SUPPLEMENTAL MATERIALS AND METHODS

Discovery Analysis

Genes, Reading, and Dyslexia Study

The Genes, Reading, and Dyslexia (GRaD) study, is a multisite case-control study of RD in minority youth across the United States, Canada, and Puerto Rico. Detailed descriptions of recruitment, inclusion, and exclusion criteria for the sample are reported elsewhere [1]. Briefly, male and female children age 8-15 years of African American and/or Hispanic American ancestry were recruited for study (Table S1). Children were excluded if they were not of African American or Hispanic American ancestry, had less than 3 years of primary educational instruction in English, placed in foster care, had a medical history or neurological condition that could affect cognitive or neural development (i.e. preterm birth, prolonged stay in the NICU, seizures, acquired brain injuries), diagnosis of any cognitive or neuropsychiatric disorder (i.e. intellectual disability, autism spectrum disorder, depression), or documented hearing or vision impairment. Inclusion criteria for participants likely to have a reading disability included either history of poor reading skills (as documented by prior school or clinical testing), report of skills falling below expected level for age or grade, and/or provision of special services in the area of reading. For inclusion of participants likely to be controls, inclusion criteria were "competent reading skills" as identified by reading skills at or above current expectations for grade and performance above the 40th percentile on standardized school or clinical testing. A total of 1,432 unrelated children were recruited into the GRaD study. Of these subjects, 1,331 children with high-quality DNA samples were included in the analysis. Informed consent was obtained for all participants and parents or legal guardians of subjects. Ethical approval of study protocols and recruitment was obtained by Institutional Review Boards at each recruiting site (University of

Colorado-Boulder, University of Denver, Tufts University, University of New Mexico, Kennedy Krieger Institute, Hospital for Sick Children-Toronto, and Yale University).

Rapid Automatized Naming (RAN) and Rapid Automatized Stimulus Measures (RAS)

RAN and RAS performance in the GRaD sample was evaluated using the RAN Objects, RAN Letters, and RAS Letters/Numbers tasks developed by Wolf and Denckla [2]. For RAN Objects and RAN Letters, subjects name aloud 50 familiar, high frequency objects or letters, respectively, arranged in a 5 × 10 array as quickly and accurately as possible. The format of RAS Letters/Numbers is similar to RAN Objects and RAN Letters, except that items in the array are from alternating stimulus categories (e.g. letters and numbers). Time to complete each task is recorded and converted to age-standardized scores. Age-standardized scores (mean of 100, SD of 15) were then converted to z-scores (mean of 0, SD of 1) used for downstream genetic analyses (Table S2).

Test of Word Reading Efficiency (TOWRE)

The TOWRE Total Word Reading Efficiency is a composite of both Sight Word Efficiency and Phonetic Decoding Efficiency scores and is an assessment of reading fluency (the ability to read words quickly and accurately) under timed conditions [3]. For this assessment, the subject is evaluated on the number of individual words (Sight Word Efficiency) and nonwords (Phonetic Decoding Efficiency) correctly read in 45 seconds. For each subtest, the total number of words read correctly is converted into a standard score based on age norms and then converted to a z-score (Table S2). The TOWRE Total Word Reading Efficiency composite was used for downstream analysis.

Woodcock–Johnson Tests of Achievement, Third Edition (WJ-III)

The WJ-III Basic Reading Score is a composite of the WJ-III Letter-Word Identification and WJ-III Word Attack subtests [4]. The WJ-III Letter Word Identification subtest is an untimed measure of reading increasingly complex English words aloud. The Word Attack subtest is a decoding measure of nonwords or pseudowords in isolation. For each subtest, the total number of words read correctly is converted into a standard score based on age norms and then converted to a z-score (Table S2). The WJ-III Basic Reading composite was used for downstream analysis.

DNA Collection, Genotyping, and Analysis

Saliva was collected from each GRaD subject using the Oragene-DNA self-collection kit (OG-500; DNA Genotek Inc, Ottawa, Ontario, Canada). DNA was extracted using prepIT-L2P (DNA Genotek Inc, Ottawa, Ontario, Canada). Subjects were successfully genotyped for 2,391,739 single nucleotide polymorphisms (SNPs) using the Illumina Infinium Omni2.5-8 BeadChip at the Yale Center for Genome Analysis (Orange, CT). Initial genotyping quality control and SNP genotype calls were conducted using GenomeStudio (Illumina, San Diego, CA) and standard Infinium genotyping data analysis parameters to optimize genotyping accuracy. SNPs were removed from downstream analysis if they had missingness greater than 5% (n=22,849), Hardy-Weinberg equilibrium $p<0.0001$ (n=116,259), were not autosomal (n=60,551), or had a minor allele frequency (MAF) less than 5% (n=926,457). Samples were removed if they were missing more than 3% of their genotypes (n=39), if there were discrepancies between reported and inferred sex based on X chromosome heterozygosity (n=52), and IBD > 0.125 calculated using REAP (n=10) [5]. After quality control, there were a total 1,331 samples genotyped with 1,265,623 SNPs.

Population Stratification

Principal components (PCs) derived from genome-wide SNP data were used to correct for genomic inflation due to allele frequency differences across different ancestries (population stratification) were computed using EIGENSTRAT [6] (Figure S1). Visual inspection of a scree plot depicting the cumulative percent explained by each successive PC was used to determine the number of PCs needed for ancestry correction (Figure S2). A bend in the plot plus the subsequent PC dictated the number of PCs used. Adequate correction for population stratification was evaluated using a genomic inflation factor (λ) calculated using the R package GenABEL and PLINK v1.9 [7]. A λ factor below the standard threshold of 1.05 indicates sufficient correction for population stratification. For the full GRaD sample, three PCs were used as covariates to correct for population stratification (Figure S1 and S2A).

Hispanic American and African American ancestry in the GRaD sample was determined based on genetic similarity to 1000 genomes reference populations ASW (Americans of African Ancestry in Southwest USA) and AMR (Ad Mixed American), which consists of individuals with Mexican Ancestry from Los Angeles USA (MXL), Puerto Ricans from Puerto Rico (PUR), Colombians from Medellin, Colombia (CLM), and Peruvians from Lima, Peru (PEL). A PC analysis in EIGENSTRAT was conducted on a combined AMR and ASW sample set using SNPs that intersected with the GRaD sample. Calculated eigenvectors were then projected on to the GRaD sample. Resultant eigenvectors from the AMR, ASW, and GRaD Study subjects were then clustered by k-means clustering into two groups using the first three PCs. Clusters were labeled by visually comparing centroids against the projection of the AMR and ASW reference populations. GRaD subjects that clustered with the AMR reference population were considered Hispanic American, while subjects clustering with the ASW reference population were African American (Figure S3).

An independent set of genetic ancestry specific PCs in the GRaD sample, stratified based on cluster identity with 1000 genomes populations AMR and ASW, were generated using EIGENSTRAT. The first three and first two PCs were used to correct for population stratification for the Hispanic American and African American clusters, respectfully, for downstream analyses (Figure S2B and S2C).

Statistical Analysis: Multivariate Genome-wide Association Study

Multivariate genetic analysis jointly analyzing RAN Objects, RAN Letters, and RAS Letters/Numbers for pleiotropic effects was conducted using the R package MultiPhen [8]. MultiPhen allows for the simultaneous examination of multiple correlated traits using a reversed gaussian regression to determine the linear combination of traits most associated with specific genotypes at each SNP. It then performs a log likelihood ratio test on the joint model against the null model to evaluate association. All statistical models using the full GRaD sample were corrected for the first three PCs to correct for population stratification, sex, age, and socioeconomic status (SES). In this study, SES was assessed as a binary variable that describes whether the subject is enrolled in at least 1 government assistance program with a gross (pre-tax) income eligibility requirement of no more than 130% of the US federal poverty level (e.g. food stamps, Medicaid, housing choice voucher program, and/or Women, Infants, and Children program). Age, sex, and SES all account for significant trait variance in RAN Objects, RAN Letters, and RAS Letters/Numbers using a linear regression model in the full GRaD sample ($p < 0.05$). Spanish language use in the home environment was considered as a potential covariate in the final model, but was removed because it did not significantly explain additional trait variance the in the full GRaD or Hispanic American subjects. To correct for multiple testing in the

genome-wide multivariate GWAS, we used the standard threshold of 5×10^{-8} (Bonferroni correction for 1 million tests) to determine genome-wide significance.

Statistical Analysis: Genome-wide Meta-analysis

Follow-up univariate GWAS on a latent variable of RAN Objects, RAN Letters, and RAS Letters/Numbers was conducted in the GRaD sample stratified by cluster identity to AMR or ASW reference populations (Hispanic American and African American, respectfully) and then used in a meta-analysis (Figure S3). The latent variable reflects an unobservable factor that influences more than one observed measure and can account for the correlation and covariation among observed measures [9]. This allows for the examination of genetic factors that contributes to a "shared" cognitive component of RAN and RAS tasks in a univariate framework. The latent RAN Objects, RAN Letters, and RAS Letters/Numbers variable was generated using a confirmatory factor analysis to fit a measurement model, with the goal of explicitly modeling and segregating measurement error to allow the formation of a single latent naming speed variable [10]. Mplus version 8 was used, and standard scores on the three RAN tasks were incorporated as indicators [11]. The model was initially fit for the entire sample, fixing factor loadings to be equal across ethnic groups. Then, model invariance across ethnic divisions within the overall sample was evaluated by freeing factor loadings across samples/groups. The resultant latent naming speed variable was used in a univariate GWAS for Hispanic American and African American samples, respectively, in PLINK v1.9, while correcting for the effects of age, sex, SES, and ethnic specific PCs (three for Hispanic American, two for African American) to correct for intra-ethnic stratification (Figure S2B and S2C). Summary statistics from the ethnic specific GWASs of the latent naming speed variable were then used in a meta-analysis using a sample

size weighted design in METAL [12]. A heterogeneity analysis was also conducted in METAL to determine whether observed effect sizes were consistent across ethnic samples.

Statistical Analysis: Univariate ANCOVA

Univariate ANCOVA testing the top SNP in the multivariate GWAS or RAN Objects, RAN Letters, and RAS Letters/Numbers, and GWAS meta-analysis of a latent naming speed variable against measures of reading fluency (TOWRE and WJ-III) in the full GRaD cohort was conducted in SPSS v.24. Models corrected for the effects of age, sex, SES, and the first three genome-wide PCs (Figure S2A).

Replication Analysis

Colorado Learning Disability Research Center (CLDRC) Cohort

Replication was conducted on samples from the CLDRC. Methods related to recruitment, ascertainment, data collection (neurocognitive and genetic), and data processing are described in detail elsewhere [13 14]. Briefly, the CLDRC sample is a selected twin cohort for RD, ADHD, and other learning disabilities recruited from 27 school districts in Colorado [14 15]. Subjects were assessed with RAN Colors, RAN Objects, RAN Letters, and RAN Numbers [16]. For this task, participants named as many items in a 15 x 5 array as quickly and accurately as possible. The number of correctly named items in 15 seconds was recorded. Raw scores were standardized and age-regressed for each of the tasks based on a separate control sample consisting of typically developing children. All experimental procedures and written informed consent forms were approved by the institutional review boards (IRB) at University of Colorado-Boulder and University of Denver.

DNA was collected and extracted from saliva and genotyped using the Illumina Human OmniExpress genotyping panel (713,599 SNPs). Initial genotyping quality control and SNP genotype calls were conducted using GenomeStudio (Illumina, San Diego, CA). Initial quality control filters included the removal of samples with a call rate <98% and SNPs with a call rate <95%, HWE <0.0001, and MAF < 5%. SNPs on chromosome 10 identified in the GRaD discovery analysis that were not genotyped were imputed with genipe—an automated genome-wide imputation pipeline that executes PLINK 1.07 [7], SHAPEIT [17], and IMPUTE2 [18] for data imputation to the 1000 Genomes Project, Phase 3 reference [19 20]. All imputed SNPs had an info score (IMPUTE2 imputation quality metric) greater than 0.9, indicating that all replication SNPs were imputed with high confidence.

The sample of twins and siblings available for this study comprised 749 participants in total, mean age 11.7 years, age range 8–19, from 343 unrelated twinships/sibships (Table S1). For the present study, only participants of European descent were analyzed and one child per twinship/sibship were randomly selected for analysis based on the availability of RAN and genetic data. The total sample size for replication analysis was 318 unrelated individuals.

Multivariate genetic analysis jointly analyzing RAN Colors, RAN Pictures, RAN Letters, and RAN Numbers standard scores for association at SNPs identified in the GRaD discovery analyses was conducted using MultiPhen while covarying for the effects of age, sex, and the first genome-wide PC calculated using EIGENSTRAT to correct for population stratification (Figure S2D).

Bioinformatic Analysis

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

GenoSkyline is an unsupervised learning framework that predicts tissue-specific functionality in non-coding regions of the genome by integrating genome-wide epigenetic data from the Roadmap Epigenomics Project and ENCODE [21-23]. Detailed description of tissue specific functionality has been previously described [21]. Briefly, GenoSkyline uses an unsupervised-learning technique that evaluates the presence of well characterized DNase 1 hypersensitivity sites and histone marks (H3k4me1, H3k4me3, H3k36me3, H3k27me3, H3k9me3, H3k27ac, H3k9ac) and calculates a posterior probability score (GS score) that a given genetic coordinate is functional. A GS score of "1" suggests that the genomic region of interest is functional within the given tissue type, while a score of "0" suggests no functional significance. Pre-calculated, genome-wide, tissue-specific GS scores for whole brain and sub-regions 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. We evaluated the 18-state model previously generated by the Roadmap Epigenomics Project. Briefly, a model was derived from a multivariate hidden Markov model that considers the combinatorial interactions of 6 histone marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, and H3K27ac) across 127 epigenomes and predicts the chromatin state within a genomic region within 18 different classifications [22]. Brain specific regions sampled were angular gyrus, prefrontal cortex, cingulate gyrus, anterior caudate, hippocampus, inferior temporal gyrus, and substantia nigra. Data were visualized using the Washington University in St. Louis (WashU) EpiGenome Browser v.42.

Linkage disequilibrium (LD) blocks were calculated using confidence intervals for CEU, AMR and YRI populations in 1000 Genomes Project Phase 3 [20 24]. SNPs with MAF < 0.1 were removed prior to calculation of LD blocks. D' values were extracted using the web program LDlink.

The 3D genome browser was used to visualize publicly available chromatin contact maps at 10 kb resolution from adult hippocampal tissue [25-27].

Neuroimaging Genetic Analysis

Pediatric, Imaging, Neurocognition, and Genetics (PING) Study

Neuroimaging genetics analysis of volumes of cortical ROIs was conducted in the PING sample. Methods related to recruitment, ascertainment, data collection (neuroimaging, neurocognitive, genetic), and data processing are described in detail elsewhere [28]. Briefly, the PING study is a cross sectional sample of typically developing children ranging in age from 3-20 years old (Table S1). Individuals were excluded from participating if they had a history of major developmental, psychiatric, or neurological disorders, brain injury, prematurity (i.e., born at less than 36 weeks gestational age), prenatal exposure to illicit drugs or alcohol, history of head trauma, or other medical conditions that could affect development. Subjects with learning disability or ADHD were not excluded. All experimental procedures and written informed consent forms were approved by the institutional review boards (IRB) at each of the 10 participating PING study recruitment sites (University of California at San Diego, University of Hawaii, University of California at Los Angeles, Children's Hospital of Los Angeles of the University of Southern California, University of California at Davis, Kennedy Krieger Institute of Johns Hopkins University, Sackler Institute of Weill Cornell Medical College, University of

Massachusetts, Massachusetts General Hospital at Harvard University, and Yale University). Parental informed consent was obtained for participants less than 18 years of age with child assent (ages 7-17). For individuals 18 years of age and older, written informed consent was obtained.

DNA was extracted from saliva and genotyped on the Illumina Human660W Quad BeadChip (655,214 SNPs) for 1391 subjects in the PING study. Initial genotyping quality control and SNP genotype calls was conducted using GenomeStudio (Illumina, San Diego, CA) by the PING genomics core at the Scripps Translational Science Institute (La Jolla, CA). Initial quality control filters included the removal of samples with a call rate <98% and SNPs with a call rate <95%, HWE <0.0001, and MAF < 5%.

In depth descriptions of methods for neuroimaging data acquisition and processing for the PING study are described elsewhere [28]. Briefly, structural MRI data were collected from all individuals using a standardized multiple-modality high-resolution structural MRI protocol involving 3D T1-weighted volumes across sites to maintain consistency across data collection. Image files in DICOM format were processed by the neuroimaging post-processing core at the University of California at San Diego using an automated processing stream written in MATLAB and C++. Cortical surface reconstruction and subcortical segmentation were performed using a fully automated set of tools available in the Freesurfer software suite. Cortical parcellation of sulci and gyri were automatically defined using the Desikan-Killiany Atlas integrated within the FreeSurfer suite [29]. All data (genetic and neuroanatomical) were obtained from the PING portal.

Cortical volume regions of interest selection and statistical analysis

Left and right hemisphere cortical regions spanning the inferior frontal gyrus, temporoparietal region, and occipitotemporal area were selected for candidate region of interest analysis (Table S3). Left hemisphere structures spanning the inferior frontal gyrus, temporoparietal region, and occipitotemporal area comprise the canonical reading network, and show atypical patterns of functional activation in reading disabled children relative to typically developing children [30]. Respective right hemisphere regions were also included in the analysis since evidence suggests that naming speed is also associated with grey matter volume differences in right hemisphere frontal, temporoparietal and occipital regions [30 31]. Although hippocampus is not part of the canonical cortical reading network, the left and right hippocampus was included in the analysis to follow-up on tissue specific epigenetic and chromatin interaction findings in the present study.

Imaging genetics analysis was conducted on 690 subjects with complete phenotype and genotype data that passed both neuroimaging and genotype QC. All analyses were corrected for the effects of age, sex, handedness, scanner device [32], intracranial volume, highest parental education, family income, and the first four genome-wide PCs calculated using EIGENSTRAT to correct for population stratification (Figure S2E). The SNP rs1555839 was genotyped using the Illumina Human660W Quad BeadChip in the PING sample and used for neuroimaging genetics analysis. Cortical volumes across 16 ROIs were tested for association with rs1555839 using linear regression under an additive genetic model in PLINK 1.9 [33].

REFERENCES

1. Jacobson LA, Koriakin T, Lipkin P, et al. Executive Functions Contribute Uniquely to Reading Competence in Minority Youth. J Learn Disabil 2016 doi: 10.1177/0022219415618501[published Online First: Epub Date]|.
2. Wolf M, Denkla MB. RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests. Austin, TX: Pro-Ed, 2005.
3. Torgesen JK, Wagner RK, Rashotte CA. Test of Word Reading Efficiency. Austin, TX: PRO-ED, 1999.
4. Woodcock RW, McGrew KS, Mather N. Woodcock-Johnson® III Normative Update Complete. Boston, MA: Houghton Mifflin Harcourt, 2001.
5. Thornton T, Tang H, Hoffmann TJ, et al. Estimating kinship in admixed populations. Am J Hum Genet 2012;**91**(1):122-38 doi: 10.1016/j.ajhg.2012.05.024[published Online First: Epub Date]|.
6. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;**38**(8):904-9 doi: 10.1038/ng1847[published Online First: Epub Date]|.
7. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;**81**(3):559-75 doi: 10.1086/519795[published Online First: Epub Date]|.
8. O'Reilly PF, Hoggart CJ, Pomyen Y, et al. MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS. PLOS ONE 2012;**7**(5):e34861 doi: 10.1371/journal.pone.0034861[published Online First: Epub Date]|.
9. Todd DL, Timothy AB.
10. Brown TA, Moore MT. Confirmatory factor analysis. Handbook of structural equation modeling. New York, NY, US: Guilford Press, 2012:361-79.
11. Muthén LK, Muthén BO. Mplus User's Guide, Eighth Edition. Los Angeles, CA: Muthén & Muthén, 1998-2017.
12. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;**26**(17):2190-1 doi: 10.1093/bioinformatics/btq340[published Online First: Epub Date]|.
13. Gialluisi A, Newbury DF, Wilcutt EG, et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav 2014;**13**(7):686-701 doi: 10.1111/gbb.12158[published Online First: Epub Date]|.
14. DeFries JC, Filipek PA, Fulker DW, et al. Colorado Learning Disabilities Research Center. Learning Disabilities: AMultidisciplinary Journal 1997;**8**:7-19
15. Willcutt EG, Pennington BF, Olson RK, et al. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. Developmental Neuropsychology 2005;**27**(1):35-78 doi: 10.1207/s15326942dn2701_3[published Online First: Epub Date]|.
16. Compton DL, Olson RK, DeFries JC, et al. Comparing the Relationships Among Two Different Versions of Alphanumeric Rapid Automatized Naming and Word Level Reading Skills. Scientific Studies of Reading 2002;**6**(4):343-68
17. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 2013;**10**(1):5-6 doi: 10.1038/nmeth.2307[published Online First: Epub Date]|.

18. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;**11**(7):499-511 doi: 10.1038/nrg2796[published Online First: Epub Date]|.
19. Perreault L-PL, Legault M-A, Asselin G, et al. genipe: An automated genome-wide imputation pipeline with automatic reporting and statistical tools. Bioinformatics 2016
20. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;**526**(7571):68-74 doi: 10.1038/nature15393[published Online First: Epub Date]|.
21. Lu Q, Powles RL, Wang Q, et al. 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 Genetics 2016;**12**(4):e1005947 doi: 10.1371/journal.pgen.1005947[published Online First: Epub Date]|.
22. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature 2015;**518**(7539):317-30 doi: 10.1038/nature14248[published Online First: Epub Date]|.
23. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;**489**(7414):57-74 doi: 10.1038/nature11247[published Online First: Epub Date]|.
24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;**296**(5576):2225-29 doi: 10.1126/science.1069424[published Online First: Epub Date]|.
25. Wang Y, Zhang B, Zhang L, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. bioRxiv 2017
26. Zhang Y, An L, Xu J, et al. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nature Communications 2018;**9**:750 doi: 10.1038/s41467-018-03113-2[published Online First: Epub Date]|.
27. Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps Reveal Spatially Active Regions in the Human Genome. Cell reports 2016;**17**(8):2042-59 doi: 10.1016/j.celrep.2016.10.061[published Online First: Epub Date]|.
28. Jernigan TL, Brown TT, Hagler DJ, Jr., et al. The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. Neuroimage 2016;**124**(Pt B):1149-54 doi: 10.1016/j.neuroimage.2015.04.057[published Online First: Epub Date]|.
29. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;**31**(3):968-80 doi: 10.1016/j.neuroimage.2006.01.021[published Online First: Epub Date]|.
30. Norton ES, Beach SD, Gabrieli JD. Neurobiology of dyslexia. Curr Opin Neurobiol 2015;**30**:73-8 doi: 10.1016/j.conb.2014.09.007[published Online First: Epub Date]|.
31. 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 doi: 10.1146/annurev-psych-120710-100431[published Online First: Epub Date]|.
32. Mehta CM, Gruen JR, Zhang H. A method for integrating neuroimaging into genetic models of learning performance. Genet Epidemiol 2017;**41**(1):4-17 doi: 10.1002/gepi.22025[published Online First: Epub Date]|.

33. Chang CC, Chow CC, Tellier LCAM, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;**4**(1):7 doi: 10.1186/s13742-015-0047-8[published Online First: Epub Date]|.

Figure S1: Results of principal components analysis (PCA) depicting the first 10 PCs plotted against one another in the full GRaD sample. Visual inspection of PC plots 1-3 reflect population structure, while PC 4-10 do not. Light grey dots = self-reported African American ethnicity; Dark grey dots = self-reported Hispanic American ethnicity; Black dots = self-reported Hispanic American and African American ethnicity.

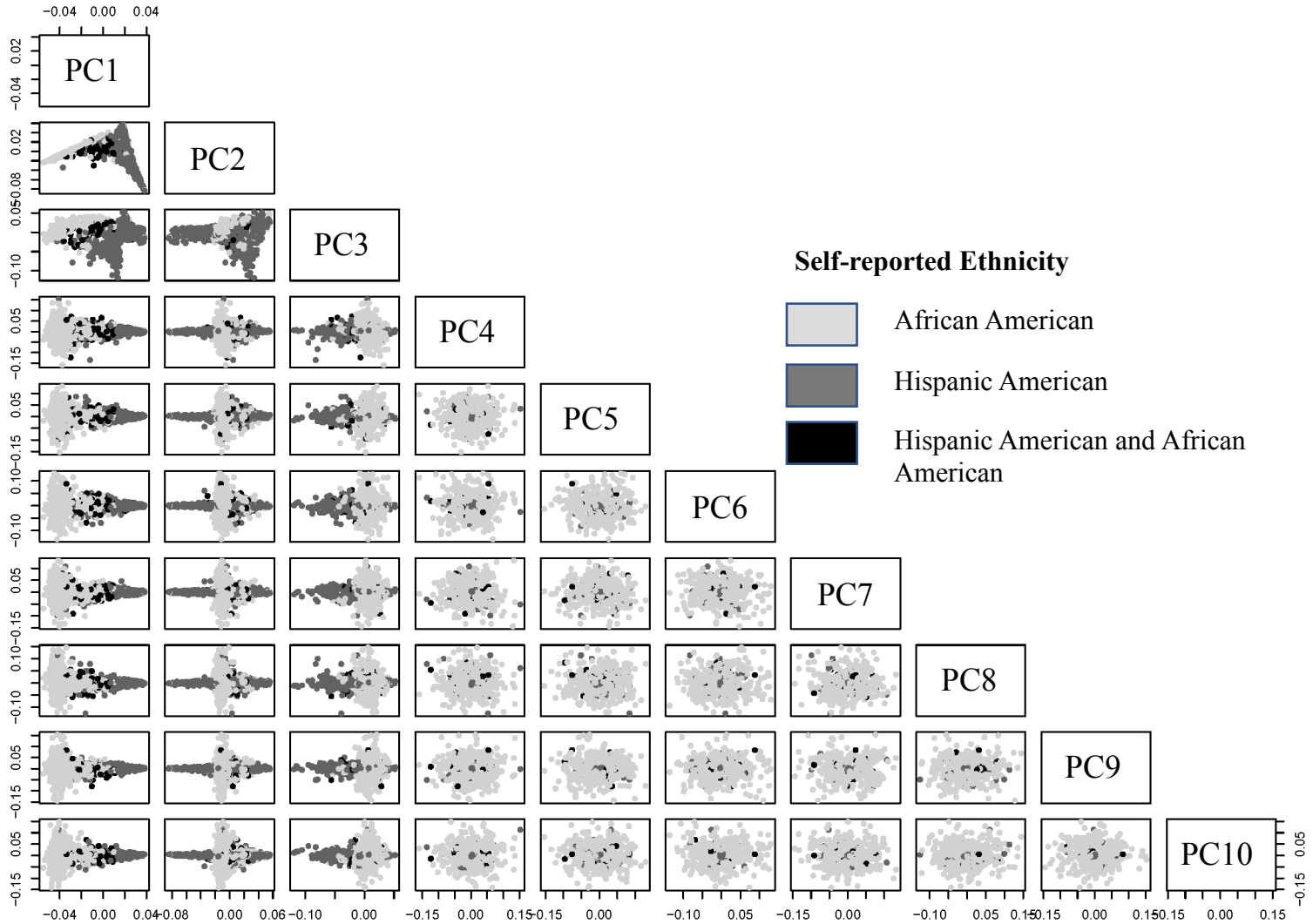

Figure S2: Scree plots depicting cumulative variance explained with each additional consecutive principal component (PC) for A) the full GRaD sample, B) Hispanic American cluster subset of GRaD, C) African American cluster subset of GRaD, D) CLDRC, and E) PING. A bend in the plot represents the point at which each additional PC contributes to incremental gains in variance explained. PCs located to the left of the bend, plus one, are included in the final statistical models as covariates to correct for population stratification. For example, in the GRaD sample, the first three PCs, accounting for 6.6% of the variance, are used as covariates in the final models to correct for population stratification.

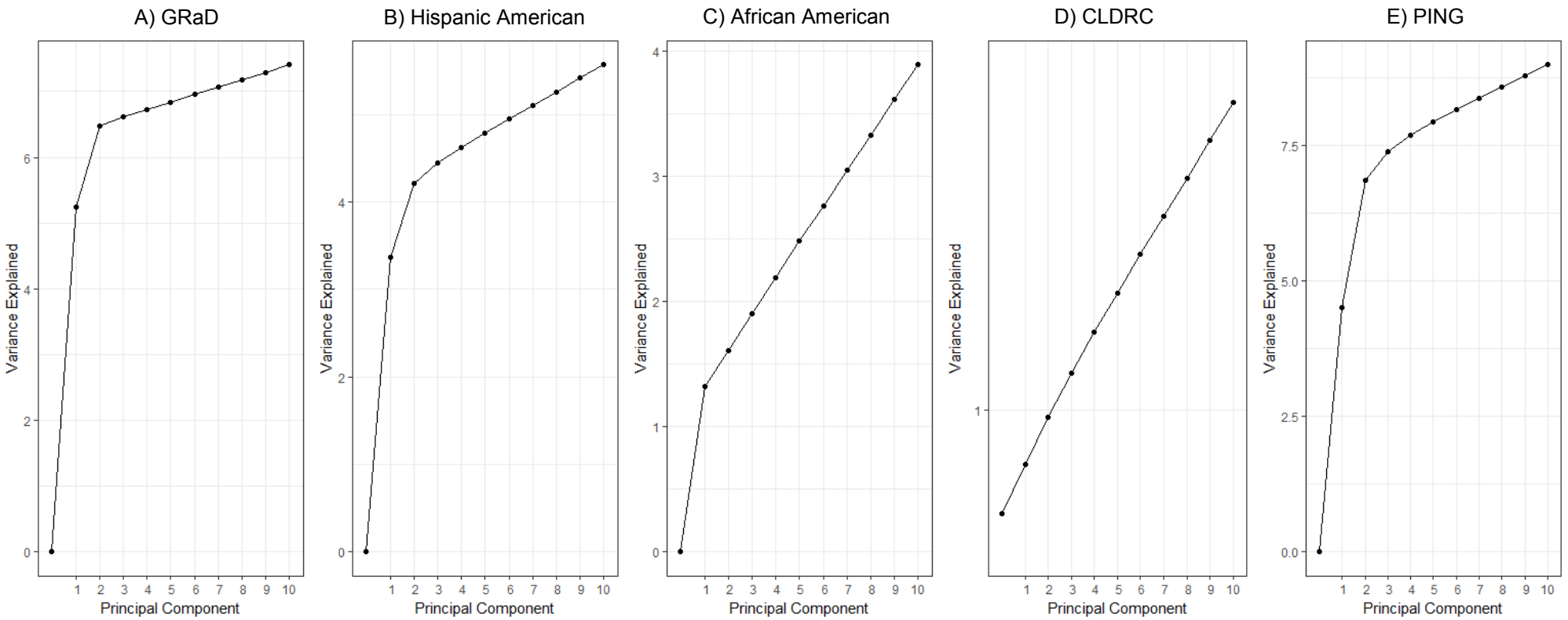

Figure S3: Results of PCA with PC projection of AMR and ASW reference populations on the full GRaD sample with K-means clustering using two centroids. PC1 plotted against PC2. GRaD subjects clustering with the majority of AMR reference samples were identified as Hispanic American, while GRaD subjects clustering with the majority of ASW reference samples were identified as African American. With K-means clustering, there was 82.4% correspondence between self-reported African Americans in GRaD and cluster identification with the ASW reference population. For self-reported Hispanic Americans in GRaD, there was 95.9% correspondence with AMR cluster identification

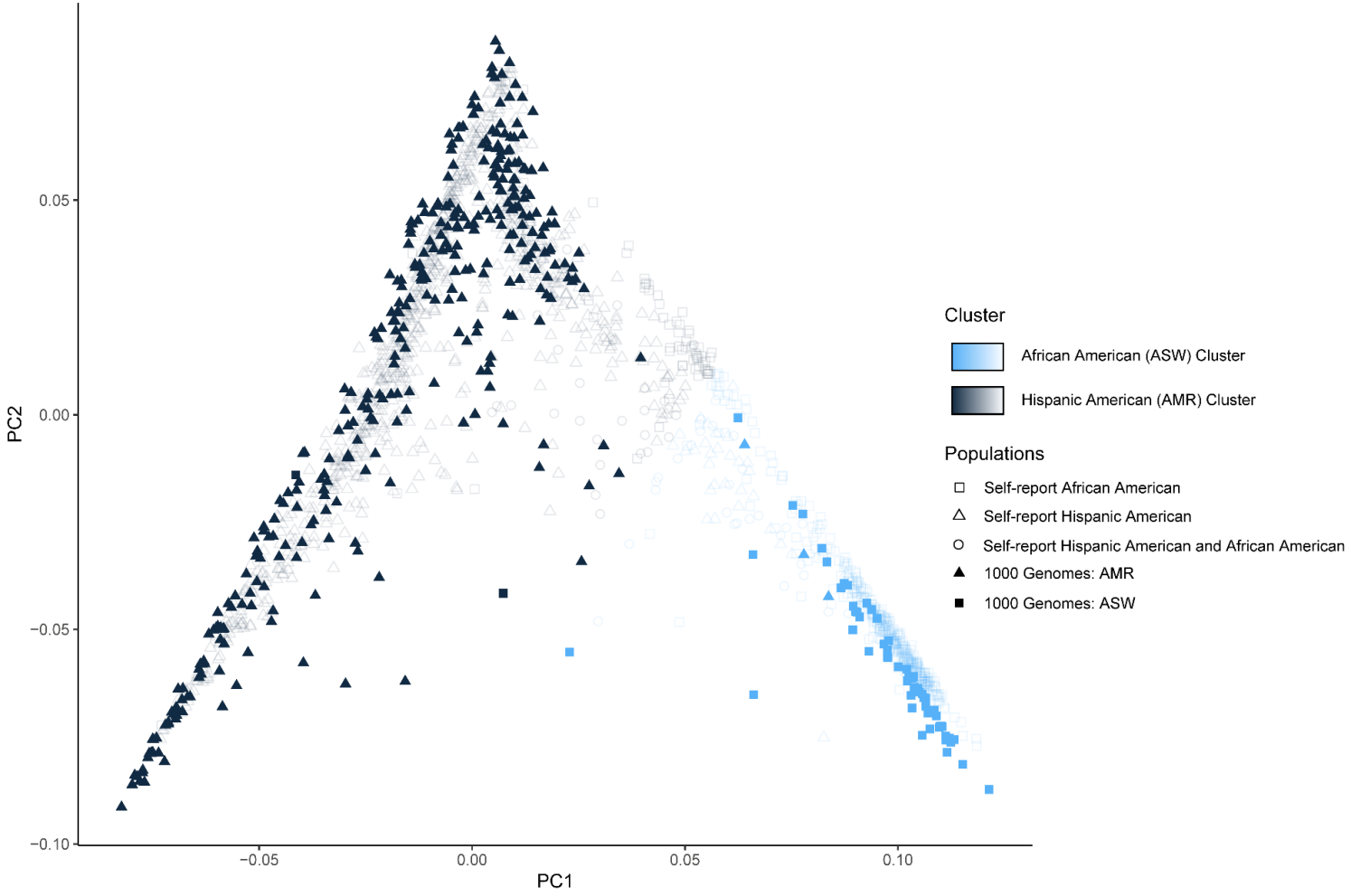

Figure S4: Q-Q plot showing expected distribution of test statistics over observed for the multivariate GWAS analysis in the GRaD sample.

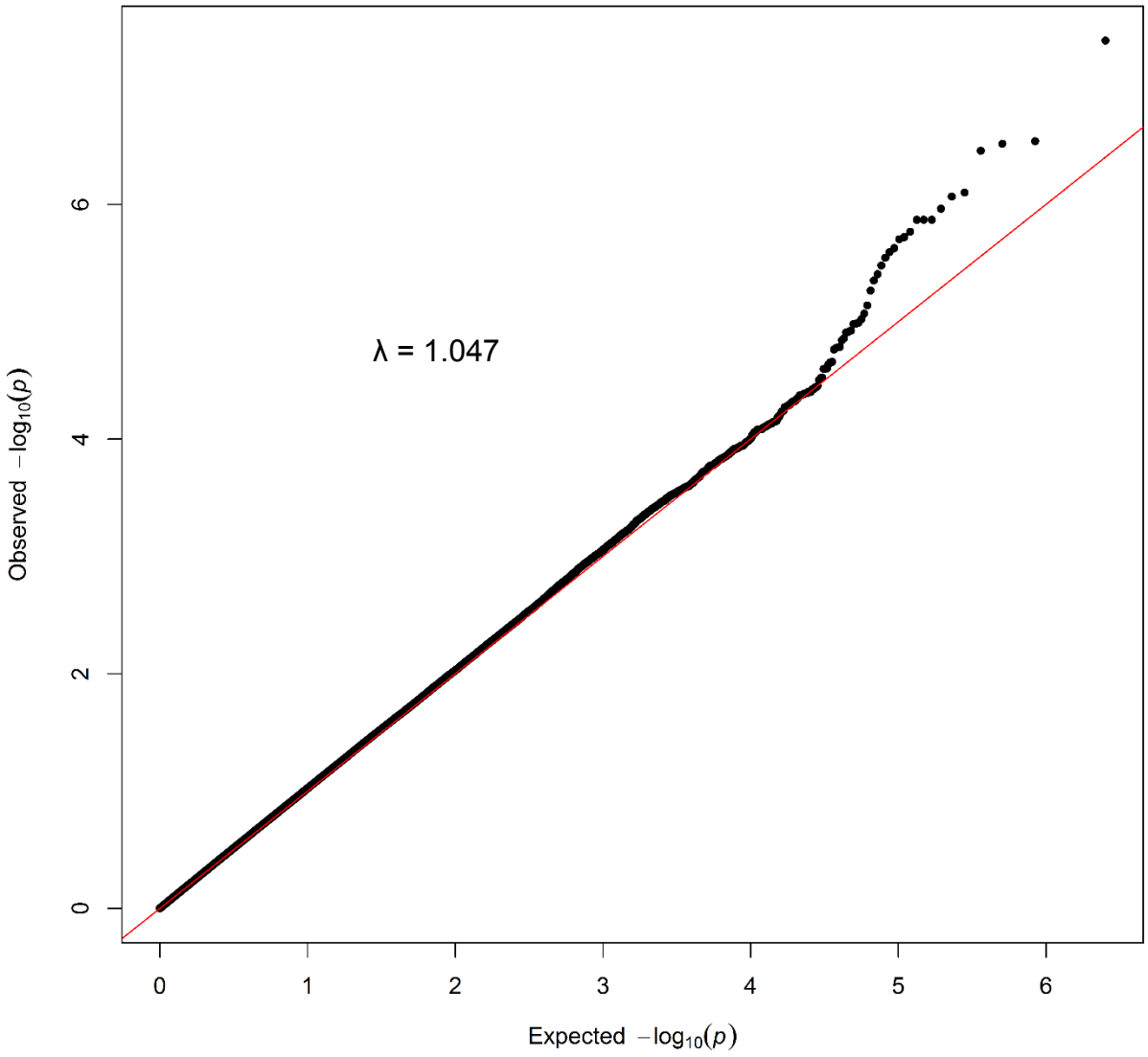

Figure S5: A) Manhattan and B) Q-Q plots summarizing results from univariate GWAS of RAN Objects.

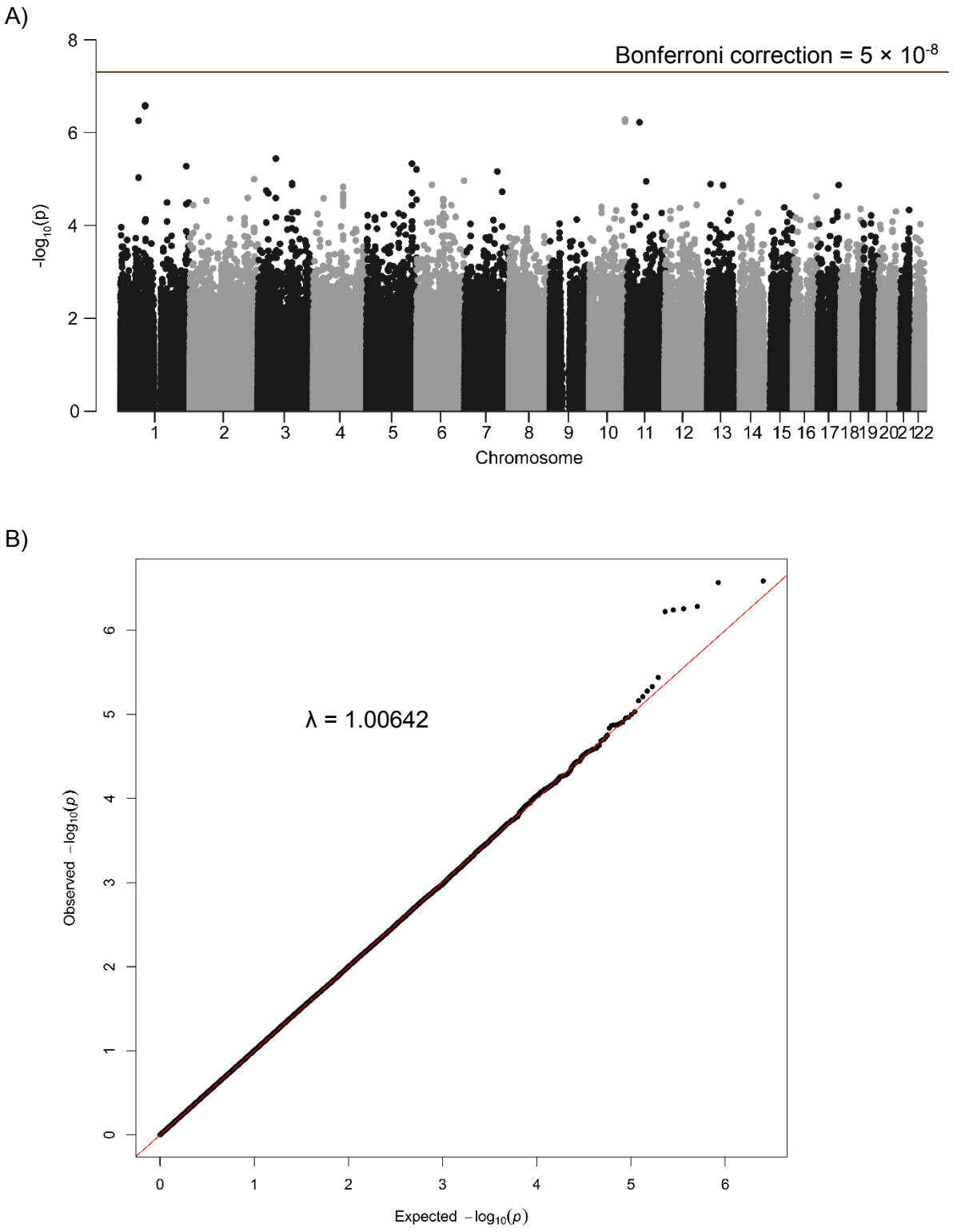

Figure S6: A) Manhattan and B) Q-Q plots summarizing results from univariate GWAS of RAN Letters

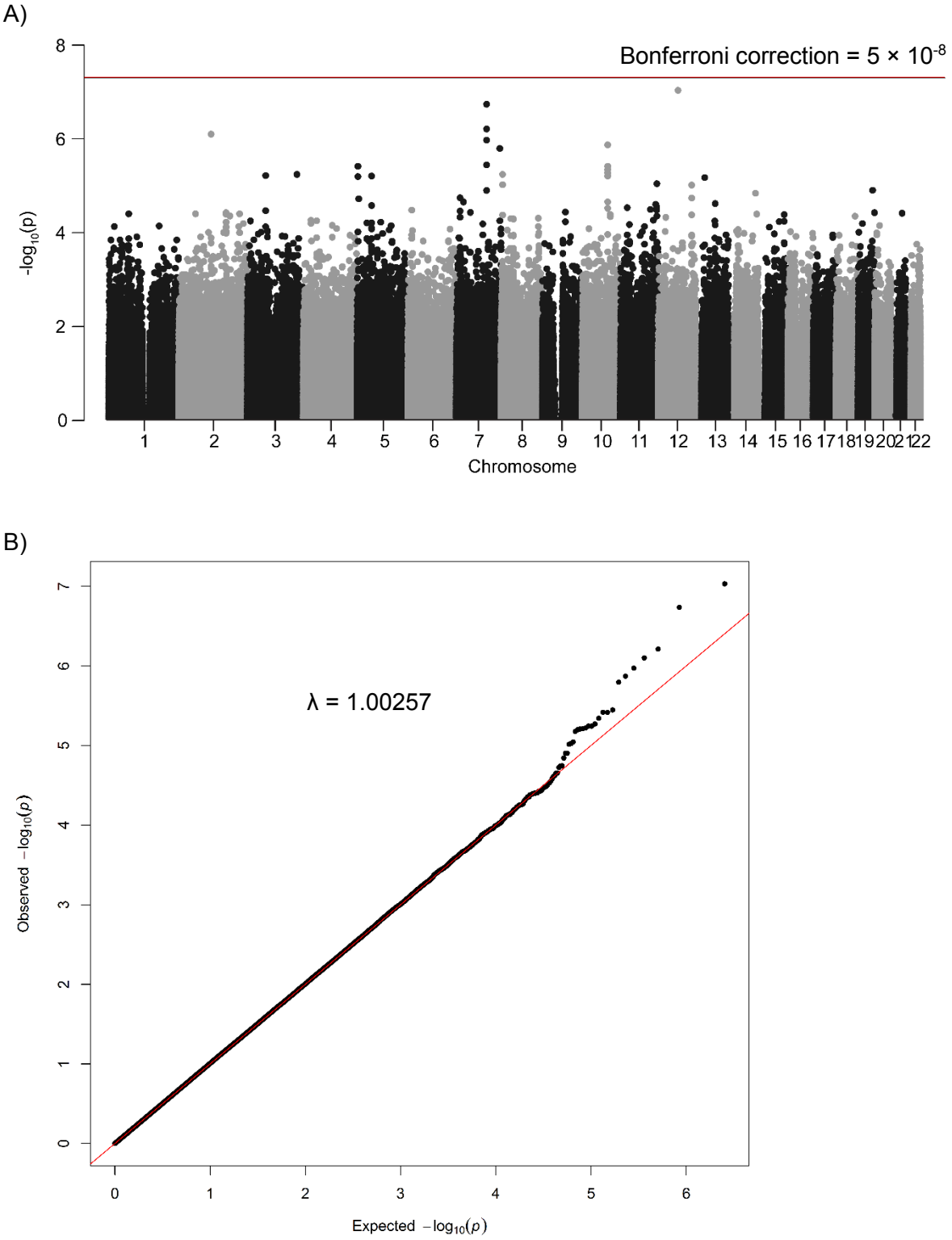

Figure S7: A) Manhattan and B) Q-Q plots summarizing results from univariate GWAS of RAS Letters/Numbers

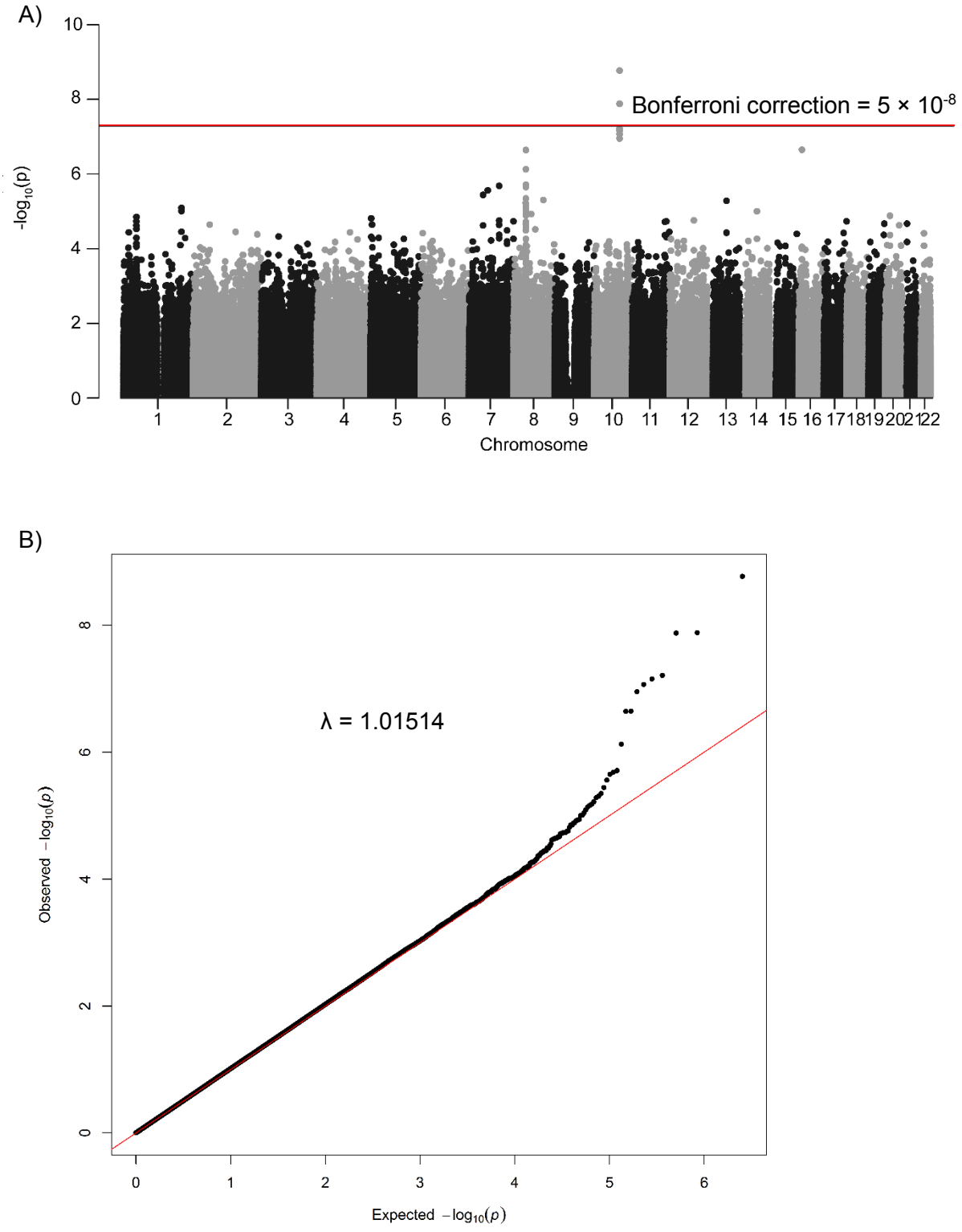

Figure S8: A) Manhattan and B) Q-Q plots summarizing results from meta-analysis of a latent naming speed variable in GRaD.

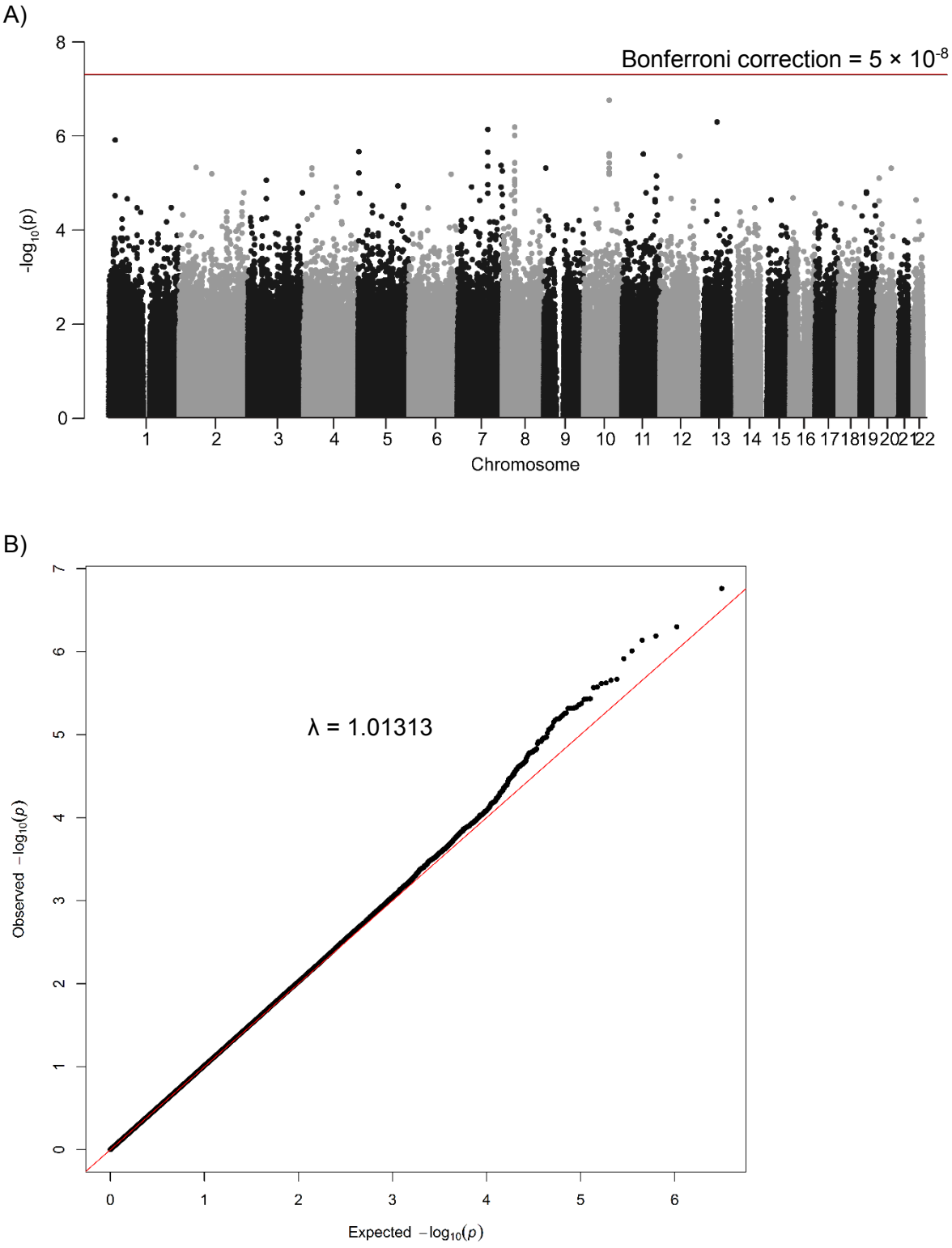

Figure S9: Pre-calculated GS scores (blue) from GenoSkyline for whole brain, angular gyrus, prefrontal cortex, cingulate gyrus, anterior caudate, hippocampus, inferior temporal gyrus, and substantia nigra. A GS score of "1" suggests that the genomic region of interest is functional within the given tissue type, while a score of "0" suggests no functional significance.

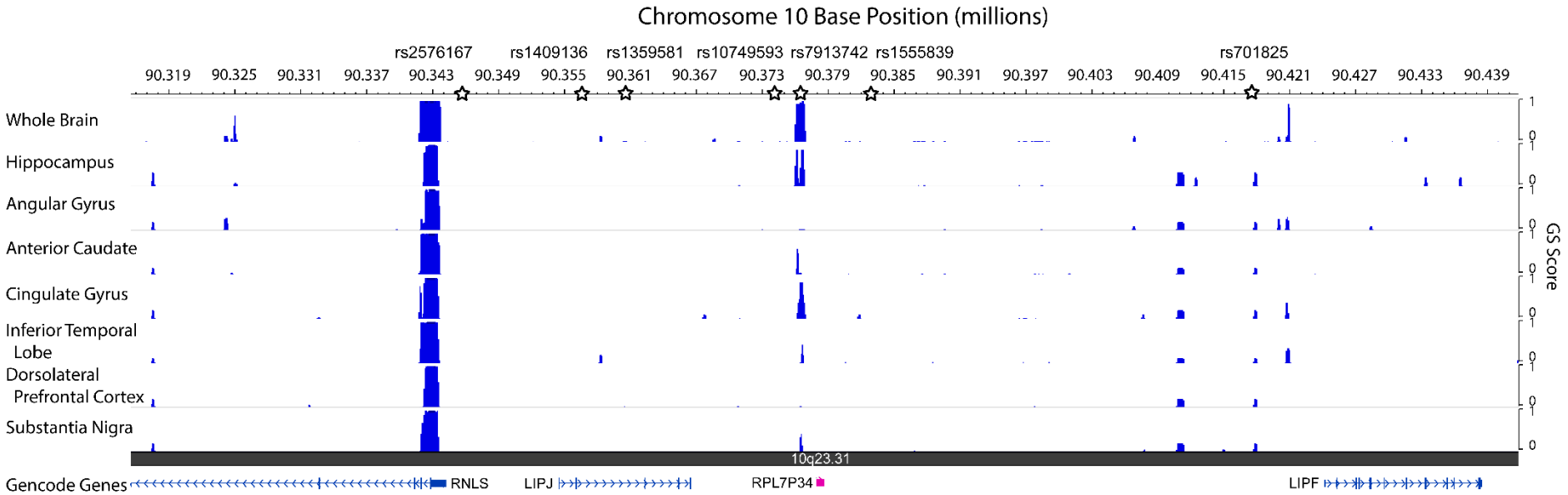

Table S1: Demographic information for GRaD, CLDRC, and PING samples

	Self-Reported Ancestry	Child's Age in Years, range (mean)	Sex Male:Female	Sample Size
GRaD	Hispanic American	7.8-15.9 (11.3)	379:433	812
	African American	8-15.8 (11.5)	212:233	445
	Hispanic and African American	8.1-15.8 (12.1)	33:34	67
	Total	7.8-15.9 (11.4)	624:700	1324
CLDRC	European	8-19 (11.6)	165:153	318
PING	[a]White	3.7-20.8 (12.1)	241:226	467
	[a]African American or Black	3.5-20.8 (13.2)	56:40	96
	[a]American Indian or Alaska Native	3.7-20.8 (12.1)	5:0	5
	[a]Asian	3.6-21 (14.5)	16:48	64
	Report More Than One	3.5-20.9 (11.8)	25:33	58
	Total	3.5-21 (12.4)	343:347	690

[a]US census category

Table S2: GRaD Assessments Descriptives: Descriptive statistics for assessments evaluated in the GRaD sample. Means reflect the z-score (mean = 0, standard deviation (SD) = 1) of the population standard score (mean = 100, SD = 15) available for each psychometric analysis.

	Mean (SD)	Skew	Kurtosis
RanObjects	-0.198 (1.044)	0.341	0.055
RanLetters	0.063 (0.938)	0.059	-0.081
RasLettNum	0.050 (0.943)	-0.128	-0.068
TOWRE	-0.489 (1.118)	0.019	-0.200
WJ-III Basic Reading	-0.365 (0.924)	-0.473	1.279

Table S3: Cortical regions of interest associated with reading ability and disability

Inferior Frontal Gyrus	Pars Opercularis Pars Orbitalis Pars Triangularis
Temporo-parietal	Supramarginal Gyrus Inferior Parietal Cortex
Occipito-temporal	Inferior Temporal Gyrus Fusiform Gyrus

Table S4: Pearson correlation coefficients across RAN Objects, RAN Letters, RAS Letters/Numbers, Test of Word Reading Efficiency (TOWRE), and Woodcock-Johnson Basic Reading (WJ-III Basic Reading).

	RAN Objects	RAN Letters	RAS Letters/Numbers	TOWRE	WJ-III Basic Reading
RAN Objects	1	0.621**	0.624**	0.413**	0.325**
RAN Letters		1	0.781**	0.547**	0.464**
RAS Letters/Numbers			1	0.578**	0.488**
TOWRE				1	0.846**
WJ-III Basic Reading					1

**$p < 1 \times 10^{-32}$ (two-tailed)

Table S5: Meta-analysis results on latent naming speed variable derived from RAN and RAS performance in GRaD.

MARKER	CHR	BP	Minor Allele	Z-score	P-value	Direction	HetISq	HetPVal	GENE
rs1555839	10	90382820	C	5.225	1.74×10^{-7}	++	31.1	0.228	*RPL7P34*
rs9540938	13	67441725	A	-5.025	5.03×10^{-7}	--	0	0.779	*PCDH9*
rs8188533	8	43586240	T	-4.976	6.48×10^{-7}	--	0	0.580	--
rs6963842	7	107634989	G	4.953	7.29×10^{-7}	++	63.1	0.099	*LAMB1*
rs7463498	8	43447235	G	4.896	9.77×10^{-7}	++	59	0.118	--
rs16870453	5	2795776	T	4.738	2.15×10^{-6}	++	0	0.943	--
rs4320486	7	107643977	T	4.733	2.21×10^{-6}	++	73.2	0.053	*LAMB1*
rs701825	10	90417547	G	4.718	2.39×10^{-6}	++	85.7	8.28×10^{-3}	--
rs11177505	12	69516642	G	-4.694	2.68×10^{-6}	--	54.4	0.139	--
rs8175494	8	43621389	C	4.628	3.69×10^{-6}	++	0	0.401	--

Top 10 markers from a meta-analysis of ethnicity specific GWAS on a latent naming speed variable for RAN Objects, RAN Letters, and RAS Letters and Numbers in Hispanic American and African American participants in the GRaD study. Markers were assigned to genes if they fell within the canonical gene body as described by 1000 Genomes Project, Phase 3 (v80 GRCh37). CHR = chromosome, BP = Base Position, MAF = Minor Allele Frequency, Z-score = combined z-statistic, P-value = meta-analysis p-value, Direction = summary of effect direction for each sample, HetISq = I^2 statistic which measures heterogeneity on scale of 0-100% across samples, HetPVal = P-value for heterogeneity statistic.

Table S6: Univariate association analysis of RAN Letters in the CLDRC cohort displaying results from top markers identified in the GRaD multivariate GWAS and latent naming speed GWAS meta-analysis.

MARKER	CHR	BP	Minor Allele	MAF	BETA	SE	STAT	P
rs6963842	7	107634989	T	0.495	0.020	0.110	0.182	0.856
rs1555839	10	90382820	C	0.319	-0.311	0.101	-3.063	2.38×10^{-3}*
rs701825	10	90417547	G	0.282	-0.345	0.103	-3.34	9.32×10^{-4}*

*Survives Bonferroni correction for multiple testing ($p < 0.0161$)

Table S7: Genetic coordinates for LD blocks within chr10:90325000-90427000

1000 genomes population	Genetic Coordinates on Chromosome 10
CEU	90325770-90334313 90334338-90418065 90422136-90426612
AMR	90325133-90339593 90340747-90341049 90343398-90418065 90422136-90426612
YRI	90327653-90329787 90331816-90334835 90340747-90341049 90342837-90381875 90386161-90417499 90422136-90425327