Identification of genotype-phenotype associations in Phelan-McDermid Syndrome using patient-sourced data from an international registry

Maxime Wack, Claire Hassen-Khodja, Megan O'Boyle, Geraldine Bliss, Liz Horn, Andria Cornell, Cartik Saravana Muthu, Catalina Betancur, Isaac Kohane, Paul Avillach

Abstract

Phelan-McDermid Syndrome (PMS) is a syndromic form of autism caused by terminal deletions of the long arm of chromosome 22 affecting at least the SHANK3 gene. It variably associates autism, global developmental delay, delayed speech, neonatal hypotonia, and mildly dysmorphic features. Isolated haploinsufficiency of SHANK3 has been shown to be responsible of a subset of PMS features. The PMS International Registry (PMSIR) compiles clinical data in the form of Patient Reported Outcomes, as well as patient-sourced genetic test results. Data from the PMSIR have been harmonized and integrated into the i2b2/tranSMART clinical & genomics data warehouse. We conducted genotype-phenotype analyses using regression models associating the deletion size as a predictor of the different clinical outcomes. 156 patients were included, with deletion sizes ranging from 10.34 kb to 9.057 Mb, 6 patients presenting small isolated SHANK3 mutations.Increased deletion size is significantly associated with delay in gross motor acquisitions, vesicoureteral reflux, socio-emotional and behavioral development delays, verbal speech, mild dysmorphic features (large fleshy hands, dysplastic toenails/fingernails and sacral dimple), and a spectrum of conditions related to poor muscle tone, suggesting the implication of genes upstream of SHANK3. In this study using data from the PMSIR, we demonstrate the use of entirely patient-sourced registry data consisting of PRO items filled by the parents, and curated genetic test reports to conduct genotype-phenotype analyses. Known results are replicated and novel findings show the ability of registry data to uncover new associations between comorbidities and deleted chromosomal regions in PMS.

# Introduction

The 22q13 deletion syndrome (OMIM 606232), also called Phelan McDermid Syndrome (PMS)1,2, is a syndromic form of Autism Spectrum Disorder (ASD)2. It associates autistic features, with rates of autism or autistic-like behavior ranging from 44% to 94%3,4, and an estimate of 84% of cases using the gold standard diagnostic instrument, the Autism Diagnostic Observation Schedule5 ; global developmental delay in more than 98% of patients2,6 ; and Intellectual Disability with delayed or absent speech in nearly 100% of cases2–8. It accounts for an estimated 0.5% of ASD cases3,4,8–10 and 2% of ID cases3,8,9, with an equal number of male and female patients4,6.  
In addition to ASD and ID, PMS patients are characterized with specific behavioral features, such as impairments in communication and social activities, chewing of non-food items (80-90%), teeth grinding (~25%) and aggressive behavior (10-15%)1–3,6,12.  
The syndrome is a contiguous gene disorder caused by random deletions of the distal long arm of chromosome 22, deleting the *SHANK3* gene located at the end of the chromosome. Diagnosis is made by genetic testing for deletion of the 22q13 chromosomal region.  
As of today there are approximately 1,100 diagnosed cases worldwide, with at least 520 cases confirmed in the US, and no treatment are yet available, although IGF1 has been shown to restore the *SHANK3* pathway13 and has been used successfully in a pilot randomized controlled trial14.

*SHANK3* (SH3 and multiple ankyrin repeat domain 3) is a gene coding for a post-synaptic scaffolding protein binding to several different neurotransmitter receptors and is considered to be of great importance in the assembly and plasticity of the postsynaptic density (PSD) in brain synapses. It has been shown to be involved in non-PMS autistic spectrum disorders7,8,10,15–19, further indicating its role in the neurobehavioral features of PMS. Deficiency (deletion or loss of function) of a single copy of *SHANK3* has been shown to be sufficient to cause the disease20, due to cases and mice models involving only *de novo* point mutations, or interstitial deletions only disrupting *SHANK3*7,16,21–24. PMS patients with isolated *SHANK3* mutations or deletions show major speech delays and profound ID, global developmental delay, ASD symptoms, slightly delayed motor milestones, mild hypotonia and minor dysmorphic features (dolichocephaly, epicanthic folds).  
*SHANK3* haploinsufficiency, with Fragile X Syndrome, are the two most frequent single gene genetic causes of autism/autism spectrum disorder.

Patients exhibit heterogeneous chromosomal abnormalities in the 22q13 region: terminal deletions, interstitial deletions, duplications and ring chromosome, resulting in randomly distributed deletions sizes from 100 kb to over 9 Mb6. Previous studies have shown that there is a gradient of severity in cognitive impairments8, associated with the size of the deletions, also linked to the presence of other phenotypic features5,12,25,26.  
However, a growing number of cases are reported sharing features with the diagnosis of 22q13 deletion syndrome without presenting the classical terminal 22q deletion present in the majority of PMS cases. Instead, they present with interstitial deletions with sizes as small as 0.72Mb beyond the range of the *SHANK3* gene, indicating that a subset of features of PMS are not caused by *SHANK3* haploinsufficiency alone27–31. These patients show severe developmental delays, hypotonia with feeding difficulties, overgrowth,, speech impairments, seizures and abnormal neuroimagery, strabismus, congenital heart defects, kidney anomalies, ASD, minor dysmorphic features (epicanthic folds, dysplastic toenails, fleshy hands) and the absence of high pain tolerance.  
The Phelan McDermid Syndrome International Registry (PMSIR)32, a patients’ family-powered registry was launched in 2011 by the Phelan McDermid Syndrome Foundation, founded by the parents of PMS patients. It is the only international registry for this disease, and today includes more than half the worldwide population of known patients. The information contained in the registry is obtained and annotated by the patients’ parents.  
We hypothesized that the retrospective, patient-reported data, including genetic test results obtained from the Phelan McDermid Syndrome International Registry32 could be used to conduct deletion analyses and uncover phenotypes and comorbidities significantly linked to genes other than *SHANK3*.

# Methods

The US Patient Centered Outcome Research Institute (PCORI) funded Phelan McDermid Syndrome Data Network (PMS\_DN) project mission is to advance knowledge of Phelan-McDermid Syndrome and related conditions by integrating 1) Patient reported outcomes from a patient registry, 2) Curated genetic test results, in the centralized i2b2/tranSMART platform to facilitate patient-centered research and 3) clinical knowledge extracted from clinical notes. The latter was not used in this study. The network promotes a culture of transparency and authentic engagement and leadership of families in network vision, governance, and generation of research priorities. PMS\_DN uses data entered on the online Phelan McDermid Syndrome International Registry by the patients' parents.  
The registry comprises 1,300 patient reported outcome (PRO) items over three distinct questionnaires, a “clinical” questionnaire with questions regarding diagnosed comorbidities, symptoms, tests and treatments for the whole 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 with specific questions aimed at patients aged 15 or more, regarding the evolution of symptoms after puberty. The latter was not used in this study. The surveys for the phenotypic data can be retaken at any point by the patient’s family, allowing for longitudinal analysis and ensuring the most complete information for each patient.  
Alongside phenotypic data, the registry contains curated genetic testing of chromosomal abnormalities from karyotypes, FISH probes, comparative genomic hybridization (aCGH), microarrays and sequencing represented using the ISCN Standard Nomenclature33. The scanned paper reports of the genetic test results are uploaded by the families on the platform then manually curated by trained genetic counselors, filling 57 structured fields to represent the genetic abnormalities.

All the phenotypic and genotypic data in the PMS registry were integrated in the i2b2/tranSMART34–36 data warehouse to facilitate their integration with future updates of the registry and additional sources of data. Also, a PMS ontology was created to tag the registry data elements in a controlled vocabulary and organizing them to allow for efficient user-selection of various stratifications of the patients in the registry. 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 PMS\_DN over time, and consistency in the way every variable is coded in the system. Variables representing each clinical item in the survey were categorized as being either historical or evolving over time, to facilitate further analysis by other researchers. All historical data was kept for each time point, and the most recent entry for historical items was kept as the most accurate value (Supplemental data 1a and 1b). This has enabled greater reliability in data entry. For example, when taking the survey at a new time point, displaying the previous entry prevents erasure or duplication of previously entered data, guaranteeing that the most recent record holds up to date information.

Because we reused already existing genetic test results, they presented a huge disparity in techniques, accuracy and references. All the curated genetic test result information for each patient were manually reviewed to extract the coordinates and genome assembly of the chromosomal abnormalities. Chromosomal coordinates for aCGH were extracted from the relevant structured fields (chromosome, gain/loss, start, end), from ISCN nomenclature and comments where necessary. Chromosomal coordinates for FISH results were directly obtained in the GRCh38/hg38 genome assembly from the NCBI Clone database. When multiple assays were available for the same region, the most recent or most precise one was kept: sequencing, over array CGH, over FISH. Chromosomal coordinates were transformed from each original human genome assembly to the latest one available at the time of this study, GRCh38/hg3837, using the UCSC liftOver tool. Coordinates that did not map to GRCh38/hg38 were kept in the original genome assembly and were not used in the analysis. Only the terminal deletions were kept in the analyses to ensure comparability and relevant use of the models and their interpretation.

For the analysis reported here, we compared the extent of deleted genetic material on the terminus of chromosome 22 with the occurrence of each phenotypic feature, in a multi-phenotype-genotype association study (aka Phenome-Wide Association study - PheWas38), using the deletion extent as the genetic anomaly of interest. Phenotypic features, both physical and behavioral, including features commonly associated with the diagnosis of PMS, were selected for analysis. Only features with strictly more than five patients in each subgroup were selected to satisfy technical requirements of the regression models. The comparisons were made using only data from patients for which deletion coordinates were available or translatable to GRCh38/hg38 human genome assembly, to ensure comparability and consistency of the results. We built multivariate models associating the phenotypic outcome with the deletion size, adjusted for the age and gender of the patients.  
Linear regression was used when the outcome was a continuous variable (birth length, weight, head circumference and APGAR scores), logistic regression when the outcome was binomial (phenotype present or not), a method used in previous deletion analyses in PMS12,25,26, and proportional odds logistic regression39 for ordinal outcomes (severity of a phenotype on a 4 point scale from “absent” to “always present”, age ranges for the acquisition of developmental milestones). The False Discovery Rate (FDR) method was used to correct the p values for multiple hypotheses testing40. Odd ratios (and their 95% confidence intervals) are given for every million deleted base pairs.

Statistical analyses were conducted using R version 3.2.4. The proportional odds logistic regression used the clm function from the "ordinal" package.  
The project was approved by Harvard Medical School Institutional Review Board (HMS IRB14-2161).

# Results

## Population and variable selection

As of December 04 2015, date of the registry data export used in this study, 364 families, representing around 50% of all US PMS patients, consented to participate in the network and PCORnet studies. 310 out of these 364 patients in the PMS\_DN entered demographic information. Among these 310 patients, 306 completed the clinical questionnaire and 234 the developmental questionnaire. A total of 310 patients had demographic and clinical or developmental information entered.  
Demographic information for the included patients is described in Table 1. There were no demographic biases at inclusion. Completion rates ranged from 10.26% to 99.36% of items, with a mean completion rate of 77.98%.

Chromosomal abnormalities span all chromosomes except 14, 17 and 18. From the test results of 162 patients we extracted 231 validated deletion/duplication ranges expressed relative to GRCh38/hg38, including 173 on chromosome 22: 10 mutations, 157 deletions and 6 duplications. 159 patients presented terminal deletion of chromosome 22, including 6 patients with point mutations of *SHANK3*. Deletion sizes ranged between 10.34 kb to 9.057 Mb, with a median deletion size of 3.587 Mb.

A total of 156 patients who entered demographic and clinical or developmental phenotypic information as well as presented terminal deletion of chromosome 22 or point mutation of *SHANK3* were included for analysis.

Of the 1,300 patient reported outcome (PRO) items present in the registry, 566 were selected as representing phenotypic outcomes. PRO items not selected were items not referring directly to a symptom or a condition, such as “Does the patient have a primary cardiologist?”, “How was the genetic test paid?”, “Has the patient had any of the following tests – Manometry testing”, etc. 302 phenotypes satisfied the inclusion criteria and were included in the analyses.

## Deletion analyses results

162 patients had at least one genetic test result for which deletion/duplication coordinates could be translated into the GRCh38/hg38 human genome assembly on chromosome 22. Figure 1 represents the chromosomal abnormalities on chromosome 22 for these 162 patients. Panel A of the figure is a representation of chromosome 22 with the zoomed region framed in red. Panel B represents the deletion sizes for each patient, one line per patient, ordered by the decreasing total size of deleted genetic material. Duplicated regions are represented as blue boxes, deleted regions as red boxes, mutations of *SHANK3* as dark red boxes. The range of the *SHANK3* gene is represented as a grey transparent box across the figure. Horizontal black lines help relate different chromosomal abnormalities for patients with more than one abnormality.

302 phenotypes were tested against deletion sizes for 156 patients using the models described in the methods section. After correction of the p values using the FDR method, 50 phenotypes were significantly associated with deletions of greater size, meaning that these phenotypes are probably associated with at least one other gene than *SHANK3* on the long arm of chromosome 22, independently or in association with *SHANK3*. Complete results are presented in Supplemental data 2, and figures for all groups of phenotypes in Supplemental data 3.

Significantly associated phenotypes with positive effect sizes are presented in Table 2, with their respective raw p value, FDR adjusted p value, and odd ratio with 95% confidence interval (odd ratios are expressed for every million base pair deleted). They include the age of acquisition for all main gross motor developmental milestones: (Figure 1) Hold head (p = 0.0023), Crawl (p = 5.99e-05), Sit (p = 2.65e-06), Walk (p = 3.67e-07), and Climb stairs (p = 0.0414) where greater deletion sizes are associated with later ages of acquisition; certain kidney ailments: (Figure 2) Vesico-ureteral reflux (p = 0.0023) and Hydronephrosis (p = 0.0455); a range of phenotypes related to gross motor and muscle tone: Balance maintaining (p = 0.0023), Floppy baby (p = 0.0023), Neonatal hypotonia (p = 0.00144), Apnea (p = 0.0399), and Fatigue (p = 0.00961); multiple items related to feeding difficulties during the first-year development (p = 0.0023); dysmorphic features commonly described in PMS: Sacral dimple (p = 0.00019), Large fleshy hands (p = 0.000201), Dysplastic toenails (p = 0.00961) and fingernails (p = 0.0463); and central nervous system anomalies: Abnormal Cranial CAT scan (p = 0.00411), Abnormal MRI (p = 0.0282), and Febrile seizures (p = 0.0044).

Significantly associated phenotypes with negative effect sizes are presented in Table 3. They include behaviors related to hyperactivity: Act as if driven by a motor (p = 0.0023), Run around and climb excessively (p = 0.0232), Symptoms of ADD/ADHD (p = 0.0156), Difficulty going back to sleep after nighttime awakening (p = 0.0441); social behaviors: Plays alongside others but not with them (p = 0.00362), Respond to others emotions (p = ), Verbal speech ability (p = 0.0114); and differential diagnoses: ADHD (p = 0.0104), Pervasive Developmental Disorder (p = 0.00578), Autism (p = 0.0228).

No allergy, anesthesia related, cardiovascular, ears nose and throat, immunologic, psychiatric, sensory, skin, and sleep conditions were significantly associated with greater deletion sizes. Other classical dysmorphic features in PMS were not found associated with deletion size, nor forms of aggressive behavior. Non-febrile seizures were not found associated with deletion size, as opposed to febrile seizures. No fine motor developmental milestones were found associated with deletion size, when all major gross motor milestones were found to be associated with greater deletions.

# Discussion

We demonstrate the feasibility of using registry data based on Patient Reported Outcome questionnaires to conduct a multi-phenotype-genotype association study and further demonstrate replication of findings of previous prospective association studies, as well as discovery of novel findings in renal malformations and gross motor development. This is one of the largest study of its kind in PMS and the first using the Phelan McDermid Syndrome International Registry. Using the PMSIR allowed us to recruit the second largest sample size for a genotype-phenotype association analysis in PMS (Soorya et al.(2013), 32 patients5, Sarasua et al.(2014), 70 patients26, Sarasua et al.(2011), 71 patients25, Sarasua et al.(2014), 201 patients12). The data provided by this patient-driven network differs from data obtained from medical records and is rich with information that only caregivers and families can provide. The information being reported by the parents themselves, as proxy of the patients, allows the registry items to be more numerous and more specific in all domains, which would be very difficult and costly in a conventional academic-institution-led prospective study. Indeed, children with PMS exhibiting hyperactive and aggressive behaviors, can be difficult to manage, and filling in extensive surveys during interviews with the pediatrician is very difficult, and subject to recall bias. Parents have more time to fill the surveys at home when the child is asleep, are more able to to make corrections to entered data, and are more likely to answer items more thoroughly and keep the surveys updated. The consistency of the results throughout the different analyses and with previous findings demonstrates the reliability of the information contained in the registry. The extended precision of the results shows the potential of analyses on the data from the PMSIR to pinpoint very specific phenotype associations in this syndrome.

Renal malformations (hydronephrosis and vesicoureteral reflux), present in nearly 30% of PMS patients3, are strongly associated with greater deletion sizes, with the distribution of deletion sizes hinting at a region of interest on chromosome 22, which had only been reported once for all renal abnormalities regardless of the precise condition5. This indicates that renal malformations in PMS, absent from patients with isolated *SHANK3* mutations7,16,21–24, but found in patients with interstitial deletions not involving *SHANK3*30, are related to different genes than *SHANK3* playing a role in the renal morphogenesis.

In addition to cognitive and social development delays and depth of intellectual disability8,28, greater deletion sizes are consistently and proportionally positively associated with gross motor development delays, specifically and differentially from fine motor acquisitions, as well as phenotypes related to muscle tonus in the early stages of development, and feeding difficulties, and negatively associated with diagnosis and symptoms related to hyperactivity. Some of these findings had been uncovered separately before12,25,26 but not yet consistently for each gross motor milestone, as previous studies only captured age at walk12,26 or age at walk and crawl25. These features are part of the ones found in PMS patients with interstitial deletions not affecting *SHANK3*28,30, but also in a lesser extent and less severity in PMS patients with isolated *SHANK3* mutations22–24. The fact that all these tonus-related conditions start to appear even with the smallest deletions, with the severity increasing proportionally to the deletion size suggests a cumulative role of multiple genes on the long arm of chromosome 22, including *SHANK3.* The previously unreported associations of deletion sizes with gastrointestinal aspiration symptoms and apnea could also be part of this muscle tonus severity gradient.

Previous results of association with large fleshy hands and dysplastic toenails/fingernails12,25,26 could be reproduced in this study, and results of increased prevalence of lymphedema5 and aggressive behaviors26 were not replicated here.

This method of analysis has previously been considered as a proxy for the number of deleted genes and severity of the genetic condition5. However, if it can indicate the participation of other genes in the upstream part of chromosome 22 in relation to *SHANK3*, it cannot identify such genes or discard the implication of *SHANK3* in the presence of the phenotypes. When used as a predictor of developmental delays and depth of cognitive or social disabilities, it can suggest cumulative effect of the interplay of multiple gene alterations. If it does not come as a surprise that more deleted genetic material would cause more severe phenotypes, the consistency of associations only for gross motor milestones delays, and not with other groups of phenotypes such as cardiovascular, dental, endocrine, fine motor skills milestones, etc. indicates that a number of genes in this region of interest spanning 9 Mb at the end of chromosome 22 could be specifically responsible for gross motor functions and muscle tone.

Interestingly, if we may compare phenotypic behaviors between (with all the limits that are needed) mice and men, in Mei et al41 *SHANK3* KO mice show signs of decreased exploratory behaviors and social interaction, increased anxiety and signs of repetitive grooming. This is concordant with our results, as in our study, tested phenotypes related to anxiety and to decreased social interactions are all found to not be associated with greater deletion sizes, indicating the role of *SHANK3* deletion alone to be sufficient.

When used as a predictor for binomial outcomes such as the occurrence of a clearly defined somatic phenotype, use of the deletion can only act as a proxy for responsibility of chromosomal regions of interest and hint at potential breakpoints. Comparisons using individual genes could help narrow down ranges of genes that might play a role in the phenotype, but this needs specific model developments as gene statuses inferred from this data would generate strong colinearity between contiguous genes affected in the same deletion ranges.

It is to be noted that PMS being caused by 22q13 deletions, a majority of patients are tested only for genetic defects on chromosome 22, specifically using array CGH or FISH techniques and providing ranges of deletions. This is a strong bias in the data as most information are for chromosome 22 defects only, ignoring possible additional chromosomal aberrations.

A third source of data consisting of the natural language processing of the patients' clinical notes, encoded as UMLS concepts, will provide us with a range of phenotypes that were not captured by the registry questionnaires and will permit even more detailed analyses.

# Conclusions

In this study using data from the Phelan McDermid Syndrome International Registry, we proceeded to demonstrate the viability of this source of data to conduct genotype-phenotype analyses. We begin by providing a method of data curation and integration dealing with the shortcomings of the registry, and which should serve as a reference for future projects on this data. We show that deletion size is associated with gross motor development, a number of diverse conditions painting a picture of a spectrum of generalized muscle tone impairments, with novel associations previously unreported. We confirm the previous findings of association between greater deletion sizes and large fleshy hands, dysplastic toenails, sacral dimple, verbal speech ability, and socio-emotional development.

# Supplemental Data description

Supplemental Data contains two figures describing the data management workflow, one table of full results from the statistical analyses, and 31 figures detailing results of the analyses by group of features.

# Acknowledgements

The Phelan-McDermid Syndrome Foundation, the patients and their families. Chris Botka and the Harvard Medical School Research Computing center.  
This work was supported by PCORI Grand Number PPRN-1306-04814 and by Research Grand EDU\_R\_FY2015\_Q2\_HarvardMedicalSchool\_Avillach-NEW from Amazon Inc.

# Web Resources

All the code for ETL and analysis is openly available on github (<https://github.com/hms-dbmi/ETL/tree/master/data_projects/pms_dn/pms_registry/R-ETL>, <https://github.com/hms-dbmi/Avillach-Lab/tree/master/PMS_DN/GenoPheno>)

# Figures titles and legends

Figure 1: Deletions (in red), *SHANK3* mutations (in dark red), and duplications (in blue) on the long arm of chromosome 22 (as shown in Panel A), ordered by decreasing size of total deleted genetic material, one line for each patient (Panel B)

Figure 2: On the left: terminal deletions (in red) and *SHANK3* mutations (in dark red) for the included patients, ordered by decreasing size of terminal deletion size. On the right: status for each patient for every renal phenotype, lined up with the respective genetic status.

Figure 3: On the left: terminal deletions (in red) and *SHANK3* mutations (in dark red) for the included patients, ordered by decreasing size of terminal deletion size. On the right: status for each patient for every developmental gross motor phenotype, lined up with the respective genetic status.

# Tables

Sample characteristics and comparison with the excluded patients

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FALSE | NA | NA | TRUE | NA | NA | p |
|  | N=154 |  |  | N=156 |  |  |  |
|  | N | %/moy | et | N | %/moy | et |  |
| Age | 154 | 11.7 | 8.9 | 156 | 11.6 | 8.3 | 0.76 (M-W) |
| Gender |  |  |  |  |  |  | 1.00 (X) |
| Female | 82 | 26.5 |  | 83 | 26.8 |  |  |
| Male | 72 | 23.2 |  | 73 | 23.5 |  |  |
| Ancestral.Background |  |  |  |  |  |  | 0.12 (f) |
| Asian | 1 | 0.4 |  | 2 | 0.8 |  |  |
| Black (African American) | 3 | 1.2 |  | 0 | 0.0 |  |  |
| Black (Not African American) | 1 | 0.4 |  | 0 | 0.0 |  |  |
| Caucasian (Latino/Hispanic) | 17 | 6.9 |  | 18 | 7.3 |  |  |
| Caucasian (Not Latino/Hispanic) | 78 | 31.6 |  | 90 | 36.4 |  |  |
| Native American | 1 | 0.4 |  | 0 | 0.0 |  |  |
| Other | 23 | 9.3 |  | 13 | 5.3 |  |  |

Significantly associated phenotypes with largest positive effect sizes

|  |  |  |  |
| --- | --- | --- | --- |
| Phenotype | p | p.adj | OR |
| Walk unassisted | 1.22e-09 | 3.67e-07 | 5.02 [ 3.04 - 8.64 ] |
| Sit when placed | 1.76e-08 | 2.65e-06 | 3.44 [ 2.27 - 5.39 ] |
| Crawl on hands and knees | 7.96e-07 | 5.99e-05 | 3.38 [ 2.12 - 5.59 ] |
| Roll over back to stomach | 7.75e-07 | 5.99e-05 | 2.9 [ 1.93 - 4.5 ] |
| Hold head up on his/her own | 1.19e-04 | 2.30e-03 | 2.14 [ 1.46 - 3.18 ] |
| Climb stairs standing up without help | 5.78e-03 | 4.14e-02 | 2.11 [ 1.25 - 3.62 ] |
| Difficulties maintaining balance | 1.12e-04 | 2.30e-03 | 2.02 [ 1.42 - 2.9 ] |
| Abnormal Cranial CAT scan | 2.87e-04 | 4.11e-03 | 1.82 [ 1.37 - 2.66 ] |
| Hip dysplasia – developmental | 4.63e-03 | 3.58e-02 | 1.66 [ 1.22 - 2.52 ] |
| Vesicoureteral Reflux | 1.19e-04 | 2.30e-03 | 1.6 [ 1.29 - 2.1 ] |
| Required special feeds | 6.18e-05 | 1.86e-03 | 1.55 [ 1.27 - 1.95 ] |
| Left foot | 2.35e-03 | 2.21e-02 | 1.52 [ 1.19 - 2.07 ] |
| Aspiration | 2.28e-05 | 8.60e-04 | 1.5 [ 1.26 - 1.84 ] |
| Does the patient have a sacral dimple? | 3.15e-06 | 1.90e-04 | 1.45 [ 1.25 - 1.71 ] |
| Floppy baby | 9.67e-05 | 2.30e-03 | 1.45 [ 1.22 - 1.77 ] |
| Right foot | 2.00e-03 | 1.94e-02 | 1.44 [ 1.16 - 1.87 ] |
| Had swallowing problems | 1.12e-05 | 4.82e-04 | 1.41 [ 1.22 - 1.65 ] |
| Hydronephrosis | 7.25e-03 | 4.55e-02 | 1.4 [ 1.11 - 1.83 ] |
| Tongue tied | 5.72e-03 | 4.14e-02 | 1.38 [ 1.11 - 1.78 ] |
| Large fleshy hands | 4.00e-06 | 2.01e-04 | 1.37 [ 1.21 - 1.58 ] |
| Apnea | 5.30e-03 | 3.99e-02 | 1.37 [ 1.11 - 1.73 ] |
| Tooth/teeth extraction | 2.44e-04 | 3.87e-03 | 1.36 [ 1.16 - 1.61 ] |
| Hypotonia | 4.31e-05 | 1.44e-03 | 1.35 [ 1.17 - 1.56 ] |
| Has the patient ever had febrile seizures? | 3.21e-04 | 4.40e-03 | 1.34 [ 1.15 - 1.58 ] |
| Fatigues easily | 8.60e-04 | 9.61e-03 | 1.34 [ 1.14 - 1.61 ] |
| Need for IV antibiotics to clear infections | 6.91e-03 | 4.49e-02 | 1.33 [ 1.09 - 1.66 ] |
| Feeding problems | 1.22e-04 | 2.30e-03 | 1.3 [ 1.14 - 1.49 ] |
| Recurring ingrown toenails | 7.79e-04 | 9.38e-03 | 1.3 [ 1.12 - 1.53 ] |
| Had poor suck | 4.25e-04 | 5.56e-03 | 1.27 [ 1.12 - 1.46 ] |
| Dysplastic or unusual toenails | 8.62e-04 | 9.61e-03 | 1.24 [ 1.09 - 1.41 ] |
| Abnormal MRI | 3.55e-03 | 2.82e-02 | 1.24 [ 1.08 - 1.44 ] |
| Had difficulty latching on the bottle | 3.38e-03 | 2.75e-02 | 1.23 [ 1.08 - 1.43 ] |
| Dysplastic or unusual fingernails | 7.53e-03 | 4.63e-02 | 1.2 [ 1.05 - 1.38 ] |
| Had difficulty latching on the breast | 8.29e-03 | 4.99e-02 | 1.19 [ 1.05 - 1.36 ] |
| Has the patient had ear tubes? | 7.01e-03 | 4.49e-02 | 1.18 [ 1.05 - 1.33 ] |

Significantly associated phenotypes with largest negative effect sizes

|  |  |  |  |
| --- | --- | --- | --- |
| Phenotype | p | p.adj | OR |
| Act as if driven by a motor | 0.000122 | 0.00230 | 0.461 [ 0.307 - 0.678 ] |
| Plays alongside others but not with them | 0.000216 | 0.00362 | 0.483 [ 0.324 - 0.704 ] |
| Run around or climb excessively when inappropriate | 0.002770 | 0.02320 | 0.529 [ 0.346 - 0.8 ] |
| Verbal speech | 0.001100 | 0.01140 | 0.536 [ 0.366 - 0.776 ] |
| Engage in pretend play | 0.002450 | 0.02230 | 0.551 [ 0.37 - 0.803 ] |
| Respond to others – emotions | 0.006590 | 0.04410 | 0.602 [ 0.413 - 0.862 ] |
| ADHD | 0.000969 | 0.01040 | 0.658 [ 0.496 - 0.823 ] |
| Pervasive Developmental Disorder | 0.000461 | 0.00578 | 0.7 [ 0.563 - 0.843 ] |
| Understanding of the use of familiar objects | 0.000286 | 0.00411 | 0.757 [ 0.646 - 0.874 ] |
| Symptoms/diagnostic of ADD/ADHD | 0.001560 | 0.01560 | 0.776 [ 0.657 - 0.901 ] |
| Exhibits sensory seeking behaviors | 0.005950 | 0.04170 | 0.795 [ 0.671 - 0.931 ] |
| Autism | 0.002580 | 0.02280 | 0.808 [ 0.699 - 0.924 ] |
| Imitate household activities during play | 0.002770 | 0.02320 | 0.81 [ 0.701 - 0.926 ] |
| Difficulty going back to sleep after nighttime awakening | 0.006520 | 0.04410 | 0.843 [ 0.743 - 0.951 ] |
| 5 minutes | 0.000136 | 0.00241 | 0.845 [ 0.778 - 0.918 ] |

# References

1. Phelan, K., and Rogers, R.C. (1993). Phelan-McDermid syndrome. In GeneReviews(®), R.A. Pagon, M.P. Adam, H.H. Ardinger, T.D. Bird, C.R. Dolan, C.-T. Fong, R.J. Smith, and K. Stephens, eds. (Seattle (WA): University of Washington, Seattle),

2. Phelan, K., and McDermid, H.E. (2012). The 22q13.3 deletion syndrome (phelan-McDermid syndrome). Mol Syndromol *2*, 186–201.

3. Kolevzon, A., Angarita, B., Bush, L., Wang, A.T., Frank, Y., Yang, A., Rapaport, R., Saland, J., Srivastava, S., Farrell, C., et al. (2014). Phelan-McDermid syndrome: A review of the literature and practice parameters for medical assessment and monitoring. Journal of Neurodevelopmental Disorders *6*, 39.

4. Costales, J.L., and Kolevzon, A. (2015). Phelan–McDermid syndrome and SHANK3: Implications for treatment. Neurotherapeutics 1–11.

5. 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.

6. Phelan, M.C. (2008). Deletion 22q13.3 syndrome. Orphanet Journal of Rare Diseases *3*, 14.

7. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsäter, H., et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet *39*, 25–27.

8. 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.

9. Gong, X., Jiang, Y.-w., Zhang, X., An, Y., Zhang, J., Wu, Y., Wang, J., Sun, Y., Liu, Y., Gao, X., et al. (2012). High proportion of 22q13 deletions and SHANK3 mutations in chinese patients with intellectual disability. PLoS ONE *7*, e34739.

10. Betancur, C., and Buxbaum, J.D. (2013). SHANK3 haploinsufficiency: A “common” but underdiagnosed highly penetrant monogenic cause of autism spectrum disorders. Molecular Autism *4*, 17.

11. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846.

12. Sarasua, S.M., Boccuto, L., Sharp, J.L., Dwivedi, A., Chen, C.-F., Rollins, J.D., Rogers, R.C., Phelan, K., and DuPont, B.R. (2014). Clinical and genomic evaluation of 201 patients with phelan–McDermid syndrome. Hum Genet *133*, 847–859.

13. Shcheglovitov, A., Shcheglovitova, O., Yazawa, M., Portmann, T., Shu, R., Sebastiano, V., Krawisz, A., Froehlich, W., Bernstein, J.A., Hallmayer, J.F., et al. (2013). SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. Nature *503*, 267–271.

14. Kolevzon, A., Bush, L., Wang, A.T., Halpern, D., Frank, Y., Grodberg, D., Rapaport, R., Tavassoli, T., Chaplin, W., Soorya, L., et al. (2014). A pilot controlled trial of insulin-like growth factor-1 in children with phelan-McDermid syndrome. Mol Autism *5*, 54.

15. Shao, S., Xu, S., Yang, J., Zhang, T., He, Z., Sun, Z., and Song, R. (2014). A commonly carried genetic variant, rs9616915, in SHANK3 gene is associated with a reduced risk of autism spectrum disorder: Replication in a chinese population. Mol. Biol. Rep. *41*, 1591–1595.

16. Moessner, R., Marshall, C.R., Sutcliffe, J.S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., et al. (2007). Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet *81*, 1289–1297.

17. 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.

18. Uchino, S., and Waga, C. (2013). SHANK3 as an autism spectrum disorder-associated gene. Brain Dev. *35*, 106–110.

19. Herbert, M.R. (2011). SHANK3, the synapse, and autism. N. Engl. J. Med. *365*, 173–175.

20. 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.

21. Bozdagi, O., Sakurai, T., Papapetrou, D., Wang, X., Dickstein, D.L., Takahashi, N., Kajiwara, Y., Yang, M., Katz, A.M., Scattoni, M.L., et al. (2010). Haploinsufficiency of the autism-associated shank3 gene leads to deficits in synaptic function, social interaction, and social communication. Molecular Autism *1*, 15.

22. Macedoni-Lukšič, M., Krgović, D., Zagradišnik, B., and Kokalj-Vokač, N. (2013). Deletion of the last exon of SHANK3 gene produces the full phelan–McDermid phenotype: A case report. Gene *524*, 386–389.

23. Bonaglia, M.C., Giorda, R., Borgatti, R., Felisari, G., Gagliardi, C., Selicorni, A., and Zuffardi, O. (2001). Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome. The American Journal of Human Genetics *69*, 261–268.

24. Delahaye, A., Toutain, A., Aboura, A., Dupont, C., Tabet, A.C., Benzacken, B., Elion, J., Verloes, A., Pipiras, E., and Drunat, S. (2009). Chromosome 22q13.3 deletion syndrome with a de novo interstitial 22q13.3 cryptic deletion disrupting SHANK3. European Journal of Medical Genetics *52*, 328–332.

25. Sarasua, S.M., Dwivedi, A., Boccuto, L., Rollins, J.D., Chen, C.-F., Rogers, R.C., Phelan, K., DuPont, B.R., and Collins, J.S. (2011). Association between deletion size and important phenotypes expands the genomic region of interest in phelan–McDermid syndrome (22q13 deletion syndrome). J Med Genet *48*, 761–766.

26. 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). 22q13.2q13.32 genomic regions associated with severity of speech delay, developmental delay, and physical features in phelan-McDermid syndrome. Genet Med *16*, 318–328.

27. Fujita, Y., Mochizuki, D., Mori, Y., Nakamoto, N., Kobayashi, M., Omi, K., Kodama, H., Yanagawa, Y., Abe, T., Tsuzuku, T., et al. (2000). Girl with accelerated growth, hearing loss, inner ear anomalies, delayed myelination of the brain, and del(22)(q13.1q13.2). Am. J. Med. Genet. *92*, 195–199.

28. Wilson, H.L., Crolla, J.A., Walker, D., Artifoni, L., Dallapiccola, B., Takano, T., Vasudevan, P., Huang, S., Maloney, V., Yobb, T., et al. (2008). Interstitial 22q13 deletions: Genes other than SHANK3 have major effects on cognitive and language development. Eur. J. Hum. Genet. *16*, 1301–1310.

29. Simenson, K., Õiglane-Shlik, E., Teek, R., Kuuse, K., and Õunap, K. (2014). 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. *164*, 806–809.

30. Vittoria Disciglio, C.L.R. (2014). Interstitial 22q13 deletions not involving SHANK3 gene: A new contiguous gene syndrome. American Journal of Medical Genetics Part A.

31. Maitz, S., Gentilin, B., Colli, A.M., Rizzuti, T., Brandolisio, E., Vetro, A., Zuffardi, O., Guerneri, S., and Lalatta, F. (2008). Expanding the phenotype of 22q13.3 deletion: Report of a case detected prenatally. Prenat. Diagn. *28*, 978–980.

32. Phelan-McDermid syndrome international registry.

33. International Standing Committee on Human Cytogenetic Nomenclature (2013). ISCN 2013 : An international system for human cytogenetic nomenclature (2013) (Basel ; Unionvillle, CT: Karger).

34. Athey, B.D., Braxenthaler, M., Haas, M., and Guo, Y. (2013). tranSMART: An open source and community-driven informatics and data sharing platform for clinical and translational research. AMIA Jt Summits Transl Sci Proc *2013*, 6–8.

35. Canuel, V., Rance, B., Avillach, P., Degoulet, P., and Burgun, A. (2015). Translational research platforms integrating clinical and omics data: A review of publicly available solutions. Brief. Bioinformatics *16*, 280–290.

36. Murphy, S.N., Weber, G., Mendis, M., Gainer, V., Chueh, H.C., Churchill, S., and Kohane, I. (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc *17*, 124–130.

37. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC genome browser database: 2015 update. Nucleic Acids Res. *43*, D670–681.

38. Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics *26*, 1205–1210.

39. Valenta, Z., Pitha, J., and Poledne, R. (2006). Proportional odds logistic regression–effective means of dealing with limited uncertainty in dichotomizing clinical outcomes. Stat Med *25*, 4227–4234.

40. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289–300.

41. Mei, Y., Monteiro, P., Zhou, Y., Kim, J.-A., Gao, X., Fu, Z., and Feng, G. (2016). Adult restoration of shank3 expression rescues selective autistic-like phenotypes. Nature *530*, 481–484.