Integration and comparison of multi-omics profiles of NGLY1 deficiency plasma and cellular models to identify clinically relevant molecular phenotypes

Songjie Chen[1,2], Guangwen Wang[1,2], Xiaotao Shen[1,2], Daniel Hornburg[1], Rene Hoffman[1], Stephanie Nevins[1], Xun Cheng[1], Michael Snyder[1,3]

[1] Department of Genetics, Stanford University, Stanford, CA, USA.

[2] These authors contributed equally: Songjie Chen, Guangwen Wang, and Xiaotao Shen.

[3] Corresponding author: Michael Synder, email: mpsnyder@stanford.edu

Author Contributions

S.C. and G.W. designed and managed the study; S.C. performed metabolomics and proteomics profiling of plasma and cell lysate samples, analysis and interpretation of data, drafting of the manuscript; G.W., R.H., S.N., X.C. generated the disease modelling iPSCs and NPCs; X.S. performed the annotation of metabolites and visualization of results. D.H. performed lipidomics profiling of plasma samples. G.W. and M.S. reviewed and edited the manuscript; M.S. supervised the project.

Ethics declarations

Competing interests

The authors declare no competing interests.

Acknowledgements

We thank the following people for their essential contributions to this paper: Jing Liang and Yael Rosenberg-Hasson at Stanford Human Immune Monitoring Center (HIMC) for cytokine quantification, Lulu Cao and Christine Yeh for the pilot test of iPSC proteomics analysis. Dan Li, Yu Zhang and Liang Liang for reviewing the manuscripts. This work was supported by Grace Science Foundation.

Abstract

NGLY1 (N-glycanase 1) deficiency is a rare congenital recessive disorder caused by a mutation in the *NGLY1* gene, which encodes the cytosol enzyme N-glycanase 1. The NGLY1 protein catalyzes the first step in protein deglycosylation, a process prerequisite for the cytosolic degradation of misfolded glycoproteins. By performing and combining metabolomics and proteomics profiles, we showed that NGLY1 deficiency induced the activation of immune response, disturbed lipid metabolism, biogenic amine synthesis and glutathione metabolism. The discovery was further validated by profiling of patient-derived induced pluripotent stem cells (iPSCs) and differentiated neural progenitor cells (NPCs), which serve as a personalized cellular model of the disease. This study provides new insights into the NGLY1 deficiency pathology and demonstrates that the upregulation of immune response and downregulation of lipid metabolism appear to be important molecular phenotypes of NGLY1 deficiency, together with the dysregulation of amino acid metabolism, biogenic amine synthesis and diverse signaling pathways, likely causing broad downstream syndromes. Collectively, such valuable multi-omics profiles identified broad molecular associations of potential pathological mechanisms during the onset of NGLY1 deficiency and suggested potential therapeutic targets for researchers and clinicians.

Key words: NGLY1, metabolomics, proteomics, immunoassay, lipidomics, iPSC, NPC, CRISPR

Main

Introduction

NGLY1 deficiency is an inherited recessive disorder, which gives rise to complex symptoms, including global developmental delay, delayed bone age, movement disorder, alacrimia, peripheral neuropathy, hepatopathy, and seizures (PMID: 24651605, 27388694, 31965062, 33673403, 32071843, 25220016,). Symptoms arise from infancy and their severity varies among individual patients. Most patients do not survive past childhood (PMID:

33497766, 31957011). A therapeutic has not been developed. As revealed by a whole exome sequencing, the disorder is caused by a mutation in *NGLY1* gene, which encodes the cytosolic enzyme N-glycanase1 (NGLY1) responsible for protein deglycosylation, a process prerequisite for the endoplasmic reticulum associated degradation (ERAD) of glycoproteins (PMID: 24651605, 26611529, 12606569, 32395402, 12711318). At the molecular level, N-linked glycosylation directs the proper folding of glycoproteins during their insertion into the endoplasmic reticulum (ER) (PMID: 12606569, 11978727, 26226420). It also plays an essential role for diverse protein functions, including the targeting of glycoproteins to appropriate destinations, and their stability, solubility, antigenicity, as well as binding with receptors (PMID: 14749736, 29202016, 29346549, 16584072). The inactivity of NGLY1 disrupts the degradation of glycoproteins and modification of transcription factors, which likely is the molecular basis of the diverse NGLY1 disorders (PMID: 25605922, 31002798, 29202016).

NGLY1 deficiency is among more than 100 congenital disorders of glycosylation (CDG), a rapidly expanding group of rare genetic metabolic disorders currently underdiagnosed (PMID: 8295395, 8286854, 16959566). Genomic analysis has suggested an unexpectedly high proportion of human genetic diseases that might be caused by additional protein glycosylation (PMID: 15924140, 17467977). It has been suggested that diagnosis of CDG should be considered for any patient with multisystem disease (PMID: 19474279, 23798010). Thus, the pathological and therapeutic investigation of NGLY1 deficiency may have a large health impact across an array of related disorders. However, the heterogeneity of NGLY1 syndromes, its low frequency and lack of study has limited our understanding and therapeutic intervention of the NGLY1 deficiency.

In the past two decades, advances in mass spectrum-based metabolomics and proteomics have enabled deep profiling of molecular phenotypes of diverse diseases (PMID: 22424236, 31142858). Metabolomics and proteomics profiling of human disease has the potential to provide detailed biological insights in disease deficiencies and alterations (PMID: 23938363,

27135364). Such information is expected to be valuable both as biomarkers of the disease and insights into therapeutic strategies.

To date, we have explored the potential of multi-omics profiling of complex diseases and health status, to identify targets for early diagnosis and prediction (PMID: 22424236, 31068711). In this study, we profiled 58 plasma samples collected from 19 patients and their healthy parents and siblings (Table 1). We also generated induced pluripotent stem cells (iPSCs) from a subset of the patients and prepared isogenic CRISPR correct lines, then further differentiated these iPSCs into neural progenitor cells (NPCs) to trace the progression of the disorder in the nervous system (PMID: 23197518, 24651067, 24500035). This panel of plasma and undifferentiated and differentiated samples were profiled using both targeted and untargeted metabolomics and proteomics. By integrating multi-omics profiles of plasma and iPSCs/NPCs samples at the molecular level, we established a new research paradigm for the study of the underlying pathology mechanism of NGLY1 deficiency.

Results

Summary of cohort and research design

With the support of Grace Science Foundation (https://gracescience.org/), Stanford University generated a cataloged biobank of NGLY1 deficient families' biological samples, genomic information, and medical records (https://med.stanford.edu/news/all-news/2017/10/biobank-entrepreneur-accelerate-research-into-rare-disease.html). This study was conducted with 58 plasma samples from the biobank, including 19 diagnosed NGLY1 deficient patients and 39 healthy immediate family members (Fig 1A, B, Table 1). To the best of our knowledge, this is the world's most updated and largest NGLY1 deficiency cohort. Our study undertook a comprehensive molecular profiling of metabolites (metabolome), lipids (lipidome), proteins (proteome) and immunoproteins (immunome) of plasma collected from patients and healthy controls.

To further validate the identified disease relevant metabolites, proteins and pathways, we generated a series of iPSCs from fibroblasts of patient CP1's family (CP1, CP1Fa, CP1Mo) and an external healthy control (MSJ), then differentiated iPSCs into NPCs (Fig 1, 2, Supplementary Table 1). CP1 holds two unique mutations of *NGLY1 gene* in each allele (exon 8: p.402_403del; exon 11: p.R524X). The patient-derived iPSCs were characterized by the expression of pluripotency markers (Supplementary Fig 1A) and karyotyping (Supplementary Fig 1B). The mutation of *NGLY1* of iPSCs was confirmed by Sanger sequencing (Fig 2A). We leveraged CRISPR/Cas9 technology to genetically edit the genotype of the cells from individual CP1 alleles. These efforts allow the evaluation whether correction in either mutation (CP1-E8C; CP1-E11C) or double mutation (CP1-E118C; CP1-E811C) are necessary or sufficient to rescue the disease molecular phenotype in the cell models. The correction of *NGLY1* mutation of CP1's iPSCs was also confirmed by sequencing (Fig 2B). Each cell line included 2-3 replicates (Fig 1C, Supplementary Table 2). iPSCs were further differentiated into NPCs with dual inhibition of Smad signaling (Supplementary Fig 2). Taken together, we analyzed the NGLY1 deficiency cohort and a series of iPSCs/NPCs derived from a patient's family (Fig 1A). The plasma metabolome, lipidome, proteome (Supplementary Table 2), cellular metabolome and proteome were integrated to demonstrate systematic molecular phenotypes.

Multi-omics profiling of NGLY1 deficiency plasma identified NGLY1 deficiency relevant molecular phenotypes

We conducted a deep profiling of plasma metabolome, lipidome and proteome of NGLY1 deficiency cohort, to identify NGLY1 deficiency relevant molecular phenotype. 396 metabolites of the NGLY1 deficiency cohort's plasma samples were quantified using Absolute IDQ-p400 kit (Biocrates Life Science AG) (PMID: 32052189) (Supplementary Table 3, 63 metabolites with more than 33% missing values were excluded from further analysis). Based on the remaining 333 quantified metabolites, patients and controls were grouped by hierarchical analysis and partial least squares (PLS) regression (Supplementary Fig 3A, 3B). A total of 36

metabolites demonstrated a fold change more than 1.5 between patients and controls ($P < 0.05$) (Supplementary Fig 3C). To further adjust the potential bias induced by differential age distributions between patients and controls, a mixed linear regression was applied to the 333 quantified metabolomics (Supplementary Fig 3D). The metabolites with significant correlation with the disorder ($P < 0.05$) were shown in Supplementary Table 4. These metabolites were classified into amino acids, biogenic amines, glycerophospholipids, glycerophospholipids, acylcarnitines (AC), ceramides (CER), sphingolipids (SP), cholesterol esters (CE) and diacylglycerols.

Among 723 plasma lipids quantified by Lypidyzer (Sciex) (Supplementary Table 5), 90 lipids demonstrated significant variation between patients and controls ($P < 0.05$), including CE, CER, diacylglycerols (DAG), dihydroceramide (DCE), free fatty acids (FFA), hexosylceramide (HCE), lysophosphatidylcholines (LPC), lysophosphatidylethanolamine (LPE), phosphatidylcholines (PC), phosphatidylethanolamines (PE), sphingomyelin (SM) and triacylglycerols (TAG) (Supplementary Fig 4A, 4B, 4D). 11 lipids demonstrated a fold change more than 2 (Supplementary Fig 4C). With the adjustment of age and gender bias by linear regression, the NGLY1 deficiency relevant lipids ($P < 0.05$) were listed in Supplementary Table 6.

We further profiled a broad spectrum of NGLY1 deficiency relevant plasma metabolites and putatively annotated 1091 metabolic features by using an in-house library and a combined public spectral database (https://doi.org/10.1101/2021.05.08.443258), among which 76 were annotated with a metabolomics standards initiative (MSI) level 1 or 2 identification (PMID: 28113135). Based on the untargeted metabolome, although the PLS clustering is not significant enough to separate patients and controls (Supplementary Fig 5A), a panel of metabolites with level 1 or 2 identification were able to demonstrate significant variation ($P < 0.05$) between controls and patients (Supplementary Fig 5B - 5D).

8920 proteins were quantified in depleted plasma samples using a standardized protocol with minor modification (PMID: 22424236, 32916130). 92 proteins demonstrated significant alternation between patients and controls (Supplementary Table 7). Although hierarchical

clustering and PLS clustering of the plasma proteome is not significant enough to separate patients and controls (Supplementary Fig 7A, 7B), a broad panel of plasma proteins demonstrated significant variation and fold change between patients and controls (Supplementary Fig 7C, 7D), including diverse immune proteins. Thus, we further quantified 62 cytokines in plasma, among which 23 demonstrated significant variation between patients and controls after the adjustment of age and gender by linear regression ($P < 0.05$) (Supplementary Table 8, Supplementary Fig 8C). However, the hierarchical clustering and PLS regression of cytokine profiles did not show clear separation between patients and controls (Supplementary Fig 8A, 8B). The plasma metabolome and proteome demonstrated a holistic dysregulation of lipid metabolism, immune responses, metabolism of amino acids and biogenic amines, which could be ascribed to NGLY1 deficiency induced the abnormal activity of key substrates such as NRF1 and BMP4 (PMID: 29202016, 33315951, 32720893) .

NGLY1 deficiency patients demonstrated global downregulation of lipid metabolism.

Both the untargeted and targeted metabolomics profiling of plasma samples demonstrated consistent downregulation of a broad panel of plasma lipids, including LPC, PC, CE, CER, SM, DAG, TAG, etc., except for PC 40.6, which is slightly elevated in patients' plasma (Supplementary Fig 3C, 4B).

PCs are the key component of the lipid bilayer of cells and involved in diverse metabolism and signaling pathways (PMID: 17704541). Despite most PCs were downregulated in patients' plasma, the role of elevated PC 40.6 level in NGLY1 deficiency affected lipid metabolism remain unclarified.

CEs, the major component of blood lipoprotein particles for transport and storage (PMID: 25682154, 33219259, 12654166). Recent study has suggested a potential correlation between cardiovascular disease and inefficient synthesis of CE from cholesterol (PMID: 30282999). The abnormal cholesterol metabolism as a typical phenotype of disrupted NRF2 pathway (PMID: 32566088), could be a compensate syndrome induced by NGLY1 deficiency (PMID: 31615832). Particularly, the downregulated plasma CEs in NGLY1 patients are mainly

with odd-chain fatty acids (CE 17.2, CE 19.2, CE 19.3, Supplementary Fig 3D), indicating a potential dysregulation in fatty acid synthesis and lipid transport. LPCs are components of oxidized low-density lipoprotein, usually involved in atherosclerosis and inflammatory diseases (PMID: 29769317, 27127201). The downregulated LPCs suggested low inflammatory syndromes and active immune response (Supplementary Fig 3C, 3D, 4A, 4D). Several SMs in patients' plasma were downregulated, with diverse fatty acid chain length. SMs are highly enriched in plasma membranes of cells for signal transduction (PMID: 9009829, 19857494). The SM-CER metabolism has been suggested with diverse diseases, such as neurodegenerative diseases, cardiovascular diseases and cancers (PMID: 25345489, 11116061, 24029234). CERs were also slightly downregulated in patients' plasma (Supplementary Fig 4A, 4D). Both SM and CER belong to SP, which is crucial in the biosynthesis of glycosphingolipids and gangliosides.

Notably, the fatty acid chains of the above downregulated lipids bear a wide range of chain length and saturation, although most of the free fatty acids did not show significant variation between patients and controls. Taken together, the fatty acid metabolism was generally downregulated in NGLY1 deficiency patients' plasma. Meanwhile, several ACs were upregulated in patients' plasma (AC 2.0, AC 14.0, AC 16.0, etc. Fig 3A, Supplementary Fig 3A), which also indicated a dysregulation of mitochondrial fatty acid beta-metabolism and a response to stress (PMID: 32057943). In patients' plasma proteome, a paralog of formylglycine-generating enzyme (SUMF2, Fig 3A) was significantly upregulated, which is involved in the metabolism of SPs and interaction with interleukins (PMID: 12757706 ,19739097). Cytoplasmic aspartate aminotransferase (GOT1) and transketolase (TKT) were also elevated in patient's plasma (Fig 3A), indicating an increasement of aspartate aminotransferase activity in hepatic glucose synthesis and in adipocyte glyceroneogenesis (PMID: 21900944, 28455244, 27259054). Thus, the transporting process of lipids or lipid precursors could be potentially inhibited in NGLY1 deficiency. As most patients were reported with abnormal liver functions (PMID: 32429657, 24651605), GOT1 and TKT could serve as potential targets for the aiding of hepatic symptom. The downregulation of cholic acid in

patients' plasma also indicates a reduced fat absorption and cholesterol excretion (PMID: 29032641, 15146238). Such metabolic disorder is relevant with the dysregulation of other carbohydrate metabolism pathways, suggested by an elevated level of lactate and deoxyribose 5-phosphate, and the reduced level of glycerol-3-phosphate (Supplementary Fig 5B-D). The reduced level of CSN2 and SERPINF2 in patients' plasma suggested a dysregulation of glucocorticoid receptor signaling pathways and coagulation system (Fig 3A, 3D) (PMID: 12788072, 25087955, 11604408, 12893771). In addition, the slashed lipid metabolism, combined with the elevated activity of glycolysis intermediates, indicated a potential mitochondrial dysfunction (PMID: 28750498), which affects the energy metabolism of NGLY1 deficiency patients (Fig 3B-D).

NGLY1 deficiency patients demonstrated dysregulation of amino acid metabolism, biogenic amine synthesis and glutathione metabolism

The targeted and untargeted profiling of plasma samples captured significant variation of amino acid and biogenic amine derivative. Creatinine as a biomarker of glomerular filtration rate, was significantly downregulated in patients' plasma (Supplementary Fig 3A, 3C, 3D). Its upstream and downstream metabolites, phosphocreatine and sarcosine, were slightly upregulated (Supplementary Fig 3A, 3D, 5B, 5D), suggesting an enhanced turnover rate of creatine metabolism. Nonetheless, there is no reported patient case with abnormal kidney function to date, while only one case with adrenal insufficiency (PMID: 30740912).

Diverse amino acids (Val, Phe, Trp, Asn, Lys, Thr, Pro) and their derivatives such as N-omega-hydroxyarginine and pyroglutamic acid were downregulated in patients' plasma, while several polyamines (putrescine, spermidine, sarcosine), acting as potential growth factors for cell division were significantly upregulated in patients' plasma (Supplementary Fig 3A, 3C, Fig 3C). More NGLY1-relevant metabolic phenotypes were characterized in untargeted metabolomics profiling, including slightly downregulation of glutathione, S-adenosylmethionine, upregulation of cysteinylglycine disulfide, pyroglutamic acid and asymmetric dimethylarginine (ADMA) (Supplementary Fig 3A, 5B, 5C). Tryptophan

metabolism plays a critical role encompassing the modulation of immune responses, the regulation of growth and neurodegeneration (PMID: 30760888, 27161289). Although we did not observe significant alternation of relevant enzymes, the ratio of kynurenine/Trp was significantly upregulated in patients' plasma (Fig 3A), suggesting an elevated activity of indoleamine-2,3-dioxygenase (IDO), an enzyme playing a role in a variety of pathophysiological processes such as antimicrobial and antitumor defense, neuropathology, immunoregulation, and antioxidant activity. The kynurenine pathway also modulates carbohydrate metabolism (PMID: 28751584, 27511838, 20872026)

The ratio of serotonin/Trp was upregulated in patients' plasma (Fig 3A), suggesting an increased activity of the hydroxylase TPH, which catalyzes the first step of serotonin synthesis. Serotonin is an important hormone and neurotransmitter, impairing diverse functions in gut-brain signaling, memory, sleep, and behavior. This pathway was associated with an elevated risk for a variety of diseases and disorders, including schizophrenia, somatic anxiety, etc. (PMID: 33671067, 33937270)

Asymmetric dimethylarginine (ADMA) is a metabolic by-product of continual protein modification processes in the cytoplasm of all human cells. It interferes with arginine in the production of nitric oxide, a key chemical to endothelial and cardiovascular health. The methyl groups transferred to create ADMA are derived from the methyl group donor S-adenosylmethionine, an intermediate in the metabolism of homocysteine and polyamines (Fig 3C, Supplementary Fig 3A, 5B-D). Remarkably, several enzymes detected in plasma proteome, SUMF2, GOT1, and TKT also play a role in the above pathways, showing a systemic metabolic response to NGLY1 deficiency (Fig 3A, 3C).

Integration of plasma metabolomics and proteomics profiles

To further investigate the insight of NGLY1 deficiency induced phenotypes, a pathway enrichment analysis was performed on metabolomics and proteomics plasma profiles. The metabolomic pathways significantly affected by NGLY1 deficiency were mainly involved in purine nucleotides degradation, skeletal muscle cell signaling, choline degradation and

phospholipases, etc. (Fig 3B). Compared with the metabolomics pathways, more proteomic pathways were affected by NGLY1 deficiency, including epithelial adherens junction signaling, endocytosis signaling, LXR/RXR activation, actin cytoskeleton signaling, and acute phase response signaling, etc. (Fig 3D). The top enriched pathways in plasma proteome and metabolome overlapped with tissue organization, lipid metabolism and immune signaling. In plasma metabolome, the dysregulation of purine nucleotide degradation suggests a disruption of purine salvage pathway, which disturbs the regeneration of ATP, thus affects rapidly prefiltering cells. Skeleton muscle signaling involves diverse signaling pathways, such as Akt hypertrophy signaling pathways, NF-κB pathways, E3 ubiquitination, TGFβ and mitochondrial oxidative metabolism. Choline degradation also occurs in mitochondria, involving diverse metabolic pathways, i.e., phosphatidylcholine biosynthesis, neurotransmitter synthesis, and protects cells from stress. In plasma proteome, the adherens junction signaling plays an essential role in the development of normal tissue organization. LXRs are important regulators of cholesterol, fatty acid, and glucose homeostasis. Actin cytoskeleton signaling is crucial for the movement regulation. The acute phase response plays an essential role in the innate immune system, cross-talking with LXR signaling in response to inflammatory cytokine signaling. By merging the top enriched pathways of plasma proteome with metabolome, we obtained a comprehensive network composed of neurotransmitters, lipid metabolites, cytoskeleton proteins, cytokines, with diverse relationship (Fig 3C, 3E). Metabolites in glycolysis, such as citric acid, were also involved in the network, suggesting an impairment of energy metabolism. Several complex and enzymes demonstrated essential roles in the merged network, such as AMPK, cytochrome C and glutathione peroxidase, suggesting an increased oxidative stress induced by NGLY1 deficiency.

NGLY1 deficiency patients demonstrated elevated activity of immune system.

Besides the metabolic enzymes, most of the quantified plasma proteins relevant with NGLY1 deficiency were involved in immune response (Supplementary Table 9). Thus, we also conducted targeted quantification of 62 plasma cytokines. Unlike the lipidomic profiles, most

of the immune relevant proteins were upregulated in patients' plasma, indicating a global activation of immune response in patients' plasma (Fig 4A). The most significant upregulated cytokine is BDNF, which plays an important role in the neuroimmune axis regulation and is closely relevant with schizophrenia, Alzheimer's disease, mood disorders and Parkinson's disease (PMID: 18408624, 18786162). EOTAXIN is also significantly elevated in patients' plasma. This cytokine is closely related with allergic reactions and skewed immune response (Fig 4A, Supplementary Fig 8). It influences neural progenitor cells and microglia, involved in the progression of aging, neurogenesis, and neurodegeneration. The increased circulating level of EOTAXIN was observed in major psychiatric disorders (schizophrenia, bipolar disorder, major depression) (PMID: 12193745). Other upregulated cytokines such as interleukins, leukemia inhibitory factor, chemokines and interferons were inflammatory mediators involved in chemotactic activation of inflammatory and NF-κB pathways (PMID: 29158945). In contrast, RANTES was downregulated in patients' plasma. RANTES mediates monocyte/macrophage infiltration into atherosclerotic lesions (PMID: 17325252). Previous studies also suggested that the dysregulation of the immune system is closely relevant with alternation of lipid metabolism (PMID: 34036900).

Cytokine profiles suggested a dysregulation of communication between immune cells, HMGB1 signaling and hepatic cholestasis, with an elevated level of most cytokines in the network (Fig 4B, C). The disrupted pathways impair diverse cellular functions such as the cell viability of neurons, differentiation of embryonic cells, growth of the brain, etc. NF-κB complex performed an essential role in the NGLY1 affected cytokine network (Fig 4C).

Multi-omics profiling of personalized NGLY1 deficiency cellular models validated the molecular phenotype during disease progress.

To further validate the NGLY1 deficiency relevant molecular phenotypes, we generated personalized cellular models from a patient CP1 (Supplementary Table 1) and applied an unbiased profiling approach to analyze the proteins and metabolites extracted from the cell lysate. For the profiling of disease-modeling cells, we captured 101973 metabolic features with putative identification. In iPSCs, 181 metabolites were annotated with level 1 or 2

identification; in NPCs, 89 metabolites were annotated with level 1 or 2 identification. Based on the untargeted metabolome of iPSC and NPC, the hierarchical clustering and principal component analysis (PCA) showed a relatively clear separation between the patient's cells, isogenic control cells and healthy adults' cells. In cellular proteomics profiles, we quantified 9550 proteins of iPSC lysate. Although the hierarchical clustering and PLS regression of iPSC proteome did not show clear separation between patients and controls, 189 proteins showed significant variation between patients and controls (Supplementary Fig 8A-C). 16 iPSC proteins showed a fold change more than 2 (Supplementary Fig 8D). In the iPSC-derived NPC, we quantified 14302 proteins of NPC lysate. Although the hierarchical clustering and PLS regression of NPC proteomes did not show clear separation between patients and controls, 51 proteins showed significant variation between patients and controls (Supplementary Fig 9A-C). Supplementary Fig 9D shows NPC proteins with most significant fold changes.

Patient-derived iPSC demonstrated personalized molecular phenotypes.

The personalized iPSCs derived from a patient's family were applied as a model to study the pathological mechanism. Besides the external control iPSC derived from patients' healthy parents and a healthy adult, we also applied CRISPR technology to correct the patient's mutations, generating isogenic controls. The metabolic phenotypes of patient's iPSCs were enormously distinctive from plasma. The iPSC metabolome and proteome demonstrated significant variation between patients and controls, relevant with diverse pathways, including nucleotide degradation, amino acid metabolism, EIF2 signaling, protein ubiquitination pathways, epithelial adherens junction signaling pathways and diverse signaling pathways (Fig 5A).

iPSC proteome demonstrated a general upregulation of stress-relevant and immune response proteins. The enrichment analysis of iPSC proteome suggested a significant impairment on EIF2 signaling, which incorporates diverse stress-related signals to regulate mRNA translation. Other affected iPSC pathways, epithelial adherens junctions and actin cytoskeleton signaling, were also dysregulated in plasma proteome. Several pathways were

directly linked with protein degradation, such as the unfolded protein response, which is upregulated, while the protein kinase A signaling was downregulated. Several metabolic pathways, such as glycolysis and purine nucleotides synthesis, overlapped with the affected pathways in plasma metabolome as well. Among the iPSC proteins illustrating most significant variation between patients and controls (Fig 5C), a majority were ribosomal proteins, scaffold proteins and heat shock proteins (SLC9A3R, STAU1, HSPA9). Several enzymes, such as HK2, DDAH2, PARP1 and PGK2, were downregulated in patient's iPSCs. HK2 is a hexokinase that phosphorylates glucose to produce glucose-6-phosphate, the first step in most glucose metabolism pathways. The downregulation of H2K may also interact with the dysfunction of mitochondria and dysregulation of lipid metabolism (PMID: 33377180). DDAH2, as a dimethylarginine dimethylaminohydrolase, regulates the cellular concentrations of methylarginines and generates nitric oxide. It is associated with chronic kidney disease and preeclampsia (PMID: 34018394, 34014421). PARP1(poly(ADP-ribosyl)transferase) is a chromatin-associated enzyme, modifies various nuclear proteins DNA and is involved in the regulation of various important cellular processes such as differentiation, proliferation, and antiviral innate immunity (PMID: 34037418). PGK2 (Phosphoglycerate kinase 2) is phosphoglycerate kinase that catalyzes the reversible conversion of 1,3-bisphosphoglycerate to 3-phosphoglycerate, affecting the carbohydrate metabolism (PMID: 29497897). H2AJ is a replication-independent histone that is a variant H2A histone. It impairs diverse signaling pathways, such as signaling by Wnt (PMID: 28489069). SLC9A3R1 is a scaffold protein that connects plasma membrane proteins and regulates their surface expression (PMID: 29491210). CKMT was slightly upregulated in patient's iPSCs. It is a mitochondrial creatine kinase responsible for the transfer of high energy phosphate from mitochondria creatine. Thus, we observed an upregulation of phosphocreatine in patients' iPSC (PMID: 26708229), which could explain the reduced plasma creatinine level.

The metabolomics profiles of iPSCs are more consistent with plasma metabolome, with most of the amino acids, amino acid derivatives, lipids, especially free fatty acids downregulated in patient's iPSCs. The modulation of nucleotides is more complex, with a downregulation of

guanine and guanosine, and an upregulation of adenine, adenosine and hypoxanthine (Supplementary Fig 6A, 6C). N-acetylmannosamine (ManNAc) is also downregulated in both patients' plasma and iPSCs. ManNAc and uridine 5'-diphosphate can be biosynthesized from uridine diphosphate-N-acetylglucosamine; which is mediated by the bifunctional enzyme UDP-N-acetyl glucosamine-2-epimerase/N-acetylmannosamine kinase, a key enzyme in the biosynthesis and glycosylation of proteins. ManNAc is the first committed biological precursor of N-acetylneuraminic acid (Neu5Ac, sialic acid), which is an essential precursor for protein glycosylation (PMID: 23842869). Other intermediates in the relevant pathways such as UMP and UDP were also downregulated in the patient's iPSCs (Supplementary Fig 6A, 6C), indicating a downregulation of glycan synthesis, which could be a consequence of the NGLY1 deficiency induced glycosylation traffic (PMID: 17460662). In addition, several intermediates in the TCA cycle and glycolysis, such as citric acid and pyruvate, are also downregulated in the patient's iPSC, suggesting a holistic downregulation of carbohydrate metabolism in NGLY1 deficiency (Supplementary Fig 6A, 6C). Ribonucleotide degradation and related pathways as the most significantly altered pathways in iPSC (Fig 5E, Supplementary Fig 6A, 6C), was mainly inhibited, although no significant variation in proliferation was observed in patient's iPSCs.

iPSC-differentiated NPC demonstrated more tissue-specific molecular phenotypes.

We further differentiated the disease-modeling iPSCs into NPCs to track the disease progressing during the development of the nervous system. NPC proteomes demonstrated more pathways with significant alternation between patients and controls. Interestingly, the most significantly dysregulated pathways in iPSC were also dysregulated in NPC, including EIF2 signaling, protein ubiquitination pathways, epithelial adherens junction signaling pathways, etc. (Fig 5B). Other relevant pathways included proteasome-mediated degradation pathways, protein ubiquitination, interleukin signaling, cellular response to stress and diverse signaling pathways. Besides the signaling and immune relevant pathways partially overlapped with the pathways identified in iPSC and plasma (Fig 5A, 3D, 4B), the NPC proteome demonstrated more alternation of pathways relevant with protein degradation (Fig 5B).

Notably, most of the quantified NPC proteins were upregulated in patient's NPCs (Fig 5D, Supplementary 9D). We found that synuclein alpha (SNCA) and synuclein beta (SNCB) were both significantly upregulated in the patient's NPC (Fig 5D). SNCA as a biomarker for neurodegenerative diseases, is abundant in the brain, and smaller amounts are found in the heart, muscles, and other tissues. Although the function of SNCA is not well understood, studies suggest that it plays a key role in maintaining an adequate supply of synaptic vesicles in presynaptic terminals. It may also help regulate the release of dopamine, a neurotransmitter that is critical for controlling voluntary and involuntary movements (PMID: 32740728). Notably, dopamine was also downregulated in patients' plasma. SNCA peptides are a major component of amyloid plaques in the brains of patients with Alzheimer's disease. Its aggregation is affected by PC membrane, whose components were also altered by NGLY1 deficiency (PMID: 29853693). SNCB is a non-amyloid component of senile plaques found in Alzheimer disease. It could also act as a regulator of the SNCA aggregation process (PMID: 29851342, 29782835). Although no direct observation of protein aggregation in NGLY1 deficiency cells has been reported, the SNCA-relevant synaptic vesicle trafficking and cellular skeleton deficiency could be an alternative explanation of the neural phenotype of NGLY1 deficiency.

NPC proteomes demonstrated dysregulation of a broad spectrum of ubiquitination-relevant proteins (Fig 5D), although some of their roles in NGLY1 deficiency remain unclarified. For instance, UBE2K as a ubiquitin-conjugating enzyme, mediates the selective degradation of short-lived and misfolded proteins. It also determines neurogenic potential during development (PMID: 32451438). The upregulated USP10 also belongs to a ubiquitin-specific protease family of cysteine proteases. The enzyme specifically cleaves ubiquitin from ubiquitin-conjugated protein substrates, thus can activate AMPK (PMID: 26876938, 33315951). Previous study also suggested connection between ubiquitination and NGLY1 deficiency (PMID: 29436721). VCP is an ATPase that emerge a central role in the ubiquitin system, involved in protein degradation, DNA repair and replication, regulation of the cell

cycle, activation of the NF-κB pathway, and also causes profound mitochondrial uncoupling (PMID: 25146396, 23498975).

A panel of enzymes involved in carbohydrate and lipid metabolism were also altered in the patient's NPC (Fig 5D). PCYT2, for instance, catalyzes the formation of CDP-ethanolamine from CTP and phosphoethanolamine in the Kennedy pathway of phospholipid synthesis (PMID: 20427062). GFPT2 (glutamine-fructose-6-phosphate transaminase 2), controls the flux of glucose into the hexosamine pathway. It is likely involved in regulating the availability of precursors for N- and O-linked glycosylation of proteins (PMID: 33257855). Like the iPSC proteome, another phosphoglycerate kinase PGK1 was downregulated in patient's NPC (PMID: 30029001). Another down-regulated protein in the patient's NPC, EEF1B2 is a translation elongation factor involved in the transfer of aminoacylated tRNAs to the ribosome. It is associated with diverse diseases and neurodevelopmental disorders (PMID: 30370994).

Among the dysregulated cytosol enzymes, N-Alpha-Acetyltransferase 35 (NAA35) is auxiliary component of the N-terminal acetyltransferase C (NatC) complex which catalyzes N-terminal acetylation, an important process for maintenance of protein function and cell viability in human cells (PMID: 19398576, 22718336). FN3K and FN3KRP catalyze the phosphorylation of fructosamines of glycated proteins, further resulting in deglycation (PMID: 20967558). They are also modulators of NRF2 activity and involved in the response to oxidative stress by mediating deglycation of NRF2 (PMID: 31398338). PLCG1 catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate, contributing to cell death and inflammatory reactions (PMID: 29937272, 23140367). GAA is an alpha glucosidase, which is essential for the degradation of glycogen to glucose in lysosomes. Its deficiency leads to the glycogen accumulation in lysosomes, also known as Pompe's disease, which is an autosomal recessive disorder with characterized by muscle weakness similar as NGLY1 deficiency (PMID: 31392190, 32745073). Its related pathways are the innate immune system and glucose metabolism, also affected by NGLY1 deficiency. SGSH is a sulfamidase, involved in the lysosomal degradation of heparan sulfate. The dysfunction are associated with the lysosomal storage disorder resulting from impaired

degradation of heparan sulfate, and relevant with early onset of neurodegeneration (PMID: 33423317).

NPC metabolome also showed significant downregulation of amino acids and derivatives, indicating a down-regulated protein metabolism. Like the iPSC metabolome, ManNAc, UMP and UDP were also downregulated. An intermediate of creatinine synthesis, phosphocreatine, was also downregulated in the patient's NPC, indicating a similar downregulation of the creatinine metabolism. Notably, Uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), a coenzyme in metabolism and a key substrate of glycosyltransferases for the synthesis of glycans (pmid: 16847058), was significantly upregulated in NPC metabolome. UDP-GlcNAc is the product of the hexosamine biosynthesis pathway starting from the fructose 6-phosphate and glutamine, both of which are downregulated in patient's NPC. Thus, the excess amount of UDP-GlcNAc could be derived from other sources than the biosynthesis from fructose 6-phosphate and glutamine. A potential source could be the extra N-glycans of accumulated glycoproteins in NPC. UDP-GlcNAc is extensively involved in intracellular signaling as a substrate for O-linked N-acetylglucosamine transferases. It is also involved in nuclear pore formation and nuclear signaling (PMID: 22522719). Comparing the NPC metabolome with iPSC metabolism, the top enriched pathways were identical, nonetheless, the upregulated and downregulated metabolites were distinct (Fig 5E, 5F). In NPC metabolome, more metabolites in xanthine and xanthosine salvage, taurine biosynthesis, adenine nucleotide degradation pathways were downregulated, while creatine-phosphate biosynthesis and choline degradation pathways displayed similar regulation as in plasma and iPSC.

Integration of plasma and cellular molecular phenotype suggested a switch of regulation pathways during disease progress

The integration of multi-omics profiles remains a challenge despite the rapid development of data analytical strategies (PMID: 33248648, 31003449). Unlike the population data driven investigations, the study of rare diseases is extremely limited by the patient cohort. Thus, the patient-derived personalized cellular models provide a complementary strategy to address the limitation. The isogenic control generated by CRISPR technology added more power for multi-

omics profiling and data analysis. By comparing the metabolomics and proteomics profiles across different sample types, we identified common molecular phenotypes and pathways relevant with NGLY1 deficiency, including the downregulation of amino acid metabolism, lipid metabolism and upregulation of biogenic amine synthesis, hexosamine pathways and immune response. The hierarchical clustering of pathways identified in plasma, iPSC and NPC demonstrated a massive shift of pathway regulations during the disease progress, as shown in the closer clustering of iPSC and plasma pathways (Fig 6A). NPC displayed more tissue-specific phenotypes, especially in the proteomics profiles, suggesting a global alternation of both metabolism pathways and signaling pathways. The overlapped pathways between iPSC, NPC, plasma, and cytokines indicate a more enhanced immune response in NPC during the development of NGLY1 deficiency, as shown with closer clustering of the NPC and plasma proteome (Fig 6A). The elevated immune response in NPC could induce more phenotypes consequently. Although the NGLY1-affected NPC pathways presented more overlapping with plasma and cytokine, the distinct regulation behavior shown in the clustering between NPC and plasma suggested differential molecular phenotypes intracellular and extracellular, which need to be evaluated when integrating multi-omics profiles. In general, multi-omics profiling of plasma samples identified significant molecular phenotypes relevant with amino acid metabolism, biogenic amine synthesis, hexosamine metabolism, and upregulation of stress response, such as the global downregulation of lipids, upregulation of immune response, etc. (Fig 6B). The alternation of related pathways was validated by the metabolome and proteome of patient-derived cellular models, especially NPC.

To summarize, NGLY1 deficiency induced comprehensive alternation of patients' metabolome and proteome, including a global downregulation of amino acid metabolism, lipid metabolism and glycan synthesis, and an upregulation of immune response and protein ubiquitination. A variety of pathways relevant with cellular signaling, cellular skeleton and intracellular communications were also disrupted in patient-derived iPSC and NPC. Compared with the plasma metabolome and proteome, the NPC profiles showed higher overlapping with plasma profiles (Fig 6B), especially for the upregulation of immune response. This indicated

a potential inducement of immune reaction by the neural degeneration. Among all the characterized pathways, the metabolic pathways provided the highest potential for therapeutic intervention. We identified several pathways that could correlate with the immune and neurodegenerative syndromes, in which some key enzymes and metabolites may serve as potential therapeutic targets, although the direct mediator between NGLY1 deficiency and the molecular phenotypes remains unclarified.

Discussion

Although the NGLY1 deficiency relevant downregulation of amino acid derivatives are a common phenotype across the plasma and cellular models, the plasma phenotypes became less significant after applying linear regression to adjust the bias of age and gender. Nonetheless, clinical records indicated that most patients took multivitamin and amino acid diet supplies. Thus, the downregulation of amino acid metabolism could be a more significant phenotype, with a potential cause of interrupted protein degradation or cell proliferation. The downregulation of lipids is also a typical phenotype across different types of samples, nevertheless, the affected lipids variated, and the overall downregulation of lipids were less significant in cellular models. Thus, the global downregulation of plasma lipidome could be induced by interruption of molecular trafficking. Fatty acids were the most common down-regulated lipid species, although the chain length and saturation in different sample types varied significantly. Taking account of the fatty acid chains in diverse lipid species, NGLY1 deficiency induced a general dysregulation of fatty synthesis, which could be induced by the disrupted carbohydrate metabolism sharing the intermediate with fatty acid synthesis, or the interruption of fatty acid metabolism and transporting.

The biogenic amine synthesis is a complicated network, with diverse cross-talking between different super-pathways. The reduced glutathione is downregulated across different sample types, indicating a common response with stress. Nonetheless, the alternation of relevant metabolites was not consistent across different sample types, meaning differential mechanisms of NGLY1-deficiency-induced stress. The downregulation of ManNAc and

relevant glycan synthesis is another common pathway alternation, which could be the consequence of the excess amount of glycans and enhanced hexoamine metabolism induced by NGLY1 deficiency. However, the pathway not only involved the synthesis of N-glycans, but also impairs with the synthesis of O-glycans. Thus, the O-glycan signaling could be also inhibited due to insufficient glycan precursors, which may affect diverse signaling pathways. Although iPSC proteome did not show significant alteration enzymes involved in the above pathways, NPC proteome demonstrated comprehensive dysregulation of relevant enzymes.

We also noticed that iPSC proteome did not demonstrate significant alteration in immune relevant pathways. Nonetheless, the NPC proteome demonstrated the upregulation of interleukin signaling, suggesting a potential immune response induced by dysregulation nerve function.

The activation of ubiquitination signaling pathway and NRF2 mediated oxidation response are specific molecular phenotypes of the patient's NPC. The upregulation of diverse ubiquitination pathways could be a consequence of the disrupted deglycosylation pathways, as ubiquitination is a prerequisite for NGLY1 catalyzed deglycosylation. NRF2 mediated oxidative response was also a consequence of the dysfunction of NGLY1. Previous studies have suggested that deglycosylation was a crucial step for the maturation of the NRF family. Nonetheless, as the relevant pathways were not significantly alternated in patients' iPSC, this could be a downstream phenotype.

NGLY1 deficiency as a rare disease with comprehensive syndromes, its therapeutic development was withdrawn by personalized phenotypes and lack of biomarkers. Our results showed that patients' plasma and patient-derived iPSC/NPC demonstrated substantial individualized molecular phenotypes. Potential reasons for such diversity could be the differential downstream pathways regulated by NGLY1 in individual patients, and such responsive reactions of the NGLY1 deficiency are also personalized. The patient-derived cellular combining with the gene-editing technology provided a potential solution to such problems. As more patients are included in the generation of personalized cellular models, more personalized pathway responses will be discovered for NGLY1 deficiency.

Conclusion

This study reported the most comprehensive and updated multi-omics study of NGLY1 deficiency to date. The cohort covers a majority of reported cases worldwide. The patient-derived cellular model assisted the identification of disease-induced pathway disruption. The identified differential metabolites and proteins in both plasma samples and cellular models revealed the alternation of diverse pathways in NGLY1 deficiency, lipid metabolism, amino acid metabolism, biogenic amine synthesis, glutathione metabolism and immune response. The plasma metabolome and proteome demonstrated significant individualization of patients' molecular phenotypes. The personalized iPSC/NPC metabolome and proteome revealed that the onset of immune response could be a consequence of other syndromes and correlated with the progression of the disease. The findings offered abundant information about the global proteomic and metabolic alterations induced by NGLY1 deficiency and will facilitate the therapeutic and diagnosis development of the disease. This approach is transferable to the study of diverse underdiagnosed rare diseases and will shed light on the pathologic and therapeutic investigation of other CDGs.

Methods

Metabolomics analysis

We profiled the plasma samples using both targeted and untargeted approaches. For targeted profiling, we used the "AbsoluteIDQ p400 HR assay" kit to quantify 406 metabolites including sugars, amino acid and biogenic amine derivatives and lipids. In parallelly, we used the Lipidyzer platform (SCIEX) to quantify around 1,000 lipid species across 10 lipid classes. The untargeted LC-MS approach provided a broader coverage of the plasma metabolome. We also used the untargeted approach to profile plasma metabolome and the personalized disease modeling cellular metabolome to validate the comprehensive molecular phenotypes of plasma and trace the progression of the disease in patient-derived cell models.

Targeted metabolomics

10 µL plasma were prepared and analyzed following the sample preparation procedures and MS/MS (Tandem Mass Spectrometry)-based analytical methods provided with "AbsoluteIDQ® p400 HR assay" kit. LC-MS and flow injection analysis (FIA) data were acquired on a Q Exactive HF mass spectrometer (Thermo Scientific). The raw data was processed by Xcarlibur (Thermo Scientific) and MetIDQ (Biocrates). The complete list of metabolites is reported in Supplementary Table 0. Metabolomic data below the limit of detection (LOD) were replaced with a random number from a uniform distribution on the interval (0, LOD/2). To control the false discovery rate (FDR), P values were adjusted applying the Benjamini-Hochberg correction within each metabolite class. All statistical tests used a two-sided 5% significance level. Statistical analyses were performed using RStudio (RStudio: Integrated Development for R. RStudio, Inc.).

Lipidomics profiling of plasma samples

Lipids were extracted and analyzed using a standard protocol (PMID: 32470399). A mixture of methyl tertiary-butyl ether, methanol and water was used to extract lipids from 40 µl of plasma by biphasic separation. Lipids were then analyzed with the Lipidyzer platform consisting in a DMS device (SelexION Technology, SCIEX) and a QTRAP 5500 (SCIEX). Lipids were semi-quantified using a mixture of 58 labeled internal standards provided with the platform.

Untargeted metabolomics

Untargeted plasma metabolomics was performed with an optimized pipeline on a broad-spectrum LC–MS platform (PMID: 25787789). In brief, complementary reverse-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) separations were applied in parallel to maximize metabolome coverage. Plasma metabolome data were acquired on a Q Exactive plus mass spectrometer (Thermo Scientific). Cellular metabolome data were acquired on Q Exactive plus mass spectrometer (Thermo Scientific)

and QTOF 6538 (Agilent). Both instruments were operated in full mass-spectroscopy scan mode. Tandem mass spectroscopy data were acquired at various collision energies on pooled quality control samples. LC–MS data were processed using Progenesis QI (Nonlinear Dynamics) and metabolic features were annotated by matching retention time and fragmentation spectra to authentic standards or to public repositories by MetID pipeline available in https://jaspershen.github.io/metID/ and validated manually (https://doi.org/10.1101/2021.05.08.443258).

Proteomics analysis

LC-MS based Tandem Mass Tag (TMT) labeled identification and quantification of proteome were applied for the proteomics profiling of both plasma samples and disease-modeling cell lysates. In brief, for patient-derived cells, cell pellets were lysed in a fresh prepared cell lysis buffer containing 6 M Gdmcl, 10 mM TCEP, 40 mM CAA, 500 mM Tris (pH 8.5). For plasma proteome, 25 μL plasma were depleted using an Agilent Mars human 14 column (4.6 mm x 50mm). The unbound fraction was collected for buffer exchange with the fresh cell lysis buffer for further proteome analysis as previously described. Data acquisition was operated on a NanoAquity 2D nanoLC (Waters) directly coupled in-line with a linear trap quadrupole (LTQ)-Orbitrap Velos instrument (Thermo Scientific) via Thermo nanoelectrospray source. The raw data were processed with the Proteome Discoverer 2.1 (Thermo), using UniProtKB/Swiss-Prot database, applying 1% FDR. Further data analysis was performed using Rstudio. P values were adjusted applying the Benjamini-Hochberg correction.

Immune protein measurements

The 62 plex-Luminex antibody-conjugated bead capture assay (Affymetrix) was used to characterize blood levels of immune proteins. The assay was performed by the Stanford Human Immune Monitoring Center. The protocol is available in https://iti.stanford.edu/himc/protocols.html.

Data analysis and integration

A linear regression was applied to plasma metabolome and proteome data to adjust potential bias induced by the unmatching age distribution between patients and controls (Supplementary Fig 2). The metabolites and proteins relevant with NGLY1 deficiency ($P < 0.05$) were selected for pathway enrichment analysis, correlation analysis and network analysis with IPA (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis), Cytoscape (PMID:14597658) and "igraph" R package [Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006].

Generation of iPSCs and NPCs

Patient's skin biopsy samples were collected in Stanford Hospital and fibroblast cells were derived in Stanford Cytogenetics Laboratory. The iPSCs were generated with the Sendaivirus delivery system (PMID: 19838014) using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (ThermoFisher Scientific). Briefly, the fibroblast cells were maintained in DMEM containing 10% fetal bovine serum. During the reprogramming, $2x10^5$ cells were transduced with Sendai virus with MOI of 5:5:3 (KOS:c-Myc:Klf4) as instructed in the manual. The cells were maintained in fibroblast media for 6 days then passaged onto Matrigel-coated dishes for further reprogramming in Essential 6 medium plus human bFGF (ThermoFisher Scientific). The iPSC colonies were manually picked onto Matrigel-coated plates to generate stable iPSC lines. The iPSCs were maintained in mTeSR1 medium (Stemcell Technologies) and routinely passaged every 4~6 days with Versene (EDTA) solution (Lonza). Selected iPSC clones were further validated for expression of pluripotency markers with immunostaining [supp data] and NGLY1 mutations were confirmed with Sanger sequencing. The karyotyping of the iPSCs were checked by G-banding analysis in WiCell institute.

Ngly1 mutation correction with CRISPR-Cas9 gene editing in iPSCs

For Ngly1 gene editing, the guide RNA sequence was cloned into pX459 vector for either exon 8 (c.1205_1207delGAA) or exon 11 (c.1624C>T) mutation correction. Patient-derived iPSCs harboring Ngly1 mutations were dissociated with Accutase into single cells, and $2x10^6$ iPSCs were mixed with pX459-sgRNA vector and ssDNA repair donor template and electroporated with Lonza 4D-Nucleofector system using the P3 Primary Cell Nucleofector Kit (Lonza). The iPSCs were then recovered in mTeSR1 medium for 48 hours before adding puromycin for 3-day selection. The survived iPSCs were dissociated with Accutase and plated at a very low density for subcloning. Each individual iPSC colony was expanded and screened with Sanger sequencing for mutation correction.

Feeder-free differentiation of iPSCs to NPCs

The feeder-free NPC differentiation was achieved by dual-SMAD signaling inhibition modified from the publication from Dr. Studer lab [Chambers et al, 2009]. iPSC monolayer cultures were incubated in KO-DMEM medium containing 20% knockout serum replacement (KSR, ThermoFisher) with 10uM SB431542, 100nM LDN-193189, Glutamax (1/100), NEAA (1/100) and beta-mercaptoethanol (0.1mM). The cells were fed with fresh media every other day and collected seven days after differentiation.

Figures & Tables

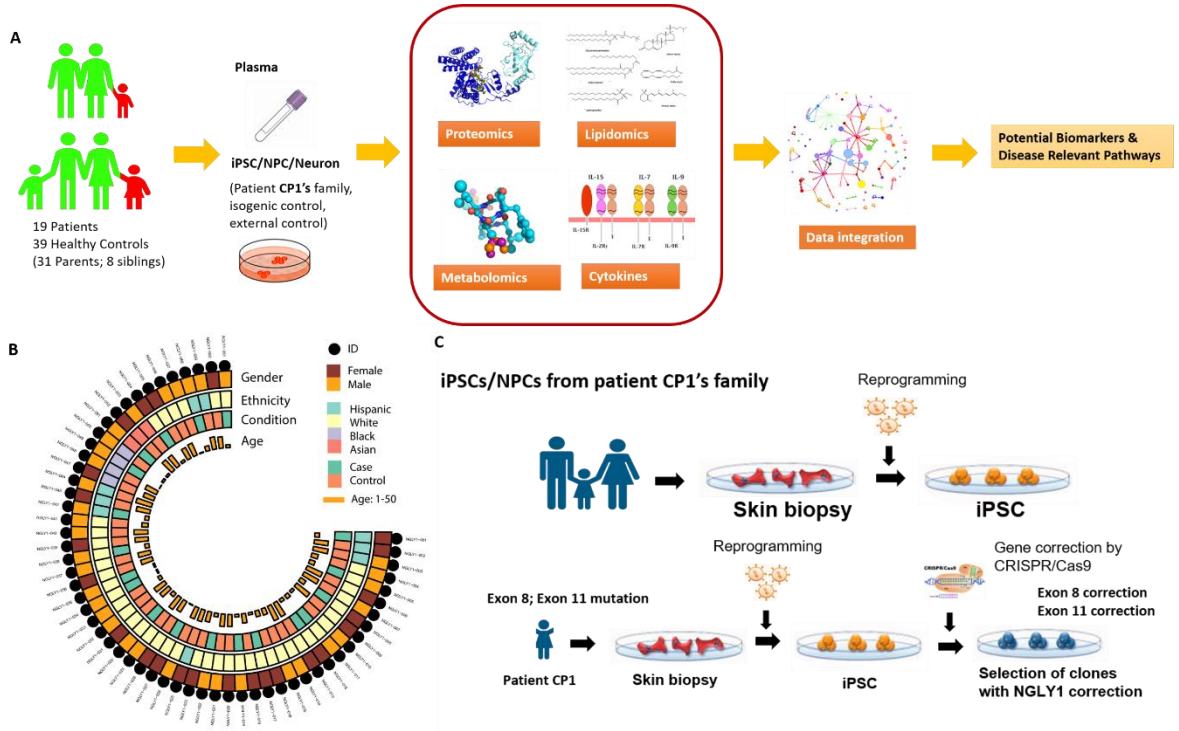

Fig 1. Overview of multi-omics profiling landscape. A) Multi-omics profiling of NGLY1 deficiency cohort and patient-derived cellular models. A unique cohort including 19 NGLY1 deficiency patients and 39 of their healthy family members were recruited for metabolomics and proteomics analysis. We collected plasma from the whole cohort and derived iPSCs from patient CP1's family. The plasma proteome, metabolome, lipidome and cytokines were quantified to identify potential therapeutic targets for NGLY1 deficiency. We also profiled the metabolome and proteome of patient CP1's iPSC/NPC and its isogenic and external controls to validate the alternation of multi-omics profiles during the onset of the disorder. B) The characteristics of the patient cohort, showing patients and controls' gender, age and ethnicity. C) The personalized cellular disease models derived from patient CP1's family and modified by CRISPR technology.

Table 1. Characteristics of the NGLY1 deficiency cohort

	Patients	Controls
Mean Age (range)	9.8 (1-19)	34.4 (2-50)

Sex	n	%	n	%
Female	6	32.5%	18	46.5%
Ethnicity				
Asian	1	5.3%	2	5.1%
African American	1	5.3%	3	7.7%
Caucasian	14	73.7%	26	66.7%
Hispanic	3	15.8%	6	15.4%

Table 1. The summary of plasma samples collected in NGLY1 deficiency biobank. We presented patients' gender, age, and ethnicity, which were further adjusted by linear regression.

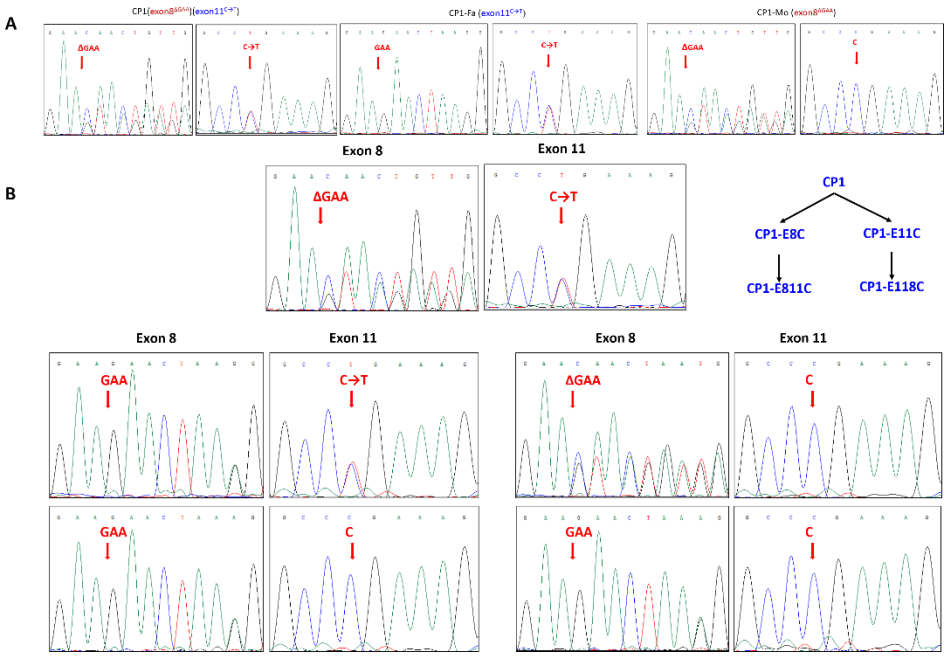

Fig 2. Characterization of patient-derived iPSCs. A) Confirming NGLY1 mutations in iPSCs by Sanger sequencing. B) Confirming the correction of NGLY1 mutations by CRISPR in iPSCs.

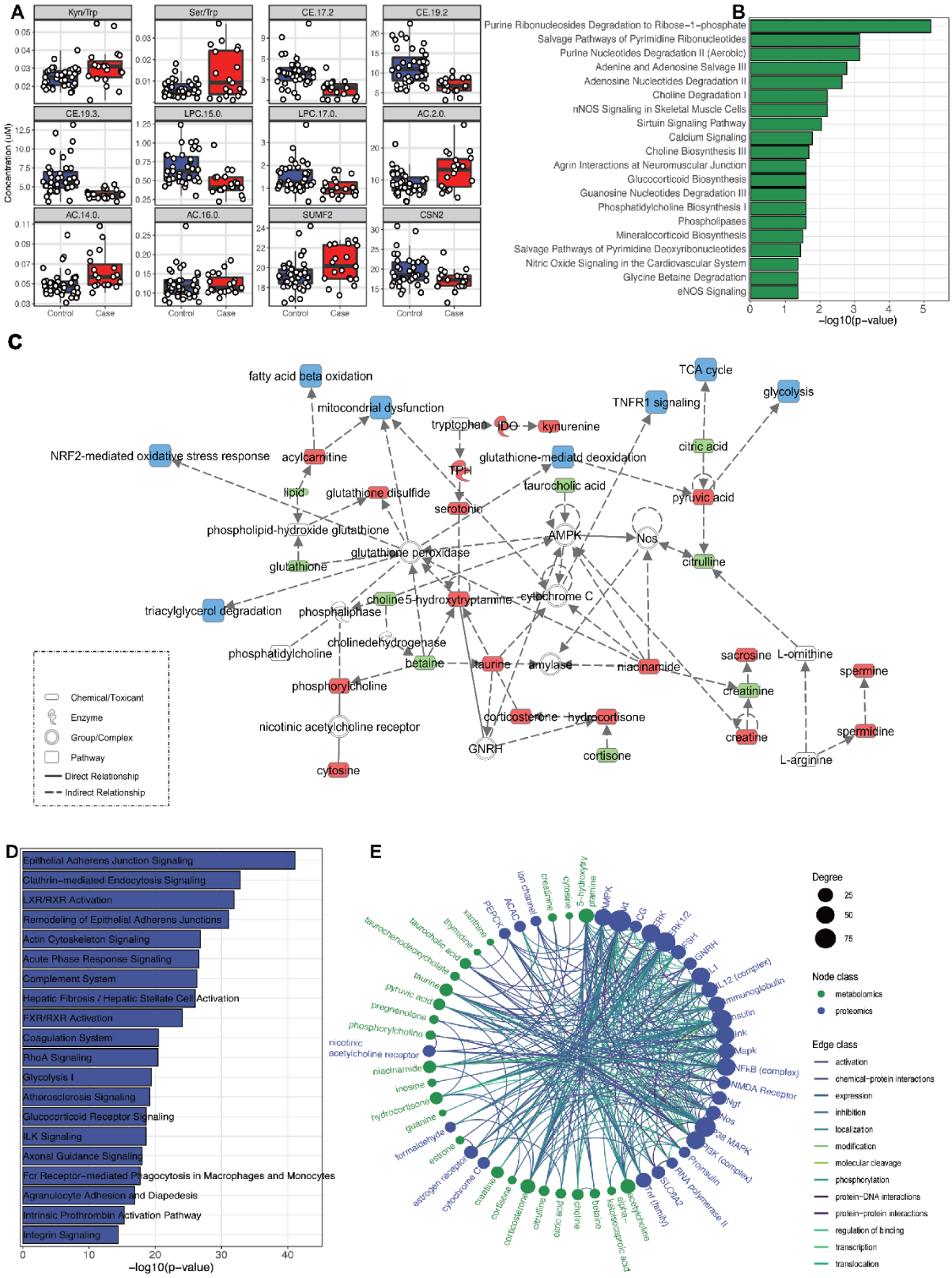

Fig 3. NGLY1 deficiency-induced metabolome and proteome dysregulation in plasma. A) The boxplot of selective plasma metabolites and proteins with significant variation between controls and patients (P < 0.05). B) The enrichment of dysregulated pathways of plasma

metabolome (top 20). C) The diagram of dysregulated plasma metabolites involved pathways. D) The enrichment of dysregulated pathways of plasma proteome (top 20). E) The networks composed of dysregulated proteins and metabolites in patients' plasma. Nodes were metabolites and proteins with significant alternation in patients' plasma. The edge shows the direct and indirect relation between each two nodes.

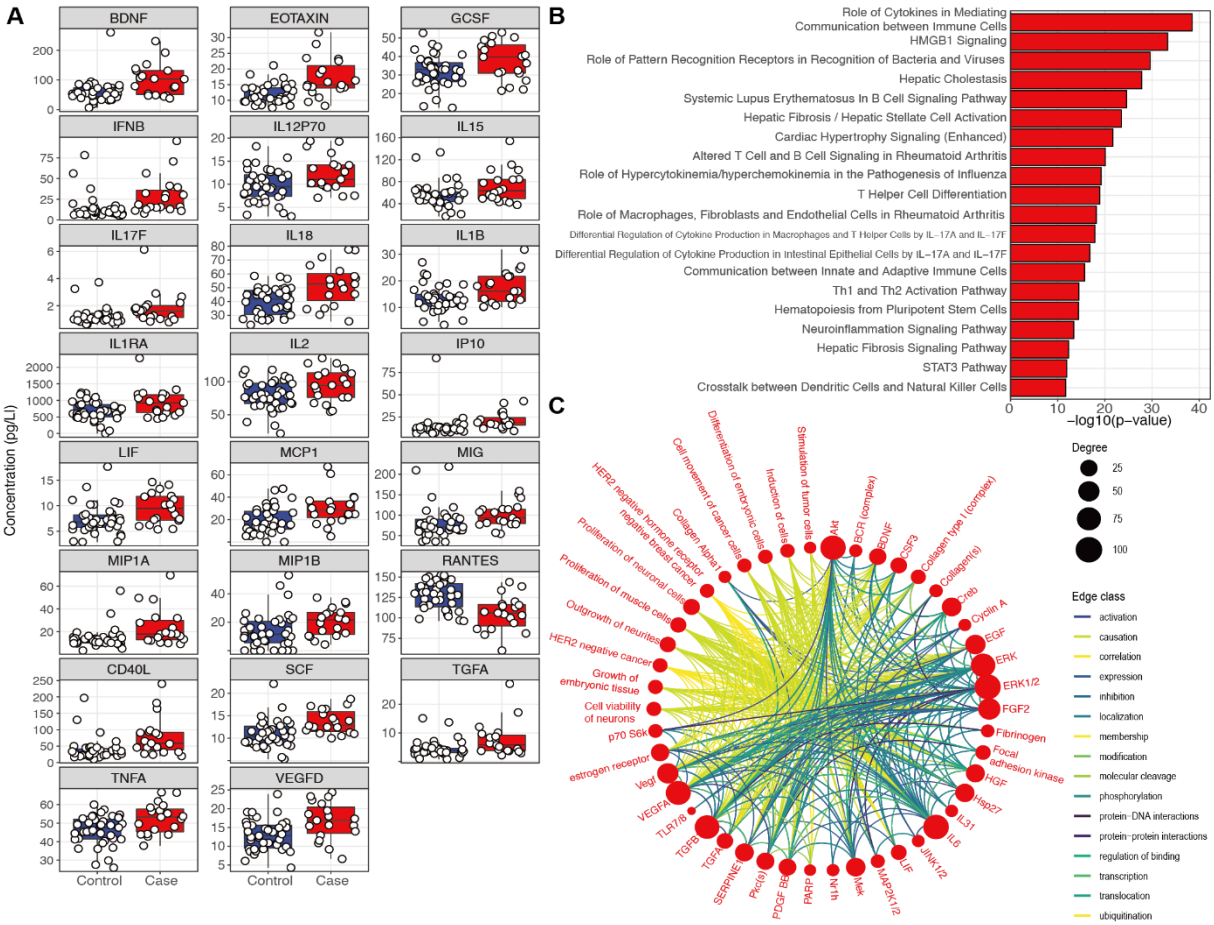

Fig 4. The NGLY1 deficiency induced dysregulation of the cytokine pathway. A) Boxplots of cytokines with most significant fold changes and variation between controls and patients. B) The enrichment of dysregulated cytokine pathways. The bar shows -$\log_{10}$ P value. C) The most upregulated cytokines and their relevant functions and diseases in the network. The edge shows the direct and indirect relation between each two cytokines.

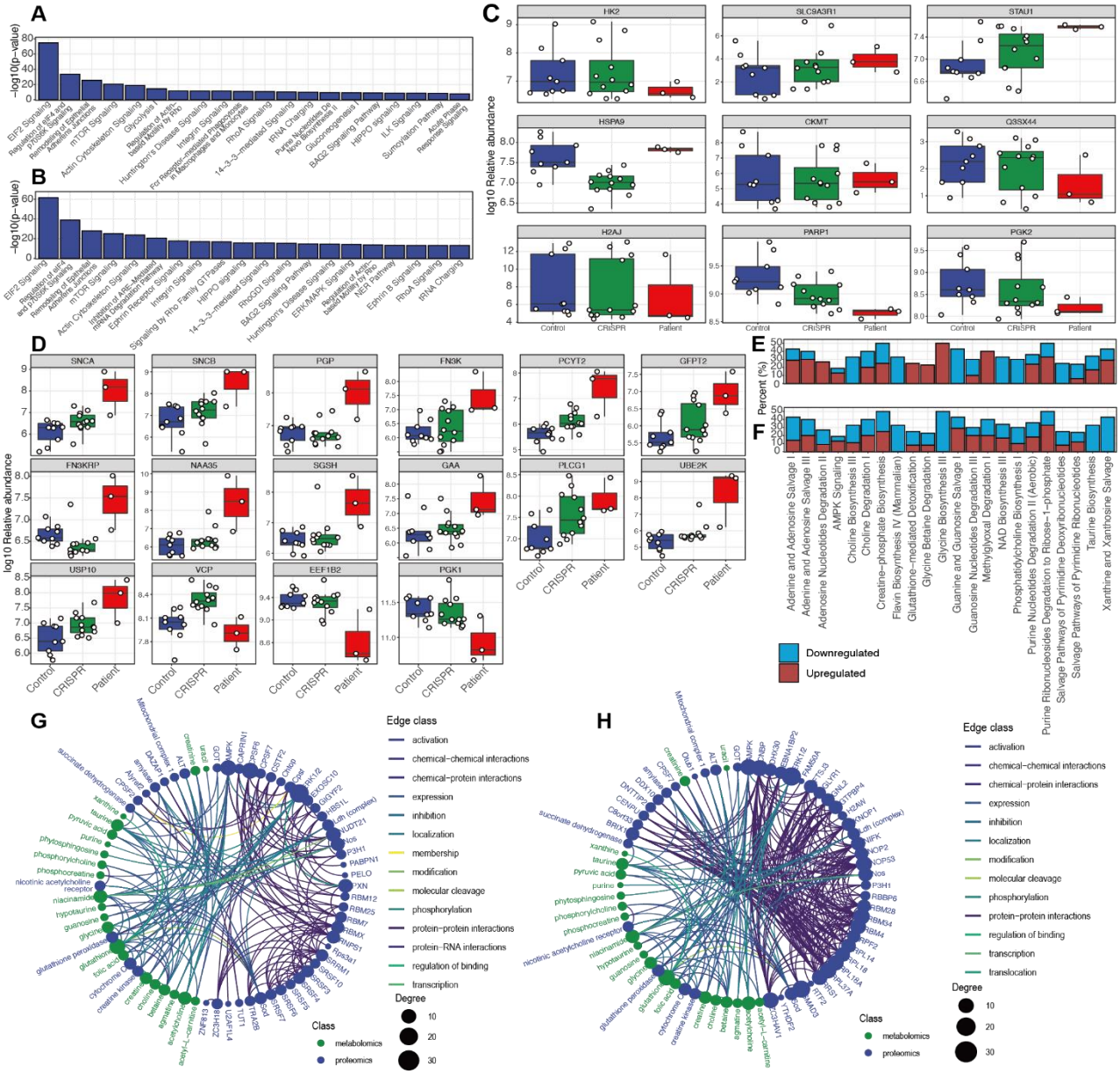

Fig 5. The NGLY1 deficiency relevant pathway dysregulation in patient-derived iPSC/NPC/neurons. A) Dysregulated proteomics pathways in patient-derived iPSC ($-\log_{10}$ P > 5). The bar shows the P value. B) Dysregulated proteomics pathways in patient-derived NPC ($-\log_{10}$ P > 5). The bar shows the $-\log_{10}$ P value. C) iPSC proteins with significant variation between patient (blue), CRISPR-corrected isogenic control (green) and external control (red). D) NPC proteins with significant variation between external control, CRISPR-corrected isogenic control and patient. E) Dysregulated metabolomics pathways in patient-derived iPSC (P < 0.05). The bar shows the percentage of downregulated and upregulated metabolites in

the pathway. F) Dysregulated metabolomics pathways in patient-derived NPC ($P < 0.05$). The bar shows the percentage of downregulated and upregulated metabolites in the pathway. G) The networks composed of dysregulated proteins (blue) and metabolites (green) in patient-derived iPSC. Nodes were metabolites and proteins with significant alternations. The edge shows the direct and indirect relation between each two nodes. H) The networks composed of dysregulated proteins and metabolites in patient-derived NPC. Nodes were metabolites and proteins with significant alterations. The edge shows the direct and indirect relation between each two nodes.

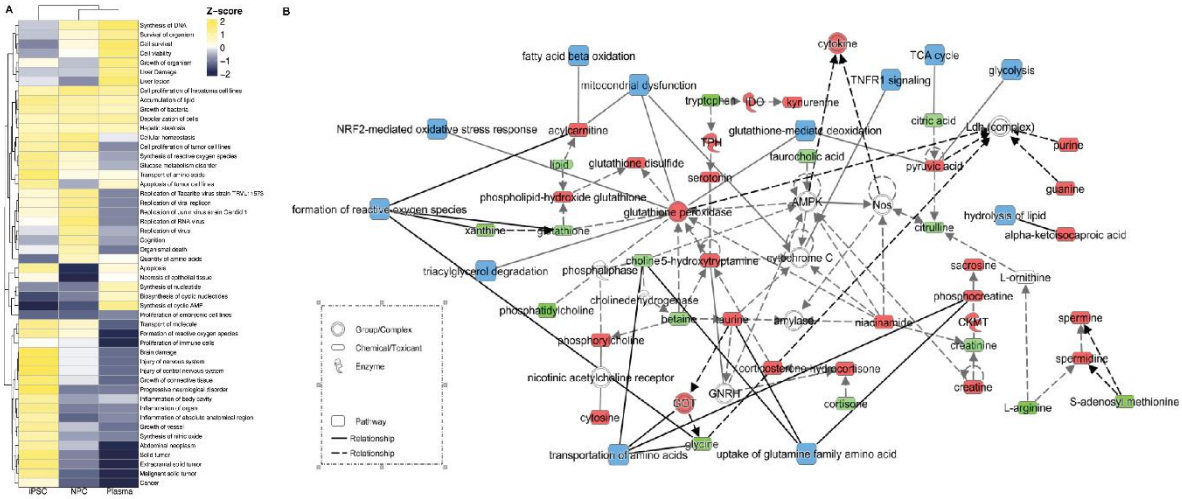

Fig 6. The summary of NGLY1 deficiency induced pathway disruption. A) The clustering of disrupted pathways in patients' plasma and patient-derived iPSC and NPC. B) Significant NGLY1 affected pathways annotated by integrating metabolome and proteome data of plasma, iPSC and NPC.

Supplementary figures:

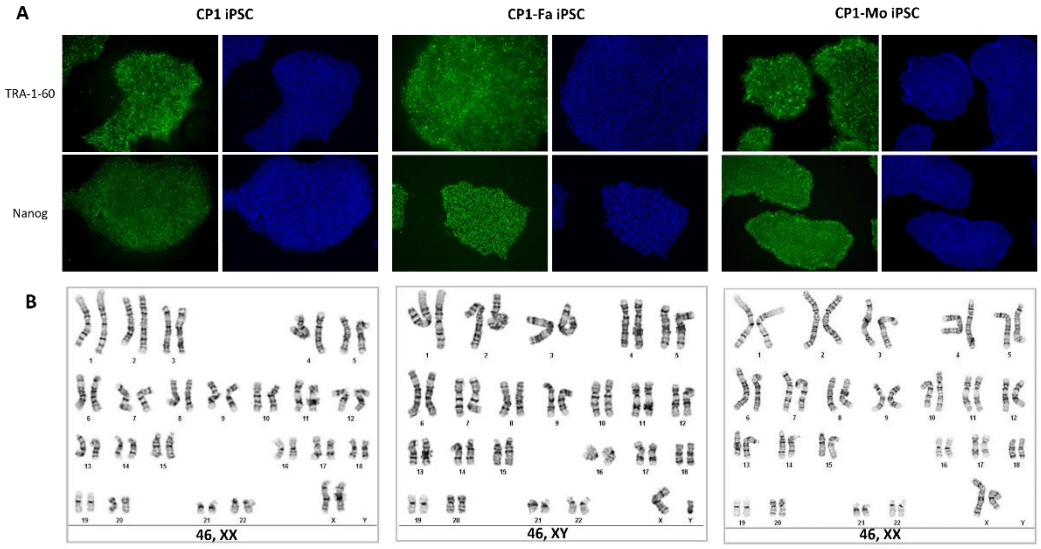

Fig S1. Characterization of patient-derived iPSCs. A) iPSCs from patient CP1's family (CP1, CP1-Fa, CP1-Mo) were confirmed with pluripotency markers expression (TRA-1-60, Nanog). B) iPSCs from CP1's family were confirmed with normal karyotyping (by G-banding).

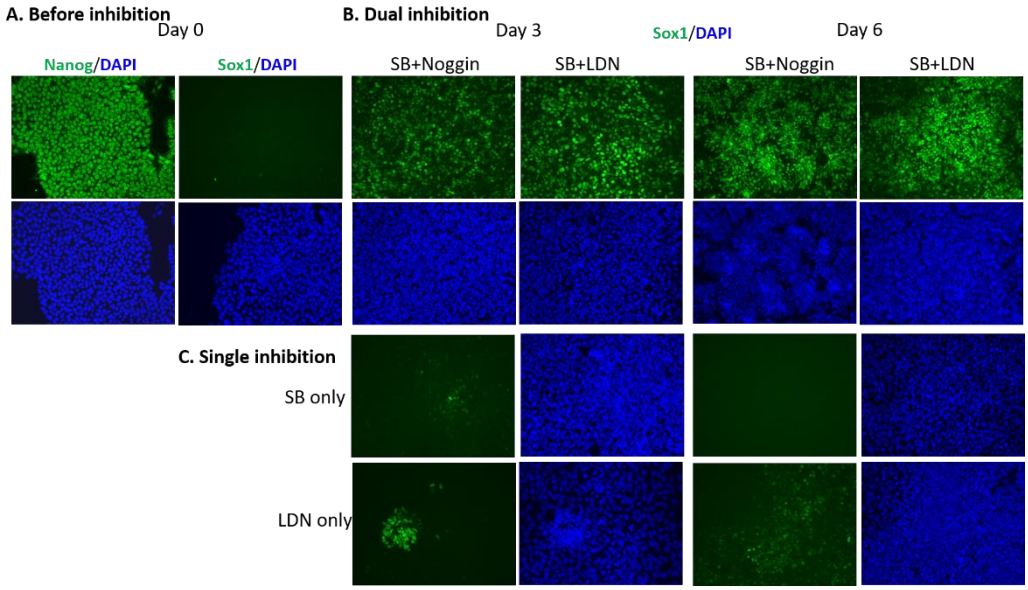

Fig S2. The induction of iPSCs by dual and single inhibition. A) Fluorescence imaging of iPSCs before inhibition on day 0. B) The induction of iPSCs to NPCs by dual inhibition (SB+Noggin, SB+LDN). C) The induction of iPSCs to NPCs by single inhibition (SB, LDN).

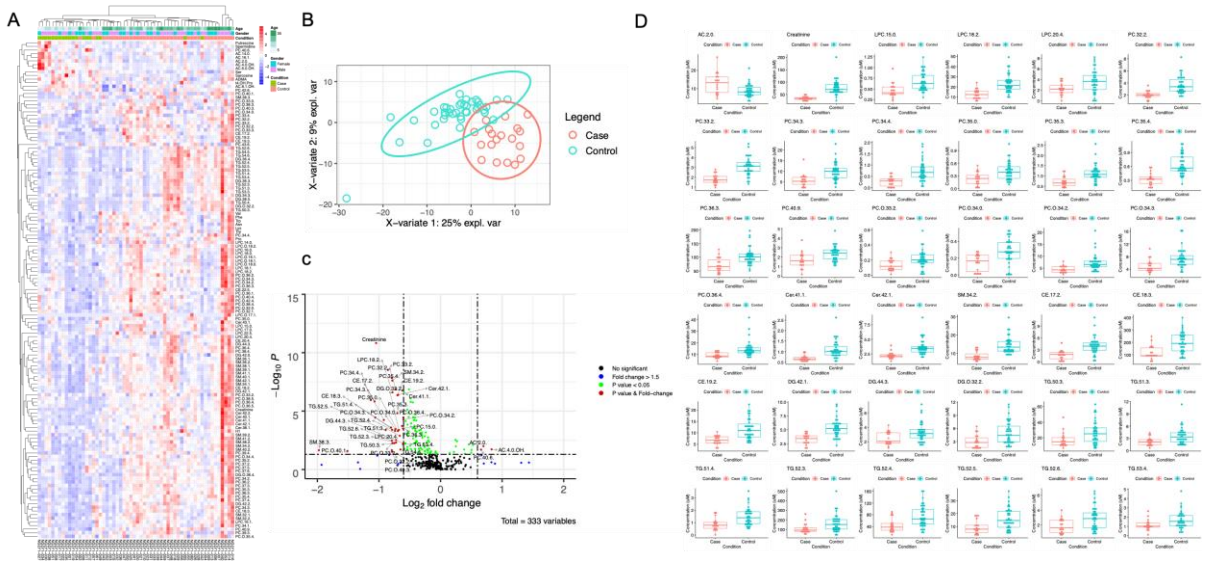

Fig S3. Summary of the targeted metabolomics profiling of NGLY1 deficiency plasma samples. A) The hierarchical clustering of metabolites with significant variation ($P < 0.05$) between patients and controls. B) PLS regression of targeted metabolomics profiles demonstrated separation between patients and controls. C) Volcano plot of plasma metabolites with significant variation and fold changes between patients and control. D) Boxplots of plasma metabolites with most significant fold changes and variation between patients and controls.

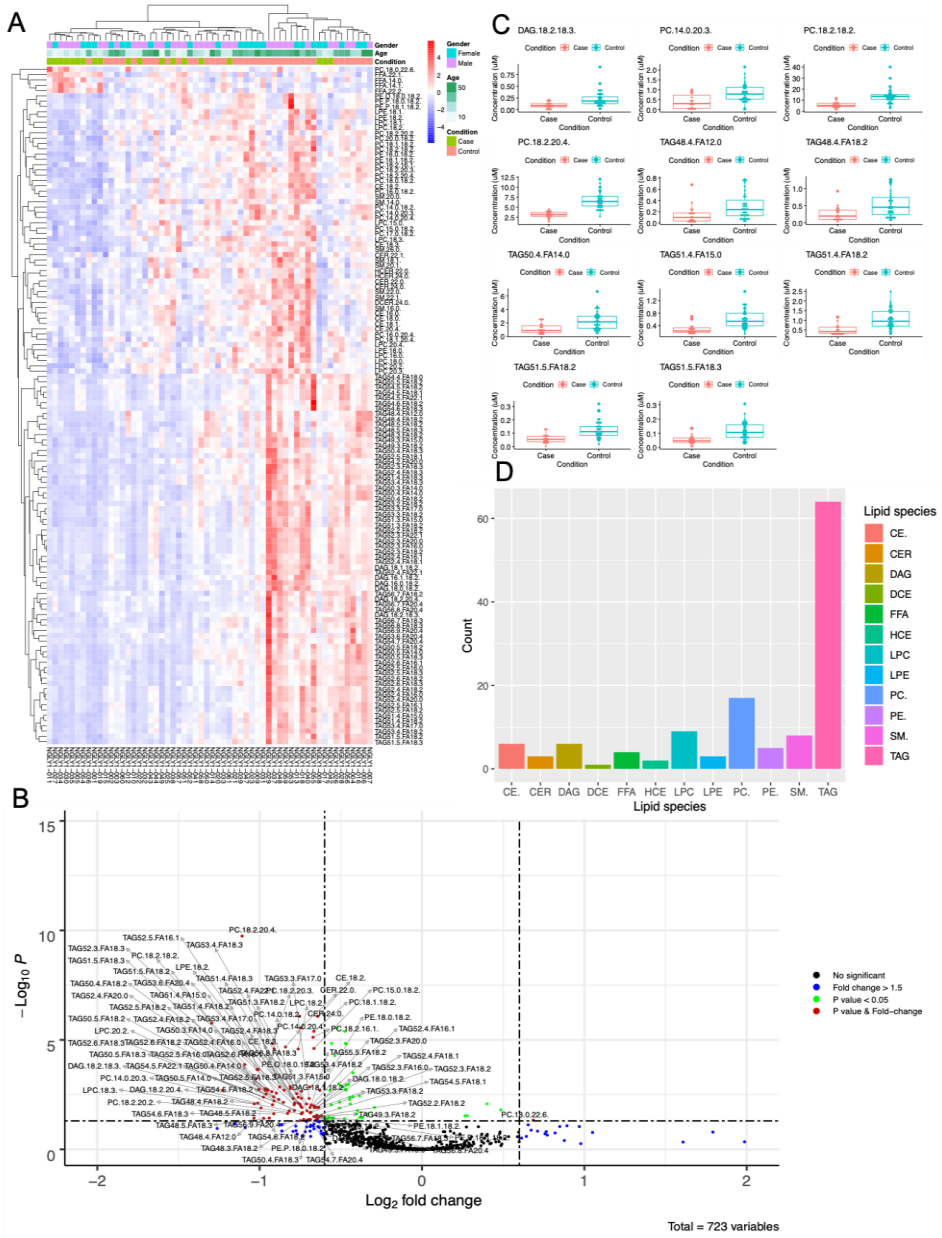

Fig S4. Summary of the lipidomics profiling of NGLY1 deficiency plasma samples. A) The hierarchical clustering of metabolites with significant variation ($P < 0.05$) between patients and controls. B) Volcano plot of plasma metabolites with significant variation and fold changes between patients and control. C) Boxplots of plasma metabolites with most significant fold changes and variation between patients and controls. D) Amounts of lipid species with significant variation between patients and controls.

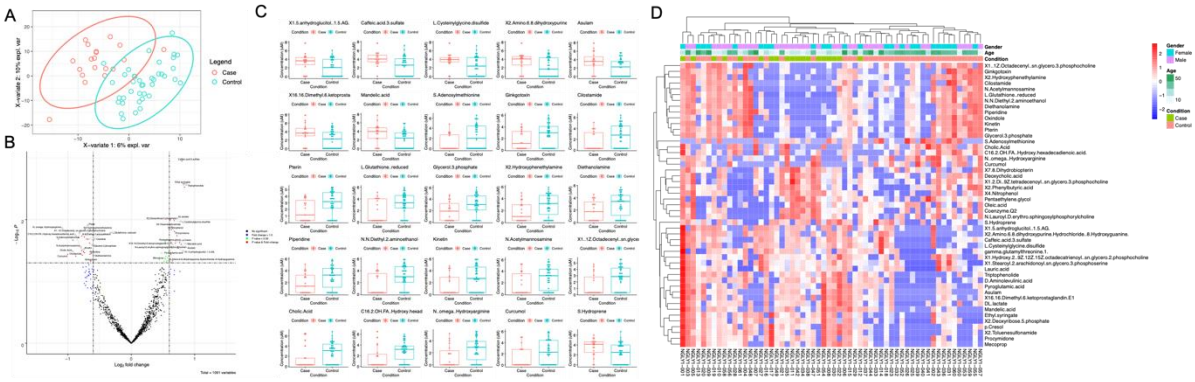

Fig S5. Summary of the untargeted metabolomics profiling of NGLY1 deficiency plasma samples. A) PLS regression of untargeted metabolomics profiles demonstrated separation between patients and controls. B) Volcano plot of plasma metabolites with significant variation and fold changes between patients and control. C) Boxplots of plasma metabolites with most significant fold changes and variation between patients and controls. D) The hierarchical clustering of metabolites with significant variation ($P < 0.05$) between patients and controls.

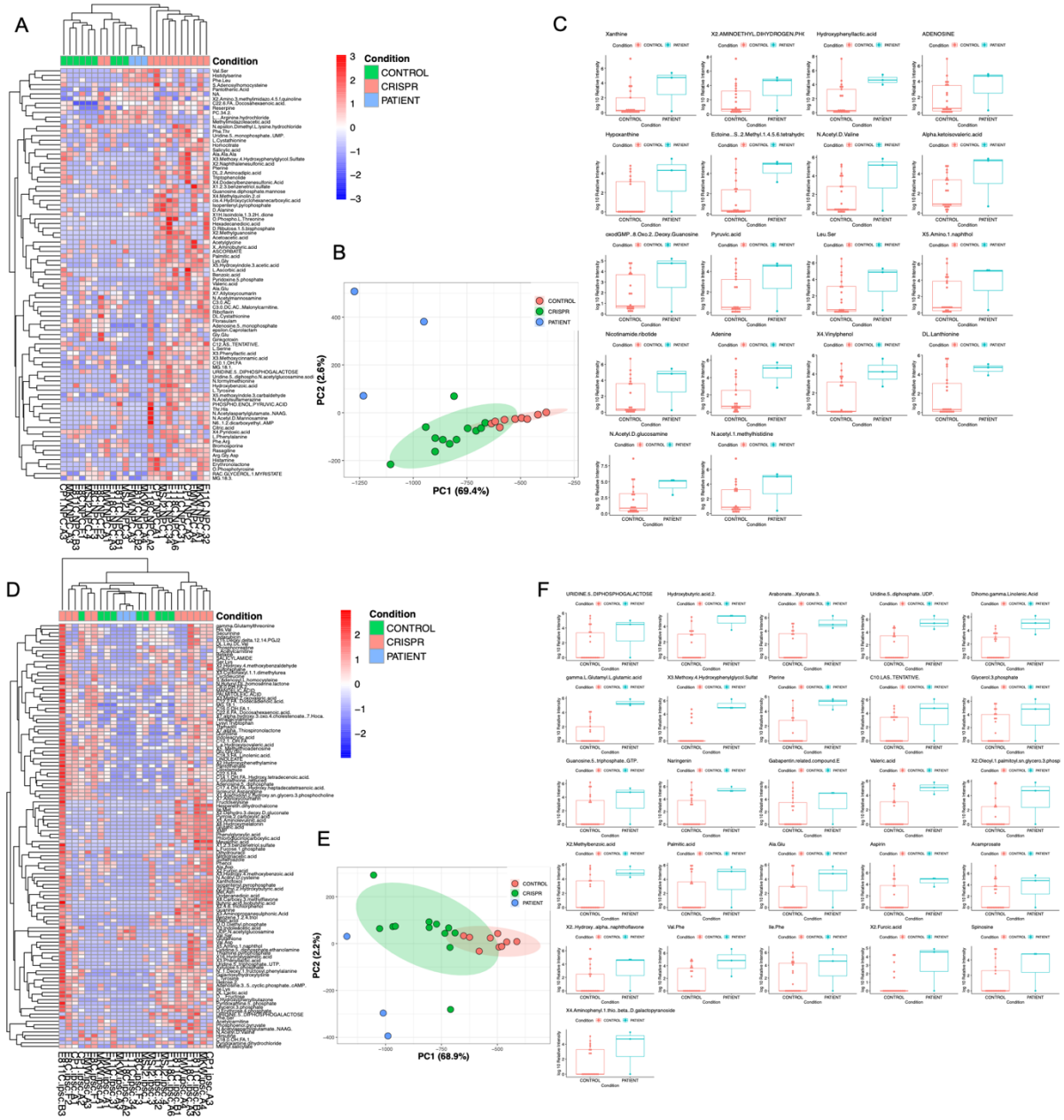

Fig S6. Summary of the untargeted metabolomics profiling of personalized disease-modeling iPSC and NPC. A) The hierarchical clustering of metabolites of level 1 and 2 ID. B) PCA of IPSC metabolomics profiles demonstrated separation between patients and controls. The external C) Boxplot of IPSC metabolites with vast fold changes between patients and control. D) The hierarchical clustering of NPC metabolites of level 1 and 2 ID. B) PCA of NPC metabolomics profiles demonstrated separation between patients and controls. The external C) Boxplot of NPC metabolites with vast fold changes between patients and control.

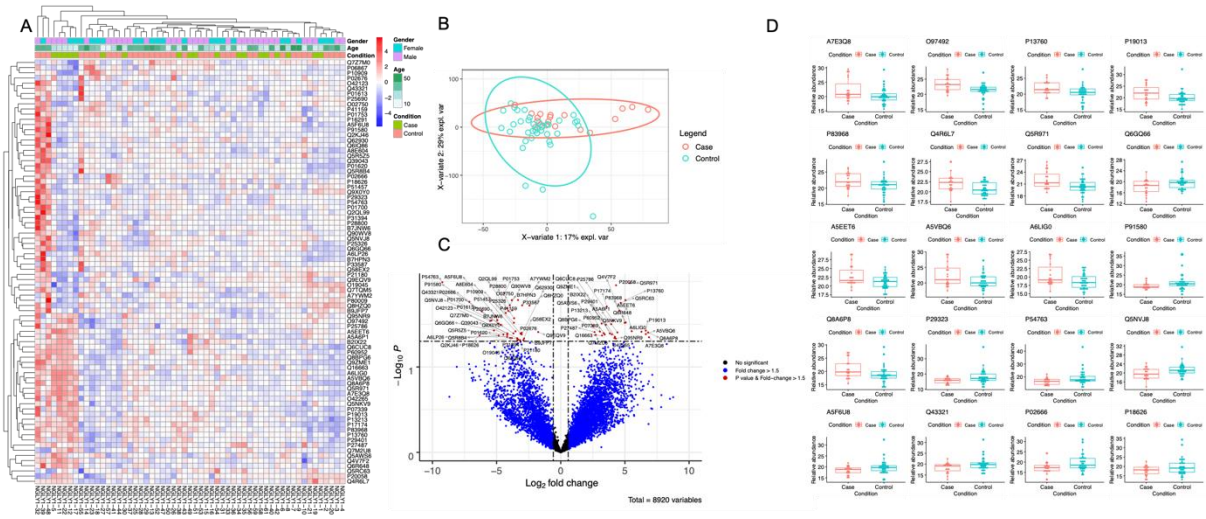

Fig S7. Summary of proteomics profiling of NGLY1 deficiency plasma samples. A) The hierarchical clustering of proteins with significant variation ($p < 0.05$) between patients and controls. B) PLS regression of proteomics profiles demonstrated separation between patients and controls. C) Volcano plot of plasma proteins with significant variation and fold changes between patients and control. D) Boxplots of plasma proteins with most significant fold changes and variation between patients and controls.

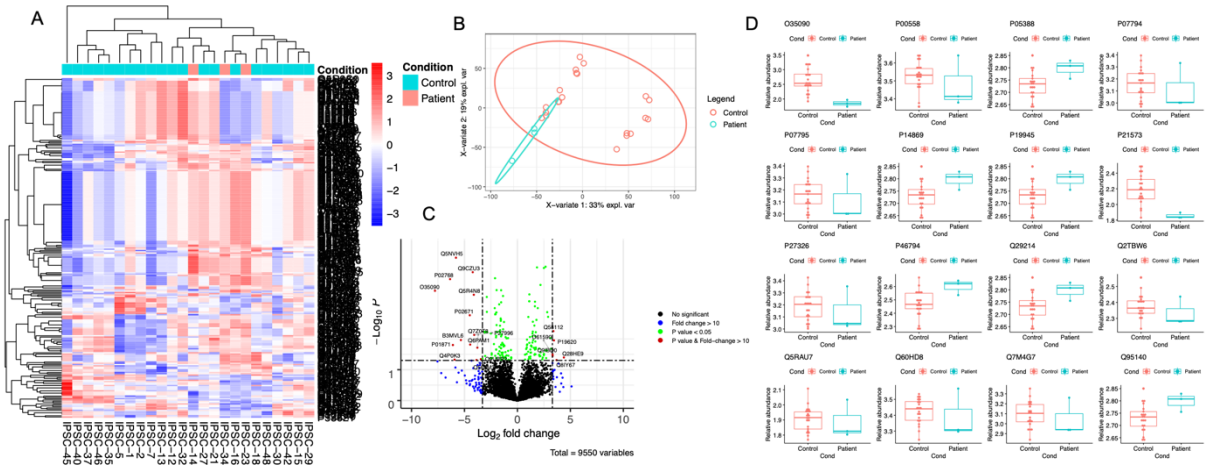

Fig S8. Summary of proteomics profiling of disease-modeling iPSC derived from Patient CP1's family and external controls. A) The hierarchical clustering of proteins with significant variation ($P < 0.05$) between patients and controls. B) PLS regression of proteomics profiles demonstrated partial separation between patients and controls. C) Volcano plot of iPSC proteins with significant variation and fold changes between patients and control. D) Boxplots

of iPSC proteins with most significant fold changes and variation between patients and controls.

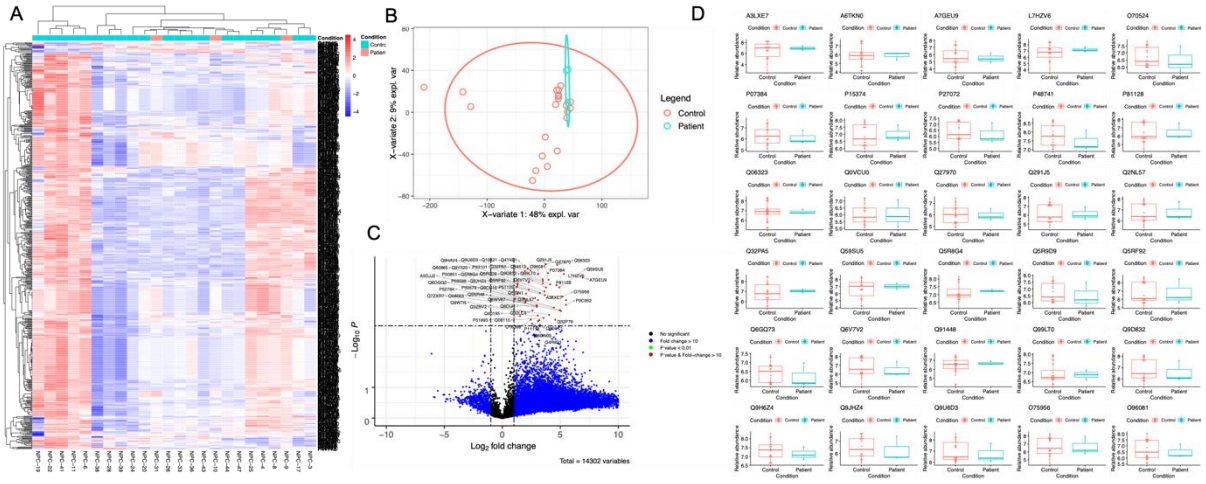

Fig S9. Summary of proteomics profiling of disease-modeling NPC derived from Patient CP1's family and external controls. A) The hierarchical clustering of proteins with significant variation ($P < 0.05$) between patients and controls. B) PLS regression of proteomics profiles demonstrated partial separation between patients and controls. C) Volcano plot of NPC proteins with significant variation and fold changes between patients and control. D) Boxplots of NPC proteins with most significant fold changes and variation between patients and controls.

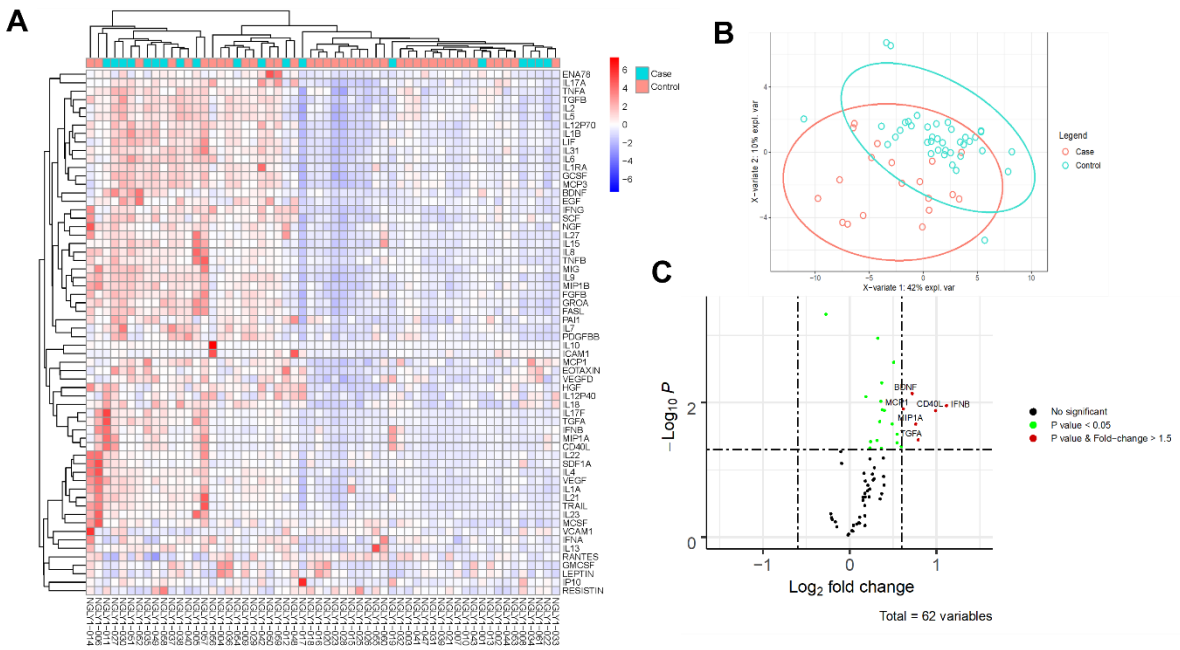

Fig S10. Summary of cytokine measurement of NGLY1 biobank plasma samples. A) The hierarchical clustering of proteins with significant variation ($P < 0.05$) between patients and controls. B) PLS regression of cytokine profiles demonstrated partial separation between patients and controls. C) Volcano plot of cytokines with significant variation and fold changes between patients and control.