Molecular Psychiatry (2020) 25:3010–3019
https://doi.org/10.1038/s41380-018-0205-3

ARTICLE

Identification of novel loci associated with infant cognitive ability

Ryan Sun[1] · Zhaoxi Wang[2] · Birgit Claus Henn[3] · Li Su[2] · Quan Lu[2] · Xihong Lin[1] · Robert O. Wright[4] · David C. Bellinger[5] · Molly Kile[6] · Maitreyi Mazumdar[7] · Martha Maria Tellez-Rojo[8] · Lourdes Schnaas[8] · David C. Christiani[2]

Received: 2 December 2017 / Revised: 21 June 2018 / Accepted: 28 June 2018 / Published online: 17 August 2018

Abstract

It is believed that genetic factors play a large role in the development of many cognitive and neurological processes; however, epidemiological evidence for the genetic basis of childhood neurodevelopment is very limited. Identification of the genetic polymorphisms associated with early-stage neurodevelopment will help elucidate biological mechanisms involved in neuro-behavior and provide a better understanding of the developing brain. To search for such variants, we performed a genome-wide association study (GWAS) for infant mental and motor ability at two years of age with mothers and children recruited from cohorts in Bangladesh and Mexico. Infant ability was assessed using mental and motor composite scores calculated with country-specific versions of the Bayley Scales of Infant Development. A missense variant (rs1055153) located in the gene *WWTR1* reached genome-wide significance in association with mental composite score (meta-analysis effect size of minor allele $\beta_{\text{meta}} = -6.04$; 95% CI: −8.13 to −3.94; $P = 1.56\times10^{-8}$). Infants carrying the minor allele reported substantially lower cognitive scores in both cohorts, and this variant is predicted to be in the top 0.3% of most deleterious substitutions in the human genome. Fine mapping and region-based association testing provided additional suggestive evidence that both *WWTR1* and a second gene, *LRP1B*, were associated with infant cognitive ability. Comparisons with recently conducted GWAS in intelligence and educational attainment indicate that our phenotypes do not possess a high genetic correlation with either adolescent or adult cognitive traits, suggesting that infant neurological assessments should be treated as an independent outcome of interest. Additional functional studies and replication efforts in other cohorts may help uncover new biological pathways and genetic architectures that are crucial to the developing brain.

Introduction

Human neurodevelopment is known to be an intricate process that involves both genetic and non-genetic factors [1]. However despite the complexity of this mechanism, certain aspects of neurodevelopment manifest as relatively stable traits [2] and have been extensively studied. For instance, general cognitive ability is acknowledged to be reliably measureable [3], and it has been linked with outcomes ranging from longevity [4] to socioeconomic status [5]. Another example is motor ability, which has been shown to predict cognitive performance [6] and is thus associated with the same range of attributes. The broad spectrum and importance

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41380-018-0205-3) contains supplementary material, which is available to authorized users.

✉ Ryan Sun
rys177@mail.harvard.edu

[1] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[2] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[3] Department of Environmental Health, Boston University School of Public Health, Boston, MA 02118, USA

[4] Department of Preventive Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA

[5] Department of Psychiatry, Harvard Medical School and Boston Children's Hospital, Boston, MA 02115, USA

[6] College of Public Health and Human Sciences, Oregon State University, Corvallis, OR 97331, USA

[7] Department of Neurology, Boston Children's Hospital, Boston, MA 02115, USA

[8] Center of Nutrition and Health Research, National Institute of Public Health, Cuernavaca, Morelos 62100, Mexico

of these outcomes motivate us to more clearly understand the processes that drive and influence neurodevelopment.

Many measures of neurodevelopment are also known to be highly heritable [7], and genetic association studies have been conducted for a variety of cognition-related traits including cortical thickness [8], language skills [9], and mathematical ability [10], although early investigations encountered difficulty in pinpointing variants with significant effects [11]. More recently, a number of increasingly larger genome-wide association studies (GWAS) have led to the discovery of many loci significantly associated with cognitive function [4, 11–16] and educational attainment [17–19]. Results from these massive studies have been used to construct genome-wide polygenic scores predicting more than 10% of the variance in adult intelligence, which comprises one-fifth of the estimated heritability [20]. However, to generate such large sample sizes, these studies focus on older populations composed mainly of Europeans. While such results are crucial to our understanding of neurological outcomes, the heritability of cognitive development is known to vary dramatically with age [21]—intelligence test results from age two predict less than 5% of the variance in intelligence during late adolescence [20]—and much less research has been performed specifically on young children or populations of non-European descent.

In this study we attempted to address the paucity of knowledge about genetic loci affecting early-stage neurodevelopment by conducting a GWAS for cognitive and motor ability in two-year-old children. Our cohorts were recruited from Bangladesh and Mexico, and ability was assessed through mental and motor composite scores, two measures derived from the Bayley Scales of Infant Development, Third Edition (BSID-III). While our data are cross-sectional in nature and do not strictly measure development over time, point estimates of mental and motor capability can still offer important insight into the mechanisms of childhood neurodevelopment; to the best of our knowledge, this study is one of the first attempts to link genetic factors with infant aptitude. We find that our traits do not demonstrate high genetic correlation with previously performed GWAS of older populations, providing additional evidence that the biological mechanisms driving cognitive ability change over time and illustrating that the phenotypes we have studied should be considered distinct from adult intelligence or educational attainment. Our study identifies two genes associated with mental composite score and provides a better understanding of how genetic influences shape the developing brain in very young children.

Methods

Recruitment and enrollment of study participants

Detailed recruitment and enrollment procedures have been described previously for both the Bangladesh [22] and Mexico [23] cohorts. The studies were approved by Dhaka Community Hospital, the National Institute of Public Health Mexico, Harvard School of Public Health, and other universities involved in the data collection. Written consent was obtained from all mothers.

Covariate and outcome data collection

In both cohorts, mothers were interviewed at enrollment to obtain demographic and other background information. Follow-up visits occurred approximately two years after the birth of each child. During these visits, trained personnel administered versions of the BSID-III specially adapted for each region. Mental composite score was calculated as the sum of cognition, expressive language, and receptive language scores, and motor composite score was calculated as the sum of fine motor and gross motor scores. Additional information on study recruitment, data collection, and exclusion of outliers is described in the Supplementary Methods. The final sample sizes were 502 mother–infant pairs in Bangladesh and 462 pairs in Mexico.

Preparation of genotype data

Genotyping was performed using commercially available Illumina arrays [24]. Broad Genomics at the Broad Institute performed genotyping in the Bangladesh cohort using the OmniExpressExome-8 BeadChip. The Children's Hospital of Pennsylvania performed genotyping in the Mexico cohort using the HumanOmni1-Quad BeadChip (San Diego, CA, USA). Details on quality control measures, correction for population stratification, and imputation are provided in the Supplementary Methods.

Genome-wide association study

Because the two cohorts possessed such different baseline genetic architectures, our analysis strategy differed slightly from the standard GWAS procedure. Instead of testing all 964 subjects in the same model, we stratified the sample by ethnicity first to generate cohort-specific test statistics for each outcome. We then used METAL [25] to perform a standard error-based meta-analysis on the cohort-specific test statistics. Thus for each outcome there was still only one final measure of association with each single-nucleotide polymorphism (SNP). This method was chosen to reduce the potential for finding spurious associations due to population stratification.

Cohort-specific test statistics were generated in PLINK v1.9 [26], using additive allelic linear models that adjusted for the confounding effects of gestational age, infant sex, maternal education level, age at time of neurodevelopmental assessment, household smoking, and the first two genotype

principal component vectors for that cohort. Maternal education level was coded as a binary variable taking the value 1 if the mother completed schooling past primary school. Household smoking was also coded as a binary variable. All other variables were recorded in continuous form. The effect of the SNP was analyzed with a Wald test; all *P*-values were two-sided. We chose a significance threshold of 4.5×10^{-8} based on a Bonferroni adjustment for 1,104,974 tests (552,487 SNPs multiplied by two outcomes), and we also only considered SNPs demonstrating a *P*-value less than 1×10^{-3} in both cohorts. Power simulations were performed in R version 3.3.1. Our threshold was slightly more stringent than the commonly applied 5×10^{-8} limit for genome-wide significance. We further checked the robustness of our top hits through resampling methods and a number of sensitivity analyses, stratifying the cohort by gender, gestational age, and other covariates (Supplementary Methods).

Finally we performed local imputation and region-based testing in the genes surrounding signals discovered above. Each gene was partitioned into 100 kb windows, resulting in two windows for *WWTR1* and 19 for *LRP1B*. This size of window was chosen to capture most long-range Linkage Disequilibrium (LD) patterns [27]. *P*-values for the association with mental and motor composite score were calculated with the Generalized Higher Criticism (GHC) statistic [28]. The significance threshold for all individual-SNP tests of association on the imputed SNPs was again set at $P = 4.5 \times 10^{-8}$ while the threshold for region-based tests was set at $P = 0.0012$ (0.05/42), corresponding to 21 regions tested with two outcomes each. All set-based analysis was performed using the "GBJ" package in R.

Gene–environment interaction effects

After initial genome-wide testing, we utilized available toxicology data to search for possible gene–environment interaction effects between the most highly associated SNPs and metal exposures. As there is evidence that lead and other metals are neurotoxicants [29, 30], and because the Bangladesh cohort in particular is a high-exposure population, it is possible that significant interaction effects could contribute to large marginal SNP effects. To test this hypothesis, we introduced measurements of lead and manganese concentration (μg/dl) in umbilical cord blood one at a time into the cohort-specific linear models. Specifically, we fit the same main effect model but also added a metal concentration covariate as well as a SNP–metal interaction term, and we then re-performed the meta-analysis for the interaction term. This procedure was performed separately for each metal exposure. Cohort-specific *P*-values were calculated using a robust inference procedure to mitigate the effects of possible exposure misspecification [31].

Table 1 Demographic, clinical, and neurological assessment data for Bangladesh and Mexico cohorts

Characteristics	Mean (SD) or *n* (%)	
	Bangladesh ($n = 502$)	Mexico ($n = 462$)
Sex, *n* (%)		
Male	256 (51)	245 (53)
Female	246 (49)	217 (47)
Concentration of metals in umbilical cord blood, μg/dl[a]		
Manganese	5.92 (4.45)	3.20 (1.11)
Lead	4.72 (4.04)	3.83 (2.70)
Gestational age, weeks	38.2 (1.7)	38.4 (1.6)
Age at exam, weeks	99.4 (18.5)	106.1 (2.4)
Mother's education >primary, *n* (%)	269 (54)	269 (58)
Smoking in household environment, *n* (%)	212 (42)	172 (37)
BSID Scores		
Mental composite	112.7 (10.5)	108.5 (9.7)
Motor composite	92.7 (5.0)	91.8 (4.6)

SD standard deviation, *CI* confidence interval

[a]There were 23 subjects in the Bangladesh cohort and 73 in the Mexico cohort who did not have any recorded values for metal concentration. Also, for each of the metals separately, we classified subjects with a measured value greater than 5 standard deviations from the median as outliers and removed them from analysis with that metal due to possible measurement error. This procedure resulted in the removal of three subjects from the manganese sample and two subjects from the lead sample, all from the Bangladesh cohort

Results

Full demographic differences between the populations are provided in Table 1. Mean exposures to potentially toxic metals were higher in Bangladesh, but the differences were relatively small compared to their corresponding standard errors. Other demographic factors were relatively similar between the two populations.

Genome-wide association analysis

After meta-analysis we found one SNP, rs1055153, to reach genome-wide significance in association with mental composite score (Table 2). rs1055153 ($\beta_{meta} = -6.04$; 95% CI: −8.13 to −3.94; $P = 1.56 \times 10^{-8}$) is a missense SNP (Pro74Gln) located on chromosome 3 in the gene *WWTR1* (NCBI Entrez Gene 25937, also referred to as *TAZ* in publications). Diagnostic quantile–quantile plots of both the cohort-specific GWAS *P*-values (Supplementary Figure 1) and the meta-analysis *P*-values (Supplementary Figure 2) appear to show almost no inflation, indicating that the association is a robust result. Additional sensitivity analyses did not produce noteworthy findings (Supplementary Tables 1–3).

Table 2 Two top SNPs from initial GWAS meta-analysis and their association with infant mental and motor composite score

Cohort	Mental composite score			Motor composite score			MAF
	Effect size[a] (SD)	95% CI	*P*-value	Effect size[a](SD)	95% CI	*P*-value	
rs1055153/Chr3:149 374 873[b]							
Meta-analysis	−6.04 (1.07)	−8.13 to −3.94	1.56×10^{-8}	−2.21 (0.52)	−3.22 to −1.19	2.17×10^{-5}	0.05
Bangladesh	−5.56 (1.24)	−7.98 to −3.14	8.68×10^{-6}	−2.26 (0.60)	−3.43 to −1.09	1.68×10^{-4}	0.08
Mexico	−7.43 (2.11)	−11.58 to −3.29	4.84×10^{-4}	−2.04 (1.06)	−4.11 to 0.04	5.54×10^{-2}	0.02
rs13013197/Chr2:141 882 963[c]							
Meta-analysis	2.01 (0.47)	1.09 to 2.93	1.86×10^{-5}	0.95 (0.23)	0.50 to 1.40	3.63×10^{-5}	0.33
Bangladesh	1.54 (0.74)	0.09 to 2.98	3.72×10^{-2}	0.74 (0.35)	0.05 to 1.44	3.52×10^{-2}	0.27
Mexico	2.34 (0.61)	1.14 to 3.53	1.47×10^{-4}	1.10 (0.30)	0.51 to 1.69	3.15×10^{-4}	0.40

SD standard error, *CI* confidence interval, *MAF* minor allele frequency

[a]Effect size refers to the estimated mean change in outcome for one additional minor allele, holding all other covariates constant

[b]SNP position recorded from NCBI Build 37/UCSC hg19 coordinates. This SNP resides in the gene *WWTR1*, and the minor allele is T

[c]SNP position recorded from NCBI Build 37/UCSC hg19 coordinates. This SNP resides in the gene *LRP1B*, and the minor allele is A

The average mental composite score for children in Bangladesh with one copy of the minor allele was substantially lower than the score for children with no copies of the minor allele: 108.1–113.6 (Fig. 1). In the Mexico cohort, the average mental composite score for children with one copy of the minor allele was 101.5, compared to 108.8 for children with no copies of the minor allele. Neither cohort contained a subject who was homozygous for the minor allele.

Although no other SNPs reached our nominal significance threshold in association with either mental composite score or motor composite score, we noted another SNP, rs13013197 in the gene *LRP1B* (NCBI Entrez Gene 53353), showed strong associations (Table 2, Supplementary Tables 4 and 5, and Supplementary Figure 3) with both outcomes (mental composite score $\beta_{meta} = 2.01$; 95% CI: 1.09–2.93; $P = 1.86 \times 10^{-5}$; motor composite score $\beta_{meta} = 0.95$; 95% CI: 0.50–1.40; $P = 3.63 \times 10^{-5}$), as higher Bayley scores were observed among carriers of the minor allele. Based on the strength of this evidence, we decided to carry forward rs13013197 as a second locus for the fine-mapping stage of our study.

Fine mapping and region-based testing

We imputed 784 SNPs common to both cohorts in the region of *WWTR1* to investigate whether the observed association with rs1055153 could be attributed to an SNP that was in LD with rs1055153 but not found on the genotyping chip. Fig. 2 shows a local Manhattan plot of both imputed and originally genotyped SNPs in the region, as well their meta-analysis *P*-values for association with mental composite score and motor composite score. The single most significant association for both outcomes was the originally documented SNP rs1055153. No other SNPs in the region came close to the genome-wide significance threshold, and only a handful showed evidence of association at $P < 0.01$ (Fig. 2 and Supplementary Figure 4).

Region-based testing (Supplementary Table 6) reinforced trends shown in the local Manhattan plot. Of the two 100 kb windows tested in *WWTR1*, only the region containing rs1055153 was associated with both outcomes (GHC $P = 1.18 \times 10^{-5}$ for mental composite score and $P = 8.60 \times 10^{-4}$ for motor composite score). The other region-outcome combinations did not meet our nominal significance threshold, suggesting that the originally genotyped SNP was driving the entirety of the association with cognitive ability.

We also imputed 7113 SNPs common to both cohorts in the region of *LRP1B*. In contrast to *WWTR1*, the imputed SNPs in *LRP1B* appeared to be much more highly associated with both outcomes (Fig. 2 and Supplementary Figure 5). The original top SNP rs13013197 was no longer the SNP that showed the strongest association with either outcome, and in fact, there appeared to be multiple loci containing an enrichment of SNPs demonstrating association with both outcomes. Region-based testing showed the most significant 100 kb window for association to be the one beginning at 140,988,996 (NCBI Build 37/UCSC hg19 coordinates) and ending at 141,088,995 (GHC $P = 1.08 \times 10^{-3}$ for mental composite score). This window was quite far from rs13013197 and could even be classified as a second locus of interest. These results appeared to indicate that SNPs not on our genotyping chip could be driving the associations at *LRP1B*.

Gene–environment interaction

The gene–environment interaction models (Table 3) indicated that there may be a significant interaction effect

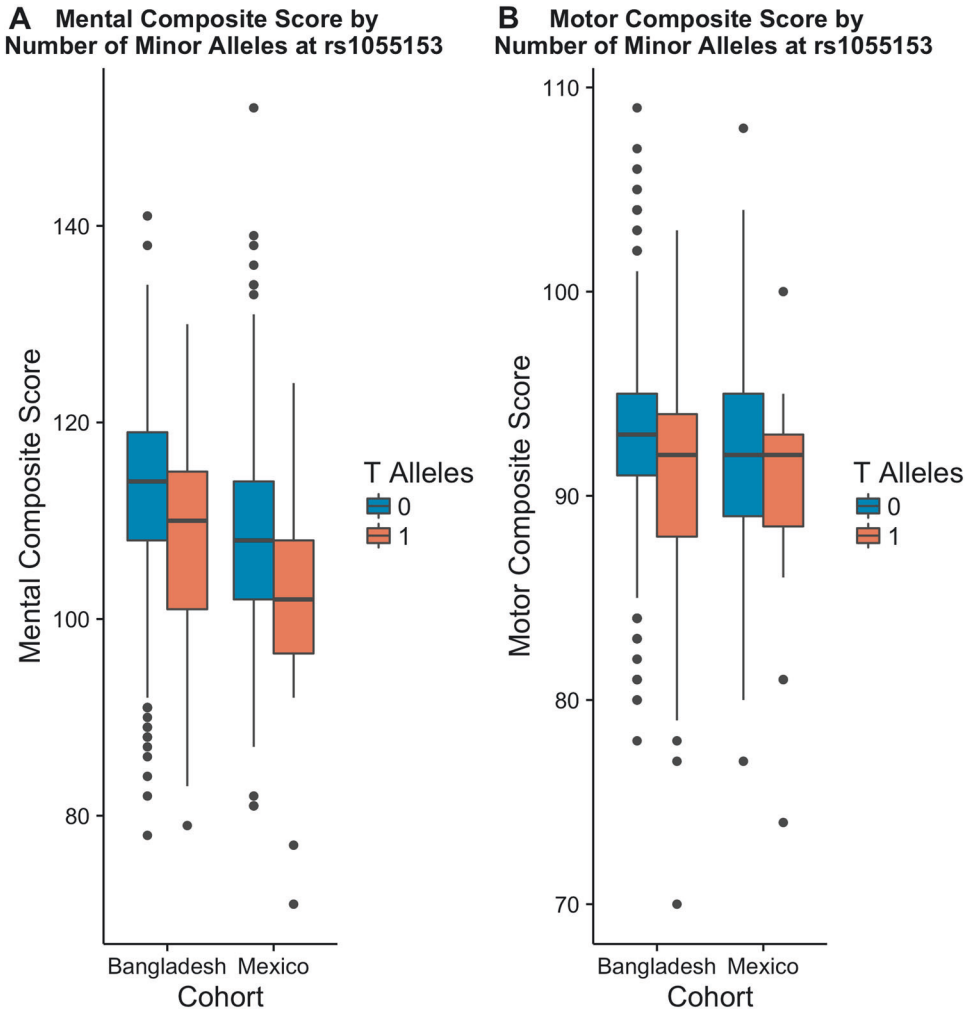

Fig. 1 Population mental and motor composite score by number of minor alleles at rs1055153. Boxplots for mental composite score (**a**) and motor composite score (**b**) within Bangladesh and Mexico. No subject in the entire study had two copies of the minor allele at this SNP. Thick black line shows the median, and the box ends at the first and third quartiles. Whiskers extend to 1.5 times the interquartile range

involving rs1055153 and manganese exposure ($P < 0.05$ for both mental composite score and motor composite score) for both outcomes. Each additional copy of the minor allele was estimated to be associated with an additional decrease in mental and motor composite score for a unit increase in blood manganese concentration. No other SNP–metal interaction terms demonstrated a similar level of association.

Functional analysis of top variants

We sought to further verify the functional link between our outcomes and the two top SNPs by investigating metrics designed to offer predictions about variant impact. One such measure is Combined Annotation Dependent Deletion (CADD) score, which incorporates a variety of different annotations to predict the genomic substitutions having the most deleterious effects in humans [32]. By this measure, the variant at rs1055153 is predicted to be in the top 0.3% of most deleterious substitutions in the entire human genome. Other similar algorithms offer slightly more conservative predictions (PolyPhen-2 [33] = 0.98, probably deleterious; SIFT [34] = 0.085, predicted to be tolerated). GERP++ [35] also indicates the variant is highly conserved (GERP++ score = 5.09). Outside of the SIFT prediction, these results demonstrate that rs1055153 is likely the *WWTR1* SNP most responsible for the observed region-based association with cognitive function.

In contrast, the variant at rs13013197 is not scored highly by algorithmic approaches, does not possess a clear biological role given its status as an intron, and lies in a high LD region next to other strongly associated SNPs, necessitating further investigation to infer the true risk-conferring variant at its locus. We thus reviewed multi-omics data from over 30 different sources for all imputed variants within 50 kb of rs13013197 that showed a similarly strong association with mental or motor composite score (Supplementary Methods). This search produced one variant, rs13035290 (Supplementary Table 7), that is both highly correlated with rs13013197 (squared Pearson correlation coefficient $r^2 > 0.9$ in both cohorts) and highly conserved in primates. In particular, rs13035290 possesses high phastCons [36] and

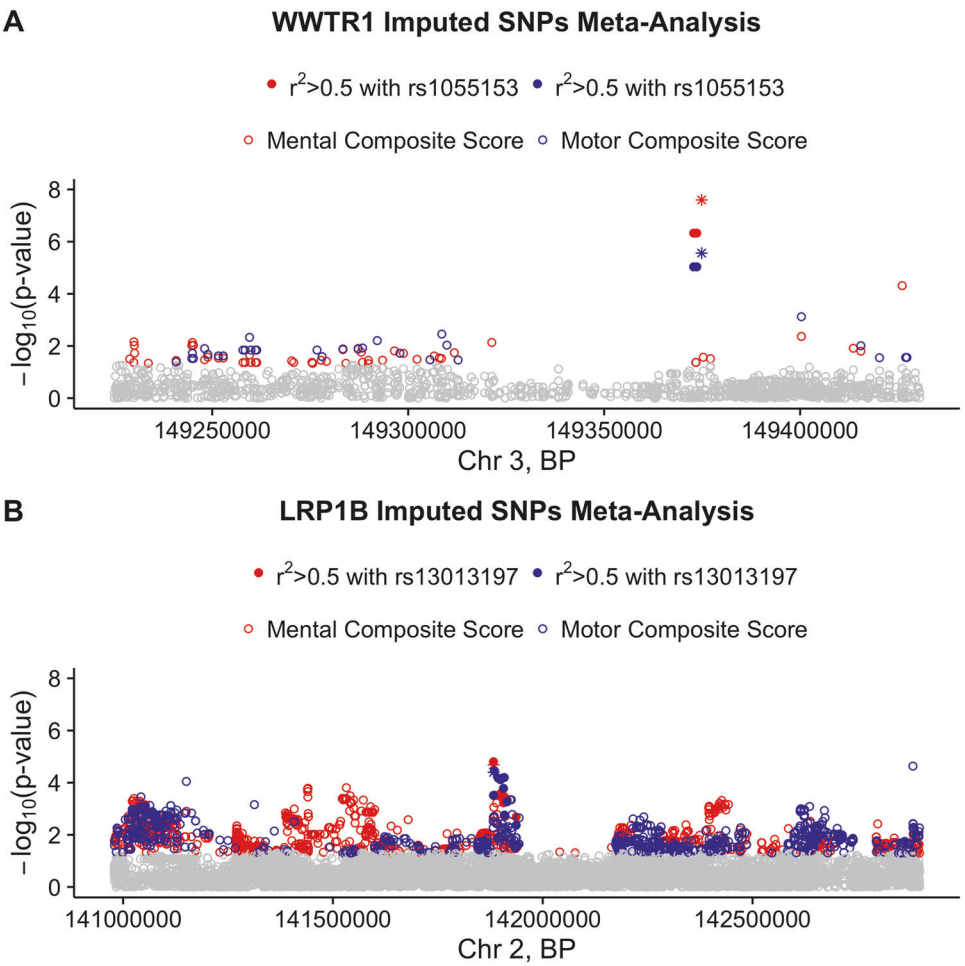

Fig. 2 Local Manhattan plot for association with mental composite score and motor composite score in *WWTR1* and *LRP1B*. Meta-analysis *P*-values for both originally genotyped SNPs and imputed SNPs at *WWTR1* (**a**) and *LRP1B* (**b**). Asterisk (*) used to denote the top SNPs from the initial analysis (rs1055153 in *WWTR1* and rs13013197 in *LRP1B*). Solid icons denote SNPs that have squared Pearson correlation coefficient $r^2 \geq 0.5$ with the top SNPs in both cohorts. SNPs with meta-analysis *P*-values less than 0.01 are shown in gray to demonstrate contrast in gene-wide association levels between *WWTR1* and *LRP1B*

phyloP [37] scores (Supplementary Table 8), indicating evolutionary pressures and an elevated level of functionality. No other SNPs at this locus demonstrate similar levels of conservation or other measures of functionality. These data suggest that a primary SNP of importance in *LRP1B* is rs13035290.

Replication, genetic correlation with other cognitive traits, and SNP-sense heritability

Several extremely large GWAS have recently been conducted for other cognitive traits, and we searched studies offering publicly available summary statistics for replications of our top associations. Specifically, we focused on GWAS of educational attainment [18] and intelligence [13], both conducted with sample sizes of over 75,000 subjects. In these studies none of rs1055153, rs13013197, or rs13035290 demonstrated any evidence of association with the main outcome ($P > 0.05$). However, further analysis indicated that replication in these GWAS and other similar studies was generally unlikely to be very successful. Using LD Score Regression [38], we estimated that the genetic correlation between mental composite score and educational attainment was only 0.12. The genetic correlation with intelligence was slightly higher at 0.39. All genetic correlations with other available cognitive GWAS [4, 14, 17] fell below 0.12, including a GWAS of intelligence among 6–18 year olds. These low correlations provide confirmation that infant neurodevelopment processes are in many ways distinct from the pathways that affect cognition in older children and adults. Finally we used GCTA [39] to estimate SNP-sense heritability in both mental and motor composite scores. For mental composite score we estimated SNP-based heritability of 0.34 in the Mexico cohort and 0.54 in the Bangladesh cohort, evaluations roughly matching the heritability of adult intelligence. Estimates for motor composite score did not converge.

Discussion

To our knowledge, this study is the first to perform a genome-wide search for variants associated with cognitive and motor ability in infants. We found that a missense

Table 3 Association of SNP–metal interaction with infant mental and motor composite score for top two SNPs

Model	Effect size[a] (SD)	95% CI	*P*-value
rs1055153			
Lead			
Mental Composite Score	0.24 (0.36)	−0.47 to 0.95	5.07×10^{-1}
Motor Composite Score	0.15 (0.16)	−0.17 to 0.46	1.67×10^{-1}
Manganese			
Mental Composite Score	−0.65 (0.26)	−1.16 to −0.14	1.30×10^{-2}
Motor Composite Score	−0.27 (0.08)	−0.43 to −0.10	1.69×10^{-3}
rs13013197			
Lead			
Mental Composite Score	−0.07 (0.10)	−0.28 to 0.13	4.77×10^{-1}
Motor Composite Score	0.01 (0.06)	−0.12 to 0.13	9.11×10^{-1}
Manganese			
Mental Composite Score	−0.12 (0.19)	−0.50 to 0.25	5.15×10^{-1}
Motor Composite Score	−0.02 (0.07)	−0.15 to 0.11	7.53×10^{-1}

CI confidence interval, *SD* standard deviation

[a] Effect size refers to the estimated additional change in Bayley score for each unit increase in concentration (μg/dl) of metal exposure in umbilical cord blood, while holding all other covariates constant, for each additional minor allele

variant in the *WWTR1* gene was associated with lower cognitive ability at a genome-wide significant level. Certain functional annotations also predicted this SNP to be an extremely deleterious substitution.

WWTR1 is a transcriptional coactivator that is a key member of the Hippo signaling pathway. The Hippo pathway regulates growth and organ size by promoting and restricting cell proliferation and apoptosis, among other functions [40, 41]. Both *WWTR1* and the Hippo pathway have been linked to a variety of oncogenic activities [42–44]. Increasingly, there is also growing evidence [45, 46] of the association between the Hippo pathway and neuronal health.

The Hippo pathway has been shown to regulate proliferation [47] in neural progenitor cells, and it has also been demonstrated to play a role in the early-stage differentiation of neuronal cells [48]. Further reports have linked Hippo to development of the corpus callosum [49], cerebral cortical development [50], and craniofacial development [51], among other neurogenesis processes. In addition to developmental duties, the pathway also possesses responsibilities in preserving neuronal health [52], for example in the performance of dendritic maintenance [53].

This existing body of literature is consistent with our finding of a genome-wide significant missense variant in *WWTR1*. The above reports suggest that the minor allele at rs1055153 could be disrupting the normal signaling and regulatory functions of the Hippo pathway, disturbing the growth and differentiation of neuronal cells in children. It is possible that altering these biological mechanisms hinders the standard neurodevelopment process, leading to poorer cognitive performance. Further research into the relationship between *WWTR1* and neural cells is necessary to corroborate this link.

We also found suggestive evidence, including a significant region-based association, that the gene *LRP1B* was associated with cognitive ability. After imputation, the originally genotyped SNP that led us to investigate this gene no longer demonstrated the highest degree of association, and functional analysis using multi-omics data appeared to pinpoint rs13035290 as a key variant. *LRP1B* encodes a member of the low-density lipoprotein family; certain lipoproteins have previously been associated with cognition and neurological diseases [54].

A recent genome-wide scan found multiple SNPs in *LRP1B* to be associated with absence of Alzheimer's disease in a sample of seniors [55]. The authors hypothesized that haplotypes in the gene were perhaps associated with a lack of cognitive decline in aging. Transcriptomic analysis has also revealed that *LRP1B* is most highly expressed in the brain [56]. These reports support the idea that the protein encoded by *LRP1B* possesses certain influences in neurodevelopment, although additional research is still needed. Our evidence suggests that further investigations into *LRP1B* could illuminate more about the neurological impacts of lipoproteins.

One other intriguing aspect of our study is the finding that infant cognitive performance does not show high genetic correlations with adult intelligence or educational attainment. It has long been recognized that human brain development occurs through continual complex interactions of genetic factors and environmental exposures, with

different patterns of development in different regions [57, 58]. For example, areas associated with complex reasoning show greater heritability in older subjects [59]. In addition, gene–environment effects vary with age [20]. Our results supplement these previous findings from a new angle. Specifically, the low observed genetic correlations suggest that the many varied aspects of neurodevelopment are attributable to different sets of genetic factors acting over different stages of maturation, as opposed to one static collection of variants wielding influence at all ages. An interesting topic for future work would be an investigation of the different genetic themes that dominate infant and adult cognitive phenotypes and how they relate to, for example, neurocircuitry development.

Our study has a few limitations. First, because of the scarcity of genetic studies on infant neurodevelopment, we were forced to search for replicated associations in GWAS that were not very similar to our study. These other studies possess much more power, but they focus on phenotypes that are genetically dissimilar from infant cognitive ability, as we confirmed by calculating genetic correlations. Second, our study was performed in two rather distinct populations. Although the diagnostic plots appeared to demonstrate sufficient control of spurious associations due to population stratification, we may have also decreased our power to detect true association signals due to the meta-analysis strategy. Thirdly, we did not attempt to measure development over time. Further longitudinal studies are necessary to fully understand how genetic factors impact the rate of development. These studies could complement our work by controlling for baseline ability. Finally, it is not clear that our results will generalize to individuals of other ethnicities, for example, those of European descent.

In conclusion, we performed a genome-wide association study of infant cognitive and motor ability at age two, and we identified one SNP in *WWTR1* to possess a genome-wide significant association with mental composite score, a measure of cognitive ability calculated from the BSID-III. Imputation and region-based association testing with the *LRP1B* gene demonstrated that a set of SNPs in this gene may also possess a combined effect on mental composite score. We believe the connections between *WWTR1*, *LRP1B*, and infant neurodevelopment are worthy of further study and can contribute significantly to the understanding of neurologic maturation in young children.

Acknowledgements We thank Maegan Harden, PhD, Broad Institute of MIT and Harvard, Boston, MA for assistance in genotyping the Bangladesh cohort. We thank Hakon Hakonarson, MD, PhD and Cecelia Kim, PhD, from the Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA for assistance in genotyping the Mexico cohort. This work was supported in part by National Institute of Environmental Health Sciences grants R01-ES007821, R01-ES014930, R01-ES013744, R01-ES022986, R01-ES015533, P42-ES016454, and P30-ES00002. Funding was also provided by Science to Achieve Results Research Assistance Agreement No. FP-91690001, awarded by the U.S. Environmental Protection Agency (EPA). The EPA has not officially endorsed this publication and the views expressed herein may not reflect the views of the EPA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Krushkal J, Murphy LE, Palmer FB, Graff JC, Sutter TR, Mozhui K, et al. Epigenetic analysis of neurocognitive development at 1 year of age in a community-based pregnancy cohort. Behav Genet. 2014;44:113–25.
2. Deary IJ, Whalley LJ, Lemmon H, Crawford JR, Starr JM. The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 Scottish Mental Survey. Intelligence. 2000;28:49–55.
3. Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. Mol Psychiatry. 2015;20:98–108.
4. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion M-J, et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Mol Psychiatry. 2014;19:253.
5. Strenze T. Intelligence and socioeconomic success: a meta-analytic review of longitudinal research. Intelligence. 2007;35:401–26.
6. Davis EE, Pitchford NJ, Limback E. The interrelation between cognitive and motor development in typically developing children aged 4–11 years is underpinned by visual processing and fine manual control. Br J Psychol. 2011;102:569–84.
7. Deary IJ, Johnson W, Houlihan LM. Genetic foundations of human intelligence. Hum Genet. 2009;126:215–32.
8. Desrivieres S, Lourdusamy A, Tao C, Toro R, Jia T, Loth E, et al. Single nucleotide polymorphism in the neuroplastin locus associates with cortical thickness and intellectual ability in adolescents. Mol Psychiatry. 2015;20:263–74.
9. Hayiou-Thomas ME. Genetic and environmental influences on early speech, language and literacy development. J Commun Disord. 2008;41:397–408.
10. Docherty SJ, Davis OS, Kovas Y, Meaburn EL, Dale PS, Petrill SA, et al. A genome-wide association study identifies multiple loci associated with mathematics ability and disability. Genes Brain Behav. 2010;9:234–47.
11. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry. 2011;16:996.
12. Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). Mol Psychiatry. 2015;20:183.
13. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JR, Krapohl E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet. 2017;49:1107.
14. Rietveld CA, Esko T, Davies G, Pers TH, Turley P, Benyamin B, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proc Natl Acad Sci. 2014;111:13790–4.

15. Trampush J, Yang M, Yu J, Knowles E, Davies G, Liewald D, et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. Mol Psychiatry. 2017;22:336.
16. Zabaneh D, Krapohl E, Gaspar HA, Curtis C, Lee SH, Patel H, et al. A genome-wide association study for extremely high intelligence. Mol Psychiatry. 2018;23:1226–32.
17. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013;340:1467–71.
18. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533:539–42.
19. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Mol Psychiatry. 2016;21:758–67.
20. Plomin R, von Stumm S. The new genetics of intelligence. Nat Rev Genet. 2018;19:148–59.
21. Davis OS, Haworth CM, Plomin R. Dramatic increase in heritability of cognitive development from early to middle childhood: an 8-year longitudinal study of 8,700 pairs of twins. Psychol Sci. 2009;20:1301–8.
22. Kile ML, Rodrigues EG, Mazumdar M, Dobson CB, Diao N, Golam M, et al. A prospective cohort study of the association between drinking water arsenic exposure and self-reported maternal health symptoms during pregnancy in Bangladesh. Environ Health. 2014;13:29.
23. Burris HH, Braun JM, Byun H-M, Tarantini L, Mercado A, Wright RJ, et al. Association between birth weight and DNA methylation of IGF2, glucocorticoid receptor and repetitive elements LINE-1 and Alu. Epigenomics. 2013; 5:271–81.
24. Wang Z, Henn BC, Wang C, Wei Y, Su L, Sun R, et al. Genome-wide gene by lead exposure interaction analysis identifies UNC5D as a candidate gene for neurodevelopment. Environ Health. 2017;16:81.
25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.
26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
27. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, et al. A first-generation linkage disequilibrium map of human chromosome 22. Nature. 2002;418:544–8.
28. Barnett I, Mukherjee R, Lin X. The generalized higher criticism for testing SNP-set effects in genetic association studies. J Am Stat Assoc. 2017;112:64–76.
29. Bellinger D, Leviton A, Waternaux C, Needleman H, Rabinowitz M. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. N Engl J Med. 1987;316:1037–43.
30. Claus Henn B, Ettinger AS, Schwartz J, Tellez-Rojo MM, Lamadrid-Figueroa H, Hernandez-Avila M, et al. Early postnatal blood manganese levels and children's neurodevelopment. Epidemiology. 2010;21:433–9.
31. Sun R, Carroll RJ, Christiani DC, Lin X. Testing for gene-environment interaction under exposure misspecification. Biometrics. 2017;74:653–62.
32. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
34. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073.
35. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010;6:e1001025.
36. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005;15:1034–50.
37. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010;20:110–21.
38. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33:272–9.
39. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Human Genet. 2011;88:76–82.
40. Genevet A, Tapon N. The Hippo pathway and apico-basal cell polarity. Biochem J. 2011;436:213–24.
41. Pan D. The hippo signaling pathway in development and cancer. Dev Cell. 2010;19:491–505.
42. Xie M, Zhang L, He CS, Hou JH, Lin SX, Hu ZH, et al. Prognostic significance of TAZ expression in resected non-small cell lung cancer. J Thorac Oncol. 2012;7:799–807.
43. Li Z, Wang Y, Zhu Y, Yuan C, Wang D, Zhang W, et al. The Hippo transducer TAZ promotes epithelial to mesenchymal transition and cancer stem cell maintenance in oral cancer. Mol Oncol. 2015;9:1091–105.
44. Moroishi T, Hansen CG, Guan KL. The emerging roles of YAP and TAZ in cancer. Nat Rev Cancer. 2015;15:73–79.
45. Sun Y, Yong KM, Villa-Diaz LG, Zhang X, Chen W, Philson R, et al. Hippo/YAP-mediated rigidity-dependent motor neuron differentiation of human pluripotent stem cells. Nat Mater. 2014;13:599–604.
46. Han D, Byun S-H, Park S, Kim J, Kim I, Ha S, et al. YAP/TAZ enhance mammalian embryonic neural stem cell characteristics in a Tead-dependent manner. Biochem Biophys Res Commun. 2015;458:110–6.
47. Cao X, Pfaff SL, Gage FH. YAP regulates neural progenitor cell number via the TEA domain transcription factor. Genes Dev. 2008;22:3320–34.
48. Ahmed AF, de Bock CE, Lincz LF, Pundavela J, Zouikr I, Sontag E, et al. FAT1 cadherin acts upstream of Hippo signalling through TAZ to regulate neuronal differentiation. Cell Mol Life Sci. 2015;72:4653–69.
49. Lavado A, He Y, Pare J, Neale G, Olson EN, Giovannini M, et al. Tumor suppressor Nf2 limits expansion of the neural progenitor pool by inhibiting Yap/Taz transcriptional coactivators. Development. 2013;140:3323–34.
50. Cappello S, Gray MJ, Badouel C, Lange S, Einsiedler M, Srour M, et al. Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development. Nat Genet. 2013;45:1300–8.
51. Wang J, Xiao Y, Hsu CW, Martinez-Traverso IM, Zhang M, Bai Y, et al. Yap and Taz play a crucial role in neural crest-derived craniofacial development. Development. 2016;143:504–15.

52. Pfleger CM. The Hippo pathway: a master regulatory network important in development and dysregulated in disease. Curr Top Dev Biol. 2017;123:181–228.
53. Emoto K, Parrish JZ, Jan LY, Jan YN. The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. Nature. 2006;443:210–3.
54. Huang X, Chen H, Miller WC, Mailman RB, Woodard JL, Chen PC, et al. Lower low-density lipoprotein cholesterol levels are associated with Parkinson's disease. Mov Disord. 2007;22:377–81.
55. Poduslo SE, Huang R, Spiro A 3rd. A genome screen of successful aging without cognitive decline identifies LRP1B by haplotype analysis. Am J Med Genet B Neuropsychiatr Genet. 2010;153B:114–9.
56. John L, Jeffrey T, Mike S, Rebecca P, Edmund L, Saboor S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580.
57. Lenroot RK, Giedd JN. The changing impact of genes and environment on brain development during childhood and adolescence: initial findings from a neuroimaging study of pediatric twins. Dev Psychopathol. 2008;20:1161–75.
58. Wright IC, Sham P, Murray RM, Weinberger DR, Bullmore ET. Genetic contributions to regional variability in human brain structure: methods and preliminary results. Neuroimage. 2002;17:256–71.
59. Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen VP, Huttunen M, et al. Genetic influences on brain structure. Nat Neurosci. 2001;4:1253–8.