Speech and language deficits are central to *SETBP1* haploinsufficiency disorder

Angela Morgan*[1-4], Ruth Braden*[1,2], Maggie M. K. Wong[5], Estelle Colin[6], David Amor[1,3,4,], Frederique Liegeois[7], Siddharth Srivastava[8], Adam Vogel[9], Varoona Bizaoui[10], Kara Ranguin[10], Simon E. Fisher[5,11], Bregje W. van Bon[12]

Affiliations

1. Speech & Language, Murdoch Children's Research Institute, Melbourne, Australia

2. Department of Paediatrics & Department of Audiology & Speech Pathology, University of Melbourne, Australia

3. Royal Children's Hospital, Melbourne, Australia

4. Victorian Clinical Genetics Service, Melbourne, Australia

5. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

6. Service de Génétique Médicale, Centre Hospitalier Universitaire d'Angers, Angers, France

7. UCL Great Ormond Street Institute of Child Health, London, UK

8. Boston Children's, Harvard Medical Centre, United States

9. Centre for Neuroscience of Speech, Department of Audiology and Speech Pathology, University of Melbourne

10. Service de Génétique, Centre Hospitalier Universitaire Caen Normandie, France

11. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

12. Radboud University Medical centre, Nijmegen, The Netherlands

*These authors contributed equally to this work.

Corresponding author:

Angela Morgan PhD

Head, Speech and Language

Murdoch Children's Research Institute

50 Flemington Rd, Parkville, 3052

Phone: +613 83416458

Email: angela.morgan@mcri.edu.au

Abstract

Expressive communication impairment is associated with haploinsufficiency of *SETBP1*, as reported in small case series. Heterozygous pathogenic loss of function (LoF) variants in *SETBP1* have also been identified in independent cohorts ascertained for childhood apraxia of speech (CAS), warranting further investigation of the roles of this gene in speech development. Thirty-one participants (20 males, aged 0;8 to 23;2 years, 28 with pathogenic *SETBP1* LoF variants, 3 with 18q12.3 deletions) were assessed for speech, language and literacy abilities. Broader development was examined with standardised motor, social, and daily life skills assessments. Gross and fine motor deficits (94%) and intellectual impairments (68%) were common. Protracted and aberrant speech development was consistently seen, regardless of motor or intellectual ability. We expand the linguistic phenotype associated with *SETBP1*-LoF syndrome (SETBP1 haploinsufficiency disorder), revealing a striking presentation that implicates both motor (CAS, dysarthria) and language (phonological errors) systems, with CAS (80%) being the most common diagnosis. In contrast to past reports, the understanding of language was rarely better preserved than language expression (29%). Language was typically low, to moderately impaired, with commensurate expression and comprehension ability. Children were sociable with a strong desire to communicate. Minimally verbal children (32%) augmented speech with sign language, gestures or digital devices. Overall, relative to general development, spoken language and literacy were poorer than social, daily living, motor and adaptive behaviour skills. Our findings show that poor communication is a central feature of *SETBP1* haploinsufficiency disorder, confirming this gene as a strong candidate for speech and language disorders.

1. Introduction

Clinical investigations of individuals with disruptions of SET binding protein 1 (*SETBP1*) on 18q12.3 have suggested the gene as a candidate for expressive speech disorder (1). Haploinsufficiency of *SETBP1* was first associated with expressive language 'delay' or 'impairment' in descriptive single cases (1, 2) and case series (3). Other presenting features were mild to severely impaired intellect, gross and fine motor delays and/or deficits, hypotonia, distinctive facial features, attention deficits, and less commonly, autistic traits (1, 3).

A novel disorder encompassing this symptomatology was later confirmed in a cohort selected for intellectual disability (ID) (4). Five individuals with LoF variants and one with a *de novo* deletion encompassing *SETBP1* were identified (4). The authors combined these cases with 2 further novel cases with truncating variants from a separate ID screen, together with the previously published small deletions (1, 3) and *de novo* variants (2) to examine the phenotype across all the identified individuals. Only retrospective clinical data were available, limiting the scope of the investigation. Most of the individuals were reported to have intelligence quotient (IQ) and language deficits, with completely absent or significantly impaired speech in 92% of them (4). The nature of the speech deficits was not described, although apraxia was noted previously in one of the cases (1). The specific speech and language phenotype in individuals with *SETBP1* LoF variants remains unclear.

From a clinical genetics perspective, there is a need to identify genes that contribute to severe and persistent communication deficits, such as childhood apraxia of speech (CAS). Parental concern for speech development is a common reason for referral to paediatricians, yet the aetiology and prognosis for CAS is poorly understood and children are largely managed with a 'watch and wait' approach (5). Until recently, few additional candidates for CAS had

been revealed, since the unearthing of *FOXP2* almost 20 years ago (6). A number of new contenders have now been identified through next-generation sequencing screens of genomes/exomes in two cohorts ascertained on the basis of CAS (7, 8). Notably, whilst modest in cohort size (n=18 and n=34 respectively), each study independently identified an individual with a heterozygous pathogenic *SETBP1* LoF variant, suggesting disruptions of this gene as a recurrent cause for CAS; which occurs at a rate of only 1 or 2 cases per 1000 in the general population (9). The findings add weight to the premise that *SETBP1* may play an important role in speech and language development.

Two studies examining common single nucleotide polymorphisms (SNPs) further support the potential relevance of *SETBP1* variation for communication abilities. Associations between *SETBP1* and scores on a test examining syntactic complexity (mean length of sentences and use of complex sentence structures) were reported in a Genome Wide Association Study of developmental language disorder in a geographically isolated Russian cohort aged 3 to 18 years (10). More tentatively, SNPs in *SETBP1* have been reported to show association with phonological working memory; just one of many reading-related traits examined across a reading-impaired cohort of modest size for complex trait analyses (n=135) (11). Whether *SETBP1* is more closely linked with variations in speech, language and/or reading ability in the general population is not yet clear.

Thus, nascent evidence from a range of sources suggests *SETBP1* as a gene of relevance to speech and language development. Yet information concerning the clinical phenotype associated with *SETBP1* haploinsufficiency has so far been drawn only from descriptive case series, relying largely on retrospective examination of medical records. Here, we performed in-depth examination of speech, language and literacy abilities in a cohort of 31 individuals

with *SETBP1* LoF variants and deletions, using standardised tests, to precisely characterise the communication phenotype of this syndrome. Linguistic performance was considered relative to other areas of neurodevelopment (eg. motor abilities, social skills), to determine whether communication was differentially affected.

Methods

Inclusion criteria were a molecular diagnosis of heterozygous *SETBP1* truncating (stop-gain or frameshift) variants or 18q12.3 deletions in individuals aged ≥6 months. Participants were recruited globally via the *SETBP1* Society (http://www.setbp1.org) and clinician referral. Ethics approval was obtained from the Royal Children's Hospital, Melbourne, Human Research Ethics Committee (HREC 37353A).

Thirty one participants were recruited (Table 1; 20 males, average age 9 years, range 0;8 to 23;2 years). Genotypes included single nucleotide variants (28 participants), a 18q12.3 intragenic deletion and two larger deletions encompassing *SETBP1* (Figure 1). Twenty four participants were novel cases, not previously reported in the literature. Nineteen participants also participated in a separate study on the broader medical phenotype (see companion manuscript Jansen et al). Deletions and phenotypic data were submitted to Decipher (https://decipher.sanger.ac.uk/) and sequence variants were submitted to Leiden Open Variation Database (Database ID: #chr18_002464-002468, #SETBP1_000018-000020, #SETBP1_000033, #SETBP1_000078, #SETBP1_000083, #SETBP1_000085, #SETBP1_000103, #SETBP1_000106, #SETBP1_000108-000111, #SETBP1_00014, #SETBP1_000116-000117, #SETBP1_000119-000120, #SETBP1_000123-000125, #SETBP1_000127, #SETBP1_000129).

Health and development

Health and medical information, including data on neurodevelopmental conditions, intellectual ability, and intervention (eg. speech therapy, physiotherapy), was collected via an established survey (12, 13) (Table 1), translated into multiple languages. Health professional reports (eg. psychology, neurology, clinical genetic, speech pathology) and telehealth consults confirmed questionnaire responses. Feeding (Child Oral and Motor Proficiency Scale (ChOMPS) (14) and drooling (Drooling Impact Scale (DIS) (15) measures were collected where age appropriate.

Speech

Verbal children completed the Diagnostic Evaluation of Articulation and Phonology to examine articulation and phonological errors (16). A 5-minute speech sample was analysed for childhood apraxia of speech (CAS) (17) and dysarthria (18, 19). The Intelligibility in Context Scale (20) examined how often the individual is understood, with a 5 point scale of responses ranging from never to always.

Language

Verbal children were assessed with the Children's Communication Checklist (CCC-2) (21). Minimally verbal children (defined as less than 50 spoken words), and those less than four years of age were assessed with the Macarthur Bates Communicative Development Inventory (MB-CDI) (22) and Communication and Symbolic Behaviour Scales Developmental Profile (CSBS-DP) (23). The MB-CDI measures understanding and use of gesture, vocabulary and sentences. The CSBS-DP provides social communication, speech and symbolic communication scores for children aged 6 to 24 months, or for chronologically older children with limited linguistic abilities (23).

Adaptive behaviour

The Vineland Adaptive Behaviour Scales- Parent/Caregiver (24), provided domain scores for communication, socialisation, daily living and motor skills, with an overall adaptive behaviour composite. Wilcoxon signed-rank sum tests were used to determine the relative involvement of language compared to other Vineland subdomain scores.

Results

Health and development

The cohort consisted of one infant, seven preschool children, 22 school-aged children and adolescents and two adults (*n*=31 (20 males, 11 females); Table 1). Participants were recruited from the US (n=12), Netherlands (n=8), United Kingdom (n=3), France (n=3), Canada (n=2), Israel (n=2) and Australia (n=1). Developmental issues relevant to speech and language development included feeding difficulties (35%), and excessive drooling (35%). Almost all participants (94%) had generalised motor delay or disorder that required occupational therapy and/or physiotherapy (87%; Table 1; Figure 2). Motor deficits included difficulties with personal care (managing buttons and zippers, teeth brushing, washing), writing, drawing, using scissors, riding a bike and toilet training. Intellectual impairment was reported in the majority (22/23, 96%) of individuals aged >4 years. In the eight individuals aged <4 years, seven parents reported their child was experiencing developmental delay compared to same-aged peers. Seven patients had seizures; four had febrile seizures, one had generalised tonic-clonic seizures, one had absence seizures and one reported a history of seizures despite a normal EEG. Hearing impairment was infrequent (10%) and all presentations were mild (25-39 dBHL) and bilateral, with two cases of mixed (IND 8, 12) and one of conductive (IND 27) hearing loss, although periodic conductive losses due to otitis media were also common (58%). Visual impairments (42%) were addressed with glasses and hypermetropia was the most common

diagnosis (62%). Palatal abnormalities included cleft lip and palate (IND 26), submucous cleft palate (IND 23) and a high arch palate (IND 31). Micrognathia (10%) was noted in few participants.

Attention issues were common (55%; including nine with formal ADHD diagnoses, see Figure 2). All but the 8-month old participant had been directly assessed for ASD, but only three (10%) received a clinical ASD diagnosis. Other diagnoses included developmental coordination disorder (19%) and sensory processing disorder (23%).

Academically, eight of the 21 (38%) school-aged participants attended mainstream schools and the remaining 13 (62%) attended special education schools. Of the eight preschool participants, three (38%) attended mainstream, and four (50%) specialised, childcare or pre-school settings, and one (13%) was cared for at home. Learning support was common (86%) across all settings. The two young adults had completed school and were engaged in supported employment. Most parents of school-aged children and young adults (21/23, 91%) reported that their child's academic progress had been limited by their speech and language difficulties.

Speech

Speech development was characterised by limited babbling and a reduced phonetic (sound) inventory relative to peers across the first seven years of life; when a full inventory is typically acquired. Most participants had acquired first spoken words by 18 months of age (52%) (Table 2). For the majority of the verbal children, short sentences were developed by six to seven years (protracted relative to typical developmental milestone of 2-3 years) (Table 2). Most (94%) had accessed speech therapy, with the exception of two young participants (aged eight months and one year three months). The dose of speech therapy increased during the preschool period;

typically once per week/fortnight, but up to five times per week where available. Intelligibility across the group ranged from never understood (32%), to rarely (23%), sometimes (61%) and usually understood (13%), based on the Intelligibility in Context Scale (20) scores.

Verbal children presented with a complex motor speech disorder, best characterised as Childhood Apraxia of Speech (CAS) (80%; Table 2) with inconsistency of production, increased errors with increasing word length; simplified syllable structures relative to age, vowel and prosodic errors. A small proportion (16%) had dysarthria, typically characterised by low pitch, hypernasality, monotonous, monoloud and flaccid, slow speech. Other speech diagnoses of phonological disorder (48%), articulation impairment (lisp) (9%), and stuttering (3%) were reported alongside CAS.

Minimally-verbal children (11/31, 35%; Table 2) had few spoken words but had communicative intent, and used gesture, sign and/or communication devices for expression. Of this group, six were young children (aged 3 to 5 years of age) and five were older (aged 7 to 15 years). Speech intervention for the younger group focused on language stimulation, non-verbal gestures and verbal speech production. The older group were producing single words or short phrases using their digital devices. Relative strengths in social and symbolic language abilities (average standard scores 9.29 and 9.57, where mean = 10, standard deviation =3) relative to speech (average composite standard score 6.57) were revealed on the CSBS.

Language (Expressive, Receptive, Written, Social)

For most participants expressive and receptive language abilities were commensurate with each other (18/28; 64%) on the VABS. Poorer expressive than receptive performance was the next most common profile (8/28; 29%). For children of reading age, written language ranged from typical (2/24; 8%) to moderately low (8/24; 33%) and low (14/24; 58%). Many had difficulty with writing tasks, such as copying letters or their name, although a few older patients were

able to write in longer sentences. A large proportion (10/23; 43%) had received a formal diagnosis of a reading and/or writing disorder from a health professional.

The CCC-2 enabled further comparison within and across general communication (eg. semantics, syntax, coherence) and social interaction domains. All children assessed with the CCC-2 (n=16), had poor communication abilities (Table 3). Pragmatic language and social skills were relative strengths overall, compared to speech, language structure, vocabulary and discourse, based on individual scaled scores (Figure 3). Autistic traits were reported in half this group (8/16, 50%), including poor social skills and restricted interests compared to peers. Yet only three had a clinical diagnosis of ASD, as noted earlier (Table 1). Participants showed a desire to communicate and share interests, with intact basic social skills and non-verbal gestures. Whilst data are limited, there seemed to be a widening gap in social skills relative to peers with increasing age (Table 2).

Adaptive behaviour: Language relative to daily functioning, social and motor skills

Overall adaptive behaviour scores were in the moderately low range (Table 4). This was commensurate with daily living skills and socialisation (Table 4, Figure 4). Motor skills were stronger than communication abilities (p=0.0021), although these data represent participants aged <9;11 only, with information about normative motor skills unavailable for older children. Fine motor skills were poorer relative to gross motor in most cases, confirming parent reported motor abilities from the questionnaire data. Performance in the communication domain was substantially lower than that for socialisation (p=0.0055) and daily living skills (p=0.0023) domains (Table 4).

Discussion

Here we report the speech and language phenotype of individuals selected for pathogenic *SETBP1* LoF variants. Our findings indicate that articulatory, spoken and written language (reading, writing) deficits are distinctive features of the broader neurodevelopmental profile. We expand the phenotype of this disorder beyond 'expressive speech' difficulties, to reveal specific sub-types of speech disorder and highlight difficulties with the understanding as well as expression of language.

Protracted and impoverished speech, language and literacy (reading, writing) development was seen across the group, regardless of cognitive ability. Although first words were developed at the typical 12-month milestone for some, the ongoing trajectory of linguistic development was markedly protracted. Verbal children displayed a complex range of speech diagnoses implicating perturbed motor (CAS, articulation impairment, dysarthria) and linguistic deficits (phonological errors) that have not previously been recognized as features of *SETBP1* haploinsufficiency disorder. Yet CAS was the most common finding in our cohort, in line with the recent identification of pathogenic *SETBP1* LoF variants in gene discovery cohorts ascertained for CAS (7, 8). The co-morbid articulation, phonological and dysarthric impairments seen alongside CAS were more notable in older children in the middle school years. Changes in speech profile across the lifespan are recognised in other neurogenetic conditions (8, 12, 25) and confirm the need for regular speech surveillance to enable precisely targeted therapies at particular ages.

A subset of children remained minimally verbal at ages 7 to 15 years. How to extricate the relative cognitive-linguistic from motor contributions in children with minimally verbal presentations is an area of ongoing debate in other neurodevelopmental conditions such as

autism (26), and no simple algorithm is available. All minimally verbal children here had communicative intent and used augmentative approaches alongside speech, such as sign language, gesture or digital devices to convey messages. This strong desire to communicate was also reported in the histories of children who became verbal, showing little differentiation between verbal and minimally verbal speakers in this regard. One hypothesis to explain the minimally verbal presentation in some is that they may have more severe involvement on the speech-motor continuum, described as anarthria and/or significant speech praxis. Early speech intervention appears to be critical for all with *SETBP1* LoF variants who present with severe speech disorder, with best evidenced approaches for speech apraxia known to involve intensive therapy as often as four sessions per week (27). Future clinical trials of intensive speech therapies in individuals with *SETBP1* LoF variants are warranted.

In terms of language performance; previous case descriptions of children with *SETBP1* LoF variants have implied that language comprehension is more intact than language production (1, 3). Yet administration of standardised language tests in our cohort revealed that understanding of language is largely commensurate with expression. This highlights potential for clinical bias in making subjective assessments of language comprehension in children with speech production disorders. Further, language deficits appeared ubiquitous without clear disparity across subdomains of vocabulary, syntax and coherence. Similarly, there was corresponding involvement of written (reading, spelling) and spoken language, without clear dissociations between these skills. Spoken and written language abilities were in turn, generally commensurate with cognitive abilities.

In terms of broader neurodevelopmental profile, clinical reports of gross and fine motor deficits affecting motor planning, programming and execution occurred with equivalent

frequency to the speech-motor deficits seen here. Attention deficits and cognitive impairment were also prevalent. These are recognisable features previously reported as concomitant with CAS (8).

Differentiating severe communication deficits from ASD can be challenging (28) and for some individuals, the negative cycle of communication breakdown leads to further social withdrawal over time (29, 30). A number of children in our cohort had 'autistic features' represented by limited social skills and restricted interests relative to peers, yet only three had a formal ASD diagnosis. Further, concern over limited speech development, rather than autistic features, was the core presenting concern for parents, and all had a strong desire to communicate, despite their recognised social skill deficits. Overall, we found limited evidence for a distinct ASD signature associated with *SETBP1* LoF variants.

Clinical implications

We show that aberrant communication development is a central feature of the *SETBP1*-LoF syndrome. Children with heterozygous pathogenic *SETBP1* LoF variants or deletions should be enrolled in speech therapy in the first year of life. Given the markedly delayed verbal communication trajectory, multi-modal communication, such as sign language or communication devices would support language acquisition prior to speech developing. The complex and widespread linguistic deficits signal that children will need speech motor therapies to develop verbal speech, but also phonological interventions focused on early literacy awareness and approaches targeting language comprehension as well as production. Whilst children demonstrate a strong desire to communicate, social skills warrant therapeutic attention. Given this pervasive communication profile, we confirm *SETBP1* as a strong candidate for speech and language disorders.

References

1.   Marseglia G, Scordo MR, Pescucci C, Nannetti G, Biagini E, Scandurra V, et al. 372 kb microdeletion in 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment. Eur J Med Genet. 2012; 55: 216-221.

2.   Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. The Lancet. 2012; 380: 1674-1682.

3.   Filges I, Shimojima K, Okamoto N, Röthlisberger B, Weber P, Huber AR, et al. Reduced expression by SETBP1 haploinsufficiency causes developmental and expressive language delay indicating a phenotype distinct from Schinzel-Giedion syndrome. J Med Genet. 2011; 48: 117-122.

4.   Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014; 46: 1063-1071.

5.   Morgan AT, Eecen KT, Pezic A, Brommeyer K, Mei C, Eadie P, et al. Who to refer for speech therapy at 4 years of age versus who to "watch and wait"? J Pediatr. 2017; 185: 200-204.

6.   Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001; 413: 519-523.

7.   Eising E, Carrion-Castillo A, Vino A, Strand EA, Jakielski KJ, Scerri TS, et al. A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. Mol Psychiatry. 2018; 1065-1078.

8.   Hildebrand MS, Jackson VE, Scerri TS, Van Reyk O, Coleman M, Braden RO, et al. Severe childhood speech disorder: Gene discovery highlights transcriptional

dysregulation. Neurology. 2020; 94: e2148-e67. doi: 10.1212/WNL.0000000000009441

9. Shriberg LD. Five subtypes of developmental phonological disorders. Clin Commun Disord. 1994; 4: 38-53.

10. Kornilov SA, Rakhlin N, Koposov R, Lee M, Yrigollen C, Caglayan AO, et al. Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. Pediatrics. 2016; 137: e20152469. doi: 10.1542/peds.2015-2469.

11. Perdue MV, Mascheretti S, Kornilov SA, Jasińska KK, Ryherd K, Mencl WE, et al. Common variation within the SETBP1 gene is associated with reading-related skills and patterns of functional neural activation. Neuropsychologia. 2019; 130: 44-51.

12. Morgan AT, van Haaften L, van Hulst K, Edley C, Mei C, Tan TY, et al. Early speech development in Koolen de Vries syndrome limited by oral praxis and hypotonia. Eur J Hum Genet. 2018; 26: 75-84.

13. Mei C, Fedorenko E, Amor DJ, Boys A, Hoeflin C, Carew P, et al. Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. Eur J Hum Genet. 2018; 26: 676-686.

14. Pados BF, Thoyre SM, Park J. Age-based norm-reference values for the Child Oral and Motor Proficiency Scale. Acta Paediatrica. 2018; 107: 1427-1432.

15. Reid SM, Johnson HM, Reddihough DS. The Drooling Impact Scale: A measure of the impact of drooling in children with developmental disabilities. Dev Med Child Neurol. 2010; 52: e23-e8.

16. Dodd, B. H., Z. Crosbie, S. Holm, A. Ozanne, A. *Diagnostic Evaluation of Articulation and Phonology*. (Psychological Corporation, London, 2002)

17. American-Speech-Language-Hearing-Association. Childhood apraxia of speech. 2007 [Available from: http://www.asha.org]

18. Duffy JR. Motor Speech Disorders: Substrates, Differential Diagnosis and Management. 3rd ed. (St. Louis, Mosby, 2013)

19. Morgan AT, Masterton R, Pigdon L, Connelly A, Liegeois FJ. Functional magnetic resonance imaging of chronic dysarthric speech after childhood brain injury: reliance on a left-hemisphere compensatory network. Brain. 2013; 136(Pt 2): 646-657.

20. McLeod S, Crowe K, Shahaeian A. Intelligibility in Context Scale: Normative and validation data for English-speaking preschoolers. Language, speech, and hearing services in schools. 2015; 46: 266-276.

21. Bishop DV. Children's Communication Checklist, Second Edition. 2nd edn. (Pearson, London, 2003).

22. Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick S, Bates E. The MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual, 2nd edn. (Brookes, Baltimore, 2006).

23. Wetherby AM, Prizant BM. Communication and Symbolic Behavior Scales Developmental Profile™ Infant-Toddler Checklist. (Brookes, Baltimore, 2002).

24. Sparrow SS, Cicchetti DV, Saulnier CA. Vineland Adaptive Behaviour Scales, Third Edition. 3rd edn. (Pearson, Bloomington, 2016).

25. Morgan AT, Fisher SE, Scheffer IE, Hildebrand MS. *FOXP2*-Related Speech and Language Disorders. In: GeneReviews(R) [Internet]. (University of Washington, Seattle, 2017). GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.

26. Chenausky KV, Brignell A, Morgan A, Gagne D, Norton A, Tager-Flusberg H, et al. Factor analysis of signs of childhood apraxia of speech. J Commun Disord. 2020: 106033. doi.org/10.1016/j.jcomdis.2020.106033.

27.   Morgan AT, Murray E, Liegeois FJ. Interventions for childhood apraxia of speech. Cochrane Database Syst Rev. 2018; 5: CD006278. doi: 10.1002/14651858.CD006278.pub3

28.   Dereu M, Roeyers H, Raymaekers R, Meirsschaut M, Warreyn P. How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills, and symptom severity in children with a false positive screen for autism spectrum disorder. Eur Child Adolesc Psychiatry. 2012; 21: 541-551.

29.   Fujiki M, Brinton B, Hart CH, Olsen J, Coombs M. Using measurement invariance to study social withdrawal in children with developmental language disorders. Language, speech, and hearing services in schools. 2019; 50: 253-266.

30.   Durkin K, Toseeb U, Botting N, Pickles A, Conti-Ramsden G. Social confidence in early adulthood among young people with and without a history of language impairment. J Speech Lang Hear Res. 2017; 60: 1635-1647.

31.   Coccaro N, Tota G, Zagaria A, Anelli L, Specchia G, Albano F. SETBP1 dysregulation in congenital disorders and myeloid neoplasms. Oncotarget. 2017; 8: 51920-51935.

32.   Minakuchi M, Kakazu N, Gorrin-Rivas MJ, Abe T, Copeland TD, Ueda K et al. Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. Eur J Biochem. 2001; 268: 1340-1351.

33.   Piazza R, Magistroni V, Redaelli S, Mauri M, Massimino L, Sessa A et al. SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. Nat Commun. 2018; 9: 2192.

34.   Piazza R, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. Nat Genet. 2013; 45: 18-24.

Figure Legends

Figure 1. Location of truncating variants in relation to SETBP1 protein.

Figure 2. Radar plot showing core neurodevelopmental features (n=29)^

Figure 3. Performance across sub-domains of language in the Children's Communication Checklist (n=16)^

Figure 4. Performance across domains of the Vineland Adaptive Behaviour Scale-3 (n=27).

Figure 1. Location of truncating variants in relation to SETBP1 protein.
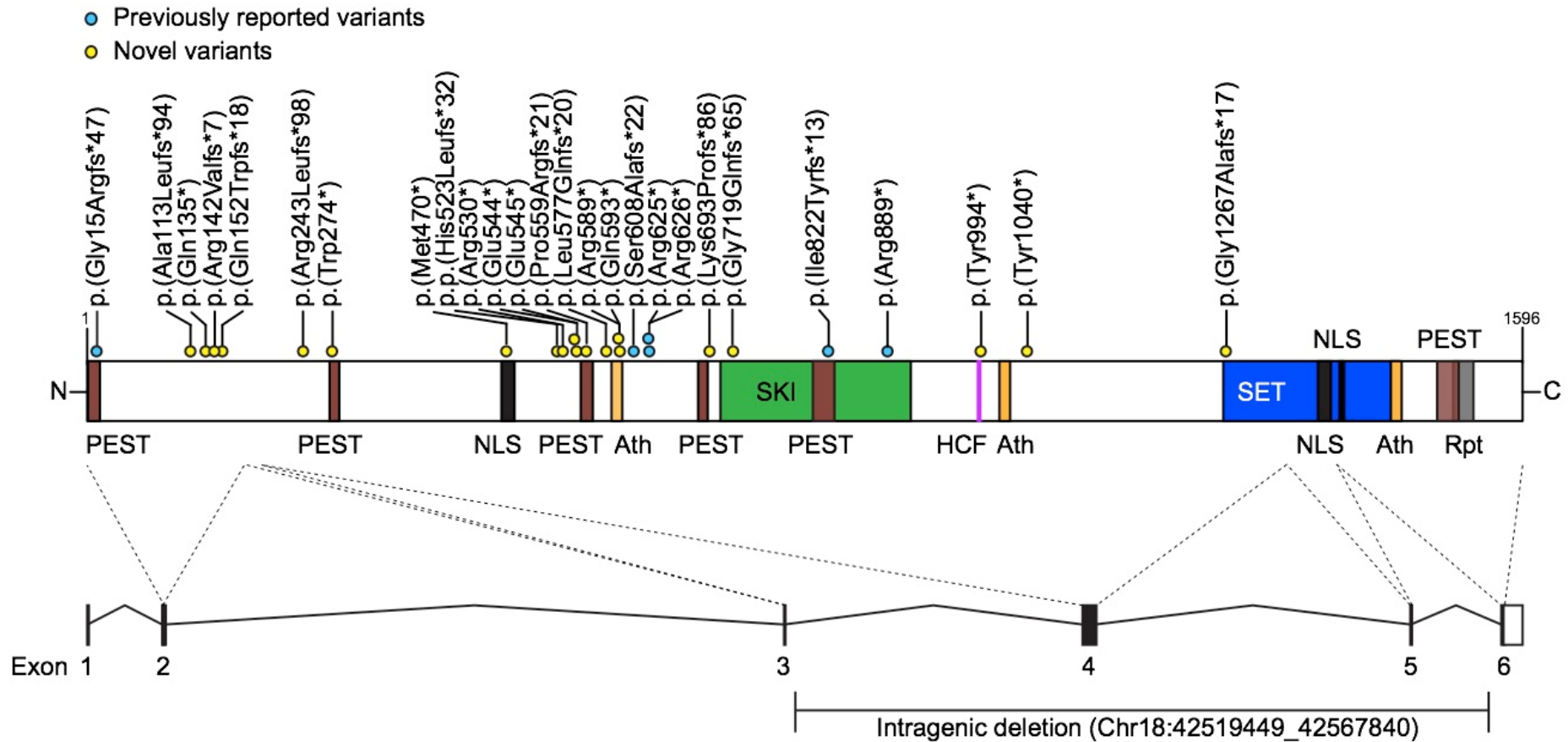
Schematic representation of the SETBP1 protein (UniProt: Q9Y6X0) indicating loss-of-function variants included in this study. Five exons (black bars) encode isoform A of the protein (1,596 amino acids). Five exons (black bars) encode isoform A of the protein (1,596 amino acid residues).

The SETBP1 protein sequence contains three AT hook domains (Ath; orange; amino acids 584–596, 1,016–1,028, 1,451–1,463), a SKI homologous region (SKI; green; amino acids 706–917), a HCF1-binding motif (HCF; magenta; amino acids 991–994), a SET-binding domain (SET; blue; amino acids 1,292–1,488), three bipartite NLS motifs (black; amino acids 462-477, 1370-1384, 1383-1399), six PEST sequences (brown; amino acids 1-13, 269-280, 548-561, 678-689, 806-830, 1502-1526) and a repeat domain (Rpt; grey; amino acids 1,520–1,543).[31-34] Blue circles represent previously reported variants and yellow circles indicate novel variants. Two individuals with larger deletions (IND 4,24) are not shown here. For cDNA annotation of the variants see Table 1.

Figure 2. Radar plot showing core neurodevelopmental features (n=29)^

Cohort Characteristics
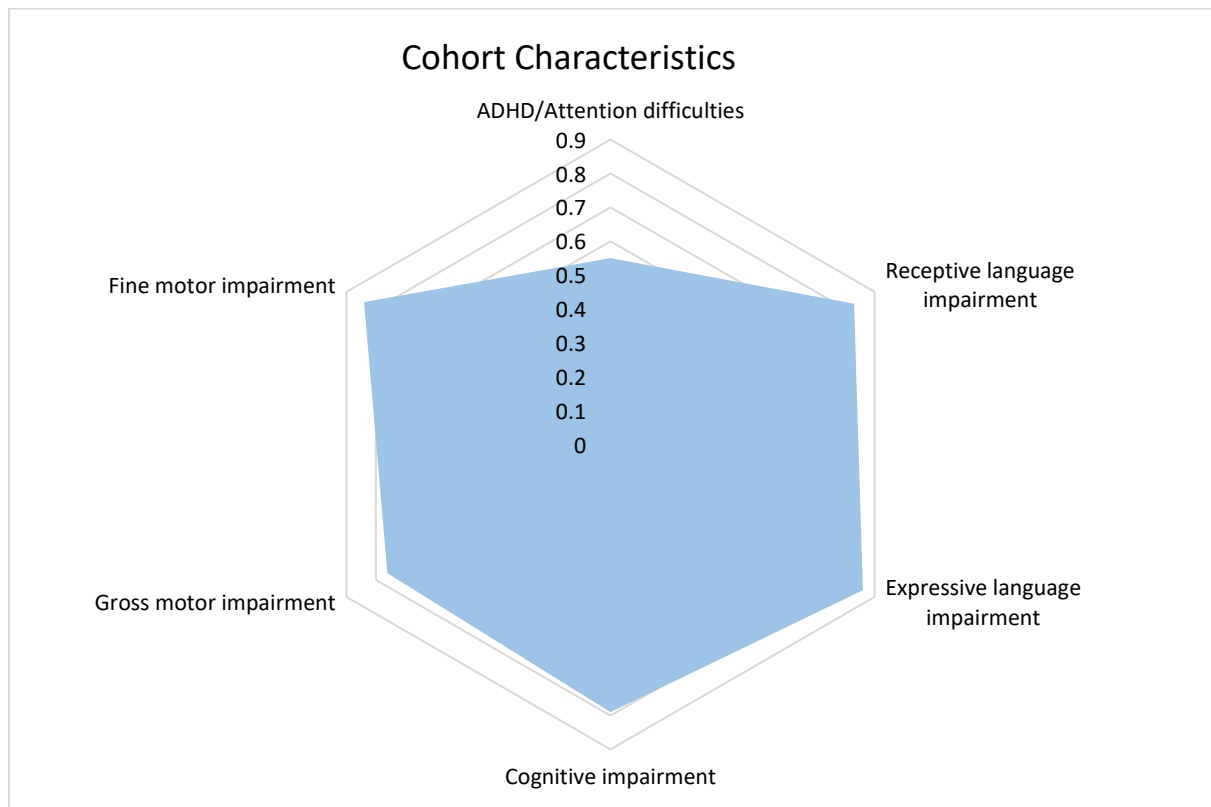
^2 children < 2 years and hence too young for reliably determining presence of these comorbid features.

Figure 3. Performance across sub-domains of language in the Children's Communication Checklist (n=16)^

Communication scores (CCC-2)
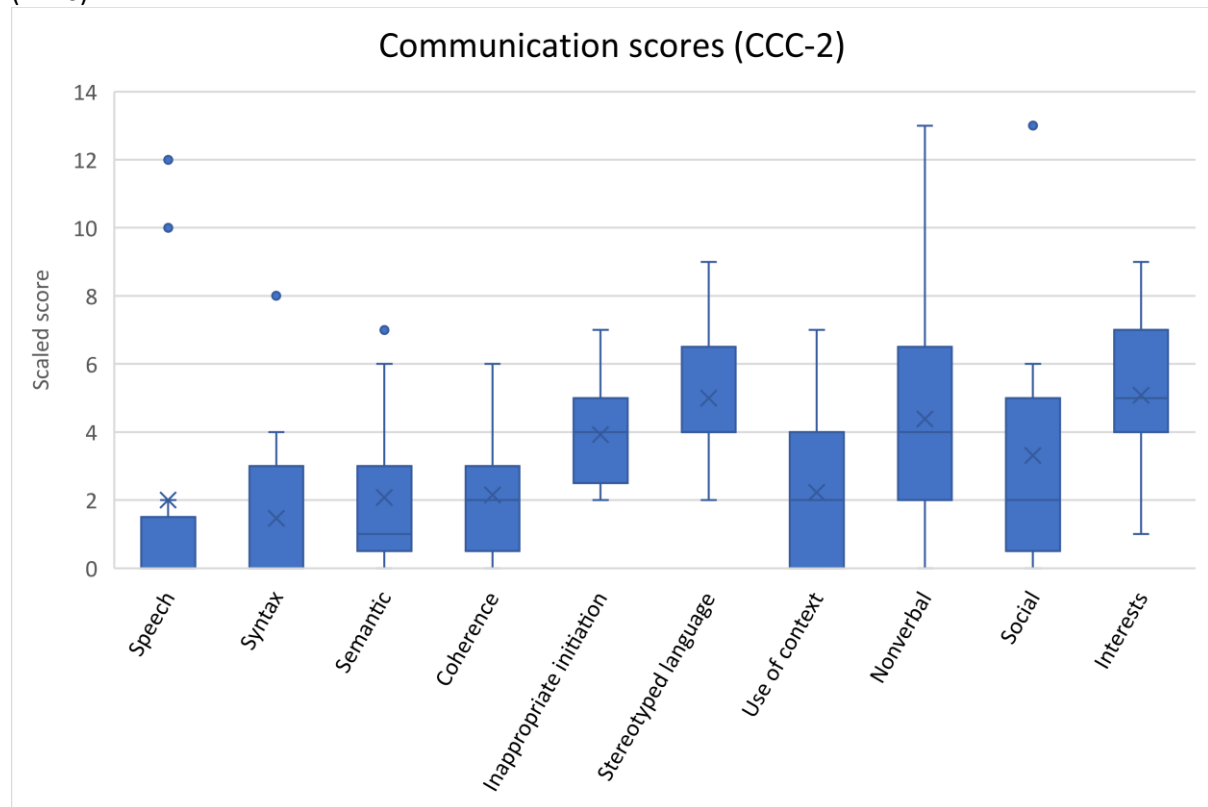
Lines denote median scores and X denotes the mean scores; • indicates an outlier. Scaled scores between 7-13 are within the average range. ^ English speaking children aged > 4 years.

Figure 4. Performance across domains of the Vineland Adaptive Behaviour Scale-3 (n=27).

Adaptive behaviour (Vineland-3)

*p<0.05
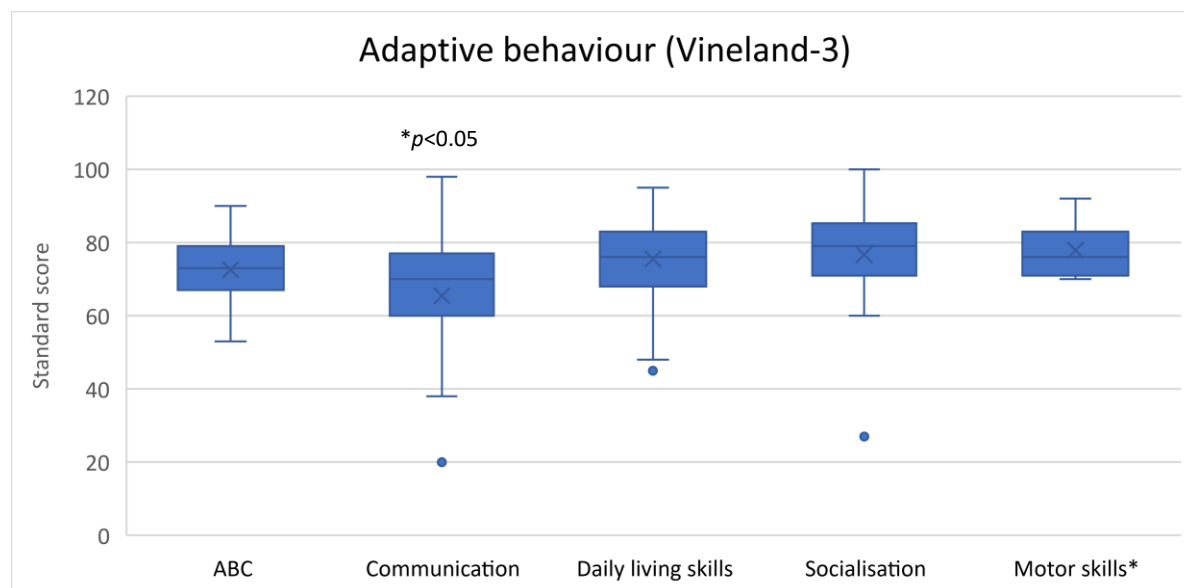
Lines denote median scores; X denotes mean scores; • indicates an outlier; ABC, Adaptive Behaviour Composite, ie. overall combined score. Standard scores between 85 and 115 are considered within the average range.

Table 1. Developmental history and medical features

IND	Age	Sex	Variant (NM_ 015559.2)	History of feeding difficulties	Intellectual impairment	Speech therapy	Fine motor ability$	Gross motor ability$	OT/PT	ADHD/ attention issues
1	0y8m	M	c.453_454insTGGG (p.(Lys152Trpfs*18))	Y	#	N	Mod. Low	Mod. Low	Y	N
2	1y3m	M	arr[GRCh37] 18q12.3(42519449_42567840)x1	Y	#	N	Mod. Low*	Mod. Low	Y	N
3	2y11m	F	c.1588C>T (p.(Arg530*))	Y	#	Y	Mod. Low	Mod. Low	Y	N
4	3y2m	M	arr 18q12.2q12.3(34576844_43015201)x1	Y	#	Y	Mod. Low	Mod. Low	N	N
5	3y4m	F	c.1633G>T (p.(Glu545*))	N	#	Y	Adequate	Mod. Low	Y	N
6	3y7m	F	c.1630C>T (p.(Arg544*))	Y	#	Y	Y~	Y~	Y	N
7	3y8m	F	c.1873C>T (p.(Arg625*))	Y	Mild	Y	Low	Mod. Low	Y	N
8	4y6m	F	c.1777C>T (p.(Gln593*))	N	Mild	Y	Adequate	Mod. Low	Y	Y
9	5y7m	F	c.1876C>T (p.(Arg626*))	Y	Moderate	Y	Y~	Y~	Y	N
10	6y1m	M	c.1408delA (p.(Met470*))	Y	Moderate~	Y	Mod. Low	Mod. Low	Y	N
11	6y2m	M	c.726_732del (p.(Arg243Leufs*98))	N	Borderline	Y	Mod. Low*	Adequate	Y	Y
12	6y10m	M	c.1676delC (p.(Pro559Argfs*21))	N	Mild	Y	Adequate	Mod. Low	Y	Y
13	7y0m	F	c.3799_3800insC (p.(Gly1267Alafs*17))	Y	Moderate	Y	Low	Mod. Low	Y	N
14	7y2m	M	c.3120C>A (p.(Tyr1040*))	N	Borderline	Y	Mod. Low	Low	Y	Y
15	7y6m	M	c.337delG (p.(Ala113Leufs*94))	Y	Mild-moderate	Y	Mod. Low	Mod. Low	Y	N
16	8y1m	M	c.1568delA (p.(His523Leufs*32))	Y	Borderline	Y	Mod. Low	Mod. Low	Y	Y
17	8y5m	M	c.821G>A (p.(Trp274*))	N	Borderline	Y	Adequate	Adequate	Y	Y
18	8y11m	F	c.2665C>T (p.(Arg889*))	Y	Borderline	Y	Y	Y	Y	N
19	9y4m	F	c.1765C>T (p.(Arg589*))	N	Mild	Y	Low	Mod. Low	Y	N
20	11y0m	M	c.1730_1749del (p.(Leu577Glnfs*20))	Y	Moderate	Y	Low *	Adequate *	Y	N
21	11y7m	M	c.1821del (p.(Ser608Alafs*22))	Y	Mild	Y	Low *	Mod. Low *	Y	Y
22	12y0m	F	c.422dup (p.(Arg142Valfs*7))	Y	Mild	Y	Mod. Low*	Mod. Low *	Y	Y
23	12y1m	M	c.1873C>T (p.(Arg625*))	N	Mild	Y	Low *	Adequate *	Y	Y
24	13y0m	M	arr[hg19] 18q12.3q21.1(39600614_45460709)x1	N	Moderate	Y	Y~	NA	N	N
25	13y5m	M	c.403C>T (p.(Gln135*))	N	Severe~	Y	Low *	Adequate *	N	N
26	13y7m	M	c.1765C>T (p.(Arg589*))	N	Mild	Y	Low *	Mod. Low *	Y	Y
27	14y2m	F	c.2076_2092delinsC (p.(Lys693Profs*86))	Y	Mild	Y	Mod. Low*	Adequate *	Y	N
28	14y10m	M	c.2156del (p.(Gly719Glufs*65))	Y	Mild-moderate	Y	Low *	Mod. Low *	Y	Y
29	15y10m	M	c.39_40del (p.(Gly15Argfs*47))	N	Moderate	Y	Low *	Mod. Low *	N	N
30	19y10m	M	c.2464del (p.(Ile822Tyrfs*13))	N	Moderate	Y	N~	N~	Y	N
31	23y2m	M	c.2982C>G (p.(Tyr994*))	Y	Average	Y	Mod. Low*	Mod. Low*	Y	N

y, years; m, months; F, Female; M, male; Y, feature present; N, feature absent; Severity of cognitive impairment was based on standard ID severity levels: borderline (71-85), mild (50/55- 70), moderate (35-49/54), severe (20-34), profound (<20); [&]vision impairment could be a familial trait unrelated to the *SETBP1* variant; ~ Parent report; # No clinical cognitive assessment due to young age; [$]Motor scores on Vineland Adaptive Behaviour Scale; OT, Occupational therapy received; PT, Physiotherapy received; ADHD, Attention Deficit Hyperactivity Disorder.

Table 2. Speech and language performance

IND	Age	Communication milestones (y;m)		Minimally verbal	Speech diagnosis~	Language ability[#]			Social skills[#]
		Spoken words	Short sentences			Expressive	Receptive	Written	
1	0y8m	^	^	^	^	Mod. Low	Mod. Low	NA	Adequate
2	1y3m	NYA^	^	^	N~	Low	Low	NA	Mod. Low
3	2y11m	NYA	NYA	Y	CAS	Low	Mod. Low	NA	Adequate
4	3y2m	15-18m	NYA	Y	NA	Low	Mod. Low	Mod. Low	Mod. Low
5	3y4m	<12m	NYA	Y	Phon, CAS	Mod. Low	Adequate	Adequate	Mod. Low
6	3y7m	<12m	NYA	Y	N~	NA	NA	NA	NA
7	3y8m	>18m	NYA	Y	CAS	Low	Low	Low	Mod. Low
8	4y6m	<12m	4-5y	N	Phon, CAS	Low	Low	Mod. Low	Adequate
9	5y7m	>18m	NYA	Y	N~	Low	Low	Low	Low
10	6y1m	15-18m	4-5y	N	Phon, CAS	Low	Adequate	Low	Adequate
11	6y2m	>18m	4-5y	N	Artic, CAS, Phon	Mod. Low	Mod. Low	Mod. Low	Mod. Low
12	6y10m	12-15m	6-7y	N	CAS	Low	Mod. Low	Low	Mod. Low
13	7y0m	>18m	NYA	Y	Phon, CAS, Dysarthria~	Low	Low	Low	Mod. Low
14	7y2m	>18m	4-5y	N	Phon, CAS, Dysarthria	Mod. Low	Mod. Low	Low	Low
15	7y6m	>18m	4-5y	N	Phon, CAS, Dysfluency	Low	Mod. Low	Low	Mod. Low
16	8y1m	12-15m	4-5y	N	Phon, CAS	Mod. Low	Adequate	Mod. Low	Mod. Low
17	8y5m	>18m	4-5y	N	CAS	Adequate	Adequate	Adequate	Adequate
18	8y11m	>18m	6-7y	N	CAS, Dysarthria	Severe$	Severe$	Severe$	-
19	9y4m	>18m	6-7y	N	Phon, CAS	Low	Low	Mod. Low	Mod. Low
20	11y0m	12-15m	NYA	Y	N~	Low	Low	Low	Mod. Low
21	11y7m	<12m	6-7y	N	Phon, CAS, Dysarthria	Low	Low	Low	Mod. Low
22	12y0m	15-18m	6-7y	N	Artic, Phon, CAS	Adequate	Mod. Low	Low	Adequate
23	12y1m	<12m	4-5y	N	Phon, CAS	Adequate	Adequate	Mod. Low	Mod. Low
24	13y0m	NYA	NYA	Y	Artic, CAS	Severe$	NA	NA	NA
25	13y5m	NYA	NYA	Y	CAS	Low	Mod. Low	Low	NA
26	13y7m	>18m	4-5y	N	CAS	Mod. Low	Mod. Low	Mod. Low	Low
27	14y2m	12-15m	6-7y	N	CAS	Mod. Low	Mod. Low	Low	Low
28	14y10m	>18m	>8y	N	CAS, Dysarthria	Mod. Low	Mod. Low	Low	Mod. Low
29	15y10m	15-18m	NYA	Y	Phon, CAS~	Mod. Low	Mod. Low	Low	Mod. Low
30	19y10m	<12m	4-5y	N	Phon, CAS	NA	NA	NA	NA
31	23y2m	<12m	2-3y	N	Phon, CAS	Adequate	Low	Mod. Low	Low

m, months; y, years; NYA, Not yet achieved; NA, Not assessed; Phon, Phonological disorder; CAS, Childhood Apraxia of Speech; Artic, Articulation disorder; [#]Rated using Vineland-3 standard scores and severity ratings; ~Based on parent report; ^ Patient is too young for these developmental milestones; $ Severity rated according to clinical language assessment results.

Table 3: Children's Communication Checklist scores (N=16)^

Children's Communication Checklist Domains	Average Standard Score Mean = 10, SD = 3
Speech	1.9
Syntax	1.4
Semantic	1.9
Coherence	2.0
Inappropriate initiation	3.6
Stereotyped (scripted language)	4.8
Use of context	2.1
Nonverbal communication	4.2
Social relations	3.1
Interests	4.9

^English speaking children aged > 4 years.

Table 4: Vineland-3 Adaptive Behaviour Scores (N=27)

Vineland-3 sub-domains	Average Standard Score Mean = 100, SD = 15
Adaptive Behaviour Composite	72.5
Communication Domain	65.4
Daily Living Skills Domain	75.6
Socialization Domain	76.7
Motor skills*	77.9

*Based on scores for patients aged <9y11m, normative data is unavailable for older ages.

Averages calculated based on data from the Vineland-3.