RESEARCH

Phenome-wide profiling identifies genotype-phenotype associations in Phelan-McDermid syndrome using family-sourced data from an international registry

Rui Yin[1,4†], Maxime Wack[1†], Claire Hassen-Khodja[1], Michael T. McDuffie[1], Geraldine Bliss[2], Elizabeth J. Horn[2], Cartik Kothari[1], Brittany McLarney[2], Rebecca Davis[2], Kristen Hanson[2], Megan O'Boyle[2], Catalina Betancur[3*†] and Paul Avillach[1*†]

Abstract

Background Phelan-McDermid syndrome (PMS) is a rare neurodevelopmental disorder caused by 22q13 deletions that include the *SHANK3* gene or pathogenic sequence variants in *SHANK3*. It is characterized by global developmental delay, intellectual disability, speech impairment, autism spectrum disorder, and hypotonia; other variable features include epilepsy, brain and renal malformations, and mild dysmorphic features. Here, we conducted genotype-phenotype correlation analyses using the PMS International Registry, a family-driven registry that compiles clinical data in the form of family-reported outcomes and family-sourced genetic test results.

Methods Data from the registry were harmonized and integrated into the i2b2/tranSMART clinical and genomics data warehouse. We gathered information from 401 individuals with 22q13 deletions including *SHANK3* (n = 350, ranging in size from 10 kb to 9.1 Mb) or pathogenic or likely pathogenic *SHANK3* sequence variants (n = 51), and used regression models with deletion size as a potential predictor of clinical outcomes for 328 phenotypes.

Results Our results showed that increased deletion size was significantly associated with delay in gross and fine motor acquisitions, a spectrum of conditions related to poor muscle tone, renal malformations, mild dysmorphic features (e.g., large fleshy hands, sacral dimple, dysplastic toenails, supernumerary teeth), lymphedema, congenital heart defects, and more frequent neuroimaging abnormalities and infections. These findings indicate that genes upstream of *SHANK3* also contribute to some of the manifestations of PMS in individuals with larger deletions. We also showed that self-help skills, verbal ability and a range of psychiatric diagnoses (e.g., autism, ADHD, anxiety disorder) were more common among individuals with smaller deletions and *SHANK3* variants.

---

†Rui Yin and Maxime Wack are co-first authors.

---

†Catalina Betancur and Paul Avillach have contributed equally to this work.

---

*Correspondence:
Catalina Betancur
catalina.betancur@inserm.fr
Paul Avillach
Paul_Avillach@hms.harvard.edu
Full list of author information is available at the end of the article

Yin *et al. Molecular Autism*   (2024) 15:40

Page 2 of 18

Limitations Some participants were tested with targeted 22q microarrays rather than genome-wide arrays, and karyotypes were unavailable in many cases, thus precluding the analysis of the effect of other copy number variants or chromosomal rearrangements on the phenotype.

Conclusions This is the largest reported case series of individuals with PMS. Overall, we demonstrate the feasibility of using data from a family-sourced registry to conduct genotype-phenotype analyses in rare genetic disorders. We replicate and strengthen previous findings, and reveal novel associations between larger 22q13 deletions and congenital heart defects, neuroimaging abnormalities and recurrent infections.

Keywords Phelan-McDermid syndrome, *SHANK3*, 22q13.3 deletion, Phenome-wide association study

Background

Phelan-McDermid syndrome (PMS, OMIM 606232) is a rare genetic condition [1, 2], with 3216 affected individuals worldwide (1739 in the US) engaged with the Phelan-McDermid Syndrome Foundation (PMSF). Clinically, affected individuals present with global developmental delay, intellectual disability (ID), hypotonia, and severe speech impairments [3–7]. Autistic features and autism spectrum disorder (ASD) are common [6, 8], with 63% of cases meeting criteria for ASD [6]. Other frequent findings include seizures, gastrointestinal problems, minor dysmorphic features, gait abnormalities, structural brain abnormalities, and renal malformations [1–4, 6–9]. In addition, some individuals develop severe neuropsychiatric illness in adolescence or early adulthood, including bipolar disorder, catatonia, and regression of skills [10, 11].

PMS is caused by deletions of varying sizes in the distal long arm of chromosome 22 affecting the *SHANK3* (SH3 and multiple ankyrin repeat domains 3) gene, or by pathogenic sequence variants in *SHANK3* [1, 2]. *SHANK3* encodes a post-synaptic scaffolding protein that plays a critical role in the development and function of excitatory synapses [12]. Haploinsufficiency of *SHANK3* causes the core features of PMS [13–16]. Individuals with *SHANK3* variants or interstitial deletions that exclusively disrupt *SHANK3* exhibit developmental delay, moderate to profound ID, motor deficits, severely impaired speech, ASD, hypotonia, seizures, brain abnormalities and minor dysmorphic features [5, 6, 9, 15, 16]. They are also more likely to have psychiatric diagnoses such as bipolar disorder, catatonia, depression, psychotic disorders, and regression [6, 10, 11].

Most reported cases of PMS are caused by 22q13.3 deletions, which can be either terminal or interstitial, and can be associated with ring chromosomes, translocations, or inversions [1, 2, 17]. Deletions range in size from < 10 kb to > 9 Mb, and usually encompass many genes in addition to *SHANK3*. Previous studies have shown that the size of the deletion is positively associated with hypotonia, developmental delay,

speech deficits, dysmorphic features and a number of medical comorbidities, including renal abnormalities and lymphedema [4, 6–9, 18, 19]. However, most of the studies exploring genotype-phenotype relationships in PMS have used very small samples (reviewed in [6]), with only the two most recent ones analyzing larger cohorts (n = 170 and 210) [6, 9].

Interstitial 22q13 deletions of variable size upstream of *SHANK3* have been reported in a small number of individuals [20–24]. These individuals share features common to PMS, including developmental delay, hypotonia, language delay, behavioral problems, dysmorphic features, renal abnormalities, congenital heart defects, and abnormal neuroimaging. These findings suggest that haploinsufficiency of other genes in the 22q13 region besides *SHANK3* contributes to the phenotype of PMS patients with larger deletions. Thus far, only two genes have been identified as contributing to the PMS phenotype: *TCF20*, involved in a syndromic neurodevelopmental disorder with ID and ASD [25, 26], and *CELSR1*, implicated in lymphedema [27, 28].

To explore the contribution of genes other than *SHANK3* to the phenotype of individuals with larger deletions, we utilized the PMS International Registry (PMSIR), a family-driven registry founded by parents of individuals with PMS and launched in 2011 by the PMSF. The international registry includes 1623 of the 3216 patients registered with PMSF. The phenotypic information is annotated by the patients' families, who are also asked to submit copies of the genetic tests. We hypothesized that although retrospective and not recorded in the context of a clinical evaluation, family-reported data (including genetic test results) obtained from the PMSIR could be used to conduct genotype-phenotype correlation analyses. Our results uncovered phenotypes significantly associated with larger deletions, implicating additional genes besides *SHANK3* in determining the phenotypic outcome. These findings have implications not only for PMS but also for those interested in using patient or family-reported data to shed new light on other rare diseases.

Yin *et al. Molecular Autism*   (2024) 15:40

Page 3 of 18

Methods

PMS International Registry and PMS Data Network

Funded by the Patient Centered Outcome Research Institute (PCORI), the PMS Data Network (PMS_DN) project aims to advance knowledge about PMS and related conditions by integrating patient-reported outcomes, curated genetic test results, and knowledge extracted from clinical records in the centralized i2b2/tranSMART platform to facilitate patient-centered research [29]. Clinical records were not used in this study. The network promotes a culture of transparency, as well as authentic engagement and leadership of families in the network's vision, governance, and generation of research priorities. The PMS_DN uses data entered into the online PMSIR by the patients' families. Only families in the registry that provided specific consent to participate in the PMS_DN were included in this study.

The registry comprises 1300 patient-reported outcome items across three distinct questionnaires: a "clinical" questionnaire, based on diagnosed comorbidities, symptoms, tests, and treatments for the entire range of known pathologies and features associated with PMS; a "developmental" questionnaire, focusing on physical, motor, behavioral, cognitive, and social development; and an "adult" questionnaire aimed at patients aged 12 or older, regarding the evolution of manifestations after puberty. Only the first two questionnaires were used in this study. Families can retake the surveys any number of times, allowing for longitudinal analysis and ensuring the most complete and accurate information. The registry is available in four languages: English, Spanish, French and Italian.

In addition to phenotypic data, the registry also contains curated genetic information, including results from karyotypes, fluorescent in situ hybridization (FISH), chromosomal microarrays, and *SHANK3* sequencing. The scanned paper reports of genetic test results are uploaded by families onto the platform, and manually curated by a team of trained genetic counselors, who fill in structured fields documenting the genetic abnormalities. The data are then reviewed by an expert in PMS genetics, who also interprets the pathogenicity of the variants.

All phenotypic and genetic data from the patients who consented to participate in the PMS_DN were integrated into the i2b2/tranSMART [30–32] data warehouse. In addition, a PMS ontology was created to tag and organize the registry data elements in a controlled vocabulary. The aim was to enable efficient user-selection of various stratifications of registry patients. This required developing an automatic data cleaning and loading pipeline (using the R language) to provide easy integration of new data from the PMSIR into the PMS_DN over time with consistent variable coding. Variables representing clinical items in the survey were categorized as being either "historical" or "evolving". "Historical" variables are immutable (e.g., birth weight), whereas "evolving" variables change over time (e.g., current weight). When a family retakes a survey in the PMSIR, we considered a change in an "evolving" variable to be an update to the current value, and a change in a "historical" variable to be a correction of a previous mistake. All "evolving" data were stored for each time point, and the most recent entry for "historical" items was considered the most accurate (Additional file 1: Figure S1). This ensures that the most recent record has the most up-to-date "historical" information, while storing the "evolving" parameter histories.

We note that some of the families enrolled in the PMSIR have also participated in previous studies investigating genotype-phenotype relationships in PMS, in the US and in other countries [4–9, 14, 16–19].

Genetic criteria

Only deletions with coordinates determined by chromosomal microarrays were included in the study. When multiple arrays were available, we selected the one with the highest resolution for the analysis. Chromosomal coordinates were transformed to the most recent assembly, GRCh38/hg38, using the LiftOver tool from the UCSC Genome Browser. In three individuals with contiguous deletions that were separated by a gap of less than 5% of their total length, the deletions were merged into a single deletion to prevent underestimation of their effect. The present study focused exclusively on deletions affecting *SHANK3*; proximal copy number variants in chromosome 22q or in other chromosomes were not included in the analysis. *SHANK3* variants were interpreted according to the American College of Medical Genetics and Genomics guidelines [33]. Only variants classified as pathogenic or likely pathogenic were included in the analyses.

In this phenome-wide association study [34], we compared the extent of deleted genetic material at the q terminus of chromosome 22 with the occurrence of each phenotypic feature. To this end, we treated patients with *SHANK3* sequence variants as patients with terminal deletions extending from the starting position of *SHANK3* to the end of the chromosome. Similarly, interstitial deletions of *SHANK3* were extended to the end of the chromosome. Note that the two coding genes distal to *SHANK3* do not contribute to the PMS phenotype: *ACR,* encoding acrosin, a sperm protein, and *RABL2B,* encoding a small Rab GTPase, are both tolerant to heterozygous loss-of-function variation in the general population (Genome Aggregation Database, gnomAD v2.1.1).

Yin *et al. Molecular Autism* (2024) 15:40

Page 4 of 18

Selection of phenotypic variables

In order to satisfy the technical requirements of the regression models, we selected for analysis physical and behavioral phenotypic features that were present in at least five individuals. To represent phenotypic outcomes, we preselected 454 out of the 1300 total family-reported outcome items available in the registry. The excluded items were those not directly referring to a symptom or a condition, such as "Does the patient have a primary cardiologist?", "How was the genetic test paid for?", and "Has the patient had any of the following tests: Manometry testing?". Questions referring to treatments were also excluded. Overall, 328 phenotypes satisfied our inclusion criteria.

Statistical analyses

We built multivariate models that associated phenotypic outcome with deletion size, adjusting for age and gender. When the outcome was a continuous variable (birth length, weight, head circumference, or Apgar score), we used linear regression to analyze the associations. To investigate the association between the phenotypes and the deletion size, we employed logistic regression, a method that has previously been used for deletion analyses in PMS [7, 8, 18]. For ordinal outcomes (severity of phenotype based on a four-point scale from "absent" to "always present," and age ranges for the acquisition of developmental milestones), we used proportional odds logistic regression [35]. We used the Benjamini-Hochberg procedure for determining the false discovery rate (FDR) globally to adjust the *p* values for multiple hypothesis testing [36]. Age was analyzed using the nonparametric Mann-Whitney test, while gender and race were analyzed using the Fisher test. Statistical analyses were conducted using R version 4.0.0. For the proportional odds logistic regression, the "clm" function from the "ordinal" package was used.

Ethics

The project was approved by the Harvard Medical School Institutional Review Board (HMS IRB14-2161). The PMSIR obtained continuing review approval from the Advarra (formerly Chesapeake) IRB as of January 21, 2021 (CR00182785). All the data were treated anonymously, from the integration into i2b2/transmart to the analysis.

Results

Study population

As of June 2020, 1132 families (661 in the US) consented to participate in the PMS_DN project. Of these, 754 patients had at least their demographic information entered into the registry, including 749 with phenotypic information from the clinical questionnaire (718 patients) and/or the developmental questionnaire (497 patients), and 491 individuals with genetic test results. We included in the analyses 401 patients with phenotypic information who had either 22q13 deletions with genomic coordinates defined by chromosomal microarray and encompassing *SHANK3* (n=350) or pathogenic or likely pathogenic sequence variants in *SHANK3* (n=51). As expected, there were no differences in the gender distribution of the participants (46.9% males, Table 1). Analysis of the demographic information (Table 1) revealed that the selected population was slightly younger and less ethnically diverse than the larger PMS_DN cohort. The majority of the individuals studied were White, which is likely a consequence of the socio-economic factors involved in the referral of individuals with ID and ASD for genetic testing. The 401 patients included in the analyses were

Table 1 Sample characteristics and comparison with individuals excluded because of missing phenotypic or genetic data

	Included (N=401)				Not included (N=352)				*p* value[a]
	N	%	Median	IQR	N	%	Median	IQR	
Age (years)	401		12.1	10.2	352		12.8	11.5	1.48e−01
Gender									8.26e−01
Female	213	53.1			191	54.3			
Male	188	46.9			161	45.7			
Ethnicity	347				289				2.11e−02
Asian	9	2.6			5	1.7			
Black (African American)	3	0.9			6	2.1			
Caucasian (Latino/Hispanic)	66	19.0			44	15.2			
Caucasian (Not Latino/Hispanic)	215	62.0			162	56.1			
Other	54	15.6			72	24.9			

[a] Age was analyzed using the Mann-Whitney test, while gender and ethnicity were analyzed using the Fisher test. IQR, interquartile range

Yin *et al. Molecular Autism*   (2024) 15:40

Page 5 of 18

distributed across countries as follows: United States 240, Spain 34, Australia 22, Canada 20, United Kingdom 19, France 15, Brazil 14, Italy 12, Belgium 3, Greece 3, Ireland 3, Chile 2, Russia 2, and one patient each from Colombia, Costa Rica, Israel, Mexico, Netherlands, New Zealand, Norway, Portugal, Romania, South Africa, Sweden, and United Arab Emirates. The completion rates of items in the clinical and developmental questionnaires ranged from 2% to 95%, with a mean completion rate of 69%. A total of 328 phenotypes, each present in at least five individuals, were included in the analyses.

Genetic findings

Figure 1 shows the 22q13.2-q13.33 region deleted in PMS and the genetic variants in 401 individuals included in the genotype-phenotype analysis. The variant types are summarized in Table 2. Deletion sizes ranged from 10 kb to 9.1 Mb, with a median deletion size of 3.7 Mb. Most deletions were terminal (97.1%), and were associated with translocations in 12.1% and with ring chromosome 22 in at least 8.5% of cases. The proportion of individuals with ring chromosome 22 is likely underestimated, given that the majority of individuals with deletions (57.1%) were diagnosed with chromosomal microarray and had not had a karyotype performed to exclude the presence of a ring chromosome. Thirteen individuals (3.7%) had mosaic deletions. *SHANK3* sequence variants included frameshift, nonsense, splice-site and missense variants; the majority were truncating variants (Table 2). Among the individuals for whom parental testing was available, all deletions and sequence variants were confirmed to have occurred de novo.

Genotype-phenotype analyses

We tested 328 phenotypes in 401 patients in relation to deletion size, using the multivariate models described in the Methods section. We found that 130 phenotypes were significantly associated with larger deletions (adjusted p value < 0.05), suggesting that additional genes on chromosome 22q contribute to these clinical manifestations, independently or in association with *SHANK3*. Tables 3 and 4 show the phenotypes positively and negatively associated with deletion size, respectively. For complete analyses and figures of all the phenotypes, see Additional file 2 (Table S1) and Additional file 3 (Other supplementary figures).

Phenotypes positively associated with deletion size

Major delays in nearly all gross motor milestones were strongly and linearly associated with larger deletion sizes (Fig. 2). These included Walks unassisted, Crawls on hands and knees, Sits when placed, Rolls over back to stomach, Climbs stairs standing up without help,

Descends stairs without help, and Holds head up on his/her own (adjusted p values ranged from 9.27e−18 to 4.01e−5, ORs ranged from 2.13 to 5.31; p values and ORs for each phenotype are listed in Table 3). Large deletions were also associated with delays in fine motor skills (Deliberately releases object to a container; Transfers object between hands; Looks at, reaches and grasps distant objects; p values from 1.74e−3 to 4.87e−3, ORs from 1.58 to 1.66), coordination deficits (Difficulties maintaining balance and Coordination difficulties, p values 1.93e−9 and 1.35e−3, ORs 2.25 and 1.56, respectively), and difficulties with motor planning (p = 3.35e−2, OR 1.37). These delays and deficits were accompanied by a range of conditions related to low muscular tone, including floppy baby, neonatal hypotonia, neonatal horeflexia, neonatal feeding problems, fatigue, 'weak muscles', and poor reflexes (p values from 4.68e−10 to 1.46e−2, ORs from 1.17 to 1.39). Specific feeding difficulties in babies observed more frequently in individuals with larger deletions included Required special feeds, Swallowing problems, Difficulty latching onto the breast, Difficulty latching onto the bottle, Poor suck, and Failure to gain weight normally (p values from 3.51e−6 to 2.50e−2, ORs from 1.14 to 1.29), as well as the need for nasogastric tube (p = 1.59e−2, OR 1.33). Moreover, breathing difficulties at birth, requiring neonatal intensive care, were more common among carriers of larger deletions.

Of note, all renal conditions in the clinical questionnaire, including Vesicoureteral reflux, Dysplastic kidney, Polycystic kidney (defined in the questionnaire as "multiple cysts present in one or both kidneys", which may include multicystic dysplastic kidney), Hydronephrosis, Recurrent urinary tract infections, and Increased kidney size, were strongly associated with deletion size (p values from 2.32e−7 to 1.18e−2, ORs from 1.30 to 1.67). The results indicate that patients with renal malformations carried larger deletions (Fig. 3). Persistent fever related to recurrent urinary tract infections was also associated with deletion size (p = 7.60e−3, OR 1.53). In addition, primary lymphedema was reported more frequently in individuals with large deletions (p = 4.55e−3, OR 1.35) (Fig. 4). Individuals with large deletions also exhibited swelling of the extremities, likely reflecting the presence of lymphedema (Persistent swelling of the hands and feet and Persistent swelling of the legs; p values 2.94e−6 and 1.05e−4, ORs 1.40 and 1.39, respectively).

In terms of cardiac abnormalities, the presence of a heart murmur (p = 3.43e−5, OR 1.32) and congenital heart defects (p = 4.91e−3, OR 1.25) were associated with larger deletions, including patent ductus arteriosus (n = 15), ventricular septal defect (n = 10), 'hole in the heart' (n = 9), bicuspid aortic valve (n = 9), persistent left superior vena cava (n = 5), total anomalous pulmonary
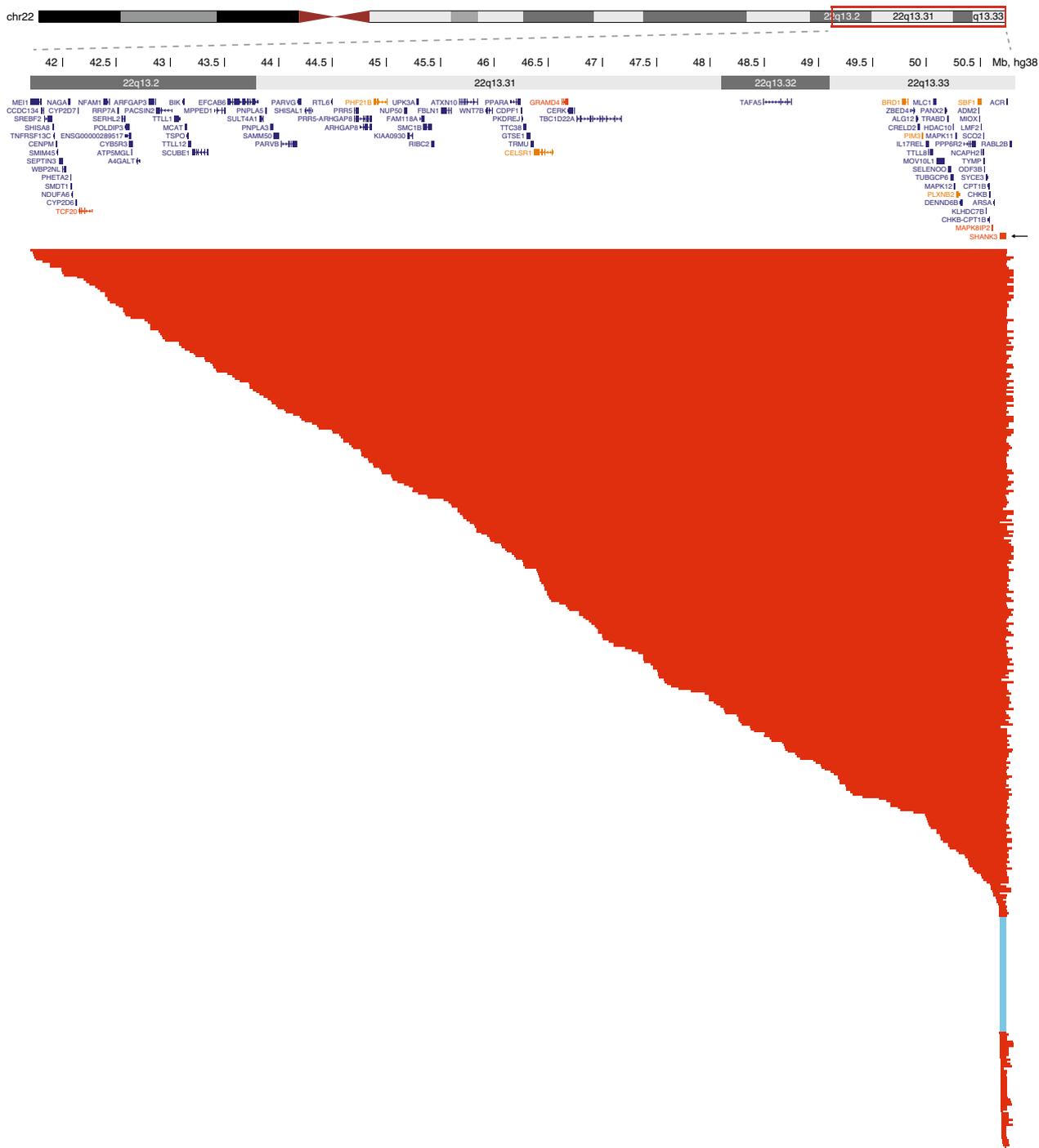
Fig. 1 Genetic variants affecting *SHANK3* in 401 PMS patients included in the study. Each line represents a patient; deletions are shown in red and *SHANK3* sequence variants in blue. For simplicity, sequence variants are represented as overlapping the whole gene. *SHANK3* is indicated by an arrow. Chromosomal coordinates are based on the GRCh38 genome build. Constrained genes intolerant to loss-of-function variants as measured by the LOEUF (loss-of-function observed/expected upper bound fraction) metric from gnomAD (v2.1.1) are indicated in orange (LOEUF < 0.2 darker orange, < 0.3 lighter orange; smaller LOEUF indicates higher constraint)

Table 2 *SHANK3* genetic variation in 401 individuals with Phelan-McDermid syndrome

Genetic variant	N (%)	Mosaic
22q13 deletion including *SHANK3*	350 (87.3%)	
Terminal deletion	340 (97.1%)	11
Simple deletion	70 (20.6%)	1
Unbalanced translocation	41 (12.1%)	1
Ring chromosome 22	29 (8.5%)	2
Other rearrangements[a]	6 (1.8%)	1
Unknown[b]	194 (57.1%)	6
Interstitial deletion	10 (2.9%)	2
SHANK3 sequence variant	51 (12.7%)	
Frameshift	40 (78.4%)	
Nonsense	5 (9.8%)	
Splice site	2 (3.9%)	
Missense	4 (7.8%)	

[a] Other rearrangements in chromosome 22 included 2 satellited chromosomes, 1 pseudodicentric chromosome, 1 unbalanced inversion, and 2 uncharacterized rearrangements

[b] Patients with terminal deletions identified by chromosomal microarray but without a karyotype to exclude the presence of a ring chromosome were classified as unknown

venous return or connection (n=4), 'heart valve problems' (n=3), coarctation of the aorta (n=2), aortic stenosis (n=2), enlarged aorta (n=2), and enlarged aortic root (n=1). Other congenital heart defects in the questionnaire (mitral valve prolapse) or cardiomyopathy were not reported in the study population.

Several dysmorphic features reported in PMS were associated with greater deletion sizes, including large fleshy hands ($p=9.47e-16$, OR 1.50), sacral dimple ($p=5.05e-12$, OR 1.45), dysplastic toenails ($p=1.56e-6$, OR 1.24) or fingernails ($p=2.47e-3$, OR 1.17), defective tooth enamel ($p=4.87e-3$, OR 1.20), supernumerary teeth ($p=5.29e-3$, OR 1.33), overgrowth (Too tall for age; $p=1.16e-3$, OR 1.20), bushy eyebrows ($p=8.69e-3$, OR 1.15), 2–3 toe syndactyly ($p=2.36e-2$, OR 1.15), and diastema (Gap between two front teeth; $p=9.83e-3$, OR 1.13). Recurring ingrown toenails were also more common in those with larger deletions ($p=8.65e-6$, OR 1.29), probably related to toenail dysplasia. Teeth extractions were also more frequent ($p=8.94e-3$, OR 1.18), possibly because of crowding and/or supernumerary teeth.

Abnormal brain magnetic resonance imaging (MRI) ($p=6.72e-5$, OR 1.23) and computed tomography (CT) scans ($p=5.83e-4$, OR 1.38), as well as the presence of specific abnormalities such as irregular brain ventricles, decreased gray matter, decreased myelination, and arachnoid cysts, were more likely to be found in patients with larger deletions (p values from $7.36e-4$ to $2.40e-2$, ORs from 1.20 to 1.32). Individuals with larger

deletions also had more current seizures ($p=2.07e-2$, OR 1.13) and more febrile seizures ($p=4.51e-6$, OR 1.30). However, other types of seizures (e.g., absence seizures, tonic-clonic seizures, simple and complex partial seizures) were not associated with deletion size. Significant associations were observed for the need for intravenous antibiotics, chronic bronchitis, >2 pneumonias in a year, >2 serious sinus infections in a year, recurrent pneumonia, and placement of ear tubes (p values from $2.86e-4$ to $3.98e-2$, ORs from 1.16 to 1.31), suggesting that recurrent infections may be more frequent among individuals with larger deletions. Except for a mild association between deletion size and chronic aspiration ($p=1.90e-4$, OR 1.26), no gastrointestinal conditions or symptoms were found to be associated with deletion size. The only ocular issue significantly associated with deletion size was strabismus ($p=1.16e-3$, OR 1.19).

Phenotypes negatively associated with deletion size
Self-help skills had some of the strongest negative associations with deletion size, including Dresses without assistance, Undresses without assistance, Toilets independently, Night-time toilet trained, Eats independently, and Drinks independently (p values from $2.43e-9$ to $2.75e-4$, ORs from 0.327 to 0.558; p values and ORs for each phenotype are listed in Table 4). In addition, abilities related to socialization and imagination were more common among individuals with smaller deletions and *SHANK3* variants: Awareness of imaginary characters, Engages in pretend play, Plays alongside others alone, Responds to others' emotions, Responds affectionately to caregivers, Hugs/kisses caregivers or dolls, Plays peek-a-boo, and Plays with peers (p values from $1.65e-5$ to $3.09e-2$, ORs from 0.327 to 0.750). Speech and language abilities (Number of words in a typical sentence, Verbal speech ability, and Ability to follow directions; p values from $9.41e-7$ to $4.50e-3$, ORs from 0.378 to 0.560) were observed more frequently in individuals with smaller deletions. Similarly, nonverbal communication skills, including Pointing and Gesturing, were used more often by individuals with smaller deletions (p values from $3.19e-3$ to $4.87e-3$, ORs from 0.855 to 0.860). Some speech disorders, such as Apraxia of speech, Receptive language disorder, and Expressive language disorder, were more frequently diagnosed in patients with smaller deletions (p values from $2.79e-4$ to $6.31e-4$, ORs from 0.759 to 0.807). Auditory processing disorder was also negatively associated with deletion size ($p=4.87e-3$, OR 0.693), whereas no significant association was observed for hearing loss.

A range of psychiatric diagnoses, including autism, pervasive developmental disorder, attention deficit-hyperactivity disorder (ADHD), anxiety disorder, obsessive-compulsive

Yin *et al. Molecular Autism*   (2024) 15:40

Page 8 of 18

Table 3 Phenotypes significantly and positively associated with deletion size[a]

Phenotype	*p* value	Adjusted *p* value[b]	OR [CI]
Walks unassisted	2.80e−20	9.27e−18	5.31 [3.76–7.64]
Crawls on hands and knees	1.55e−16	1.29e−14	4.23 [3.03–6.02]
Sits when placed	8.58e−18	9.47e−16	3.49 [2.64–4.68]
Climbs stairs standing up without help	1.28e−07	2.13e−06	2.69 [1.87–3.92]
Rolls over back to stomach	1.34e−12	7.38e−11	2.66 [2.04–3.51]
Descends stairs without help	4.24e−06	4.01e−05	2.54 [1.71–3.80]
Difficulties maintaining balance	4.65e−11	1.93e−09	2.25 [1.77–2.88]
Holds head up on his/her own	1.77e−09	5.30e−08	2.13 [1.67–2.74]
Vesicoureteral reflux	9.81e−09	2.32e−07	1.67 [1.42–2.03]
Deliberately releases object to a container	1.28e−03	4.87e−03	1.66 [1.22–2.26]
Transfers object between hands	4.81e−04	2.31e−03	1.62 [1.24–2.13]
Looks at, reaches and grasps distant objects	3.58e−04	1.74e−03	1.58 [1.23–2.04]
Dysplastic kidney	4.29e−05	2.79e−04	1.57 [1.29–1.99]
Coordination difficulties	2.74e−04	1.35e−03	1.56 [1.23–1.99]
Persistent fever related to kidney issues	2.18e−03	7.60e−03	1.53 [1.20–2.10]
Large fleshy hands	6.50e−18	9.47e−16	1.50 [1.37–1.65]
Sacral dimple	7.62e−14	5.05e−12	1.45 [1.32–1.60]
Persistent swelling of the hands and feet[c]	1.95e−07	2.94e−06	1.40 [1.24–1.60]
Floppy baby	1.27e−08	2.79e−07	1.39 [1.24–1.56]
Persistent swelling of the legs[c]	1.39e−05	1.05e−04	1.39 [1.21–1.64]
Polycystic kidney	2.52e−03	8.59e−03	1.39 [1.14–1.77]
Hydronephrosis	7.15e−06	6.07e−05	1.38 [1.21–1.60]
Abnormal cranial CT scan	9.69e−05	5.83e−04	1.38 [1.18–1.64]
Difficulties with motor planning	1.25e−02	3.35e−02	1.37 [1.07–1.76]
Neonatal hypotonia	9.90e−12	4.68e−10	1.35 [1.24–1.48]
Primary lymphedema	1.13e−03	4.55e−03	1.35 [1.14–1.64]
Recurrent urinary tract infections	7.80e−03	3.44e−03	1.34 [1.14–1.61]
Supernumerary teeth	1.42e−03	5.29e−03	1.33 [1.12–1.60]
Nasogastric tube (as a baby)	5.39e−03	1.59e−02	1.33 [1.10–1.65]
Heart murmur	3.53e−06	3.43e−05	1.32 [1.18–1.49]
Decreased gray matter	6.34e−04	2.87e−03	1.32 [1.13–1.57]
IV antibiotics needed	4.50e−05	2.86e−04	1.31 [1.15–1.49]
Febrile seizures	3.39e−07	4.51e−06	1.30 [1.18–1.45]
Increased kidney size	3.71e−03	1.18e−02	1.30 [1.10–1.56]
Neonatal feeding problems	1.92e−09	5.30e−08	1.29 [1.19–1.40]
Recurring ingrown toenails	7.32e−07	8.65e−06	1.29 [1.17–1.43]
Required special feeds as a baby	1.37e−05	1.05e−04	1.29 [1.16–1.46]
Irregular brain ventricles	1.29e−04	7.36e−04	1.29 [1.13–1.47]
Swallowing problems as a baby	7.24e−07	8.65e−06	1.27 [1.16–1.40]
Difficulty latching onto the breast	2.44e−07	3.51e−06	1.26 [1.16–1.38]
Aspiration	2.69e−05	1.90e−04	1.26 [1.13–1.40]
Congenital heart defects	3.63e−02	8.24e−02	1.17 [1.01–1.36]
Dysplastic or unusual toenails	8.68e−08	1.56e−06	1.24 [1.14–1.34]
Chronic bronchitis	9.70e−04	4.06e−03	1.24 [1.09–1.41]
Poor suck	5.34e−06	4.68e−05	1.23 [1.13–1.34]
Abnormal brain MRI	8.33e−06	6.72e−05	1.23 [1.12–1.35]
Decreased myelination	1.36e−03	5.11e−03	1.23 [1.09–1.40]
Apnea (obstructive)	2.36e−03	8.14e−03	1.23 [1.08–1.42]
Neonatal hyporeflexia	4.89e−03	1.46e−02	1.22 [1.07–1.41]

Table 3 (continued)

Phenotype	*p* value	Adjusted *p* value[b]	OR [CI]
2 + pneumonias in a year	4.27e−03	1.31e−02	1.22 [1.07–1.40]
Poor reflexes	2.09e−04	1.13e−03	1.21 [1.09–1.33]
Neonatal breathing difficulties (NICU)	4.17e−03	1.30e−02	1.21 [1.07–1.39]
Congenital torticollis	1.65e−02	4.26e−02	1.21 [1.04–1.43]
Difficulty latching onto the bottle	1.39e−04	7.69e−04	1.20 [1.10–1.33]
Too tall for age	2.21e−04	1.16e−03	1.20 [1.09–1.32]
Weak muscles	1.06e−03	4.38e−03	1.20 [1.08–1.34]
Weak/soft tooth enamel	1.28e−03	4.87e−03	1.20 [1.07–1.34]
Arachnoid cyst	8.56e−03	2.40e−02	1.20 [1.05–1.38]
Strabismus	2.18e−04	1.16e−03	1.19 [1.09–1.31]
2 + serious sinus infections in a year	2.87e−03	9.49e−03	1.18 [1.06–1.31]
Tooth/teeth extraction	2.67e−03	8.94e−03	1.18 [1.06–1.31]
Dysplastic or unusual fingernails	5.38e−04	2.47e−03	1.17 [1.07–1.28]
Fatigues easily	1.93e−03	6.95e−03	1.17 [1.06–1.29]
Torticollis	1.25e−02	3.35e−02	1.17 [1.04–1.33]
Recurrent pneumonia	1.51e−02	3.98e−02	1.17 [1.03–1.34]
Ear tubes	1.22e−03	7.06e−04	1.16 [1.08–1.25]
Scoliosis	1.25e−02	3.35e−02	1.16 [1.03–1.31]
Bushy eyebrows	2.57e−03	8.69e−03	1.15 [1.05–1.26]
Teeth grinding (day or night)	1.90e−03	6.92e−03	1.15 [1.05–1.25]
2–3 toe syndactyly	8.34e−03	2.36e−02	1.15 [1.04–1.28]
Failure to gain weight normally as a baby	9.06e−03	2.50e−02	1.14 [1.03–1.27]
Gap between two front teeth	3.00e−03	9.83e−03	1.13 [1.04–1.23]
Seizures (currently)	7.23e−03	2.07e−02	1.13 [1.03–1.23]
Daytime teeth grinding	4.66e−03	1.42e−02	1.12 [1.04–1.22]
Difficulty regulating body temperature	4.10e−03	1.29e−02	1.12 [1.04–1.21]

CI, confidence interval; CT, computerized tomography; IV, intravenous, MRI, magnetic resonance imaging; NICU, neonatal intensive care unit; OR, odds ratio

[a] For developmental phenotypes, positive associations indicate that larger deletions are associated with delayed acquisition

[b] Benjamini-Hochberg false discovery rate (FDR) adjusted *p* values

[c] These questions refer to individuals with lymphedema

disorder, and pica, were negatively associated with deletion size (*p* values from 4.51e−6 to 7.35e−3, ORs from 0.643 to 0.820). Attention deficit and hyperactivity symptoms were also less prevalent in individuals with larger deletions; specifically, Symptoms/diagnosis of ADD/ADHD, Acts as if driven by a motor, Runs around or climbs excessively when inappropriate, Appears to act without thinking, Difficulty playing quietly, Leaves seat when remaining seated is expected, and Difficulty sustaining attention in tasks (*p* values from 8.97e−8 to 4.06e−2, ORs from 0.486 to 0.721). Sleep disturbances (Difficulty falling asleep, Difficulty going back to sleep, Frequent nighttime awakenings, and Short nighttime sleep; *p* values from 9.40e−6 to 4.58e−2, ORs from 0.823 to 0.908) were all less frequent in individuals with larger deletions.

Interestingly, nearly all the cognitive development milestones surveyed showed significant impairment proportional to the size of the deletion. These included, Understands the use of familiar objects, Imitates household activities during play, Uses functional toys appropriately, Acts out recent experiences using gestures or words, Observes actions and imitates them later, and Understands the difference between food and objects (*p* values from 4.51e−6 to 4.87e−3, ORs from 0.786 to 0.864). Accordingly, regression in cognitive skills was observed less frequently in individuals with larger deletions (*p* value 1.99e−2, OR 0.862). Furthermore, Apgar scores at 1 and 5 min were significantly lower in individuals with larger deletions (*p* values 1.31e−2 and 1.35e−3, ORs 0.870 and 0.901, respectively), indicating an adverse impact on newborn health. Gestational age and birth weight were not associated with deletion size.

Finally, there were no associations with any endocrine (including hypothyroidism, too tall or too short for age, or puberty milestones), orthopedic (except scoliosis and torticollis), nose and throat, vision (except strabismus), or allergy-related conditions (Additional file 2: Table S1). Responses to sounds, movement, touch, oral input, or

Yin *et al. Molecular Autism*   (2024) 15:40

Page 10 of 18

Table 4 Phenotypes significantly and negatively associated with deletion size[a]

Phenotype	*p* value	Adjusted *p* value[b]	OR [CI]
Awareness of imaginary characters	1.63e−06	1.68e−05	0.327 [0.199–0.501]
Dresses without assistance	2.06e−08	4.27e−07	0.327 [0.217–0.474]
Toilets independently	8.93e−08	1.56e−06	0.353 [0.236–0.508]
Number of words in a typical sentence	1.10e−03	4.50e−03	0.378 [0.201–0.654]
Undresses without assistance	1.51e−10	5.00e−09	0.391 [0.290–0.517]
Eats independently	6.61e−11	2.43e−09	0.456 [0.359–0.575]
Night-time toilet trained	4.07e−05	2.75e−04	0.465 [0.317–0.661]
Acts as if driven by a motor	1.65e−07	2.60e−06	0.486 [0.369–0.634]
Verbal speech ability	4.83e−08	9.41e−07	0.489 [0.377–0.630]
Engages in pretend play	1.54e−06	1.65e−05	0.530 [0.407–0.684]
Plays alongside others, alone	2.58e−06	2.59e−05	0.553 [0.431–0.707]
Drinks independently	4.57e−06	4.21e−05	0.558 [0.434–0.715]
Ability to follow directions	5.37e−06	4.68e−05	0.560 [0.435–0.717]
Responds to others' emotions	1.59e−05	1.15e−04	0.572 [0.442–0.734]
Appears to act without thinking	2.88e−05	1.99e−04	0.575 [0.443–0.744]
Hugs/kisses caregivers, dolls	7.97e−06	6.60e−05	0.590 [0.467–0.742]
Responds affectionately to caregivers	5.28e−05	3.30e−04	0.623 [0.495–0.783]
Runs around or climbs excessively when inappropriate	1.66e−03	6.09e−03	0.635 [0.477–0.842]
Anxiety disorder	2.39e−04	1.24e−03	0.643 [0.491–0.792]
Obsessive compulsive disorder	2.09e−03	7.35e−03	0.658 [0.481–0.830]
ADHD	1.59e−05	1.15e−04	0.674 [0.552–0.793]
Difficulty playing quietly	3.53e−03	1.14e−02	0.681 [0.525–0.881]
Auditory processing disorder	1.28e−03	4.87e−03	0.693 [0.539–0.848]
Symptoms/diagnosis of ADD/ADHD	3.52e−09	8.97e−08	0.704 [0.624–0.787]
Leaves seat when remaining seated is expected	1.56e−02	4.06e−02	0.706 [0.531–0.935]
Symptoms/diagnosis of obsessive–compulsive disorder	1.98e−03	7.03e−03	0.720 [0.570–0.870]
Difficulty sustaining attention in tasks	8.79e−03	2.44e−02	0.721 [0.563–0.920]
Plays peek-a-boo	5.84e−03	1.71e−02	0.729 [0.581–0.912]
Response to sound	1.30e−02	3.44e−02	0.739 [0.580–0.937]
Plays with peers	1.13e−02	3.09e−02	0.750 [0.599–0.936]
Symptoms/diagnosis of generalized anxiety disorder	7.95e−04	3.46e−03	0.751 [0.627–0.879]
Receptive language disorder	7.76e−05	4.76e−04	0.759 [0.657–0.865]
Smiles at others	1.99e−02	5.00e−02	0.760 [0.602–0.957]
Apraxia of speech	4.30e−05	2.79e−04	0.767 [0.671–0.866]
Pica	8.26e−04	3.55e−03	0.773 [0.657–0.891]
Imitates household activities during play	3.48e−07	4.51e−06	0.786 [0.715–0.861]
Autism	3.54e−07	4.51e−06	0.794 [0.724–0.866]
Understands the use of familiar objects	9.98e−07	1.10e−05	0.798 [0.727–0.872]
Expressive language disorder	1.07e−04	6.31e−04	0.807 [0.721–0.897]
Uses functional toys appropriately	8.88e−06	7.00e−05	0.815 [0.743–0.890]
Pervasive developmental disorder	5.38e−04	2.47e−03	0.820 [0.729–0.913]
Difficulty falling asleep	8.24e−07	9.40e−06	0.823 [0.761–0.888]
Acts out recent experiences using gestures, etc	2.44e−04	1.24e−03	0.825 [0.742–0.912]
Exhibits sensory seeking behaviors	3.29e−03	1.07e−02	0.854 [0.767–0.947]
Pointing	7.12e−04	3.19e−03	0.855 [0.779–0.935]
Observes actions and imitates them later	5.23e−04	2.47e−03	0.859 [0.787–0.935]
Gesturing	1.22e−03	4.87e−03	0.860 [0.784–0.941]
Regression in cognitive abilities	6.55e−03	1.99e−02	0.862 [0.772–0.957]
Understands the difference between food and objects	1.25e−03	4.87e−03	0.864 [0.789–0.943]

Yin *et al. Molecular Autism*   (2024) 15:40

Page 11 of 18

Table 4 (continued)

Phenotype	*p* value	Adjusted *p* value[b]	OR [CI]
Enuresis	4.77e−03	1.44e−02	0.865 [0.779–0.954]
Difficulty going back to sleep	1.32e−04	7.43e−04	0.865 [0.803–0.931]
1 min Apgar score	4.28e−03	1.31e−02	0.870 [0.792–0.957]
Frequent nighttime awakenings	8.68e−04	3.68e−03	0.881 [0.817–0.949]
Sometimes aggressive behavior	1.91e−02	4.83e−02	0.897 [0.818–0.981]
5 min Apgar score	2.69e−04	1.35e−03	0.901 [0.852–0.952]
Plays with interactive games/toys	1.72e−02	4.41e−02	0.905 [0.832–0.982]
Short nighttime sleep	1.80e−02	4.58e−02	0.908 [0.838–0.983]

ADD, attention deficit disorder; ADHD, attention deficit-hyperactivity disorder; CI, confidence interval; OR, odds ratio

[a] For developmental phenotypes, negative associations indicate that acquisition of the skill was observed less frequently among individuals with larger deletions

[b] Benjamini-Hochberg false discovery rate (FDR) adjusted *p* values

temperature variations were not associated with deletion size. Similarly, tolerance to pain was not associated with deletion size.

Genotype-phenotype associations in individuals with sequence variants or Class I or Class II deletions

Individuals with PMS can be classified as having *SHANK3* sequence variants; small Class I deletions, including *SHANK3* only or *SHANK3* with *ARSA* and/or *ACR* and *RABL2B*; or larger Class II deletions, which include all other deletions [6]. The *ARSA*, *ACR* and *RABL2B* genes are not sensitive to haploinsufficiency and are therefore not expected to contribute to the phenotype of PMS. Comparison of individuals with sequence variants or Class I deletions versus those with Class II deletions, shown in Table 5, confirmed our findings based on the analysis of deletion size. Kidney abnormalities, lymphedema, congenital heart disease, abnormal brain imaging, hypotonia, feeding difficulties, and certain dysmorphic features were more common in individuals with Class II deletions than in those with sequence variants or Class I deletions. Individuals with sequence variants and Class I deletions were more likely to have verbal speech, self-help skills, and psychiatric diagnoses, including ASD, ADHD, depression, bipolar disorder and anxiety. Due to the nature of the parent-reported data used in this study, the frequencies in Table 5 are not directly comparable to those reported in studies based on clinical assessment or direct review of medical records [6]. In particular, the total number of individuals considered for each characteristic listed in the table includes the cases with a "No" response, but this does not necessarily mean that the clinical feature was analyzed in all individuals. There were also many unanswered questions, raising the possibility that some parents may have prioritized questions about their children's problems. For these reasons, we opted not to perform statistical analyses on these data.

Discussion

We demonstrated the feasibility of using registry data based on family-reported outcome questionnaires to conduct a multiphenotype-genotype association study. We further replicated findings from previous association studies, confirmed recent findings regarding renal malformations, lymphedema, and gross motor development, and suggested novel associations with congenital heart defects, neuroimaging abnormalities and recurrent infections. This is the largest study of its kind in PMS and the first to use the PMSIR. Access to the PMSIR allowed us to recruit the largest sample size (401 patients) for a genotype-phenotype association analysis in PMS, with deletion breakpoints identified by microarray or *SHANK3* sequence variants. The sample sizes of previous studies of individuals with PMS were considerably smaller: 29 [19, 37], 30 [38], 32 [4], 51 [14], 67 [39], 70 [7], 71 [18], 98 [8], 170 [6], and 210 [9]. Most previous studies focused on the analysis of deletions, with only the two most recent and largest studies including a significant proportion of *SHANK3* variants, between 10 and 20% [6, 9], compared to 12.7% in the current study.

The data provided by this family-driven registry differ from those obtained from medical records. Patient-driven data are rich in information that only caregivers and families can provide, and because the information is provided by the parents themselves, as a proxy for the patients, this approach allows for greater volume and specificity of registry items in many domains. Such a repository would be very difficult and costly to amass through a conventional academic institution-led prospective study. In addition, individuals with PMS can exhibit challenging or hyperactive behaviors so it may be difficult for parents to complete extensive surveys during interviews with their physicians. Furthermore, the information is subject to recall bias. At home, parents have more time and flexibility to complete the
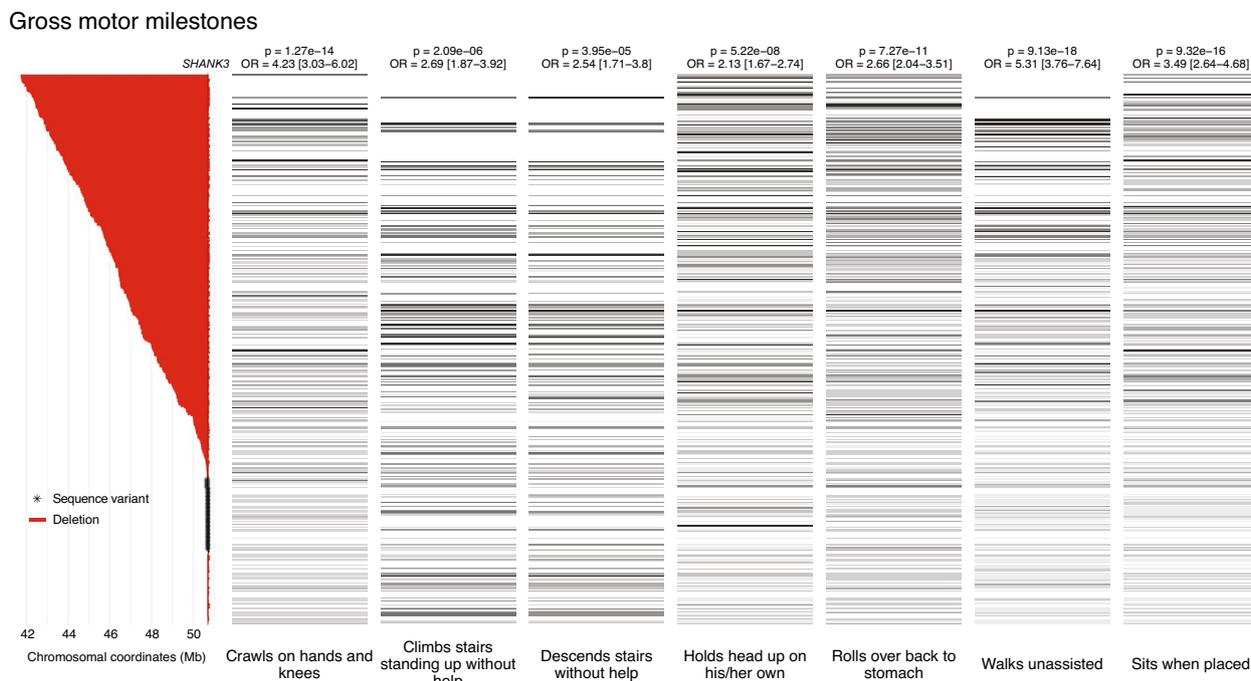
Yin *et al. Molecular Autism*   (2024) 15:40

Page 12 of 18

Fig. 2 Association between deletion size and age of acquisition of gross motor milestones. On the left: deletions (in red) and *SHANK3* variants (asterisks), ordered by decreasing deletion size/position. The position of *SHANK3* is indicated by a vertical line. On the right: status of each patient for developmental gross motor phenotypes, lined up with the respective genetic status. The shades of gray and black represent the age range at which the patients achieved gross motor milestones, with lighter shades indicating younger ages and darker shades indicating older ages. The color white indicates that the answer was either 'not applicable' or 'unsure'

surveys; find it easier to make corrections to the data entered, making it more accurate; and are more likely to answer items more thoroughly and to update the surveys, as they are able to refer to the child's medical records when answering questions. We assert that the reliability of the registry information is demonstrated by the consistency of the results throughout different analyses and with previous findings. Overall, the results based on a larger sample size demonstrate the potential of analyses using data from the PMSIR to pinpoint specific phenotype associations in this syndrome.

Larger deletion sizes are consistently and proportionally associated with more severe gross motor development delays, as well as phenotypes related to low muscle tone in the early stages of development and feeding difficulties. Some of these findings have been reported previously [6–8, 18]. However, previous studies only reported "age at walking" [6–8] or "age at walking and crawling" [18], and therefore did not capture positive associations in a consistent manner for each gross motor milestone. Some of these features have also been reported in individuals with interstitial deletions not affecting *SHANK3* [22–24] and are also present to a lesser extent and severity in PMS patients with *SHANK3* sequence variants [6, 16]. The fact that these

tone-related conditions were observed even in individuals with the smallest deletions, and that the severity of the features increased in proportion to the deletion size, suggests a cumulative role of multiple genes on the long arm of chromosome 22, including *SHANK3.*

Renal malformations, reported in 15–40% of PMS patients with deletions [4, 6, 9], are strongly associated with larger deletion sizes, with the distribution of deletion sizes hinting at a potentially causative gene (or genes) on chromosome 22q. This association has been reported in four previous studies, which analyzed all renal and urogenital abnormalities, regardless of the specific condition, in smaller samples [4, 6, 9, 19]. This finding indicates that renal malformations in PMS, which are absent (or extremely rare) in patients with *SHANK3* sequence variants [6, 9, 16] but are found in some individuals with interstitial deletions not involving *SHANK3* [22–24], are related to genes other than *SHANK3* and have incomplete penetrance.

Our results also confirm previous studies showing an association of deletion size with large fleshy hands, dysplastic toenails and sacral dimple [7, 8, 18, 38], and an increased prevalence of lymphedema and swelling of the extremities [4, 9, 19]. Larger deletions were significantly associated with heart conditions, including heart

Yin *et al. Molecular Autism*   (2024) 15:40
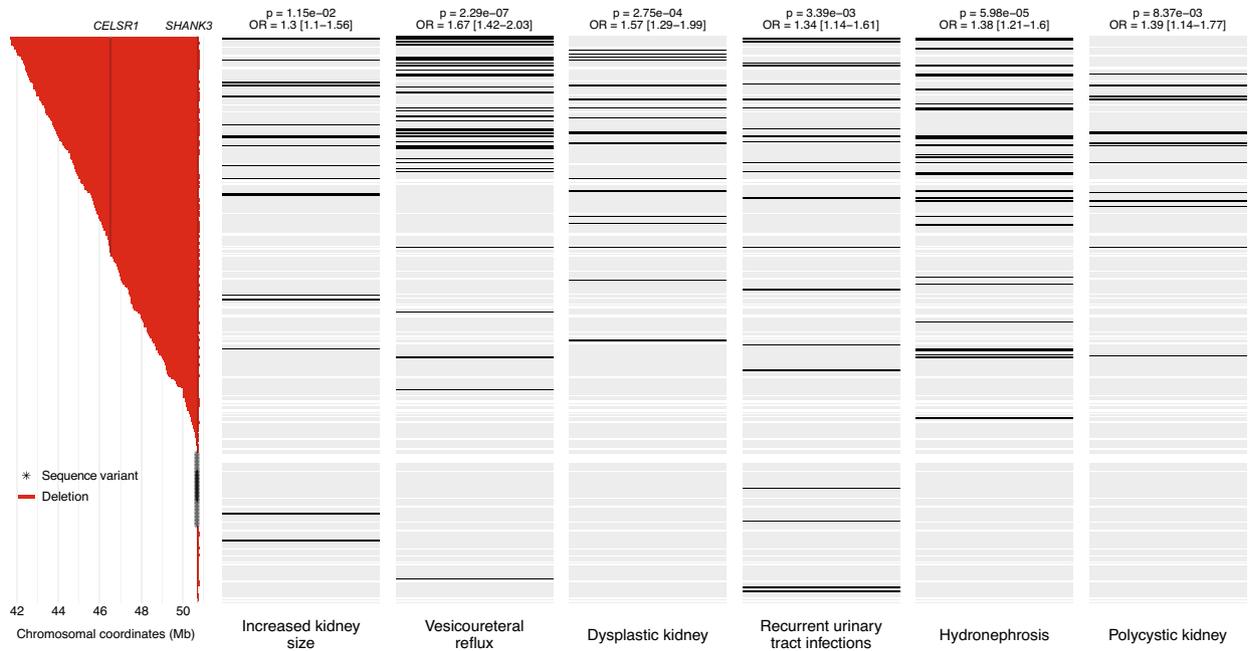
Page 13 of 18

Renal conditions

Fig. 3 Association between deletion size and renal abnormalities. On the left: deletions (in red) and *SHANK3* variants (asterisks), ordered by decreasing deletion size/position. The positions of *SHANK3* and *CELSR1*, recently implicated in renal defects, are indicated by vertical lines. On the right: status of each patient for renal conditions, lined up with the respective genetic status. The color black indicates the presence of the phenotype, gray indicates the absence of the phenotype, and white indicates missing information
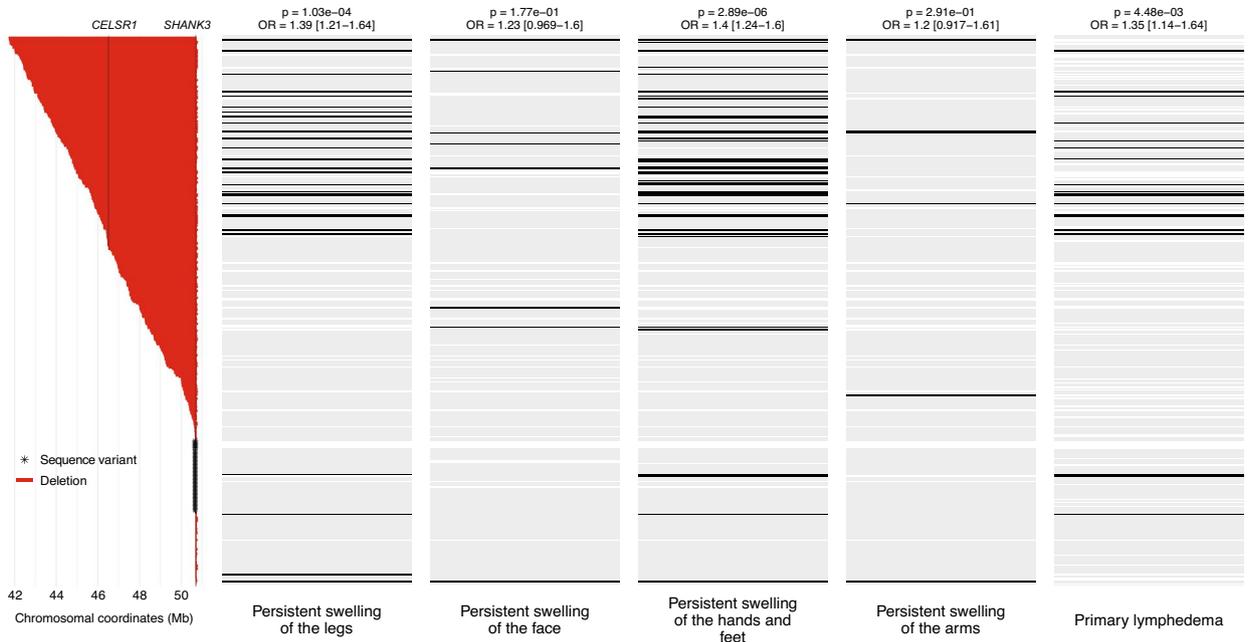
Lymphatic–related conditions

Fig. 4 Association between deletion size and lymphatic-related conditions. On the left: deletions (in red) and *SHANK3* variants (asterisks), ordered by decreasing deletion size/position. The positions of *SHANK3* and *CELSR1*, recently implicated in lymphedema, are indicated by vertical lines. On the right: status of each patient for lymphatic-related conditions, lined up with the respective genetic status. The color black indicates the presence of the phenotype, gray indicates the absence of the phenotype, and white indicates missing information

Table 5 Clinical features of individuals with *SHANK3* sequence variants, Class I deletions and Class II deletions

Variable	Sequence variants	Class I deletions	Sequence variants + Class I deletions	Class II deletions
Total N	51	64	115	286
Gender	F 21 (41.2%), M 30	F 29 (45.3%), M 35	F 50 (43.5%), M 65	F 163 (57.0%), M 123
Age (mean ± SD)	11.47 ± 8.20	12.10 ± 9.45	11.82 ± 8.89	8.43 ± 8.56
Medical features				
Kidney abnormalities[a]	1/40 (2.5%)	2/50 (4.0%)	3/90 (3.3%)	80/232 (34.5%)
Lymphedema	1/39 (2.6%)	1/54 (1.9%)	2/93 (2.2%)	20/226 (8.8%)
Congenital heart disease[b]	2/42 (4.8%)	2/53 (3.8%)	4/95 (4.2%)	42/241 (17.4%)
Recurrent infections[c]	10/37 (27.0%)	10/49 (20.4%)	20/86 (23.3%)	86/223 (38.6%)
Abnormal brain imaging	10/38 (26.3%)	14/45 (31.1%)	24/83 (28.9%)	122/235 (51.9%)
GI abnormalities[d]	27/41 (65.9%)	27/53 (50.9%)	54/94 (57.4%)	142/237 (59.9%)
Febrile seizures	6/42 (14.3%)	3/51 (5.9%)	9/93 (9.7%)	60/222 (27.0%)
Non-febrile seizures	14/41 (34.1%)	13/45 (28.9%)	27/86 (31.4%)	46/187 (24.6%)
Neonatal hypotonia	1/42 (2.4%)	9/54 (16.7%)	10/96 (10.4%)	112/255 (43.9%)
Feeding problems at birth	6/42 (14.3%)	8/54 (14.8%)	14/96 (14.6%)	125/255 (49.0%)
Hypotonia (diagnosis at any age)	27/40 (67.5%)	32/50 (64.0%)	59/90 (65.6%)	213/237 (89.9%)
Vision impairments				
Refractive errors[e]	7/37 (18.9%)	9/51 (17.6%)	16/88 (18.2%)	54/229 (23.6%)
Other visual problems[f]	2/37 (5.4%)	5/51 (9.8%)	7/88 (8.0%)	36/229 (15.7%)
Ocular abnormalities[g]	6/37 (16.2%)	7/51 (13.7%)	13/88 (14.8%)	76/229 (33.2%)
Sensorineural hearing loss	0/37 (0%)	1/48 (2.1%)	1/85 (1.2%)	10/218 (4.6%)
Dysmorphic features				
Sacral dimple	5/33 (15.2%)	6/44 (13.6%)	11/77 (14.3%)	105/218 (48.2%)
Large fleshy hands	3/28 (10.7%)	5/38 (13.2%)	8/66 (12.1%)	143/232 (61.6%)
Dysplastic or unusual fingernails	2/28 (7.1%)	5/38 (13.2%)	7/66 (10.6%)	72/232 (31.0%)
Dysplastic or unusual toenails	10/28 (35.7%)	12/38 (31.6%)	22/66 (33.3%)	139/232 (59.9%)
Supernumerary teeth	0/29 (0%)	2/34 (5.9%)	2/63 (3.2%)	16/116 (13.8%)
Spine abnormalities[h]	3/35 (8.6%)	6/53 (11.3%)	9/88 (10.2%)	35/237 (14.8%)
Torticollis	1/35 (2.9%)	3/53 (5.7%)	4/88 (4.5%)	27/237 (11.4%)
Thyroid dysfunction[i]	2/39 (5.1%)	3/58 (5.2%)	5/97 (5.2%)	9/247 (3.6%)
Too tall for age (> 95th percentile)	3/25 (12.0%)	3/38 (7.9%)	6/63 (9.5%)	57/164 (34.8%)
Too short for age (< 5th percentile)	7/25 (28.0%)	6/38 (15.8%)	13/63 (20.6%)	13/164 (7.9%)
Sleep disturbances	38/46 (82.6%)	45/58 (77.6%)	83/104 (79.8%)	176/254 (69.3%)
Psychiatric disorders				
ASD[j]	31/42 (73.8%)	31/47 (66.0%)	62/89 (69.7%)	72/159 (45.3%)
ADHD[j]	10/43 (23.3%)	14/48 (29.2%)	24/91 (26.4%)	18/176 (10.2%)
Depression[k]	3/35 (8.6%)	5/38 (13.2%)	8/73 (11.0%)	3/127 (2.4%)
Bipolar disorder[k]	5/35 (14.3%)	10/38 (26.3%)	15/73 (20.5%)	6/127 (4.7%)
Anxiety disorder[k]	8/35 (22.9%)	13/38 (34.2%)	21/73 (28.8%)	13/127 (10.2%)
Obsessive compulsive disorder[k]	3/35 (8.6%)	8/38 (21.1%)	11/73 (15.1%)	10/127 (7.9%)
Schizophrenia[k]	0/35 (0%)	0/38 (0%)	0/73 (0%)	2/127 (1.6%)
Neurodevelopment				
Verbal speech (if > 2 years)[l]	12/29 (41.4%)	25/47 (53.2%)	37/76 (48.7%)	32/170 (18.8%)
Able to dress without assistance[m]				
Always	4/27 (14.8%)	6/42 (14.3%)	10/69 (14.5%)	9/130 (6.9%)
Sometimes	4/27 (14.8%)	19/42 (45.2%)	23/69 (33.3%)	17/130 (13.1%)

Yin *et al. Molecular Autism* (2024) 15:40

Page 15 of 18

Table 5 (continued)

Variable	Sequence variants	Class I deletions	Sequence variants + Class I deletions	Class II deletions
Able to toilet independently (daytime)[m]				
Always	3/27 (11.1%)	19/43 (44.2%)	22/70 (31.4%)	9/130 (6.9%)
Sometimes	5/27 (18.5%)	6/43 (14.0%)	11/70 (15.7%)	19/130 (14.6%)
Coordination difficulties				
Always	11/32 (34.4%)	16/50 (32.0%)	27/82 (32.9%)	105/206 (51.0%)
Sometimes	19/32 (59.4%)	30/50 (60.0%)	49/82 (59.8%)	86/206 (41.7%)

Class I deletions include only *SHANK3* or *SHANK3* in combination with *ARSA*, *ACR* or *RABL2B*. Class II deletions include all deletions that do not qualify as Class I deletions

ADHD, attention deficit-hyperactivity disorder; ASD, autism spectrum disorder; F, females; GI, gastrointestinal; M, males

[a] Kidney abnormalities include Dysplastic kidney(s), Polycystic kidney(s), Hydronephrosis, Increased kidney size, Duplicate or extra kidney, and Vesicoureteral reflux. (Recurrent urinary tract infections with no reported malformations, and Kidney stones were not included.)

[b] Congenital heart disease includes Patent ductus arteriosus, Ventricular septal defect, 'Hole in the heart', Bicuspid aortic valve, Heart valve problems, Persistent left superior vena cava, Total anomalous pulmonary venous return or connection, Enlarged aorta, Enlarged aortic root, Coarctation of the aorta, and Aortic stenosis

[c] Recurrent infections include Four or more new ear infections within one year; Two or more serious sinus infections within 1 year; 2 or more months on antibiotics with little effect; Two or more pneumonias within 1 year; Recurrent, deep skin or organ abscesses; Persistent thrush in mouth or fungal infection on skin; Need for intravenous antibiotics to clear infections; Two or more deep-seated infections including septicemia; Recurrent pneumonia; and Recurrent cellulitis

[d] GI abnormalities include Gastroesophageal reflux, Silent reflux, Chronic constipation, Chronic diarrhea, Irritable bowel syndrome, and Inflammatory bowel disease (Crohn's disease, Ulcerative colitis)

[e] Refractive errors include Farsighted, Nearsighted, and Astigmatism

[f] Other visual problems include Depth perception problems, Cortical vision impairment, and Blindness

[g] Ocular abnormalities include Strabismus, Eye movement abnormalities, and Cataract

[h] Spine abnormalities include Scoliosis, Kyphosis, and Lordosis

[i] Thyroid dysfunction includes Hypothyroidism, Hyperthyroidism, Hashimoto's disease, and Graves' disease

[j] Only individuals aged 3 years and older were included in the analysis

[k] Only individuals aged 5 years and older were included in the analysis

[l] Uses meaningful single words or phrases that are understandable to others

[m] Only individuals aged 4 years and older were included in the analysis

murmur and congenital heart defects; each was present in 11% (45/401) of the patients. Cardiac abnormalities had been associated with larger deletions in a small study of 30 individuals with ring chromosome 22 [38], but this finding had not been replicated in other cohorts with larger samples. Similarly, recurrent ear infections were correlated with deletion size in a study of 51 individuals with 22q13 deletion syndrome [14], but not in subsequent studies with larger sample sizes. Here, we showed that larger deletions were associated with recurrent ear and sinus infections, chronic bronchitis, recurrent pneumonia, and the need for intravenous antibiotics. These new associations were identified by analyzing a much larger dataset compared to previous studies.

Additionally, larger deletion sizes were associated with abnormalities in brain MRI and CT scans, including abnormal ventricles, decreased gray matter, and arachnoid cysts, which are well documented in individuals with PMS [4, 40] but had not been previously associated with deletion size. No other neurological morphological abnormalities or symptoms were associated with larger deletion sizes, except for current seizures and febrile

seizures. Specific subtypes of seizures (e.g., absence seizures, tonic–clonic seizures, simple and complex partial seizures) were not associated with deletion size. Analyses also revealed that individuals with small (Class I) deletions or *SHANK3* sequence variants had fewer febrile seizures compared to those with Class II deletions (Table 5), whereas no difference was observed for non-febrile seizures. The difference between febrile and non-febrile seizures is difficult to explain and needs to be confirmed in future studies.

We found that individuals with larger deletions were less likely to be diagnosed with ADHD and ASD, or to present with behavioral features of these disorders. As previously suggested [7], this finding might reflect the severity of the disability due to the larger deletion size "masking" autistic or ADHD manifestations. The negative association between deletion size and ADHD and hyperactivity symptoms might further be linked to the severity of low muscle tone and its association with larger deletion sizes. Social, cognitive and communication development, representing the depth and breadth of intellectual disability, were inversely associated with

Yin *et al. Molecular Autism* (2024) 15:40

Page 16 of 18

deletion size. The abilities most significantly impacted were those related to understanding social conventions and interacting with others, whereas individual abilities and anxiety were evenly distributed throughout the sample.

Our method had previously been used as a proxy for the number of deleted genes, which is linked to the severity of the genetic condition [4]. Although our findings indicate the involvement of other genes upstream of *SHANK3*, our approach cannot determine their identity. The proportional association of deletion size with developmental delays and severity of cognitive or social disabilities suggests a cumulative effect of multiple gene alterations. Specifically, the consistency of associations between deletion size and gross motor milestone delays, as opposed to other features such as endocrine or sensory issues, for instance, indicates that a number of haploinsufficient genes spanning the terminal 9 Mb of the long arm of chromosome 22 affect gross motor functions and muscle tone when deleted. For those phenotypes not associated with larger deletion sizes, we suggest that some may not actually be associated with PMS. This would apply to phenotypes with a prevalence in our cohort similar to that observed in the general population, as is the case for the majority of phenotypes (e.g., painful urination, night terrors, or enuresis). Phenotypes present in nearly all affected individuals could be part of the core features of PMS caused by defects in *SHANK3*. Haploinsufficiency of *SHANK3* resulting from sequence variants is sufficient to cause intellectual disability, ASD or autistic features, severe speech deficits, hypotonia, motor skill deficits, regression, seizures, brain abnormalities, mild dysmorphic features, and feeding and gastrointestinal problems [5, 6, 9, 16]. The present data suggest that in addition to the role of *SHANK3* in these phenotypes, the loss of other genes in individuals with larger deletions can also contribute to some of these manifestations. To date, only two genes with autosomal dominant inheritance and loss-of-function mechanism have been identified in the interval deleted in PMS. *TCF20*, located in the 22q13.2 region and deleted in individuals with the largest terminal deletions (>8.6 Mb), is implicated in a novel syndrome characterized by motor and language delay, ID, ASD, hypotonia, and variable dysmorphic features [25, 26]. Large terminal deletions involving *TCF20* are rare, suggesting that there are other genes contributing to the phenotype that remain to be identified. *CELSR1,* located in the 22q13.31 region and deleted in individuals with deletions >4.4 Mb, has recently been implicated in lymphedema [27, 28]. Interestingly, of the 22 individuals with lymphedema included in this study, 20 (91%) have a >4.4 Mb deletion including *CELSR1*, lending further support to this gene playing a role in the lymphedema observed in individuals with PMS. The report of lymphedema in two individuals without loss of *CELSR1* is unexpected and needs to be confirmed. More recently, *CELSR1* loss of function was also associated with renal defects, including ureteropelvic junction obstruction and renal atrophy [41]. This finding indicates that loss of *CELSR1* may be responsible for at least some of the kidney abnormalities in individuals with PMS with larger deletions, although loss of a more distal gene (or genes) may also increase the risk of renal defects in those with deletions that do not include *CELSR1* (see Fig. 3).

Limitations

There are some limitations that warrant consideration. One caveat of the current analyses is that, when used as a predictor of binomial outcomes, such as the occurrence of a clearly defined somatic phenotype, deletions can only act as a proxy for the contribution of chromosomal regions of interest and hint at potential locations of the genes involved in specific phenotypes. Comparisons between individual genes could be used to narrow down the range of genetic contributors to phenotypes. However, since the majority of genetic abnormalities studied are contiguous deletions, there is a strong correlation among the deleted genes that would require the development of specific models. We also note that not all participants diagnosed with PMS caused by 22q13.33 deletions were tested with a genome-wide array; some were only tested for copy number variants on chromosome 22q12.3-terminus using a custom microarray in a research setting [8, 18], and karyotypes were not available in many cases. Because of this missing information in a significant proportion of individuals, we did not take into account the possible contribution of additional chromosomal aberrations to the phenotype. Another limitation stems from the reporting of data by families, with the potential for error, particularly concerning complex medical terms. In the future, novel associations, such as those with congenital heart defects in the present study, would ideally be validated by asking families reporting these defects to provide relevant medical records. Additionally, in terms of future directions, there exists a third source of data based upon natural language processing of the patients' clinical notes, encoded as Unified Medical Language System concepts. We expect that this will provide a range of phenotypes that were not captured by the registry questionnaires and enable further detailed analyses. Finally, it is worth noting that due to missing genetic and/or phenotypic data, we were unable to analyze three-quarters of the individuals in the registry.

Yin *et al. Molecular Autism*   (2024) 15:40

Page 17 of 18

Conclusions

In this study, we presented a method for the curation and integration of data from a patient-driven registry and applied it to the PMSIR. We used this data source to demonstrate that deletion size is associated with the severity of delays in gross motor development and a spectrum of generalized muscle tone impairments. Furthermore, we corroborated previous studies indicating that larger deletion sizes are associated with renal defects, lymphedema, large fleshy hands, dysplastic toenails, sacral dimple, and more severe impairments in verbal ability and socio-cognitive development. We also showed novel associations, including a higher prevalence of congenital heart defects, neuroimaging abnormalities, and infections among individuals with larger deletions. Overall, the results demonstrated the feasibility of utilizing this data source to perform genotype-phenotype analyses. Our method can serve as a reference for future research projects related to this and similar data sources.

Abbreviations

ADD	Attention deficit disorder
ADHD	Attention deficit-hyperactivity disorder
ASD	Autism spectrum disorder
CI	Confidence interval
CT	Computed tomography
FDR	False discovery rate
FISH	Fluorescent in situ hybridization
ID	Intellectual disability
MRI	Magnetic resonance imaging
Mb	Megabase
OR	Odds ratio
PCORI	Patient Centered Outcome Research Institute
PMS	Phelan-McDermid syndrome
PMS_DN	Phelan-McDermid Syndrome Data Network
PMSF	Phelan-McDermid Syndrome Foundation
PMSIR	Phelan-McDermid Syndrome International Registry
US	United States

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13229-024-00619-z.

> Additional file 1: Figure S1. Data management workflow.
>
> Additional file 2: Table S1. Results of the statistical analyses for all phenotypes.
>
> Additional file 3: Supplementary figures illustrating the results of the analyses by group of phenotypic features.

Author contributions

RY, MW, CB, and PA analyzed and interpreted the data and wrote the manuscript. CHK, MTM, and CK participated in data collection and integration. GB and MO are founders of the PMS International Registry of the PMS Foundation and played an essential role in the initiation and conceptualization of this project. EJH coordinated the PMS Data Network. BM, RD, and KH entered and curated the genetic data in the PMS International Registry. CB was responsible for the final curation and interpretation of the genetic variants. All authors read and approved the final manuscript.

Availability of data and materials

The PMS International Registry dataset analyzed in this study is available by application to the Phelan-McDermid Syndrome Foundation. All the code for ETL (extract, transform, load) and analysis is openly available on GitHub (https://github.com/hms-dbmi/PMSDN_R-ETL, https://github.com/hms-dbmi/PMSDN_GenoPheno).

Declarations

Ethics approval and consent to participate

The project was approved by the Harvard Medical School Institutional Review Board (HMS IRB14-2161). The PMS International Registry obtained continuing review approval from the Advarra (formerly Chesapeake) IRB as of January 21, 2021 (CR00182785). All participants in the registry provided written informed consent to participate.

Consent for publication

Not applicable.

Competing interests

C. Betancur is Editor-in-Chief of Molecular Autism. The other authors declare no competing interests.

Author details

¹Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA. ²Phelan-McDermid Syndrome Foundation, Osprey, FL 34229, USA. ³INSERM, CNRS, Neuroscience Paris Seine, Institut de Biologie Paris Seine, Sorbonne Université, 75005 Paris, France. ⁴Present Address: Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, USA.

References

1. Phelan K, Rogers RC, Boccuto L. Phelan-McDermid syndrome. GeneReviews®. 2018. https://www.ncbi.nlm.nih.gov/sites/books/NBK1198/.
2. Phelan K, McDermid HE. The 22q13.3 deletion syndrome (Phelan-McDermid syndrome). Mol Syndromol. 2012;2:186–201.
3. Kolevzon A, Angarita B, Bush L, Wang AT, Frank Y, Yang A, et al. Phelan-McDermid syndrome: a review of the literature and practice parameters for medical assessment and monitoring. J Neurodev Disord. 2014;6:39.
4. Soorya L, Kolevzon A, Zweifach J, Lim T, Dobry Y, Schwartz L, et al. Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. Mol Autism. 2013;4:18.
5. Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S, et al. Meta-analysis of SHANK mutations in autism spectrum disorders: a gradient of severity in cognitive impairments. PLoS Genet. 2014;10: e1004580.
6. Levy T, Foss-Feig JH, Betancur C, Siper PM, del Pilar M, Halpern D, et al. Strong evidence for genotype–phenotype correlations in Phelan-McDermid syndrome: results from the developmental synaptopathies consortium. Hum Mol Genet. 2021;31:625–37.

Yin *et al. Molecular Autism*   (2024) 15:40

Page 18 of 18

7. Sarasua SM, Dwivedi A, Boccuto L, Chen C-F, Sharp JL, Rollins JD, et al. 22q13.2q13.32 genomic regions associated with severity of speech delay, developmental delay, and physical features in Phelan-McDermid syndrome. Genet Med. 2014;16:318–28.

8. Sarasua SM, Boccuto L, Sharp JL, Dwivedi A, Chen C-F, Rollins JD, et al. Clinical and genomic evaluation of 201 patients with Phelan-McDermid syndrome. Hum Genet. 2014;133:847–59.

9. Nevado J, García-Miñaúr S, Palomares-Bralo M, Vallespín E, Guillén-Navarro E, Rosell J, et al. Variability in Phelan-McDermid syndrome in a cohort of 210 individuals. Front Genet. 2022;13: 652454.

10. Kolevzon A, Delaby E, Berry-Kravis E, Buxbaum JD, Betancur C. Neuropsychiatric decompensation in adolescents and adults with Phelan-McDermid syndrome: a systematic review of the literature. Mol Autism. 2019;10:50.

11. Kohlenberg TM, Trelles MP, McLarney B, Betancur C, Thurm A, Kolevzon A. Psychiatric illness and regression in individuals with Phelan-McDermid syndrome. J Neurodev Disord. 2020;12:7.

12. Monteiro P, Feng G. SHANK proteins: roles at the synapse and in autism spectrum disorder. Nat Rev Neurosci. 2017;18:147–57.

13. Bonaglia MC, Giorda R, Borgatti R, Felisari G, Gagliardi C, Selicorni A, et al. Disruption of the ProSAP2 gene in a t(12;22)(q241;q13.3) is associated with the 22q13.3 deletion syndrome. Am J Hum Genet. 2001;69:261–8.

14. Wilson HL, Wong ACC, Shaw SR, Tse W-Y, Stapleton GA, Phelan MC, et al. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. J Med Genet. 2003;40:575–84.

15. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet. 2007;39:25–7.

16. De Rubeis S, Siper PM, Durkin A, Weissman J, Muratet F, Halpern D, et al. Delineation of the genetic and clinical spectrum of Phelan-McDermid syndrome caused by SHANK3 point mutations. Mol Autism. 2018;9:31.

17. Bonaglia MC, Giorda R, Beri S, De Agostini C, Novara F, Fichera M, et al. Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. PLoS Genet. 2011;7: e1002173.

18. Sarasua SM, Dwivedi A, Boccuto L, Rollins JD, Chen C-F, Curtis Rogers R, et al. Association between deletion size and important phenotypes expands the genomic region of interest in Phelan–McDermid syndrome (22q13 deletion syndrome). J Med Genet. 2011;48:761–6.

19. Samogy-Costa CI, Varella-Branco E, Monfardini F, Ferraz H, Fock RA, Barbosa RHA, et al. A Brazilian cohort of individuals with Phelan-McDermid syndrome: genotype-phenotype correlation and identification of an atypical case. J Neurodev Disord. 2019;11:13.

20. Upadia J, Gonzales PR, Atkinson TP, Schroeder HW, Robin NH, Rudy NL, et al. A previously unrecognized 22q13.2 microdeletion syndrome that encompasses TCF20 and TNFRSF13C. Am J Med Genet A. 2018;176:2791–7.

21. Simenson K, Õiglane-Shlik E, Teek R, Kuuse K, Õunap K. A patient with the classic features of Phelan-McDermid syndrome and a high immunoglobulin E level caused by a cryptic interstitial 0.72-Mb deletion in the 22q13.2 region. Am J Med Genet A. 2014;164A:806–9.

22. Disciglio V, Lo Rizzo C, Mencarelli MA, Mucciolo M, Marozza A, Di Marco C, et al. Interstitial 22q13 deletions not involving SHANK3 gene: a new contiguous gene syndrome. Am J Med Genet A. 2014;164A:1666–76.

23. Palumbo P, Accadia M, Leone MP, Palladino T, Stallone R, Carella M, et al. Clinical and molecular characterization of an emerging chromosome 22q13.31 microdeletion syndrome. Am J Med Genet A. 2018;176:391–8.

24. Ha JF, Ahmad A, Lesperance MM. Clinical characterization of novel chromosome 22q13 microdeletions. Int J Pediatr Otorhinolaryngol. 2017;95:121–6.

25. Torti E, Keren B, Palmer EE, Zhu Z, Afenjar A, Anderson IJ, et al. Variants in TCF20 in neurodevelopmental disability: description of 27 new patients and review of literature. Genet Med. 2019;21:2036–42.

26. Vetrini F, McKee S, Rosenfeld JA, Suri M, Lewis AM, Nugent KM, et al. De novo and inherited TCF20 pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological impairments with similarities to Smith-Magenis syndrome. Genome Med. 2019;11:12.

27. Gonzalez-Garay ML, Aldrich MB, Rasmussen JC, Guilliod R, Lapinski PE, King PD, et al. A novel mutation in CELSR1 is associated with hereditary lymphedema. Vasc Cell. 2016;8:1.

28. Maltese PE, Michelini S, Ricci M, Maitz S, Fiorentino A, Serrani R, et al. Increasing evidence of hereditary lymphedema caused by CELSR1 loss-of-function variants. Am J Med Genet A. 2019;179:1718–24.

29. Kothari C, Wack M, Hassen-Khodja C, Finan S, Savova G, O'Boyle M, et al. Phelan-McDermid syndrome data network: integrating patient reported outcomes with clinical notes and curated genetic reports. Am J Med Genet B Neuropsychiatr Genet. 2018;177:613–24.

30. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. AMIA Jt Summits Transl Sci Proc. 2013;2013:6–8.

31. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. Brief Bioinform. 2015;16:280–90.

32. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010;17:124–30.

33. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.

34. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. Bioinformatics. 2010;26:1205–10.

35. Valenta Z, Pitha J, Poledne R. Proportional odds logistic regression–effective means of dealing with limited uncertainty in dichotomizing clinical outcomes. Stat Med. 2006;25:4227–34.

36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.

37. Xu N, Lv H, Yang T, Du X, Sun Y, Xiao B, et al. A 29 Mainland Chinese cohort of patients with Phelan-McDermid syndrome: genotype-phenotype correlations and the role of SHANK3 haploinsufficiency in the important phenotypes. Orphanet J Rare Dis. 2020;15:335.

38. Jeffries AR, Curran S, Elmslie F, Sharma A, Wenger S, Hummel M, et al. Molecular and phenotypic characterization of ring chromosome 22. Am J Med Genet A. 2005;137:139–47.

39. Tabet A-C, Rolland T, Ducloy M, Lévy J, Buratti J, Mathieu A, et al. A framework to identify contributing genes in patients with Phelan-McDermid syndrome. NPJ Genom Med. 2017;2:32.

40. Holder JL Jr, Quach MM. The spectrum of epilepsy and electroencephalographic abnormalities due to SHANK3 loss-of-function mutations. Epilepsia. 2016;57:1651–9.

41. Alpaslan M, Mestré-Godin S, Lay A, Giacalone G, Helaers R, Adham S, et al. Ureteropelvic junction obstruction with primary lymphoedema associated with CELSR1 variants. J Med Genet. 2023;60:1161–8.

Publisher's Note