

Research Article

Genome-Wide Association Study of Receptive Language Ability of 12-Year-Olds

Nicole Harlaar,^a Emma L. Meaburn,^b Marianna E. Hayiou-Thomas,^c
Wellcome Trust Case Control Consortium 2,^d Oliver S. P. Davis,^e Sophia Docherty,^f
Ken B. Hanscombe,^f Claire M. A. Haworth,^f Thomas S. Price,^g Maciej Trzaskowski,^f
Philip S. Dale,^h and Robert Plomin^f

Purpose: Researchers have previously shown that individual differences in measures of receptive language ability at age 12 are highly heritable. In the current study, the authors attempted to identify some of the genes responsible for the heritability of receptive language ability using a *genome-wide association* approach.

Method: The authors administered 4 Internet-based measures of receptive language (vocabulary, semantics, syntax, and pragmatics) to a sample of 2,329 twelve-year-olds for whom DNA and genome-wide genotyping were available. Nearly 700,000 single-nucleotide polymorphisms (SNPs) and 1 million imputed SNPs were included in a genome-wide association analysis of receptive language composite scores.

Results: No SNP associations met the demanding criterion of genome-wide significance that corrects for multiple testing

across the genome ($p < 5 \times 10^{-8}$). The strongest SNP association did not replicate in an additional sample of 2,639 twelve-year-olds.

Conclusions: These results indicate that individual differences in receptive language ability in the general population do not reflect common genetic variants that account for more than 3% of the phenotypic variance. The search for genetic variants associated with language skill will require larger samples and additional methods to identify and functionally characterize the full spectrum of risk variants.

Key Words: receptive language, adolescents, genome-wide association study, genetics

Studies of twins have provided an extensive body of evidence demonstrating that genetic factors partly account for individual differences in language, speech, and literacy development (Hayiou-Thomas, 2008; Plomin, DeFries, Knopik, & Neiderhiser, 2013). Heritability estimates

vary widely depending on the phenotype under assessment and, in the case of language impairment, whether diagnoses are based on population screening or by measures of “clinical concern” (Bishop & Hayiou-Thomas, 2008). In most cases, however, estimates are nonzero and often substantial. Twin studies have also led to some more surprising conclusions. In particular, statistical modeling analyses have revealed that a continuum of genetic risk underlies both typical and delayed or atypical language development, such that many of the genes that influence the risk for delayed language also likely influence variability in normal language development (Plomin, Haworth, & Davis, 2009).

Over the past two decades, there have been vigorous efforts in the field of molecular genetics to identify some of the specific DNA variants responsible for the heritability of language development and language disorders (reviewed in Graham & Fisher, 2013; Paracchini, 2011). The first studies used linkage designs in families with multiple affected members. The goal of these studies was to identify chromosome regions inherited by affected family members at a frequency above chance, based on the expectation that these regions may harbor causal genetic variants. Using this approach, an

^aUniversity of Colorado at Boulder

^bUniversity of London, London, United Kingdom

^cUniversity of York, York, United Kingdom

^dWellcome Trust Case Control Consortium 2 (a full list of membership of the Wellcome Trust Case Consortium 2 is included in the Acknowledgments)

^eUCL Genetics Institute, London, United Kingdom

^fKing’s College London, London, United Kingdom

^gUniversity of Pennsylvania School of Medicine

^hUniversity of New Mexico

Correspondence to Nicole Harlaar: nicole.harlaar@colorado.edu or Emma L. Meaburn: e.meaburn@bbk.ac.uk.

Nicole Harlaar and Emma L. Meaburn are joint first authors.

Editor: Janna Oetting

Associate Editor: Sean Redmond

Received September 17, 2012

Revision received February 18, 2013

Accepted April 22, 2013

DOI: 10.1044/1092-4388(2013/12-0303)

Disclosure: The authors have declared that no competing interests existed at the time of publication.

early milestone was the discovery of the missense mutation in the forkhead-box protein (*FOXP2*) gene (chromosome 7q31), which was found to account for a severe and unusual form of developmental verbal dyspraxia in the KE family (Vargha-Khadem et al., 1998; Vargha-Khadem, Gadian, Copp, & Mishkin, 2005). Further studies have identified other genetic variants in the *FOXP2* gene in pedigrees or cases with dyspraxia (e.g., Lennon et al., 2007; Tomblin et al., 2009; Zeesman et al., 2006), although genetic variants in *FOXP2* have not been linked to language impairments in general population samples (Meaburn, Dale, Craig, & Plomin, 2002; Newbury et al., 2002; O'Brien, Zhang, Nishimura, Tomblin, & Murray, 2003). Subsequent linkage studies have implicated additional genetic regions in language disorders (reviewed in N. Li & Bartlett, 2012), and some of these findings have been successfully replicated (e.g., see Bartlett et al., 2002, 2004, for the *SLI3* loci on chromosome 13 and SLI Consortium, 2002, 2004, for *SLI1* on chromosome 16 and *SLI2* on chromosome 19).

Notwithstanding the importance of these early discoveries, a weakness of linkage-based designs is that they have low resolution: The chromosome regions they identify are often millions of base pairs long (Risch & Merikangas, 1996). An alternative approach is *allelic association*, which involves correlating trait variation in a population-based sample with allele frequencies of genetic variants, typically single-nucleotide polymorphisms (SNPs). A significant association may arise if the SNP itself is the causal genetic variant or if the SNP is correlated (in linkage disequilibrium [LD]) with the true causal allele. The first applications of allelic association in the language field were fine-mapping studies conducted within the context of linkage studies. The goal of these studies was to identify SNPs within the region "tagged" by an observed linkage signal. For example, an effort to identify the specific genetic variants in the *SLI1* linkage region yielded positive results for two genes, encoding c-maf-inducing protein (*CMIP*) and calcium-transporting ATPase, type2C, member C (*ATP2C2*), respectively; these associations were reported in families with language impairment as well as in a sample selected for low language performance from a population cohort (Newbury et al., 2009).

An alternative to nesting association designs within a linkage-based study is to examine allelic associations directly, either at a gene of interest (i.e., a candidate gene study) or across the genome (i.e., a genome-wide association [GWA] study). GWA studies are particularly useful if the goal is to identify novel candidate genetic variants—that is, SNPs that have not previously been associated with a phenotype. A GWA study is typically performed using DNA arrays, which permit cost-effective, high-throughput genotyping of common SNPs (typically 100,000–2,000,000 SNPs in total). The density of genetic markers assayed in GWA studies is usually sufficient to capture a large proportion of the common variation in the human genome. For quantitative traits, linear regression or Spearman's rank correlation is then used to test each SNP for an association between genotype and trait values on the phenotype of interest. The first major GWA studies of common medical disorders were reported in 2007 (Wellcome Trust Case Control Consortium [WTCCC],

2007). Significant results have since been reported for more than 200 disorders in 1,500 GWA studies (Hindorff et al., 2012; Visscher, Brown, McCarthy, & Yang, 2012), although replication of significant findings is often challenging (Ioannidis, Thomas, & Daly, 2009).

In the present article, we report the results of a GWA study of receptive language skill in a population-based sample. We focused on language skill, rather than language disorder, because common forms of language disorder (in contrast to *FOXP2*-associated language problems) are likely to be multifactorial, reflecting the effects of many genes, each with a small effect size. Quantitative genetic theory predicts that, if many genes affect a disorder, the disorder will reflect genetic variants that are relatively common in the population and that influence variation in language across all skill levels (Plomin et al., 2009). More specifically, we focused on language in early adolescence because research has previously shown that individual differences in language at this developmental stage show moderate to high heritability (Dale, Harlaar, Hayiou-Thomas, & Plomin, 2010; Hayiou-Thomas, Dale, & Plomin, 2012). Heritability estimates do not provide any insight into the effect sizes for individual genes, and therefore they cannot be used to predict how successful gene-discovery efforts are likely to be (except in the theoretical case where heritability is zero). Nonetheless, a relatively large and significant heritability estimate is an attractive starting point when selecting a trait for a GWA study because it implies that the total genetic variation (the sum of all genetic variants) in the sample studied makes a greater contribution to phenotypic variation compared with environmental or other nongenetic factors.

Examining genetic influences on individual differences in language development in early adolescence is also of general scientific interest (Nippold, 2007). During the transition to adolescence, demands on language grow in complexity and abstractness. There typically are gradual and subtle improvements in vocabulary and syntax. Sentence length slowly increases, and low-frequency structures, such as participial phrases and adverbial conjuncts, are used with increasing proficiency. There are also improvements in verbal reasoning and the ability to understand figurative language, such as words and expression that have abstract or multiple meanings. These skills enable the individual to engage in social interactions effectively and to use language as a means of analysis and self-control. There is substantial variation in these language skills across individuals, and this variation partly reflects genetic factors (Dale et al., 2010).

Given the number of statistical tests performed in GWA, probability values that are very small by traditional standards are to be expected merely by chance (Hirschhorn & Daly, 2005). As a consequence, standards of evidence for a GWA study are rigorous. Any identified association between an SNP and a phenotype must withstand a Bonferroni-type correction for over 1 million correlations, and it must be exactly replicated in one or more independent samples (i.e., the same SNP, allele, and direction of association). Accordingly, our study included both an initial discovery stage and a replication stage. We report GWA results for 1.7 million

common SNPs for receptive language ability in a representative sample of 2,329 twelve-year-olds for whom genome-wide genotyping of DNA was available. We sought to replicate the top hit emerging from the discovery sample in a replication sample of 2,639 twelve-year-olds for whom DNA was available but who were not included in the genome-wide genotyping.

Method

Participants

The sample was drawn from the Twins Early Development Study (TEDS), a longitudinal study of twins born in England and Wales between January 1994 and December 1996 (Haworth, Davis, & Plomin, 2013; Kovas, Haworth, Dale, & Plomin, 2007). Parental consent was obtained prior to data collection, and the project received approval from the Institute of Psychiatry ethics committee.

The discovery sample was drawn from the entire TEDS sample of over 11,000 twin pairs for whom DNA was available from saliva samples. Twins with severe birth complications, with medical problems, or whose first language was not English were excluded from the sample. To reduce possible confounding as a result of ancestry effects, the sample was restricted to families who identified themselves as Caucasian. After we had implemented perinatal, medical, language, and ethnicity exclusions, one member of each twin pair was selected for the discovery sample. Children were selected if they had more than 5 μ g DNA available and if they had participated in web-based cognitive testing at age 12 (described in Haworth et al., 2007). If both members of a twin pair fulfilled these two criteria, then the twin with the most DNA available was selected. This resulted in a sample of 4,442 children. Of this sample, 2,329 passed genotyping quality control (QC) procedures (detailed in the online Supplementary Materials) and had complete data on four receptive language measures included in the cognitive test battery. These 2,329 children formed the discovery sample.

The replication sample was drawn from the remaining TEDS sample, after excluding the 4,442 children selected for the discovery sample plus their co-twin if the twin pair was monozygotic (MZ). Children were selected if they had more than 3 μ g DNA available and if they had taken part in the web-based cognitive testing at age 12. To maximize the replication sample size and maintain power, both members of dizygotic twin pairs were included if they passed the selection criteria, and the dizygotic co-twins of discovery sample individuals were also included if eligible. Only one member of each MZ twin pair was selected, and if both members of an MZ twin pair fulfilled the selection criteria, then the twin with the most DNA was selected. These selection criteria resulted in a sample of 2,750 children. A subset of this sample, consisting of 2,639 children, passed the genotyping QC procedures and had complete data on the four receptive language measures. These 2,639 children formed our primary replication sample. We also identified a subsample of 1,010 unrelated individuals from the primary replication sample. This subsample, which we used as a

secondary replication sample, excluded individuals with twin siblings in the discovery sample and included only one member from each twin pair.

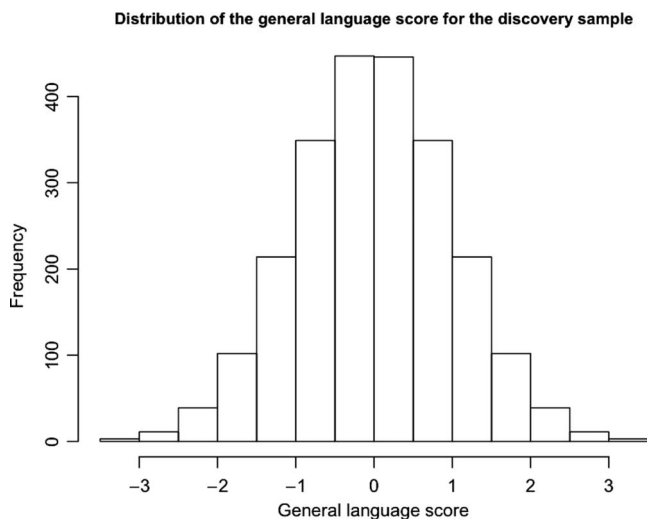
Our replication approach is somewhat unorthodox because any observed convergence between the discovery sample and the primary replication sample may be spuriously inflated by the nonindependence of these samples. On the other hand, the primary replication sample provides maximum power for replication. If agreement is observed between the results for the discovery sample and the primary replication sample, then additional replication in our fully independent sample of 1,010 unrelated individuals would be required for us to have confidence in the results. However, if the results from the discovery sample do not replicate in the primary replication sample, this is strong evidence of failure to replicate because this replication sample is highly similar to the discovery sample.

Materials and Procedures

Reliance on Internet-based testing, necessary for assessment of a large sample, led to our focus on measures of receptive language (Haworth et al., 2007). We used four measures of receptive language skill: (a) vocabulary, assessed using the Vocabulary Multiple Choice subtest of the Wechsler Intelligence Scale for Children (3rd ed., UK version; Wechsler, 1992); (b) semantics, assessed using Level 2 of the Figurative Language subtest of the Test of Language Competence—Expanded Edition (Wigg, Secord, & Sabers, 1989); (c) syntax, assessed using the Listening Grammar subtest of the Test of Adolescent and Adult Language—Third Edition (Hammill, Brown, Larsen, & Wiederholt, 1994); and (d) pragmatics, assessed using Level 2 of the Making Inferences subtest of the Test of Language Competence—Expanded Edition (Wigg et al., 1989). Details of the measures were described in detail in Dale et al.'s (2010) article. Sample statistics (*M*s and *SD*s) for the individual measures are shown for males and females in Supplementary Table 1; these did not differ significantly from those reported by Dale et al.

A previous multivariate genetic analysis showed that the four language measures were substantially correlated at a genetic level ($r_s = .74-.97$), indicating that genetic factors that contribute to variation in these measures largely overlap. A general language latent factor, reflecting the common variance among all four measures, free from measure-specific error, was highly heritable ($h^2 = 0.59$; Dale et al., 2010). Because it is not possible to obtain latent language factor scores (free from measurement error) for individual participants, we computed simple composite scores for the present analysis. These composite scores yielded a heritability estimate of 0.39 (95% confidence interval [CI] [0.34, 0.44]; Plomin, Haworth, et al., 2013). Only participants with valid data for all four language measures were genotyped. We adjusted scores for the linear effects of age at time of testing using the residuals from a least-squares linear regression as the phenotype. The distribution of test scores on the receptive language composite for the 2,329 individuals in the discovery sample is shown in Figure 1.

Figure 1. Distribution of receptive language composite scores in the discovery sample ($n = 2,329$).



Genotyping on the Affymetrix 6.0 GeneChip and subsequent QC was carried out as part of the WTCCC2 project (UK IBD Genetics Consortium et al., 2009). Nearly 700,000 genotyped SNPs met QC criteria. In addition, because genotyped SNPs are thought to “tag” causal variants, more than 1 million other SNPs were imputed using IMPUTE (Version 2) software (Howie, Donnelly, & Marchini, 2009) in order to increase the chances that common causal variants are represented. Details about the genotyping, QC procedures, and imputation method are included in the Supplemental Material.

We conducted GWA analyses using a linear regression approach implemented in SNPTEST (Version 2.0; WTCCC, 2007) under an additive model. This approach uses a frequentist method to account for uncertainty of genotype information (Marcini, Howie, Myers, McVean, & Donnelly, 2007). Because even small differences in allelic frequency within subgroups in the population can generate false-positive results, we used eight principal components representing population ancestry to control for population stratification. Sex and DNA sample plate number were also included as covariates. Details about the statistical analyses are given in the Supplemental Material. We visualized results using Manhattan plots, quantile–quantile (Q–Q) plots, and genotype–phenotype plots, generated in R (R Core Team, 2012). We also created a regional association plot, using LocusZoom (Prium et al., 2010).

Replication

The strongest SNP association from the GWA analysis of the discovery sample was selected for genotyping in the replication sample using the TaqMan SNP Genotyping assay. Linear regression was implemented in SNPTEST under an additive model, with sex added as a covariate. In addition, a

family-based test of association that accounts for sibling relatedness for the 377 sibling pairs within the primary replication sample of 2,639 individuals was performed in Plink (Version 1.07; Purcell et al., 2007).

Results

GWA Discovery

Because a GWA study generates a very large number of associations (each with its own p value), it is useful to compare the distribution of the actual p values derived from the GWA analyses with the distribution to be expected by chance. A Q–Q plot for the general language factor, which summarizes this comparison, is presented in Figure 2. This plot shows the expected distribution of association test statistics across SNPs on the x -axis compared to the observed values on the y -axis (negative log base 10 of the p values). The straight line at $x = y$ represents chance association, and the gray areas represent 95% concentration bands that approximate CIs on the null. One can see in Figure 1 that few associations fall outside the concentration bands, indicating little evidence of true association.

An alternative approach to visualizing these results is to use a Manhattan plot, shown in Figure 3. Each point represents a different SNP, laid out across the 22 human autosomes on the x -axis. The negative log base 10 p values are plotted on the y -axis. Evidence for a significant association would be indicated if we were able to detect a cluster of SNPs that form

Figure 2. Quantile–quantile plot for general language ability at age 12. Negative log base 10 ($-\log_{10}$) of the p values from a mixed-effects model likelihood ratio test are plotted against theoretical quantiles from the null distribution. The straight line at $x = y$ represents the null distribution, and the gray area represents a 95% confidence band on the null.

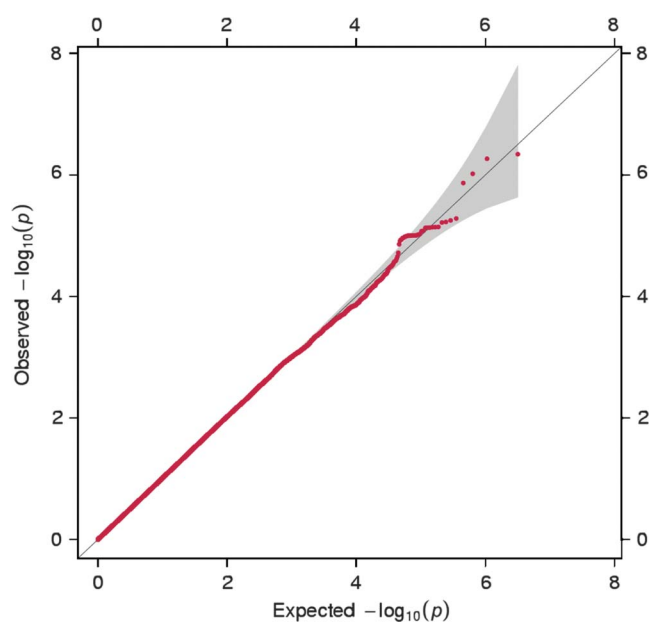
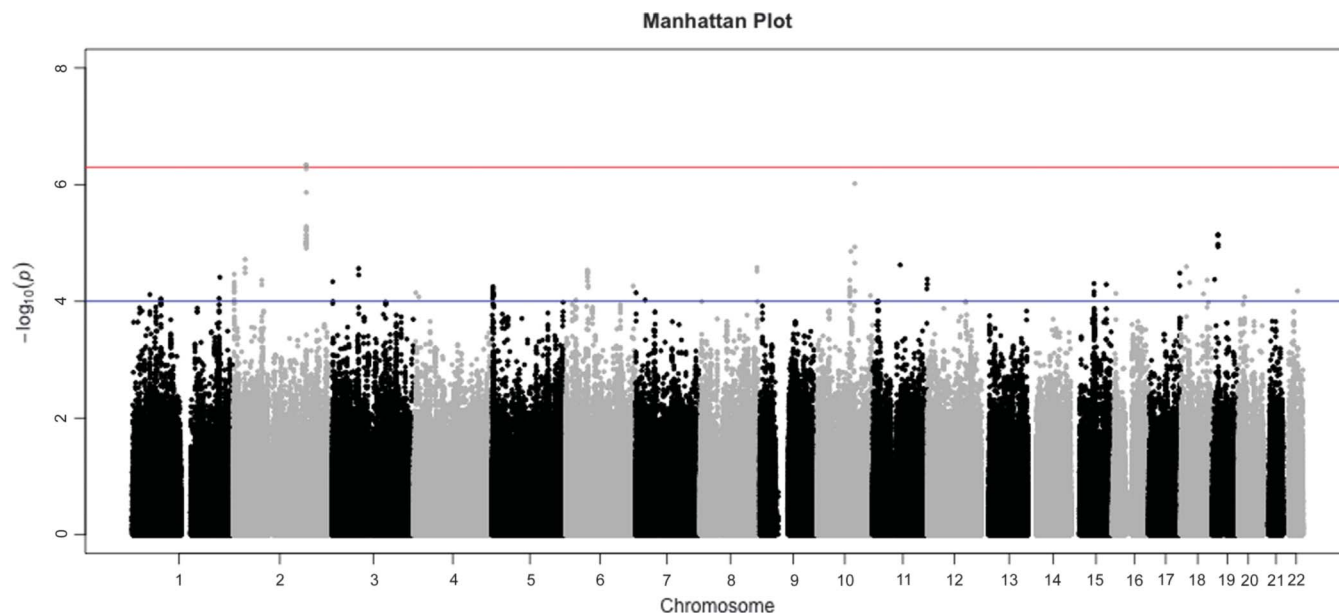


Figure 3. Manhattan plot for general language ability at age 12. Negative log base 10 p values from a mixed-effects model likelihood ratio test are plotted against genomic position.



a “tower” (resembling a Manhattan skyscraper), the result of nearby SNPs being in LD with one another and thus all marking the same association signal. In the discovery sample, none of the SNPs reached the conventional genome-wide significance threshold of $p < 5 \times 10^{-8}$. One SNP (rs12474600) on chromosome 2 showed an association just below this threshold ($p = 4.57 \times 10^{-7}$, $B = -0.24$, $SE = 0.05$, $n = 2329$; the solid red line in Figure 3). Clusters of low p values were also observed on chromosomes 10 and 19. The 114 strongest associations ($p \leq 1 \times 10^{-4}$; the solid blue line in Figure 3) are detailed in Supplementary Table 2.

The regional association plot in Figure 4 provides a more in-depth view of the chromosome 2 signal. This plot illustrates the associated region in the context of local patterns of LD and nearby genes. Specifically, the figure highlights a cluster of 24 SNPs that are strongly correlated and have a $p \leq 1 \times 10^{-4}$. The strongest associated SNP (rs12474600, $p = 4.57 \times 10^{-7}$) is an imputed SNP; however, six of the 24 SNPs with a $p \leq 1 \times 10^{-4}$ in this cluster were genotyped, confirming that the signal in this region is not based purely on imputation.

Figure 5 is a genotype–phenotype plot in which mean standardized language scores and standard errors are shown for the three genotypes for rs12474600 for the discovery sample of more than 2,300 individuals. For bi-allelic SNPs, the two alleles are designated “A” or “B” alphabetically. Under an additive model of association, the sign of the unstandardized beta indicates the direction of the association in relation to the number of copies of allele B. So, for rs12474600, an A/G SNP, the effect size ($B = -0.24$) indicates that allele B (G) is associated with lower language scores compared with allele A (A); that is, GG homozygotes have

lower language scores than AA homozygotes. This is illustrated in Figure 5, which shows that the AA homozygotes have an average standardized language score of 0.62, more than 0.5 SD higher than the GG homozygotes. The AG heterozygotes’ scores are significantly lower than the AA homozygotes and significantly higher than the GG homozygotes, as suggested by the nonoverlapping standard error bars for each of the genotypic points. This pattern indicates that the A and G alleles have additive effects. The relatively large standard error for the AA genotype reflects its relatively small sample size, caused by the *minor allele frequency* (the frequency at which the less common allele occurs in a population) of 0.10. In other words, for rs12474600, the relatively rare A allele contributes to higher language scores. However, we reiterate that even though receptive language scores differ significantly as a function of genotype (AA homozygote, AG heterozygote, GG homozygote), the overall effect of this SNP did not reach genome-wide significance.

Replication

We attempted to replicate the most significant SNP association, rs12474600 on chromosome 2. Although this SNP was imputed in the discovery sample, for the replication we used a validated Taqman assay to genotype the SNP. The rs12474600 association in the primary replication sample of 2,639 twelve-year-olds was not significant ($p = .357$), and its effect size was negligible ($B = -0.02$, $SE = 0.04$). Sibling relatedness that is not accounted for in association analyses may bias standard errors, and so we repeated the analysis taking family structure into account. The result remained nonsignificant ($p = .358$, $B = 0.02$). As would be expected, the

Figure 4. Regional association plot showing the 409 single-nucleotide polymorphisms (SNPs, filled circles) located in the 400-kb flanking the SNP rs12474600 (purple diamond). The plot shows association p values ($-\log_{10}(p)$) on the y -axis and the chromosomal position in base pairs on the x -axis. The strength of pairwise linkage disequilibrium of each SNP with rs12474600 is indicated by the color of the filled circles.

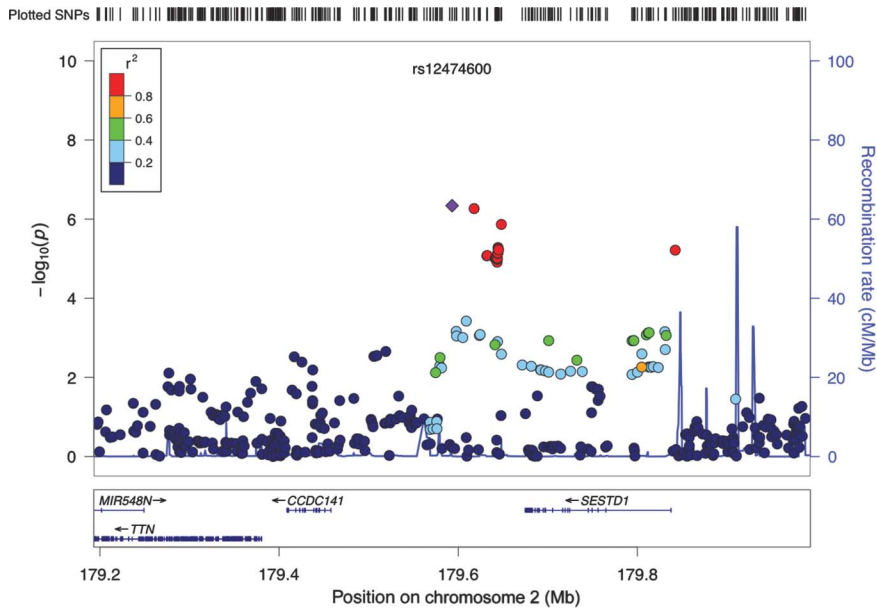
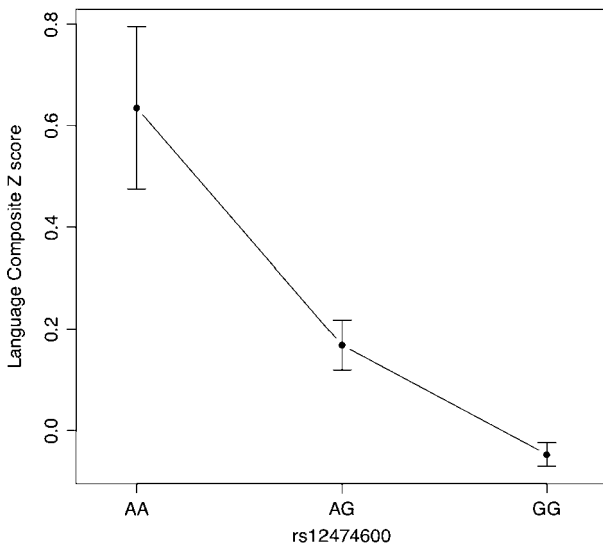


Figure 5. Results for rs12474600, the SNP with the lowest p value in the discovery sample ($n = 2,329$). The figure shows the mean standardized language score and standard errors for the three genotypes. Also shown are the chromosome, p value, unstandardized regression coefficient (B) and minor allele frequency.

rs12474600; chromosome 2; $p = 4.57E-07$; beta = -0.23908 ; MAF = 0.11



secondary replication sample, consisting of 1,010 individuals unrelated to individuals in the discovery sample, yielded similarly negative results ($p = .27$, $B = -0.043$, $SE = 0.07$). Because rs12474600 was imputed in the discovery sample, it is possible that this SNP would not have shown the lowest p value in the discovery sample if it had been genotyped directly. However, this seems unlikely because other SNPs in LD with rs12474600 that were genotyped directly showed similarly low p values (see Figure 4). In any case, the SNP with the lowest p value in the discovery sample did not show any association in our replication sample.

Discussion

This GWA study of receptive language ability in early adolescence found no evidence for genome-wide significant associations. The SNP closest to the conventional genome-wide significance level, rs12474600, failed to replicate, even though the replication sample consisted of a highly similar sample tested at the same age and with the same measures. In the discovery sample we had 92% power at the $p < 5 \times 10^{-8}$ level to detect an association for a causal variant with a minor allele frequency of 20% and a 2% effect size (Purcell, Cherny, & Sham, 2003). Given the estimated power, the results are consistent with a view that there are no detectable common SNPs associated with receptive language that account for more than 2% of the variance.

How can we reconcile the current findings with the robust and relatively high heritability estimates for language? A parochial explanation is that heritability estimates from

twin studies are simply wrong. For example, they may be overinflated due to violations of the equal-environments assumption (Plomin, DeFries, et al., 2013). However, similar heritability estimates for language have been reported in other designs, such as pedigree studies (e.g., Logan et al., 2011), which have different assumptions and problems. The hypothesis we favor is that the genetic architecture of receptive language, similar to height, weight, and IQ, reflects many common SNPs, each with a very small effect size. Support for this view comes from statistical methods that estimate the net effect of genetic influence using genotyped SNPs in samples of unrelated individuals. The first application of this approach was included in the software package Genome-wide Complex Trait Analysis (GCTA; Visscher, Yang, & Goddard, 2010; Yang, Lee, Goddard, & Visscher, 2011). GCTA uses random genetic similarity between each pair of unrelated individuals to estimate the variance in a phenotype accounted for by the genotyped SNPs (Visscher et al., 2010; Yang et al., 2011). Although GCTA does not identify specific SNPs that contribute to phenotypic variance, it does provide an estimate of how much variance the relevant SNPs would account for if one could identify each of them and add up their effects. It also provides some insight into genetic architecture. GCTA detects additive effects only of the common variants that are included on commercial genotyping arrays used in GWA studies. Because GWA studies are also limited to detecting additive effects of common variants, GCTA estimates of genetic influence mark the ceiling for GWA studies; that is, to the extent that associations with rare variants or gene–gene interactions (epistasis) are important, neither GCTA nor GWA studies will detect them. However, a recent GCTA analysis of cognitive phenotypes in TEDS yielded a significant SNP-based heritability estimate of .29 for the 12-year receptive language composite (95% CI [0.05, 0.53], which accounted for three quarters of the twin study heritability estimate of .39 (95% CI [0.34, 0.44]; Plomin, Haworth, et al., 2013); that is, about three quarters of the additive genetic variation in receptive language in early adolescence is tagged by common SNPs on commercial genotyping arrays.

The results from the current study are also unsurprising in the light of results for a range of complex genetic disorders and quantitative traits published since this study was conceived in 2007. Even in studies with tens of thousands of participants, research has shown that the largest detectable genetic effect sizes account for less than 1% of the phenotypic variance (e.g., for height, Lango Allen et al., 2010; and for weight, Walley, Asher, & Froguel, 2009). For behavioral traits, the largest effect sizes in the first GWA studies of reading, mathematics, and general cognitive ability assessed as quantitative traits in children comprised less than 0.5% of the variance (Butcher, Davis, Craig, & Plomin, 2008; Docherty, Davis, et al., 2010; Meaburn, Harlaar, Craig, Schalkwyk, & Plomin, 2008). It follows that extremely large samples will be needed in order to reveal significant genetic associations for language skill, given the stringent thresholds of statistical significance used to establish association in GWA studies at large (Plomin, 2013). For example, Chabris et al. (2012)

proposed that a sample size of 100,000 individuals has statistical power of 80% to discover genetic variants accounting for as little as 0.04% of the variance in a trait at a genome-wide significance level of $p < 5 \times 10^{-8}$. It is unlikely that any single laboratory would be able to attain a sample of this size. Carefully designed meta-analyses (combining p value results) and mega-analyses (combining data) will therefore be crucial in efforts to increase sample size and statistical power.

Although we have stressed the importance of common SNPs with tiny effects, another direction for molecular genetic studies of language is to study low-frequency polymorphisms: variants that are not rare but are less common than those tagged by commercially available microarrays (e.g., minor allele frequencies between 1% and 5%; Plomin, 2013). These variants may have a spectrum of effect sizes, from very small to intermediate or even large for some individuals, even though their effect overall in the population is miniscule. Although the jury remains out on the relative importance of uncommon variants for complex quantitative traits, linkage study findings for language suggest that such variants may be important, especially at the extremes of the language distribution (e.g., the *CMIP* and *ATP2C2* variants; Newbury et al., 2009). DNA arrays for a new wave of GWA studies, which will account for these less common variants, are already being designed. Use of such arrays will not be a panacea to standard GWA analysis, however; very large samples will still be required if individual variants explain a very small proportion of variance in the population (Visscher, Goddard, Derks, & Wray, 2012).

Finally, whole-genome sequencing methods that determine the complete DNA sequence of 3 billion nucleotide base pairs of an individual is the “next big thing” in genomics (Cirulli & Goldstein, 2010; Pasaniuc et al., 2012; Plomin, 2013; Plomin & Simpson, in press). Whole-genome sequencing means that DNA variants of any kind—not just common SNPs—can be detected. For studies of language skill and disorder, maximal power may be gained by oversampling individuals at one or both ends of the extremes, that is, sequencing individuals with very poor language scores, who may be the most likely to carry a high-risk allele burden, and individuals with very good language scores, who may be most likely to carry alleles conferring protection (Guey et al., 2011; D. Li, Lewinger, Gauderman, Murcray, & Conti, 2011).

Several caveats should be noted. One limitation of the current study, already mentioned, is the relatively small sample size; in order to have adequate power to detect common variants that account for 1% of the variance in language abilities, a discovery sample of more than 6,000 individuals is required. A second limitation is that we studied a single facet of language, namely, receptive language skill. Although the use of multiple tests augments the reliability of our receptive language scores, molecular genetic studies of language would benefit from multiple adequately normed indices of specific language skills, such as pragmatics, vocabulary, and syntactic skill (McCardle, Cooper, & Freund, 2005). Third, we did not include the sex chromosome in our analyses because of previous agreements with the WTCCC2, and so any associations in this region will have been missed.

The current results notwithstanding, we remain optimistic about the future of molecular genetic research on language skill, although this will require larger samples and new methods. Genetic variants that are robustly associated with language will provide essential biological leads for subsequent functional studies that aim to improve understanding of the molecular mechanisms involved in language development. In addition, if we are able eventually to identify a number of genetic variants that are associated with language, composites of these SNPs could be leveraged to test research questions raised by quantitative genetics, such as the extent to which genetic influences for language disorders overlap for commonly comorbid disorders, such as dyslexia (the *generalist genes hypothesis*; Plomin & Kovas, 2005; see, e.g., Docherty, Kovas, Petrill, & Plomin, 2010). The coming decade will likely be an exciting one for researchers interested in understanding the contribution of genetic factors to language abilities and disabilities.

Acknowledgments

We gratefully acknowledge the ongoing contribution of the parents and children in the Twins Early Development Study (TEDS). TEDS is supported by a program grant [G0901245; and previously G0500079] from the U.K. Medical Research Council (MRC), with additional support from the US National Institutes of Health [HD044454; HD059215]. R.P. is supported by a Medical Research Council Research Professorship award [G19/2] and a European Advanced Investigator award [295366]; O.S.P.D. is supported by a Sir Henry Wellcome Fellowship from the Wellcome Trust [WT088984]; C.M.A.H. is supported by a research fellowship from the British Academy. Genome-wide genotyping was made possible by grants from the Wellcome Trust Case Control Consortium 2 project [085475/B/08/Z; 085475/Z/08/Z]; see below for membership of WTCCC2 and member affiliations.

WTCCC2 membership is as follows:

Management Committee: Peter Donnelly (Chair),^{1,2} Ines Barroso (Deputy Chair),³ Jenefer M. Blackwell,^{4,5} Elvira Bramon,⁶ Matthew A. Brown,⁷ Juan P. Casas,⁸ Aiden Corvin,⁹ Panos Deloukas,³ Audrey Duncanson,¹⁰ Janusz Jankowski,¹¹ Hugh S. Markus,¹² Christopher G. Mathew,¹³ Colin N. A. Palmer,¹⁴ Robert Plomin,¹⁵ Anna Rautanen,¹ Stephen J. Sawcer,¹⁶ Richard C. Trembath,¹³ Ananth C. Viswanathan,¹⁷ and Nicholas W. Wood.¹⁸

Data and Analysis Group: Chris C. A. Spencer,¹ Gavin Band,¹ Céline Bellenguez,¹ Colin Freeman,¹ Garrett Hellenthal,¹ Eleni Giannoulatou,¹ Matti Pirinen,¹ Richard Pearson,¹ Amy Strange,¹ Zhan Su,¹ Damjan Vukcevic,¹ and Peter Donnelly.^{1,2}

DNA, Genotyping, Data Quality Control, and Informatics Group: Cordelia Langford,³ Sarah E. Hunt,³ Sarah Edkins,³ Rhian Gwilliam,³ Hannah Blackburn,³ Suzannah J. Bumpstead,³ Serge Dronov,³ Matthew Gillman,³ Emma Gray,³ Naomi Hammond,³ Alagurevathi Jayakumar,³ Owen T. McCann,³ Jennifer Liddle,³ Simon C. Potter,³ Radhi Ravindrarajah,³ Michelle Ricketts,³ Matthew Waller,³ Paul Weston,³ Sara Widaa,³ Pamela Whittaker,³ Ines Barroso,³ and Panos Deloukas.³

Publications Committee: Christopher G. Mathew (Chair),¹³ Jenefer M. Blackwell,^{4,5} Matthew A. Brown,⁷ Aiden Corvin,⁹ and Chris C. A. Spencer.¹

WTCCC2 member affiliations are as follows:

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. ²Department of Statistics, University of Oxford, Oxford, United Kingdom. ³Wellcome Trust

Sanger Institute, Cambridge, United Kingdom. ⁴Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Subiaco. ⁵Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom. ⁶Department of Psychosis Studies, National Institute for Health Research Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London, and The South London and Maudsley National Health Service Foundation Trust, London, United Kingdom. ⁷University of Queensland Diamantina Institute, Brisbane, Queensland, Australia. ⁸Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, and Department of Epidemiology and Public Health, UCL, London, United Kingdom. ⁹Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland. ¹⁰Molecular and Physiological Sciences, The Wellcome Trust, London, United Kingdom. ¹¹Department of Oncology, University of Oxford, Oxford, United Kingdom; Digestive Diseases Centre, Leicester Royal Infirmary, Leicester, United Kingdom; and Centre for Digestive Diseases, Queen Mary University of London, United Kingdom. ¹²Department of Clinical Neurosciences, St. George's University of London, United Kingdom. ¹³Department of Medical and Molecular Genetics, King's College London, United Kingdom, and King's Health Partners, Guy's Hospital, London, United Kingdom. ¹⁴Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee, United Kingdom. ¹⁵King's College London Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, United Kingdom. ¹⁶University of Cambridge Department of Clinical Neurosciences, Addenbrooke's Hospital, Cambridge, United Kingdom. ¹⁷National Institute for Health Research, Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital, London, United Kingdom, and National Health Service Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom. ¹⁸Department of Molecular Neuroscience, Institute of Neurology, London, United Kingdom.

References

- Bartlett, C. W., Flax, J. F., Logue, M. W., Smith, B. K., Vieland, V. J., Bassett, A. S., ... Brzustowicz, L. M.** (2004). Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Human Heredity*, *57*, 10–20. doi:10.1159/000077385
- Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P., & Brzustowicz, L. M.** (2002). A major susceptibility locus for specific language impairment is located on 13q21. *American Journal of Human Genetics*, *71*, 45–55. doi:10.1086/341095
- Bishop, D. V., & Hayiou-Thomas, M. E.** (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain and Behavior*, *7*, 365–372. doi:10.1111/j.1601-183X.2007.00360.x
- Butcher, L. M., Davis, O. S. P., Craig, I. W., & Plomin, R.** (2008). Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500 K single nucleotide polymorphism microarrays. *Genes, Brain and Behavior*, *7*, 435–446. doi:10.1111/j.1601-183X.2007.00368.x
- Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J. P., Cesarini, D., van der Loos, M., ... Laibson, D.** (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, *23*, 1314–1323. doi:10.1177/0956797611435528

- Cirulli, E. T., & Goldstein, D. B.** (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, *11*, 415–425. doi:10.1038/nrg2779
- Dale, P. S., Harlaar, N., Hayiou-Thomas, M. E., & Plomin, R.** (2010). The etiology of diverse receptive language skills at 12 years. *Journal of Speech, Language, and Hearing Research*, *53*, 982–992. doi:10.1044/1092-4388(2009/09-0108)
- Docherty, S. J., Davis, O. S., Kovas, Y., Meaburn, E. L., Dale, P. S., Petrill, S. A., ... Plomin, R.** (2010). A genome-wide association study identifies multiple loci associated with mathematics ability and disability. *Genes, Brain and Behavior*, *9*, 234–247. doi:10.1111/j.1601-183X.2009.00553.x
- Docherty, S. J., Kovas, Y., Petrill, S. A., & Plomin, R.** (2010). Generalist genes analysis of DNA markers associated with mathematical ability and disability reveals shared influence across ages and abilities. *BMC Genetics*, *11*, 61. doi:10.1186/1471-2156-11-61
- Graham, S. A., & Fisher, S. E.** (2013). Decoding the genetics of speech and language. *Current Opinion in Neurobiology*, *23*, 43–51. doi:10.1016/j.conb.2012.11.006
- Guey, L. T., Kravic, J., Melander, O., Burt, N. P., Laramie, J. M., Jonsson, A., ... Voight, B. F.** (2011). Power in the phenotypic extremes: A simulation study of power in discovery and replication of rare variants. *Genetic Epidemiology*. Advance online publication. doi:10.1002/gepi.20572
- Hamill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L.** (1994). *Test of Adolescent and Adult Language*. Austin, TX: Pro-Ed.
- Haworth, C. M., Davis, O. S., & Plomin, R.** (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*, 117–125. doi:10.1017/thg.2012.91
- Haworth, C. M. A., Harlaar, N., Kovas, Y., Davis, O. S. P., Oliver, B. R., Hayiou-Thomas, M. E., ... Plomin, R.** (2007). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*, *10*, 554–563.
- Hayiou-Thomas, M. E.** (2008). Genetic and environmental influences on early speech, language and literacy development. *Journal of Communication Disorders*, *41*, 397–408. doi:dx.doi.org/10.1016/j.jcomdis.2008.03.002
- Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R.** (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science*, *15*, 233–244. doi:10.1111/j.1467-7687.2011.01119.x
- Hindorf, L. A., MacArthur, J., Morales, J., Junkins, H. A., Hall, P. N., Klemm, A. K., & Manolio, T. A.** (2012). *A catalog of published genome-wide association studies*. Retrieved from www.genome.gov/gwastudies
- Hirschhorn, J. N., & Daly, M. J.** (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*, 95–108. doi:10.1038/nrg1521
- Howie, B. N., Donnelly, P., & Marchini, J.** (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, *5*(6), e1000529.
- Ioannidis, J. P., Thomas, G., & Daly, M. J.** (2009). Validating, augmenting and refining genome-wide association signals. *Nature Reviews Genetics*, *10*, 318–329. doi:10.1038/nrg2544
- Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R.** (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, *72*, 1–144. doi:10.1111/j.1540-5834.2007.00453.x
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... Hirschhorn, J. N.** (2010, October 14). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*, 832–838. doi:10.1038/nature09410
- Lennon, P. A., Cooper, M. L., Peiffer, D. A., Gunderson, K. L., Patel, A., Peters, S., ... Bacino, C. A.** (2007). Deletion of 7q31.1 supports involvement of *FOXP2* in language impairment: Clinical report and review. *American Journal of Medical Genetics Part A*, *143A*, 791–798. doi:10.1002/ajmg.a.31632
- Li, D., Lewinger, J. P., Gauderman, W. J., Murcray, C. E., & Conti, D.** (2011). Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genetic Epidemiology*, *35*, 790–799. doi:10.1002/gepi.20628
- Li, N., & Bartlett, C. W.** (2012). Defining the genetic architecture of human developmental language impairment. *Life Sciences*, *90*, 469–475. doi:10.1016/j.lfs.2012.01.016
- Logan, J., Petrill, S. A., Flax, J., Justice, L. M., Hou, L., Bassett, A. S., ... Bartlett, C. W.** (2011). Genetic covariation underlying reading, language and related measures in a sample selected for specific language impairment. *Behavior Genetics*, *41*, 651–659. doi:10.1007/s10519-010-9435-0
- Marcini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P.** (2007). A new multipoint method for genome-wide association studies via imputation of genotypes. *Nature Genetics*, *38*, 906–913. doi:10.1038/ng2088
- McCardle, P., Cooper, J., & Freund, L.** (2005). Language and genetics: Needs and opportunities. *Applied Psycholinguistics*, *26*, 129–135. doi:10.1017/S0142716405050101
- Meaburn, E., Dale, P. S., Craig, I. W., & Plomin, R.** (2002). Language-impaired children: No sign of the *FOXP2* mutation. *NeuroReport*, *13*, 1075–1077.
- Meaburn, E. L., Harlaar, N., Craig, I. W., Schalkwyk, L. C., & Plomin, R.** (2008). Quantitative trait locus association scan of early reading ability and disability using pooled DNA and 100 K SNP microarrays in a sample of 5760 children. *Molecular Psychiatry*, *13*, 729–740. doi:10.1038/sj.mp.4002063
- Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S., Baird, G., ... International Molecular Genetic Study of Autism Consortium.** (2002). *FOXP2* is not a major susceptibility gene for autism of specific language impairment. *American Journal of Human Genetics*, *70*, 1318–1327. doi:10.1086/339931
- Newbury, D. F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L. L., Clark, A., ... Monaco, A. P.** (2009). *CMIP* and *ATP2C2* modulate phonological short-term memory in language impairment. *American Journal of Human Genetics*, *85*, 264–272. doi:10.1016/j.ajhg.2009.07.004
- Nippold, M. A.** (2007). *Later language development: The school-age and adolescent years* (3rd ed.). Austin, TX: Pro-Ed.
- O'Brien, E. K., Zhang, X., Nishimura, C., Tomblin, J. B., & Murray, J. C.** (2003). Association of specific language impairment (SLI) to the region of 7q31. *American Journal of Human Genetics*, *72*, 1536–1543. doi:10.1086/375403
- Paracchini, S.** (2011). Dissection of genetic associations with language-related traits in population-based cohorts. *Journal of Neurodevelopmental Disorders*, *3*, 365–373. doi:10.1007/s11689-011-9091-6
- Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., ... Price, A. L.** (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics*, *44*, 631–635. doi:10.1038/ng.2283
- Plomin, R.** (2013). Child development and molecular genetics: 14 years later. *Child Development*, *84*, 104–120. doi:10.1111/j.1467-8624.2012.01757.x

- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M.** (2013). *Behavioral genetics* (6th ed.). New York, NY: Worth.
- Plomin, R., Haworth, C. M. A., & Davis, O. S. P.** (2009). Common disorders are quantitative traits. *Nature Reviews Genetics, 10*, 872–878. doi:10.1038/nrg2670
- Plomin, R., Haworth, C. M. A., Meaburn, E. L., Price, T., Wellcome Trust Case Control Consortium, & Davis, O. S. P.** (2013). Common DNA markers can account for at least half of the genetic influence on cognitive abilities. *Psychological Science, 24*, 562–568. doi:10.1177/0956797612457952
- Plomin, R., & Kovas, Y.** (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*, 592–617. doi:10.1037/0033-2909.131.4.592
- Plomin, R., & Simpson, M. A.** (in press). The future of genomics for developmentalists. *Development and Psychopathology*.
- Priem, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., ... Willer, C. J.** (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics, 26*, 2336–2337. doi:10.1093/bioinformatics/btq419
- Purcell, S., Cherny, S. S., & Sham, P. C.** (2003). Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics, 19*, 149–150. doi:10.1093/bioinformatics/19.1.149
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C.** (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics, 81*, 559–575. doi:10.1086/519795
- R Core Team.** (2012). *R: A language and environment for statistical computing*. Retrieved from www.R-project.org
- Risch, N., & Merikangas, K.** (1996, September 13). The future of genetic studies of complex human diseases. *Science, 13*, 1516–1517. doi:10.1126/science.273.5281.1516
- SLI Consortium.** (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics, 70*, 384–398. doi:10.1086/338649
- SLI Consortium.** (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *American Journal of Human Genetics, 74*, 1225–1238. doi:10.1086/421529
- Tomblin, J. B., O'Brien, M., Shriberg, L. D., Williams, C., Murray, J., Patil, S., ... Ballard, K.** (2009). Language features in a mother and daughter of a chromosome 7;13 translocation involving *FOXP2*. *Journal of Speech, Language, and Hearing Research, 52*, 1157–1174. doi:10.1044/1092-4388(2009/07-0162)
- UK IBD Genetics Consortium, Barrett, J. C., Lee, J. C., Lees, C. W., Prescott, N. J., Anderson, C. A., ... Strachan, D. P.** (2009). Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nature Genetics, 41*, 1330–1334. doi:10.1038/ng.483
- Vargha-Khadem, F., Gadian, D. G., Copp, A., & Mishkin, M.** (2005). *FOXP2* and the neuroanatomy of speech and language. *Nature Reviews Neuroscience, 6*, 131–138. doi:10.1038/nrn1605
- Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., ... Passingham, R. E.** (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences USA, 95*, 12695–12700. doi:10.1073/pnas.95.21.12695
- Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J.** (2012). Five years of GWAS discovery. *American Journal of Human Genetics, 90*, 7–24. doi:10.1016/j.ajhg.2011.11.029
- Visscher, P. M., Goddard, M. E., Derks, E. M., & Wray, N. R.** (2012). Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Molecular Psychiatry, 17*, 474–485. doi:10.1038/mp.2011.65
- Visscher, P. M., Yang, J., & Goddard, M. E.** (2010). A commentary on “Common SNPs Explain a Large Proportion of the Heritability for Human Height” by Yang et al. (2010). *Twin Research and Human Genetics, 13*, 517–524.
- Walley, A. J., Asher, J. E., & Froguel, P.** (2009). The genetic contribution to non-syndromic human obesity. *Nature Reviews Genetics, 10*, 431–442. doi:10.1038/nrg2594
- Wechsler, D.** (1992). *Wechsler Intelligence Scale for Children manual* (UK version [3rd ed.]). London, United Kingdom: The Psychological Corporation.
- Wellcome Trust Case Control Consortium.** (2007, June 7). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*, 661–678. doi:10.1038/nature05911
- Wiig, E. H., Secord, W., & Sabers, D.** (1989). *Test of Language Competence* (expanded ed.). San Antonio, TX: The Psychological Corporation.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M.** (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics, 88*, 76–82. doi:10.1016/j.ajhg.2010.11.011
- Zeesman, S., Nowaczyk, M. J., Teshima, I., Roberts, W., Cardy, J. O., Brian, J., ... Scherer, S. W.** (2006). Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves *FOXP2*. *American Journal of Medical Genetics Part A, 140A*, 509–514. doi:10.1002/ajmg.a.31110