OPEN ACCESS

ORIGINAL ARTICLE

Multivariate genome-wide association study of rapid automatised naming and rapid alternating stimulus in Hispanic American and African–American youth

Dongnhu Thuy Truong,[1] Andrew Kenneth Adams,[2] Steven Paniagua,[2] Jan C Frijters,[3] Richard Boada,[4] Dina E Hill,[5] Maureen W Lovett,[6] E Mark Mahone,[7] Erik G Willcutt,[8,9] Maryanne Wolf,[10] John C Defries,[8,9] Alessandro Gialluisi,[11] Clyde Francks,[12,13] Simon E Fisher, [11,13] Richard K Olson,[8,9] Bruce F Pennington,[14] Shelley D Smith,[15] Joan Bosson-Heenan,[1] Jeffrey R Gruen, [1,2,16] On behalf of the Pediatric, Imaging, Neurocognition, and Genetics Consortium

BMJ

ABSTRACT

Background Rapid automatised naming (RAN) and rapid alternating stimulus (RAS) are reliable predictors of reading disability. The underlying biology of reading disability is poorly understood. However, the high correlation among RAN, RAS and reading could be attributable to shared genetic factors that contribute to common biological mechanisms.

Objective To identify shared genetic factors that contribute to RAN and RAS performance using a multivariate approach.

Methods We conducted a multivariate genome-wide association analysis of RAN Objects, RAN Letters and RAS Letters/Numbers in a sample of 1331 Hispanic American and African–American youth. Follow-up neuroimaging genetic analysis of cortical regions associated with reading ability in an independent sample and epigenetic examination of extant data predicting tissue-specific functionality in the brain were also conducted.

Results Genome-wide significant effects were observed at rs1555839 (p=4.03×10^{-8}) and replicated in an independent sample of 318 children of European ancestry. Epigenetic analysis and chromatin state models of the implicated 70 kb region of 10q23.31 support active transcription of the gene *RNLS* in the brain, which encodes a catecholamine metabolising protein. Chromatin contact maps of adult hippocampal tissue indicate a potential enhancer–promoter interaction regulating *RNLS* expression. Neuroimaging genetic analysis in an independent, multiethnic sample (n=690) showed that rs1555839 is associated with structural variation in the right inferior parietal lobule.

Conclusion This study provides support for a novel trait locus at chromosome 10q23.31 and proposes a potential gene–brain–behaviour relationship for targeted future functional analysis to understand underlying biological mechanisms for reading disability.

INTRODUCTION

Reading disability (RD, also known as developmental dyslexia) is a complex neurodevelopmental disorder characterised by difficulties in reading despite educational opportunity and normal intelligence, and is the most common neurodevelopmental disorder diagnosed in children ages 5–18 years, with a prevalence of 5%–17%.[1] The genetic component of RD is strong, with family and twin-based studies estimating moderate to high heritabilities >0.50.[2] To date, at least nine susceptibility loci for RD (DYX1-9) have been identified through linkage mapping and replicated in several independent populations, along with several candidate risk genes including *KIAA0319*, *DCDC2* and *DYX1C1*.[2] However, variants in these candidate genes and loci do not account for a substantial portion of the estimated heritability.[3] In addition, genome-wide analyses predicated on quantitative measures of reading ability and case–control status for RD have identified novel variants but have had little success attaining genome-wide significance.[4] This could be explained by low sample sizes in the studies, but another reason may be phenotypic heterogeneity of the RD phenotype.

One approach to studying the genetics of complex trait disorders such as RD is to examine relevant endophenotypes. An endophenotype is a quantitative trait measure that is correlated with a disorder or trait of interest due, in part, to shared (pleiotropic) underlying genetic influences. For a phenotype to be considered an endophenotype for a complex trait disorder, it must have a genetic component, be independent of clinical state (affected or unaffected), cosegregate with disorder status in a family and have reproducible measurements.[5] Endophenotypes are conceived as being closer to the underlying biology than the corresponding complex disorder. This could improve the statistical power to identify genetic variants through larger effect sizes. Two potential endophenotypes for reading proficiency that satisfy these criteria are rapid automatised naming (RAN) and rapid alternating stimulus (RAS).

RAN tasks require sequential naming of visually presented, familiar items (eg, objects, letters, numbers or colours) as quickly and accurately as possible. RAN has been described as capturing a 'mini-circuit' of the reading network that taps

into cognitive subprocesses critical for reading, such as automatic attentional processing.[6] Reading and RAN performance are moderately correlated (0.28–0.57), with deficits in RAN performance observed in 60%–75% of individuals with RD.[7–9] Furthermore, performance on RAN in kindergarten is predictive of later reading fluency and is stable through elementary school.[10 11] RAN deficits even persist into adulthood.[12 13]

RAS tasks are similar in overall design to RAN tasks but involve sequential naming of items across *alternating* stimulus categories as quickly and accurately as possible (ie, alternating letters and numbers). RAS tasks evaluate the ability to direct attention while performing an automatic task like sequential naming.[14] RAN and RAS are highly correlated (0.85), and like RAN, RAS performance can differentiate poor from typically developing readers and is highly predictive of later reading ability.[14–16]

RAN has a strong genetic component with heritability estimates ranging from 0.46 to 0.65.[17 18] To date, several family-based studies have identified novel genetic loci linked to RAN and RAS performance, with some studies also supporting linkage to other previously identified DYX loci.[19] The majority of studies focused on linkage mapping, while the limited number of association studies on RAN and RAS directly targeted the strongest candidate genes and/or variants identified for RD from the available literature. However, a recent genome-wide meta-analysis and multivariate analysis of RAN Digits, RAN Letters and RAN Pictures found significant associations with rs17663182 in chromosome 18q12.2 in a sample of 2562–3468 European individuals.[20]

Notably, genetic studies of RAN, RAS and RD were previously conducted on samples of largely European descent. This is problematic because for causal variants, differences in observed allele frequencies and linkage disequilibrium (LD) between populations could lead to associations observed in one population but not in another. Furthermore, lack of diversity in association studies leads to missed opportunities to identify novel trait loci present in certain populations, thus limiting the identification of biological clues surrounding the aetiology of a complex disorder.[21] Using a single population to identify disease-associating regions constrains the ability to generalise relevant findings to other populations, and severely limits the benefits of biomedical research to only those populations with selected demographics.[22 23]

The present study aims to identify shared genetic factors that contribute to covariance across RAN Objects, RAN Letters and RAS Letters/Numbers using a multivariate genome-wide association study in 1331 unrelated children in the Genes, Reading, and Dyslexia (GRaD) study, a case–control sample of RD in Hispanic American and African–American youth. It builds on previous studies that investigated the genetic underpinnings of RAN and RAS performance as univariate traits. However, a univariate design can miss underlying covariance shared across two or more correlated traits and therefore has low sensitivity for detecting shared genetic factors.[24] We hypothesise that the correlations between RAN and RAS performance can partially be attributed to shared genetic factors.[17] To test this hypothesis, we apply a multivariate genetic analysis to leverage covariance across phenotypes and to increase statistical power to identify the presence of genetic pleiotropy across different RAN and RAS tasks.[24]

METHODS
Subjects, recruitment and DNA collection
The GRaD study is a multisite, case–control study of RD in minority youth across the USA, Canada and Puerto Rico. Subject recruitment and collection of phenotype data and DNA were completed for 1432 unrelated children. Of these subjects, 1331

children who passed genotyping quality control were included in the analysis.

The Colorado Learning Disability Research Center (CLDRC) cohort is a family-based sample selected for RD, Attention-Deficit/Hyperactivity Disorder (ADHD) and other learning disabilities. DNA and phenotype data from 318 participants, one from each twinship/sibship, were randomly selected for analysis from a total of 749 participants.

The Pediatric Imaging, Neurocognition, and Genetics (PING) study is a cross-sectional sample of typically developing children ranging in age from 3 to 20 years old. Imaging genetics analysis was conducted on 690 subjects with complete phenotype and genotype data.

Detailed descriptions of phenotypes, recruitment, inclusion and exclusion criteria, self-reported demographic breakdown, and DNA collection for each of the three samples are reported in the online supplementary methods and tables S1–S2.

Statistical analysis
A multivariate genome-wide association study (GWAS) of RAN Objects, RAN Letters and RAS Letters/Numbers was conducted using the R package MultiPhen.[25] Joint models were corrected for the first three principal components (PC; online supplementary figure S1 and S2A) to correct for population stratification, sex, age and socioeconomic status (SES). A genome-wide significance threshold of $p < 5 \times 10^{-8}$ was used to correct for multiple testing.[26]

Multivariate GWAS results were verified using a follow-up univariate GWAS on a latent variable composite of RAN Objects, RAN Letters and RAS Letters/Numbers in PLINK V.1.9 in the GRaD sample stratified by assignment to a 'Hispanic American' and 'African American' cluster based on genetic similarity to 1000 genomes reference populations AMR (Ad Mixed American) and ASW (African–Americans in the Southwest USA), respectively (online supplementary figure S3). For the Hispanic American cluster, the univariate GWAS model was corrected for the effects of age, sex, SES and the first three ethnic-specific PCs to correct for intraethnic stratification (online supplementary figure S2B). For the African–American cluster, the model was corrected for age, sex, SES and the first two ethnic-specific PCs (online supplementary figure S2C). A meta-analysis was then conducted in METAL using summary statistics from the latent naming speed GWAS.[27] Univariate analysis of covariance (ANCOVA) testing the top SNP against measures of reading fluency (Test of Word Reading Efficiency (TOWRE) and Woodcock-Johnson 3 (WJ-III)) in the full GRaD sample was conducted in SPSS V.24. Models corrected for the effects of age, sex, SES and the first three genome-wide PCs (online supplementary figure S2A).

Replication was conducted using the CLDRC sample through joint analysis of RAN Colours, RAN Pictures, RAN Letters and RAN Numbers at SNPs identified in the GRaD discovery analyses. Models were corrected for the effects of age, sex and the first genome-wide PC (online supplementary table S2).

Detailed descriptions of statistical analyses are reported in the online supplementary methods.

Bioinformatic analysis
All analyses were conducted using positions mapping to genome build GRCh37/hg19.

Predicted tissue-specific functionality in the genome was estimated through the presence of well-characterised DNase 1 hypersensitivity sites and histone marks and using GenoSkyline.[28] Precalculated, genome-wide, tissue-specific, posterior

probability scores (GenoSkyline (GS) scores) for the whole brain and subregions of the brain (angular gyrus, prefrontal cortex, cingulate gyrus, anterior caudate, hippocampus, inferior temporal gyrus and substantia nigra) were obtained from the GenoSkyline database.

Follow-up analysis of epigenetic data from the Roadmap Epigenomics Project was conducted to identify predicted chromatin state in a genomic region of interest across specific tissue types. The 18-state model was evaluated in angular gyrus, prefrontal cortex, cingulate gyrus, anterior caudate, hippocampus, inferior temporal gyrus and substantia nigra. Data were visualised using the Washington University in St Louis EpiGenome Browser V.42.

The 3D Genome Browser was used to visualise publicly available chromatin contact maps at 10 kb resolution from adult hippocampal tissue.[29 30]

Detailed descriptions of the bioinformatic analysis are reported in the online supplementary methods.

Neuroimaging genetic analysis

Neuroimaging genetics analysis in the PING sample was conducted using a linear regression association test in PLINK V.1.9. All analyses were corrected for the effects of age, sex, handedness, scanner device,[31] intracranial volume, highest parental education, family income and the first four genome-wide PCs to correct for population stratification (online supplementary figure S2E). Left and right cortical volumes across seven regions of interest associated with the canonical reading network were tested for genetic association (online supplementary table S3). The left and right hippocampus was also included based on follow-up analyses, for a total of 16 assessed regions.

RESULTS

Multivariate and univariate associations for RAN and RAS

There are significant Pearson correlations across RAN Objects, RAN Letters and RAS Letters/Numbers in the GRaD sample ranging from 0.621 to 0.781 ($p<1\times10^{-140}$; online supplementary table S4), providing the rationale to use a multivariate GWAS. Multivariate GWAS of RAS Letter/Numbers, RAN Objects and RAN Letters reveals a genome-wide significant effect for rs1555839 ($p=4.03\times10^{-8}$; genomic inflation factor [λ]=1.047; figure 1, online supplementary figure S4, table 1) located approximately 5 kb upstream from *ribosomal protein L7 pseudogene 34* (*RPL7P34*). *RPL7P34* is a long non-coding RNA (lncRNA) on chromosome 10 mapping between the genes *LIPF* and *LIPJ*. Additional SNPs approaching significance ($p<1\times10^{-5}$) cluster within a 70 kb region of chromosome 10q23.31 that spans *RPL7P34* and a gene called renalase (*RNLS*; MIM: 609360) (figure 1B). To determine whether specific variables may be driving the multivariate signal, we conducted an independent, post-hoc, univariate examination of RAN Objects, RAN Letters and RAS Letters/Numbers on the top SNPs identified in the multivariate analysis. The independent analysis shows that the SNPs located on chromosome 10q23.31 have an effect across all RAN and RAS tasks (table 1, online supplementary figure S5–S7). The strongest association in the chromosome 10q23.31 region is with SNP rs1555839 ($\beta=-0.31$, $p=1.71\times10^{-9}$) for RAS Letters/Numbers.

To further validate the findings from the multivariate GWAS of the full GRaD cohort, we performed a GWAS meta-analysis of Hispanic American (AMR cluster; n=883) and African–American (ASW cluster; n=441) participants in the GRaD sample on a latent naming speed variable (online supplementary figure S3). rs1555839 and rs6963842, previously implicated in the multivariate GWAS

of RAN Objects, RAN Letters and RAS Letters/Numbers, show suggestive association ($p<1\times10^{-6}$; λ=1.013; online supplementary figure S8, table S5). rs701825 on chromosome 10, also implicated in the discovery multivariate GWAS, is also highly associated with naming speed latent variable in the meta-analysis. However, the significant heterogeneity statistic indicates a difference in the magnitude of the effect between Hispanic Americans and African–Americans on naming speed, but the direction of effect remains the same (online supplementary table S5). Both the multivariate GWAS and the latent naming speed meta-analysis implicate the same regions of chromosome 7q31.1 and 10q23.31, providing independent evidence that findings are not from hidden population stratification.

Due to significant correlations observed between RAN/RAS tasks and measures of word reading (r=0.325–578, $p<1\times10^{-32}$; online supplementary table S4), we evaluated whether allelic variation at rs1555839 was associated with mean differences in performance for TOWRE and WJ-III Basic Reading. Univariate ANCOVA shows a significant effect of rs1555839 with TOWRE ($F[1,1281]=12.50$, $p=4.21\times10^{-4}$, $\eta_p^2=0.01$) and WJ-III Basic Reading ($F[1,1278]=13.47$, $p=2.53\times10^{-4}$, $\eta_p^2=0.01$) (figure 2). Overall, performance on reading fluency and word reading is worse in the presence of the C allele at rs1555839, which is the same allele and direction of effect associated with reduced scores on RAN and RAS.

Replication of multivariate GWAS and latent naming speed GWAS meta-analysis

Next, we brought forward all three SNPs that achieved genome-wide significance or were suggestive in both the discovery multivariate GWAS and the latent naming speed GWAS meta-analysis for replication in the CLDRC. Assessments of RAN Colours, RAN Letters, RAN Pictures and RAN Numbers were available in 318 unrelated participants. RAN Objects and RAS Letters/Numbers were not assessed in the CLDRC. Multivariate analyses of RAN Colours, RAN Letters, RAN Pictures and RAN Numbers show a significant replication of rs701825 on chromosome 10, while we observe marginal significance with rs1555839. rs6963842 on chromosome 7 does not replicate (table 2). Post-hoc univariate association analysis shows a significant replication of rs1555839 and rs701825 for RAN Letters (table 2, online supplementary table S6). The direction of effect on RAN Letters is the same for both cohorts (online supplementary table S6).

Bioinformatic analysis

RNLS, a gene located within 50 kB of rs1555839, is expressed in all regions of the brain (Genotype-Tissue Expression (GTEx) median >1 transcripts per million (TPM))[32] and throughout the human lifespan (BrainSpan).[33] *LIPJ* is expressed at low levels in the cerebellar hemisphere and cerebellum (median <0.7 TPM). *LIPF* is not expressed in the brain in either the GTEx or BrainSpan databases.

We used GenoSkyline to evaluate the presence of predicted functional regions of the genome surrounding the 70 kb region of chromosome 10q23.31 spanning rs2576167–rs701825 implicated in the discovery multivariate GWAS using epigenomic annotations from the National Institutes of Health Roadmap Epigenomics Project and Encyclopedia of DNA Elements (ENCODE). Within all sampled tissues, GenoSkyline predicts tissue-specific functionality surrounding our top SNPs. The first major peak (chr10:90341871–90343706) encompasses the promoter region of *RNLS*, reaching the maximum GS score of 1 for all tissues (table 3, online supplementary figure S9). The
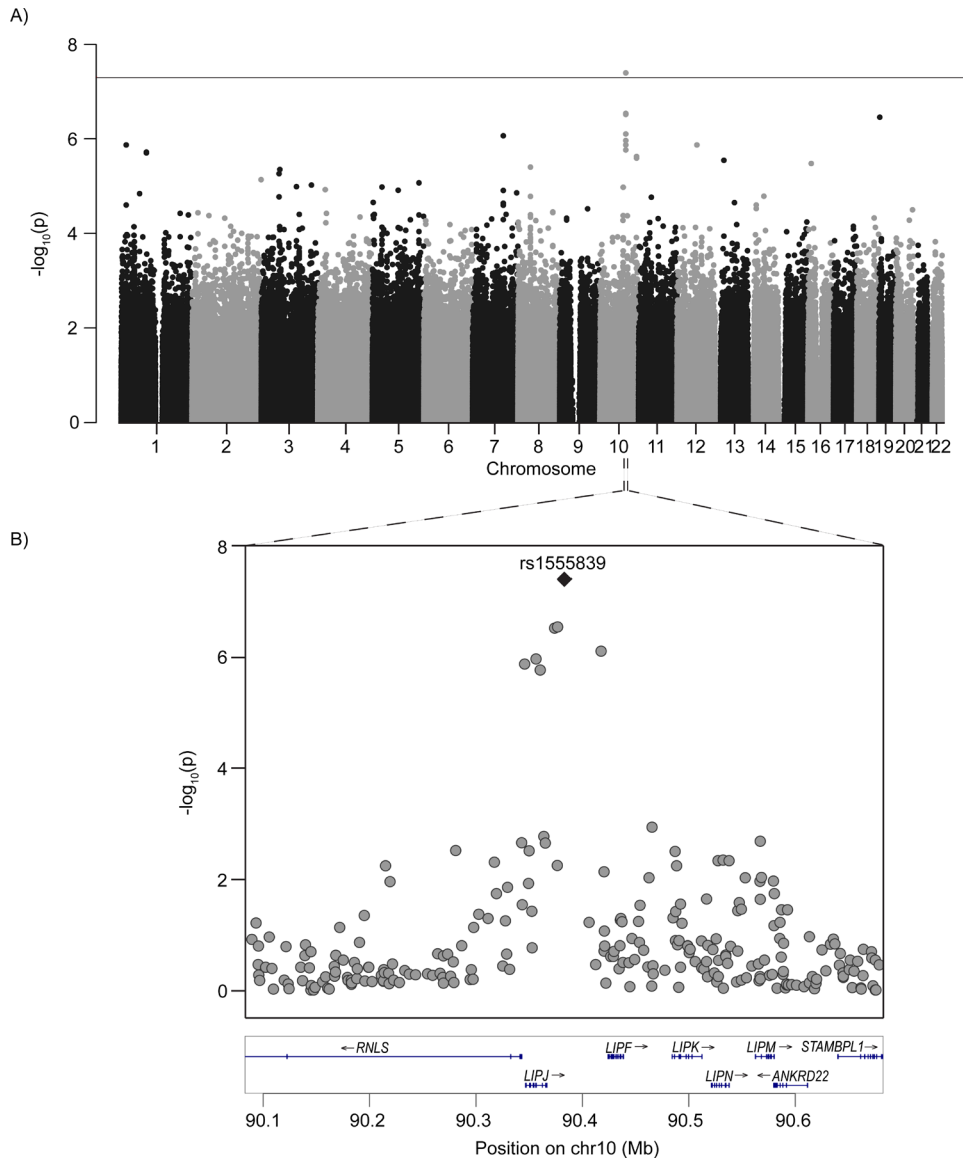
Figure 1 (A) Manhattan plot summarising the results of the multivariate genome-wide association study (GWAS) for RAN Objects, RAN Letters and RAS Letters/Numbers and (B) LocusZoom plot of genomic region surrounding genome-wide significant SNP, rs1555839. The genome-wide-associated SNP, rs1555839, is marked by a black diamond. lncRNA pseudogene, *RPL7P34* (chr10: 90,377,980–90,378,691), is not represented in the LocusZoom plot. Grey line: Bonferroni correction for multiple testing, $p < 5 \times 10^{-8}$. lncRNA, long non-coding RNA; RAN, rapid automatised naming.

**Table 1** Multivariate genome-wide association study results

					P value				
Marker	CHR	BP	Minor allele	MAF	Joint model	RAN Objects	RAN Letters	RAS Letters/Numbers	Gene
rs1555839	10	90 382 820	C	0.16*	4.03×10^{-8}†	7.06×10^{-5}	6.20×10^{-6}	1.71×10^{-9}†	*RPL7P34*
rs7913742	10	90 376 864	C	0.149	2.89×10^{-7}	2.94×10^{-4}	3.01×10^{-5}	1.33×10^{-8}†	*RPL7P34*
rs10749593	10	90 374 057	C	0.153	3.04×10^{-7}	4.30×10^{-4}	2.21×10^{-5}	1.31×10^{-8}†	*RPL7P34*
rs113424746	19	2 920 806	A	0.056	3.49×10^{-7}	8.83×10^{-5}	9.73×10^{-5}	0.334	*lnc-ZNF77-1*
rs701825	10	90 417 547	G	0.132	7.92×10^{-7}	5.47×10^{-4}	1.35×10^{-6}	6.15×10^{-8}	–
rs6963842	7	107 634 989	G	0.471	8.57×10^{-7}	2.43×10^{-4}	1.84×10^{-7}	2.08×10^{-6}	*LAMB1*

Top associated markers with $p < 1 \times 10^{-6}$ in the GRaD study that were identified using a joint model of RAN Objects, RAN Letters and RAS Letters/Numbers. Results from follow-up, independent univariate analysis of RAN Objects, RAN Letters and RAS Letters/Numbers at each marker are also represented. Markers were assigned to genes if they fell within the canonical gene body as described by 1000 Genomes Project Phase 3 (V.80 GRCh37).

*For self-reported African–Americans in the GRaD sample, rs1555839 has a MAF of 0.129. For self-reported Hispanic Americans, the MAF of rs1555839 is 0.178.

†Genome-wide significant ($p < 5 \times 10^{-8}$).

BP, base position; CHR, chromosome; GRaD, Genes, Reading, and Dyslexia study; MAF, minor allele frequency; RAN, rapid automatised naming.
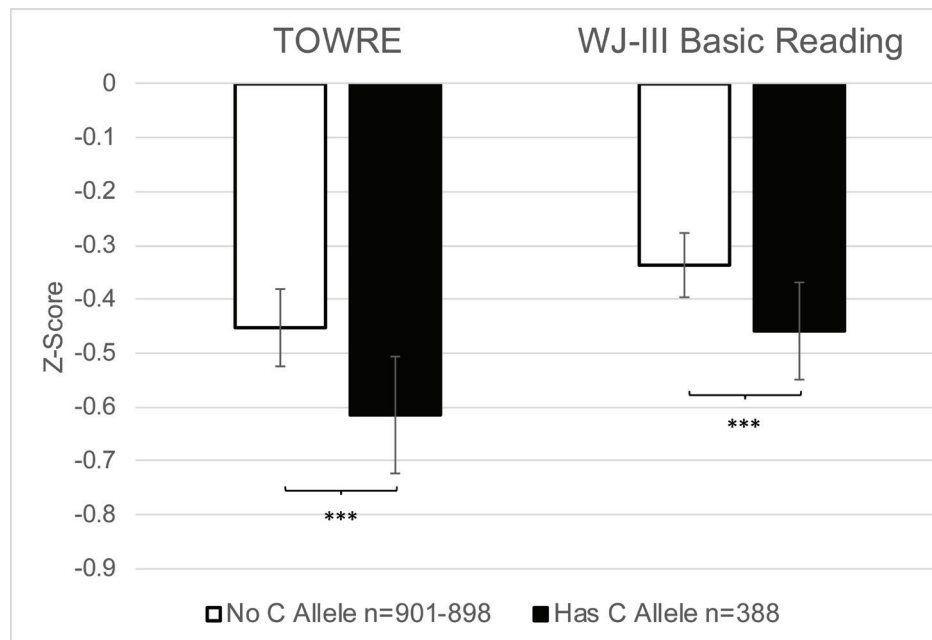
Figure 2 Mean differences in TOWRE and WJ-III Basic Reading performance by allelic variation at rs1555839 in the GRaD study. In the presence of the C allele, there is a significant reduction in TOWRE and WJ-III Basic Reading performance, while correcting for the effects of age, sex and population stratification (the first three principal components). Error bars represent 95% CI. TOWRE n=901, WJ-III n=898; ***p<0.001. GRaD, Genes, Reading, and Dyslexia study.

second major peak (chr10: 90376471–90376828) is within 2 kB of *RPL7P34*, reaching a maximum GS score of 1 in the whole brain, 0.886 in the hippocampus, 0.629 in the anterior caudate and 0.880 in the cingulate gyrus (figure 3A, online supplementary figure S9). GS scores for the angular gyrus, dorsolateral prefrontal cortex, inferior temporal gyrus and substantia nigra are less than 0.5 (online supplementary figure S9).

To determine the chromatin state of the 70 kb region of chromosome 10q23.31 that contributes to high posterior probability scores for tissue-specific functionality in the genome, we evaluated the 18-state model previously generated by the Roadmap Epigenomics Project. The region closely associated with the most significant SNPs in the GRaD discovery analysis (rs10749593, rs7913742, rs1555839) flanks an active enhancer site in the hippocampus and cingulate gyrus. These SNPs also flank regions of heterochromatin containing ZNF genes and repeats in the cingulate gyrus and inferior temporal lobe (figure 3B). In addition, an active transcription start site for *RNLS* is predicted in all brain regions (figure 3B).

Examination of LD blocks across different ethnic groups sampled in 1000 Genomes Phase 3 that are represented in the GRaD sample (Utah Residents with Northern and Western European Ancestry (CEU), Ad Mixed American (AMR) and Yoruba in Ibadan, Nigeria (YRI)) shows a single LD block spanning approximately 39 kb encompassing the segment that spans rs2576167–rs7913742 but not including rs1555839 or rs701825 (figure 3C; online supplementary table S7). However, in CEU and AMR, a single LD block spans the 70 kb segment that spans all SNPs of interest (figure 3C; online supplementary table S7). Values for D' for CEU, AMR and YRI populations range from 1 to 0.87, indicating high correlation across these SNPs in chromosome 10q23.31.[34 35] Taken together, these data suggest a similar genetic architecture across ethnic populations within this region.

To evaluate whether a potential enhancer–promoter interaction exists between the predicted enhancer site near *RPL7P34* and active transcription start site for *RNLS* in the hippocampus, we examined extant Hi-C contact maps publicly available through the 3D Genome Browser.[29 30] There is evidence that a chromatin interaction is present between a region containing the predicted enhancer site and the active transcription start site at a 10 kb resolution in the adult hippocampus (figure 3D). No extant Hi-C contact maps are publicly available for the cingulate gyrus to evaluate a potential enhancer–promoter interaction in this region.

Table 2 CLDRC multivariate replication

Marker	CHR	BP	Minor allele	MAF*	P value				
					Joint model	RAN Pictures	RAN Colours	RAN Letters	RAN Numbers
rs6963842†	7	107 634 989	T	0.495	0.360	0.963	0.644	0.856	0.244
rs1555839†	10	90 382 820	C	0.319	0.018	0.256	0.019	2.38×10⁻³‡	0.086
rs701825†	10	90 417 547	G	0.282	4.60×10⁻³‡	0.562	4.09×10⁻³‡	9.32×10⁻⁴‡	0.04

Results from the joint analysis of RAN Colours, RAN Pictures, RAN Letters and RAN Numbers across the top associated markers from the GRaD multivariate and meta-analytic discovery analyses. Results from follow-up, independent univariate analysis of RAN Pictures, RAN Colours, RAN Letters and RAN Numbers at each marker are also represented.
*MAF based on allele frequencies within the CLDRC sample.
†Imputed markers.
‡Survives Bonferroni correction for multiple testing (p<0.05/3=0.0161).
BP, base position; CHR, chromosome; CLDRC, Colorado Learning Disability Research Center; GRaD, Genes, Reading, and Dyslexia study; MAF, minor allele frequency; RAN, rapid automatised naming.

Table 3 Association results of rs1555839 tested against reading-related regions of interest in the PING study

	Left			Right		
	Beta	SE	P value	Beta	SE	P value
Pars opercularis	−33.65	66.87	0.615	−10.56	59.98	0.860
Pars orbitalis	8.777	26.69	0.742	−23.26	32	0.468
Pars triangularis	−68.95	52.30	0.188	−73.55	60.86	0.227
Supramarginal gyrus	−186.1	119.4	0.120	67.43	117.2	0.565
Inferior parietal cortex	−68.74	138.1	0.619	−437	143.8	$2.47×10^{-3}$*
Inferior temporal gyrus	−191.6	117.7	0.104	−53.74	119	0.652
Fusiform gyrus	−118.9	95	0.211	−185	85.32	0.030
Hippocampus†	−53.64	20.6	$9.42×10^{-3}$	−34.57	21.54	0.109

Statistical models were corrected for population stratification (first four principal components), age, sex, scanner device, handedness, intracranial volume, parental education and family income.
*Survives Bonferroni correction for multiple testing ($p<0.05/16=3.13×10^{-3}$).
†Structure not part of the canonical cortical reading network.

PING neuroimaging genetic analysis

Following replication of rs1555839 and epigenetic analysis showing predicted functionality in brain tissue within the region of chromosome 10q23.31, we conducted a targeted neuroimaging genetics analysis in 690 typically developing children of various ancestral backgrounds from the PING sample. The C allele of rs1555839 is associated with lower cortical volume in the right inferior parietal cortex, after correcting for multiple testing (β=−437, p=$2.43×10^{-3}$; table 3). There is a suggestive association between rs1555839 and lower cortical volumes in the left hippocampus (β=−53.64, p=$9.42×10^{-3}$; table 3).

DISCUSSION

The present study is one of the first to examine the genetics of reading-related traits in an admixed population of Hispanic American and African–American children. Here, we describe a region of chromosome 10q23.31 that exceeds genome-wide significance with pleiotropic effects across RAN Objects, RAN
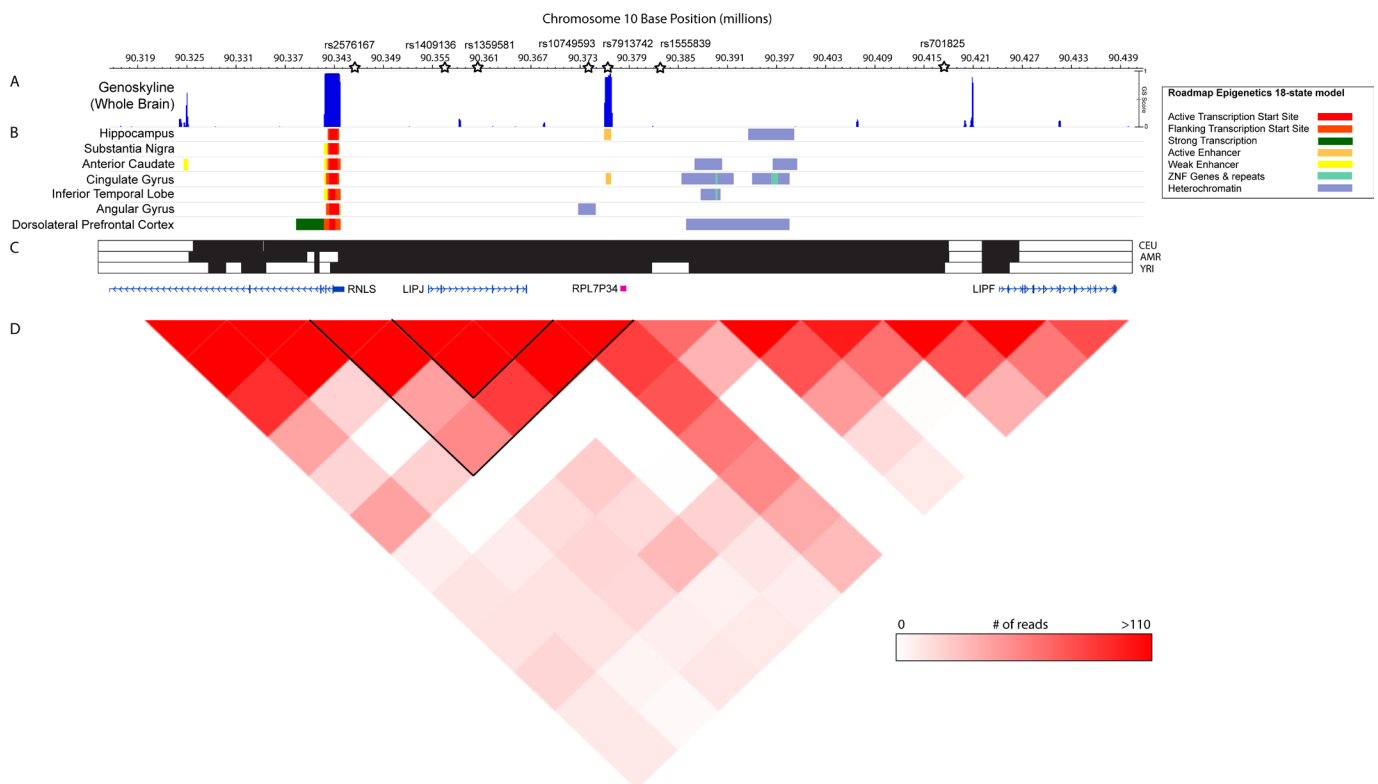
Figure 3 Epigenetic examination of chromosome 10q23.31 containing the top-performing SNPs in the GRaD discovery analysis. (A) Plot of GS scores obtained from GenoSkyline indicating the posterior probability for functionality in the brain at each genomic locus. A GenoSkyline (GS) score of 1 indicates a high probability for functionality, while a GS score of 0 suggests no functionality. (B) 18-state chromatin model from Roadmap Epigenomics Project showing predicted chromatin states based on the presence of H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3 and H3K27ac sampled across different regions of the brain. (C) Location of genes implicated in the GRaD discovery analysis, and location of linkage disequilibrium blocks (black bars) in CEU, AMR and YRI populations in 1000 Genome Project Phase 3. (D) Chromatin contact map of adult hippocampal tissue showing regions of chromatin interactions at 10 kB resolution. Open stars represent the position of top chromosome 10q23.31 SNPs in the GRaD multivariate genome-wide association study (GWAS) analysis. AMR, Ad Mixed American; GRaD, Genes, Reading, and Dyslexia study.

Letters and RAS Letters/Numbers—highly correlated tasks that are predictive of later reading outcome and reading ability in children and adults.[6] We validated the GWAS results with a genome-wide meta-analysis of a univariate latent naming speed variable that implicates the same region of chromosome 10. Although results from the genome-wide meta-analysis did not attain genome-wide significance, this could reflect the loss of power by stratifying the GRaD cohort into smaller sample sizes, or perhaps the loss of power to identify genetic factors with pleiotropic effects using a univariate approach.[24]

Multivariate replication in a European sample further supports a shared effect from a small segment of chromosome 10 across RAN tasks. Within the CLDRC sample, only one endophenotype, RAN Letters, overlapped with the GRaD sample in the present analysis. Although this is not a perfect replication of the primary analysis, the high correlation across different RAN and RAS tasks[16 17] and the underlying hypothesis of the overall study to identify shared genetic factors contributing to general RAN and RAS performance justify replicating in the CLDRC sample. Furthermore, the significant and same direction of effect we see with chromosome 10q23.31 SNPs, rs1555839 and rs701825, on RAN Letters in a different ethnic group suggests that these findings can be generalised across more than one population. This is supported by LD maps showing a similar underlying genetic architecture of this region across representative 1000 Genomes Project populations in CEU and AMR, as well as the high correlation across SNPs in CEU, AMR and YRI.

The most highly associated SNP in the RAS/RAN genome-wide screen, rs1555839, is located 5 kb upstream from the lncRNA, *RPL7P34*. The function of *RPL7P34* is unknown, and the roles of lncRNAs in the genome are poorly understood. lncRNAs form a class of non-protein coding transcripts over 200 nucleotide bases long, and have characteristics that suggest functionality including tissue-specific expression, regulated expression, and regulation of gene expression and their networks.[36] lncRNAs are reported to recruit transcription factors and interact with chromatin modifiers, suggesting that they facilitate epigenetic regulation of the genome.[37] Approximately 40% of lncRNAs are expressed in the brain and are hypothesised to play critical roles in neural development such as neural proliferation and differentiation.[38]

Epigenetic examination of the region surrounding *RPL7P34* shows a predicted active enhancer site based on the presence of H3K4me1 and H3K27ac histone modifiers specifically in the hippocampus and cingulate gyrus. It is possible that *RPL7P34* plays a role in the recruitment of proteins that bind to enhancer sites, which promote transcription of nearby genes. The nearby gene *RNLS*, approximately 30 kb away, encodes the closest predicted transcription start site within the brain. Examination of chromatin contact maps shows that *RPL7P34* and the promoter region of *RNLS* interact in the hippocampus, suggesting a potential enhancer–promoter interaction. Taken together, these results suggest that the enhancer region downstream from *RPL7P34* could regulate *RNLS*. However, further targeted functional analyses to experimentally validate the molecular and epigenetic function of *RPL7P34* and the nearby enhancer site would have to be conducted. It is also important to note that the epigenetic and Hi-C data in the brain analysed by the Roadmap Epigenomics Project were obtained from only two individuals.[39] Although these data offer some clues to potential functionality, they should be interpreted with caution, considering the number of subjects sampled in the Roadmap Epigenomics Project, and lack of substantiation by a cell-based assay directly showing enhancer function in this region.

RNLS, implicated in our GWAS and epigenetic analysis, encodes a flavin adenine dinucleotide-dependent amine oxidase, called renalase, which metabolises catecholamines such as norepinephrine and dopamine.[40 41] Renalase is known to be secreted by the kidneys, circulates in blood, and modulates cardiac function and blood pressure.[41] Gene expression data from GTEx and BrainSpan also suggest that renalase is present in the human central nervous system, including the cortex, hippocampus and cingulate gyrus.[40] In vitro analysis of monoamine oxidase activity of renalase shows that it is most efficient in metabolising dopamine, but is also effective in metabolising epinephrine and norepinephrine.[41] Two major neuromodulatory systems in the brain, dopaminergic and noradrenergic, play important functions in motivation, attention, learning and memory formation—especially in the hippocampus and cingulate gyrus—but the functional role of renalase in metabolising neurotransmitters in the brain is currently unclear.[40 42] However, genetic variants in *RNLS* have been associated with schizophrenia, a neuropsychiatric disorder that may be caused by an imbalance in neurotransmission. Specifically, a recent study showed that human-induced pluripotent stem cell (hiPSC)-derived neurons from patients with schizophrenia had altered catecholamine release relative to control hiPSC-derived neurons.[43]

This is the first study to implicate *RNLS* in reading-related domains. However, it is not the first gene associated with the metabolism of catecholamines. *COMT* encodes catechol-O-methyltransferase, which also degrades catecholamines in the brain, and has been associated with variation in reading-related tasks as well as functional networks associated with reading ability.[44 45] There is also evidence showing neurochemical differences between poor readers and typically developing controls, suggesting that there could be alterations in how neurotransmitters are metabolised in reading impaired individuals.[46]

Neuroimaging-genetic analysis indicates that rs1555839 is associated with variation in cortical volume in the right inferior parietal cortex. Limited studies have been conducted on structural neuroanatomical correlates of RAN, but there is evidence that RAN performance is associated with grey matter volumes in bilateral occipital-temporal and parietal-frontal regions, which include the right inferior parietal cortex.[47] In addition, functional variation in the right inferior parietal cortex has also been implicated in several reading-related tasks and RD.[48]

The small sample size is a limitation of this study for GWAS analyses. However, based on power calculations for a multivariate GWAS, we had modest power of 0.68 to identify a genetic variant with a minor allele frequency of 0.16, effect size ranging from 0.03 to 0.013, and cross-phenotype correlation of 0.7, in a sample with 1263 individuals (both phenotype and genotype data) and $\alpha = 5 \times 10^{-8}$—all parameters applicable to this study and the observed SNPs showing most significant associations.[49] While hidden population admixture could be confounding our analyses, the calculated genomic inflation factor is within acceptable ranges ($\lambda < 1.05$) after correction, suggesting adequate control for population stratification. Furthermore, we split the GRaD cohort by genetic ancestry using k-means clustering with AMR and ASW reference populations and conducted a genome-wide meta-analysis on a latent naming speed variable derived from performance across RAN Objects, RAN Letters and RAS Letters/Numbers, and we replicated the findings in an independent sample of European ancestry, attaining similar results across all modes of analysis.

In conclusion, by leveraging different data sources and types across neuroimaging and epigenetic data, we identified and replicated association of a region of chromosome 10q23.31

with pleiotropic effects across RAN, RAS and reading abilities in a sample of Hispanic American and African–American youth. There is converging evidence that non-coding regions of the genome have an impact on reading-related traits, and the identification of an lncRNA associated with a reading endophenotype lends additional support.[50] The association of *RNLS* also reinforces the hypothesis that alterations in neurochemical modulation in the brain could contribute to impairments in reading performance. However, further functional assays must be conducted to shed mechanistic light on the pathways involved.

This study highlights the importance of studying the genetic architecture of RD across diverse ethnicities and how genetic effects and variants may differ (or be similar) across populations. This is critical in our understanding of the biological mechanisms that contribute to RD, and is necessary for presymptomatic identification and development of precision intervention strategies that are relevant to all populations.

Electronic database information

- ► 1000 Genomes Project (http://www.internationalgenome.org/).
- ► GenoSkyline (http://genocanyon.med.yale.edu/GenoSkyline).
- ► National Institutes of Health Roadmap Epigenomics Project (http://egg2.wustl.edu/roadmap/web_portal/index.html).
- ► 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/view.php).
- ► Pediatric Imaging, Neurocognition, and Genetics Portal (http://ping.chd.ucsd.edu/).
- ► FreeSurfer (http://surfer.nmr.mgh.harvard.edu/).
- ► LDlink (https://analysistools.nci.nih.gov/LDlink/?tab=home).

Author affiliations
[1]Pediatrics, Yale University School of Medicine, New Haven, Connecticut, USA
[2]Genetics, Yale University School of Medicine, New Haven, Connecticut, USA
[3]Department of Child and Youth Studies, Brock University, St Catharines, Ontario, Canada
[4]Department of Pediatrics-Neurology, University of Colorado School of Medicine, Aurora, Colorado, USA
[5]Department of Psychiatry and Behavioral Sciences, University of New Mexico, Albuquerque, New Mexico, USA
[6]Neurosciences & Mental Health Program, The Hospital for Sick Children, Toronto, Ontario, Canada
[7]Department of Neuropsychology, Kennedy Krieger Institute, Baltimore, Maryland, USA
[8]Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA
[9]Department of Psychology and Neuroscience, University of Colorado, Boulder, Colorado, USA
[10]Eliot-Pearson Department of Child Study and Human Development, Tufts University, Medford, Massachusetts, USA
[11]Language and Genetics, Max-Planck-Institut fur Psycholinguistik, Nijmegen, The Netherlands
[12]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[13]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[14]Department of Psychology, University of Denver, Denver, Colorado, USA
[15]Pediatrics, University of Nebraska Medical Center, Omaha, Nebraska, USA
[16]Investigative Medicine, Yale University School of Medicine, New Haven, Connecticut, USA

REFERENCES

1 Shaywitz SE, Dyslexia SBA. Specific reading disability). *Biological psychiatry* 2005;57:1301–9.
2 Carrion-Castillo A, Franke B, Fisher SE. Molecular genetics of dyslexia: an overview. *Dyslexia* 2013;19:214–40.
3 Plomin R. Child Development and Molecular Genetics: 14 Years Later. *Child Dev* 2013;84:104–20.
4 Carrion-Castillo A, van Bergen E, Vino A, van Zuijen T, de Jong PF, Francks C, Fisher SE. Evaluation of results from genome-wide studies of language and reading in a novel independent dataset. *Genes Brain Behav* 2016;15:531–41.
5 Gottesman II, Gould TD. The endophenotype concept in psychiatry: Etymology and strategic intentions. *AJP* 2003;160:636–45.
6 Norton ES, Wolf M. Rapid automatized naming (Ran) and reading fluency: implications for understanding and treatment of reading disabilities. *Annu Rev Psychol* 2012;63:427–52.
7 Waber DP, Forbes PW, Wolff PH, Weiler MD. Neurodevelopmental characteristics of children with learning impairments classified according to the double-deficit hypothesis. *J Learn Disabil* 2004;37:451–61.
8 Wolf M, O'Rourke AG, Gidney C, Lovett M, Cirino P, Morris R. The second deficit: an investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Read Writ* 2002;15:43–72.
9 Katzir T, Kim YS, Wolf M, Morris R, Lovett MW. The varieties of pathways to dysfluent reading: comparing subtypes of children with dyslexia at letter, word, and connected text levels of reading. *J Learn Disabil* 2008;41:47–66.
10 Torppa M, Lyytinen P, Erskine J, Eklund K, Lyytinen H, Development L. Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *J Learn Disabil* 2010;43:308–21.
11 Schatschneider C, Fletcher JM, Francis DJ, Carlson CD, Foorman BR. Kindergarten prediction of reading skills: a longitudinal comparative analysis. *J Educ Psychol* 2004;96:265–82.
12 Vukovic RK, Wilson AM, Nash KK. Naming speed deficits in adults with reading disabilities: a test of the double-deficit hypothesis. *J Learn Disabil* 2004;37:440–50.
13 van den bos KP, Zijlstra BJH, lutje Spelberg HC. Life-span data on Continuous-Naming speeds of numbers, letters, colors, and Pictured objects, and Word-Reading speed. *Scientific Studies of Reading* 2002;6:25–49.
14 Wolf M. Rapid alternating stimulus naming in the developmental dyslexias*1. *Brain and Language* 1986;27:360–79.
15 Altemeier LE, Abbott RD, Berninger VW. Executive functions for reading and writing in typical literacy development and dyslexia. *J Clin Exp Neuropsychol* 2008;30:588–606.

16 Compton DL, Olson RK, DeFries JC, Pennington BF. Comparing the relationships among two different versions of alphanumeric rapid Automatized naming and word level reading skills. *Scientific Studies of Reading* 2002;6:343–68.

17 Davis CJ, Gayán J, Knopik VS, Smith SD, Cardon LR, Pennington BF, Olson RK, DeFries JC. Etiology of reading difficulties and rapid naming: the Colorado twin study of reading disability. *Behav Genet* 2001;31:625–35.

18 Petrill SA, Thompson LA, Deater-Deckard K, Dethorne LS, Schatschneider C. Genetic and environmental effects of serial naming and phonological awareness on early reading outcomes. *J Educ Psychol* 2006;98:112–21.

19 Rubenstein KB, Raskind WH, Berninger VW, Matsushita MM, Wijsman EM. Genome scan for cognitive trait loci of dyslexia: rapid naming and rapid switching of letters, numbers, and colors. *Am. J. Med. Genet.* 2014;165:345–56.

20 Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, Ludwig KU, Czamara D, St Pourcain B, Brandler W, Honbolygó F, Tóth D, Csépe V, Huguet G, Morris AP, Hulslander J, Willcutt EG, DeFries JC, Olson RK, Smith SD, Pennington BF, Vaessen A, Maurer U, Lyytinen H, Peyrard-Janvid M, Leppänen PHT, Brandeis D, Bonte M, Stein JF, Talcott JB, Fauchereau F, Wilcke A, Francks C, Bourgeron T, Monaco AP, Ramus F, Landerl K, Kere J, Scerri TS, Paracchini S, Fisher SE, Schumacher J, Nöthen MM, Müller-Myhsok B, Schulte-Körne G. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Transl Psychiatry* 2019;9.

21 Wojcik G, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, Belbin GM, Bien SA, Cheng I, Cullina S, Hodonsky CJ, Hu Y, Huckins LM, Jeff J, Justice AE, Kocarnik JM, Lim U, Lin BM, Lu Y, Nelson SC, Park S-SL, Poisner H, Preuss MH, Richard MA, Schurmann C, Setiawan VW, Sockell A, Vahi K, Vishnu A, Verbanck M, Walker R, Young KL, Zubair N, Acuna-Alonso V, Ambite JL, Barnes KC, Boerwinkle E, Bottinger E, Bustamante CD, Caberto C, Canizales-Quinteros S, Conomos MP, Deelman E, Do R, Doheny K, Fernandez-Rhodes L, Fornage M, Heiss G, Henn BM, Hindorff LA, Jackson RD, Hailu B, Laurie CA, Laurie CC, Li Y, Lin D-Y, Moreno-Estrada A, Nadkarni G, Norman PJ, Pooler LC, Reiner AP, Romm J, Sabati C, Sandoval K, Sheng X, Stahl EA, Stram DO, Thornton TA, Wassel CL, Wilkens LR, Winkler CA, Yoneyama S, Buyske S, Haiman C, Kooperberg C, Le Marchand L, Loos RJF, Matise TC, North KE, Peters U, Kenny EE, Carlson CS. The page study: how genetic diversity improves our understanding of the architecture of complex traits. *bioRxiv* 2018.

22 Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016;538:161–4.

23 Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. *Nature* 2011;475:163–5.

24 Galesloot TE, van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. *PLoS ONE* 2014;9:e95923.

25 O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, Coin LJM. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* 2012;7:e34861.

26 Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology* 2008;32:381–5.

27 Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.

28 Lu Q, Powles RL, Wang Q, He BJ, Zhao H. Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS Genet* 2016;12:e1005947.

29 Wang Y, Zhang B, Zhang L, An L, Xu J, Li D, Choudhary MNK, Li Y, Hu M, Hardison R, Wang T, Yue F. The 3D genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *bioRxiv* 2017.

30 Zhang Y, An L, Xu J, Zhang B, Zheng WJ, Hu M, Tang J, Yue F. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nature Communications* 2018;9.

31 Mehta CM, Gruen JR, Zhang H. A method for integrating neuroimaging into genetic models of learning performance. *Genet. Epidemiol.* 2017;41:4–17.

32 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF. The Genotype-Tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.

33 BrainSpan. *BrainSpan: Atlas of the developing human brain*. Secondary BrainSpan: Atlas of the Developing Human Brain, 2011.

34 Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C, Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman L, Sirotkin K, Slotta D, Zhang H, McVean GA, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton A, Campbell CL, Kong Y, Marcketta A, Gibbs RA, Yu F, Antunes L, Bainbridge M, Muzny D, Sabo A, Huang Z, Wang J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ, DePristo MA, Handsaker RE, Altshuler DM, Banks E, Bhatia G, del Angel G, Gabriel SB, Genovese G, Gupta N, Li H, Kashin S, Lander ES, McCarroll SA, Nemesh JC, Poplin RE, Yoon SC, Lihm J, Makarov V, Clark AG, Gottipati S, Keinan A, Rodriguez-Flores JL, Korbel JO, Rausch T, Fritz MH, Stütz AM, Flicek P, Beal K, Clarke L, Datta A, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zerbino D, Zheng-Bradley X, Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball EV, Stenson PD, Bentley DR, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Sudbrak R, Amstislavskiy VS, Herwig R, Mardis ER, Ding L, Koboldt DC, Larson D, Ye K, Gravel S, Swaroop A, Chew E, Lappalainen T, Erlich Y, Gymrek M, Frederick Willems T, Simpson JT, Shriver MD, Rosenfeld JA, Bustamante CD, Montgomery SB, De La Vega FM, Byrnes JK, Carroll AW, DeGorter MK, Lacroute P, Maples BK, Martin AR, Moreno-Estrada A, Shringarpure SS, Zakharia F, Halperin E, Baran Y, Lee C, Cerveira E, Hwang J, Malhotra A, Plewczynski D, Radew K, Romanovitch M, Zhang C, Hyland FCL, Craig DW, Christoforides A, Homer N, Izatt T, Kurdoglu AA, Sinari SA, Squire K, Sherry ST, Xiao C, Sebat J, Antaki D, Gujral M, Noor A, Ye K, Burchard EG, Hernandez RD, Gignoux CR, Haussler D, Katzman SJ, James Kent W, Howie B, Ruiz-Linares A, Dermitzakis ET, Devine SE, Abecasis GR, Min Kang H, Kidd JM, Blackwell T, Caron S, Chen W, Emery S, Fritsche L, Fuchsberger C, Jun G, Li B, Lyons R, Scheller C, Sidore C, Song S, Sliwerska E, Taliun D, Tan A, Welch R, Kate Wing M, Zhan X, Awadalla P, Hodgkinson A, Li Y, Shi X, Quitadamo A, Lunter G, McVean GA, Marchini JL, Myers S, Churchhouse C, Delaneau O, Gupta-Hinch A, Kretzschmar W, Iqbal Z, Mathieson I, Menelaou A, Rimmer A, Xifara DK, Oleksyk TK, Fu Y, Liu X, Xiong M, Jorde L, Witherspoon D, Xing J, Eichler EE, Browning BL, Browning SR, Hormozdiari F, Sudmant PH, Khurana E, Durbin RM, Hurles ME, Tyler-Smith C, Albers CA, Ayub Q, Balasubramaniam S, Chen Y, Colonna V, Danecek P, Jostins L, Keane TM, McCarthy S, Walter K, Xue Y, Gerstein MB, Abyzov A, Balasubramanian S, Chen J, Clarke D, Fu Y, Harmanci AO, Jin M, Lee D, Liu J, Jasmine Mu X, Zhang J, Zhang Y, Li Y, Luo R, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Ward AN, Wu J, Zhang M, McCarroll SA, Handsaker RE, Altshuler DM, Banks E, del Angel G, Genovese G, Hartl C, Li H, Kashin S, Nemesh JC, Shakir K, Yoon SC, Lihm J, Makarov V, Degenhardt J, Korbel JO, Fritz MH, Meiers S, Raeder B, Rausch T, Stütz AM, Flicek P, Paolo Casale F, Clarke L, Smith RE, Stegle O, Zheng-Bradley X, Bentley DR, Barnes B, Keira Cheetham R, Eberle M, Humphray S, Kahn S, Murray L, Shaw R, Lameijer E-W, Batzer MA, Konkel MK, Walker JA, Ding L, Hall I, Ye K, Lacroute P, Lee C, Cerveira E, Malhotra A, Hwang J, Plewczynski D, Radew K, Romanovitch M, Zhang C, Craig DW, Homer N, Church D, Xiao C, Sebat J, Antaki D, Bafna V, Michaelson J, Ye K, Devine SE, Gardner EJ, Abecasis GR, Kidd JM, Mills RE, Dayama G, Emery S, Jun G, Shi X, Quitadamo A, Lunter G, McVean GA, Chen K, Fan X, Chong Z, Chen T, Witherspoon D, Xing J, Eichler EE, Chaisson MJ, Hormozdiari F, Huddleston J, Malig M, Nelson BJ, Sudmant PH, Parrish NF, Khurana E, Hurles ME, Blackburne B, Lindsay SJ, Ning Z, Walter K, Zhang Y, Gerstein MB, Abyzov A, Chen J, Clarke D, Lam H, Jasmine Mu X, Sisu C, Zhang J, Zhang Y, Gibbs RA, Yu F, Bainbridge M, Challis D, Evani US, Kovar C, Lu J, Muzny D, Nagaswamy U, Reid JG, Sabo A, Yu J, Guo X, Li W, Li Y, Wu R, Marth GT, Garrison EP, Fung Leong W, Ward AN, del Angel G, DePristo MA, Gabriel SB, Gupta N, Hartl C, Poplin RE, Clark AG, Rodriguez-Flores JL, Flicek P, Clarke L, Smith RE, Zheng-Bradley X, MacArthur DG, Mardis ER, Fulton R, Koboldt DC, Gravel S, Bustamante CD, Craig DW, Christoforides A, Homer N, Izatt T, Sherry ST, Xiao C, Dermitzakis ET, Abecasis GR, Min Kang H, McVean GA, Gerstein MB, Balasubramanian S, Habegger L, Yu H, Flicek P, Clarke L, Cunningham F, Dunham I, Zerbino D, Zheng-Bradley X, Lage K, Berg Jespersen J, Horn H, Montgomery SB, DeGorter MK, Khurana E, Tyler-Smith C, Chen Y, Colonna V, Xue Y, Gerstein MB, Balasubramanian S, Fu Y, Kim D, Auton A, Marcketta A, Desalle R, Narechania A, Wilson Sayres MA, Garrison EP, Handsaker RE, Kashin S, McCarroll SA, Rodriguez-Flores JL, Flicek P, Clarke L, Zheng-Bradley X, Erlich Y, Gymrek M, Frederick Willems T, Bustamante CD, Mendez FL, David Poznik G, Underhill PA, Lee C, Cerveira E, Malhotra A, Romanovitch M, Zhang C, Abecasis GR, Coin L, Shao H, Mittelman D, Tyler-Smith C, Ayub Q, Banerjee R, Cerezo M, Chen Y, Fitzgerald TW, Louzada S, Massaia A, McCarthy S, Ritchie GR, Xue Y, Yang F, Gibbs RA, Kovar C, Kalra D, Hale W, Muzny D, Reid JG, Wang J, Dan X, Guo X, Li G, Li Y, Ye C, Zheng X, Altshuler DM, Flicek

P, Clarke L, Zheng-Bradley X, Bentley DR, Cox A, Humphray S, Kahn S, Sudbrak R, Albrecht MW, Lienhard M, Larson D, Craig DW, Izatt T, Kurdoglu AA, Sherry ST, Xiao C, Haussler D, Abecasis GR, McVean GA, Durbin RM, Balasubramaniam S, Keane TM, McCarthy S, Stalker J, Bodmer W, Bedoya G, Ruiz-Linares A, Cai Z, Gao Y, Chu J, Peltonen L, Garcia-Montero A, Orfao A, Dutil J, Martinez-Cruzado JC, Oleksyk TK, Barnes KC, Mathias RA, Hennis A, Watson H, McKenzie C, Qadri F, LaRocque R, Sabeti PC, Zhu J, Deng X, Sabeti PC, Asogun D, Folarin O, Happi C, Omoniwa O, Stremlau M, Tariyal R, Jallow M, Sisay Joof F, Corrah T, Rockett K, Kwiatkowski D, Kooner J, Tịnh Hiê'n Trâ'n, Dunstan SJ, Thuy Hang N, Fonnie R, Garry R, Kanneh L, Moses L, Sabeti PC, Schieffelin J, Grant DS, Gallo C, Poletti G, Saleheen D, Rasheed A. A global reference for human genetic variation. *Nature* 2015;526:68–74.

35 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants: fig. 1. *Bioinformatics* 2015;31:3555–7.

36 Wu H, Yang L, Chen L-L. The diversity of long noncoding RNAs and their generation. Trends in genetics.

37 Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Coller J, Baker KE. Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. *Cell Reports* 2014;7:1858–66.

38 Roberts TC, Morris KV, Wood MJA. The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2014;369.

39 Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.

40 Hennebry SC, Eikelis N, Socratous F, Desir G, Lambert G, Schlaich M, Renalase SM. Renalase, a novel soluble FAD-dependent protein, is synthesized in the brain and peripheral nerves. *Mol Psychiatry* 2010;15:234–6.

41 Xu J, Li G, Wang P, Velazquez H, Yao X, Li Y, Wu Y, Peixoto A, Crowley S, Desir GV. Renalase is a novel, soluble monoamine oxidase that regulates cardiac function and blood pressure. *J Clin Invest* 2005;115:1275–80.

42 Kobayashi K. Role of catecholamine signaling in brain and nervous system functions: new insights from mouse molecular genetic study. *J Investig Dermatol Symp Proc* 2001;6:115–21.

43 Hook V, Brennand KJ, Kim Y, Toneff T, Funkelstein L, Lee KC, Ziegler M, Gage FH. Human iPSC neurons display activity-dependent neurotransmitter secretion: aberrant catecholamine levels in schizophrenia neurons. *Stem Cell Reports* 2014;3:531–8.

44 Landi N, Frost SJ, Mencl WE, Preston JL, Jacobsen LK, Lee M, Yrigollen C, Pugh KR, Grigorenko EL. The *COMT* Val/Met polymorphism is associated with reading-related skills and consistent patterns of functional neural activation. *Dev Sci* 2013;16:13–23.

45 Grigorenko EL, Deyoung CG, Getchell M, Haeffel GJ, Klinteberg BAF, Koposov RA, Oreland L, Pakstis AJ, Ruchkin VV, Yrigollen CM. Exploring interactive effects of genes and environments in etiology of individual differences in reading comprehension. *Develop Psychopathol* 2007;19:1089–103.

46 Pugh KR, Frost SJ, Rothman DL, Hoeft F, Del Tufo SN, Mason GF, Molfese PJ, Mencl WE, Grigorenko EL, Landi N, Preston JL, Jacobsen L, Seidenberg MS, Fulbright RK. Glutamate and choline levels predict individual differences in reading ability in emergent readers. *J Neurosci* 2014;34:4082–9.

47 He Q, Xue G, Chen C, Chen C, Lu Z-L, Dong Q. Decoding the neuroanatomical basis of reading ability: a multivoxel morphometric study. *J Neurosci* 2013;33:12835–43.

48 Hoeft F, Meyler A, Hernandez A, Juel C, Taylor-Hill H, Martindale JL, McMillon G, Kolchugina G, Black JM, Faizi A, Deutsch GK, Siok WT, Reiss AL, Whitfield-Gabrieli S, Gabrieli JDE. Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences* 2007;104:4234–9.

49 Porter HF, O'Reilly PF. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci Rep* 2017;7.

50 Devanna P, Chen XS, Ho J, Gajewski D, Smith SD, Gialluisi A, Francks C, Fisher SE, Newbury DF, Vernes SC. Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry* 2017.