, J. D., BROADHURST, D., WILLIAMS, R., ASHWORTH, J. J., CAMERON, M. & KELL, D. B. 2007. Closed-loop, multiobjective optimization of two-dimensional gas chromatography/mass spectrometry for serum metabolomics. *Anal Chem,* 79**,** 464-76.

ORPHANET. 2024. *Orphanet: The portal for rare diseases and orphan drugs* [Online]. Available: https://www.orpha.net/ [Accessed 11 February 2024].

PANG, Z., CHONG, J., ZHOU, G., DE LIMA MORAIS, D. A., CHANG, L., BARRETTE, M., GAUTHIER, C., JACQUES, P., LI, S. & XIA, J. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res,* 49**,** W388-w396.

PARKER, L. A., CHILET-ROSELL, E., HERNÁNDEZ-AGUADO, I., PASTOR-VALERO, M., GEA, S. & LUMBRERAS, B. 2018. Diagnostic Biomarkers: Are We Moving from Discovery to Clinical Application? *Clinical Chemistry,* 64**,** 1657-1667.

PATTI, G. J., YANES, O. & SIUZDAK, G. 2012a. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol,* 13**,** 263-9.

PATTI, G. J., YANES, O. & SIUZDAK, G. 2012b. Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology,* 13**,** 263-269.

PENCE, H. E. & WILLIAMS, A. 2010. ChemSpider: An Online Chemical Information Resource. *Journal of Chemical Education,* 87**,** 1123-1124.

PEPE, M. S., LI, C. I. & FENG, Z. 2015. Improving the quality of biomarker discovery research: the right samples and enough of them. *Cancer Epidemiol Biomarkers Prev,* 24**,** 944-50.

PERLIS, R. H. 2011. Translating biomarkers to clinical practice. *Mol Psychiatry,* 16**,** 1076-87.

PERRONE, F., CACACE, R., VAN DER ZEE, J. & VAN BROECKHOVEN, C. 2021. Emerging genetic complexity and rare genetic variants in neurodegenerative brain diseases. *Genome Medicine,* 13**,** 59.

PFEUFFER, J., SACHSENBERG, T., ALKA, O., WALZER, M., FILLBRUNN, A., NILSE, L., SCHILLING, O., REINERT, K. & KOHLBACHER, O. 2017. OpenMS - A platform for reproducible analysis of mass spectrometry data. *J Biotechnol,* 261**,** 142-148.

PLAIASU, V., NANU, M. & MATEI, D. 2010. Rare Disease Day - at a glance. *Mædica,* 5**,** 65-6.

PORTER, K. J. B. J. O. E. P. 1957. Effect of homologous bone marrow injections in x-irradiated rabbits. 38**,** 401.

QIU, S., CAI, Y., YAO, H., LIN, C., XIE, Y., TANG, S. & ZHANG, A. 2023. Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduction and Targeted Therapy,* 8**,** 132.

QUINONES, M. P. & KADDURAH-DAOUK, R. J. N. O. D. 2009. Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. 35**,** 165-176.

RACHAKONDA, V., PAN, T. H. & LE, W. D. 2004. Biomarkers of neurodegenerative disorders: How good are they? *Cell Research,* 14**,** 349-358.

RANTA, S., ZHANG, Y., ROSS, B., LONKA, L., TAKKUNEN, E., MESSER, A., SHARP, J., WHEELER, R., KUSUMI, K., MOLE, S., LIU, W., SOARES, M. B., DE FATIMA BONALDO, M., HIRVASNIEMI, A., CHAPELLE, A. D. L., GILLIAM, T. C. & LEHESJOKI, A.-E. 1999. The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nature Genetics,* 23**,** 233-236.

RHO, J. H., BAUMAN, A. J., BOETTGER, H. G. & YEN, T. F. 1973. A search for porphyrin biomarkers in Nonesuch Shale and extraterrestrial samples. *Space Life Sci,* 4**,** 69-77.

RIBBENSTEDT, A., ZIARRUSTA, H. & BENSKIN, J. P. 2018. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLoS One,* 13**,** e0207082.

RIDER, J. A. & RIDER, D. L. 1988. Batten disease: past, present, and future. *Am J Med Genet Suppl,* 5**,** 21-6.

ROSENBERG, J. B., CHEN, A., KAMINSKY, S. M., CRYSTAL, R. G. & SONDHI, D. 2019. Advances in the Treatment of Neuronal Ceroid Lipofuscinosis. *Expert Opin Orphan Drugs,* 7**,** 473-500.

SANTAVUORI, P. 1988. Neuronal ceroid-lipofuscinoses in childhood. *Brain Dev,* 10**,** 80-3.

SAVUKOSKI, M., KESTILÄ, M., WILLIAMS, R., JÄRVELÄ, I., SHARP, J., HARRIS, J., SANTAVUORI, P., GARDINER, M. & PELTONEN, L. 1994. Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. *Am J Hum Genet,* 55**,** 695-701.

SAYERS, E. W., BOLTON, E. E., BRISTER, J R., CANESE, K., CHAN, J., COMEAU, DONALD C., FARRELL, CATHERINE M., FELDGARDEN, M., FINE, A. M., FUNK, K., HATCHER, E., KANNAN, S., KELLY, C., KIM, S., KLIMKE, W., LANDRUM, MELISSA J., LATHROP, S., LU, Z., MADDEN, THOMAS L., MALHEIRO, A., MARCHLER-BAUER, A., MURPHY, TERENCE D., PHAN, L., PUJAR, S., RANGWALA, SANJIDA H., SCHNEIDER, VALERIE A., TSE, T., WANG, J., YE, J., TRAWICK, BARTON W., PRUITT, KIM D. & SHERRY, STEPHEN T. 2022. Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Research*.

SCHNEIDER, P., WALTERS, W. P., PLOWRIGHT, A. T., SIEROKA, N., LISTGARTEN, J., GOODNOW, R. A., FISHER, J., JANSEN, J. M., DUCA, J. S., RUSH, T. S., ZENTGRAF, M., HILL, J. E., KRUTOHOLOW, E., KOHLER, M., BLANEY, J., FUNATSU, K., LUEBKEMANN, C. & SCHNEIDER, G. 2020. Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery,* 19**,** 353-364.

SCHRIMPE-RUTLEDGE, A. C., CODREANU, S. G., SHERROD, S. D. & MCLEAN, J. A. 2016. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom,* 27**,** 1897-1905.

SCHULZ, A. & KOHLSCHÜTTER, A. 2013. NCL Disorders: Frequent Causes of Childhood Dementia. *Iranian journal of child neurology,* 7**,** 1-8.

SCHULZ, A., KOHLSCHUTTER, A., MINK, J., SIMONATI, A. & WILLIAMS, R. 2013. NCL diseases - clinical perspectives. *Biochim Biophys Acta,* 1832**,** 1801-6.

SEYHAN, A. A. 2019. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. *Translational Medicine Communications,* 4**,** 18.

SHEPHERD, C. E., ALVENDIA, H. & HALLIDAY, G. M. 2019. Brain Banking for Research into Neurodegenerative Disorders and Ageing. *Neurosci Bull,* 35**,** 283-288.

SHIN, S.-Y., FAUMAN, E. B., PETERSEN, A.-K., KRUMSIEK, J., SANTOS, R., HUANG, J., ARNOLD, M., ERTE, I., FORGETTA, V., YANG, T.-P., WALTER, K., MENNI, C., CHEN, L., VASQUEZ, L., VALDES, A. M., HYDE, C. L., WANG, V., ZIEMEK, D., ROBERTS, P., XI, L., GRUNDBERG, E., WALDENBERGER, M., RICHARDS, J. B., MOHNEY, R. P., MILBURN, M. V., JOHN, S. L., TRIMMER, J., THEIS, F. J., OVERINGTON, J. P., SUHRE, K., BROSNAN, M. J., GIEGER, C., KASTENMÜLLER, G., SPECTOR, T. D., SORANZO, N. & THE MULTIPLE TISSUE HUMAN EXPRESSION RESOURCE, C. 2014. An atlas of genetic influences on human blood metabolites. *Nature Genetics,* 46**,** 543-550.

SHULAEV, V. 2006. Metabolomics technology and bioinformatics. *Brief Bioinform,* 7**,** 128-39.

SIINTOLA, E., PARTANEN, S., STRÖMME, P., HAAPANEN, A., HALTIA, M., MAEHLEN, J., LEHESJOKI, A.-E. & TYYNELÄ, J. 2006. Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain : a journal of neurology,* 129**,** 1438-1445.

SIINTOLA, E., TOPCU, M., AULA, N., LOHI, H., MINASSIAN, B. A., PATERSON, A. D., LIU, X. Q., WILSON, C., LAHTINEN, U., ANTTONEN, A. K. & LEHESJOKI, A. E. 2007. The novel neuronal ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter. *Am J Hum Genet,* 81**,** 136-46.

SIMONATI, A. & WILLIAMS, R. E. J. F. I. N. 2022. Neuronal ceroid lipofuscinosis: the multifaceted approach to the clinical issues, an overview. 13**,** 87.

SINDELAR, M., DYKE, J. P., DEEB, R. S., SONDHI, D., KAMINSKY, S. M., KOSOFSKY, B. E., BALLON, D. J., CRYSTAL, R. G. & GROSS, S. S. 2018. Untargeted Metabolite Profiling of Cerebrospinal Fluid Uncovers Biomarkers for Severity of Late Infantile Neuronal Ceroid Lipofuscinosis (CLN2, Batten Disease). *Scientific Reports,* 8**,** 15229.

SIRBU, C. A., IVAN, R., AUTHIER, F. J., IONITA-RADU, F., JIANU, D. C., VASILIU, O., CONSTANTIN, C. & TUȚĂ, S. 2023. Orphan Drugs in Neurology-A Narrative Review. *J Pers Med,* 13.

SLEAT, D. E., DONNELLY, R. J., LACKLAND, H., LIU, C. G., SOHAR, I., PULLARKAT, R. K. & LOBEL, P. 1997. Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science,* 277**,** 1802-5.

SLEAT, D. E., GEDVILAITE, E., ZHANG, Y., LOBEL, P. & XING, J. 2016. Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid lipofuscinosis. *Gene,* 593**,** 284-291.

SLEAT, D. E., TANNOUS, A., SOHAR, I., WISEMAN, J. A., ZHENG, H., QIAN, M., ZHAO, C., XIN, W., BARONE, R. & SIMS, K. B. J. J. O. P. R. 2017. Proteomic analysis of brain and cerebrospinal fluid from the three major forms of neuronal ceroid lipofuscinosis reveals potential biomarkers. 16**,** 3787-3804.

SMEDLEY, D., SMITH, K. R., MARTIN, A., THOMAS, E. A., MCDONAGH, E. M., CIPRIANI, V., ELLINGFORD, J. M., ARNO, G., TUCCI, A., VANDROVCOVA, J., CHAN, G., WILLIAMS, H. J., RATNAIKE, T., WEI, W.,

STIRRUPS, K., IBANEZ, K., MOUTSIANAS, L., WIELSCHER, M., NEED, A., BARNES, M. R., VESTITO, L., BUCHANAN, J., WORDSWORTH, S., ASHFORD, S., REHMSTRÖM, K., LI, E., FULLER, G., TWISS, P., SPASIC-BOSKOVIC, O., HALSALL, S., FLOTO, R. A., POOLE, K., WAGNER, A., MEHTA, S. G., GURNELL, M., BURROWS, N., JAMES, R., PENKETT, C., DEWHURST, E., GRÄF, S., MAPETA, R., KASANICKI, M., HAWORTH, A., SAVAGE, H., BABCOCK, M., REESE, M. G., BALE, M., BAPLE, E., BOUSTRED, C., BRITTAIN, H., DE BURCA, A., BLEDA, M., DEVEREAU, A., HALAI, D., HARALDSDOTTIR, E., HYDER, Z., KASPERAVICIUTE, D., PATCH, C., POLYCHRONOPOULOS, D., MATCHAN, A., SULTANA, R., RYTEN, M., TAVARES, A. L. T., TREGIDGO, C., TURNBULL, C., WELLAND, M., WOOD, S., SNOW, C., WILLIAMS, E., LEIGH, S., FOULGER, R. E., DAUGHERTY, L. C., NIBLOCK, O., LEONG, I. U. S., WRIGHT, C. F., DAVIES, J., CRICHTON, C., WELCH, J., WOODS, K., ABULHOUL, L., AURORA, P., BOCKENHAUER, D., BROOMFIELD, A., CLEARY, M. A., LAM, T., DATTANI, M., FOOTITT, E., GANESAN, V., GRUNEWALD, S., COMPEYROT-LACASSAGNE, S., MUNTONI, F., PILKINGTON, C., QUINLIVAN, R., THAPAR, N., WALLIS, C., WEDDERBURN, L. R., WORTH, A., BUESER, T., COMPTON, C., DESHPANDE, C., et al. 2021. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med,* 385**,** 1868-1880.

SMITH, K. R., DAHL, H. H., CANAFOGLIA, L., ANDERMANN, E., DAMIANO, J., MORBIN, M., BRUNI, A. C., GIACCONE, G., COSSETTE, P., SAFTIG, P., GRÖTZINGER, J., SCHWAKE, M., ANDERMANN, F., STAROPOLI, J. F., SIMS, K. B., MOLE, S. E., FRANCESCHETTI, S., ALEXANDER, N. A., COOPER, J. D., CHAPMAN, H. A., CARPENTER, S., BERKOVIC, S. F. & BAHLO, M. 2013. Cathepsin F mutations cause Type B Kufs disease, an adult-onset neuronal ceroid lipofuscinosis. *Hum Mol Genet,* 22**,** 1417-23.

SON, M. Y., KIM, Y. D., SEOL, B., LEE, M. O., NA, H. J., YOO, B., CHANG, J. S. & CHO, Y. S. 2017. Biomarker Discovery by Modeling Behçet's Disease with Patient-Specific Human Induced Pluripotent Stem Cells. *Stem Cells Dev,* 26**,** 133-145.

SPECCHIO, N., PIETRAFUSA, N., TRIVISANO, M. J. T. & MANAGEMENT, C. R. 2020. Changing times for CLN2 disease: the era of enzyme replacement therapy. 213-222.

STARK, Z. & SCOTT, R. H. 2023. Genomic newborn screening for rare diseases. *Nature Reviews Genetics,* 24**,** 755-766.

STAROPOLI, J. F., KARAA, A., LIM, E. T., KIRBY, A., ELBALALESY, N., ROMANSKY, S. G., LEYDIKER, K. B., COPPEL, S. H., BARONE, R., XIN, W., MACDONALD, M. E., ABDENUR, J. E., DALY, M. J., SIMS, K. B. & COTMAN, S. L. 2012. A homozygous mutation in KCTD7 links neuronal ceroid lipofuscinosis to the ubiquitin-proteasome system. *American journal of human genetics,* 91**,** 202-208.

STEINFELD, R., REINHARDT, K., SCHREIBER, K., HILLEBRAND, M., KRAETZNER, R., BRUCK, W., SAFTIG, P. & GARTNER, J. 2006. Cathepsin D deficiency is associated with a human neurodegenerative disorder. *American journal of human genetics,* 78**,** 988-998.

STENGEL, C. 1982. Account of a singular illness among four siblings in the vicinity of Røraas. *Ceroid-lipofuscinosis (Batten's disease).* Elsevier Biomedical Press, Amsterdam.

STRICKER, T., BONNER, R., LISACEK, F. & HOPFGARTNER, G. 2021. Adduct annotation in liquid chromatography/high-resolution mass spectrometry to enhance compound identification. *Analytical and Bioanalytical Chemistry,* 413**,** 503-517.

STURGEON, C., HILL, R., HORTIN, G. L. & THOMPSON, D. 2010. Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory. *Proteomics Clin Appl,* 4**,** 892-903.

SUD, M., FAHY, E., COTTER, D., BROWN, A., DENNIS, E. A., GLASS, C. K., MERRILL, A. H., JR., MURPHY, R. C., RAETZ, C. R., RUSSELL, D. W. & SUBRAMANIAM, S. 2007. LMSD: LIPID MAPS structure database. *Nucleic Acids Res,* 35**,** D527-32.

SUGIMOTO, M., KAWAKAMI, M., ROBERT, M., SOGA, T. & TOMITA, M. 2012. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Curr Bioinform,* 7**,** 96-108.

TAMBUYZER, E., VANDENDRIESSCHE, B., AUSTIN, C. P., BROOKS, P. J., LARSSON, K., MILLER NEEDLEMAN, K. I., VALENTINE, J., DAVIES, K., GROFT, S. C., PRETI, R., OPREA, T. I. & PRUNOTTO, M. 2020.

Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. *Nat Rev Drug Discov,* 19**,** 93-111.
TAUTENHAHN, R., CHO, K., URITBOONTHAI, W., ZHU, Z., PATTI, G. J. & SIUZDAK, G. 2012. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat Biotechnol,* 30**,** 826-8.
TEAM, R. C. 2021. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.
TEIXEIRA, C. A., ESPINOLA, J., HUO, L., KOHLSCHÜTTER, J., PERSAUD SAWIN, D. A., MINASSIAN, B., BESSA, C. J., GUIMARÃES, A., STEPHAN, D. A. & SA MIRANDA, M. C. J. H. M. 2003. Novel mutations in the CLN6 gene causing a variant late infantile neuronal ceroid lipofuscinosis. 21**,** 502-508.
THE FOOD AND DRUG ADMINISTRATION, U. S. 2024. *Rare Diseases at FDA* [Online]. Available: https://www.fda.gov/patients/rare-diseases-fda [Accessed February 11 2024].
TORGRIP, R. J., ÅBERG, M., KARLBERG, B. & JACOBSSON, S. P. J. J. O. C. A. J. O. T. C. S. 2003. Peak alignment using reduced set mapping. 17**,** 573-582.
TORRES MORAL, T., SANCHEZ-NIUBO, A., MONISTROL-MULA, A., GERARDI, C., BANZI, R., GARCIA, P., DEMOTES-MAINARD, J., HARO, J. M. & GROUP, T. P. 2022. Methods for Stratification and Validation Cohorts: A Scoping Review. 12**,** 688.
TRIFONOVA, O. P., MASLOV, D. L., BALASHOVA, E. E. & LOKHOV, P. G. 2019. Evaluation of Dried Blood Spot Sampling for Clinical Metabolomics: Effects of Different Papers and Sample Storage Stability. 9**,** 277.
TRIVISANO, M., FERRETTI, A., CALABRESE, C., PIETRAFUSA, N., PISCITELLO, L., CARFI' PAVIA, G., VIGEVANO, F. & SPECCHIO, N. 2022. Neurophysiological Findings in Neuronal Ceroid Lipofuscinoses. *Front Neurol,* 13**,** 845877.
TRUJILLANO, D., OPREA, G.-E., SCHMITZ, Y., BERTOLI-AVELLA, A. M., ABOU JAMRA, R. & ROLFS, A. 2017. A comprehensive global genotype-phenotype database for rare diseases. *Molecular genetics & genomic medicine* [Online], 5. [Accessed 2017/01//].
TRUSHEIM, M. R., BERNDT, E. R. & DOUGLAS, F. L. 2007. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews Drug Discovery,* 6**,** 287-293.
UHLENBUSCH, N., LÖWE, B. & DEPPING, M. K. 2019. Perceived burden in dealing with different rare diseases: a qualitative focus group study. *BMJ Open,* 9**,** e033353.
VERBERK, I. M. W., LAARHUIS, M. B., VAN DEN BOSCH, K. A., EBENAU, J. L., VAN LEEUWENSTIJN, M., PRINS, N. D., SCHELTENS, P., TEUNISSEN, C. E. & VAN DER FLIER, W. M. 2021. Serum markers glial fibrillary acidic protein and neurofilament light for prognosis and monitoring in cognitively normal older people: a prospective memory clinic-based cohort study. *The Lancet Healthy Longevity,* 2**,** e87-e95.
VESA, J., HELLSTEN, E., VERKRUYSE, L. A., CAMP, L. A., RAPOLA, J., SANTAVUORI, P., HOFMANN, S. L. & PELTONEN, L. 1995. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature,* 376**,** 584-7.
VINAIXA, M., SCHYMANSKI, E. L., NEUMANN, S., NAVARRO, M., SALEK, R. M. & YANES, O. J. T. T. I. A. C. 2016. Mass spectral databases for LC/MS-and GC/MS-based metabolomics: State of the field and future prospects. 78**,** 23-35.
WANG, D., WANG, X., KONG, J., WU, J. & LAI, M. 2016. GC-MS-Based metabolomics discovers a shared serum metabolic characteristic among three types of epileptic seizures. *Epilepsy Res,* 126**,** 83-9.
WANG, Y., ALELYUNAS, Y., SMITH, K., CLELAND, G., MORTISHIRE-SMITH, R., & WRONA, M. D. 2017. Building a collision cross section library of pharmaceutical drugs using the Vion IMS QTof platform: Verification of system performance precision and deviation of CCS measurements. *Waters Corporation. .* Waters Corporation.
WARRIER, V., VIEIRA, M. & MOLE, S. E. 2013. Genetic basis and phenotypic correlations of the neuronal ceroid lipofusinoses. *Biochim Biophys Acta,* 1832**,** 1827-30.

WEI, H., PASMAN, W., RUBINGH, C., WOPEREIS, S., TIENSTRA, M., SCHROEN, J., WANG, M., VERHEIJ, E. & VAN DER GREEF, J. 2012. Urine metabolomics combined with the personalized diagnosis guided by Chinese medicine reveals subtypes of pre-diabetes. *Mol Biosyst,* 8**,** 1482-91.

WHEELER, R. B., SHARP, J. D., SCHULTZ, R. A., JOSLIN, J. M., WILLIAMS, R. E. & MOLE, S. E. 2002. The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (CLN6) and in nclf mutant mice encodes a novel predicted transmembrane protein. *Am J Hum Genet,* 70**,** 537-42.

WILLIAMS, R. E. 2011. 361Appendix 1: NCL Incidence and Prevalence Data. *In:* MOLE, S., WILLIAMS, R. & GOEBEL, H. (eds.) *The Neuronal Ceroid Lipofuscinoses (Batten Disease).* Oxford University Press.

WISHART, D. S. 2007. Human Metabolome Database: completing the 'human parts list'. *Pharmacogenomics,* 8**,** 683-6.

WISHART, D. S., GUO, A., OLER, E., WANG, F., ANJUM, A., PETERS, H., DIZON, R., SAYEEDA, Z., TIAN, S., LEE, B. L., BERJANSKII, M., MAH, R., YAMAMOTO, M., JOVEL, J., TORRES-CALZADA, C., HIEBERT-GIESBRECHT, M., LUI, V. W., VARSHAVI, D., VARSHAVI, D., ALLEN, D., ARNDT, D., KHETARPAL, N., SIVAKUMARAN, A., HARFORD, K., SANFORD, S., YEE, K., CAO, X., BUDINSKI, Z., LIIGAND, J., ZHANG, L., ZHENG, J., MANDAL, R., KARU, N., DAMBROVA, M., SCHIÖTH, H. B., GREINER, R. & GAUTAM, V. 2022. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res,* 50**,** D622-d631.

WITT, S., SCHUETT, K., WIEGAND-GREFE, S., BOETTCHER, J. & QUITMANN, J. 2023. Living with a rare disease - experiences and needs in pediatric patients and their parents. *Orphanet J Rare Dis,* 18**,** 242.

WU, Y. & LI, L. 2016. Sample normalization methods in quantitative metabolomics. *J Chromatogr A,* 1430**,** 80-95.

XIA, J., PSYCHOGIOS, N., YOUNG, N. & WISHART, D. S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res,* 37**,** W652-60.

XIAO, Q., MOORE, S. C., BOCA, S. M., MATTHEWS, C. E., ROTHMAN, N., STOLZENBERG-SOLOMON, R. Z., SINHA, R., CROSS, A. J. & SAMPSON, J. N. 2014. Sources of variability in metabolite measurements from urinary samples. *PLoS One,* 9**,** e95749.

XIAO, Y., BI, M., GUO, H. & LI, M. 2022. Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis. *EBioMedicine,* 79**,** 104001.

XUE, J., GUIJAS, C., BENTON, H. P., WARTH, B. & SIUZDAK, G. 2020. METLIN MS(2) molecular standards database: a broad chemical and biological resource. *Nat Methods,* 17**,** 953-954.

YIN, P., LEHMANN, R. & XU, G. 2015. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal Bioanal Chem,* 407**,** 4879-92.

YIN, P., PETER, A., FRANKEN, H., ZHAO, X., NEUKAMM, S. S., ROSENBAUM, L., LUCIO, M., ZELL, A., HÄRING, H. U., XU, G. & LEHMANN, R. 2013. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin Chem,* 59**,** 833-45.

ZEMAN, W. & DYKEN, P. R. J. P. 1969. Neuronal ceroid-lipofuscinosis (Batten's disease): relationship to amaurotic family idiocy? 44 4**,** 570-83.

ZHANG, A., SUN, H., YAN, G., WANG, P. & WANG, X. 2016. Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed Chromatogr,* 30**,** 7-12.

ZHENG, Y. 2018. Study Design Considerations for Cancer Biomarker Discoveries. *J Appl Lab Med,* 3**,** 282-289.

APPENDIX

Curriculum vitae

Contact information

Name:	Corina-Marcela Rus
Date of birth:	13 December 1985
Place of birth:	Sannicolau Mare, Romania

Education

since 11.2019	**Industrial Ph.D. student**	Rostock University & Centogene GmbH
2008 - 2010	**Student (Master, Biology)**	West University of Timisoara
2004 - 2008	**Student (Bachelor, Biology)**	West University of Timisoara

metabolites

MDPI

Article

Dried Blood Spot (DBS) Methodology Study for Biomarker Discovery in Lysosomal Storage Disease (LSD)

Corina-Marcela Rus [1,2,*], Sebastiano Di Bucchianico [3], Claudia Cozma [1], Ralf Zimmermann [2,3] and Peter Bauer [1]

[1] Centogene GmbH, Am Strande 7, 18055 Rostock, Germany; Claudia.Cozma@centogene.com (C.C.); Peter.Bauer@centogene.com (P.B.)
[2] Institute of Chemistry, University of Rostock, Dr.-Lorenz-Weg 1, 18051 Rostock, Germany; ralf.zimmermann@helmholtz-muenchen.de
[3] Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany; dibucchianico@helmholtz-muenchen.de
* Correspondence: corina-marcela.rus@centogene.com

Abstract: Lysosomal storage diseases (LSDs) are a heterogeneous group of inherited metabolic diseases caused by mutations in genes encoding for proteins involved in the lysosomal degradation of macromolecules. They occur in approximately 1 in 5000 live births and pose a lifelong risk. Therefore, to achieve the maximum benefit from LSDs therapies, a fast and early diagnosis of the disease is required. In this framework, biomarker discovery is a significant factor in disease diagnosis and in predicting its outcomes. On the other hand, the dried blood spot (DBS) based metabolomics platform can open up new pathways for studying non-directional hypothesis approaches to biomarker discovery. This study aims to increase the efficiency of the developed methods for biomarker development in the context of rare diseases, with an improved impact on the reliability of the detected compounds. Thereby, we conducted two independent experiments and integrated them into the screening of the human blood metabolome: (1) comparison of EDTA blood and filter cards in terms of their suitability for metabolomics studies; (2) optimization of the extraction method: a side-by-side comparison of a series of buffers to the best utility to the disease of interest. The findings were compared to previous studies across parameters such as metabolite coverage, sample type suitability, and stability. The results indicate that measurements of metabolites are susceptible to differences in pre-analytical conditions and extraction solvents. This proposed approach can increase the positive rate of the future development of biomarkers. Altogether, the procedure can be easily adapted and applied to other studies, where the limited number of samples is a common barrier.

Keywords: mass spectrometry; metabolomics; biomarkers; lysosomal storage diseases (LSD); dried blood spot (DBS)

Citation: Rus, C.-M.; Di Bucchianico, S.; Cozma, C.; Zimmermann, R.; Bauer, P. Dried Blood Spot (DBS) Methodology Study for Biomarker Discovery in Lysosomal Storage Disease (LSD). *Metabolites* **2021**, *11*, 382. https://doi.org/10.3390/metabo11060382

Academic Editor: Peter Meikle

Received: 30 April 2021
Accepted: 10 June 2021
Published: 13 June 2021

1. Introduction

After five decades of comprehensive research on the sample quality in metabolomics study, the criteria needed for quality sampling and their influence on the research outcome are still not resolved to complete satisfaction [1]. Hundreds of scientific publications provide a set of guidelines to help select the sample types for different studies [2]. Since there is no universal approach, a clear understanding of patient samples' characteristics is essential to select the appropriate sample matrix that derives meaningful findings [3]. Over the past decades, dried blood spot (DBS) technology has become a convenient tool in both qualitative and quantitative laboratory analysis [4]. Its applicability saw a significant expansion in recent decades, with a shift from basic to clinical research and medicine [5]. The advantage of DBS technology over other sampling techniques has been extensively tested and published [6]. There are a number of characteristics that make DBS "easy": it is easy to prepare, easy to receive, easy to use, and easy to store. These key elements made them the method of choice in research [7].

To date, over 6000 distinct types of rare diseases have been described in the literature [8,9], and the number is updated every year with 250–280 new conditions [10]. Of these, lysosomal storage diseases, also referred to herein as LSDs, are a group of more than 70 different inborn metabolic errors with a combined occurrence of around 1 in 5000 live births [11]. This research focused on LSDs due to their monogenic origins, as the majority of metabolic defects are detectable in the metabolomic profile. Novel analytical technologies are needed [12] to advance knowledge and speed up progress towards treatment options for rare diseases [13]. The field of translational metabolomics helps to boost biomarker development [14]. Biomarkers play a pivotal role in preclinical studies [15] as key indicators that allow for the early diagnosis and monitoring of disease [16,17].

The need for disease-specific biomarkers is high [18], and many putative biomarkers are identified in publications every year [19]. Nonetheless, only a few of them have made their transition from bench to bedside [20], and the approval rate remains too low [21]. Successful biomarker discovery requires extensive research, yet the process is slow [18], and some shortages occur at various stages of discovery [22]. There are several reasons for the shortfall in the biomarker pipeline [23], and lack of standardized methodology was cited as the number one purported reason.

Numerous studies have linked the sample selection and extraction protocols as factors that provide the best outcomes [24]. However, there is no consensus on the optimal experimental design, and the approaches are usually correlated with immediate availability and depend on personal aims, triggering deviating results [25,26]. Here, we report the establishment of a proper stratification of the samples used in metabolomics studies to identify the processing methods that bring accurate results in the field of biomarker discovery. This approach helped to minimize the extraneous variable disunity and reduced the errors of analytical experiments to acquire specificity.

2. Results

2.1. Stability Study

The stability study was divided into (1) study evaluating the stability of the metabolites in DBS samples; (2) study testing the stability of the metabolites in DBS cards prepared from fresh and frozen blood.

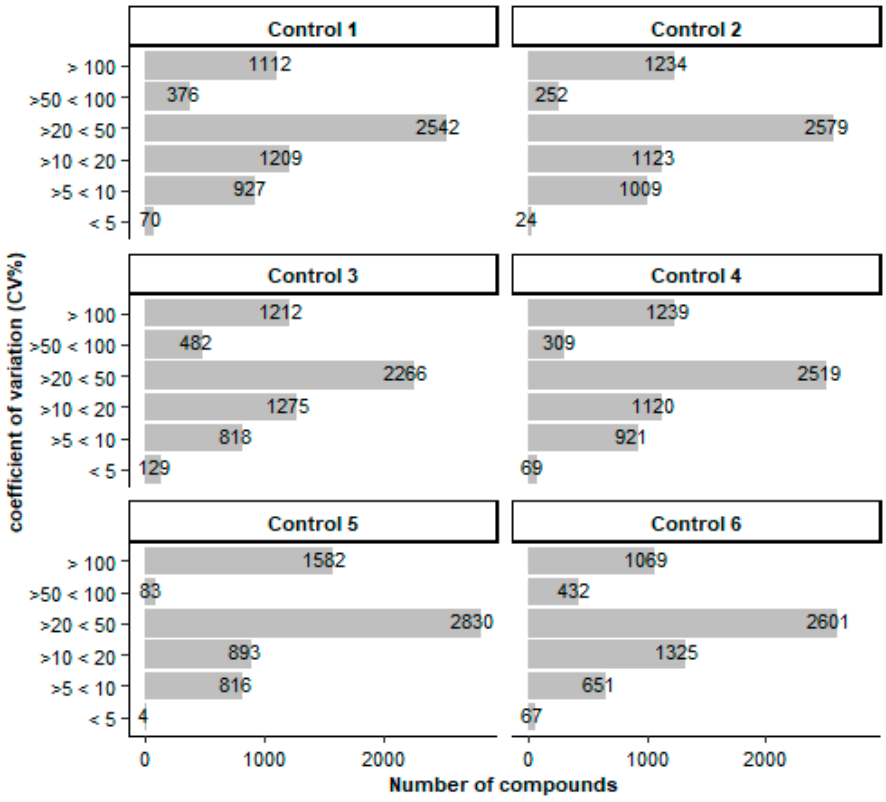
2.1.1. Stability Study 1

Storage at −20 °C over an Extended Period Has a Major Impact on Metabolite Stability

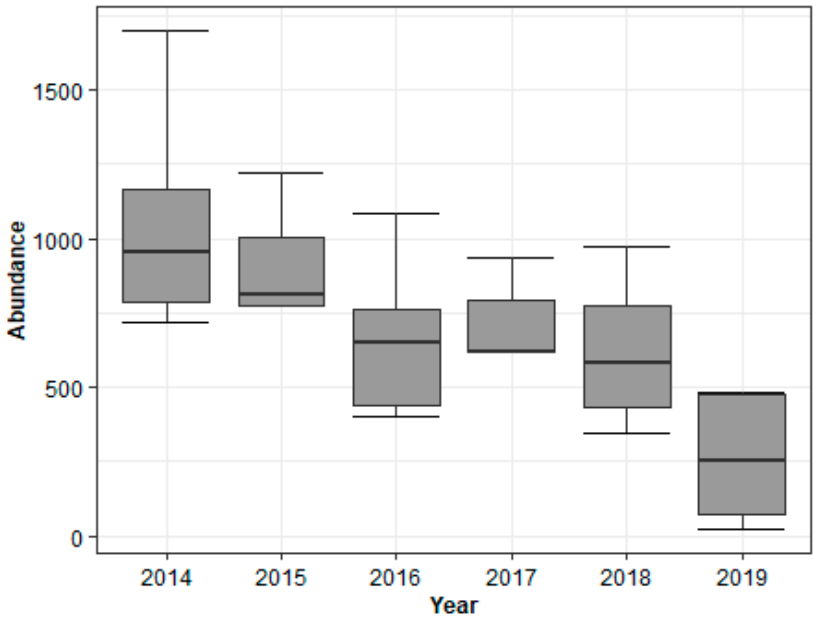
For the samples used in the stability study, the extraction was carried out using the method described by Cozma et al. 2017 [27]. First, metabolite variation was studied using samples collected from the same individuals (n = 6) over six years and stored at −20 °C. Figure 1a shows that a low coefficient of variation (<5%) was present in only 1% of the metabolites, whereas more than 40% of them had a coefficient of variation in the range of 20% to 50%. Additionally, Figure 1b exemplifies the year-to-year variation of the metabolites in DBS stored at −20 °C. These findings indicate that filter cards, even when kept at low temperatures (−20 °C), are unstable from one year to the next.

The Stability of the Metabolites Is Affected by Short-Term Storage in Various Conditions

To validate the prior findings, another test was performed to evaluate the inter-day variation of metabolite yield. This test required the analysis of fresh blood samples stored at RT and −20 °C. Blood was drawn from the same control subjects as in the previous experiment. Following blood collection, duplicate filter cards were prepared. For three days, one duplicate was kept at room temperature (RT), while the other was stored in the freezer (−20 °C). The samples were prepared, extracted, and analyzed in a batch to exclude any deviations. The results showed that three-day storage was sufficient to detect variations in the stability of the analytes when comparing DBS samples stored at RT to those stored at −20 °C (Figure 2).



(**a**)



(**b**)

Figure 1. Investigation of the DBS stability. (**a**) The effect of six years of storage on the stability of DBS cards. Per year, blood was drawn from the same individuals (six controls), dripped onto DBS cards within two hours of collection, and stored at −20 °C. As seen in the figure, the optimal variation (5%) was present in few metabolites, while most of them had variations of up to 50%. (**b**) Box plots show an example of a random compound (7.30_439.3416 *m/z*) in DBS samples from the six control subjects collected from 2014 to 2019 and its variation in abundance throughout the years. Box = 25th and 75th percentiles; bars = min and max values.

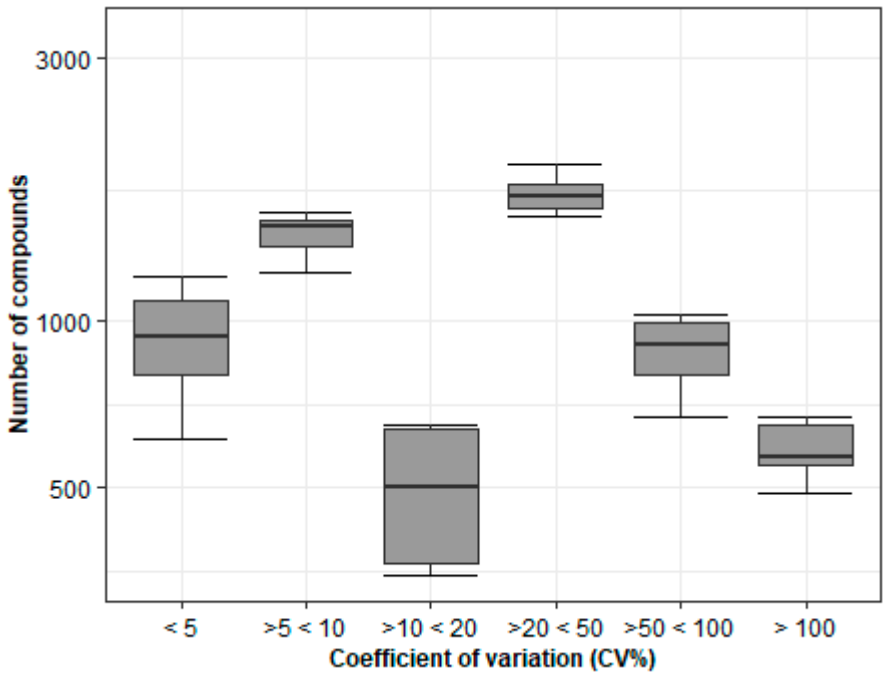


Figure 2. Inter-day influence (three days) of the storage conditions on the human blood metabolome. When DBS samples stored at room temperature were compared to DBS samples maintained at −20 °C, majority of the compounds showed a CV ranging from 20% to 50%.

2.1.2. Stability Study 2

Metabolite Yield Is Influenced by Storage Conditions, Sample Types, and Card Age

The next approach was to compare the DBS samples obtained by spotting fresh EDTA whole blood and frozen blood, after long and short-term storage. Here, the cohort comprised subjects with different types of LSDs, and the patients selected were the ones from whom we received both DBS and EDTA whole blood samples for analysis (Figure 3a). From the EDTA blood, the cards were prepared in duplicates at separate points in time. Therefore, one set of DBS cards was prepared within one day of receiving the blood (fresh whole blood), and the other set was prepared after the storage of the EDTA blood at −20 °C for a long period of time (frozen blood). Figure 3b represents a detailed overview of the sample selection and preparation before the analytical study. The results once again revealed a high fluctuation of metabolites (Figure 4). These findings reiterate the importance of sample storage conditions, sample age, and type regarding their suitability for metabolomics studies. These parameters accounted for the differences in the metabolite yield when the assessed batch was made up of heterogeneous samples.

2.2. Optimization of Extraction Solvent

2.2.1. The Extraction Solvent Methanol: Acetonitrile Produced Metabolites with the Highest Peak Intensity

Four different buffers were tested and compared for the LC-qTOF/MS-based metabolomics analysis. These were as follows: dimethyl sulfoxide: water, methanol: acetonitrile, isopropanol: acetonitrile: water, and ammonium acetate: water (Table 1a). The data were evaluated based on the total number of metabolites and their abundances.

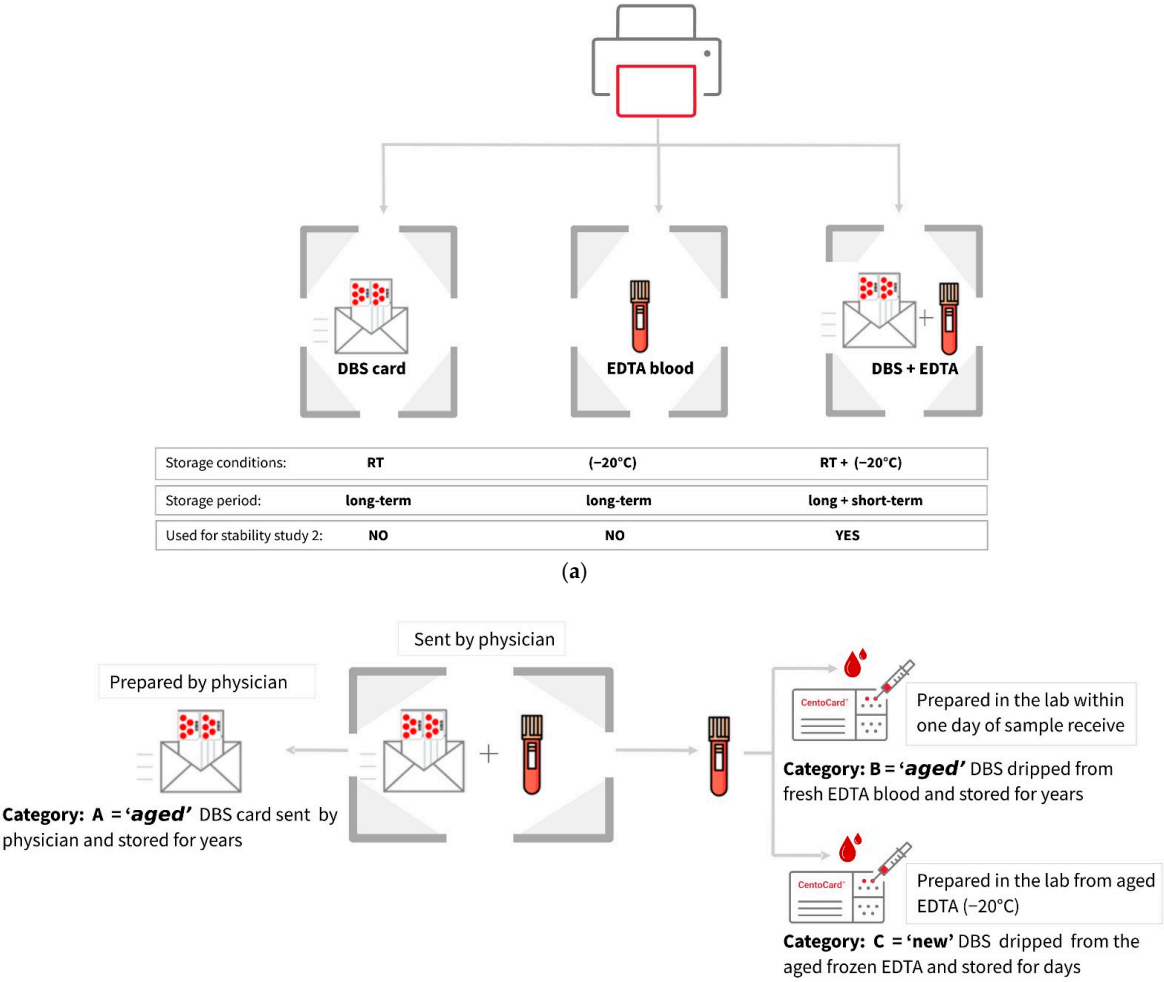
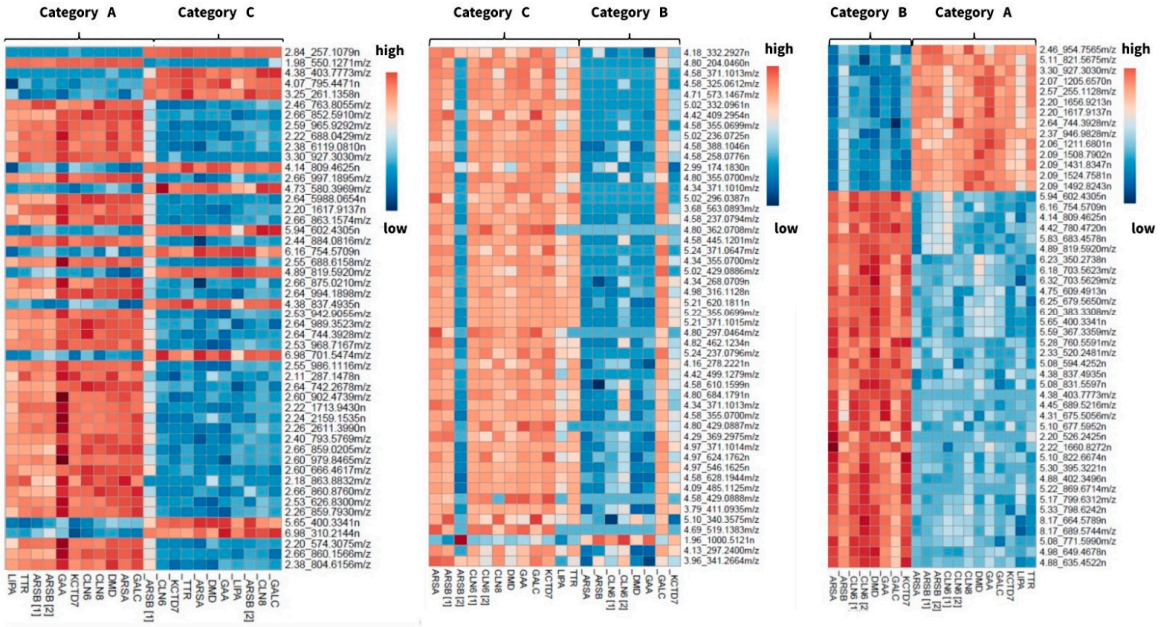


Figure 3. Overview of sample selection, storage and DBS preparation. (**a**) Type of samples received from physician per patient. The samples received from each patient were in the form of either DBS card, or whole EDTA blood, or DBS and EDTA. Stability study 2 comprised only patients with DBS and EDTA samples combined. (**b**) Three different categories of samples were used in stability study 2.

Table 1. Summary of the amount and type of solvent used and its performance on DBS extraction.

(a)					
Extraction	**Ratio *v*/*v***	**Features Detected**	**Metabolites Abundance > 10.000**	**Metabolome Coverage**	**Metabolites Abundance**
dimethyl sulphoxide: water	3:2	9867	16	++++	+
isopropanol: acetonitrile: water	3:3:2	9290	49	+++	++
methanol: acetonitrile	3:1	7759	70	++	+++
ammonium acetate: water	2 mM	5970	95	+	++++
(b)					
Methanol Mixtures	**Ratio *v*/*v***	**Features Detected**	**Metabolites Abundance > 10.000**	**Metabolome Coverage**	**Metabolites Abundance**
methanol: acetonitrile	3:1	7759	70	+	++++
methanol	100%	9120	39	+++	+++
methanol: acetonitrile	1:1	9964	29	++++	++
methanol: water	3:1	7833	25	++	+

(+) indicates the solvents performance ranking from weakest + to strongest ++++.



Experiment 1. Category: A + C Experiment 2. Category: C + B Experiment 3. Category: B + A

Figure 4. The heatmaps show differences in the abundance of identified compounds across various types of samples. Only the 50 most significant compounds were selected. The colors indicate the abundance of the metabolites: brown indicating the higher level and blue the lower level. The darker the square, the more significant the difference. The data were log-transformed and auto-scaled. The individual samples (columns) and compounds (rows) are separated using hierarchical clustering (Ward's algorithm).

The raw abundances detected by untargeted MS were normalized in Progenesis using the default parameters. It is important to note that a filter was applied for the selection of compounds. This means that specific values for *m/z*, retention time, charge, fold change, coefficient of variation, and compound abundance were chosen. Only the compounds that passed the filter were considered trustworthy for further investigations.

The results revealed that dimethyl sulfoxide: water was the most effective extraction buffer in terms of metabolome coverage (Figure 5), whereas the mixture of methanol and acetonitrile (3:1, *v/v*) provided the highest number of metabolites with intensities exceeding one thousand (Figure 6).

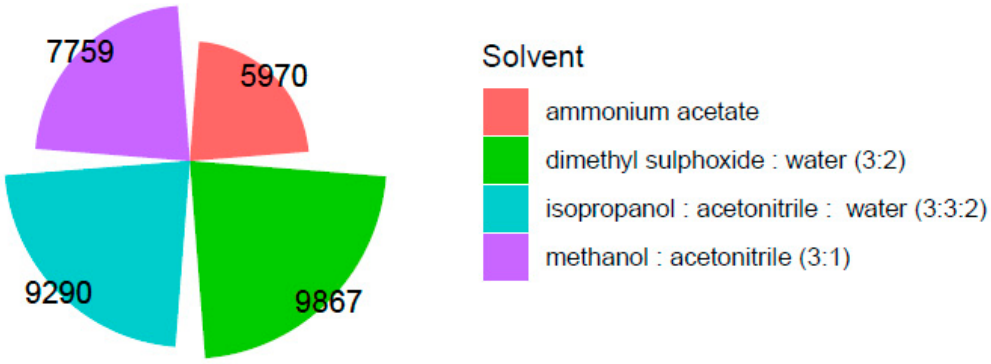


Figure 5. A comparison of the total number of compounds extracted from each of the four extraction solvents.

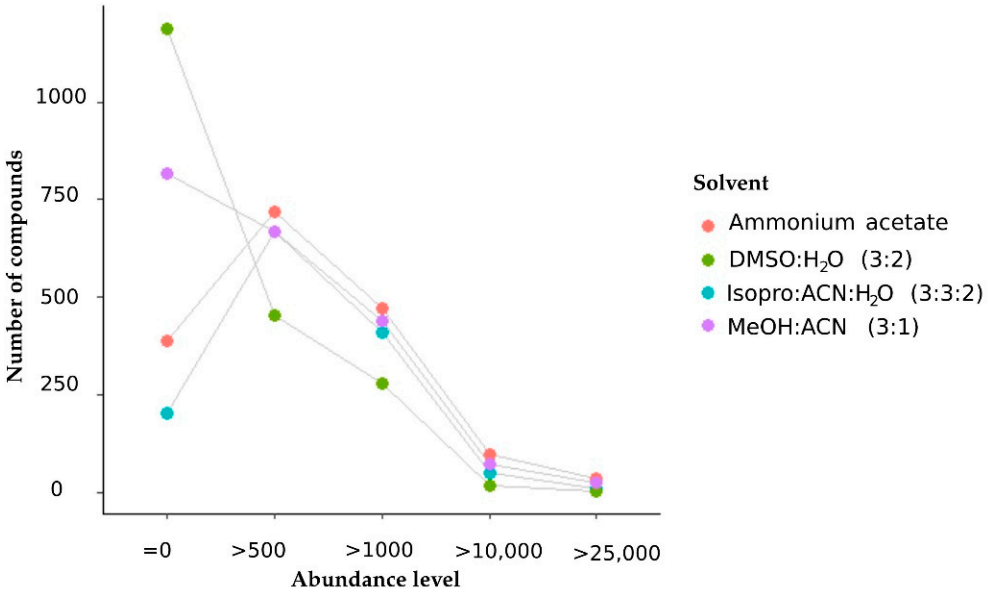


Figure 6. Dotted lines represent the number of features detected and their abundances across the four different extraction solvents. Each dot represents a solvent type.

Extraction with 2 mM ammonium acetate in water (pH = 8) produced metabolites with high concentrations, but the global metabolome coverage was the lowest of the four solvents studied. The isopropanol: acetonitrile: water ranked second in the metabolome coverage, but the number of compounds with high intensity was lower than the other mixtures.

Given the preliminary results, we assume that methanol is a better complement for detecting high-abundance metabolites (Figure 7). Furthermore, this study demonstrated that the extraction solvent for LC-MS-based metabolomics has a visible effect on biomarker discovery projects (Figure 8).

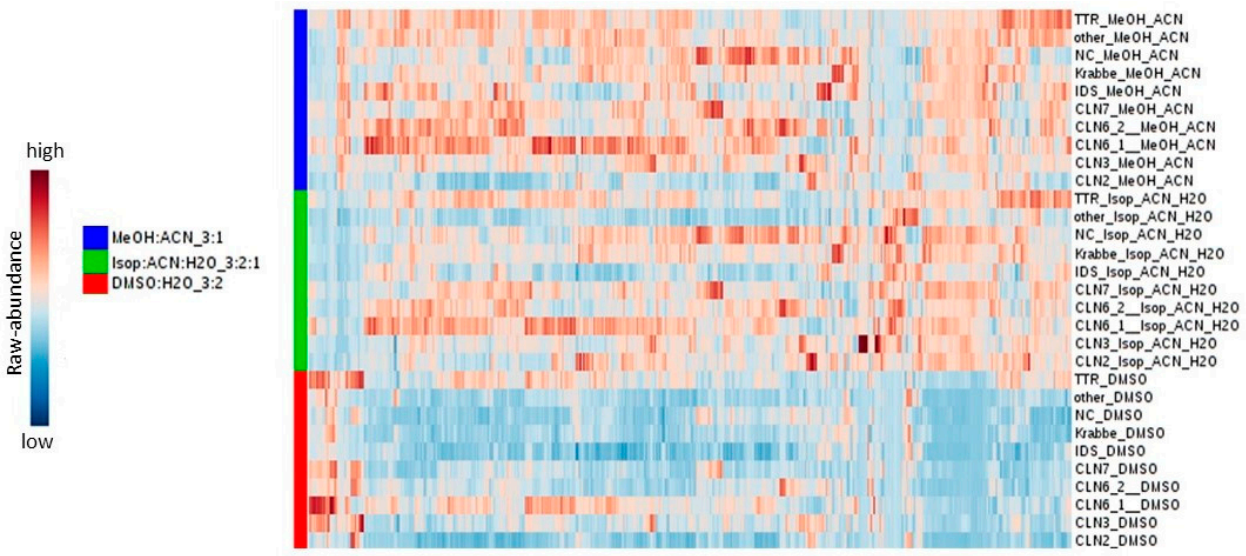


Figure 7. Heatmap showing the reproducibility of the extraction procedure.

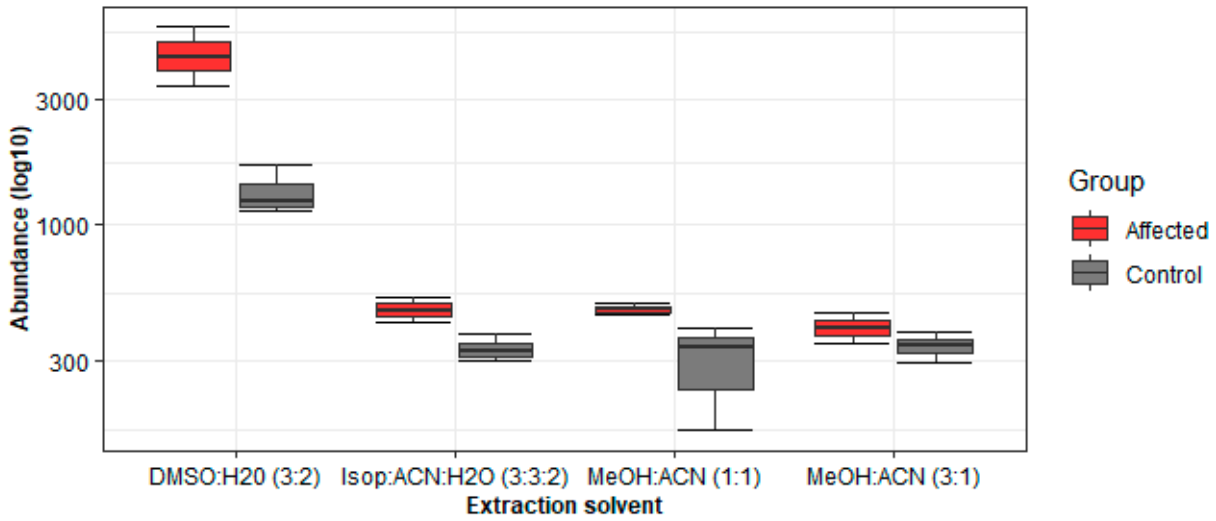


Figure 8. Boxplots showing the distribution of a random metabolite (4.40_421.2811 *m/z*) across different extraction solvents. Control–asymptomatic LSD subjects (*n* = 12); affected–LSD symptomatic patients (*n* = 8). Box = 25th and 75th percentiles; bars = min and max values.

2.2.2. Methanol–Acetonitrile (3:1) Ranks Well in Terms of Metabolome Coverage and Metabolite Intensity

The next step was a comparative study of methanol–acetonitrile (3:1, *v/v*) with other methanol mixtures at various ratios, such as methanol 100%, methanol–water (1:1, *v/v*), and methanol–acetonitrile (1:1, *v/v*). The results were evaluated based on the total number of features and their abundances. Methanol–acetonitrile (3:1, *v/v*) proved to be the most effective mixture for extracting metabolites with high intensity. Table 1a shows the subtle but significant differences among the four different extraction solvents and the comparison of the various methanol mixtures. The considerable overlap between methanol 100% and methanol: acetonitrile (1:1, *v/v*), Table 1b, is noteworthy in terms of metabolome coverage and the number of features with high intensity.

2.2.3. The Application of Learned Principles to the CLN6 Metabolomics Study Helps in the Discovery of Disease-Specific Metabolites

The reliability of the selected buffer, namely methanol–acetonitrile (3:1, *v/v*), was assessed on a full batch of 95 samples performed on the LC-MS QToF Vion. The batch included, among the CLN6 and the control group, a group made of eight types of neuronal ceroid lipofuscinoses diseases (CLN2, CLN3, CLN5, CLN7, CLN8, CLN12, and CLN14), and one that comprised ten types of LSDs (Mucopolysaccharidosis type I, Mucopolysaccharidosis type II, Mucopolysaccharidosis type IIIa, Mucopolysaccharidosis type IV, Spinal muscular atrophy, Fabry disease, Gaucher disease, Krabbe disease, Metachromatic leukodystrophy, and GM1 gangliosidosis disease). Due to the utility of this extraction, we were able to retrieve a larger number of metabolites with high intensity (over 1000), as well as identify compounds that differentiated between diseases and could be used in later stages of the discovery phase (Figure 9).

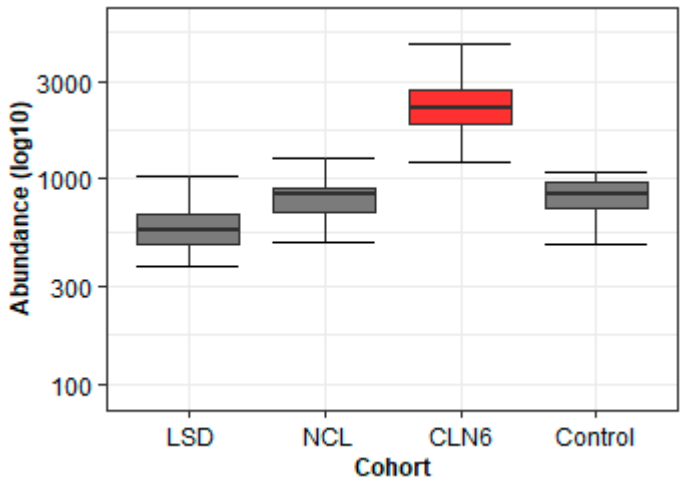


Figure 9. Boxplots showing the "behavior":abundance distribution of a CLN6 (red color) specific biomarker candidate (4.42_384.3240n) in four cohorts (LSD n = 27, NCL n = 19, CLN6 n = 30, control n = 20) using methanol extraction. The circles represent the outliers. Boxes represent the 25th and 75th percentiles, while bars represent minimum and maximum values.

3. Discussion

There are numerous studies conducted on the factors that contribute to the stability of the metabolites in human samples [3]. Still, there is a big gap between the number of stability studies performed on DBS samples and non-DBS samples, such as serum, plasma, urine, and tissue [28].

In the current study, we explored the profile variation of the human metabolome using DBS from two different angles: (1) assessing the stability of metabolites over time; (2) investigating the extraction solvents suitable for metabolomics studies.

It is well known that storage conditions for a given biological material varies depending on the type of sample and the predetermined storage time. Typically, one of three temperatures is considered adequate to store blood samples: room temperature, refrigerated storage (4 °C), or freezer storage (below −20 °C) [29].

Trifonova et al. 2019 [30] examined the stability of several types of single patient-based filter cards stored for four weeks at room temperature. The obtained results showed no significant impact on DBS stability during the four-week storage period. In contrast, the study designed by Drolet et al. 2017 [22], namely a short-term stability study (maximum two weeks) of DBS and urine, showed that DBS cards are unstable at room temperature. According to our results, a limited storage time at −20 °C is crucial in maintaining the reliability of metabolomics studies, while at room temperature even short-term storage can affect the stability of the metabolites. This contradicts the study reported by Prentice et al. 2013 [31], who found that filter cards are stable for weeks at room temperature and up to a year if kept at −20 °C.

Furthermore, we observed that sample types and the card age have an impact on metabolite yield. To ensure the reliability of the data obtained from the metabolomics study, using samples prepared in an identical way and under similar storage conditions is recommended. Several studies reported on the tactics used to improve the extraction efficiencies in metabolomics [32]. Nonetheless, relatively, very few attempted to examine the global metabolome coverage of DBS, as most of them addressed either specific metabolites or other biofluids [33]. Considering the chemical diversity of the human blood metabolome, capturing all its features is challenging [34]. In principle, it is preferable to determine the extraction method that best suits a purpose and optimize it to maximize the number of metabolites. However, it is a daunting task to choose one that applies to such a large scale and fits such a broad spectrum of diseases.

Comparison of fourteen extraction methods on serum samples revealed that methanol was the most effective extraction and provided the highest metabolome coverage [35]. Similar to our results and in agreement with the previous study done by Alshammari et al. 2015 [36], methanol is a better addition to identify high-intensity metabolites. Therefore, our data highlighted the importance of the solvent on untargeted metabolomics and demonstrated that the extraction solvent for LC-MS-based metabolomics has a visible effect on biomarker-focused studies.

Although knowledge of the stability of the analytes in DBS is of crucial importance for biomarker discovery, data available on the assessment of human DBS stability for metabolomics analysis are still scarce. Existing publications on untargeted metabolomics mention several protocols intended to improve the metabolome coverage [37,38]. Yet, few targeted large cohorts of patients and, to our knowledge, none focused on disease-specific metabolites. Moreover, due to the scarce data available on the assessment of DBS applicability to metabolomics studies, the final goal is to incorporate this approach into future studies in biomarker research.

4. Materials and Methods

4.1. Chemicals and Reagents

The solvents used in metabolite extraction, such as methanol, acetonitrile, isopropanol, and formic acid 99% (eluent additive for LC-MS), were all UPLC-MS grades from Biosolve (Dieuze, France). Water LC-MS grade was purchased from VWR (Darmstadt, Germany), ammonium acetate was obtained from Sigma Aldrich (St. Louis, MO, USA), and ethanol 96% from ROTH.

4.2. Mass Spectrometric Analysis

The extract analysis was performed on Waters i-Class ACQUITY UPLC (Waters, Borehamwood, UK) coupled with a Vion IMS-QToF mass spectrometer (Waters, Borehamwood, UK) equipped with an ESI ion source. The chromatographic run was in the positive ionization mode in the mass range of 100–1000 *m/z*. From each extract, 10 µL was injected into a Kinetex EVO (C18, 2.1 × 150 mm, 5 µm) LC column (Phenomenex, Aschaffenburg, Germany) preheated at 50 °C with a flow rate of 0.5 mL/min. The analytes were eluted by using a linear gradient in a range from 1% to 100% B (50 mM formic acid in methanol: acetonitrile vol. 1:1) and A (50 mM formic acid in water). The following parameters were used for mass spectrometric acquisition: high-definition mass spectrometry (HDMSE), capillary voltage 1.2 kV, source temperature 150 °C, desolvation temperature 600 °C, desolvation gas 1000 L/h, cone gas 50 L/h, low collision energy 6 eV, high collision energy ramp 20–40 eV, scan mass 50–1000 *m/z*, scan time 0.5 s.

4.3. Data Acquisition and Analysis

The acquisition was carried out using the Unifi software v1.9 (Waters, Borehamwood, UK), and the results were exported as Unify export packages (.uep). The file was imported into the Progenesis QI software v2.3 (Nonlinear Dynamics, Newcastle upon Tyne, UK) for normalization, metabolites filtering and exported as a .csv file for statistical analysis. The analysis was conducted using MetaboAnalyst tool 4.0 [39], R version 3.6.2 [40], and the figures were produced using the package ggplot2 [41].

4.4. Blood Sample Collection and Preparation

In the current study, the participants were divided into two groups: (i) a control group with no LSD symptoms and DBS cards prepared in-house; (ii) an LSD group with samples (DBS card and/or EDTA blood) shipped by the physician. Following that, the LSDs samples were divided into two cohorts: one set of eighteen types of LSDs, and one of NCL disorder with its nine types of ceroid-lipofuscinosis neuronal diseases, also known as CLNs diseases (Supplementary Table S1). Blood samples derived from the control subjects were initially collected for routine metabolic research and processed within two hours of

blood withdrawal. For the DBS preparation, the EDTA blood tube was gently inverted five times, and a 60 µL aliquot was spotted onto each spot of the CentoCard® (Centogene GmbH, Rostock, Germany). After drying for at least four hours at room temperature, the cards were sealed into plastic bags and stored at −20 °C until further processing. The DBS samples from the LSD-affected patient were prepared by the physician and shipped to us for routine diagnostic analysis. The samples were kept at room temperature after they arrived, pending further analysis. Per analysis, five (3.2 mm diameter) center punches were taken from each card. The spots were cut with the PerkinElmer puncher (PerkinElmer LAS, Rodgau, Germany), which were then collected into a 96-well microtitration plate.

4.5. Dried Blood Spot Extraction

All samples were prepared under the same conditions. The extraction of the metabolites for the stability study has previously been described [30]. Briefly, an extraction mixture of 50 µL DMSO in water (3:2 *v/v*), and 100 µL internal standard 200 ng/mL (lyso-Gb2, Matreya LLC, State College, PA, USA) dissolved in ethanol was added to each well. The aforementioned protocol was modified solely for the extraction–optimization study. Here, the extraction was performed by adding 50 µL of the extraction solution (Table 1) and 100 µL internal standard solution 200 ng/mL dissolved in methanol. After adding the corresponding buffer to the DBS punches, the plate was sonicated for 10 min (Sonoswiss Ultrasonic Cleaner SW12H, Ramsen, Switzerland) and incubated for 60 min at 37 °C with 700 rpm (Heidolph, Schwabach, Germany). Following incubation, the samples were sonicated for 10 min then centrifuged for 5 min at 3500 rpm in a Hermle Z300 plate centrifuge (Hermle Labortehnik, Wehingen, Germany).

4.6. Patients Inclusion

The samples varied in age (children and adults), sex, time of sampling, and storage duration, thereby having considerable variations in factors that could influence the metabolome. All samples analyzed were anonymized, so there is no overall breach of data privacy. The study included 27 controls (13 male and 14 female) ranging in age from 23 to 65 years old, as well as 39 LSD patients (18 male and 21 female) ranging in age from 2 to 65 years old (Supplementary Table S2(a–c)). They were divided as follows: (1) stability study, (2) extraction study, and (3) cumulative study (Table 2).

Table 2. Distribution of the individuals by experiments.

	(1) Stability Study			(2) Extraction Study		(3) Cumulative Study		
Subjects	**Control**	**CLN6**	**LSD**	**Control**	**LSD**	**Control**	**NCL**	**LSD**
male/female	3/3	2/8	6/5	10/11	4/7	10/11	19/30	10/15
age (mean ± SD)	36.5	6.5	15.8	32	6	33	6	6.5

5. Conclusions

The sample grouping procedure was done based on three metrics: sample types, sample age, and sample storage conditions. These key metrics are not commonly used in the existing biomarker-focused metabolomic studies. According to our findings, differences in these parameters have a considerable impact on the stability of the metabolites. In addition, the use of different solvents for the extraction has shown variable results on metabolite intensity and abundance scale. We conclude that the reliability of metabolites is higher if the following conditions are met: similar sample types, similar sample age, and similar storage conditions are obtained; methanol is used as the extraction solvent.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/metabo11060382/s1. Table S1: Overview of lysosomal storage diseases (LSDs); Table S2(a–c): List of the total number of patients and controls used in the study.

Author Contributions: C.-M.R. organized and performed the experimental work, analyzed the data, and wrote the paper; C.C. contributed to the concept of the study; S.D.B. and R.Z. reviewed and commented on the manuscript; P.B. supervised the research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study because it included samples sent to Centogene GmbH (Germany) for diagnostic purposes.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. We used anonymized samples with written consent for both diagnosis and research, given by all participants or their legal guardians. There is no identifying information about the patients included in the study. The data that supports the reported samples is not publicly available because it contains the personal information of the research participants.

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

References

1. Faber, J.; Fonseca, L.M. How sample size influences research outcomes. *Dent. Press J. Orthod.* **2014**, *19*, 27–29. [CrossRef] [PubMed]
2. Broadhurst, D.; Goodacre, R.; Reinke, S.N.; Kuligowski, J.; Wilson, I.D.; Lewis, M.R.; Dunn, W.B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018**, *14*, 72. [CrossRef]
3. Stevens, V.L.; Hoover, E.; Wang, Y.; Zanetti, K.A. Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review. *Metabolites* **2019**, *9*, 156. [CrossRef]
4. Balashova, E.; Trifonova, O.P.; Maslov, D.L.; Lokhov, P.G. Application of dried blood spot for analysis of low molecular weight fraction (metabolome) of blood. *Health Prim. Care* **2018**, *2*, 1–11. [CrossRef]
5. Zakaria, R.; Allen, K.J.; Koplin, J.; Roche, P.; Greaves, R.F. Advantages and Challenges of Dried Blood Spot Analysis by Mass Spectrometry across the Total Testing Process. *EJIFCC* **2016**, *27*, 288–317.
6. Lim, M.D. Dried Blood Spots for Global Health Diagnostics and Surveillance: Opportunities and Challenges. *Am. J. Trop. Med. Hyg.* **2018**, *99*, 256–265. [CrossRef] [PubMed]
7. Freeman, J.D.; Rosman, L.M.; Ratcliff, J.; Strickland, P.T.; Graham, D.R.; Silbergeld, E.K. State of the Science in Dried Blood Spots. *Clin. Chem.* **2018**, *64*, 656–679. [CrossRef]
8. ORPHANET. *Prevalence and Incidence of Rare Diseases: Bibliographic—Prevalence, Incidence or Number of Published Cases Listed by Diseases (in Alphabetical Order)*. Orphanet Reports Series N1. January 2021. Available online: https://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf (accessed on 20 January 2021).
9. Wakap, S.N.; Lambert, D.M.; Olry, A.; Rodwell, C.; Gueydan, C.; Lanneau, V.; Murphy, D.; Le Cam, Y.; Rath, A. Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database. *Eur. J. Hum. Genet.* **2020**, *28*, 165–173. [CrossRef]
10. Dawkins, H.J.; Draghia-Akli, R.; Lasko, P.; Lau, L.P.; Jonker, A.H.; Cutillo, C.M.; Rath, A.; Boycott, K.M.; Baynam, G.; Lochmüller, H.; et al. Progress in Rare Diseases Research 2010–2016: An IRDiRC Perspective. *Clin. Transl. Sci.* **2018**, *11*, 11–20. [CrossRef] [PubMed]
11. Fuller, M.; Meikle, P.J.; Hopwood, J.J. *Epidemiology of Lysosomal Storage Diseases: An Overview*; Oxford PharmaGenesis: Oxford, UK, 2006.
12. Gagne, J.J.; Thompson, L.; O'Keefe, K.; Kesselheim, A.S. Innovative research methods for studying treatments for rare diseases: Methodological review. *BMJ* **2014**, *349*, g6802. [CrossRef]
13. Southall, N.T.; Natarajan, M.; Lau, L.P.L.; Jonker, A.H.; Deprez, B.; Guilliams, T.; Hunter, L.; Rademaker, C.M.; Hivert, V.; Ardigò, D.; et al. The use or generation of biomedical data and existing medicines to discover and establish new treatments for patients with rare diseases—Recommendations of the IRDiRC Data Mining and Repurposing Task Force. *Orphanet. J. Rare Dis.* **2019**, *14*, 1–8. [CrossRef] [PubMed]
14. Pinu, F.R.; Goldansaz, S.A.; Jaine, J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites* **2019**, *9*, 108. [CrossRef]
15. Kohler, I.; Hankemeier, T.; van der Graaf, P.H.; Knibbe, C.A.; van Hasselt, J.C. Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. *Eur. J. Pharm. Sci.* **2017**, *109*, S15–S21. [CrossRef]
16. Gülbakan, B.; Özgül, R.K.; Yüzbaşıoğlu, A.; Kohl, M.; Deigner, H.-P.; Özgüç, M. Discovery of biomarkers in rare diseases: Innovative approaches by predictive and personalized medicine. *EPMA J.* **2016**, *7*, 1–6. [CrossRef] [PubMed]
17. Wang, X.; A Ward, P. Opportunities and challenges of disease biomarkers: A new section in the journal of translational medicine. *J. Transl. Med.* **2012**, *10*, 240. [CrossRef]

18. Considine, E.C. The Search for Clinically Useful Biomarkers of Complex Disease: A Data Analysis Perspective. *Metabolites* **2019**, *9*, 126. [CrossRef]
19. Paulovich, A.G.; Whiteaker, J.R.; Hoofnagle, A.N.; Wang, P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteom. Clin. Appl.* **2008**, *2*, 1386–1402. [CrossRef]
20. Marchand, C.R.; Farshidfar, F.; Rattner, J.; Bathe, O.F. A Framework for Development of Useful Metabolomic Biomarkers and Their Effective Knowledge Translation. *Metabolites* **2018**, *8*, 59. [CrossRef]
21. Zhang, Z.; Chan, D.W. The Road from Discovery to Clinical Diagnostics: Lessons Learned from the First FDA-Cleared In Vitro Diagnostic Multivariate Index Assay of Proteomic Biomarkers. *Cancer Epidemiol. Biomark. Prev.* **2010**, *19*, 2995–2999. [CrossRef]
22. Drolet, J.; Tolstikov, V.; Williams, B.A.; Greenwood, B.P.; Hill, C.; Vishnudas, V.K.; Sarangarajan, R.; Narain, N.R.; Kiebish, M.A. Integrated Metabolomics Assessment of Human Dried Blood Spots and Urine Strips. *Metabolites* **2017**, *7*, 35. [CrossRef] [PubMed]
23. Ioannidis, J.P.A.; Bossuyt, P.M.M. Waste, Leaks, and Failures in the Biomarker Pipeline. *Clin. Chem.* **2017**, *63*, 963–972. [CrossRef] [PubMed]
24. Gika, H.; Theodoridis, G. Sample preparation prior to the LC–MS-based metabolomics/metabonomics of blood-derived samples. *Bioanalysis* **2011**, *3*, 1647–1661. [CrossRef]
25. Barnes, S.; Benton, H.P.; Casazza, K.; Cooper, S.; Cui, X.; Du, X.; Engler, J.; Kabarowski, J.H.; Li, S.; Pathmasiri, W.; et al. Training in metabolomics research. I. Designing the experiment, collecting and extracting samples and generating metabolomics data. *J. Mass Spectrom.* **2016**, *51*, 461–475. [CrossRef]
26. Surowiec, I.; Johansson, E.; Torell, F.; Idborg, H.; Gunnarsson, I.; Svenungsson, E.; Jakobsson, P.-J.; Trygg, J. Multivariate strategy for the sample selection and integration of multi-batch data in metabolomics. *Metabolomics* **2017**, *13*, 114. [CrossRef]
27. Cozma, C.; Iurașcu, M.-I.; Eichler, S.; Hovakimyan, M.; Brandau, O.; Zielke, S.; Böttcher, T.; Giese, A.-K.; Lukas, J.; Rolfs, A. C26-Ceramide as highly sensitive biomarker for the diagnosis of Farber Disease. *Sci. Rep.* **2017**, *7*, 6149. [CrossRef]
28. Wu, Z.E.; Kruger, M.C.; Cooper, G.J.; Poppitt, S.D.; Fraser, K. Tissue-Specific Sample Dilution: An Important Parameter to Optimise Prior to Untargeted LC-MS Metabolomics. *Metabolites* **2019**, *9*, 124. [CrossRef]
29. Wu, D.-W.; Li, Y.-M.; Wang, F. How Long Can We Store Blood Samples: A Systematic Review and Meta-Analysis. *EBioMedicine* **2017**, *24*, 277–285. [CrossRef]
30. Trifonova, O.P.; Maslov, D.L.; Balashova, E.E.; Lokhov, P.G. Evaluation of Dried Blood Spot Sampling for Clinical Metabolomics: Effects of Different Papers and Sample Storage Stability. *Metabolites* **2019**, *9*, 277. [CrossRef] [PubMed]
31. Prentice, P.; Turner, C.; Wong, M.C.; Dalton, R.N. Stability of metabolites in dried blood spots stored at different temperatures over a 2-year period. *Bioanalysis* **2013**, *5*, 1507–1514. [CrossRef] [PubMed]
32. Vuckovic, D. Improving metabolome coverage and data quality: Advancing metabolomics and lipidomics for biomarker discovery. *Chem. Commun.* **2018**, *54*, 6728–6749. [CrossRef] [PubMed]
33. Sitnikov, D.G.; Monnin, C.S.; Vuckovic, D. Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS. *Sci. Rep.* **2016**, *6*, 38885. [CrossRef]
34. Fiehn, O.; Kind, T. Metabolite Profiling in Blood Plasma. In *Metabolomics: Methods and Protocols*; Weckwerth, W., Ed.; Humana Press: Totowa, NJ, USA, 2007; pp. 3–17. [CrossRef]
35. Want, E.J.; O'Maille, G.; Smith, C.A.; Brandon, T.R.; Uritboonthai, W.; Qin, C.; Trauger, A.S.A.; Siuzdak, G. Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Anal. Chem.* **2006**, *78*, 743–752. [CrossRef] [PubMed]
36. Alshammari, T.M.; Al-Hassan, A.A.; Ben Hadda, T.; Aljofan, M. Comparison of different serum sample extraction methods and their suitability for mass spectrometry analysis. *Saudi Pharm. J.* **2015**, *23*, 689–697. [CrossRef] [PubMed]
37. Yang, Y.; Yin, Y.; Chen, X.; Chen, C.; Xia, Y.; Qi, H.; Baker, P.N.; Zhang, H.; Han, T.-L. Evaluating different extraction solvents for GC-MS based metabolomic analysis of the fecal metabolome of adult and baby giant pandas. *Sci. Rep.* **2019**, *9*, 12017–12019. [CrossRef] [PubMed]
38. Lindahl, A.; Sääf, S.; Lehtiö, J.; Nordström, A. Tuning Metabolome Coverage in Reversed Phase LC–MS Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition. *Anal. Chem.* **2017**, *89*, 7356–7364. [CrossRef] [PubMed]
39. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **2018**, *46*, W486–W494. [CrossRef]
40. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
41. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, UY, USA, 2016.

Orphanet Journal of Rare Diseases

RESEARCH

Open Access

Clinical and genetic characterization of a cohort of 97 CLN6 patients tested at a single center

Corina-Marcela Rus[1,2*], Thomas Weissensteiner[1], Catarina Pereira[1], Iuliana Susnea[1], Bright D. Danquah[1], Galina Morales Torres[1], Maria Eugenia Rocha[1], Claudia Cozma[1], Deepa Saravanakumar[1], Sumanth Mannepalli[1], Krishna K. Kandaswamy[1], Sebastiano Di Bucchianico[3], Ralf Zimmermann[2,3], Arndt Rolfs[1,4], Peter Bauer[1,5] and Christian Beetz[1*]

Abstract

Background: Ceroid lipofuscinoses neuronal 6 (CLN6) disease belongs to the neuronal ceroid lipofuscinoses (NCLs), complex and genetically heterogeneous disorders with wide geographical and phenotypic variation. The first clinical signs usually appear between 18 months and 8 years, but examples of later-onset have also been reported. Common manifestations include ataxia, seizures, vision impairment, and developmental regression. Because these are shared by other neurological diseases, identification of *CLN6* genetic variants is imperative for early diagnosis.

Results: We present one of the largest cohorts to date of genetically diagnosed CLN6 patients screened at a single center. In total 97 subjects, originating from 20 countries were screened between 2010 and 2020. They comprised 86 late-infantile, eight juvenile, and three adult-onset cases (two patients with Kufs disease type A, and one with teenage progressive myoclonic epilepsy). The male to female ratio was 1.06: 1.00. The age at referral was between six months and 33 years. The time from disease onset to referral ranged from less than 1 month to 8.3 years. The clinical phenotype consisted of a combination of symptoms, as reported before. We characterized a total of 45 distinct variants defining 45 distinct genotypes. Twenty-four were novel variants, some with distinct geographic associations. Remarkably, c.257A > G (p.H86R) was present in five out of 23 unrelated Egyptian individuals but in no patients from other countries. The most common genotype was homozygosity for the c.794_796del in-frame deletion. It was present in about one-third of CLN6 patients (28 unrelated cases, and 2 familial cases), all with late-infantile onset. Variants with a high likelihood of causing loss of CLN6 function were found in 21% of cases and made up 33% of all distinct variants. Forty-four percent of variants were classified as pathogenic or likely pathogenic.

Conclusions: Our study significantly expands the number of published clinical cases and the mutational spectrum of disease-associated *CLN6* variants, especially for the Middle Eastern and North African regions. We confirm previous observations regarding the most prevalent symptoms and recommend including *CLN6* in the genetic diagnosis of patients presenting with early-onset abnormalities of the nervous system, musculoskeletal system, and eye.

Keywords: Rare disease, Lysosomal storage disorder, Batten disease, Neuronal ceroid lipofuscinoses, *CLN6*, New variant, Genotype, Phenotype

*Correspondence: corina-marcela.rus@uni-rostock.de; christian.beetz@centogene.com
[1] Centogene GmbH, Am Strande 7, 18057 Rostock, Germany
Full list of author information is available at the end of the article

Background

Neuronal ceroid lipofuscinoses (NCLs) are a genetically heterogeneous group of inherited lysosomal storage disorders [1, 2]. Together, they constitute the most prevalent class of rare childhood-onset neurodegenerative diseases [3–5]. The estimated total incidence of all NCLs ranges from 0.01 to 9 per 100 000 live births [6–8] but varies between countries and geographical regions [5]. NCLs are characterized by an accumulation of intracellular auto-fluorescent storage material (ceroid) and neurodegeneration [9]. The clinical spectrum consists of a combination of symptoms including intellectual and motor deterioration, visual impairment, seizures, psycho-motor decline, and loss of neurons [2]. The order in which these symptoms appear differs between the disease subtypes, but the outcome is always fatal. NCLs have been subclassified according to age at onset and clinical features into congenital (CNCL), infantile (INCL), late-infantile (LINCL), juvenile (JNCL), and adult (ANCL) neuronal ceroid lipofuscinoses. Thirteen genes with NCL associated variants have been established to date, named *CLN1-CLN8* and *CLN10-CLN14* [10, 11]. Additionally, a new subtype of NCL (CLN15) has been proposed [12]. The general pattern of inheritance is autosomal recessive, except for ANCL which can be either autosomal recessive or dominant [13]. Genetic heterogeneity and overlapping clinical features make the diagnosis of NCL disease challenging.

Ceroid lipofuscinosis neuronal protein 6, encoded by the *CLN6* gene, forms complexes with other proteins which act as key-regulators of vesicular sorting and trafficking. Defects can have a variety of consequences, from diminished lysosomal function to impaired neurotransmitter secretion and neurite outgrowth [14–16]. The worldwide incidence of CLN6 disease is currently not accurately known. The classical clinical subtypes are late-infantile and juvenile (OMIM# 601780). In addition, atypical phenotypes such as Kufs disease type A (OMIM# 204300), with or without teenage progressive myoclonic epilepsy, have been reported [17]. The major clinical manifestations of classical-onset CLN6 are similar to that of other NCLs and premature death typically occurs between five and 12 years of age [18]. Recent research has improved knowledge of the pathogenic mechanisms but therapy to delay disease progression exists so far only for patients with defects in *CLN2* [19]. However, *CLN6* gene therapy has shown encouraging results in mice and primates [20], and is currently being trialed in humans (NCT02725580) [20–22].

Here, we present an analysis of a large cohort of patients (n = 97) who were referred to us for molecular genetic testing and diagnosed with disease-associated *CLN6* variants (Fig. 1). Our aims were a better

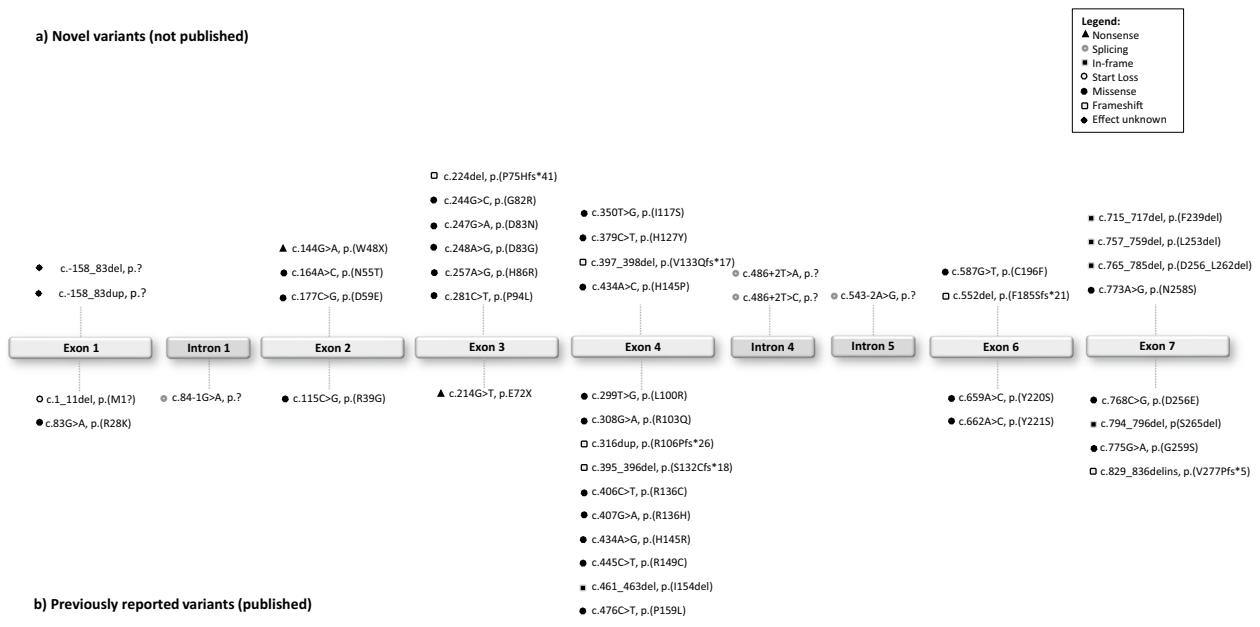


Fig. 1 Schematic representation of the variant type and exon–intron distribution of the variants in CLN6. Each symbol represents the corresponding variant type (see legend). Novel variants identified in this study are shown above the grey boxes (**a**) and previously reported variants from this study are presented below the grey boxes (**b**). The majority of variants are missense variants. Most of the novel reported variants are located in exons 3, 4, and 7

understanding of the diversity of clinical symptoms and the characterization of *CLN6* variants, especially in geographical regions which were underrepresented in public datasets.

Results

Demographic and clinical characteristics of the cohort

At the time of writing, CentoMD® 5.8 [23] stored curated data for 97 individuals genetically diagnosed with CLN6 disease. These included 85 (87%) unrelated individuals and 12 (13%) familial cases. The familial cases were six sibling pairs with late-infantile disease and variable age of referral. Three sibling pairs were from Egypt (North Africa), two from Lebanon, and one from Saudi Arabia (Middle East). The following analysis was performed without accounting for family membership, except when noted and implicitly when calculating the frequency of distinct alleles in the cohort.

Patients originated from 20 countries, grouped into six geographical regions. Most cases were from the Middle East (59%) and North Africa (32%). Within these regions, respectively, Saudi Arabia (22%) and Egypt (27%) were the countries that contributed the greatest number of cases in the total cohort (Table 1). The ratio of males (n = 47) to females (n = 44) was nearly equal. According to categorical disease subtype information, 86 (89%) of patients had late-infantile, 8 (8%) juvenile, and 3 (3%) adult-onset disease [24]. The adult-onset cases comprised two Kufs disease type A, and one teenage progressive myoclonic epilepsy.

In addition, the ages of onset and at referral for genetic testing were provided for a subset of patients (Fig. 2A). The median age of onset was 3.8 years, with a range from less than a month to 15 years and an interquartile range (IQR) of 3.0–5.0 years (n = 34). The median age at referral was 6.1 years (range 7–33 years, IQR: 5.3–8.7 years, n = 88). The median time from disease onset to referral was 2.7 years and ranged from less than 1 month–8.3 years (IQR 2.0–3.6 years, n = 32) (Fig. 2B).

Clinical symptoms were provided for 86 patients and catalogued into 213 Human Phenotype Ontology (HPO) terms. The most frequent were "Developmental regression" (n = 46, 53%), "Seizure" (n = 37, 43%), "Ataxia" (n = 28, 33%), and "Intellectual disability" (n = 26, 30%). Figure 3 shows the 31 terms that were used in five or more patients. Alternatively, grouping HPO terms at the "Phenotypic abnormality" level provided a low-resolution overview that revealed that after nervous and musculoskeletal systems, the eye was indeed the third most affected organ in CLN6 patients. The symptoms were diverse and included various degrees of visual impairment in 12 patients, macular and other retinal abnormalities in eight and abnormalities of eye movement in five.

Age of onset and time to referral

The ages at which patients were referred for genetic testing differed according to disease onset type as would be expected (Fig. 4A). The median referral age for patients with late-infantile.

CLN6 disease was 5.8 years, but with a wide range of from six months to 17.2 years, and IQR of 5.2–7.8 years. The median age at referral for juvenile-onset patients was 9.8 years (range 7.8–18.9 years, IQR 8.9–16.9 years). The age of referral for the three adult-onset patients was 22.5, 31.8, and 33.2 years. A weak correlation existed between the ages of onset and referral (Pearson $R^2 = 0.23$, $p < 0.01$) when disregarding the adult-onset outlier (Fig. 4B).

Known and novel *CLN6* transcript variants discovered in this study

Sequencing of the cohort identified 45 distinct *CLN6* transcript variants of which 24 (53%) have not been described previously (Fig. 5A). Frameshift, splicing, nonsense, start loss and gross deletions or duplications are genetic alterations with a high likelihood of causing complete loss of function (LoF) of a gene. This type accounted for 15 (33%) distinct variants, the remaining were 25 missense substitutions and five in-frame deletions (Fig. 5B). Among the novel mutations, 9 were LoF, 12 missense, and three in-frame. According to clinical significance, seven variants were classified as pathogenic (P), 12 as likely pathogenic (LP), and 26 as a variant of uncertain significance (VUS) (Fig. 5C). Novel variants comprised one P, seven LP, and 16 VUS.

By far the most prevalent variant in our cohort was the in-frame deletion c.794_796del (p.265Sdel) which was present in 30 out of 97 patients (Fig. 6, further details in Table 1). The most frequent new variant, c.257A > G (p.H86R), was identified in five unrelated cases. Thirty-three variants were represented by a single case only.

Association of *CLN6* variant classes with disease subtype and age of onset in homozygous cases

There were only missense cases in the adult-onset and 87% of juvenile cases. (Fig. 7A). Moreover, the age of onset in patients homozygous for a missense variant displayed a wider range and slightly higher median than the ages of onset in the other two variant categories (Fig. 7B).

Association of individual *CLN6* variants with disease subtype and age of onset

A detailed overview of the variants and the associated patient characteristics is given in Table 1. The most prevalent variant in the cohort, the in-frame deletion c.794_796del (p.S265del), was exclusively found in 30 (36%) (26 unrelated and 2 familial cases) homozygous

Table 1 *CLN6* gene variants and associated patient characteristics; entries are in bold for variants that have not previously been described

cDNA	Protein	Predicted effect	Clinical significance	Evidence (ACMG)	Disease subtype	Age(s) of onset [years][†]	No. of times observed (sibling cases)	Patient origin
c.1_11del	p.M1?	LoF	VUS	PM2_P, PM2_P, PVS1_P	Late-infantile		1	Saudi Arabia
c.83G > A	p.R28K	Missense	VUS	PP3, PM2_P, PM3	Juvenile	2, 8.5	4	Oman (3), Saudi Arabia
c.84-1G > A	p.?	LoF	LP	PVS1_S, PM2_P, PM3_P	Juvenile		1	Sri Lanka
c.115C > G	p.R39G	Missense	VUS	PM2_P	Juvenile		1	Lebanon
c.144G > A	**p.W48***	**LoF**	**LP**	**PVS1, PM2_P**	**Late-infantile**		**1**	**Jordan**
c.-158_83del	**p.?**	**LoF**	**LP**	**PVS1_S, PM2_P, PM3_P**	**Late-infantile**		**1**	**United Arab Emirates**
c.-158_83dup	**p.?**	**LoF**	**VUS**	**PM2_P, PVS1_S**	**Late-infantile**		**1**	**Saudi Arabia**
c.164A > C	**p.N55T**	**Missense**	**VUS**	**PM2_P, PP3**	**Juvenile**	**13**	**1**	**Egypt**
c.177C > G	**p.D59E**	**Missense**	**VUS**	**PM2_P, PP3**	**Juvenile**	**7.0**	**1**	**Morocco**
c.214G > T	p.E72*	LoF	P	PM2_P, PVS1_VS, PM2_M, PS3_P	Late-infantile		1	Lebanon
c.224del	**p.P75Hfs*41**	**LoF**	**LP**	**PVS1, PM2_P**	**Late-infantile**		**1**	**Iran**
c.244G > C [‡]	**p.G82R**	**Missense**	**VUS**	**PM2_P, PP3, PP1_P**	**Late-infantile**		**1**	**Mexico**
c.247G > A	**p.D83N**	**Missense**	**VUS**	**PM2_P, PP3, PM3_P**	**Late-infantile**	**4.7**	**2 (1)**	**Egypt**
c.248A > G	**p.D83G**	**Missense**	**VUS**	**PM2_P, PM3_P**	**Late-infantile**		**1**	**Saudi Arabia**
c.257A > G	**p.H86R**	**Missense**	**VUS**	**PM3, PP3, PM2_P**	**Late-infantile**	**3.0, 4.0**	**5**	**Egypt**
c.281C > T	**p.P94L**	**Missense**	**VUS**	**PM2_P**	**Late-infantile**		**1**	**Egypt**
c.299 T > G	p.L100R	Missense	VUS	PM2_P, PP3, PM3_P	Late-infantile		1	Egypt
c.308G > A [§]	p.R103Q	Missense	LP	PM5, PM3, PM2_P, PP1, PP3	Adult		1	Colombia
c.316dup	p.R106Pfs*26	LoF	P	PVS1_VS, PM2P_P, PM3	Late-infantile	3.1, 5.1	2	Pakistan, Saudi Arabia
c.350 T > G	**p.I117S**	**Missense**	**VUS**	**PM2_P, PP3, PM3_P**	**Adult**		**1**	**Lebanon**
c.379C > T	**p.H127Y**	**Missense**	**VUS**	**PM2_P, PP3, PM3_P**	**Late-infantile**	**5.5**	**1**	**Turkey**
c.395_396del	p.S132Cfs*18	LoF	P	PVS1_VS, PM3, PM2_P	Late-infantile	3.0, 3.1, 4.0, 4.1	6 (1)	Saudi Arabia (3), Egypt (2), Turkey
c.397_398del	**p.V133Qfs*17**	**LoF**	**P**	**PVS1_VS, PM2_P, PM3**	**Late-infantile**	**0.1**	**2**	**Iran**
c.406C > T	p.R136C	Missense	LP	PM2_P, PP3, PM3, PP1_P, PM5	Late-infantile	3.5	1	Egypt
c.407G > A	p.R136H	Missense	LP	PM2_P, PP3, PM3_S, PM5	Late-infantile	3.4	2	Tunisia, Libya

Table 1 (continued)

cDNA	Protein	Predicted effect	Clinical significance	Evidence (ACMG)	Disease subtype	Age(s) of onset [years][†]	No. of times observed (sibling cases)	Patient origin
c.434A > C	**p.H145P**	**Missense**	**VUS**	**PM2_P, PP3**	**Late-infantile**		**1**	**Egypt**
c.434A > G	p.H145R	Missense	VUS	PM2_P, PP3	Late-infantile		1	Saudi Arabia
c.445C > T	p.R149C	Missense	VUS	PM2_P, PP3, PM5, PM3_P	Late-infantile		1	Egypt
c.461_463del	p.I154del	In-frame	P	PM4, PM2_P, PS3_P, PM3_S, PP1_M	Late-infantile		4	Brazil (2), Turkey, Algeria
c.476C > T	p.P159L	Missense	VUS	PM2_P, PP3, PM3	Late-infantile	2.1, 3.0	3	Iran (2), Turkey
c.486 + 2 T > A	**p.?**	**LoF**	**LP**	**PVS1_S, PM2_P, PM3_P**	**Late-infantile**	**6.1**	**1**	**Egypt**
c.486 + 2 T > C	**p.?**	**LoF**	**LP**	**PVS1_S, PM2_P, PM3_P**	**Late-infantile**	**3.0**	**1**	**Tunisia**
c.543-2A > G	**p.?**	**LoF**	**LP**	**PVS1_S, PM2_P, PM3_P**	**Late-infantile**	**2.9**	**1**	**Egypt**
c.552del[‡]	**p.F185Sfs*21**	**LoF**	**LP**	**PVS1_VS, PM2_P**	**Late-infantile**		**1**	**Mexico**
c.587G > T	**p.C196F**	**Missense**	**VUS**	**PM2_P**	**Late-infantile**		**1**	**Iran**
c.659A > C	p.Y220S	Missense	VUS	PM2_P, PP3, PM3_M	Late-infantile		1	Iran
c.662A > C	p.Y221S	In-frame	VUS	PM3, PP1, PM2_P	Late-infantile	5.0	4 (1)	Lebanon (3), Pakistan
c.715_717del	**p.F239del**	**In-frame**	**VUS**	**PM2_P, PM4**	**Late-infantile**		**1**	**Egypt**
c.757_759del	**p.L253del**	**In-frame**	**VUS**	**PM2_P, PM3_P, PM4**	**Late-infantile**		**1**	**Egypt**
c.765_785del	**p.D256_L262del**	**Missense**	**VUS**	**PM4, PM2_P, PM3_P**	**Late-infantile**	**4.0**	**1**	**Pakistan**
c.768C > G	p.D256E	Missense	LP	PP1_S, PM2_P, PP3, PM3	Teenage	15	1	Pakistan
c.773A > G	**p.N258S**	**Missense**	**VUS**	**PM2_P, PP3, PM3_P, PP1_M**	**Late-infantile**	**4.0**	**3 (1)**	**Saudi Arabia**
c.775G > A[§]	p.G259S	In-frame	VUS	PM2_P, PP3, PM3_P	Adult		1	Colombia
c.794_796del	p.S265del	Missense	LP	PM2_P, PP1, PM3, PM4	Late-infantile	3.0, 3.0, 3.0, 3.5, 3.5, 4.0, 5.0, 5.6	30 (2)	Egypt (7), Iraq, Jordan (2), Kuwait (4), Lebanon (7), Saudi Arabia (9)
c.829_836delinsCCT[¶]	p.V277Pfs*5	LoF	P	PM2_P, PVS1_S, PM3_S	Late-infantile		1	Brazil

[†] Where data were available

[‡] Heterozygote c.244G > C/c.552del

[§] Heterozygote c.308G > A / c.775G > A

[¶] Heterozygote c.461_463del / c.829_836delinsCCT

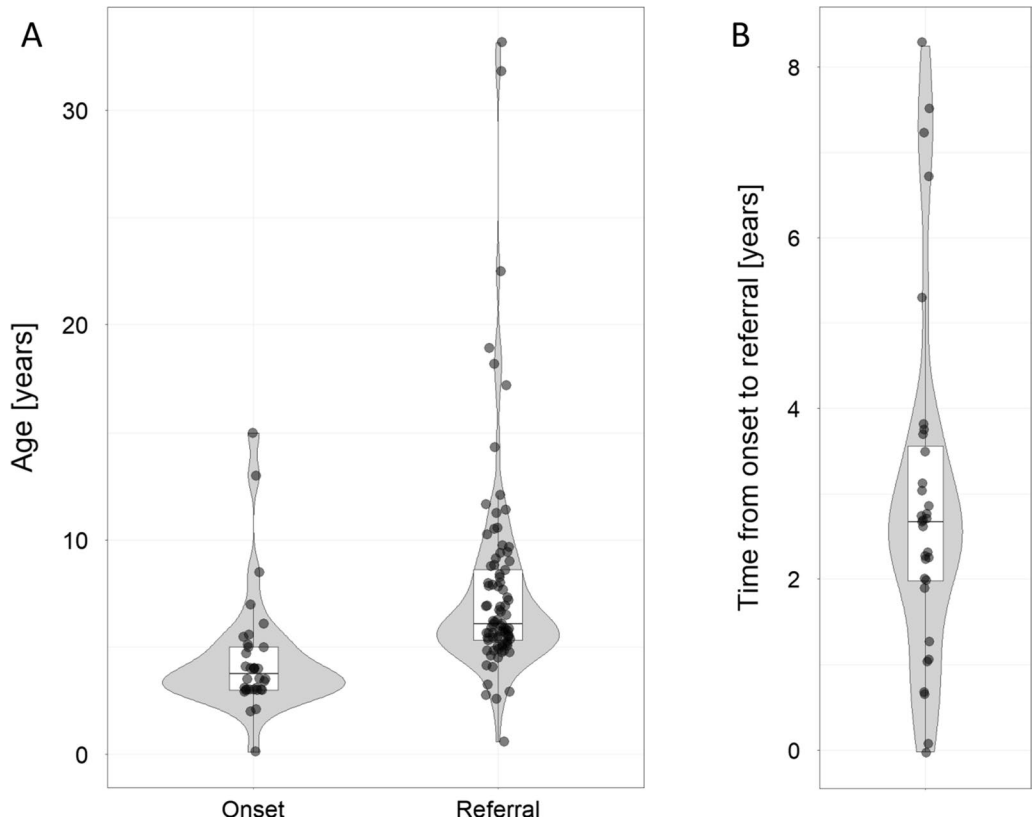


Fig. 2 Patient age at disease onset and referral. Each dot represents one of 34 (**A**) or 32 patients (**B**) where sufficient information was available. Grey areas show the smoothed distribution of ages or years, white boxes the interquartile ranges, and lines inside each box the medians. **A** Ages at onset and referral. **B** Time from disease onset to referral for genetic testing for NCL disease

patients with late-infantile onset. Likewise, none of the other alleles were shared by cases across disease onset groups. The missense variant c.83G > A (p.R28K) was exclusively present in three of the eight homozygous juvenile-onset patients (unrelated) and could be characteristic of this disease onset type. The four variants which were found only in adult-onset patients were c.768C > G, (p.D256E) and c.350 T > G (p.I117S) in homozygous, and c.308G > A (p.R103Q) and c.775G > A (p.G259S) as compound heterozygous genotype.

Association of individual *CLN6* variants with geographical region

CLN6 variant c.794 796del (p.S265del) was present in 22 out of 57 Middle Eastern, and in seven of the 31 North African patients from our cohort. Discounting sibling cases, the frequency was 39% and 21%, respectively. By contrast, this variant was absent in the nine patients from the Indian Subcontinent and Latin America. Some variants appeared to be more prevalent in specific countries. For example, all three unrelated patients from Oman, and one from Saudi Arabia had the c.83G > A (p.R28K) mutation. Of the novel described variants six originated from North Africa, four from the Middle East, one each from the Indian Subcontinent and Latin America. Cases with novel discovered variants were most prevalent in subjects from Egypt and Iran where they were present in over half of the cases (14/26 and 4 / 7, respectively). Specifically, c.257A > G(p.H86R) was found in five unrelated Egyptian individuals but none of the patients from other countries. New variants were much less frequent in Saudi Arabia (five of 21 cases) and Lebanon (one of 13 cases), the countries with the second and third-largest numbers of patients.

Discussion

We undertook a comprehensive analysis of the demographic, genetic, and clinical data collected from 97 diagnostic cases (85 unrelated and 6 pairs of siblings). To our knowledge, this is the largest cohort of CLN6 patients described so far for which genetic testing was performed at a single center. At the time of writing, the NCL Mutation and Patient Database [10, 25] listed 145 cases and 73 disease-associated CLN6 protein variants. We found

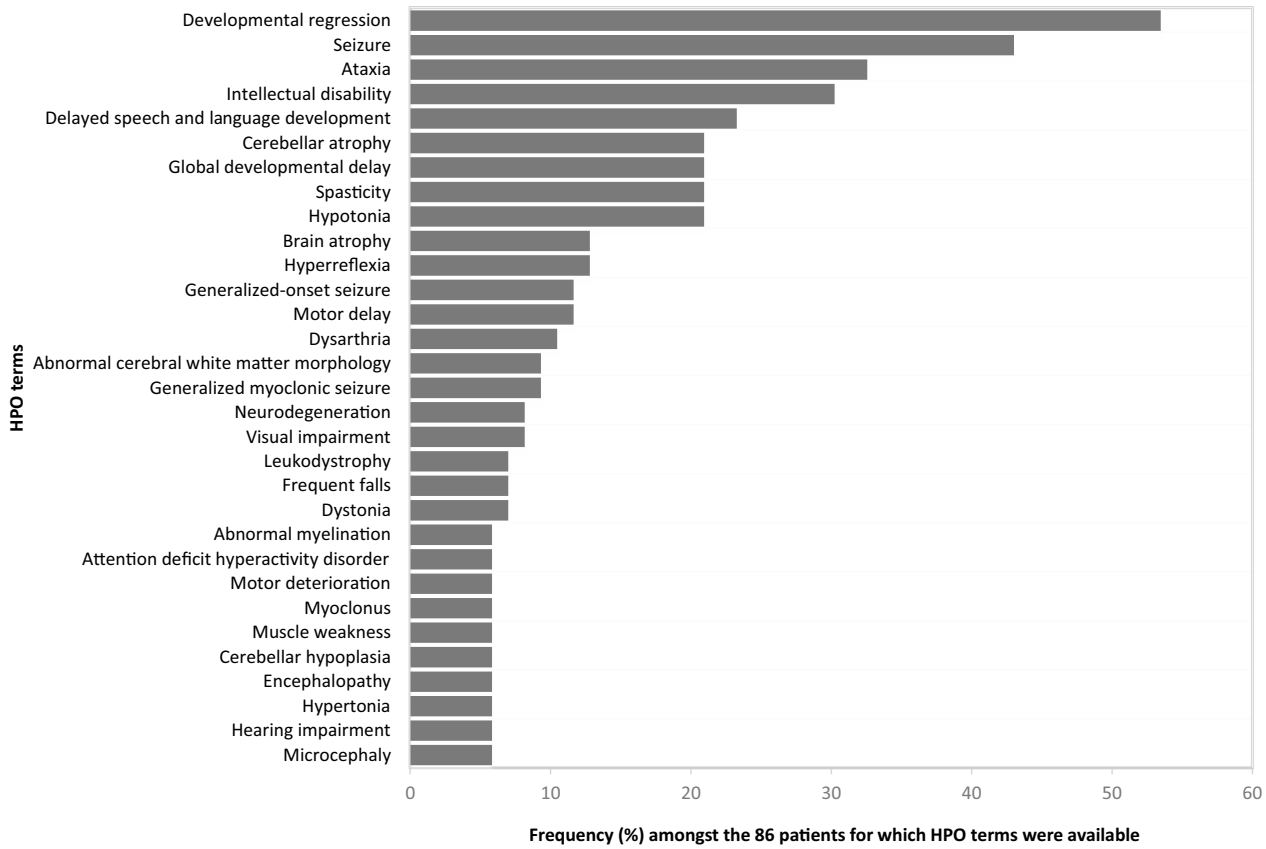


Fig. 3 Clinical features and their frequencies in the cohort. HPO terms that were used in five or more patient diagnoses. Note terms with different levels of specificity are shown regardless of whether they might represent the same symptom. Consequently, data from many patients described in highly specific but rarely used terms could not be included

45 distinct variants, including 24 that were novel identified. Our report therefore significantly expands the published dataset. Of particular interest, we describe *CLN6* disease and variants which are prevalent in regions that were underrepresented in the previous dataset compared to Europe and North America (Fig. 8A, B). Causative variants of the *CLN6* gene have been described in a wide range of ethnic groups [26–28]. These studies suggested a regional predominance of certain variants such as c.214G > T (p.E72*) in Costa Rican patients [29, 30]. This variant was observed in our cohort of mainly Middle Eastern and North African patients only once. Instead, the late-infantile onset associated with c.794 796del (p.S265del) was predominant (Fig. 6). While our patients originated from twenty different countries, by far the highest numbers were from Egypt (27%), Saudi Arabia (22%), and Lebanon (13%) (Table 1). Egypt was also the country with the highest number of novel discovered alleles, which represented more than half of the distinct variants in this population. Interestingly, the c.257A > G (p.H86R) variant was found in five out of 23 unrelated Egyptian individuals but not in patients from any other countries. Together, these data confirm the genetic and ethnic heterogeneity of CLN6 disease [10]. They also highlight a need that has been motivating the CentoMD® database: Better characterization of rare diseases in populations that are currently underrepresented in public data, to improve patient care, and to generate novel insights into genetics and disease mechanisms [23].

Percentages of missense, in-frame, and LoF mutations were similar in novel and previously described variants. A limitation of most studies, including ours, is that the sequencing method can fail to detect LoF due to large-scale genomic alterations: an example is a rare 12 kb deletion involving exon 1 of the *CLN6* gene recently described in two Japanese patients [31]. One of the new variants could be classified as pathogenic due to being an amino acid sequence deletion of considerable size.

Among the 31 cases in our cohort with novel discovered variants 91% were late-infantile, 6% juvenile, and 3% adult onset / Kufs type A. Ignoring additional disease subtypes and missing information, the NCL Mutation

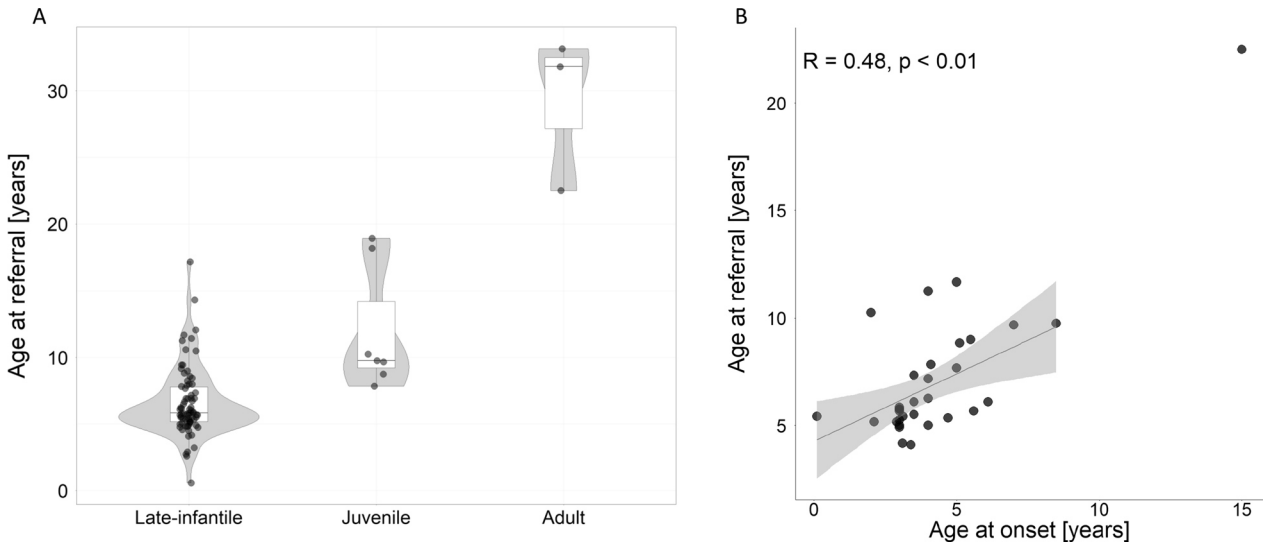


Fig. 4 Association between age at onset and age at referral. **A** Age at referral for patients diagnosed with late-infantile, juvenile, and adult-onset disease. The adult-onset group comprised two cases of Kufs disease type A, and one case of teenage progressive myoclonic epilepsy. Each dot represents one of 88 patients for which age was provided. Grey areas show the smoothed distribution of ages or years, white boxes interquartile ranges, and lines inside them the medians. **B** Correlation between age of onset and age of referral in 32 patients for which both ages were recorded. Also shown are a least square regression line with a 95% confidence interval (grey area), and the Pearson correlation coefficient (R) with significance level. A linear model was fitted without considering the outlier in the upper right corner of the plot

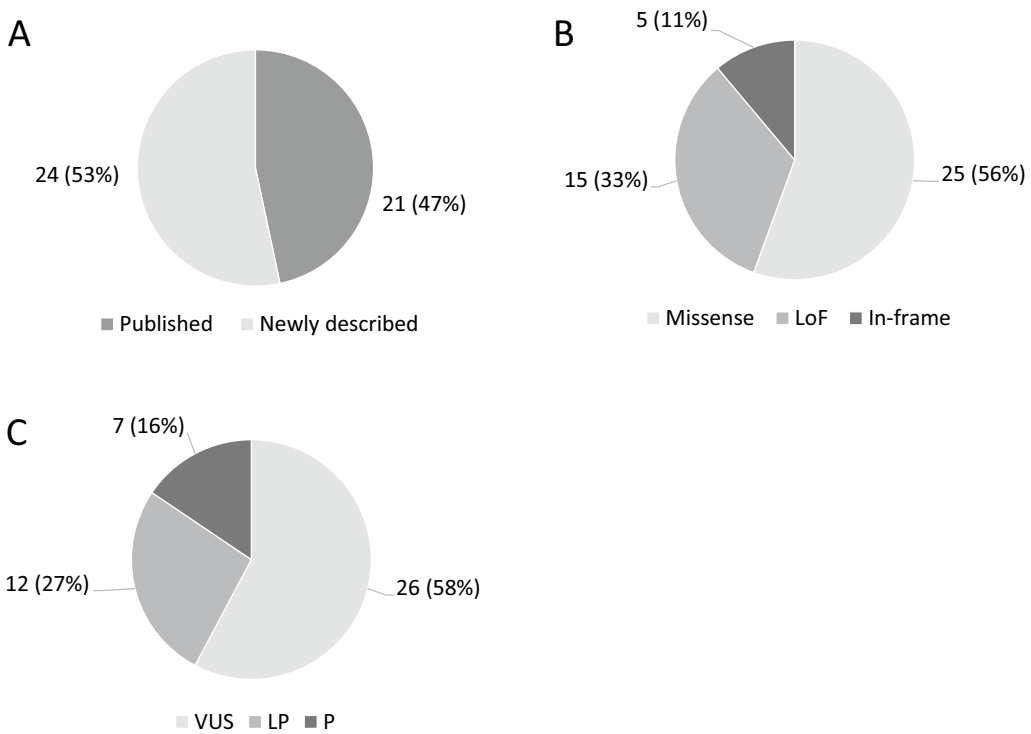


Fig. 5 Characterization of the *CLN6* variant spectrum. Shown are the total numbers of distinct variants within each class in 94 homozygous mono-allelic, and 3 compound heterozygous bi-allelic cases (relative frequencies in brackets). **A** Distribution of distinct variants classified according to clinical significance as pathogenic (P), likely pathogenic (LP), or a variant of uncertain significance (VUS). **B** Distribution of the causative variants at the protein level. The loss of function (LoF) group comprises frameshift, splicing, nonsense, start loss mutations, and gross deletions or duplications. **C** Numbers of previously published vs. novel described variants in this study

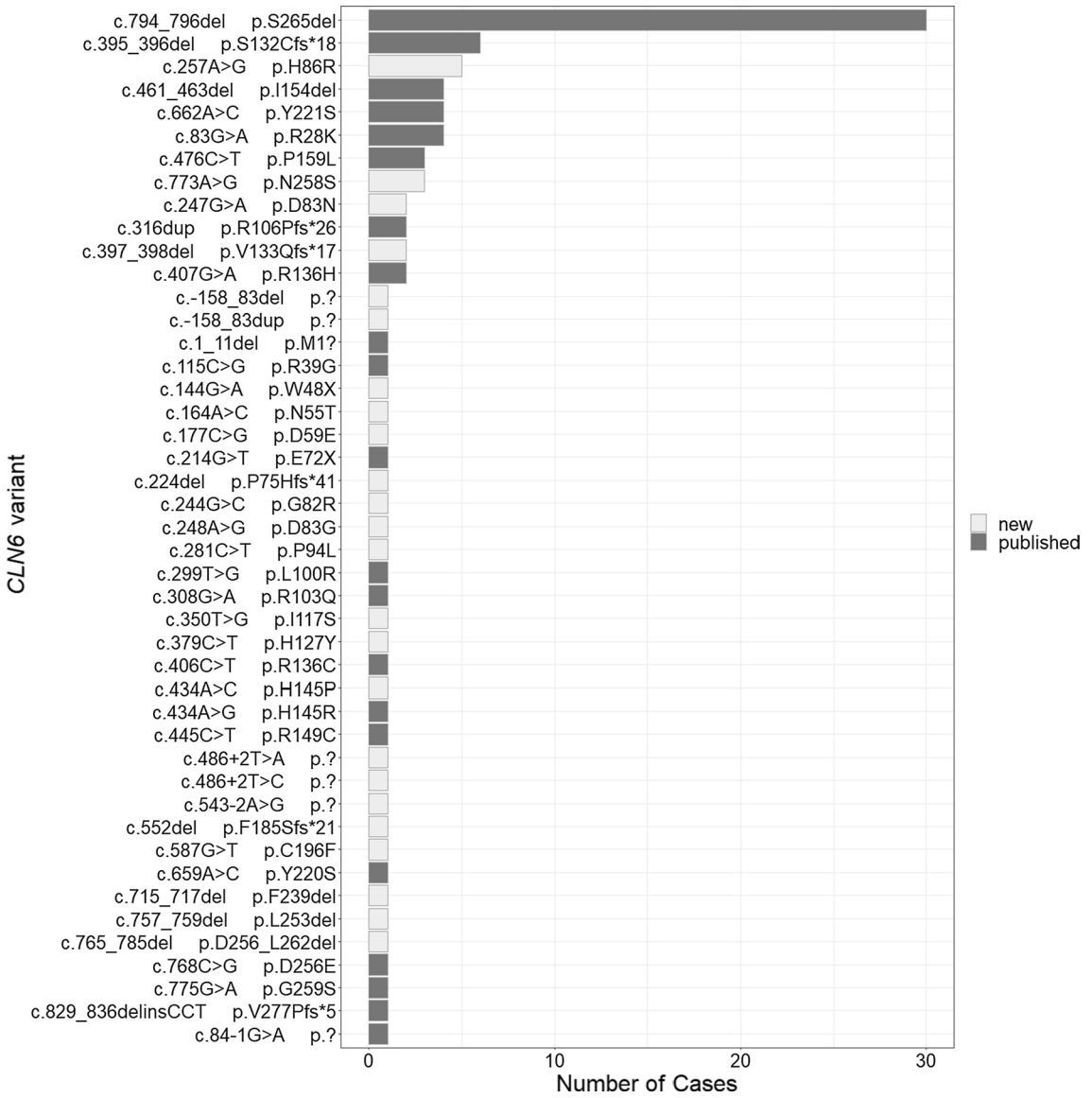


Fig. 6 Frequency of individual *CLN6* variants among patient cases. All variants are shown, including those found in three compound heterozygous individuals. Lighter shaded bars indicate the number of cases with variants novel discovered by this study. See Table 1 for numbers and associated patient characteristics.

and Patient Database [10] lists 110 cases with these clinical descriptions. Of these, 85% were late-infantile, 3% juvenile, and 13% adult onset / Kufs type A or B. Therefore, the distribution of onset types in cases with the novel discovered variants were similar to the published frequencies. The correlation between the ages of onset and referral was weak, and the time from first symptoms to referral for genetic diagnosis ranged from less than a month to up to eight years (Fig. 4).

Symptoms involving the nervous system, the musculoskeletal system, and the eye were the most common. In agreement with the current literature, we found no significant gender disparities in our cohort for patient numbers, age of onset, and occurrence of clinical symptoms (not shown). An interesting question is how the described variants can help to predict the onset and clinical symptoms of the disease. Our preliminary results suggest a trend for missense variants to be associated with a later onset (Fig. 7) but a detailed analysis was beyond the scope of this study. However, we hope that by increasing knowledge of the mutational spectrum and raising awareness of the disease this study

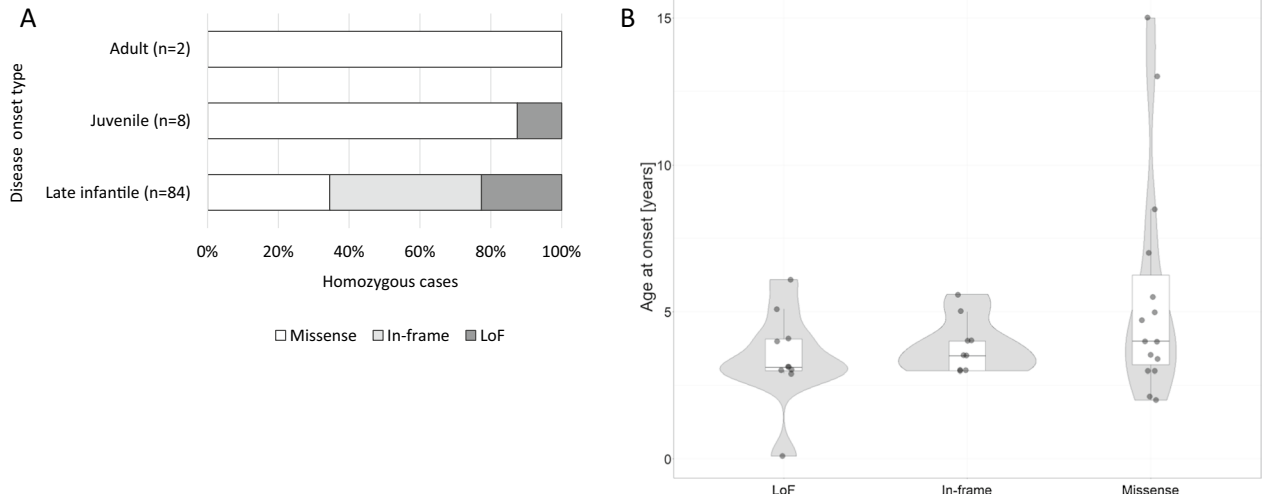


Fig. 7 Association between age at onset and *CLN6* predicted coding effect types. **A** Alleles in each variant class, stratified by disease diagnosis. Shown are the percentages of distinct alleles in homozygote cases. The total numbers are shown in the y-axis labels for each disease type; no variant was found in more than one. White: missense, grey: in-frame, black: loss of function (LoF) mutations. **B** Age at onset in 34 cases grouped by *CLN6* variant type, plotted as in Figs. 2 and 4A. All patients in this dataset were homozygotes

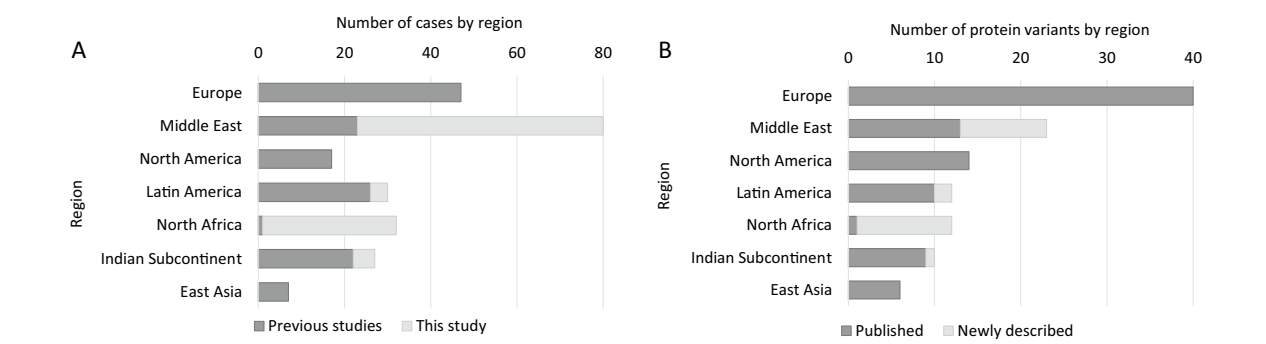


Fig. 8 Size and representation of geographical regions in datasets. Dark grey fill: NCL Mutation and Patient Database (25), light grey fill: this study. Europe: Czech Republic, Greece, Ireland, Italy, Malta, Portugal, Spain, UK. Middle East: Arabic origin, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Saudi Arabia, Turkey, United Arab Emirates. North America: Canada, USA. Latin America: Argentina, Brazil, Colombia, Costa Rica, Mexico, Venezuela. North Africa: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia. Indian Subcontinent: Afghanistan, India, Pakistan, Sri Lanka. East Asia: China, Japan. **A** Patient numbers stratified by region. **B** Numbers of distinct variants reported for each region

will contribute to earlier diagnosis for CLN6 patients worldwide. Early and accurate identification of the genetic cause will be critical for effective treatment, including the gene therapy approaches as have been recently started [21] for this devastating progressive disease.

Conclusions

We report the largest single-center cohort of CLN6 patients analyzed so far. It considerably expands the public data on CLN6 disease and *CLN6* mutational spectrum, especially for North Africa and the Middle East. It is hoped that this study will raise awareness for CLN6 disease and reduce the time from first symptoms to diagnosis for patients and their relatives worldwide. Including *CLN6* in the genetic diagnosis is recommended for individuals presenting with developmental regression, seizures, ataxia, intellectual disability, and ocular symptoms.

Methods

Patients and study design

A retrospective cross-sectional study was performed to investigate the clinical and mutational spectrum of

CLN6 disease. It involved 97 subjects submitted for routine genetic diagnosis of CLN genes between January 2010 and October 2020 at Centogene GmbH (Rostock, Germany). Clinical symptoms were the cause of referral in 72 cases, for the rest no information was provided.

Sample preparation and genetic analysis

All procedures were undertaken according to the provisions of the German Gene Diagnostic Act (Gendiagnostikgesetz) and the General Data Protection Act (Bundesdatenschutzgesetz) to guarantee confidentiality and data protection. The samples were processed at Centogene GmbH (Rostock, Germany) in a facility certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), accredited by the College of American Pathologists (CAP). Samples were provided either as extracted DNA, EDTA blood, dried blood spots (DBS) on filter cards (CentoCard®, Centogene GmbH, Rostock, Germany), amniotic fluid, or saliva. DNA extraction was done on a QIAsymphony instrument using reagents and kits recommended by the manufacturer (Qiagen, Hilden, Germany). Procedures used by us for variant screening have been described previously [23, 32]. Depending on the referring physician´s request, sequencing was performed either as whole-exome sequencing (WES), as gene panel in the CentoMetabolic® or Ceroid lipofuscinosis panels or for CLN6 alone. For gene panel sequencing, a custom double-stranded DNA capture bait pool was used to selectively enrich the coding regions, including 10 bp of flanking intronic sequences and known relevant variants beyond the coding regions, based on HGMD® and an in-house databank. Libraries were generated with Illumina compatible adaptors and sequenced on an Illumina platform (Illumina, San Diego, CA) to obtain $\geq 20\times$ coverage depth for > 99.5% of the targeted bases. Missing fragments were completed by Sanger sequencing when necessary. For WES, human consensus coding sequences were enriched from fragmented genomic DNA using the Nextera Rapid Capture Exome kit (Illumina) / SureSelect Human All Exon V6(Agilent)/ TWIST Human Core Exome (Twist Bioscience). The generated libraries were sequenced on an Illumina platform to an average coverage depth 70–100 $\times$. Any relevant variants detected by WES were validated by Sanger sequencing in both directions.

An in-house bioinformatics pipeline was applied for read alignment to the GRCh37/hg19 genome assembly, variant calling, annotation, and comprehensive variant filtering. The investigation focused on coding exons and flanking ± 10 intronic bases. Results were reviewed, interpreted, and reported by our scientific and medical experts. All potential disease-causing variants, including those reported in HGMD®, ClinVar, and in our databank were considered. Detected variants were classified according to published ACMG guidelines as pathogenic (P), likely pathogenic (LP), and variant of unknown significance (VUS) [33–35]. Clinical data provided by the referring physician were annotated in conformity with the Human Phenotype Ontology (HPO) nomenclature [36].

Statistical analysis

Medians, median-unbiased quartile ranges, and the correlation coefficient (Pearson´s R) and its significance were calculated using the stats package in R.

Acknowledgements

We thank Carolin Meier and Emir Zonic (both Centogene) for help with the clinical annotation of variants.

Authors' contributions

Conceptualization: CB and CMR, methodology: CB, software: KKK, investigation: CMR, data curation: DS, SM, original draft preparation: CMR, data analysis, figures and writing of the final draft: TW, review, and editing: CB, SDB, CP, SI, GMT, MER and BDD, supervision: RZ and CC, project administration: PB and AR. All authors have read and agreed to the published version of the manuscript. All authors were involved in revising the manuscript, gave final approval of the version to be published, and agreed to be accountable for all aspects of their work. All authors read and approved the final manuscript.

Funding

This study received no specific funding.

Availability of data and materials

Data on individual samples and research participants are not publicly available because of data privacy.

Declarations

Ethics approval and consent to participate

This project has been conducted within a diagnostic setting, and as a second step, utilized de-identified data and samples. Thus, this did not require Institutional Review Board (IRB) approval in our jurisdiction. Written informed consent was obtained from all patients for genetic studies as well as for the scientific publication of anonymized clinical data and clinical photographs. Additionally, the consent declaration included information regarding the storage of the data and further processing for research purposes. The informed consent form is available in English and several other languages at https://www.centogene.com/downloads.html.

Consent for publication

Not applicable.

Competing interests

CMR, TW, CB, CP, IS, BDD, GMT, MER, KKK, DS, SM, and PB are current employees, AR is a former employee of Centogene AG. All other authors declared no competing interests.

Author details

[1]Centogene GmbH, Am Strande 7, 18057 Rostock, Germany. [2]Institute of Chemistry, University of Rostock, Dr.-Lorenz-Weg 2, 18059 Rostock, Germany. [3]Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [4]Arcensus GmbH, Goethestrasse 20, 18055 Rostock, Germany. [5]Department of Medicine, Clinic III, Hematology, Oncology, Palliative Medicine, University of Rostock, Rostock, Germany.

Received: 21 December 2021 Accepted: 14 March 2022
Published online: 03 May 2022

References

1. Zeman W, Dyken P. Neuronal ceroid-lipofuscinosis (Batten's disease): relationship to amaurotic family idiocy? Pediatrics. 1969;44(4):570–83.
2. Mole SE, Williams RE, Goebel HH. Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. Neurogenetics. 2005;6(3):107–26.
3. Goebel HH. The neuronal ceroid-lipofuscinoses. J Child Neurol. 1995;10(6):424–37.
4. Haltia M. The neuronal ceroid-lipofuscinoses: from past to present. Biochem Biophys Acta. 2006;1762(10):850–6.
5. Williams RE, Mole SE. New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. Neurology. 2012;79(2):183–91.
6. Uvebrant P, Hagberg B. Neuronal ceroid lipofuscinoses in Scandinavia. Epidemiology and clinical pictures. Neuropediatrics. 1997;28(1):6–8.
7. Moore SJ, Buckley DJ, MacMillan A, Marshall HD, Steele L, Ray PN, et al. The clinical and genetic epidemiology of neuronal ceroid lipofuscinosis in Newfoundland. Clin Genet. 2008;74(3):213–22.
8. Santorelli FM, Garavaglia B, Cardona F, Nardocci N, Bernardina BD, Sartori S, et al. Molecular epidemiology of childhood neuronal ceroid-lipofuscinosis in Italy. Orphanet J Rare Dis. 2013;8:19.
9. Jalanko A, Braulke T. Neuronal ceroid lipofuscinoses. Biochim Biophys Acta. 2009;1793(4):697–709.
10. Mole SE, Cotman SL. Genetics of the neuronal ceroid lipofuscinoses (Batten disease). Biochim Biophys Acta. 2015;1852(10 Pt B):2237–41.
11. Brudvig JJ, Weimer JM. On the cusp of cures: breakthroughs in Batten disease research. Curr Opin Neurobiol. 2021;72:48–54.
12. Beck-Wodl S, Harzer K, Sturm M, Buchert R, Riess O, Mennel HD, et al. Homozygous TBC1 domain-containing kinase (TBCK) mutation causes a novel lysosomal storage disease—a new type of neuronal ceroid lipofuscinosis (CLN15)? Acta Neuropathol Commun. 2018;6(1):145.
13. Noskova L, Stranecky V, Hartmannova H, Pristoupilova A, Baresova V, Ivanek R, et al. Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. Am J Hum Genet. 2011;89(2):241–52.
14. Rechtzigel MJ, Meyerink BL, Leppert H, Johnson TB, Cain JT, Ferrandino G, et al. Transmembrane Batten disease proteins interact with a shared network of vesicle sorting proteins to regulate synaptic composition and function. 2021:2021.09.16.460691.
15. Benedict JW, Getty AL, Wishart TM, Gillingwater TH, Pearce DA. Protein product of CLN6 gene responsible for variant late-onset infantile neuronal ceroid lipofuscinosis interacts with CRMP-2. J Neurosci Res. 2009;87(9):2157–66.
16. Koh S, Cain J, Magee H, White K, Rechtzigel M, Meyerink B, et al. A CLN6-CRMP2-KLC4 complex regulates anterograde ER-derived vesicle trafficking in cortical neurites. 2021:2021.09.16.460653.
17. Arsov T, Smith KR, Damiano J, Franceschetti S, Canafoglia L, Bromhead CJ, et al. Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. Am J Hum Genet. 2011;88(5):566–73.
18. Tyynela J, Suopanki J, Santavuori P, Baumann M, Haltia M. Variant late infantile neuronal ceroid-lipofuscinosis: pathology and biochemistry. J Neuropathol Exp Neurol. 1997;56(4):369–75.
19. Johnson TB, Cain JT, White KA, Ramirez-Montealegre D, Pearce DA, Weimer JM. Therapeutic landscape for Batten disease: current treatments and future prospects. Nat Rev Neurol. 2019;15(3):161–78.
20. Cain JT, Likhite S, White KA, Timm DJ, Davis SS, Johnson TB, et al. Gene Therapy corrects brain and behavioral pathologies in CLN6-batten disease. Mol Ther J Am Soc Gene Ther. 2019;27(10):1836–47.
21. de los Reyes E, Aylward S, Meyer K, Lehwald L, Albright C, Rogers DL, et al. Single-dose AAV9-CLN6 gene transfer slows the decline in motor and language function in variant late infantile neuronal ceroid lipofuscinosis 6: Interim results from phase 1/2 trial. Mol Genet Metab. 2021;132(2):S32–3.
22. Chin JJ, Behnam B, Davids M, Sharma P, Zein WM, Wang C, et al. Novel mutations in CLN6 cause late-infantile neuronal ceroid lipofuscinosis without visual impairment in two unrelated patients. Mol Genet Metab. 2019;126(2):188–95.
23. Trujillano D, Oprea GE, Schmitz Y, Bertoli-Avella AM, Abou Jamra R, Rolfs A. A comprehensive global genotype-phenotype database for rare diseases. Mol Genet Genom Med. 2017;5(1):66–75.
24. Mink JW, Augustine EF, Adams HR, Marshall FJ, Kwon JM. Classification and natural history of the neuronal ceroid lipofuscinoses. J Child Neurol. 2013;28(9):1101–5.
25. Team RC. R: A Language and Environment for Statistical Computing. 2021.
26. Sharp JD, Wheeler RB, Parker KA, Gardiner RM, Williams RE, Mole SE. Spectrum of CLN6 mutations in variant late infantile neuronal ceroid lipofuscinosis. Hum Mutat. 2003;22(1):35–42.
27. Al-Muhaizea MA, Al-Hassnan ZN, Chedrawi A. Variant late infantile neuronal ceroid lipofuscinosis (CLN6 gene) in Saudi Arabia. Pediatr Neurol. 2009;41(1):74–6.
28. Nafi O, Ramadan B, Riess O, Buchert R, Froukh T. Two cases of variant late infantile ceroid lipofuscinosis in Jordan. World J Clin Cases. 2019;7(2):203–8.
29. Gao H, Boustany RM, Espinola JA, Cotman SL, Srinidhi L, Antonellis KA, et al. Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. Am J Hum Genet. 2002;70(2):324–35.
30. Wheeler RB, Sharp JD, Schultz RA, Joslin JM, Williams RE, Mole SE. The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (CLN6) and in nclf mutant mice encodes a novel predicted transmembrane protein. Am J Hum Genet. 2002;70(2):537–42.
31. Mizuguchi T, Suzuki T, Abe C, Umemura A, Tokunaga K, Kawai Y, et al. A 12-kb structural variation in progressive myoclonic epilepsy was newly identified by long-read whole-genome sequencing. J Hum Genet. 2019;64(5):359–68.
32. Bertoli-Avella AM, Kandaswamy KK, Khan S, Ordonez-Herrera N, Tripolszki K, Beetz C, et al. Combining exome/genome sequencing with data repository analysis reveals novel gene–disease associations for a wide range of genetic disorders. Genet Med. 2021;23(8):1551–68.
33. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical G, et al. ACMG standards and guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision-2013. Genet Med. 2013;15(11):901–9.
34. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Off J Am Coll Med Genet. 2015;17(5):405–24.
35. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med Off J Am Coll Med Genet. 2017;19(2):249–55.
36. Kohler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine JP, et al. Expansion of the human phenotype ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47(D1):D1018–27.

Publisher's Note

scientific reports

Check for updates

OPEN

Neuronal progenitor cells-based metabolomics study reveals dysregulated lipid metabolism and identifies putative biomarkers for CLN6 disease

Corina-Marcela Rus[1,2]✉, Daniel L. Polla[1], Sebastiano Di Bucchianico[2,3,4], Steffen Fischer[1], Jörg Hartkamp[1], Guido Hartmann[1], Yunus Alpagu[1], Claudia Cozma[1], Ralf Zimmermann[2,3,4] & Peter Bauer[1,5]

Neuronal ceroid lipofuscinosis 6 (CLN6) is a rare and fatal autosomal recessive disease primarily affecting the nervous system in children. It is caused by a pathogenic mutation in the *CLN6* gene for which no therapy is available. Employing an untargeted metabolomics approach, we analyzed the metabolic changes in CLN6 subjects to see if this system could potentially yield biomarkers for diagnosis and monitoring disease progression. Neuronal-like cells were derived from human fibroblast lines from CLN6-affected subjects (n = 3) and controls (wild type, n = 3). These were used to assess the potential of a neuronal-like cell-based metabolomics approach to identify CLN6 distinctive and specific biomarkers. The most impacted metabolic profile is associated with sphingolipids, glycerophospholipids metabolism, and calcium signaling. Over 2700 spectral features were screened, and fifteen metabolites were identified that differed significantly between both groups, including the sphingolipids C16 GlcCer, C24 GlcCer, C24:1 GlcCer and glycerophospholipids PG 40:6 and PG 40:7. Of note, these fifteen metabolites were downregulated in the CLN6 disease group. This study is the first to analyze the metabolome of neuronal-like cells with a pathogenic mutation in the *CLN6* gene and to provide insights into their metabolomic alterations. This could allow for the development of novel biomarkers for monitoring CLN6 disease.

Neuronal ceroid lipofuscinoses (NCLs) are among the most frequently encountered groups of rare, inherited neurodegenerative lysosomal storage disorders in children[1–3]. The worldwide prevalence of NCLs varies based on the region and the variant type. Their frequency is estimated at 0.01 to 9 per 100,000 live births[4,5]. NCLs are caused by mutations in one of the thirteen CLN genes (*CLN1-CLN8, CLN10-CLN14*)[6,7] that impact different proteins, one of which is the transmembrane protein CLN6[6]. Neuronal ceroid lipofuscinosis—type 6 (CLN6) [OMIM# 601780] is an autosomal recessive disease caused by pathogenic mutations in the *CLN6* gene that encodes for the CLN6 protein, whose function is not yet fully understood[8,9]. The CLN6-affected subjects develop symptoms between 18 months and 8 years of age, and among the first clinical signs are ataxia, seizures, and progressive mental deterioration[10,11].

Early diagnosis of CLN6 is essential for developing treatment and managing disease prognosis[10,12]. However, diagnosis is based on combined clinical symptoms and genetic testing and is often made at an advanced disease stage, which brings an unfavorable prognosis. Research on CLN6 has intensified over the past decade as extensive attempts were made to develop therapies and understand the disease pathology[12–14]. Even so, the disease mechanism underlying CLN6 pathogenesis remains unclear, and the development of an appropriate treatment

[1]Centogene GmbH, Am Strande 7, 18057 Rostock, Germany. [2]Joint Mass Spectrometry Center, Chair of Analytical Chemistry, University of Rostock, Albert-Einstein Straße 27, 18059 Rostock, Germany. [3]Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [4]Department Life, Light & Matter, University of Rostock, Albert-Einstein Straße 25, 18059 Rostock, Germany. [5]Department of Medicine, Clinic III, Hematology, Oncology, Palliative Medicine, Rostock University Medical Center, Ernst-Heydemann-Str. 6, 18057 Rostock, Germany. ✉email: corina-marcela.rus@uni-rostock.de

is pending. Among the clinical needs awaiting to be achieved is the discovery of specific biomarkers for disease screening, prognosis, and monitoring[15,16].

Liquid chromatography-mass spectrometry (LC–MS) is a robust platform that may uncover intricate metabolic pathways, deepen our understanding of the biochemical processes, and aid in biomarkers discovery in clinical and translational research[17,18]. Likewise, non-targeted metabolite profiling is a valuable tool for comparing metabolic changes between pathological and non-affected subjects. These together may assist in identifying a wide range of critical metabolites whose changes are linked to specific diseases[19,20].

Only a few LC–MS-based metabolomics studies in NCLs have been conducted to identify potential biomarkers of disease progression[21,22]. Furthermore, CLN6 disease metabolomics studies are scarce and mainly related to animal models[23,24]. Hence, this implies that the metabolome of human CLN6 disease in neuronal-like cells is largely unexplored.

Here, we sought to compare the metabolomic profile in neuronal-like cells generated from fibroblasts of CLN6-affected and unaffected subjects (wild type). First, we set out to differentiate human CLN6 fibroblast lines into chemical-induced neuronal progenitor cells (ciNPC). Next, this model was used for biomarker discovery, which may aid in a rapid and inexpensive diagnosis and prognosis of the neurodegenerative disease[25].

Considering that no human CLN6 studies have been published that addressed the metabolomic changes in cell lines of CLN6 patients, we aimed to analyze and compare the changes in the global metabolome of the induced neuronal-like cell lines from subjects with CLN6 disease to that of healthy subjects. The present cell-based study employed an untargeted LC–MS approach that, combined with in-depth data analysis, helped identify metabolic alterations linked to CLN6 disease.

Results

The study was based on cell lines derived from subjects with CLN6 disease (carrying bi-allelic pathogenic variants, n = 3) and controls (wild type, n = 3). Inclusion criteria for the CLN6 subject were: (1) diagnosis of CLN6 based on genetic analysis, (2) both male and female subjects, (3) under 18 years at the time of sampling, and (4) unrelated individuals. The only difference in the inclusion criteria for control subjects was that they must not be diagnosed with any NCL disease and should not have any pathogenic or potentially harmful genetic variations. Detailed information about the included subjects is provided in Table 1.

Generation of neuronal progenitor cells (NPCs) from human dermal fibroblasts

We used a simple and previously reported technique introduced by Dai et al. to directly generate ciNPCs from fibroblasts[26]. Figure 1a visually presents the sequential stages of the differentiation process using a cocktail of six small molecules, as outlined in the aforementioned protocol. As early as day 9 of differentiation, neuronal-like cell clusters were observed, and the small colonies of ciNPCs were depicted by day 20 post-induction (Fig. 1b). Three weeks after differentiation, the cells were harvested and checked for NPC markers via immunofluorescence staining (GABA, GFAP, TUJ1, and MAP2) and the metabolites extracted for LC–MS analysis. The differentiation resulted in the expression of neuronal markers, such as GABA (GABAergic neuron marker), GFAP (glial fibrillary acidic protein), TUJ 1 (Neuron-specific class III beta-tubulin), and MAP-2 (Microtubule-associated protein 2), as confirmed by immunofluorescence (Fig. 2). The CLN6 and WT NPCs had similar expression levels of the neuronal markers listed above. Immunofluorescence images revealed that ciNPCs express neuronal markers at day 20 of neuronal differentiation.

Group	Cell line	Sex	[a]Age at onset|sampling	[b]cDNA protein	Predicted effect	Type on DNA	Coding effect	Clinical significance	Evidence ACMG	Clinical symptoms
CLN6	I	M	5|16	c.-158_83del p.?	Deletion	Gross deletion	Effect unknown	Likely pathogenic	PVS1_S, PM2_P, PM3_P	Movement disorders, epilepsy, cognitive impairment
	II	F	5|5½	c.896C>T p.Pro299Leu	Missense	Substitution	Probably damaging	Pathogenic	PP3, PS3_P, PM2_P, PM3_S	Behavioural abnormalities, weight loss, cognitive regression, pyramidal and extrapyramidal signs, myoclonus, focal motor seizures
	III	F	4|5	c.794_796del p.Ser265del	In-frame	Deletion	In-frame	Likely pathogenic	PM2_P, PP1, PM3, PM4	Movement disorders, epilepsy, cognitive impairment
WT	I	M	–|8	NA						
	II	M	–|1	NA						
	III	F	–|11	NA						

Table 1. Clinical and genetic characteristics of subjects used in this study. WT wild type/control, [a]year, [b]homozygous, NA not applicable.

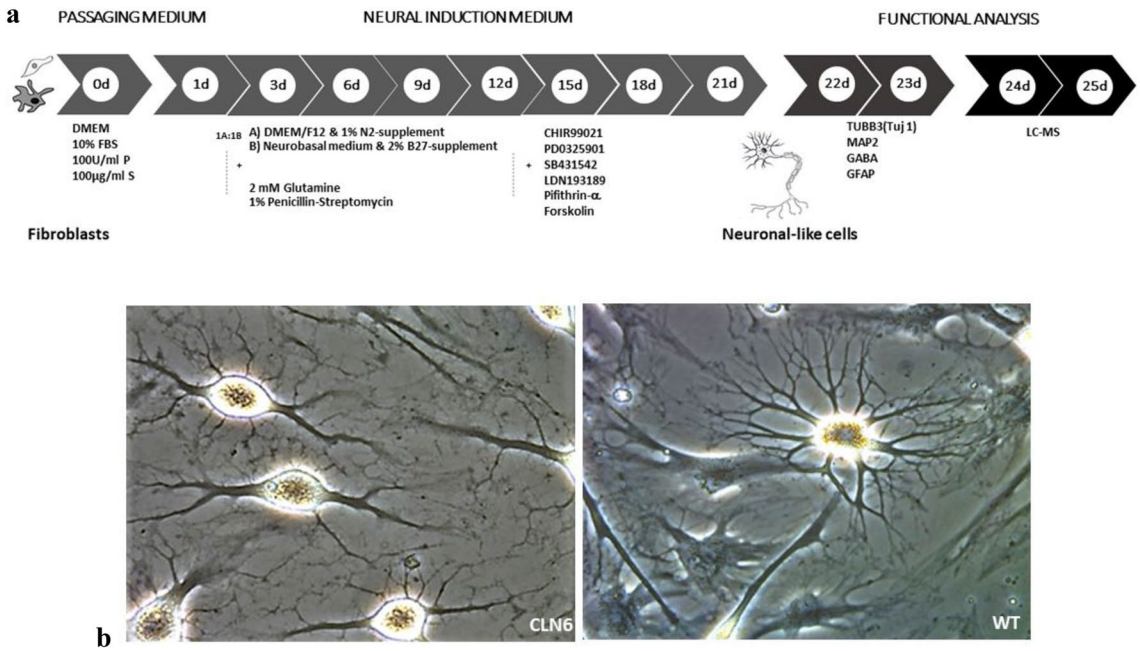


Figure 1. (**a**) Scheme of direct conversion of human dermal fibroblasts into neuronal-like cells. The experiment starts by plating the fibroblasts in the DMEM medium, which was designated as "Day"0." After 1 day, the cells were transferred to an induction medium containing chemical compounds and supplementary chemicals to promote differentiation into neuronal progenitor cells. (**b**) Representative microscope images of human ciNPCs morphologies at day 20 of development in the induction medium: (**a**) CLN6 group and (**b**) wild type/control group.

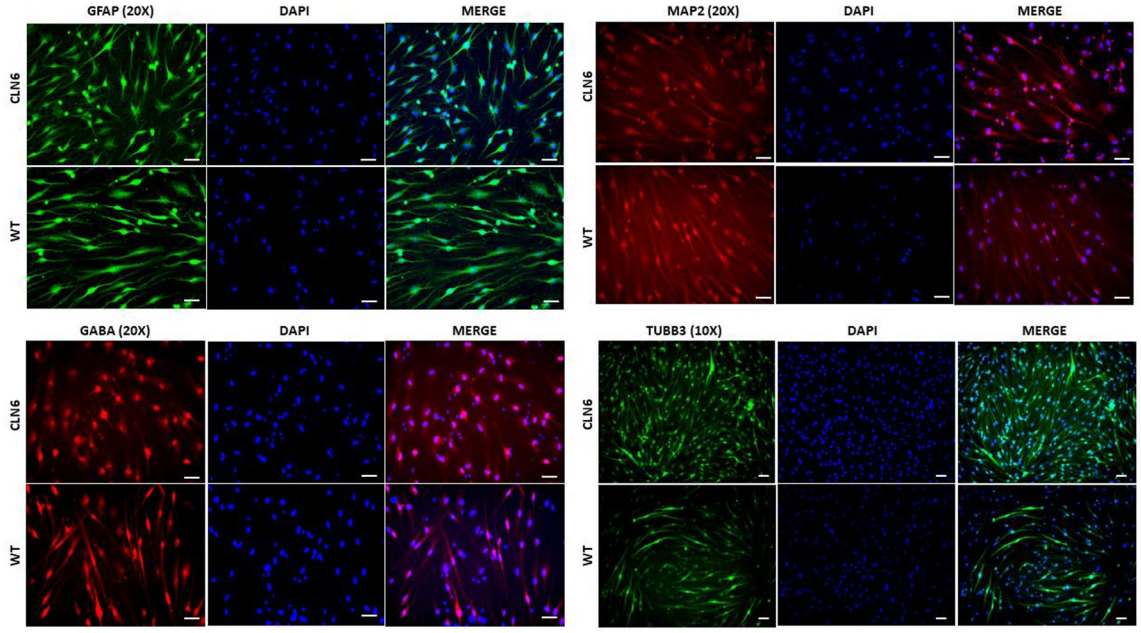


Figure 2. Immunofluorescence staining of neuronal cultures with anti-MAP2 antibody (mature neurons), anti-GABA (GABAergic neurons), anti-TUB βIII (immature neurons), and anti-GFAP antibody (astrocytes). Nuclei were stained with DAPI (blue). Scale bars = 50 µm. MAP, microtubule-associated protein 2; GFAP, glial fibrillary acidic protein; DAPI, 49,6-diamidino-2-phenylindole.

Metabolomics analysis reveals distinct metabolic profiling in CLN6 subjects

Global metabolomic profiling of fibroblast-derived ciNPCs from CLN6 and healthy subjects was performed using a non-targeted mass spectrometry (MS) approach. After conducting the sample quality check, two replicates from each group [CLN6-1 (b3-t1, b3-t2) and control-3 (b2-t1; b2-t2)] were excluded from the batch. The exclusion was necessary because there was no signal in the total ion chromatogram, likely caused by sample

evaporation prior to the LC–MS measurement. As a result, 16 replicates (data points) per group were available for further data analysis.

There were 2720 spectral features detected (Supplementary Table 1) and defined as molecular entities with a unique retention time (RT) and mass value (m/z). The coefficient of variation (CV) across all cell lines and cohorts was under 30% for more than 64% of the variables, meaning that both groups have a similar degree of variability in their concentration across all cell lines and cohorts.

The datasets were then subjected to univariate and multivariate data analysis to assess the spectral features' alteration in the two groups. Based on the entire metabolome datasets, we generated a principal component analysis (PCA) score plot that revealed a significant separation between CLN6 and the control group. As depicted in Fig. 3, PC1 contains the metabolites responsible for most between-group variations (36.8%), followed by PC2 (12.4%). As indicated by their high loadings in PC1, the glycerophospholipids such as PG 40:6, PG 40:7, PG 34:2, PG 32:1, and PG 40:4 contributed significantly to the overall variance captured by PC1.

The individual selection of discriminating metabolites associated with CLN6 disease was based on folds change of at least two and $p \leq 0.05$. Figure 4a displays the volcano plot depicting the fold changes in metabolites' abundance and highlights the most significantly altered metabolites in terms of fold change and discriminatory power between the two groups. This analysis used the comprehensive dataset from Supplementary Table 1, incorporating the raw data obtained through LC–MS. The compounds chosen for further analysis were limited to the top 20 from the list of upregulated and downregulated compounds shown in the volcano plot. They were selected based on their intensity and discriminative power to differentiate between groups. It should be emphasized that the selected compounds were exclusively downregulated upon this features screening.

The metabolites were putatively annotated, which involved matching their mass-to-charge ratio (m/z) and retention time (RT) values to internal and external databases. Supplementary Table 2 lists the fold changes and Student's t-test p-values (not adjusted) for these metabolites. Additionally, the metabolites with a fold change threshold of at least two are shown in a heatmap colored based on the actual peak intensity values to delineate the differences between the two groups (Fig. 4b, Supplementary Table 3).

Two additional multivariate analyses were conducted to investigate the differences between the CLN6 and control subjects: unsupervised hierarchical cluster analysis (HCA) and ortho partial least squares-discriminant analysis (OPLS-DA). The unsupervised hierarchical clustering analysis based on the HCA technique was performed to group the data into clusters (Fig. 5a). The OPLS-DA analysis, on the other hand, was carried out to differentiate between the two cohorts and identify CLN6 dysregulated metabolites (Fig. 5b).

The metabolites with the highest discriminating power were chosen according to the Variable Importance in Projection (VIP). A VIP score ≥ 1.00 from PLS-DA was considered significant. Fifteen metabolites, all downregulated in the CLN6 group, were differentially expressed in the CLN6 with significant discriminative power from the control group (Fig. 6, Supplementary Table 2).

Five of all metabolites analyzed, met the stringent quality and quantity criteria in our detection process, as depicted in Fig. 7. These criteria entail a significant difference between the control and targeted disease group, a minimum fold change of twofold, and a high median normalized abundance. Furthermore, the compounds were ranked based on ion alignment, peak picking inspection, chromatogram, and intensity visualization. Consequently, these metabolites have been considered eligible in terms of quality and quantity for downstream

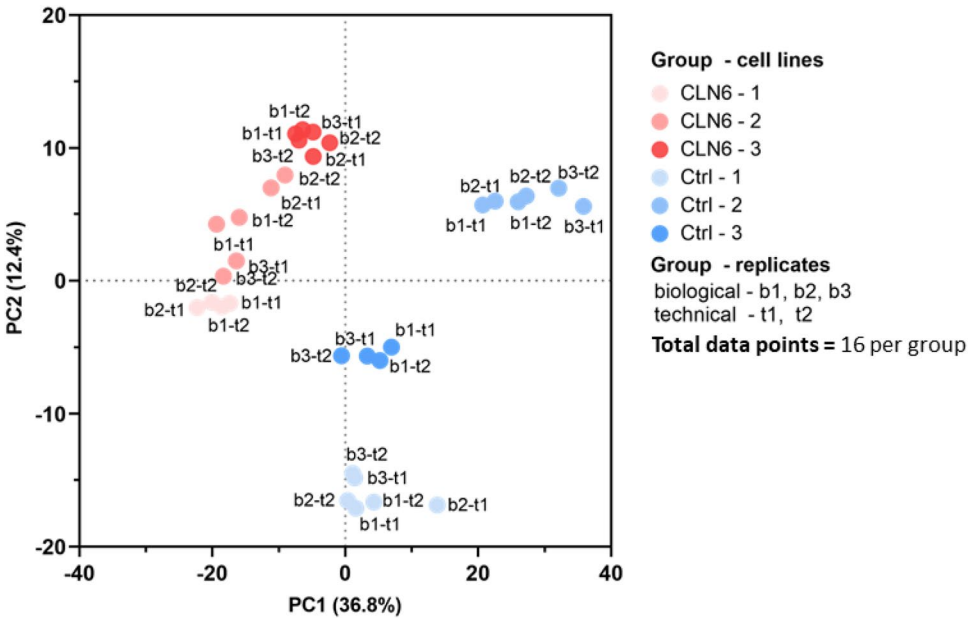


Figure 3. Principal component analysis (PCA) score plot of metabolites shows a separation of the CLN6 group from the control group, based on PC1 and PC2. The analysis was conducted using 16 data points (technical replicates) per group.

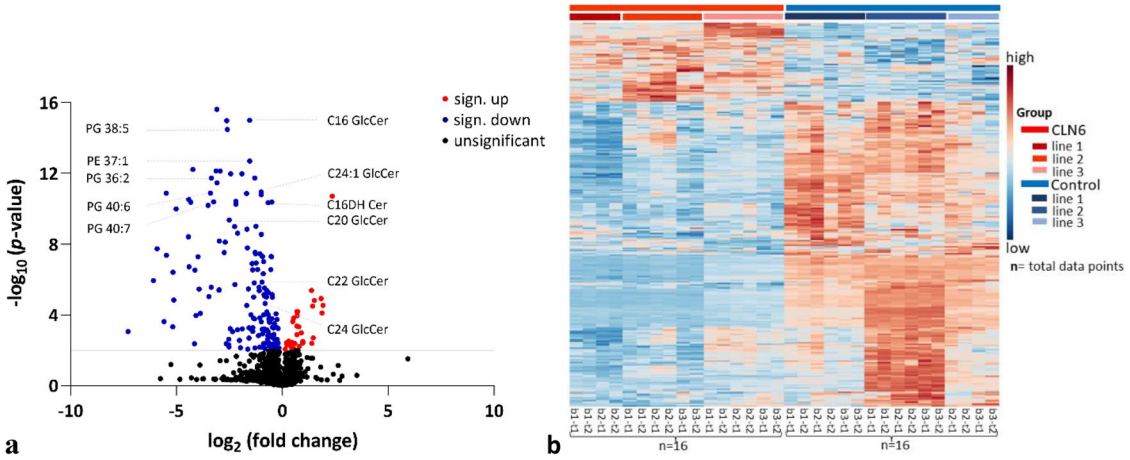


Figure 4. (**a**) Volcano plot displays the metabolites with a significant threshold of at least 2 × difference. The values are log-transformed. The threshold is shown as a grey line. The metabolites highlighted in red are upregulated for the CLN6 group, in blue are downregulated, and in black are not significantly changed ($p \leq 0.05$). (**b**) Heatmap showing the abundance levels of the 10% most distinct metabolites screened across the two groups. The colors indicate increased (red) and decreased (blue) abundance for each metabolite across the samples. The letters (bottom) represent the biological replicas (b) and the technical replicas (t). Total number of replicates (data points) per group (n) = 16. Observation: most of the metabolites are decreased in CLN6 samples.

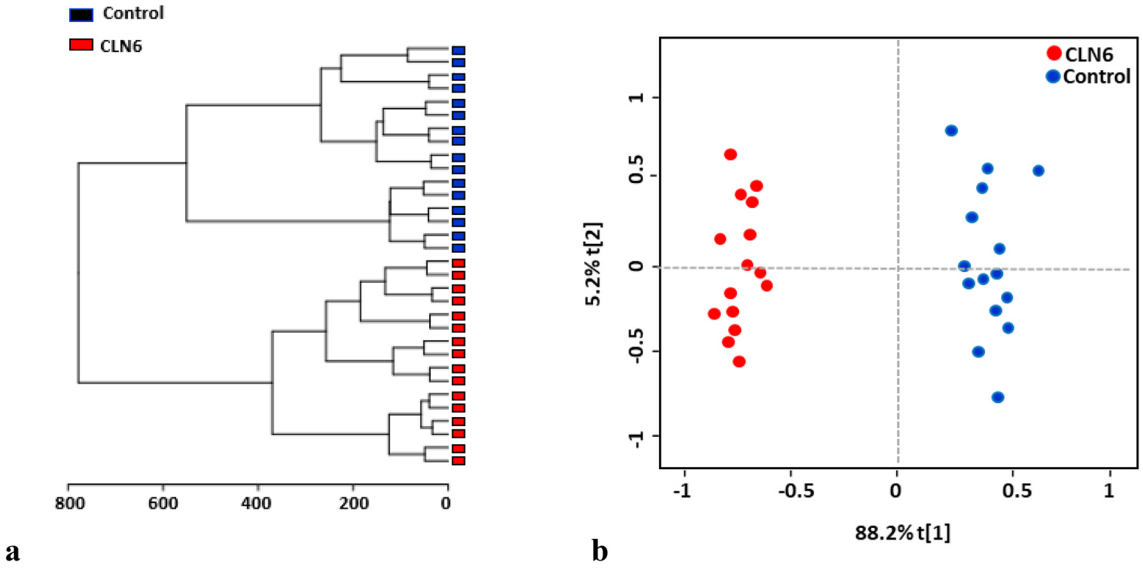


Figure 5. (**a**) Hierarchical clustering (Euclidean distance, Ward's clustering algorithm) confirms the biochemical differences between the two groups. (**b**) Ortho partial least squares-discriminant analysis (OPLS-DA) score plot from CLN6 and control. The clear separation between the two groups indicates that their metabolomic profile is distinct.

characterization. The identified metabolites, likely biologically relevant within CLN6 disease, belong to the glycerophosphoglycerols and glycosphingolipids class. Specifically, they are represented by phosphatidylglycerols PG (40:7) and PG (40:6), as well as glucosylceramides C16 GlcCer, C24 GlcCer, and C24:1 GlcCer. The key characteristics covered by these metabolites are listed in Table 2.

They were selected as potential candidate biomarkers due to their differentiating power between the CLN6 disease and control groups (Fig. 8).

The five metabolites' overlaid receiver operating characteristic (ROC) curves reveal that these compounds delineate between the two groups with high precision and accuracy, indicating that they could be potential biomarker candidates for the early disease prognosis of CLN6 disease (Fig. 9).

Pathway and network analysis of the differential expressed metabolites

To investigate the biological pathways and networks involved in the pathogenesis of CLN6 disease, we performed metabolomics data enrichment on selected metabolites using two software tools: MetaboAnalyst *v.* 5.0[27] and

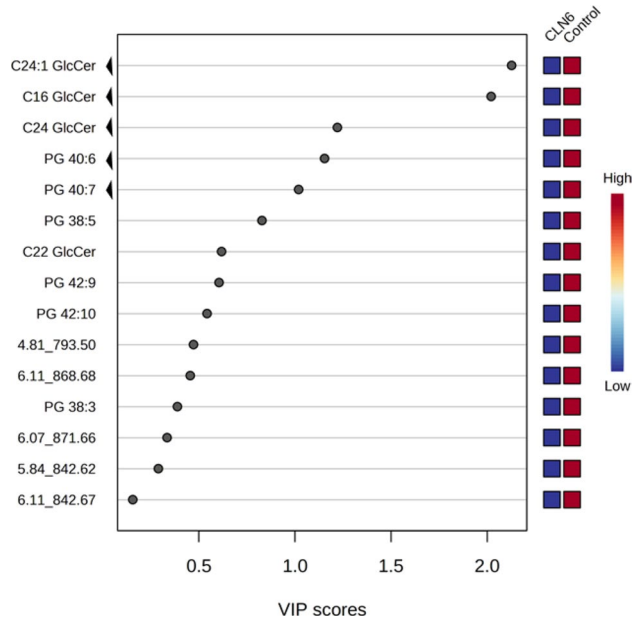


Figure 6 . Variable importance in projection (VIP) score plot displays the metabolites with discriminating power derived from the PLS-DA model. Only the metabolites with the strongest discriminating power are represented here. These metabolites have contributed most to the variance between CLN6 and the control group. The non-annotated metabolites were named based on the retention time (rt) and *m/z* values; 'arrows' indicate the five most discriminating metabolites.

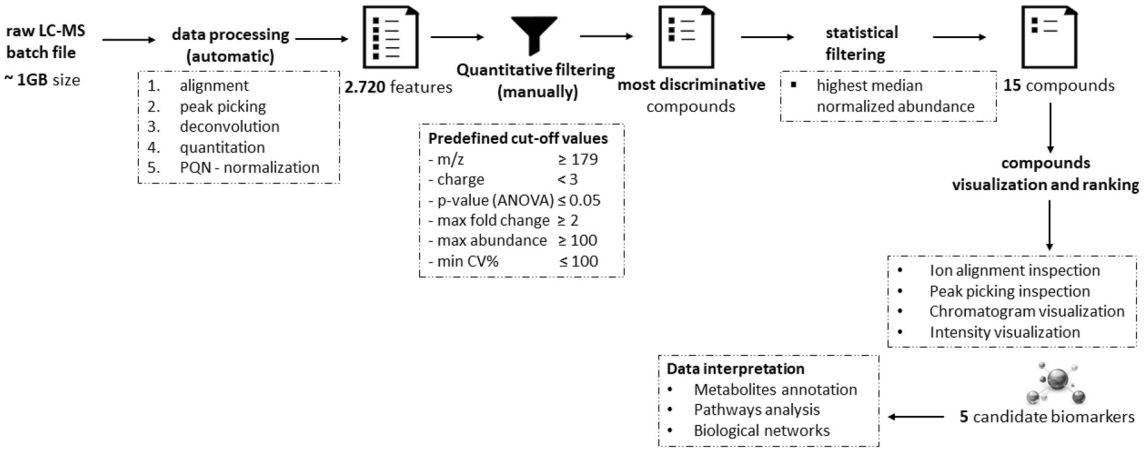


Figure 7. Metabolomics data analysis scheme for biomarker identification.

Class	Synonim	Mass (*m/z*, n)	CCS (Å)	RT (min)	*P* value	Fold change	Adducts	Formula
Glycerophosphoglycerols	PG 40:7	819.5 *m/z*	294.9	4.88	1.02E−10	-8	M−H	$C_{46}H_{77}O_{10}P$
	PG 40:6	822.5n	299.1	4.97	2.51E−11	-3	M−H, M+Cl	$C_{46}H_{79}O_{10}P$
Glycosphingolipids	C16 GlcCer	699.6n	285.0	5.31	1.39E−15	-3	M−H, M+Cl, [M+HCOO]-	$C_{40}H_{77}NO_8$
	C24 GlcCer	811.7n	309.0	6.45	4.72E−06	-2	M−H, M+Cl	$C_{48}H_{93}NO_8$
	C24:1 GlcCer	809.7 n	316.0	6.09	1.84E−13	-2	M−H, M+Cl, [M+HCOO]-	$C_{48}H_{91}NO_8$

Table 2. Characteristics of the top five putatively annotated metabolites.

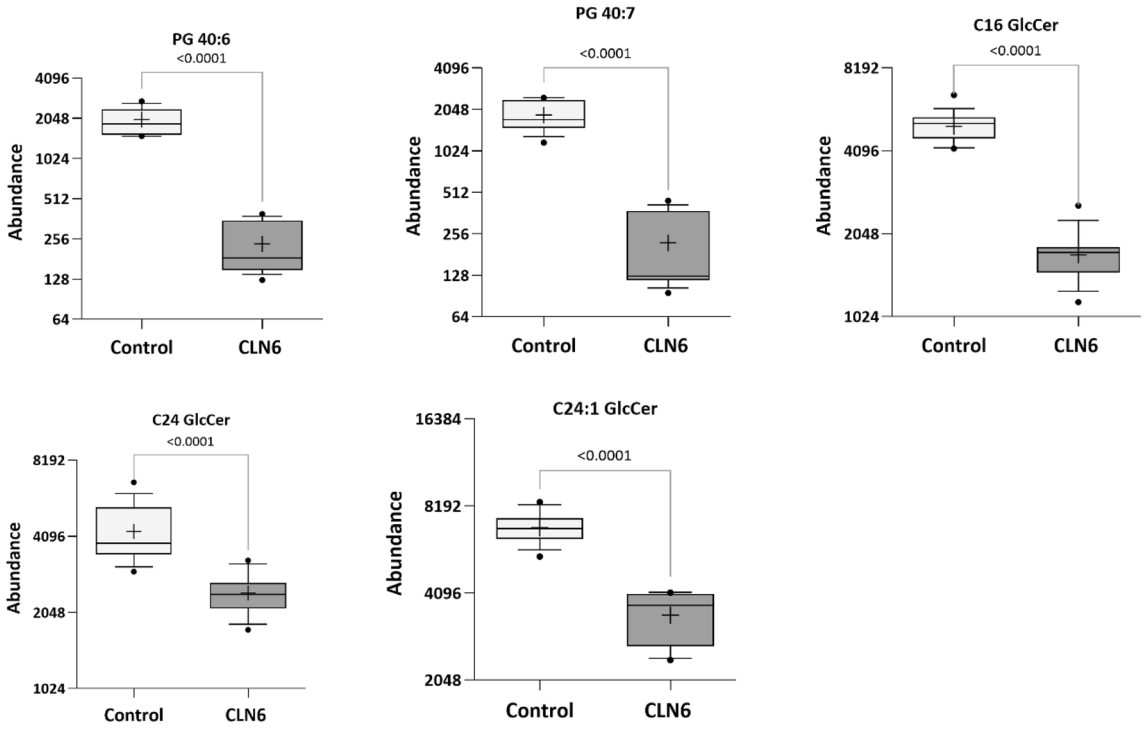


Figure 8. Potential metabolomic biomarkers for diagnosis of CLN6 disease. Representative Box plots showing the intensity of the top five metabolites in the two groups of neuronal-like cells [cell lines (n = 3) × biological replicas (n = 2–3) × technical replicas (n = 2)]. Whisker's end = the 10th and 90th percentile, bars = min and max values, horizontal line in the boxes = median value, '+' = mean. Dots represent the outliers.

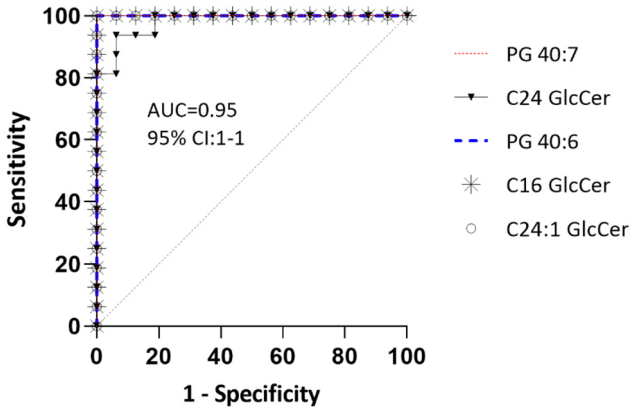


Figure 9 . The overlaid ROC (receiver operating characteristic) curves were used to visualize the metabolites with the maximum sensitivity for differentiating CLN6 disease subjects from controls. Four compounds had 100% overlapping ROC curves.

Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA)[28]. The pathway analysis module in the MetaboAnalyst software illustrates the most affected metabolic pathways of the screened metabolites, as shown in Fig. 10a, with the sphingolipid and glycerophospholipid metabolism pathways being the most significantly altered. Additionally, we utilized IPA to generate networks that allowed us to further explore and understand the biological networks involved in disease pathogenesis. The IPA analysis on differentiated neuronal-like cells from CLN6 patients was performed with 298 mapped metabolites, of which 158 were downregulated, and 140 were upregulated. Among them, 39 molecules showed significant regulation when applying a cut-off of ± 1.5-fold changes and a p-value ≤ 0.05 (28 downregulated and 11 upregulated). The list of mapped molecules by IPA can be found in Supplementary Table 4, while a summary of the obtained results is presented in Table 3.

The NAD signaling pathway showed significant association with the differentially expressed metabolites (p-value = 0.048, 5.3% overlap), and the deregulation of cyclic ADP-ribose (fold change = − 1.56, p-value = 0.00025). A regulator effect network was also identified (score 15) with associated cell signaling,

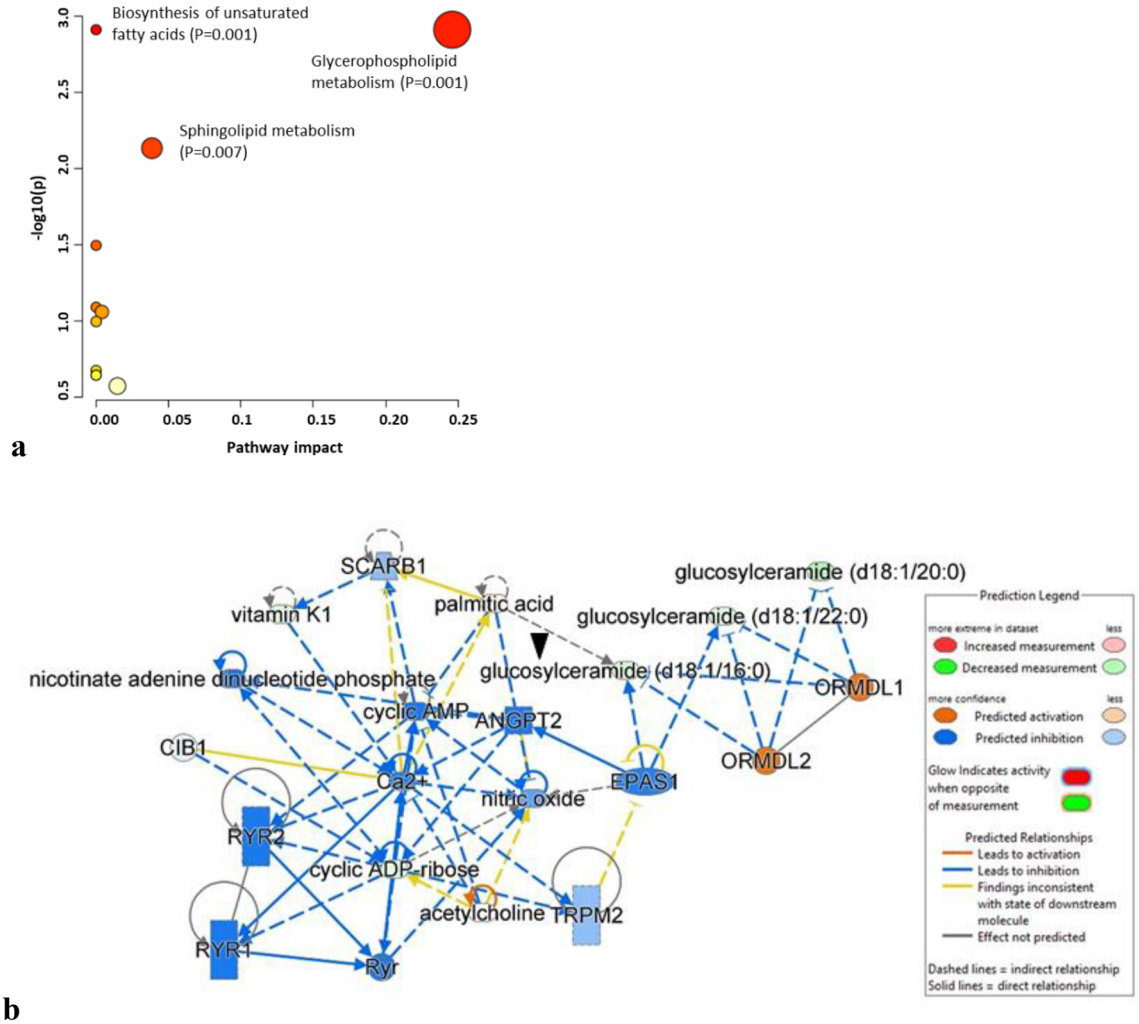


Figure 10. (**a**) The affected pathway and (**b**) pathway analysis in CLN6 patient samples as determined by MetaboAnalyst and Ingenuity Pathway Analysis software. The analysis reveals significant alterations in sphingolipid and glycerophospholipid metabolism and provides a network visualization of the interactions between different metabolic pathways and networks. The 'arrow' points to one of the metabolites among the five differentially expressed that is interconnected within the network.

molecular transport, vitamin, and mineral metabolism functions (Fig. 10b). Causal relationships were connected to the downregulated glucosylceramides via the upstream *ORMDL1* (p-value 8.08E−05) and *ORMDL2* (6.33E−05) sphingolipid biosynthesis master regulators linked to annotated functions such as the quantity of ceramide (p-value 1.25E−14), and synthesis of lipid (p-value 2.73E−08). Furthermore, the master regulator *EPAS1,* a transcription regulator often known as *HIF2A*, was also identified by IPA (p-value 2.54E−03, z-score − 1.41) in participating to the accumulation of nitric oxide (p-value 1.56E−05) both via the downregulation of glucosylceramides and cyclic ADP-ribose leading to the regulation of functions annotation like propagation of signaling of Ca^{2+} (p-value 0.00261) with the predicted downregulation of quantity of Ca^{2+} (p-value 5.91E−14), mobilization of Ca^{2+} (8.53E−07), and release of Ca^{2+} (p-value 5.11E−11). The participation of several upstream regulators, including the transporter *SCARB1* (p-value 0.0302), targeting the downregulated vitamin k1 in our dataset, the calcium and integrin binding 1 gene (*CIB1*, p-value 0.005), and the angiopoietin-2 protein coding gene (p-value 0.0252) have been identified by IPA as regulators of calcium homeostasis functions.

Discussion

To our knowledge, this study is the first to use neuronal progenitor-like cells differentiated from human CLN6 fibroblast lines to identify differentiating metabolites that can distinguish CLN6 disease from the healthy control group.

Research on CLN6 disease aiming to understand the disease pathophysiology and develop therapies has expanded significantly over the past decade. However, according to a literature review survey on the PubMed® database (https://pubmed.ncbi.nlm.nih.gov)[29], few studies aimed explicitly at metabolomics investigation of CLN6 disease, and the ones reported were model organisms-based using, for example, sheep and mouse models[23,24]. Nevertheless, these used GC–MS and NMR platforms and discovered an alteration of the

Canonical pathway	p-value	Ratio		
NAD signaling pathway	4.87E−02	0.0526		
Master regulators	**p-value**	**Target molecules in dataset**	**Activation**	**Network bias-corrected p-value**
ORMDL2	6.33E−05	Glucosylceramide (d18:1/16:0) glucosylceramide (d18:1/20:0) glucosylceramide (d18:1/22:0)	1732	4.00E−04
ORMDL1	8.08E−05	glucosylceramide (d18:1/16:0) glucosylceramide (d18:1/20:0) glucosylceramide (d18:1/22:0)	1732	4.00E−04
EPAS1	2.54E−03	Glucosylceramide (d18:1/16:0) glucosylceramide (d18:1/22:0)	− 1414	2.00E−04
SCARB1	3.02E−02	Vitamin K1	− 1	1.52E−02
ANGPT2	2.52E−02	Cyclic ADP-ribose	− 1	6.60E−03
CIB1	5.05E−03	Cyclic ADP-ribose	− 1	6.60E−03
acetylcholine	2.02E−02	Cyclic ADP-ribose	− 1	6.60E−03
Functions annotation	**Molecules**	**p-value**		
Quantity of ceramide	Ca^{2+}, EPAS1, GBA1, ORMDL1, ORMDL2, palmitic acid, PSAP, SCARB1	1.25E−14		
Quantity of Ca^{2+}	Acetylcholine, ANGPT2, Ca^{2+}, CIB1, cyclic ADP-ribose, cyclic AMP, nicotinate adenine dinucleotide phosphate, nitric oxide, palmitic acid, Ryr, RYR1, RYR2, vitamin K1	5.91E−14		
Release of Ca^{2+}	Ca^{2+}, cyclic ADP-ribose, cyclic AMP, nicotinate adenine dinucleotide phosphate, nitric oxide, Ryr, RYR1, RYR2, TRPM2	5.11E−11		
Mobilization of Ca^{2+}	Acetylcholine, Ca^{2+}, cyclic ADP-ribose, nicotinate adenine dinucleotide phosphate, palmitic acid, Ryr, SCARB1	8.53E−07		
Synthesis of lipid	Acetylcholine, ANGPT2, Ca^{2+}, cyclic AMP, GBA1, nitric oxide, ORMDL1, ORMDL2, palmitic acid, SCARB1	2.37E−08		
Accumulation of nitric oxide	Cyclic ADP-ribose, EPAS1	1.56E−05		
Propagation of signaling of Ca^{2+}	Cyclic ADP-ribose	2.61E−03		

Table 3. Summary of IPA analysis.

glutamine-glutamate metabolism and a decrease of GABA in their quest for altered metabolic pathways that lead to neuronal degeneration.

Given that no human CLN6 studies have been published that addressed the metabolomic changes in the cell lines of CLN6 patients, we aimed to analyze and compare the changes in the global metabolome of the induced neuronal-like cell lines from humans with CLN6 disease to those of healthy subjects. For this, an LC–MS metabolomics approach was employed to identify the metabolic pathways altered in CLN6 subjects and acquire further insights into potential markers of disease pathogenesis. The metabolic profiling analyses were performed on induced neuronal-like cell lines from subjects with CLN6 disease and human controls. Fifteen metabolites were markedly downregulated in CLN6 subjects and showed a robust discriminatory power between the CLN6 and the control group.

Results of pathway identification performed with the MetaboAnalyst *v* 5.0 software and IPA analysis revealed alterations in molecules relevant to sphingolipid and glycerophospholipid metabolism. Although no research has yet addressed the role of sphingolipids in the CLN6 disease, our findings imply that disruptions in sphingolipid metabolism are a feature of the CLN6 disease, which could represent a source for future biomarker discovery. Among the dysregulated metabolites related to CLN6 disease were five distinct metabolites with a VIP score greater than 1. They were represented by the two glycerophospholipids, PG 40:6 and PG 40:7, and the three sphingolipids, C16 GlcCer, C24 GlcCer, and C24:1 GlcCer.

Studies on Alzheimer's disease subjects proved that altered glycerophospholipids levels might lead to neuronal damage, neuroinflammation[30,31], and neurodegeneration[32,33], features that resemble the NCL disorder[3]. Additionally, glycerophospholipids were proposed as putative biomarkers in neurodegenerative diseases[34]. Besides glycerophospholipids, dysregulated glucosylceramide levels have also been linked with neurodegeneration in lysosomal storage disorders (LSDs)[35]. Other investigations have shown a connection between increased glucosylceramide levels, neuroinflammation, and neuronal loss[36].

Similarly, our study identified two distinct glycerophospholipids components (PG 40:6 and PG 40:7) with statistically low abundance in the CLN6 group. Moreover, the compounds with the most discriminatory power between CLN6 and the control group were C24:1 GlcCer and C16 GlcCer, highlighting their distinctive importance in the dysregulation of the sphingolipid metabolic pathways for CLN6 disease. However, in contrast to the studies mentioned above, where the levels were elevated, our results showed that glucosylceramides were downregulated in CLN6 compared to the healthy control group. Equivalent results were reported in a CLN9-deficient

cells-based study, where diminished levels of ceramide, glucosylceramide, and other sphingolipids components were observed[37].

Concerning the NCLs group, perturbed sphingolipid metabolism was described in various NCLs forms. One of the first studies that mentioned dysregulation in phospholipid metabolism was done on CLN1 and CLN3 disease patients[38]. Later investigations considered that the perturbed sphingolipid metabolism may represent a link between some forms of NCLs[39]. According to the study conducted in CLN3-defective cells, a decrease in various sphingolipids, such as lactosylceramides and glycosphingolipids, and up to a 60% reduction in the level of various HexCer components as compared to the control group was revealed[40]. A more recent study on CLN5 disease reported a similar outcome by exploring the connection between CLN5 disease and the degradation of sphingolipid metabolism[41]. These findings infer that CLN3 and CLN5 play a critical role in the changes in sphingolipid metabolism. Altogether, several studies on infantile and late infantile variants mentioned abnormal lipid metabolism[3], and others noted changes in the composition of various phospholipid and ceramides classes[38,42,43] (Table 4), yet our work is the first to link the sphingolipid metabolism to CLN6. Furthermore, the role of calcium signaling has been investigated in several models showing an important role of calcium homeostasis in NCL pathology[44,45] and elevated calcium-binding protein calbindin 1 (CALB1) levels in cerebrospinal fluid from CLN2 and CLN3 disease patients[46]. Similarly, our study found that the downregulation of glucosylceramides, cyclic ADP-ribose, and vitamin K1 was associated with the inhibition of calcium signaling in CLN6 disease.

In summary, the current study compared the metabolomic profile of CLN6-neuronal progenitor-like cells derived from fibroblast to the control group. Our findings showed that five metabolites were significantly dysregulated in the cell lines from CLN6 subjects and may be considered potential candidate biomarkers for CLN6 disease. Additionally, the metabolic pathway analysis suggests the involvement of the sphingolipid, glycerophospholipid metabolic pathway, and calcium signaling in the mechanism behind the CLN6 disease progression, which is oriented toward the downregulation of sphingolipids and that of glycerophospholipids metabolism.

While our findings enhanced our understanding of the metabolomics of the CLN6 disease, scale-up research involving additional cell lines and diverse patient cohorts is needed to validate the observed pattern in our data. The corroboration of our findings would pave the way for advanced metabolomics studies of CLN6 disease that may uncover potential therapeutic targets of CLN6 disease.

Given the significant milestones achieved in our study, which underscore the potential of ciNPCs within an LC–MS-based metabolomics approach for biomarker discovery, we recognize the need to address the following key research directions: (i) expand and diversify the cohort to encompass a broader range of ages and genders to ensure a more comprehensive dataset and greater generalizability of our findings; (ii) incorporate additional

NCL type	Lipid class	Lipid species	Level	Refs
CLN1	Phospholipids	LBPA (36:2; 34:1); PC (32:1, 34:1)	High	37–39,42
		PC (32:1, 34:1); PI (36:3, 38:5, 38:3); PC (32:1, 34:1)		
		PE (38:5, 36:4, 38:5, 36:2, 34:1, 34:2)		
		PS (34:4, 38:3, 40:3, 34:1, 36:1, 36:2, 38:2)		
		LBPA (38:1); PC (38:4, 38:6)	Low	
		PE (40: 6, 38:4, 40:4); PI (38:4, 40:6); PS (40:6)		
	Sphingolipids	SM (16:0, 16:1, 18:1)	High	38
		SM (20:0)	Low	
CLN2	Phospholipids	GPE (18:1)	High	43
		GPE (16:0, 18:0)	Low	
CLN3	Phospholipids	LPE (20:3); PC (32:1, 34:1, 36:3)	High	37,40
		PS (38:3); PI (38:3); PE (38:3, 40:3)		
		PA (36:1, 36:2); PC (38:4); LPI (18:0, 20:4); LPC (20:4)	Low	
		PG (34:2, 34:1); PE (40:4); PI (36:1, 36:2, 36:3, 36:4, 36:5)		
	Sphingolipids	Cer (16:0, 24:0, 24:1); SM (24:1)	High	39,40
		GM3 (d18:1/24:1, d18:1/16:0, d18:1/24:0)		
		GM3 (d18:1/24:1, d18:1/16:0, d18:1/24:0)		
		GD1 (d18:1/25:0, d18:1/16:0, d18:1/22:0); SM (14:0, 15:0)	Low	
		HexCer (22:0, 20:0, 18:0, 23:0); LacCer (16:0, 24:0)		
	Sterols	CE (18:2, 18:3)	Low	40
CLN9	Sphingolipids	Cer (16:0, 24:0, 24:1)	Low	37,39
		dhCer (16:0, 24:0, 24:1)		

Table 4. List of lipid species involved in dysregulated lipid metabolism in NCL disease. Cer: ceramide; CholE: cholesterol Ester; dhCer: dihydroceramides; GD1: monosialoganglioside D1; GM3: monosialoganglioside GM3; GPE: glycerophosphoethanolamine; HexCer: hexosylceramide; LacCer: lactosylceramide; LBPA: lysobisphosphatidic acid; LPC: lysophosphatidylcholine; LPE: lysophosphatidylethanolamine; LPI: lysophosphatidylinositol; PA: phosphatidic acid; PC: phosphatidylcholine; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; PI: phosphatidylinositol; PS: phosphatidylserine; SM: sphingomyelin.

NCL disease subtypes for comparative analysis thus assessing the specificity and relevance of potential biomarkers across diverse subtypes; (iii) undertake targeted MS/MS research, building upon the spectral features reported in this study to deepen our understanding of the involved metabolites and pathways, ultimately amplifying the precision and impact of our research.

The findings presented in this study offer valuable insights into the metabolic changes associated with CLN6 disease. However, further research is recommended to understand our results' broader implications fully. To this end, we propose several directions for future investigations:

1. Incorporating cellular vitality and stability assessments into upcoming research to enhance metabolomics data precision and the reliability of biomarker discovery studies. While the primary goal of this study was to assess the feasibility of ciNPC cells for identifying CLN6 biomarkers through metabolomics analysis, it is crucial to consider the potential impact of cellular vitality and health on metabolomics profiling, potentially affecting the detection of subtle disease- or treatment-related effects. Therefore, future study endeavors should expand their scope to include an assessment of cellular health and stability, such as assessing DAPI+ spots to evaluate vitality and cell death rates and employing Western blot analysis to evaluate protein expression changes as well as quantitative staining. These methods provide valuable insights into ciNPCs, ultimately enhancing understanding of cellular pathophysiology, particularly within the context of biomarker discovery.
2. Validation studies involving fibroblast cells under non-induced conditions are essential to ensure the reliability of the identified metabolites in CLN6 disease. This additional validation step will offer valuable insights into the accuracy and specificity of the identified metabolites as potential biomarkers, both in cost-effective cells and in minimal-invasive specimens. Thereby, it will significantly enhance the potential of our research findings for future applications.
3. Conducting targeted experiments, such as mass spectrometry-based proteomics, to identify and validate the putative interactors and regulators predicted by in-silico analysis will yield robust evidence to support the functional significance in the context of CLN6 disease.

To conclude, this study is the first to examine the metabolome of human CLN6 ciNPCs, which provides metabolomics insight into the pathogenesis of CLN6 disease. Over 2700 spectral features were relatively quantified, two altered pathways were determined, and five putative biomarker candidates were identified. Altogether, this demonstrates a solid basis for the applicability of LC–MS-based metabolomics, which ultimately, could lead to an earlier diagnosis and prognosis of CLN6 disease and ease monitoring the effectiveness of upcoming therapeutic trials.

Materials and methods

Cell lines

The biobank "Cell Line and DNA Bank of Genetic Movement Disorders and Mitochondrial Diseases", a member of the Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy, and the EuroBioBank Network[47] provided us with the CLN6 fibroblast lines [CLF064, CLF121, and CLF210]. The following cell lines (wild type) were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: [GM0839, GM0565, and GM0203].

Chemicals and reagents

Ultra-high-performance liquid chromatography (UHPLC), grade acetonitrile (ACN), formic acid (FA), and methanol (MeOH) were purchased from Biosolve (Dieuze, France). Water LC–MS grade was purchased from VWR (Darmstadt, Germany). DMEM high-glucose medium (Gibco, Grand Island, NY), DMEM/F12 medium (12634010, Gibco), neurobasal medium (21103049), fetal bovine serum (FBS) (26140079), and 1% penicillin–streptomycin solution (15140122) were purchased from Thermo Fisher Scientific. Phosphate-buffered saline (10×) (PBS) (AM9624) was purchased from Invitrogen. B-27™ Supplement (50×) (10828010, Gibco), N2 Supplement (100×) (17502048, Gibco), and L-Glutamine (100×) (25030081, Gibco) were obtained from Life Technologies (Grand Island, New York, USA). CHIR99021 (130106539), SB431542 (131106275) were from Miltenyi Biotec (Teterow, Germany) and PD0325901 (PZ0162), LDN193189 (SML0559), Pifithrin-a (P4359) and Forskolin (F3917) were all from Sigma Aldrich (Taufkirchen, Germany).

Cell culture

The fibroblast lines were maintained at 37 °C and 5% CO_2 in high glucose Dulbecco's modified Eagle medium (DMEM) (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), supplemented with 10% fetal bovine serum (FBS) (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), 100 U/mL penicillin, and 10 µg/ml streptomycin until they reached 90% confluence.

Generation of chemical-induced neuronal progenitor cells

The direct chemical conversion of fibroblasts into neuronal-like cells was performed according to a previously published methodology by Dai et al.[26]. According to the protocol, once the cells reach the desired confluency, they are further switched into a neuronal medium made of a mixture of one-part DMEM/F12 (1% N2 supplement, Gibco) and neurobasal medium (2% B27 supplement, Gibco), and another part was a cocktail made of six chemicals (*v*1: *v*1). The chemicals known to aid in the NPC differentiation were represented by SB431542 (2 µM, TGF-β inhibitor), CHIR99021 (1 µM, GSK3b inhibitor), PD0325901 (1 µM, MAPK inhibitor), LDN193189

(1 µM, BMP inhibitor), Pifithrin-α (5 µM, p53 inhibitor), and Forskolin (7 µM, cAMP activator). The cells were cultured until day 21, and the medium changed every third day.

Immunofluorescence

Cells were washed with PBS and fixed with 4% paraformaldehyde (Sigma-Aldrich, St. Louis, MO) for 15 min, then permeabilized in PBS containing 0.3% (w/v) Triton X-100 for 10 min, lastly blocked in PBS containing 4% (w/v) BSA for 2 h. Cells were incubated with the following primary antibodies (1:250 dilution) for 2 h at room temperature): anti-βIII-tub (Biolegend, cat. 801201), anti-MAP2ab (Merck Millipore, cat. AB5622), anti-GFAP (Biolegend, cat. 83721), and anti-GABA (Sigma Aldrich, cat. A2052). Subsequently, the cells were rinsed three times with 0.1% (w/v) BSA in PBS-Tr and incubated with the secondary antibody (1:500 dilution) for 1 h at room temperature in the dark (Alexa Fluor 488 A11029, and Alexa Fluor 568, A11036, Invitrogen). Nuclei were counterstained with DAPI (Invitrogen, cat. 1:10000) for 45 min. Cell images were acquired using a Keyence fluorescence microscope BZ-X710E equipped with the BZ-X800 Analyzer software (Keyence, Osaka, Japan) with a 20X Plan-Apo Gamma NA 0.75 objective and fluorescence filter set for GFP, TRITC, and DAPI.

Sample preparation and metabolites extraction

While the adherent cell plates were kept on dry ice, the medium was removed, and the cells were quickly rinsed with 1 mL of 0.9% NaCl (4 °C) (Baxter, Sydney, Australia) to remove extracellular metabolites. A 600 µL extraction solvent (methanol: water, 3:1 *v/v*), prechilled in a − 80 °C freezer for at least 1 h, was added to the cells, and the cells detached using a scraper while the plates were kept on dry ice. The cell suspension was transferred into a 1.5 mL Eppendorf tube containing 10 µL internal standard prepared using 200 ng/mL Lyso-Gb2 (Matreya LLC, State College, PA, USA) dissolved in methanol. The mixture of cells, extraction solvent, and internal standard (IS) was vortexed vigorously. It was then centrifuged at 14,000×*g* for 20 min at 4 °C. A volume of 150 µL supernatant was transferred into an LC–MS glass vial. The quality control (QC) was generated by pooling 5 µL of each sample. Blank samples consisted of 100% LC–MS water. Before injecting the standard samples, blank and pooled samples were injected five times each in the beginning to establish system equilibrium. Throughout the batch, 5 µL of the pool, blanks, and standard samples were injected intermittently during the run to ensure the stability of the LC–MS system.

Chromatographic and mass spectrometric conditions

Mass spectrometry was performed on a Waters® i-Class ACQUITY UPLC (Waters, Borehamwood, UK) coupled to a Vion™ IMS Q-Tof™mass spectrometer (Waters, Borehamwood, UK) equipped with an ESI ion source, system operating in negative (ESI −) ionization mode. The LC–MS method was previously reported[48]. It was based on a 5 µL aliquot extract injected into a Kinetex EVO (C18, 2.1 × 150 mm, 5 µm) LC column (Phenomenex, Aschaffenburg, Germany) preheated to 50 °C at a flow rate of 0.5 mL/min. Analytes were eluted using a linear gradient ranging from 1 to 100% B (50 mM formic acid in methanol: acetonitrile 1:1 *v/v*) and A (50 mM formic acid in water). The following settings were used for mass spectrometric acquisition: High Definition MS^E ($HDMS^E$), capillary voltage 1.2 kV, source temperature 150 °C, desolvation temperature 600 °C, desolvation gas 1000 L/h, cone gas 50 L/h, low collision energy 6 eV, high collision energy ramp 20–40 eV, scan mass 50–1000 m/*z*, scan time 0.5 s. Each signal had three identifiers: retention time in min (RT), ion mass (m/z), and CCS (collision cross-section). Leucine-enkephalin (Sigma-Aldrich, Taufkirchen, Germany) (1 ng/µL) was used as a lock mass reference compound *([M−H]− =* 554.2615, negative ion mode).

Metabolomic data processing

The raw MS data were acquired using Unifi software v1.9 (Waters, Borehamwood, UK) and exported as Unify export packages (.uep). The generated datasets were imported to Progenesis QI software v 3 (Nonlinear Dynamics, Newcastle upon Tyne, UK) for automatic data processing. The following steps were part of the data processing and analysis workflow: retention time correction, experimental design setup, peak picking, probabilistic quotient normalization (PQN)[49], deconvolution, and compound identification. The metabolites were individually assessed for statistical relevance and robustness. Only the variables that met the following quality filters were selected: significant difference between the control and CLN6 disease group ($p \le 0.05$), fold change at least twofold, charge ≤ 3, mass-to-charge ratio (m/z) ≥ 179, and a median normalized abundance ≥ 100 counts relative to the reference compound in at least one of the cohorts. The peak intensities of the selected compounds were transformed into .csv files and uploaded into the 'Statistical Analysis' toolbox of MetaboAnalyst v5.0 at http://www.metaboanalyst.ca[27]. Canonical pathway analysis was conducted using Ingenuity Pathway Analysis (IPA) software from QIAGEN (Ingenuity Systems, QIAGEN, Redwood City, CA, USA) with 298 mapped molecules by IPA using either Human Metabolome database (HMBD) or CAS registry number, or PubChem CIS IDs. A cut-off of ± 1.5-fold changes and p-value ≤ 0.05 was applied. The Euclidean distance metric and the 'Ward' clustering algorithm were used to create dendrograms. Heatmap with enforced sample grouping displayed value distributions and ranges.

Metabolite database searching

Metabolites were identified based on monoisotopic mass, retention time, and collision cross-section. The obtained features were matched against several metabolite databases. Our in-house compound library, Human Metabolome Database[50], PubChem[51], ChemSpider[52], and LIPID MAPS® Structure Database (LMSD) were among the databases used in this study.

Statistical analysis

Multivariate analysis of LC–MS data and pathway analysis were performed using the open-source software MetaboAnalyst 5.0[27]. The box-and-whisker plots, ROC, and volcano plots were generated using the GraphPad Prism (version 9.5.0) software (GraphPad Software, Inc., San Diego, CA, http://www.graphpad.com). A student's t-test was applied to identify with a 95% confidence level and 5% false positive (false discovery rate, FDR). The level of significance was set at $p \leq 0.05$.

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 14 March 2023; Accepted: 24 October 2023
Published online: 29 October 2023

References

1. Mole, S., Williams, R. & Goebel, H. *The Neuronal Ceroid Lipofuscinoses (Batten Disease)* (Oxford University Press, 2011).
2. Haltia, M. & Goebel, H. H. The neuronal ceroid-lipofuscinoses: A historical introduction. *Biochim. Biophys. Acta* **1795–1800**, 2013. https://doi.org/10.1016/j.bbadis.2012.08.012 (1832).
3. Palmer, D. N., Barry, L. A., Tyynela, J. & Cooper, J. D. NCL disease mechanisms. *Biochim. Biophys. Acta* **1832**, 1882–1893. https://doi.org/10.1016/j.bbadis.2013.05.014 (2013).
4. Santorelli, F. M. *et al.* Molecular epidemiology of childhood neuronal ceroid-lipofuscinosis in Italy. *Orphanet. J. Rare Dis.* **8**, 19. https://doi.org/10.1186/1750-1172-8-19 (2013).
5. Sleat, D. E., Gedvilaite, E., Zhang, Y., Lobel, P. & Xing, J. Analysis of large-scale whole exome sequencing data to determine the prevalence of genetically-distinct forms of neuronal ceroid lipofuscinosis. *Gene* **593**, 284–291. https://doi.org/10.1016/j.gene.2016.08.031 (2016).
6. Williams, R. E. & Mole, S. E. New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. *Neurology* **79**, 183–191. https://doi.org/10.1212/WNL.0b013e31825f0547 (2012).
7. Butz, E. S., Chandrachud, U., Mole, S. E. & Cotman, S. L. Moving towards a new era of genomics in the neuronal ceroid lipofuscinoses. *Biochim. Biophys. Acta. Mol. Basis Dis.* **1866**, 165571. https://doi.org/10.1016/j.bbadis.2019.165571 (2020).
8. Wheeler, R. B. *et al.* The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (CLN6) and in nclf mutant mice encodes a novel predicted transmembrane protein. *Am. J. Hum. Genet.* **70**, 537–542. https://doi.org/10.1086/338708 (2002).
9. Mole, S. E. *et al.* CLN6, which is associated with a lysosomal storage disease, is an endoplasmic reticulum protein. *Exp. Cell Res.* **298**, 399–406. https://doi.org/10.1016/j.yexcr.2004.04.042 (2004).
10. Schulz, A., Kohlschutter, A., Mink, J., Simonati, A. & Williams, R. NCL diseases—Clinical perspectives. *Biochim. Biophys. Acta* **1801–1806**, 2013. https://doi.org/10.1016/j.bbadis.2013.04.008 (1832).
11. Dozieres-Puyravel, B. *et al.* Paediatric-onset neuronal ceroid lipofuscinosis: First symptoms and presentation at diagnosis. *Dev. Med. Child Neurol.* **62**, 528–530. https://doi.org/10.1111/dmcn.14346 (2020).
12. Rosenberg, J. B., Chen, A., Kaminsky, S. M., Crystal, R. G. & Sondhi, D. Advances in the treatment of neuronal ceroid lipofuscinosis. *Expert Opin. Orphan Drugs* **7**, 473–500. https://doi.org/10.1080/21678707.2019.1684258 (2019).
13. Johnson, T. B. *et al.* Therapeutic landscape for Batten disease: Current treatments and future prospects. *Nat. Rev. Neurol.* **15**, 161–178. https://doi.org/10.1038/s41582-019-0138-8 (2019).
14. Kohlschutter, A., Schulz, A., Bartsch, U. & Storch, S. Current and emerging treatment strategies for neuronal ceroid lipofuscinoses. *CNS Drugs* **33**, 315–325. https://doi.org/10.1007/s40263-019-00620-8 (2019).
15. Schulte, P. A. The use of biomarkers in surveillance, medical screening, and intervention. *Mutat. Res.* **592**, 155–163. https://doi.org/10.1016/j.mrfmmm.2005.06.019 (2005).
16. Ozkan, S. A., Bakirhan, N. K. & Mollarasouli, F. *The Detection of Biomarkers: Past, Present, and the Future Prospects* (Elsevier Science, 2021).
17. Trivedi, D. K., Hollywood, K. A. & Goodacre, R. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horiz. Transl. Med.* **3**, 294–305. https://doi.org/10.1016/j.nhtm.2017.06.001 (2017).
18. Yang, Q. *et al.* Metabolomics biotechnology, applications, and future trends: A systematic review. *RSC Adv.* **9**, 37245–37257. https://doi.org/10.1039/c9ra06697g (2019).
19. Johnson, C. H. & Gonzalez, F. J. Challenges and opportunities of metabolomics. *J. Cell Physiol.* **227**, 2975–2981. https://doi.org/10.1002/jcp.24002 (2012).
20. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459. https://doi.org/10.1038/nrm.2016.25 (2016).
21. Kline, R. A., Wishart, T. M., Mills, K. & Heywood, W. E. Applying modern Omic technologies to the Neuronal Ceroid Lipofuscinoses. *Biochim. Biophys. Acta Mol. Basis Dis.* **1866**, 165498. https://doi.org/10.1016/j.bbadis.2019.06.012 (2020).
22. Sindelar, M. *et al.* Untargeted metabolite profiling of cerebrospinal fluid uncovers biomarkers for severity of late infantile neuronal ceroid lipofuscinosis (CLN2, Batten disease). *Sci. Rep.* **8**, 15229. https://doi.org/10.1038/s41598-018-33449-0 (2018).
23. Pears, M. R. *et al.* Metabolomic investigation of CLN6 neuronal ceroid lipofuscinosis in affected South Hampshire sheep. *J. Neurosci. Res.* **85**, 3494–3504. https://doi.org/10.1002/jnr.21343 (2007).
24. Salek, R. M. *et al.* A metabolomic comparison of mouse models of the Neuronal Ceroid Lipofuscinoses. *J. Biomol. NMR* **49**, 175–184. https://doi.org/10.1007/s10858-011-9491-7 (2011).
25. Hansson, O. Biomarkers for neurodegenerative diseases. *Nat. Med.* **27**, 954–963. https://doi.org/10.1038/s41591-021-01382-x (2021).
26. Dai, P., Harada, Y. & Takamatsu, T. Highly efficient direct conversion of human fibroblasts to neuronal cells by chemical compounds. *J. Clin. Biochem. Nutr.* **56**, 166–170. https://doi.org/10.3164/jcbn.15-39 (2015).
27. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, W388–W396. https://doi.org/10.1093/nar/gkab382 (2021).
28. Kramer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530. https://doi.org/10.1093/bioinformatics/btt703 (2014).
29. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **49**, D10–D17. https://doi.org/10.1093/nar/gkaa892 (2021).
30. Raszewski, G., Chwedorowicz, R., Chwedorowicz, A. & Gustaw Rothenberg, K. Homocysteine, antioxidant vitamins and lipids as biomarkers of neurodegeneration in Alzheimer's disease versus non-Alzheimer's dementia. *Ann. Agric. Environ. Med.* **23**, 193–196. https://doi.org/10.5604/12321966.1196878 (2016).
31. Russo, R. *et al.* Gut–brain axis: Role of lipids in the regulation of inflammation, pain and CNS diseases. *Curr. Med. Chem.* **25**, 3930–3952. https://doi.org/10.2174/0929867324666170216113756 (2018).

32. Dorninger, F., Forss-Petter, S., Wimmer, I. & Berger, J. Plasmalogens, platelet-activating factor and beyond—Ether lipids in signaling and neurodegeneration. *Neurobiol. Dis.* **145**, 105061. https://doi.org/10.1016/j.nbd.2020.105061 (2020).
33. Hernandez-Diaz, S. & Soukup, S. F. The role of lipids in autophagy and its implication in neurodegeneration. *Cell Stress* **4**, 167–186. https://doi.org/10.15698/cst2020.07.225 (2020).
34. Huh, Y. E. *et al.* Glucosylceramide in cerebrospinal fluid of patients with GBA-associated and idiopathic Parkinson's disease enrolled in PPMI. *NPJ Parkinsons Dis.* **7**, 102. https://doi.org/10.1038/s41531-021-00241-3 (2021).
35. Farfel-Becker, T. *et al.* Neuronal accumulation of glucosylceramide in a mouse model of neuronopathic Gaucher disease leads to neurodegeneration. *Hum. Mol. Genet.* **23**, 843–854. https://doi.org/10.1093/hmg/ddt468 (2014).
36. Korkotian, E. *et al.* Elevation of intracellular glucosylceramide levels results in an increase in endoplasmic reticulum density and in functional calcium stores in cultured neurons. *J. Biol. Chem.* **274**, 21673–21678. https://doi.org/10.1074/jbc.274.31.21673 (1999).
37. Schulz, A. *et al.* The CLN9 protein, a regulator of dihydroceramide synthase. *J. Biol. Chem.* **281**, 2784–2794. https://doi.org/10.1074/jbc.M509483200 (2006).
38. Kakela, R., Somerharju, P. & Tyynela, J. Analysis of phospholipid molecular species in brains from patients with infantile and juvenile neuronal-ceroid lipofuscinosis using liquid chromatography-electrospray ionization mass spectrometry. *J. Neurochem.* **84**, 1051–1065. https://doi.org/10.1046/j.1471-4159.2003.01602.x (2003).
39. Persaud-Sawin, D. A. *et al.* Neuronal ceroid lipofuscinosis: A common pathway?. *Pediatr. Res.* **61**, 146–152. https://doi.org/10.1203/pdr.0b013e31802d8a4a (2007).
40. Schmidtke, C. *et al.* Lysosomal proteome analysis reveals that CLN3-defective cells have multiple enzyme deficiencies associated with changes in intracellular trafficking. *J. Biol. Chem.* **294**, 9592–9604. https://doi.org/10.1074/jbc.RA119.008852 (2019).
41. Doccini, S. *et al.* Lysosomal proteomics links disturbances in lipid homeostasis and sphingolipid metabolism to CLN5 disease. *Cells* https://doi.org/10.3390/cells11111840 (2022).
42. Goswami, R. *et al.* Differential regulation of ceramide in lipid-rich microdomains (rafts): Antagonistic role of palmitoyl:protein thioesterase and neutral sphingomyelinase 2. *J. Neurosci. Res.* **81**, 208–217. https://doi.org/10.1002/jnr.20549 (2005).
43. Granier, L. A., Langley, K., Leray, C. & Sarlieve, L. L. Phospholipid composition in late infantile neuronal ceroid lipofuscinosis. *Eur. J. Clin. Investig.* **30**, 1011–1017. https://doi.org/10.1046/j.1365-2362.2000.00757.x (2000).
44. Bosch, M. E. & Kielian, T. Astrocytes in juvenile neuronal ceroid lipofuscinosis (CLN3) display metabolic and calcium signaling abnormalities. *J. Neurochem.* **148**, 612–624. https://doi.org/10.1111/jnc.14545 (2019).
45. Mathavarajah, S., O'Day, D. H. & Huber, R. J. Neuronal ceroid lipofuscinoses: Connecting calcium signalling through calmodulin. *Cells* https://doi.org/10.3390/cells7110188 (2018).
46. Sleat, D. E. *et al.* Proteomic analysis of brain and cerebrospinal fluid from the three major forms of neuronal ceroid lipofuscinosis reveals potential biomarkers. *J. Proteome Res.* **16**, 3787–3804. https://doi.org/10.1021/acs.jproteome.7b00460 (2017).
47. Filocamo, M. *et al.* Telethon Network of Genetic Biobanks: A key service for diagnosis and research on rare diseases. *Orphanet. J. Rare Dis.* **8**, 129. https://doi.org/10.1186/1750-1172-8-129 (2013).
48. Cozma, C. *et al.* C26-Ceramide as highly sensitive biomarker for the diagnosis of Farber disease. *Sci. Rep.* **7**, 6149. https://doi.org/10.1038/s41598-017-06604-2 (2017).
49. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **78**, 4281–4290. https://doi.org/10.1021/ac051632c (2006).
50. Wishart, D. S. *et al.* HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **50**, D622-d631. https://doi.org/10.1093/nar/gkab1062 (2022).
51. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkac1032 (2022).
52. Pence, H. E. & Williams, A. ChemSpider: An online chemical information resource. *J. Chem. Educ.* **87**, 1123–1124. https://doi.org/10.1021/ed100697w (2010).

Acknowledgements

The authors would like to thank the Cell Line and DNA Bank of Genetic Movement Disorders and Mitochondrial Diseases, a member of the Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy, and the EuroBioBank Network for providing the CLN6 human fibroblasts lines. We also thank the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research for providing the control human fibroblasts. Furthermore, we acknowledge Sebastian Joost and Zhu Guanchen (Centogene GmbH) for laboratory and data access support.

Author contributions

C.M.R., Y.A. and D.L.P. conception and design, C.M.R. and D.L.P. experimental work, C.M.R. and S.D.B. data analysis and figures creation, S.F., J.K. and G.H. critical revision, C.C. and R.Z. supervision, P.B. project administration and final approval of the manuscript. C.M.R. wrote the manuscript, which was improved by all co-authors. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of their work.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-45789-7.

Correspondence and requests for materials should be addressed to C.-M.R.