Common Genetic Model and Developmental Language Disorder

Monogenic causes of language disorders remain comparatively rare, and do not fully account for the developmental language disorders (DLD) prevalence rate of >7%. It is widely accepted that common risk variants confer a genetic susceptibility for DLDs. Termed 'complex genetic model', each variant contributes incrementally to an overall level of risk of developing a language disorder. Studies to identify these risk variants within a complex genetic model fall into two main approaches: linkage studies and genome-wide association studies (GWASs).

1. Linkage Analyses

Using genetic linkage studies to identify regions of the genome shared between affected individuals were the linchpin of neurodevelopmental genetics in the 2000s. As the principal method to detect common variants, they were used to generate many important findings to elucidate the genetics of complex disease. Linkage studies were particularly well suited to detect common variants with a moderate or large effect size and present in more than 10% of population. The Specific Language Impairment Consortium (SLIC) found regions strongly associated with SLI on 16q24 (SLI1) and 19q13 (SLI2) [1] (**Table 1**). Fine-mapping of these regions implicated two specific genes; C-mad inducing protein (*CMIP*) and calcium-transporting ATPase type 2C member 2 (*ATP2C2*) [2]. Both *CMIP* and *ATP2C2* were found to contain common risk variants with a moderate effect size and have additional evidence through the identification of monogenic cases of language disorder (see **Table 1**). Newbury et al. (2009) [2] discovered that *CMIP* was associated with language, reading and spelling in both the SLI cohort and in the general population. This may suggest a contribution to phonological language skills in language ability more generally. In contrast, *ATP2C2* was associated with phonological memory in the SLI cohort, but only showed association within the language-impaired group in the general population, suggesting a possible role specifically in individuals with language disorders. Recently, Martinelli et al. (2021) [3] characterised the functional effects of a rare variant in *ATP2C2* and its role in language disorders.

Table 1. Summary of findings from genome-wide association studies of language disorders.

Study	Sample No.	Cohort Type	Chr. Assoc.
Luciano et al. (2013) [4]	~6500	Population	21
Eicher et al. (2013) [5]	~170	Selected reading and language impaired	3, 4, 13
Nudel et al. (2014) [6]	~250	Selected (parent of origin)	5, 14
St Pourcain et al. (2014) [7]	~10,000	Population	3
Gialluisi et al. (2014) [8]	~1800	Selected reading and language impaired	7, 21
Harlaar et al. (2014) [9]	~2000	Population	2, 10
Kornilov et al. (2016) [10]	~400	Isolated population	9, 21
Eising et al. (2021) [11]	33,959	Selected and population Meta-analysis using 19 cohorts	1

Study	Sample No.	Cohort Type	Chr. Assoc.
Doust et al. (2021) [12]	51,800 dyslexia cases, 1,087,070 controls	Selected and population Meta-analysis using binary case/control self-reported measure of dyslexia	1, 2, 3, 6, 7, 11, 17, X

Bartlett et al. (2002) [13] showed a strong association to chromosomal region 13q21 (SLI3) on reading specific measures, and more modestly to regions 2p22 and 17q23 using more general measures of delayed language. In this case, they utilised a family-based linkage approach using five large Canadian families where several family members had a diagnosis of SLI. Evans et al. (2015) [14] identified two associated regions using 147 pairs of siblings, where at least one had an SLI diagnosis. They reported two regions associated with phonological memory, 10q23.33 and 13q33.3.

In many of these examples, the exact genes and contributory variants remain unelucidated. This is a particularly difficult issue with linkage studies, which tend to identify very large regions containing many hundreds of genes, making fine mapping difficult.

As cohort sizes have increased, and genomic data at a population level becomes more available, the linkage study has been replaced by the genome-wide association study except in large families or highly related populations. GWASs provide a much higher resolution of variants and allow for more efficient finer mapping of contributory variants. For example, Andres et al. (2019) [15] used linkage to identify a region on chromosome 2q associated with SLI in fourteen consanguineous Pakistani families, totalling 156 individuals.

Linkage studies are difficult to replicate in other populations, and there is rarely overlap between studies [16].

2. Genome-Wide Association Studies

The current methodology for detecting genomic variants associated with a complex condition is to perform a genome-wide association study (GWAS). GWAS uses advances in genetic marker technology to simultaneously assess more than 4 million sites of known common variation across the entire genome, providing higher resolution. In addition to more variants, the number of individuals has increased into the tens or hundreds of thousands. A pivotal study in the broader field of psychiatric genetics identified more than 100 regions associated with schizophrenia using 37,000 cases and 113,000 unaffected controls [17]. Studies focused on DLD and language-related phenotypes are only very recently beginning to achieve sample size on this scale (**Table 1**). A number of GWASs have been performed on SLI/DLD and related phenotypes, and identified genetic regions are summarised in **Table 1**. As with the linkage studies described in the previous section, there is little consistency between the genomic regions found to be associated between studies. This can be partially explained by differences in phenotyping used between studies, exacerbated by the lack of robust consensus criteria for diagnosing DLDs. Secondly, the genetic aetiology of DLDs are such that it likely involves many contributing variants across many different genes (and environmental factors) each of a small effect size.

Very recently, two research studies have presented meta-analyses in which multiple different GWAS cohorts are combined into one large study [11][12]. Meta-GWAS is a method whereby the GWAS summary statistics from more modest cohort sizes can be pooled together to increase power and is a cost-efficient means of gene identification. Eising et al. (2021) [11] utilised 22 different cohorts and five measures: word reading, non-word reading, spelling, phoneme awareness and non-word spelling. They found that word reading associated with a variant (rs11208009) using a subset of 19 cohorts and 33,959 individuals. The variant lies outside of a genic region, but is located near to (and in linkage disequilibrium with) three potential candidate genes: *DOCK7*, *ANGPTL3*, and *USP1*. They went on to show that both reading and language traits have a genetic basis that is largely separate to that of Performance IQ.

Even larger still, Doust et al. (2021) [12] utilised population and genetic data from 23 and Me, totalling 51,800 adults who self-reported that they had dyslexia and over a million controls without dyslexia. The authors identified 42 individual genomic regions that associated with diagnosis of dyslexia. Of these 42, 17 had been previously reported as associated with either education attainment or cognitive ability, and 25 were novel [12]. A total of 12 of the 25 novel regions went on to be independently replicated in separate cohorts. It is important to note the trade off in these studies between the sample size and phenotyping information, which was a yes/no question and self-reported.

Both Eising et al. (2021) [11] and Doust et al. (2021) [12] represent substantial leaps forward in understanding the genetic contribution to DLDs and sample size is now large enough to detect some of the missing heritability of language disorders. As the list of candidate genes grows, so does our knowledge of the biology of DLD risk. Polygenic profiles, in which an overall risk score is generated for each risk allele associated with a phenotype, can be generated which capture

the genetic differences and similarities between neurodevelopmental disorders. Early studies suggest that this is a promising area of research. Shared genetic effects have been shown to exist between cognitive ability, educational attainment, language development and psychosocial outcomes; however, this pilot study was based on very small sample sizes [18]. As polygenic profiles are updated using summary statistics from increasingly large GWASs, they become more sensitive and specific, allowing for improving inference accuracy. For example, the first profiling of educational attainment explained 2% of variance [19] while more recent scores explain 13% [20]. Polygenic methods are being developed in language disorders [18] and dyslexia [21]. Polygenic risk scores of clinical conditions indicate that polygenic profiles can be informative for the extremes of the population (who carry a high burden of risk or protective variants). So even if they are not useful for capturing individual variation in the middle of the distribution at the extremes, they can be clinically meaningful.

References

1. SLI Consortium. A Genomewide Scan Identifies Two Novel Loci Involved in Specific Language Impairment**Members of the consortium are listed in the Appendix. Am. J. Hum. Genet. 2002, 70, 384–398.

2. Newbury, D.F.; Winchester, L.; Addis, L.; Paracchini, S.; Buckingham, L.-L.; Clark, A.; Cohen, W.; Cowie, H.; Dworzynski, K.; Everitt, A.; et al. CMIP and ATP2C2 Modulate Phonological Short-Term Memory in Language Impairment. Am. J. Hum. Genet. 2009, 85, 264–272.

3. Martinelli, A.; Rice, M.L.; Talcott, J.B.; Diaz, R.; Smith, S.; Raza, M.H.; Snowling, M.J.; Hulme, C.; Stein, J.; Hayiou-Thomas, M.E.; et al. A rare missense variant in the ATP2C2 gene is associated with language impairment and related measures. Hum. Mol. Genet. 2021, 30, 1160–1171.

4. Luciano, M.; Evans, D.; Hansell, N.; Medland, S.; Montgomery, G.; Martin, N.; Wright, M.; Bates, T.C. A genome-wide association study for reading and language abilities in two population cohorts. Genes Brain Behav. 2013, 12, 645–652.

5. Eicher, J.D.; Powers, N.R.; Miller, L.L.; Akshoomoff, N.; Amaral, D.G.; Bloss, C.S.; Libiger, O.; Schork, N.J.; Darst, B.F.; Casey, B.J.; et al. Genome-wide association study of shared components of reading disability and language impairment. Genes Brain Behav. 2013, 12, 792–801.

6. Nudel, R.; Simpson, N.H.; Baird, G.; O'Hare, A.; Conti-Ramsden, G.; Bolton, P.F.; Hennessy, E.R.; SLI Consortium; Ring, S.M.; Davey Smith, G.D.; et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. Genes Brain Behav. 2014, 13, 418–429.

7. St Pourcain, B.S.; Cents, R.A.; Whitehouse, A.J.; Haworth, C.M.; Davis, O.S.; O'Reilly, P.F.; Roulstone, S.; Wren, Y.; Ang, Q.W.; Velders, F.P.; et al. Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat. Commun. 2014, 5, 4831.

8. Gialluisi, A.; Newbury, D.F.; Wilcutt, E.G.; Olson, R.K.; DeFries, J.C.; Brandler, W.M.; Pennington, B.F.; Smith, S.D.; Scerri, T.S.; Simpson, N.; et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav. 2014, 13, 686–701.

9. Harlaar, N.; Meaburn, E.L.; Hayiou-Thomas, M.E.; Davis, O.; Docherty, S.; Hanscombe, K.B.; Haworth, C.; Price, T.; Trzaskowski, M.; Dale, P.S.; et al. Genome-Wide Association Study of Receptive Language Ability of 12-Year-Olds. J. Speech Lang. Hear. Res. 2014, 57, 96–105.

10. Kornilov, S.A.; Rakhlin, N.; Koposov, R.; Lee, M.; Yrigollen, C.; Caglayan, A.O.; Magnuson, J.S.; Mane, S.; Chang, J.T.; Grigorenko, E.L. Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. Pediatrics 2016, 137, e20152469.

11. Eising, E.; Mirza-Schreiber, N.; de Zeeuw, E.L.; Wang, C.A.; Truong, D.T.; Allegrini, A.G.; Shapland, C.Y.; Zhu, G.; Wigg, K.G.; Fisher, S.E.; et al. Genome-wide association analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. bioRxiv 2021.

12. Doust, C.; Fontanillas, P.; Eising, E.; Gordon, S.D.; Wang, Z.; Alagöz, G.; Molz, B.; 23 and Me Research Team; Quantitative Trait Working Group of the GenLang Consortium; Luciano, M.; et al. Discovery of 42 Genome-Wide Significant Loci Associated with Dyslexia. medRxiv 2021.

13. Bartlett, C.W.; Flax, J.F.; Logue, M.W.; Vieland, V.J.; Bassett, A.S.; Tallal, P.; Brzustowicz, L.M. A Major Susceptibility Locus for Specific Language Impairment Is Located on 13q21. Am. J. Hum. Genet. 2002, 71, 45–55.

14. Evans, P.D.; Mueller, K.L.; Gamazon, E.R.; Cox, N.J.; Tomblin, J.B. A genome-wide sib-pair scan for quantitative language traits reveals linkage to chromosomes 10 and 13. Genes Brain Behav. 2015, 14, 387–397.

15. Andres, E.M.; Hafeez, H.; Yousaf, A.; Riazuddin, S.; Rice, M.L.; Basra, M.A.R.; Raza, M.H. A genome-wide analysis in consanguineous families reveals new chromosomal loci in specific language impairment (SLI). Eur. J. Hum. Genet.

2019, 27, 1274–1285.

16. Reader, R.H.; Covill, L.; Nudel, R.; Newbury, D.F. Genome-Wide Studies of Specific Language Impairment. Curr. Behav. Neurosci. Rep. 2014, 1, 242–250.

17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014, 511, 421–427.

18. Newbury, D.F.; Gibson, J.L.; Conti-Ramsden, G.; Pickles, A.; Durkin, K.; Toseeb, U. Using Polygenic Profiles to Predict Variation in Language and Psychosocial Outcomes in Early and Middle Childhood. J. Speech Lang. Hear. Res. 2019, 62, 3381–3396.

19. Rietveld, C.A.; Medland, S.E.; Derringer, J.; Yang, J.; Esko, T.; Martin, N.W.; Westra, H.-J.; Shakhbazov, K.; Abdellaoui, A.; Agrawal, A.; et al. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. Science 2013, 340, 1467–1471.

20. Lee, J.J.; Wedow, R.; Okbay, A.; Kong, E.; Maghzian, O.; Zacher, M.; Nguyen-Viet, T.A.; Bowers, P.; Sidorenko, J.; Karlsson Linnér, R.R.; et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 2018, 50, 1112–1121.

21. Gialluisi, A.; Andlauer, T.F.M.; Mirza-Schreiber, N.; Moll, K.; Becker, J.; Hoffmann, P.; Ludwig, K.U.; Czamara, D.; Pourcain, B.S.; Brandler, W.; et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Transl. Psychiatry 2019, 9, 77.