HGG Advances

ARTICLE

Large 22q13.3 deletions perturb peripheral transcriptomic and metabolomic profiles in Phelan-McDermid syndrome

Michael S. Breen,[1,2,3,4,18,*] Xuanjia Fan,[1,2] Tess Levy,[1,2] Rebecca M. Pollak,[1,2] Brett Collins,[1,2] Aya Osman,[1,2] Anna S. Tocheva,[3,5] Mustafa Sahin,[6,7] Elizabeth Berry-Kravis,[8,9] Latha Soorya,[10] Audrey Thurm,[11] Craig M. Powell,[12,13] Jonathan A. Bernstein,[14] Alexander Kolevzon,[1,2,15] Joseph D. Buxbaum,[1,2,3,4,16,17,*] and on behalf of the Developmental Synaptopathies Consortium

Summary

Phelan-McDermid syndrome (PMS) is a rare neurodevelopmental disorder caused at least in part by haploinsufficiency of the SHANK3 gene, due to sequence variants in SHANK3 or subtelomeric 22q13.3 deletions. Phenotypic differences have been reported between PMS participants carrying small "class I" mutations and large "class II" mutations; however, the molecular perturbations underlying these divergent phenotypes remain obscure. Using peripheral blood transcriptome and serum metabolome profiling, we examined the molecular perturbations in the peripheral circulation associated with a full spectrum of PMS genotypes spanning class I (n = 37) and class II mutations (n = 39). Transcriptomic data revealed 52 genes with blood expression profiles that tightly scale with 22q.13.3 deletion size. Furthermore, we uncover 208 underexpressed genes in PMS participants with class II mutations, which were unchanged in class I mutations. These genes were not linked to 22q13.3 and were strongly enriched for glycosphingolipid metabolism, NCAM1 interactions, and cytotoxic natural killer (NK) immune cell signatures. *In silico* predictions estimated a reduction in CD56+ CD16– NK cell proportions in class II mutations, which was validated by mass cytometry time of flight. Global metabolomics profiling identified 24 metabolites that were significantly altered in PMS participants with class II mutations and confirmed a general reduction in sphingolipid metabolism. Collectively, these results provide new evidence linking PMS participants carrying class II mutations with decreased expression of cytotoxic cell signatures, reduced relative proportions of NK cells, and lower sphingolipid metabolism. These findings highlight alternative avenues for therapeutic development and offer new mechanistic insights supporting genotype-to-phenotype associations in PMS.

Introduction

Phelan-McDermid syndrome (PMS) is one of the most penetrant and common single-locus causes of autism spectrum disorder (ASD) and accounts for approximately 1% of ASD diagnoses.[1–3] PMS is caused by heterozygous 22q13.3 deletions or *SHANK3* sequence variants leading to haploinsufficiency of the *SHANK3* gene.[2–6] Participants with PMS present with a constellation of clinical and neurobehavioral phenotypes, including neonatal hypotonia, global developmental delay, intellectual disability, severely delayed or absent speech, and/or frequent ASD.[6–8] Additional features can also include seizures, motor skill deficits, and structural brain abnormalities.[8] Heterogeneity in the clinical presentation of PMS is not fully explained by sequence variants or deletions limited to the *SHANK3* locus, emphasizing the importance of understanding the broader genetic landscape of PMS.

The majority of reported cases of PMS are caused by large 22q13.3 deletions, which encompass additional genes and can extend up to 9.2 Mb.[6–9] Given the variable nature of the deletions, it is useful to classify PMS genotypes as either class I mutations (including SHANK3 sequence variants or deletions in *SHANK3* only or *SHANK3* with *ARSA* and/or *ACR* and *RABL2B*), or class II mutations (all other deletions).[8] The largest genotype-phenotype association analysis indicates that PMS participants with class II mutations display increased rates of early developmental delays, intellectual disability, minimally verbal status, and various medical features.[8] Notably, individuals with class I mutations attained more advanced developmental milestones, which were reached at a younger age compared with those

[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [4]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [5]Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [6]Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; [7]Rosamund Stone Zander Translational Neuroscience Center and F.M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; [8]Department of Pediatrics, Rush University Medical Center, Chicago, IL, USA; [9]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA; [10]Department of Psychiatry, Rush University Medical Center, Chicago, IL, USA; [11]Neurodevelopmental and Behavioral Phenotyping Service, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; [12]Department of Neurobiology, University of Alabama at Birmingham Heersink School of Medicine, Birmingham, AL, USA; [13]Civitan International Research Center, University of Alabama at Birmingham Heersink School of Medicine, Birmingham, AL, USA; [14]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA; [15]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [16]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [17]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[18]Lead contact
*Correspondence: michael.breen@mssm.edu (M.S.B.), joseph.buxbaum@mssm.edu (J.D.B.)
https://doi.org/10.1016/j.xhgg.2022.100145.

with class II mutations, and were more likely to exhibit higher language and communication skills.[8] These results are largely consistent with smaller independent reports,[6,7,10] and together emphasize that the frequency and severity of PMS phenotypes is likely caused by haploinsufficiency of multiple additional candidate genes. A next practical step would be to identify consistent molecular changes resulting from these specific genetic alterations in individuals with PMS.

SHANK3 is a scaffolding protein of the postsynaptic density of glutamatergic synapses[11–13]; additional disrupted genes within larger class II mutations have been implicated in processes related to stress and inflammation, mitochondrial function, neuronal differentiation, and cellular metabolism.[5,14–16] Molecular profiling of tissues derived from PMS participants, albeit scarce, confirm these functional categories and support the notion of unique molecular programs underlying distinct clinical subtypes. For example, increased severity of PMS phenotypes and larger 22q13.3 deletions have been associated with alterations in mitochondrial complex I and IV activity,[14] changes in peripheral blood epi-signatures enriched for neuronal development and intracellular signaling,[15] and metabolomic changes implicated in metabolic stress and response to cytokines regulating inflammation,[15,16] all poised to influence neurodevelopment. Notably, even transcriptomics of peripheral blood and postmortem brain tissue from participants with idiopathic ASD implicate changes related to inflammation, cellular proliferation/metabolism and immune dysfunction,[17–19] supporting the notion that ongoing dysregulation of the immune system echoes alterations in the central nervous system (CNS). Despite these advances, studies examining the molecular changes underlying specific genetic alterations in PMS commonly employ modest sample sizes, implement variable assessment methods for measuring clinical phenotypes, and apply differing thresholds for defining large and small deletions, making it challenging to elucidate the full spectrum of genes and pathways associated with genes disrupted on 22q13.3.

Given the success of blood transcriptome profiling to identify novel mechanisms and high-confidence targets for several rare CNS disorders,[20–23] including idiopathic ASD[19] more broadly, we hypothesized that unbiased peripheral blood transcriptomic and metabolomic profiling across a large spectrum of genotypes would shed light on the molecular changes underlying specific genetic alterations in PMS.

The objective of this study was to examine the molecular perturbations in the peripheral circulation associated with a full spectrum of PMS genotypes. A total of 76 PMS probands were included in this study. Peripheral blood transcriptomic data were generated across 68 PMS participants, including class I mutations (n = 33) and class II mutations (n = 35), as well as an age- and sex-matched control group (n = 24). In addition, global metabolomic data were generated across a partially overlapping subset of 25 PMS participants, comprised of class I mutations (n = 11), class II mutations (n = 14), and an age- and sex-matched control group (n = 29). Using a combination of genotypic, transcriptomic, and metabolomics data, we sought to: (1) elucidate key genes, pathways and cell types altered in PMS participants with class I and class II mutations; (2) explore molecular relationships between gene expression patterns and clinical features of PMS; and (3) identify core sets of differentially abundant metabolites in PMS participants with class I and class II mutations. We identify a molecular footprint of class II mutations, which informs pathobiological mechanisms in PMS and suggests approaches for interventions.

Material and methods

Ascertainment of PMS participants and collection of clinical phenotypes

Informed consent was obtained from participants' caregivers for study participation, as described previously.[8] The cohort included 76 PMS participants (38 females, 38 males) between the ages of 1 and 42 years (8.9 ± 6.5). Forty-six participants were enrolled in studies at the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai. An additional 30 participants were enrolled by partner sites through the Rare Disease Clinical Research Network Developmental Synaptopathies Consortium (DSC), as part of a PMS phenotyping and natural history study. For each participant, a comprehensive battery of standardized assessments, semi-structured interviews, and caregiver report questionnaires was used to examine medical comorbidities, intellectual and adaptive functioning, expressive and receptive language, ASD symptomatology, and behavioral comorbidities, as described previously.[8] Studies were approved by the Institutional Review Board (IRB) for the protection of human subjects at Mount Sinai (study IDs: 98-0436, 10-0527, 12-1718) and Boston Children's Hospital (study ID: P00013300), which serves as the central IRB for the DSC.

Peripheral blood RNA isolation, library preparation, and quantification of gene expression

Peripheral blood was collected in PAXgene Blood RNA tubes (QIAGEN, Valencia, CA) for 68 PMS participants. Peripheral blood was also collected from 24 unaffected control subjects (12 females, 12 males) between the ages of 1 and 24 years (9.5 ± 4.9), 21 of whom were unaffected familial siblings. Total RNA was extracted and purified in accordance with the PAXgene Blood RNA Kit instructions (QIAGEN). Globin mRNA was depleted from samples using the GLOBINclear Human Kit (Life Technologies, Carlsbad, CA). The quantity of purified RNA was measured on a NanoDrop 2000 Spectrophotomerter (Thermo Scientific; 61.4 ± 24.1 ng μL^{-1}) and RNA integrity numbers measured with the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA; 8.0 ± 0.3). The Illumina TruSeq Total RNA kit (Illumina, San Diego, CA) was used for library preparation according to manufacturer's instructions without any modifications. Indexed RNA libraries were pooled and sequenced using long paired-end chemistry (2 × 150 bp) at an average read depth of ~11M reads per sample using the Illumina HiSeq2500. All high-quality trimmed reads were mapped to UCSC *Homo sapiens* reference genome (build hg37) using default STAR v.2.5.3 parameters.[23] Samtools was used to convert

bamfiles to samfiles, and featureCounts[24] was used to quantify gene expression levels for each individual sample using default paired-end parameters.

RNA sequencing data quality control

Raw count data measured 56,632 genes across 92 participants. Unspecific filtering removed lowly expressed genes that did not meet the requirement of a minimum of 1 count per million in at least 15 subjects (~16% of subjects). A total of 16,285 genes were retained and defined as stably expressed in peripheral blood. These genes were subjected to limma VOOM normalization[25] and inspected for outlying samples using unsupervised hierarchical clustering of subjects (based on Pearson coefficient and average distance metric) and principal-component analysis to identify potential outliers outside two standard deviations from these averages. No such outliers were identified in the current dataset.

Gene-based annotations for loss-of-function intolerance

We collected probability of loss-of-function (LoF) intolerance (pLI) scores from the gnomAD project. pLI scores indicate whether a gene is intolerant for either heterozygous or homozygous LoF variants, and was used to classify disrupted genes on 22q13.3 as either definitely LoF intolerant ($pLI \geq 0.9$), possible LoF intolerant ($0.5 \geq pLI < 0.9$) or definitely LoF tolerant ($pLI \leq 0.1$).

Differential gene expression and association testing

A moderated t test implemented through the *limma* package[25] was used to assess differential gene expression between unaffected controls and three different groupings of PMS participants: (1) all PMS participants; (2) class I mutations only; and (3) class II mutations only. We also (4) tested for differences between participants with class II mutations and class I mutations. These analyses tested PMS genotypes as the primary main outcome. Subsequently, we performed a secondary exploratory analysis examining relationships between gene expression and 19 clinical phenotypes within PMS participants only. Each clinical phenotype was tested separately and any participant with missing data would be dropped from the analysis, respectively. All analysis described here covaried for the possible influence of sex and age on gene expression differences. Significance threshold was set to a Benjamini-Hochberg (BH) multiple test corrected $p < 0.05$ to control the false discovery rate (FDR), unless specified otherwise.

Functional annotation of differentially expressed genes

Correlation adjusted mean rank (CAMERA) gene set enrichment was performed using the resulting sets of summary statistics.[25,26] CAMERA performs a competitive gene set rank test to assess whether the genes in a given gene set are highly ranked in terms of differential expression relative to genes that are not in the gene set. For example, the test ranks gene expression differences in PMS participants with class II mutations relative to unaffected controls to test whether gene sets are over-represented toward the extreme ends of this ranked list. After adjusting the variance of the resulting gene set test statistic by a variance inflation factor that depends on the gene-wise correlation (which we set to default parameters, 0.01) and the size of the set, a p value is returned and adjusted for multiple testing. We used this function to test two aims: we examined each resulting set of PMS-associated changes in gene expression for enrichment of biological processes and pathways using a well-curated collection of REACTOME pathways and gene ontology molecular factors. We specifically focused on functional annotation of differentially expressed genes (1) across all PMS participants, (2) participants with class I mutations only; (3) participants with class II mutations only; and (4) changes between class I and class II mutations.

Cell type-specific gene set enrichment analysis using single-cell RNA sequencing data

Three single-cell RNA sequencing (scRNA-seq) experiments were downloaded and incorporated in this study: the first dataset comprised 10,975 peripheral blood mononuclear cells (PBMCs) (v.2 Chemistry) and the second dataset comprised 33,227 PBMCs (v.2 Chemistry), both were downloaded from the list of publicly available 10X Genomic datasets; the third data set comprised 67,272 PBMCs and was obtained from Zheng et al.[27] These three scRNA-seq PBMC datasets were leveraged due to their differences in the total number of sequenced cells, chemistries, batches, and laboratories, thus ensuring robustness and reproducibility of our enrichment results. For each dataset, we used pre-computed filtered, normalized, and scaled data together with pre-existing cell-type classifications as originally described and deposited for each dataset. Thus, no additional data processing was performed as each experiment was pre-processed and quality controlled. Next, cell-type marker genes were curated across all three experiments using the FindAllMarkers function in the Seruat R package[28] with the following specifications: min.pct = 0.25, logfc threshold = 0.01, FDR $p < 0.05$. These resulting lists of cell-type markers were compiled into cell-type-specific gene sets and used as input to perform CAMERA gene set enrichment analysis (as described above) to determine if a rank ordered list of PMS-related differentially expressed genes contained an over-representation of cell-type-specific genes toward either extreme end. A separate independent cell-type enrichment analysis was performed using the three datasets. Rather than testing for the distribution of cell-specific marker genes along a ranked list of PMS-related genes, we directly queried the expression of a given list of PMS-related genes within and across all individual single cells using singular value decomposition. Thus, the expression of each PMS-related gene set was aggregated into one singular eigengene value, which was plotted within and across all single cells as a global representative of gene expression for a given gene set of interest.

In silico cytometry estimates the proportions of peripheral blood immune cells

The frequencies of circulating blood immune cells were estimated for each individual with transcriptomic data using CIBERSORTx cell-type de-convolution.[29] CIBERSORTx relies on known cell-subset-specific marker genes and applies linear support vector regression, a machine learning approach highly robust compared with other methods with respect to noise, unknown mixture content and closely related cell types. As input, we used the LM22 signature matrix to distinguish nine main leukocytes subtypes: B cells (CD19+), T cells (CD3+), natural killer (NK) cells (CD56+), monocytes (CD14+), dendritic cells, mast cells, macrophages, eosinophils, and neutrophils. The means of the resulting estimates were compared between PMS participants and unaffected controls and tested for significance using a Student's t test.

Cytometry by time of flight: Data acquisition, preprocessing, and analysis

High-dimensional immuno-phenotyping using cytometry by time of flight (CyTOF) was performed on frozen stabilized

PBMCs from five PMS participants with class II mutations and four age-matched pediatric control participants. Thawed PBMCs were delivered to the Human Immune Monitoring Core at the Icahn School of Medicine at Mount Sinai in fresh RPMI medium. Samples were washed in cell staining buffer (CSB; Fluidigm, San Francisco, CA) and re-suspended in fresh CSB. Fc Receptor blocking (BioLegend, San Diego, CA), Rh103 viability staining (Fluidigm), and live-cell barcoding were all performed simultaneously at room temperature. After a 30-min incubation at room temperature, samples were washed twice in CSB, pooled, and stained with surface markers for 30 min at room temperature. Two CSB washes were performed. Samples were then fixed with 2.4% PFA and subsequently labeled with iridium and osmium for 30 min at room temperature. Samples were washed twice in CSB and stored in CSB until acquisition.

Prior to data acquisition, samples were washed in cell acquisition solution (Fluidigm) and resuspended at a concentration of 1 million cells per mL in cell acquisition solution containing a 1:20 dilution of EQ normalization beads (Fluidigm). The samples were then acquired on a Helios mass cytometer equipped with a wide-bore sample injector at an event rate of <400 events per second. After acquisition, repeat acquisitions of the same sample were concatenated and normalized using Fluidigm software and uploaded to Cytobank for data analysis.

Cells were first identified based on Ir-193 DNA intensity and CD45 expression; Ce140+ normalization beads, CD45-low/Ir-193-low debris and cross-sample, and Gaussian ion-cloud multiplets were excluded from downstream analysis. After this data cleanup, manual gating was utilized to debarcode the multiplexed live-cell barcoded sample. The FCS files were split by debarcoded population to complete debarcoding and data cleanup. The cell counts and frequencies of the annotated cell subsets, excluding debris and known cell-cell multiplets, were exported for downstream statistical analyses. To identify changes in cellular populations we performed differential abundance analysis using a moderated t test implemented through limma.[25] The annotated cell frequencies were used as input into a model fit using class II mutations as the outcome variable.

Global plasma metabolomics profiling and data preprocessing

Plasma was isolated from 54 participants (n = 29 unaffected controls; n = 11 class I mutations; n = 14 class II mutations) by centrifugation of blood samples in EDTA tubes for 30 min at 1,500 × *g*. Notably, 17 PMS participants and 12 unaffected controls had paired peripheral blood transcriptomic data. Separated plasma aliquots of 0.5 mL were stored immediately at −80°C until transport in dry ice for global metabolomic profiling using the analytical DiscoveryHD4 platform by Metabolon, as described previously.[30,31] Raw data were extracted and signature chromatographic peaks and relative ion concentrations for metabolites detected were identified for each sample. Spectrometry data were analyzed using the quantify individual components in a sample method.[31] Metabolite identification was performed by matching each metabolite aggregate to an annotated reference chemical library containing >4,000 metabolites with well-defined chemical profiles. Peaks were quantified using the area under the curve. Metabolite data were then normalized in terms of raw peak area counts and re-scaled to set the median equal to one. Subsequently, any missing values, which constituted ~8% of the entire data frame, were imputed with the minimum. Finally, we removed metabolites with low standard deviation (SD < 0.01) across the entire cohort, yielding 1,045 metabolites.

Metabolomics statistical analyses

A moderated t test from limma[25] was used to assess differential abundance of metabolites between unaffected controls and three different groupings of PMS participants: (1) all PMS participants; (2) class I mutations only; and (3) class II mutations only. We also (4) tested for differences between participants with class II mutations and class I mutations. These analyses adjusted for the possible influence of sex and age on metabolite profiles and a significance threshold was set to a BH multiple test corrected $p < 0.1$ to control the FDR. Differentially abundant metabolites were subjected to pathway annotation using MetaboAnalyst5.0.[32] We applied a joint pathway analysis to integrate our transcriptomic and metabolomic data and interpret them at a pathway level. To do so, the mass of each metabolite detected was queried against the Human Metabolome Database (HMDB).[33] Once identified, a list of differentially expressed genes and differentially abundant metabolites identified by the HMDB was imported into MetaboAnalyst5.0 along with their direction of effect ($\log_2$ fold changes). These results mapped to well-curated molecular pathways for over-representation analysis using hypergeometric tests, and p values were adjusted using Holm-Bonferroni correction.

Results

Clinical features of PMS participants with class I and class II mutations

A total of 76 PMS probands were included in this study (Table 1). Across the full cohort, 17 participants had sequence variants in *SHANK3*, including 13 frameshift, 2 nonsense, 1 splice site, and 1 *de novo* missense variant (Figure S1). Participants were parsed into two groups: (1) class I mutations: sequence variants or small deletions including only *SHANK3 or SHANK3* in combination with *ARSA* and/or *ACR* and *RABL2B*; and (2) class II mutations: all larger deletions that did not qualify as class I mutations. Participants with class II mutations exhibit significantly lower full-scale, verbal and nonverbal IQ/DQ relative to class I mutations ($p = 0.045$, $p = 0.023$, $p = 0.019$, respectively), consistent with existing evidence for genotype-phenotype associations in PMS.[8] Notably, class II mutations also display significantly reduced motor skills on the Vineland-2 Adaptive Behavior Scale (VABS Motor) but were also younger when compared with participants with class I mutations ($p = 0.006$, $p = 0.02$, respectively). A subset of PMS participants in this study underwent peripheral blood transcriptome profiling (n = 68) and/or serum metabolomic profiling (n = 25) (Table S1), which were the focus of the subsequent analyses.

Class II mutations, but not class I mutations, alter transcriptional profiles in peripheral blood

Peripheral blood transcriptomic data were generated across 68 PMS participants, including class I sequence variants and mutations (n = 33) and class II mutations (n = 35), as well as an age- and sex-matched control group

Table 1. Clinical features of PMS participants harboring class II and class I mutations in this study

	Entire cohort of 76 PMS probands				Transcriptome subset (68 probands)		Metabolome subset (25 probands)	
	Class II (n = 39)	Class I (n = 37)	p value	Effect size	p value	Effect size	p value	Effect size
Sex, M/F	16/23	22/15	0.169	−0.180	1.000	0.010	0.877	−0.120
Caucasian/other	33/6 (84%)	33/4 (89%)	0.802	−0.070	1.000	−0.010	0.565	0.250
Hispanic/other	4/35 (11%)	4/33 (12%)	1.000	−0.010	0.215	−0.180	0.363	−0.250
Age (months)	85.821 ± 47.781	129.432 ± 97.734	0.026	−0.579	0.039	−0.564	0.317	−0.180
Verbal IQ/DQ	22.354 ± 17.494	33.617 ± 24.181	0.045	−0.546	0.034	−0.645	0.305	−0.490
Nonverbal IQ/DQ	28.688 ± 17.237	39.973 ± 21.916	0.023	−0.583	0.017	−0.699	0.222	−0.383
Full scale IQ/DQ	24.739 ± 16.003	34.218 ± 19.943	0.019	−0.535	0.016	−0.623	0.272	−0.437
ADOS total	15.912 ± 6.122	15.821 ± 8.094	0.854	0.013	0.891	0.085	0.717	0.745
ADOS SA	12.941 ± 5.116	11.882 ± 6.741	0.676	0.180	0.464	0.276	0.802	0.880
ADOS RRB	3 ± 1.969	3.939 ± 2.318	0.066	−0.444	0.063	−0.463	0.431	−0.475
ADI communication	12.697 ± 4.355	12.065 ± 4.761	0.425	0.141	0.425	0.070	0.410	1.401
ADI social	19.529 ± 6.872	17.031 ± 8.656	0.313	0.326	0.359	0.308	0.094	1.880
ADI RRB	4.771 ± 2.426	4.594 ± 3.12	0.743	0.065	0.822	0.050	0.451	1.241
VABS communication	50.564 ± 15.441	52.722 ± 19.823	0.531	−0.124	0.268	−0.333	0.795	−0.949
VABS DLS	51.821 ± 13.483	53.278 ± 16.77	0.702	−0.098	0.505	−0.154	0.897	−0.725
VABS social	56.513 ± 12.941	60.75 ± 17.977	0.313	−0.276	0.230	−0.361	0.938	−0.756
VABS motor	55.324 ± 11.954	63.188 ± 10.187	0.006	−0.714	0.004	−0.774	0.051	−0.117
VABS comp	51.308 ± 12.404	55.306 ± 16.603	0.190	−0.278	0.171	−0.327	0.203	−0.503
ABC irritability	10.114 ± 9.536	8.286 ± 11.184	0.158	0.179	0.117	0.312	0.560	0.707
ABC social withdrawal	10.943 ± 9.142	9.029 ± 7.86	0.485	0.228	0.328	0.270	0.950	0.992
ABC stereotypic behavior	4.857 ± 4.846	5.371 ± 6.193	0.887	−0.094	0.714	0.096	0.269	0.239
ABC hyperactivity	19.265 ± 12.425	19 ± 13.083	0.876	0.021	0.979	−0.027	0.829	−0.741
ABC inappropriate speech	1.457 ± 2.091	3.371 ± 5.719	0.113	−0.451	0.259	−0.343	0.055	−0.071
Recurrent infections, Y/N	13/24 (54%)	11/22 (50%)	1.000	0.012	1.000	0.010	1.000	0.010

Continuous measures were analyzed using a Mann-Whitney U test and effect sizes computed using Cohen's D. Discrete measures were analyzed using a chi-square test and effect sizes computed using the phi coefficient. ABC, Aberrant Behavior Checklist; ADI, Autism Diagnostic Interview; ADOS, Autism Diagnostic Observation Schedule; DLS, daily living skills; DQ, developmental quotient; IQ, intellectual quotient; RRB, restrictive and repetitive behavior; SA, Social Affect; VABS, Vineland Adaptive Behavior Scales.
Recurrent infections were clinician graded and defined as more than two pneumonia or sinus infections per year.

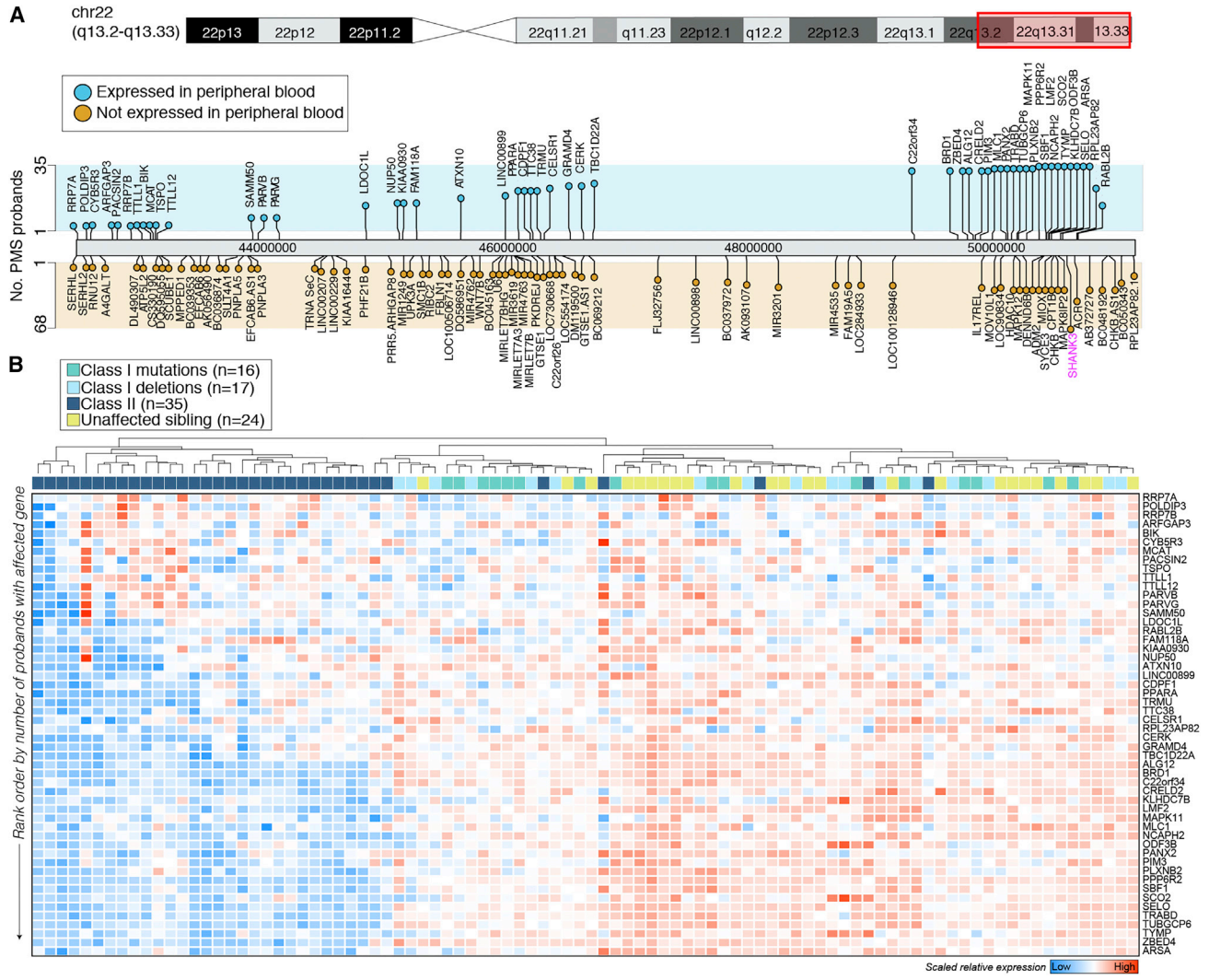

Figure 1. The landscape of class I and class II mutations in PMS

(A) Lollipop plot of genes affected by class I mutations and class II mutations across the terminal end of the long arm of chromosome 22 (22q13.3) in the 68 PMS probands included in the study. Genes are displayed as either expressed (blue; n = 52 genes) in peripheral blood or not (orange; n = 76 genes) and ranked by the number of probands harboring the affected gene (y axis). *SHANK3* is highlighted in pink. (B) Unsupervised hierarchal clustering and heatmap (blue, low; red, high) depiction of the 52 genes on 22q13.3 that are expressed in peripheral blood affected by class I and class II mutations. Note that class I mutations are parsed into two groups: sequence variants (n = 16; green) and deletions (n = 17; light blue). Clustering distinguishes probands with class II mutations from those with class I mutations and unaffected controls. Genes were rank ordered by the number of PMS participants with the affected gene (y axis; rare to more frequent).

(n = 24), which largely consisted of unaffected siblings (~91%). Given the breadth of genes affected by large class II mutations on the terminal end of the long arm of chromosome 22 (22q13.3) (Figure S2A), we queried which of these genes are stably expressed in peripheral blood. Of 128 genes affected by large class II mutations, 52 genes were stably expressed and detected in peripheral blood (~40%), including genes *ARSA*, *RABL2B*, and *BRD1*, which were affected by the majority of class II mutations (Figure 1A). Unsupervised hierarchical clustering applied to these 52 blood-expressed genes on 22q13.3 accurately distinguished 85% of class II mutations (30/35) from all other participants, based on reduced expression levels of these genes (Figure 1B). Participants with class I mutations and unaffected controls clustered together and displayed higher expression levels on average for this subset of 22q13.3 genes. To determine which of the disrupted genes on 22q13.3 are intolerant to heterozygous and homozygous LoF, we computed pLI scores. Using this metric, we classified eight of the 52 blood-expressed genes as "possibly LoF intolerant" ($0.5 \geq pLI < 0.9$) and eight additional genes as "definitely LoF intolerant" ($pLI \geq 0.9$), including genes *TRABD*, *PIM3*, *TBC1D22A*, *ZBED4*, *PLXNB2*, *BRD1*, *GRAMD4*, and *CELSR1* (Figure S2B). Notably, all 52 genes are broadly expressed across 30 distinct human tissues from the GTEx project (Figure S2C), suggesting that their disruption may affect diverse biological systems across a range of tissues.

Transcriptome-wide differences in gene expression were modelled for (1) all PMS participants, (2) class I mutations,

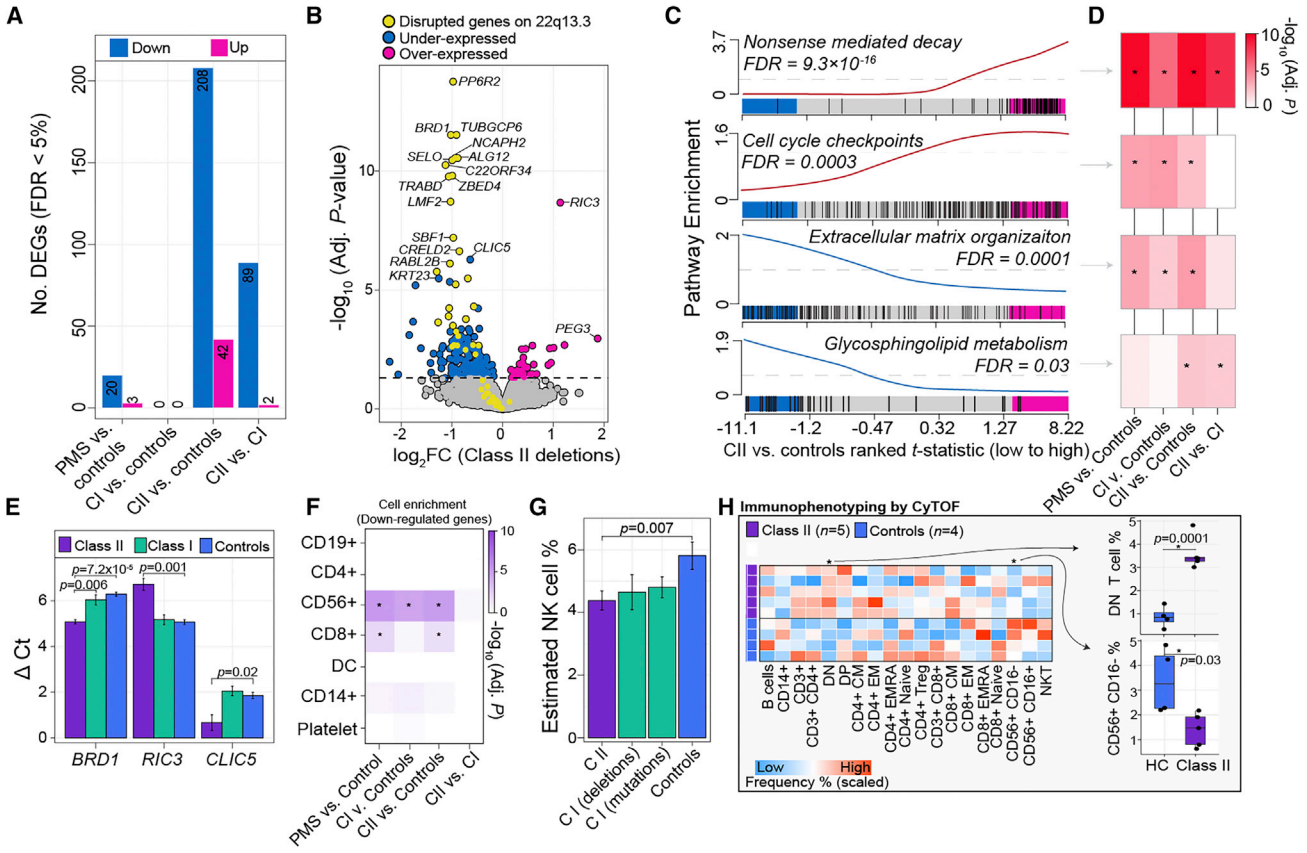

Figure 2. Altered peripheral blood gene expression profiles in class II mutations

(A) The total number of differentially expressed genes (DEGs) (y axis) for each comparison (x axis). Each analysis adjusted for sex and age as covariates.

(B) Volcano plot of class II DEGs relative to unaffected controls depicting $\log_2$ fold change ($\log_2$FC; x axis) and $-\log_{10}$ FDR adjusted p value (y axis). The horizontal line indicates FDR < 5%. Genes in yellow are the 52 genes expressed in blood that are affected by large class II mutations in PMS. Genes in blue are all other downregulated genes and those in pink are all other upregulated genes.

(C) Four representative pathway enrichment scores (y axis) of class II DEGs according to ranked t statistics, high (pink) to low (blue) (x axis). All enrichment results can be found in Table S2.

(D) The resulting FDR adjusted p value enrichment for all differential comparisons reveals shared and unique gene set enrichment among class II and class I mutations.

(E) qRT-PCR validation of three target genes across four technical replicates (used to generate standard error bars) per group: *BRD1* (a downregulated gene on chr 22); *RIC3* (an upregulated gene on chr 11); and *CLIC5* (a downregulated gene on chr 6). A Student's t test was used to test delta CT values significant differences.

(F) CAMERA cell-type enrichment of underexpressed DEGs (x axis) according to seven immune cell types (y axis) reveals strong enrichment of CD56+ genes.

(G) CIBERSORTx cell-type predictions reveal a significant reduction in the frequency CD56+ cells among class II mutations. Standard error bars summarize variability across each respective group.

(H) CyTOF validates estimated cell-type proportions on a subset of controls and participants with class II mutations. Scaled frequencies across all participants for major and minor immune cell populations are presented in heatmap form (right). Boxplots of the two immune populations with significant differences ($p < 0.05$, linear model) associated with class II mutations (left).

and (3) class II mutations, each relative to unaffected controls. We also modelled for (4) differences between class II mutations and class I mutations. Overall, the largest effect was observed between participants with class II mutations and unaffected controls, uncovering 208 underexpressed genes and 42 overexpressed genes associated with class II mutations (FDR < 5%) (Figure 2A; Table S2). Similarly, differences between class II and class I mutations revealed 89 underexpressed genes and 2 overexpressed associated with class II mutations; ~84% of these genes were also differentially expressed in class II mutations compared with unaffected controls (Figure S3). There were no significant changes observed between class I mutations compared with unaffected controls. Likewise, few genes were significantly differentially expressed when comparing all PMS participants (class I and class II mutations) to unaffected controls (n = 23 genes). These results indicate that gene expression profiles are similar between participants with class I mutations and unaffected controls, whereas participants with class II mutations are associated with unique peripheral blood transcriptional signatures.

Of the significantly underexpressed genes associated with class II mutations, 31 genes were located on 22q13.3 while the remaining (~82%) were not linked to

this genomic region (Figure 2B). We identified genes chloride intracellular channel 5 (*CLIC5*) and keratin type I cytoskeletal 23 (*KRT23*) to be among the most significant underexpressed genes in class II mutations not located on 22q13.3. We also identified RIC3 acetylcholine receptor chaperone (*RIC3*) and paternally expressed 3 (*PEG3*) to be among the most significant overexpressed genes in class II mutations. A competitive gene set ranking approach was used to functionally annotate class II-related genes, revealing overexpressed genes enriched for processes related to nonsense-mediated decay (FDR $p = 9.3 \times 10^{-16}$), protein translation (FDR $p = 4.0 \times 10^{-14}$), and cell-cycle check points (FDR $p = 0.003$), while underexpressed genes enriched for extracellular matrix organization (FDR $p = 0.0001$), NCAM1 (also known as CD56) interactions (FDR $p = 0.004$), voltage-gated calcium channel activity (FDR $p = 0.006$), and glycosphingolipid metabolism (FDR $p = 0.03$) (Figures 2C and 2D), among other processes (Table S2). To ensure confidence of our results, we performed technical validation of three genes of interest by qRT-PCR, which confirmed significant under expression of *BRD1* and *CLIC4*, as well as significant overexpression of *RIC3* in class II participants relative to class I mutations and unaffected controls (Figure 2E).

To support these functional enrichment observations, we tested whether the candidate dysregulated genes indeed interact with each other at the protein level. A significant overrepresentation of direct protein-protein interactions (PPIs) was identified among differentially expressed genes in PMS participants with class II mutations ($p < 1.0e{-}16$, observed edges = 228, expected edges = 112) (Figure S4). As expected, disrupted genes on 22q13.3 displayed a higher average number of interactions (average node degree = 2.32) relative to under- and overexpressed genes in PMS participants with class II mutations (average node degree = 1.77 and 0.50, respectively). The peripheral blood PPI network derived from participants class II mutations was again enriched for components related to NCAM1 signaling and cytotoxic immune cell signatures, and featured several underexpressed hub genes, including *NCAM1*, perforin 1 (*PRF1*), and interleukin-2 receptor subunit beta (*IL2RB*), as well as genes T-box transcription factor 21 (*TBX21*) and sphingosine-1-phosphate receptor 5 (*S1PR5*), which are critical for the maturation and recruitment of CD56+ NK cells into the periphery.[34,35]

Predicting reduced NK cell-specific expression and cellular proportions in class II mutations

A multi-step approach explored the cellular origins of the differentially expressed genes in class II mutations. First, we collected genes that are significantly and highly expressed across seven main immune cell types leveraging an existing scRNA-seq experiment (see Material and methods). Using the same gene set ranking approach as above, we performed cell-type enrichment analysis and identified a significant enrichment of CD56+ NK cell genes among underexpressed genes in class II mutations (FDR $p = 6.8 \times 10^{-10}$) (Figure 2F). Notably, we also observed an enrichment for CD56+ NK cell genes among nominally significant underexpressed genes in participants' class I mutations (FDR $p = 1.5 \times 10^{-5}$) (Figures S5A and S5B). Second, we performed the reverse approach by querying the expression of the 208 underexpressed genes in class II mutations within thousands of single peripheral blood mononuclear cells across three independent experiments (see Material and methods). These analyses confirm that underexpressed genes associated class II mutations are consistently and highly expressed in CD56+ NK cells (Figure S6). Third, we performed cell-type deconvolution analysis of the bulk peripheral blood transcriptome data using an independent cell-type-specific reference marker list and confirmed a significant reduction in the proportion of estimated CD56+ NK cells in class II mutations compared with unaffected controls ($p = 0.007$) (Figure 2G). Notably, a general reduction in the proportion of CD56+ NK cells was also observed among class I mutations, albeit non-significant ($p = 0.104$). Fourth, we re-computed our differential gene expression analyses for (1) all PMS participants and (2) class II mutations relative to unaffected controls covarying for CD56+ NK cellular proportions. Adjusting for CD56+ NK cells had the largest effect on differential gene expression, and removed ~69% of differentially expressed genes in class II mutations relative to unaffected controls (Figure S5C). Fifth, we identified 25 genes, including *S1PR5*, that were highly expressed in CD56+ NK cells via scRNA-seq that were also significantly underexpressed among class II mutations and performed unsupervised hierarchical clustering, which accurately classified 82% (29/35) of participants with class II deletion from remaining samples (Figure S5D). Finally, given the critical role of S1PR5 to recruit NK cells into the peripheral circulation and to bind lipid signaling molecule sphingosine-1-phosphate (S1P),[34,35] we asked which of the disrupted genes on 22q13.3 might play a role in regulating *S1PR5* expression and/or sphingolipid metabolism. While the vast majority the 52 blood-expressed genes on 22q13.3 were highly correlated with *S1PR5*, we found that the expression of ceramide kinase (*CERK*), and parsing class II mutations according to those with the disruption of CERK relative to the remainder of class II mutations, was moderately predictive of *S1PR5* expression levels (Figure S7). This observation was strengthened by the direct PPIs observed between CERK and SIPR5 (Figure S4).

Mass cytometry validates reduced peripheral CD56+ NK cells and egress to the periphery in class II mutations

To validate these predictions, we performed CyTOF-based immunophenotyping on a subset of PMS participants with class II mutations (n = 5) and an age- and sex-matched control group (n = 4). While both controls and class II mutations had similar distributions of major immune cell subsets in peripheral blood, the frequencies of finer immune cell types were significantly altered

(Figure 2H). Specifically, we observed a significant increase in the proportions of CD3+ CD4– CD8– (double-negative) T cells and a significant reduction in the proportions of CD56+ CD16– NK cells in class II mutations ($p = 0.001$, $p = 0.03$, respectively) (Figure 2H), validating our *in silico* predictions. Collectively, these results indicate that class II mutations are associated with unique peripheral blood transcriptional changes, which might be explained by alterations in the underlying cellular composition of CD56+ NK cells and/or related cell-specific gene expression programs.

Secondary exploratory analysis reveals transcriptomic predictors of ABC-SW

A secondary exploratory analysis examined relationships between collected clinical phenotypes and peripheral blood gene expression across all 68 PMS participants. While few significant gene-trait associations ($n = 3$) were observed at FDR < 5%, relaxing the statistical assumption of significance to FDR < 10% uncovered 534 genes positively associated with differences in the Aberrant Behavior Checklist Social Withdrawal (ABC-SW) subscale[36] and 479 genes negatively associated with ABC-SW (Figures S8A and S8B; Table S3). No other clinical phenotype was strongly associated with gene expression profiles. Genes positively associated with ABC-SW were significantly enriched for RNA binding, splicing, and protein translation, while negatively associated genes were implicated in transcription coregulatory activity, chromatin organization, and histone modifications (Figure S8C; Table S3). Genes negatively associated with ABC-SW were also significantly enriched for genes that implicate genetic risk for intellectual disability, ASD, developmental delay, and educational attainment, as well as differentially expressed genes in postmortem brain tissue from individuals with ASD (Figures S8D and S8E). Notably, ABC-SW-related genes were not enriched for an immune cell-type signature nor were associated with differences in estimated cell-type proportions (Figure S8A).

Class II mutations, but not class I mutations, reduce sphingolipid metabolism

Global metabolomic data were generated across a subset of 54 participants, comprised of class I mutations ($n = 11$), class II mutations ($n = 14$), and an age- and sex-matched control group ($n = 29$), half of which were familial related unaffected siblings (~51%). Global metabolomic profiling identified 1,045 high confidence metabolites across all 54 participants, which were largely made of lipids (37%), amino acids (19%), xenobiotics (13%), an unknown category (15%), and six other less-frequent super pathways (Figure 3A). We modelled for differential changes in metabolite abundance in PMS participants as described above, and identified 10 metabolites significantly associated with all PMS participants, 9 metabolites associated with class I mutations, and 24 metabolites associated with class II mutations relative to unaffected controls (Figure 3A; Table S4). Notably, the pattern of metabolomic effect sizes observed for class II mutations was consistent with transcriptome-wide effect sizes (Figures 3B and 2B), in that the majority of metabolites were less abundant in participants with class II mutations. Of the 24 altered metabolites associated with class II mutations, 21 were less abundant relative to unaffected controls, including 10 metabolites catalogued as part the of the sphingomyelin lipid family (Figure S9). These findings also provide validation the observed transcriptomic alterations of reduced expression of genes enriched for glycosphingolipid metabolism (Figure 2). Unsupervised hierarchical clustering of these 24 metabolites distinguished 85% (12/14) of participants with class II mutations from the remaining samples (Figure 3C). Finally, we performed an integrated analysis of differentially expressed genes and differentially abundant metabolites to elucidate their combined effect on key metabolomic pathways in participants with class II mutations. This analysis confirmed significant changes in sphingolipid metabolism followed by alterations in arginine and proline metabolism and linoleic acid metabolism (Figure 3D; Table S4).

Discussion

While increased frequency and severity of PMS phenotypes are associated with larger deletions, the molecular perturbations that result from specific genetic alteration remain poorly understood. This study presents the largest set of PMS genotypes associated with peripheral blood gene expression and global serum metabolites conducted to date, and highlights several candidate genes, pathways, metabolites, and cell types uniquely linked to PMS cases with class II mutations, despite the disruption of *SHANK3* in all participants. This suggests that SHANK3 alone is not responsible for the molecular alterations observed in the peripheral circulation. Specifically, these findings reveal that PMS participants with class II mutations display decreased expression of key cytotoxic immune cell signatures and related processes, reductions in the proportions of cytotoxic cell types, and reduced sphingolipid metabolism. Below we discuss the biological and clinical implications of our results.

Of the disrupted genes in the 22q13.3 region, 52 genes (~40%) were detected in peripheral blood, reduced in expression patterns, and largely predictive of deletion size, classifying 30/35 PMS participants with class II mutations from all other participants by unsupervised hierarchical clustering (Figure 1). Many of these disrupted, underexpressed genes are individually linked to independent rare disorders and are known to partake in diverse cellular signaling systems, including inflammatory responses (e.g., *MAPK11*),[37] glycosylation (e.g., *ALG12*),[38] mitochondrial translation (e.g., *TRMU*),[39] kinase activity (e.g., *PARVB*, *PARVG*, *CERK*, *PIM3*),[40–42] tubulin ligase activity (e.g., *TTLL12*, *TTL1*),[43] histone acetyltransferase activity

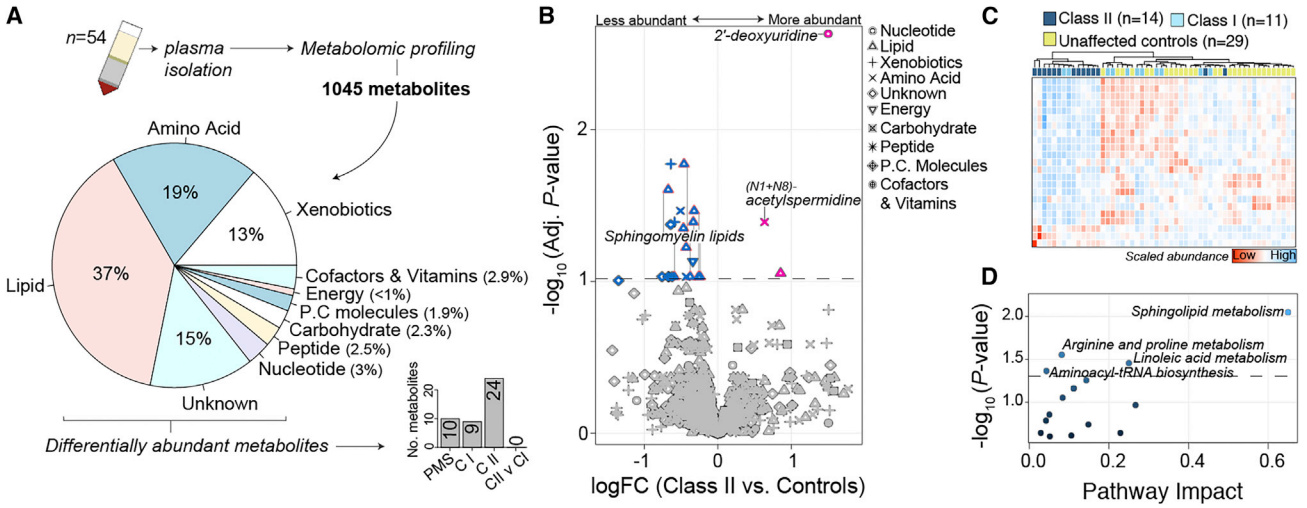

Figure 3. Plasma metabolomic profiling and alterations in participants with class II mutations

(A) Top inset: plasma was collected from 54 participants and subjected to unbiased metabolomic profiling, which generated 1,045 high-confidence metabolites for subsequent analysis. The majority of detected metabolites classified as lipids (37%), amino acids (19%), xenobiotics (13%), unknown (15%), or six other less frequent categories. Bottom inset: differential abundance of metabolites was tested and the number of significant metabolites for each comparison are displayed.

(B) Volcano plot of class II differentially abundant metabolites (DAMs) relative to unaffected controls depicting log fold change (logFC) (x axis) and $-\log_{10}$ FDR adjusted p value. The dotted horizontal line indicates a cut-off of FDR < 0.1. Metabolites are uniquely shaped according to ten super pathway categories and colored by more (red) or less (blue) abundant in participants with class II mutations. Ten sphingomyelin metabolites are outlined in pink borders.

(C) Unsupervised clustering of 24 metabolites significantly altered by class II mutations correctly classify 85% ($n = 12$) of class II mutations from the remaining samples. Heatmap depicts high (red) and low (blue) relative scaled abundance for each metabolite.

(D) Pathway analysis of metabolites altered in class II mutations reveals significant pathway enrichment (y axis; $-\log_{10}$ p value) for sphingolipid metabolism and three other metabolism pathways relative to pathway impact (x axis). Pathway impact is a combination of the centrality and pathway enrichment results computed by adding the importance measures of each of matched metabolite and dividing by the sum of the importance measures of all metabolites in each pathway.

(e.g., *BRD1*),[44,45] and sphingolipid metabolism (e.g., *ARSA*, *CERK*),[46,47] among others. For example, underexpression of modulator of VRAC current 1 (*MLC1*) and peroxisome proliferator activated receptor alpha (*PPARA*) were also observed in class II mutations, and these genes are linked to megalencephalic leukoencephalopathy with subcortical cysts disease and are known to interact with several ion channels/transporters and accessory proteins.[48,49] While these genes were also expressed across a broad collection of other human tissues (Figure S2), our results highlight the utility of peripheral blood transcriptome profiling as an accessible, alternative diagnostic read out to validate genotypic variation on 22q13.3 for PMS and several other monogenic disorders with a locus on 22q13.3.

Beyond alterations of disrupted genes on 22q13.3, the most significant transcriptomic changes were for those underexpressed genes in PMS participants with class II mutations, related to CD56+ NK cell signatures. Analysis by CyTOF confirmed a reduction in CD56+ CD16– cells in class II mutations, which are less cytotoxic than CD56+ CD16+ cells but produce greater amounts of cytokines in response to environmental cues. Underexpression of *S1PR5* was among one of many NK cell-related genes that was underexpressed in class II mutations (Figures 2, S4, and S5), and this gene is well known to promote recirculation and recruitment of CD56+ NK cells into the periphery.[34,35,50] NK cells express *S1PR5* and, in mice deficient in this receptor, NK cell distribution was altered, with reduced NK cell numbers in blood and spleen and increased numbers in the lymph node and bone marrow.[50,51] We also observed decreased expression of *TBX21* in PMS participants with class II mutations, which is required for the final maturation of NK cells[52] and is also known to induce *S1PR5*,[34,35] supporting a mechanism whereby reduced *TBX21* lends to lower expression of *S1PR5* and ultimately reduced CD56+ NK cell proportions (Figure 2G), which may increase susceptibility to viral infections in these individuals. While the rate of viral infections has not been deeply studied in PMS, the largest PMS phenotypic study to date describes that 38% of PMS participants with class I mutations (29/76) and 51% of those with class II mutations (44/87) report recurrent infections,[8] supportive of a smaller independent study where larger deletion sizes were associated with increased frequency of recurrent infections.[10] Notably, there are at least 46 single gene primary immunodeficiencies that feature NK cell deficiencies,[53] defined as either the absence of NK cells and their functions or the presences of defective NK cells within peripheral blood lymphocytes. While various therapeutics have been applied to treat individuals with NK cell deficiencies, most approaches have focused on treating susceptibility to viral infections

via the application of prophylactic antiviral drugs. Nevertheless, anecdotal cases have described apparent success using acyclovir, ganciclovir, and cytokine therapies,[54–57] such as IFN-α, to induce NK cell cytotoxic functions. However, these approaches require further investigation and consideration is currently done on a case-by-case basis.

Importantly, SHANK3 is not expressed in peripheral blood cells nor does it appear to contribute to the profound peripheral changes in PMS participants with class II mutations. Indeed, larger 22q13.3 deletions have been linked to increased rates of developmental delay and intellectual disability and moderately increased rates of recurrent infection, whereas individuals with class I mutations attain more advanced developmental milestones.[6–8,10] These clinical reports, together with the current set of findings, indicate that the frequency and severity of PMS phenotypes is indeed driven by haploinsufficiency of other candidate genes.

In an effort to resolve which of the disrupted genes in the extended 22q13.3 region might have the most direct effect on cytotoxic cell recruitment and/or sphingolipid metabolism, we predicted that the disrupted gene ceramide kinase (*CERK*), may also play a role (Figure S6). CERK is required for the phosphorylation of ceramide, which is the centerpiece of the sphingolipid metabolism and generates ceramide-1-phosphate (C1P).[47,58] Both CERK and C1P have been implicated in cellular proliferation, apoptosis, and inflammation.[59,60] Our findings from serum metabolomic profiling support these results by directly implicating reductions of sphingolipid metabolism in PMS participants with class II mutations (Figure 3). In addition to phosphorylation by CERK, ceramide can be hydrolyzed to sphingosine, which is phosphorylated to S1P by sphingosine kinases.[46,47,58] Both C1P and S1P are bioactive molecules critical for immune function and inflammation,[46,47,58–60] but also play an important role in neurotransmitter release and synaptic transmission in the brain.[61] Whereas ceramide and sphingosine are associated with cellular growth arrest and apoptosis, S1P is associated with cellular survival and suppression of apoptosis.[58–61] To this end, we anticipate that reduced sphingolipid metabolism in PMS participants with class II mutations is associated with down-regulation of ceramide biosynthesis and/or S1P synthesis. Notably, S1P signaling via S1PR5 is particularly important for regulating NK cell migration and cytotoxicity and gradients of S1P drive NK cell chemotaxis,[34,35,62,63] essential for the mobilization of NK cells to inflamed organs, supporting our transcriptomic results. Taken together, our data provide preliminary evidence for a mechanistic model linking large 22q13.3 deletions to reductions in NK cell-related gene expression signatures (*TBX21*, *S1PR5*, *NCAM1*, and partners) and CD56+ cellular proportions in the periphery, as well as reductions in sphingolipid metabolism, which would otherwise lend to the recruitment and survival of these cells in the peripheral circulation. Notably, alterations of lipid metabolism, including S1P, have been reported in serum and postmortem brain tissues from individuals with ASD,[64,65] suggesting that impairment of lipid metabolism pathways may contribute to the pathology of ASD more broadly. Several therapeutic agents have been developed to modulate sphingolipid metabolism, including stress-signaling molecule tumor necrosis factor alpha and interleukin-1β, to induce activation of sphingomyelinases,[66,67] which can also increase ceramide and subsequent ceramide-dependent responses (i.e., cell death and/or arrest).

The largest genotype-phenotype association analysis indicates that PMS participants with class II mutations display increased rates of early developmental delays, intellectual disability, minimally verbal status, and various medical features.[8] Notably, individuals with class I mutations attained more advanced developmental milestones, which were reached at a younger age compared with those with class II mutations and were more likely to exhibit higher language and communication skills.[8] These results are largely consistent with smaller independent reports,[6,7,10] and together emphasize that the frequency and severity of PMS phenotypes is likely caused by haploinsufficiency of multiple additional candidate genes.

Our secondary exploratory analysis also identified several genes that are both positively and negatively associated with variations in ABC-SW across all PMS cases (FDR < 10%) (Figure S7). Notably, genes negatively associated with ABC-SW were implicated in transcription coregulatory activity, chromatin organization, and histone modifications, and were significantly enriched for genes that implicate genetic risk for neurodevelopmental disorders. Thus, individuals with high ABC-SW scores display reduced levels of expression for these genes. Given the interest in the ABC-SW as a clinical outcome assessment of treatment efficacy in clinical trials of PMS,[68,69] peripheral biomarkers that scale ABC-SW severity may serve as a valuable resource to monitor treatment responses and outcomes in PMS and other disorders that present with social withdrawal phenotypes. However, further follow-up of these genes and their dynamic expression profiles following administration of such therapeutic agents is warranted.

Our study does present some limitations. First, in this report, a clinician-made assessment was used to characterize "recurrent infections" in PMS participants, defined as more than two pneumonia or sinus infections per year. Under this measure, ∼54% of PMS participants with class II mutations (13/25) and ∼50% of those with class I mutations (11/22) report recurrent infections (Table 1). This broad definition does not encompass viral infections, nor specifically delineate the types of observed infections, severities, annual frequencies, or medication(s) described; thus limiting our ability to causally link 22q13.3 deletion sizes and the reported

transcriptomic and metabolomic alterations with specific immune phenotypes. Nevertheless, in addition to immune function, NK cells and NCAM1 are also present in the human brain and are implicated in several brain-related mechanisms, including neuronal migration, synaptic plasticity, and clearance of α-syn aggregates through the lysosomal pathway.[70–72] Thus, more work is needed to fully dissect the relationship between the molecular and clinical expressivity observed in PMS. Second, the subset of individuals with serum metabolomics profiling (n^{total} = 54) was predominately independent from those with peripheral blood transcriptomic data (n^{total} = 92), with 29 overlapping individuals. While this variation can increase synergy and confidence of the reported alterations across transcriptomic and metabolomics datasets, increasing the availability of paired data will better power discovery and interpretation of the reported cytotoxic cell signatures and sphingolipid metabolomics changes in PMS. Third, while we were able to validate the estimated reductions in NK cells in class II mutations using CyTOF-based immunophenotyping, our sample size was indeed restricted to a subset of class II mutations (n = 5) and a matched control group (n = 4) due to the prevalence of PMS.[1] Future work would benefit from a larger collection of PMS genotypes for validation. Finally, while sphingolipids are well-known mediators of cell fate, they may also change in response to drug treatment, and such alterations might also reflect differential responses to existing treatment regimens.

In conclusion, we show that participants with class II mutations present significant peripheral transcriptomic and metabolomics alterations implicating reductions in cytotoxic immune cell signatures, CD56+ CD16– cell proportions, and sphingolipid metabolism, which may contribute to a more severe and variable phenotype in PMS. More broadly, this work demonstrates the utility of studying molecules in the peripheral blood of individuals with PMS, which is a readily available specimen type in clinical practice. It is worth noting that this combination of data is not expected to successfully shed light on disrupted genes and pathways if the affected region(s) is not expressed in the analyzed tissue or if the effects of the causal variants do not affect the expression of the gene. Therefore, expert evaluation is required when prioritizing candidate genes using RNA-seq data. We can expect that combining information from multiple "omics" sources will only further improve diagnosis and define molecular subtypes of PMS and other rare disease cases in the future.

Data and code availability

The RNA-seq datasets generated during this study are available at the Gene Expression Omnibus under accession no. GSE212096. CyTOF raw data files are available at Synapse under accession no. syn35283700. Metabolomics data are available in Table S4.

Consortia

Developmental Synaptopathies Consortium: Simon K. Warfield, Benoit Scherrer, Rajna Filip-Dhima, Kira Dies, Paige Siper, Ellen Hanson, Jennifer M. Phillips.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100145.

Acknowledgments

We are sincerely indebted to the generosity of the families and patients in PMS clinics across the United States who contributed their time and effort to this study. We would also like to thank the Phelan-McDermid Syndrome Foundation for their continued support in PMS research. This study was approved by the Institutional Review Board (IRB) for the protection of human subjects at Mount Sinai (study IDs: 98–0436, 10–0527, 12–1718) and Boston Children's Hospital (study ID: P00013300), which serves as the central IRB for the Developmental Synaptopathies Consortium (DSC). The DSC (U54NS092090) is part of the Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), National Center for Advancing Translational Sciences (NCATS). Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (NINDS), the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), the National Institute of Mental Health (NIMH), and NCATS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

Declaration of interests

A.K. receives research support from AMO Pharma and consults to Acadia, Alkermes, Neuren, and GW Pharma. He serves on scientific advisory boards for Ovid Therapeutics, Jaguar Therapeutics, and Ritrova Therapeutics. M.S. reports grant support from Novartis, Biogen, Astellas, Aeovian, Bridgebio, and Aucta. He has served on scientific advisory boards for Novartis, Roche, Regenxbio, and Alkermes. The remaining authors declare that they have no competing interests.

Received: March 3, 2022
Accepted: September 23, 2022

Web resources

CIBERSORTx, https://cibersortx.stanford.edu/
Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/
gnomAD, https://gnomad.broadinstitute.org/
MetaboAnalyst5.0, https://www.metaboanalyst.ca
Synapse, https://www.synapse.org

References

1. Betancur, C., and Buxbaum, J.D. (2013). SHANK3 haploinsufficiency: a "common" but underdiagnosed highly penetrant

monogenic cause of autism spectrum disorders. Mol. Autism. *4*, 17.
2. Leblond, C.S., Nava, C., Polge, A., Gauthier, J., Huguet, G., Lumbroso, S., Giuliano, F., Stordeur, C., Depienne, C., Mouzat, K., et al. (2014). Meta-analysis of SHANK mutations in autism spectrum disorders: a gradient of severity in cognitive impairments. PLoS Genet. *10*, e1004580.
3. Boccuto, L., Lauri, M., Sarasua, S.M., Skinner, C.D., Buccella, D., Dwivedi, A., Orteschi, D., Collins, J.S., Zollino, M., Visconti, P., et al. (2013). Prevalence of SHANK3 variants in patients with different subtypes of autism spectrum disorders. Eur. J. Hum. Genet. *21*, 310–316.
4. De Rubeis, S., Siper, P.M., Durkin, A., Weissman, J., Muratet, F., Halpern, D., Trelles, M.D.P., Frank, Y., Lozano, R., Wang, A.T., et al. (2018). Delineation of the genetic and clinical spectrum of Phelan-McDermid syndrome caused by SHANK3 point mutations. Mol. Autism. *9*, 31.
5. Mitz, A.R., Philyaw, T.J., Boccuto, L., Shcheglovitov, A., Sarasua, S.M., Kaufmann, W.E., and Thurm, A. (2018). Identification of 22q13 genes most likely to contribute to Phelan McDermid syndrome. Eur. J. Hum. Genet. *26*, 293–302.
6. Soorya, L., Kolevzon, A., Zweifach, J., Lim, T., Dobry, Y., Schwartz, L., Frank, Y., Wang, A.T., Cai, G., Parkhomenko, E., et al. (2013). Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. Mol. Autism. *4*, 18.
7. Sarasua, S.M., Dwivedi, A., Boccuto, L., Chen, C.F., Sharp, J.L., Rollins, J.D., Collins, J.S., Rogers, R.C., Phelan, K., and DuPont, B.R. (2014). 2q13.32 genomic regions associated with severity of speech delay, developmental delay, and physical features in Phelan-McDermid syndrome. Genet. Med. *16*, 318–328.
8. Levy, T., Foss-Feig, J.H., Betancur, C., Siper, P.M., Trelles-Thorne, M.D.P., Halpern, D., Frank, Y., Lozano, R., Layton, C., Britvan, B., et al.; Developmental Synaptopathies Consortium (2022). Strong evidence for genotype–phenotype correlations in Phelan- McDermid syndrome: results from the developmental synaptopathies consortium. Hum. Mol. Genet. *31*, 625–637.
9. Bonaglia, M.C., Giorda, R., Beri, S., De Agostini, C., Novara, F., Fichera, M., Grillo, L., Galesi, O., Vetro, A., Ciccone, R., et al. (2011). Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. PLoS Genet. *7*, e1002173.
10. Wilson, H.L., Wong, A.C.C., Shaw, S.R., Tse, W.Y., Stapleton, G.A., Phelan, M.C., Hu, S., Marshall, J., and McDermid, H.E. (2003). Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. J. Med. Genet. *40*, 575–584.
11. Durand, C.M., Perroy, J., Loll, F., Perrais, D., Fagni, L., Bourgeron, T., Montcouquiol, M., and Sans, N. (2012). SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin- dependent mechanism. Mol. Psychiatry *17*, 71–84.
12. Halbedl, S., Schoen, M., Feiler, M.S., Boeckers, T.M., and Schmeisser, M.J. (2016). Shank3 is localized in axons and presynaptic specializations of developing hippocampal neurons and involved in the modulation of NMDA receptor levels at axon terminals. J. Neurochem. *137*, 26–32.
13. Lee, K., Vyas, Y., Garner, C.C., and Montgomery, J.M. (2019). Autism-associated Shank3 mutations alter mGluR expression and mGluR-dependent but not NMDA receptor-dependent long-term depression. Synapse *73*, e22097.
14. Frye, R.E., Cox, D., Slattery, J., Tippett, M., Kahler, S., Granpeesheh, D., Damle, S., Legido, A., and Goldenthal, M.J. (2016). Mitochondrial dysfunction may explain symptom variation in Phelan-McDermid syndrome. Sci. Rep. *6*, 19544–19552.
15. Schenkel, L.C., Aref-Eshghi, E., Rooney, K., Kerkhof, J., Levy, M.A., McConkey, H., Rogers, R.C., Phelan, K., Sarasua, S.M., Jain, L., et al. (2021). DNA methylation epi-signature is associated with two molecularly and phenotypically distinct clinical subtypes of Phelan-McDermid syndrome. Clin. Epigenetics *13*, 2–7.
16. Jain, L., Oberman, L.M., Beamer, L., Cascio, L., May, M., Srikanth, S., Skinner, C., Jones, K., Allen, B., Rogers, C., et al. (2022). Genetic and metabolic profiling of individuals with Phelan-McDermid syndrome presenting with seizures. Clin. Genet. *101*, 87–100.
17. Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.
18. Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B., and Arking, D.E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat. Commun. *5*, 5748.
19. Tylee, D.S., Hess, J.L., Quinn, T.P., Barve, R., Huang, H., Zhang-James, Y., Chang, J., Stamova, B.S., Sharp, F.R., Hertz-Picciotto, I., et al. (2017). Blood transcriptomic comparison of individuals with and without autism spectrum disorder: a combined-samples mega-analysis. Am. J. Med. Genet. B Neuropsychiatr. Genet. *174*, 181–201.
20. Frésard, L., Smail, C., Ferraro, N.M., Teran, N.A., Li, X., Smith, K.S., Bonner, D., Kernohan, K.D., Marwaha, S., Zappala, Z., Balliu, B., Davis, J.R., Liu, B., Prybol, C.J., Kohler, J.N., Zastrow, D.B., Reuter, C.M., Fisk, D.G., Grove, M.E., Davidson, J.M., Hartley, T., Joshi, R., Strober, B.J., Utiramerur, S., Undiagnosed Diseases Network; and Care4Rare Canada Consortium, Lind, L., Ingelsson, E., Battle, A., Bejerano, G., Bernstein, J.A., Ashley, E.A., Boycott, K.M., Merker, J.D., Wheeler, M.T., and Montgomery, S.B. (2019). Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat. Med. *25*, 911–919.
21. Signorelli, M., Ebrahimpoor, M., Veth, O., Hettne, K., Verwey, N., García-Rodríguez, R., Tanganyika- deWinter, C.L., Lopez Hernandez, L.B., Escobar Cedillo, R., Gómez Díaz, B., et al. (2021). Peripheral blood transcriptome profiling enables monitoring disease progression in dystrophic mice and patients. EMBO Mol. Med. *13*, e13328.
22. Breen, M.S., Stein, D.J., and Baldwin, D.S. (2016). Systematic review of blood transcriptome profiling in neuropsychiatric disorders: guidelines for biomarker discovery. Hum. Psychopharmacol. *31*, 373–381.
23. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
24. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
25. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression

analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
26. Wu, D., and Smyth, G.K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic acids research *40*, e133.
27. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049–14052.
28. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.
29. Steen, C.B., Liu, C.L., Alizadeh, A.A., and Newman, A.M. (2020). Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. Methods Mol. Biol., 135–157. In-Stem Cell Transcriptional Networks.
30. Evans, A.M., DeHaven, C.D., Barrett, T., Mitchell, M., and Milgram, E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal. Chem. *81*, 6656–6667.
31. Dehaven, C.D., Evans, A.M., Dai, H., and Lawton, K.A. (2010). Organization of GC/MS and LC/MS metabolomics data into chemical libraries. J. Cheminform. *2*, 9.
32. Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.É., Li, S., and Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic acids research *49*, W388–W396.
33. Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., et al. (2007). HMDB: the human metabolome database. Nucleic acids research *35*, D521–D526.
34. Jenne, C.N., Enders, A., Rivera, R., Watson, S.R., Bankovich, A.J., Pereira, J.P., Xu, Y., Roots, C.M., Beilke, J.N., Banerjee, A., et al. (2009). T-bet–dependent S1P5 expression in NK cells promotes egress from lymph nodes and bone marrow. J. Exp. Med. *206*, 2469–2481.
35. Evrard, M., Wynne-Jones, E., Peng, C., Kato, Y., Christo, S.N., Fonseca, R., Park, S.L., Burn, T.N., Osman, M., Devi, S., et al. (2022). Sphingosine 1-phosphate receptor 5 (S1PR5) regulates the peripheral retention of tissue-resident lymphocytes. J. Exp. Med. *219*, e20210116.
36. Aman, M.G., Singh, N.N., Stewart, A.W., and Field, C.J. (1985). Psychometric characteristics of the aberrant behavior checklist. Am. J. Ment. Defic. *89*, 492–502.
37. Arthur, J.S.C., and Ley, S.C. (2013). Mitogen-activated protein kinases in innate immunity. Nat. Rev. Immunol. *13*, 679–692.
38. Grubenmann, C.E., Frank, C.G., Kjaergaard, S., Berger, E.G., Aebi, M., and Hennet, T. (2002). ALG12 mannosyltransferase defect in congenital disorder of glycosylation type lg. Hum. Mol. Genet. *11*, 2331–2339.
39. Zeharia, A., Shaag, A., Pappo, O., Mager-Heckel, A.M., Saada, A., Beinat, M., Karicheva, O., Mandel, H., Ofek, N., Segel, R., et al. (2009). Acute infantile liver failure due to mutations in the TRMU gene. Am. J. Hum. Genet. *85*, 401–407.
40. Yen, C.F., Wang, H.S., Lee, C.L., and Liao, S.K. (2014). Roles of integrin-linked kinase in cell signaling and its perspectives as a therapeutic target. Gynecology and Minimally Invasive Therapy *3*, 67–72.
41. Sugiura, M., Kono, K., Liu, H., Shimizugawa, T., Minekura, H., Spiegel, S., and Kohama, T. (2002). Ceramide kinase, a novel lipid kinase: molecular cloning and functional characterization. J. Biol. Chem. *277*, 23294–23300.
42. Beharry, Z., Mahajan, S., Zemskova, M., Lin, Y.W., Tholanikunnel, B.G., Xia, Z., Smith, C.D., and Kraft, A.S. (2011). The Pim protein kinases regulate energy metabolism and cell growth. Proc. Natl. Acad. Sci. USA *108*, 528–533.
43. Brants, J., Semenchenko, K., Wasylyk, C., Robert, A., Carles, A., Zambrano, A., Pradeau-Aubreton, K., Birck, C., Schalken, J.A., Poch, O., et al. (2012). Tubulin tyrosine ligase like 12, a TTLL family member with SET- and TTL-like domains and roles in histone and tubulin modifications and mitosis. PLoS One *7*, e51258.
44. Christensen, J.H., Elfving, B., Müller, H.K., Fryland, T., Nyegaard, M., Corydon, T.J., Nielsen, A.L., Mors, O., Wegener, G., and Børglum, A.D. (2012). The Schizophrenia and Bipolar Disorder associated BRD1 gene is regulated upon chronic restraint stress. Eur. Neuropsychopharmacol *22*, 651–656.
45. Fryland, T., Christensen, J.H., Pallesen, J., Mattheisen, M., Palmfeldt, J., Bak, M., Grove, J., Demontis, D., Blechingberg, J., Ooi, H.S., et al. (2016). Identification of the BRD1 interaction network and its impact on mental disorder risk. Genome Med. *8*, 53.
46. Pralhada Rao, R., Vaidyanathan, N., Rengasamy, M., Mammen Oommen, A., Somaiya, N., and Jagannath, M.R. (2013). Sphingolipid metabolic pathway: an overview of major roles played in human diseases. J. Lipids *2013*, 178910.
47. Hernández-Corbacho, M.J., Salama, M.F., Canals, D., Senkal, C.E., and Obeid, L.M. (2017). Sphingolipids in mitochondria. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids *1862*, 56–68.
48. Brignone, M.S., Lanciotti, A., Camerini, S., De Nuccio, C., Petrucci, T.C., Visentin, S., and Ambrosini, E. (2015). MLC1 protein: a likely link between leukodystrophies and brain channelopathies. Front. Cell. Neurosci. *9*, 66.
49. Yang, Y., Gocke, A.R., Lovett-Racke, A., Drew, P.D., and Racke, M.K. (2008). PPAR alpha regulation of the immune response and autoimmune encephalomyelitis. PPAR Res. *2008*.
50. Walzer, T., Chiossone, L., Chaix, J., Calver, A., Carozzo, C., Garrigue-Antar, L., Jacques, Y., Baratin, M., Tomasello, E., and Vivier, E. (2007). Natural killer cell trafficking in vivo requires a dedicated sphingosine 1- phosphate receptor. Nat. Immunol. *8*, 1337–1344.
51. Mayol, K., Biajoux, V., Marvel, J., Balabanian, K., and Walzer, T. (2011). Sequential desensitization of CXCR4 and S1P5 controls natural killer cell trafficking. Blood. Blood *118*, 4863–4871.
52. Townsend, M.J., Weinmann, A.S., Matsuda, J.L., Salomon, R., Farnham, P.J., Biron, C.A., Gapin, L., and Glimcher, L.H. (2004). T-bet regulates the terminal maturation and homeostasis of NK and Vα14i NKT cells. Immunity *20*, 477–494.
53. Orange, J.S. (2013). Natural killer cell deficiency. J. Allergy Clin. Immunol. *132*, 515–525.
54. Biron, C.A., Byron, K.S., and Sullivan, J.L. (1989). Severe herpesvirus infections in an adolescent without natural killer cells. N. Engl. J. Med. Overseas. Ed. *320*, 1731–1735.
55. Etzioni, A., Eidenschenk, C., Katz, R., Beck, R., Casanova, J.L., and Pollack, S. (2005). Fatal varicella associated with selective natural killer cell deficiency. J. Pediatr. *146*, 423–425.
56. Ornstein, B.W., Hill, E.B., Geurs, T.L., and French, A.R. (2013). Natural killer cell functional defects in pediatric patients with

severe and recurrent herpesvirus infections. J. Infect. Dis. *207*, 458–468.
57. Mace, E.M., Hsu, A.P., Monaco-Shawver, L., Makedonas, G., Rosen, J.B., Dropulic, L., Cohen, J.I., Frenkel, E.P., Bagwell, J.C., Sullivan, J.L., et al. (2013). Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56bright subset. Blood *121*, 2669–2677.
58. Meacci, E., and Garcia-Gil, M. (2019). S1P/S1P receptor signaling in neuromuscolar disorders. Int. J. Mol. Sci. *20*, 6364.
59. Kim, T.J., Mitsutake, S., and Igarashi, Y. (2006). The interaction between the pleckstrin homology domain of ceramide kinase and phosphatidylinositol 4, 5-bisphosphate regulates the plasma membrane targeting and ceramide 1-phosphate levels. Biochem. Biophys. Res. Commun. *342*, 611–617.
60. Kitatani, K., Iwabuchi, K., Snider, A., and Riboni, L. (2016). Sphingolipids in inflammation: from bench to bedside. Mediators Inflamm. *2016*, 7602526.
61. Karunakaran, I., and van Echten-Deckert, G. (2017). Sphingosine 1-phosphate–A double edged sword in the brain. Biochim. Biophys. Acta Biomembr. *1859*, 1573–1582.
62. Drouillard, A., Mathieu, A.L., Marçais, A., Belot, A., Viel, S., Mingueneau, M., Guckian, K., and Walzer, T. (2018). S1PR5 is essential for human natural killer cell migration toward sphingosine-1 phosphate. J. Allergy Clin. Immunol. *141*, 2265–2268.e1.
63. Dickinson, A.J., Meyer, M., Pawlak, E.A., Gomez, S., Jaspers, I., and Allbritton, N.L. (2015). Analysis of sphingosine kinase activity in single natural killer cells from peripheral blood. Integr. Biol. *7*, 392–401.
64. Wang, H., Liang, S., Wang, M., Gao, J., Sun, C., Wang, J., Xia, W., Wu, S., Sumner, S.J., Zhang, F., et al. (2016). Potential serum biomarkers from a metabolomics study of autism. J. Psychiatry Neurosci. *41*, 27–37.
65. Kurochkin, I., Khrameeva, E., Tkachev, A., Stepanova, V., Vanyushkina, A., Stekolshchikova, E., Li, Q., Zubkov, D., Shichkova, P., Halene, T., et al. (2019). Metabolome signature of autism in the human prefrontal cortex. Commun. Biol. *2*, 234–240.
66. Canals, D., Perry, D.M., Jenkins, R.W., and Hannun, Y.A. (2011). Drug targeting of sphingolipid metabolism: sphingomyelinases and ceramidases. Br. J. Pharmacol. *163*, 694–712.
67. Adam, D., Wiegmann, K., Adam-Klages, S., Ruff, A., and Krönke, M. (1996). A novel cytoplasmic domain of the p55 tumor necrosis factor receptor initiates the neutral sphingomyelinase pathway. J. Biol. Chem. *271*, 14617–14622.
68. Kolevzon, A., Bush, L., Wang, A.T., Halpern, D., Frank, Y., Grodberg, D., Rapaport, R., Tavassoli, T., Chaplin, W., Soorya, L., and Buxbaum, J.D. (2014). A pilot controlled trial of insulin-like growth factor-1 in children with Phelan-McDermid syndrome. Mol. Autism. *5*, 54–59.
69. Fastman, J., Foss-Feig, J., Frank, Y., Halpern, D., Harony-Nicolas, H., Layton, C., Sandin, S., Siper, P., Tang, L., Trelles, P., and Zweifach, J. (2021). A randomized controlled trial of intranasal oxytocin in phelan-McDermid syndrome. Mol. Autism. *12*, 62. https://doi.org/10.1186/s13229-021-00459-1.
70. Earls, R.H., Menees, K.B., Chung, J., Gutekunst, C.A., Lee, H.J., Hazim, M.G., Rada, B., Wood, L.B., and Lee, J.K. (2020). NK cells clear α-synuclein and the depletion of NK cells exacerbates synuclein pathology in a mouse model of α-synucleinopathy. Proc. Natl. Acad. Sci. USA *117*, 1762–1771.
71. Togashi, H., Sakisaka, T., and Takai, Y. (2009). Cell adhesion molecules in the central nervous system. Cell Adhes. Migrat. *3*, 29–35.
72. Muller, D., Mendez, P., DeRoo, M., Klauser, P., Steen, S., and Poglia, L. (2010). Role of NCAM in spine dynamics and synaptogenesis. Struct. Funct. Neural Cell Adhesion Mol. *663*, 245–256.

HGGA, Volume *4*

Supplemental information

Large 22q13.3 deletions perturb peripheral transcriptomic and metabolomic profiles in Phelan-McDermid syndrome

Michael S. Breen, Xuanjia Fan, Tess Levy, Rebecca M. Pollak, Brett Collins, Aya Osman, Anna S. Tocheva, Mustafa Sahin, Elizabeth Berry-Kravis, Latha Soorya, Audrey Thurm, Craig M. Powell, Jonathan A. Bernstein, Alexander Kolevzon, Joseph D. Buxbaum, and on behalf of the Developmental Synaptopathies Consortium

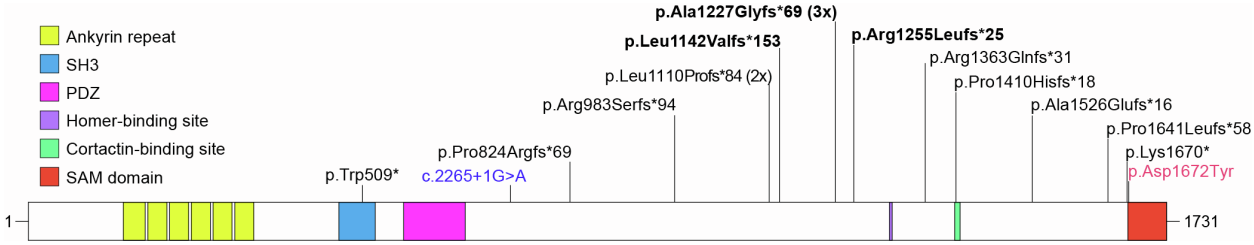

Figure S1. Landscape of *SHANK3* sequence variants in the current study. Recurrent mutations are indicated in black, missense in red and splice site variants in blue. Protein domains are from UniProt; the homer and cortactin binding sites are indicated as previously reported.

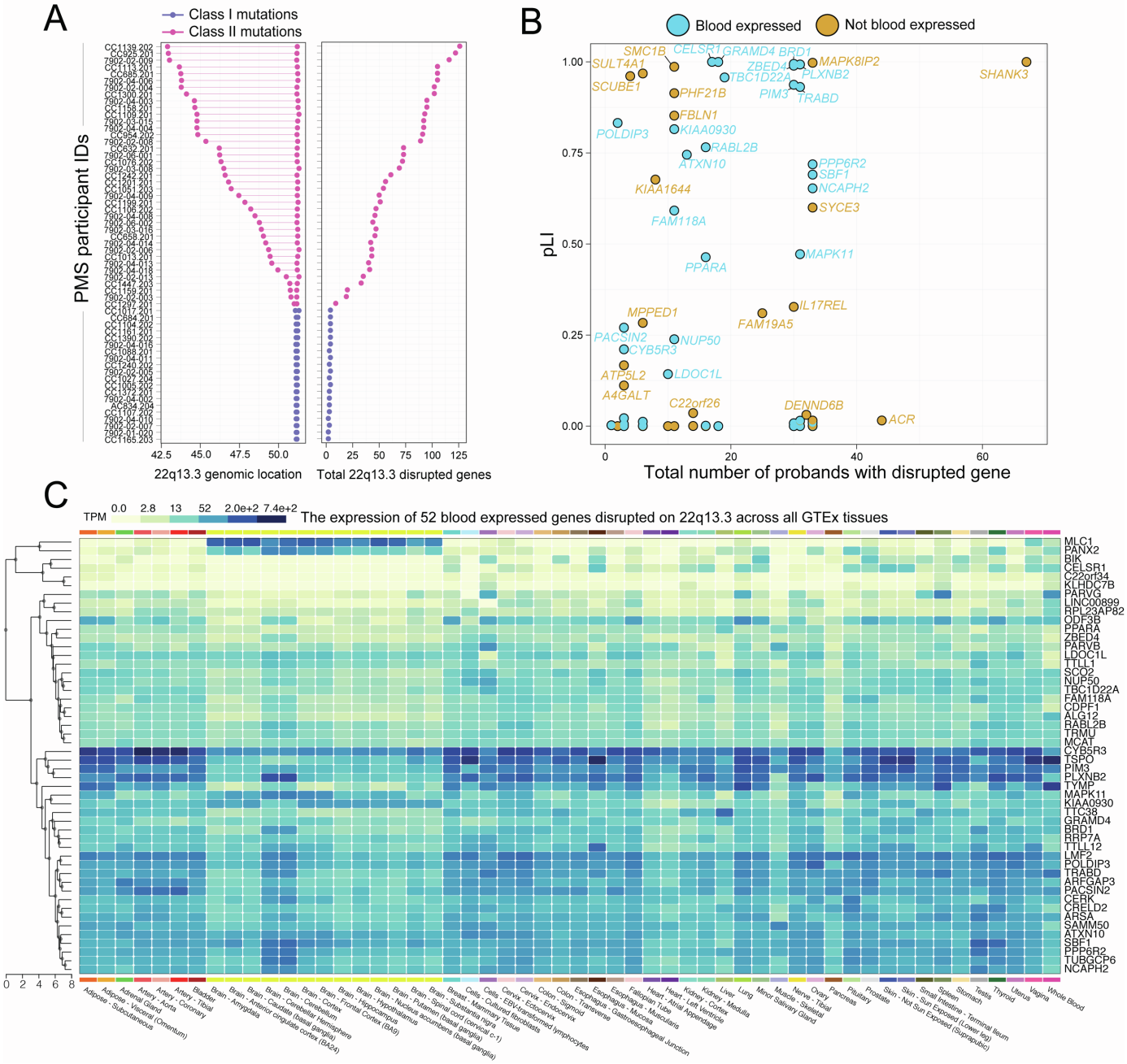

Figure S2. Characterization of 22q13.3 breakpoints and disrupted genes. (A) 22q13.3 deletion sizes (left; x-axis) and the total number of disrupted genes (right; x-axis) for each PMS participant (y-axis) designated as Class I mutations (without sequence variants) and Class II mutations. (**B**) Probability of loss of function intolerance (pLI) scores computed for each of the disrupted genes (y-axis) relative to the frequency of a gene to be disrupted across all PMS participants with transcriptome data (x-axis, max value=68 probands). pLI scores close to 1 indicate gene intolerance to heterozygous and homozygous loss of function. A total of 16 blood expressed displayed pLI > 0.5 and 8 genes had a pLI > 0.9. (**C**) We identified 52 blood expressed genes spanning the single largest 22q13.3 deletion in the current study. The average expression values (TPM) of these 52 genes were plotted across 30 distinct tissues from the Geneotype-Tissue Expression (GTEx) consortium, and in some instances, covering multiple regions per tissue. White indicates low expression and dark blue indicates high expression. This plot was generated using the multi-gene query function in the GTEx browser.

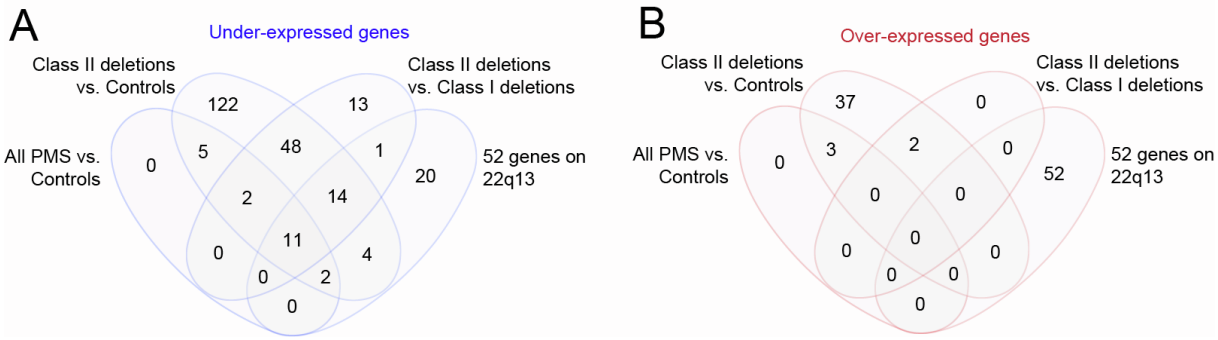

Figure S3. Overlap of differentially expressed genes at FDR < 5%. The overlap of (**A**) under-expressed and (**B**) over-expressed genes for i) all PMS participants relative to controls, ii) Class II mutations relative to controls, and iii) Class II mutations relative to Class I mutations. We also examined the overlap of the 52 peripheral blood expressed genes on 22q13 encompassed within large Class II mutations in the current study.

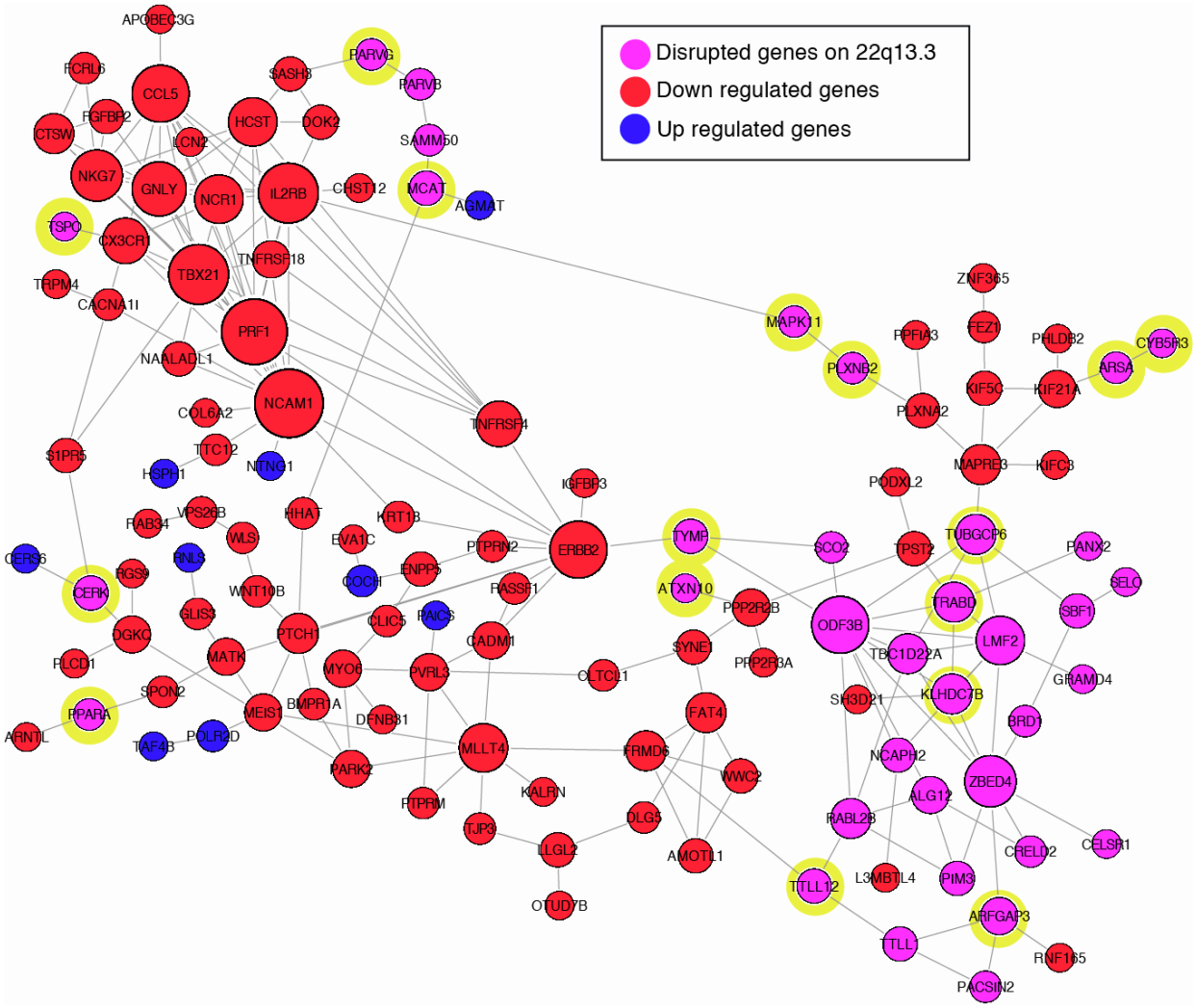

Figure S4. Direct protein-protein interaction (PPI) network. All 52 genes on 22q13.3 and differentially expressed genes (FDR < 5%) associated with PMS participants with Class II mutations were tested for enrichment of direct PPIs. The network contained significantly higher connectivity than expected by chance (p <1.0e-16). Nodes are colored by under-expressed genes (red), over-expressed genes (blue), and disrupted genes on 22q13.3 (pink). Yellow background is given to genes on 22q13.3 found to interact with differentially expressed genes.

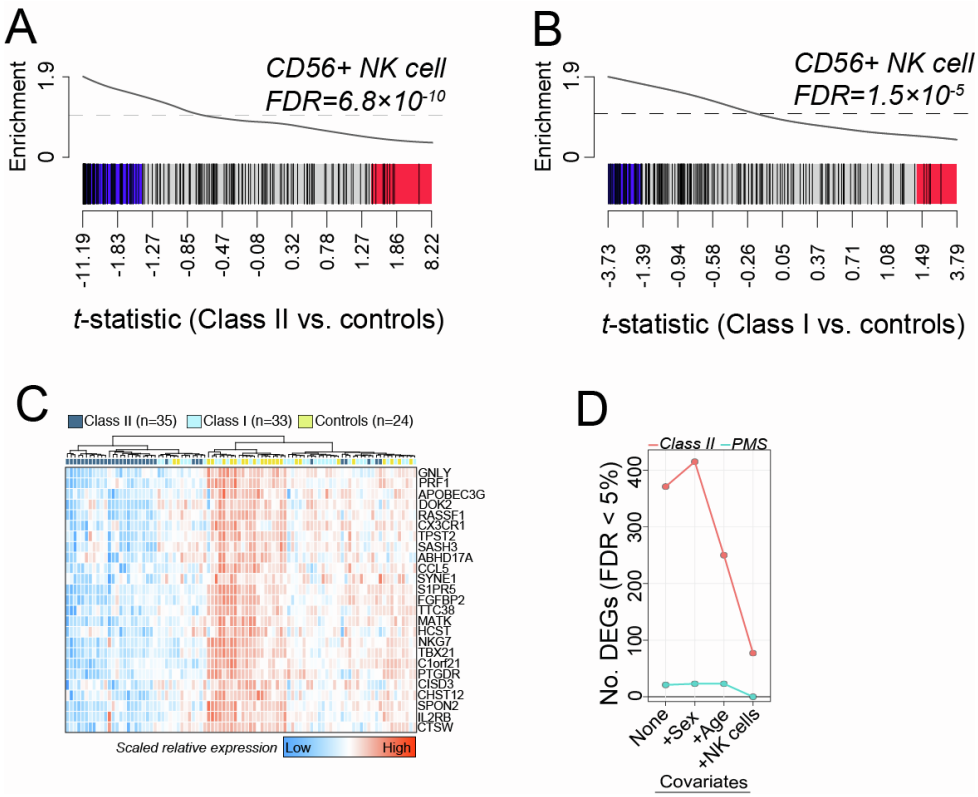

Figure S5. CD56+ NK cell enrichment gene set enrichment. CAMERA gene-set enrichment results for differentially expressed genes associated with (**A**) Class II mutations and (**B**) Class I mutations. Enrichment was tested for 190 genes that are differentially expressed CD56+ NK cells compared to all other cell types in the scRNA-seq experiment. (**C**) Unsupervised clustering of 25 CD56+ NK cell-specific genes distinguishes 82% (n=29) of Class II mutations from remaining samples. (**D**) The total number of significant differentially expressed genes in participants with Class II mutations (FDR < 5%) after adjusting for different covariates, reveals adjusting for CD56+ NK cell frequencies results in loss of ~69% of Class II-related DEGs.

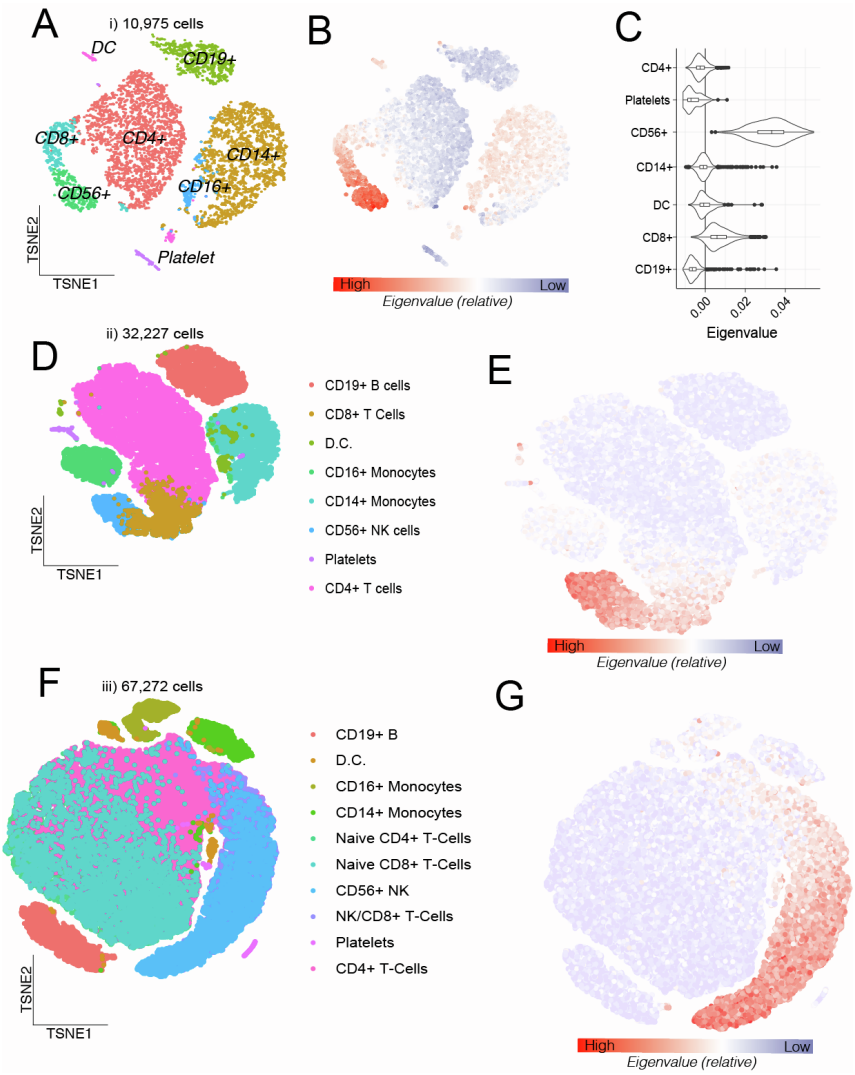

Figure S6. CD56+ NK cell-specific expression via scRNA-seq. TSNE clustering and cell type identification of eight main immune cell types across three independent studies: (**A**) the first dataset comprised of 10,975 PBMCs (v2 Chemistry); (**D**) the second dataset comprised of 33,227 PBMCs (v2 Chemistry), both were downloaded from the list of publically available 10X Genomic Inc. datasets; (**F**) third data set was comprised of 67,272 PBMCs and was obtained from Zheng et al., 2017[28]. Next, the normalized and scaled scRNA-seq expression data was used to create an eigenvalue (per cell) of 208 significantly under-expressed genes in participants with Class II mutations, which was projected onto each TSNE and color coded to illustrate high expression of these genes in CD56+ NK cells (blue=low, red=high) (**B, E, G**, respectively).(**C**) For clarity, eigenvalues (x-axis) were plotted as a violin plot for each cell type (y-axis) to illustrate strength of enrichment (merging CD14+ and CD16+ monocytes).

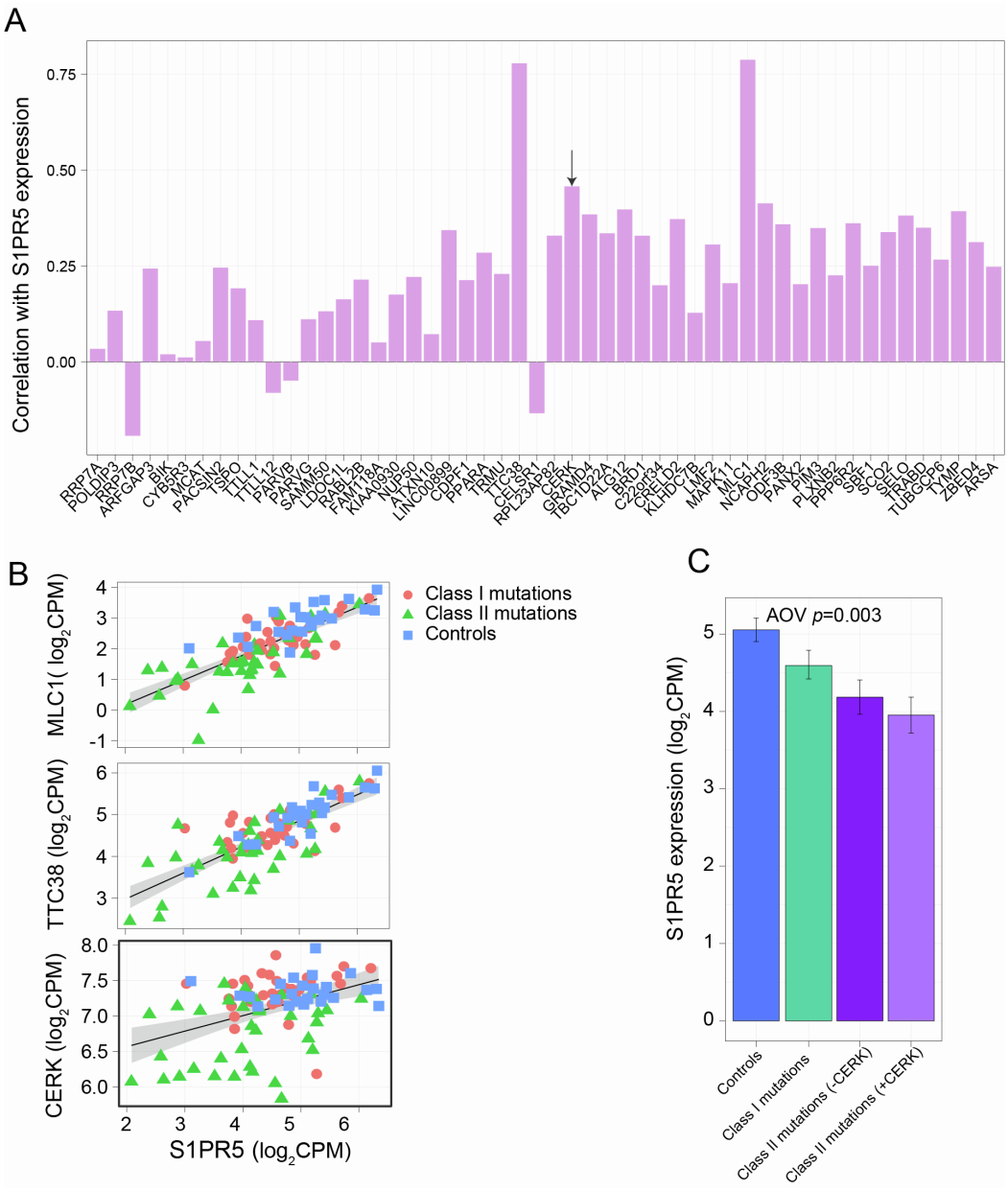

Figure S7. Gene expression on 22q13.3 that predicts *S1PR5* expression. (**A**) Barplots depicting the Pearson's correlation coefficient (y-axis) between gene expression of the 52 blood expressed genes on 22q13.3 number relative to *S1PR5* expression. *CERK* is denoted with an arrow. (**B**) The top three genes on 22q13.3 with the highest associations (y-axis) with *S1PR5* expression (x-axis) are depicted. (**C**) We anticipated that by parsing PMS participants with Class II mutations spanning *MLC1*, *TTC38*, and *CERK*, respectively, that those individuals would display lower expression of *S1PR5* relative to the remaining of individuals with Class II mutations. We found that only participants with Class II mutations spanning *CERK* were predictive of *S1PR5* expression, in that reduced expression of this gene was evident when compared with the remaining Class II mutations. An analysis of variance (AOV) was used to test for significance.

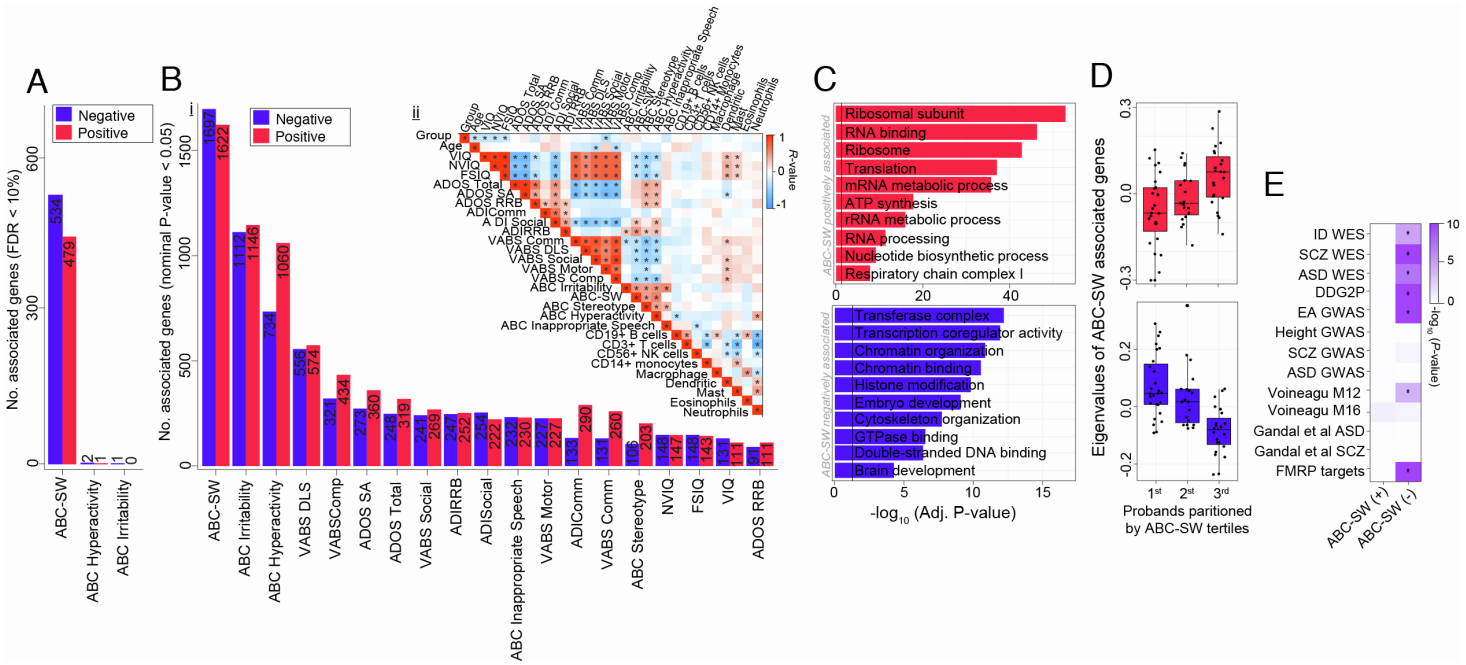

Figure S8. Exploratory analysis of phenotype-transcriptome associations. Barplots depicting the total number of genes positively (red) and negatively (blue) associated with each clinical measure presented in **Table 1** according to (**A**) a FDR < 10% and (**Bi**) a nominal *p*-value < 0.05. (**Bii**) Pearson's correlation matrix among all clinical traits in the current study (red=high; blue=low; *=significant association). (**C**) Functional annotation of genes positively and negatively associated with ABC-lethargy (social withdrawal). (**D**) To conceptualize these associations, all positively and negatively associated genes were summarized into one singular value using singular value decomposition, respectively. Probands were partitioned into tertiles according to ABC-lethargy scores and the resulting eigenvalues were plotted across low (1st tertile) to high (3rd tertile) scores confirming significant positive and negative associations. (**E**) Gene set enrichment analysis shows a significant enrichment of disease risk genes for intellectual disability (ID), schizophrenia (SCZ), autism spectrum disorder (ASD) and educational attainment (EA) among genes negatively associated with ABC-lethargy. Significance was calculated using a Fisher's exact test relative to a genome background of genes expressed in the current study.

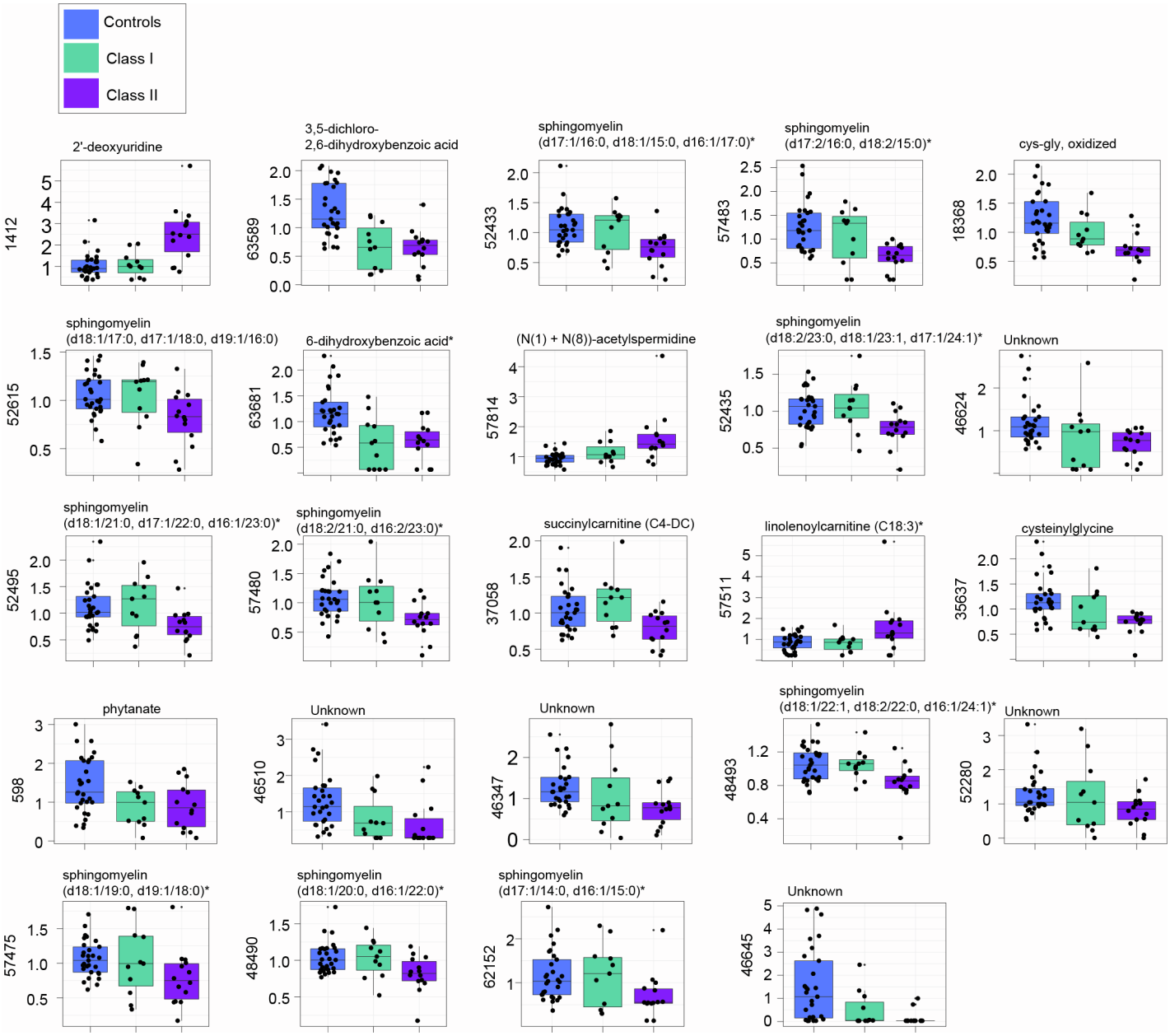

Figure S9. Metabolites associated with Class II mutations. Twenty-four differentially abundant metabolites significantly associated with Class II mutations relative to controls (FDR < 10%) are displayed. Scaled metabolite abundance (y-axes) was partitioned by deletion group (x-axes). The y-axis labels indicate compound identifiers and the main titles indicate the biochemical identifiers.