Dyslexia-Related Loci Are Significantly Associated With Language And Literacy In Chinese-English Bilingual Hong Kong Chinese Twins

Cheuk Yan Chung[1†], Dora Jue Pan[2†], Silvia Paracchini[3], Wenxuan Jiang[1], Hon-Cheong So[4,5,6,7], Catherine McBride[8], Urs Maurer[9,10,11], Mo Zheng[12†], Kwong Wai Choy[1,13*†]

[1*] Department of Obstetrics & Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China.
[2] School of Humanities and Social Science, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China.
[3] School of Medicine, University of St Andrews, St. Andrews, Scotland.
[4] School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China.
[5] KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research of Common Diseases, Kunming Institute of Zoology and The Chinese University of Hong Kong, Hong Kong SAR, China.
[6] Department of Psychiatry, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China.
[7] Hong Kong Branch of the Chinese Academy of Sciences Center for Excellence in Animal Evolution and Genetics, The Chinese University of Hong Kong, Hong Kong SAR, China.
[8] College of Health and Human Sciences, Purdue University, Indiana, the United States.
[9] Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China.
[10] Centre for Developmental Psychology, The Chinese University of Hong Kong, Hong Kong, China.
[11] Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong, China.
[12] Division of Science & Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China.
[13] Joint BCM-CUHK Center of Medical Genetics, Hong Kong SAR, China.

*Corresponding author(s). E-mail(s): richardchoy@cuhk.edu.hk;
Contributing authors: tchung19g@link.cuhk.edu.hk; panjue@cuhk.edu.cn; sp58@st-andrews.ac.uk; jiangwx@link.cuhk.edu.hk; hcso@cuhk.edu.hk; mcbridca@purdue.edu; umaurer@psy.cuhk.edu.hk; zhengmo.hk@gmail.com;
† These authors contributed equally to this work.

Abstract

A recent genome-wide association study on dyslexia in 51,800 affected European adults and 1,087,070 controls detected 42 genome-wide significant single nucleotide variants (SNPs). The association between rs2624839 in *SEMA3F* and reading fluency was replicated in a Chinese cohort. This study explores the genetic overlap between Chinese and English word reading, vocabulary knowledge and spelling, and aims at replicating the association in a unique cohort of bilingual (Chinese-English) Hong Kong Chinese twins. Our result showed an almost complete genetic overlap in vocabulary knowledge (r^2=.995), and some genetic overlaps in word reading and spelling (r^2=.846, .687) across the languages. To investigate the region near rs2624839, we tested proxy SNPs (rs1005678, rs12632110 and rs12494414) at the population level (n=305-308) and the within-twin level (n=342-344 [171-172 twin pairs]). All the three SNPs showed significant associations with quantitative Chinese and English vocabulary knowledge (p<.05). The strongest association after multiple testing correction was between rs12494414 and English vocabulary knowledge at the within-twin level (p=.004). There was a trend of associations with word reading and spelling in English but not in Chinese. Our result suggested that the region near rs2624839 is one of the common genetic factors across English and Chinese vocabulary knowledge and unique factors of English word reading and English spelling in bilingual Chinese twins. A larger sample size is required to validate our findings. Further studies on the relationship between variable expression of *SEMA3F*, which is important to neurodevelopment, and language and literacy are encouraged.

Keywords: Dyslexia, Word reading, Word spelling, Vocabulary knowledge, Twin study, Chinese-English bilinguals

Introduction

Early language and literacy development is associated with school achievement. Difficulties in language and literacy may decrease the quality of life, including psychosocial function, physiological and mental health, and living environment and satisfaction (Zou et al., 2022). Early prediction of language and literacy variability is a global concern. Many studies focused on dyslexia but not normal variations in word reading. However, it is important to examine the potential genetic factors of word reading in the unselected population. It is well known that genetics play an important role in word reading. Additive genetic variance explains 73% of Chinese reading fluency (Chow et al., 2011) and 64-73% of English reading fluency (Tosto et al., 2017). In addition, both word spelling and vocabulary knowledge are significantly associated with word reading across languages (Kim, 2020; Perfetti, 2010) and these are all important language and literacy skills for children's academic development. Therefore, it is interesting to explore whether the potential factors that are associated with word reading also predict word writing and

vocabulary knowledge. As most of the studies on language and literacy were performed in European populations, there is a gap in knowledge in different ethnicities and language systems. It calls for replication studies in non-European populations to have an overall view.

A study of bilingual twins using the bivariate ACE model suggested that there are both genetic overlap and non-overlap between first language (L1) and second language (L2) (Wong et al., 2014). It is important to investigate if there are and what are the common genetic associations of different languages. To date, there is only one study that revealed a positive result of little genetic overlap. rs6456593 in *DCDC2* is associated with both Chinese (L1) and French, German, or Spanish (third language, L3). It is yet very likely that the association of *CNTNAP2* with English is language-specific (Wong et al., 2022). Further investigation of the common genetic factors between languages is needed. A limitation of genetic studies is that the results are likely to be influenced by environmental factors. A meta-analysis of the heritability of reading demonstrated remarkable shared and unique environmental factors besides genetic factors (Andreola et al., 2021). For example, family socioeconomic status affects children's reading (Chen et al., 2018). To facilitate the study of purely genetic effects, the multilevel model of co-twin causality is a well-designed approach to separate genetic factors from environmental influences (Friedman et al., 2021). It can measure more accurately direct genetic factors that are adjusted for demographic and indirect genetic effects of parents (Howe et al., 2022).

Recently, a GWAS on self-reported dyslexia was performed on 51,800 affected European adults and 1,087,070 controls (Doust et al., 2022). Of the 42 genome-wide significant loci, five SNPs, including rs12939690:G>C (a proxy for rs34349354:C>CAG), rs2624839:T>C, rs2820882:A>T (a proxy for rs35570426:T>TG), rs62453457:A>G, and rs12453682:C>T, showed significant associations with Chinese reading accuracy and/or fluency ($p < .05$) in a replication cohort of 2,476 Northern Chinese although none of them passed the multiple testing correction. Notably, only two SNPs, rs62453457:A>G and rs2624839:T>C, had a consistent allelic direction, which indicated replicable results. The allele G of rs62453457 and the allele C of rs2624839 were consistently the risk alleles for dyslexia in both cohorts. The minor allele frequency of rs2624839 in East Asian population is .372; that of rs62453457 in East Asian population is .100 (*dbSNP*, n.d.). rs62453457 was not reported to be associated with language-related phenotypes and neurodevelopment. Due to the low minor allele frequency of rs62453457 and our limited sample size, rs62453457 would not be the main interest of this study.

rs2624839, located at intron 2 of *SEMA3F* (NM_004186.5), seats at a DNase hypersensitive site and a promoter region of *SEMA3F*. It is supported by the *GeneHancer*, the *ENCODE Registry of candidate cis-Regulatory Elements (cCREs)* and the *H3K27AC Mark* (Abascal et al., 2020; Fishilevich et al., 2017) (**Supplementary figure 1**). rs2624839 was associated with other language-related and neurocognitive traits. The T allele of rs2624839 was associated with increased general cognitive function (Davies et al., 2018) and variation in intelligence (Hill et al., 2018; Savage et al., 2018) in European populations. Semaphorin 3F (Sema3F), the protein coded by *SEMA3F*, is an important gene for neurodevelopment. It interacts with Npn2 and PlexA3 to induce homeostatic scaling in cortical neurons (Wang et al., 2017) and spine pruning in dendritic spines (Duncan et al., 2021). It also organizes cranial nerve projections in the rostral forebrain, midbrain, and hippocampus to establish limbic tracts and controls the development of amygdaloid circuitry (Sahay et al., 2003). The association between rs2624839 and

dyslexia or reading-related traits is yet to be validated. Therefore, it is of interest to investigate the region near rs2624839 in more detail.

In this study, we aimed at replicating the association of dyslexia-related rs2624839 locus with quantitative word reading in unselected Hong Kong Chinese twins at the population level (i.e. one twin from each family) and the within-twin level (i.e. co-twin causality model). As the core symptoms of dyslexia are deficits in word reading that often co-occur with problems of word spelling (Ho et al., 2004; Maughan et al., 2009; McBride-Chang et al., 2011), we measured children's word reading, word spelling, and oral vocabulary knowledge in both Chinese (L1) and English (L2). First, we hypothesized that there is a genetic overlap between Chinese and English language and literacy. We measured the genetic overlap across Chinese and English using a bivariate ACE model. Second, we hypothesized that the region/genes near rs2624839 but not the direct functional effect of rs2624839 are associated with dyslexia. Therefore, we expanded the target to the SNPs with high linkage equilibrium with rs2624839 in Southern Han Chinese population from the *Ensembl* database. With the genotypes available in our database, we tested three proxy SNPs for rs2624839 (rs1005678:C>G, rs12632110:A>G, and rs12494414:C>T) that are in linkage disequilibrium with rs2624839 in Southern Han Chinese population (**Supplementary figure 2 and 3**) (Cunningham et al., 2022). The two levels of genetic association analysis (population level and within-twin level) allow us to determine the degree to which demographic and/or indirect genetic effects contribute to the phenotypes. Third, we hypothesized that the region near rs2624839 contributed to the common genetic factor of language and literacy. Hereby, we tested if it is associated with language across Chinese and English. Besides, we tested if the association is common or specific for written language (i.e. word reading and spelling) and spoken language (i.e. oral vocabulary knowledge). A deeper level of genetic influence across languages was investigated. Studying the proxies for rs2624839 enables us to confirm the association of the region near rs2624839 with different language and literacy skills, providing insight into possible genotype-phenotype relationships.

Materials and Methods

Participants

Participants were recruited for the Chinese-English Twin Study of Biliteracy, an ongoing longitudinal twin study on the genetic and environmental influences on bilingual development of Chinese children, since 2015. Participants had Cantonese as their native language and studied in Hong Kong primary schools. Ethics approval by the Chinese University of Hong Kong (CREC Ref. No.: 2021.237) and informed written consent from participants and their parents were obtained. In total, 618 subjects were recruited and genotyped by SNP array, including 172 pairs of dizygotic twins (38 male dizygotic twin pairs; 36 female dizygotic pairs; 98 opposite-sex dizygotic pairs) and 137 pairs of monozygotic twins (59 male monozygotic twin pairs; 78 female monozygotic twin pairs). The age range was from 5.4 to 10.8 years.

Phenotypic measurements

Word reading in Chinese and English

The Chinese word reading task was a subtest of the Hong Kong Test of Specific Learning Difficulties in Reading and Writing for Primary School Students-Second Edition (Ho et al., 2007). This is a widely used diagnostic test to assess children's reading abilities with local norms. One hundred-fifty Chinese two-character words were shown to children, and they needed to read them out in order. One point was awarded for each correctly read word. Testing was discontinued when children encountered 15 consecutive failures. The maximum score was 150.

The English word reading task was adapted from the test used in a previous study by Tong and McBride-Chang (Tong &McBride-Chang, 2010). Ten easy items were deleted from the original list due to the floor effect. The current test consisted of 50 English words in total. All words were selected from English textbooks used in Hong Kong kindergartens and primary schools. The test items were arranged in increasing reading difficulty. One point was given for each correct answer. The maximum possible score for this measure was 50.

Vocabulary knowledge in Chinese and English

The Chinese measure consisted of 48 items presented in ascending difficulty (Tong et al., 2018) The first 10 items on receptive vocabulary were presented with four pictures along with an orally presented Chinese word. The participants were asked to choose the picture that matched the Chinese word. In the following 12 items on expressive vocabulary, the children were asked to name a given picture in Chinese (e.g., compass, wrist). Each item was scored either 0 or 1 in these 22 items. The last 26 items on vocabulary definitions were orally presented with a Chinese word that represented a concept or an object. The participants were asked to define the word. A three-point scale (ranging from 0-2) was used to score their answers. Participants were given a score of either 0, 1, or 2 for each item according to how close their response was to the model answer. The maximum possible score for this measure was 74.

The English task was similar to the Chinese one in structure and scoring. It included 45 items presented with ascending difficulty (Tong et al., 2018). For the first 15 items on receptive vocabulary, the participants were asked to choose the right picture among four alternatives that best represent an orally presented English word. In the next 15 items on expressive vocabulary, the participants were asked to name a picture in English (e.g., writing, globe). Each item was scored either 0 or 1 in these 30 items. The last 15 items were on vocabulary definitions. The participants were asked to give the definition of an orally presented English word. A three-point scale was used to score these items. Participants were given a score of either 0, 1, or 2 for each item according to how close their response was to the correct answer. The maximum possible score for this measure was 60.

Spelling in Chinese and English

Two writing-to-dictation tasks were used to measure participants' spelling ability in Chinese and English, respectively (Lo et al., 2018; Tong et al., 2009).

In the Chinese task, participants were asked to write a number of Chinese two-character words. It included 20 words in the task. Words were presented in ascending

difficulty. One point was given if children could write down either character correctly, and two points were given if they could write the whole word correctly. The maximum possible score for this task was 40.

In the English task, participants were asked to write 12 English words (e.g., pancake) in the English spelling task. The words were composed of one to four syllables and arranged in order of increasing difficulty. One point was given for each correctly spelled morpheme or word within the compound words, and two points were given if children could spell the whole word correctly. The maximum possible score for this task was 24.

Considering the wide age range of children in our study, all the participants' Chinese and English phenotypic scores were regressed on age and standardized residuals were used as age-adjusted scores.

Bivariate ACE model

We conducted a bivariate genetic analysis using the classic twin design in which phenotype variance was partitioned into that due to additive genetic (A), shared environmental (C) and non-shared environmental influence (E). The proportions of variance explained by each component of influence can be estimated with the structural equation modeling technique, and the proportion of variance explained by the additive genetic influence (A) gives heritability. The bivariate ACE model was set up for the Chinese and English phenotype using the standard Cholesky decomposition method (Neale &Cardon, 1992). This method simultaneously decomposed the variance and covariance of the two phenotypes into separate components which can be represented in a path diagram with six latent factors A1, C1, and E1, and A2, C2, and E2 (**Figure 1**). This procedure assumes that the latent variables A1, C1, and E1 are the sole causes of Chinese phenotype as well as partial causes of English phenotype. The factors A2, C2, and E2 account for the residual variance in English phenotype that is not shared with Chinese phenotype. We specified the six latent factors in standardized metrics and allowed the nine paths from latent factors to phenotype (a11, a12, a22, c11, c12, c22, e11, e12, and e22) freely to vary. Our analysis started with the full bivariate ACE model and estimated all nine parameters. Since we are interested in the genetic influences, we focused on the estimation of the three genetic effects a11, a22, and a12, which enable us to calculate the heritability of Chinese phenotype, the heritability of English phenotype, and the genetic correlation between Chinese and English. Then we tested whether the full model could be modified to a more parsimonious model by constraining one of the genetic effects to be zero. The test in model fit statistics between the restricted model and the full model would indicate whether the constrained parameter is significant or not. Specifically, we tested [1] whether the genetic correlation between Chinese and English phenotypes is significant by constraining a12=0, and [2] whether the Chinese and English phenotypes share the same genetic influence and whether there is no unique genetic effect for either Chinese or English by constraining a22=0.

Because our study contains three phenotypes in Chinese and English, we conducted three bivariate ACE analyses respectively for word reading, vocabulary knowledge, and spelling. Our analysis was performed using the OpenMx software package 2.20.6 (Boker et al., 2011) with the full information maximum likelihood estimation. The program to estimate heritability was adapted from the OpenMX scripts distributed at the International Workshop on Statistical Genetic Methods for Human Complex Traits (Neale et al., 2017).

DNA extraction and genotyping

Two ml of saliva were collected with Oragene•DNA (OG-500) collection kit (DNA Genotek Inc, Canada) from each subject. DNA extraction was performed using prepIT-L2P reagent (DNA Genotek Inc, Canada). DNA concentration and A260/280 indicating purity were measured by NanoDrop™ One (Thermo Scientific, USA).

DNA samples were genotyped by Infinium OmniZhongHua-8 v1.3 BeadChip (Illumina, USA), which covers 878,291 markers specific to the Chinese population according to genome-wide association studies. 200 ng DNA was required for the Infinium HD Super Assay workflow. The hybridized chips were scanned by the iScan system (Illumina, USA), where the performance of the array was ensured by a call rate of more than 99% and a Log R deviation of less than .30. The genotyping data were visualized by Illumina GenomeStudio (Illumina, USA). Only one twin from each monozygotic twin pair was genotyped; both twins from each dizygotic twin pair were genotyped.

The Phase 2 HapMap dataset of Han Chinese in Beijing, China plus Japanese in Tokyo, Japan was used for finding proxy SNPs of rs2624839 (NC_000003.11:g.50202231T>C). SNPs in r^2 > .8 with rs2624839 and minor allele frequency (MAF) > .05 were included. After screening available SNP data in our SNP array, rs1005678 (NC_000003.12:g.50172856C>G), rs12632110 (NC_000003.12:g.50186792A>G), and rs12494414 (NC_000003.12:g.50285741C>T) were eventually selected for genetic association analysis. Tests for Hardy-Weinberg equilibrium and allele, genotype frequency were performed by PLINK 1.9 (Purcell et al., 2007). The three SNPs had a minor allele frequency (MAF) > .05 and no deviation from the Hardy-Weinberg equilibrium (HWE) (p > .05) in our cohort (**Supplementary table 1**).

Statistical analyses

Test for relatedness between samples was performed by PLINK 1.9 (Purcell et al., 2007). A pair of samples from different families had a PIHAT > .1, and therefore, one of the families was randomly removed. Statistical analysis of association was done by IBM Statistical Product and Service Solutions (SPSS) Statistics 26 (IBM, USA). Age-adjusted, standardized scores were obtained for all the phenotypic scores. Gender differences on the performance were checked by t-test, which males performed significantly worse than females in Chinese spelling. Regression was done against six phenotypes respectively, including Chinese word reading, Chinese vocabulary knowledge, Chinese spelling, English word reading, English vocabulary knowledge, and English spelling. For the population level, linear regression was performed fitting an additive model and gender as a covariate. One twin of each family was randomly selected for analysis (n = 308). For the within-twin level, all dizygotic twins were included (n = 342 [171 twin pairs]); mean sibship genotype and deviation of the individual's genotype from mean sibship genotype were regressed on phenotypes (Friedman et al., 2021; Howe et al., 2022) and gender as a covariate. Mean sibship genotype is the average of the number of risk allele(s) that a twin pair carries; deviation of the individual's genotype from mean sibship genotype is the difference between the number of risk allele(s) that a twin member carries and the mean sibship genotype. The allele C of rs1005678, the allele G of rs12632110, and the allele C of rs12494414 that were in high correlation with the allele T of rs2624839 were defined as the reference alleles. Since the three SNPs are highly correlated (r^2 > .8, p < .01) (**Supplementary table 2**), the significance was not corrected for the number of SNPs.

The three phenotypes of English language and literacy were highly correlated ($r^2 > .8$, $p < .01$) (**Table 2**) while the three phenotypes of Chinese language and literacy were not. Therefore, the significance was only corrected for multiple testing of three Chinese phenotypes and English phenotypes as one by Bonferroni correction ($p = .05/4 = .0125$).

Results

Heritability and genetic overlap between Chinese and English language and literacy

The descriptive statistics of phenotypes are in **Table 1** and **Supplementary figure 4(a)-(f)**. Since the measures were standardized, the means were close to 0. They were in a normal distribution as the skewness of the phenotypes was less than ±1. Correlations between Chinese measures ranged from .375 to .678, which were moderate (**Table 2**). Correlations between English measures ranged from .825 to .942, which were strong. However, correlations between measures in L1 Chinese and L2 English were weak to moderate, ranging from .025 to .353. Therefore, it is of interest to investigate the degree of genetic correlations between Chinese and English language and literacy.

Three bivariate ACE models were tested. The standardized parameter estimates and the calculated heritability for the three phenotypes in Chinese and English were presented separately for each phenotype (**Table 3**). Chinese word reading has the highest heritability of .524, followed by Chinese vocabulary knowledge at .425, and Chinese spelling with the lowest heritability of .194. In addition, 38.4% of the variation in Chinese word reading, 29.0% of the variation in Chinese vocabulary knowledge, and 54.1% of the variation in Chinese spelling were explained by shared environmental influence. With regards to English language, the heritability for word reading, vocabulary knowledge, and spelling was .281, .104, and .272 respectively. Most of the variation in English phenotypes was explained by shared environmental influences, which accounted for 64.8% of the variation in English word reading, 82.2% in English vocabulary knowledge, and 64.2% in English spelling. Finally, the bivariate ACE model enabled us to estimate the correlation between the genetic factors for Chinese and English phenotypes. The correlation between Chinese and English genetic factors was .846 for word reading, .995 for vocabulary knowledge, and .687 for spelling respectively, suggesting a high genetic overlap between the two languages. In particular, the genetic factors for Chinese and English Vocabulary Knowledge had an almost perfect correlation. This indicates that the genetic underpinnings that influence vocabulary knowledge might be identical for Chinese and English.

To further understand whether Chinese and English languages have shared or unique genetic influences, we tested two restricted ACE models for each of the three phenotypes. In Mode 1, we constrained the shared genetic effect to be zero (a12=0) which means Chinees and English have different genetic influences and they are completely independent. Compared to the full ACE model, the restricted model resulted in a significant loss in likelihood function statistics (-2LL) for all three phenotypes and is therefore rejected (**Table 4**)**.** This suggested that a significant correlation exists between the genetic factors that influence Chinese and English language.

Next in Model 2, we tested whether English language has a unique genetic influence that is not shared with Chinese by constraining the genetic effect for English to be zero (a22 = 0). This resulted in a significant loss in model fit for word reading ($p <$

.001), but not for vocabulary knowledge (p = .976) and spelling (p = .222). This indicated that although genetic influence for Chinese and English word reading are highly correlated (.846), they are not exactly the same. On the other hand, the non-significant result for vocabulary knowledge confirmed our observation in **Table 3**, which shows that the genetic effect for Chinese and English vocabulary knowledge has a nearly perfect correlation. We could conclude that there is no unique genetic effect that influences English Vocabulary Knowledge. Finally, the non-significant result for spelling seems to suggest that Chinese and English spelling share common genetic influence, and there is no unique genetic factor for English Spelling. However, considering the low heritability of both Chinese and English Spelling (.194 for Chinese and .272 for English), we need to take caution in concluding that genetic underpinnings that influence Chinese and English spelling are identical. The non-significant unique genetic effect for English Spelling may be simply because only a small proportion of variance in English Spelling was explained by the genetic effect.

To summarize, our model fit results show that genetic underpinnings that influence Chinese and English word reading are highly correlated but distinct. In contrast, genetic factors that influence Chinese and English vocabulary knowledge are nearly identical. Chinese and English spelling also had shared genetic effect but caution need to be taken in explaining the overlap. Since some genetic overlaps in word reading, vocabulary knowledge and spelling were identified, we tested if the region near rs2624839 explains the genetic overlaps.

Association analysis with Chinese and English language and literacy

The strongest association was with vocabulary knowledge, which all three SNPs showed association (**Table 5**). At the population level, rs12632110 was nominally significantly associated with Chinese vocabulary knowledge (p = .040). At the within-twin level, rs1005678 and rs12494414 were significantly associated with English vocabulary knowledge (p = .005, p = .004) after multiple testing correction; rs12632110 showed nominally significant association with English vocabulary knowledge (p = .029). rs1005678 and rs12632110 showed nominally significant association with Chinese vocabulary knowledge (p = .025, p = .020). Our results showed that the allele G of rs1005678, the allele A of rs12632110, and the allele T of rs12494414 were the risk alleles for worse performance of vocabulary knowledge.

There was a trend of association of rs1005678 and rs12494414 with English word reading at the within-twin level (p = .081, p = .075) but none of them reached the significance level (**Supplementary table 3**). Similarly, there was no significant association for spelling, yet a trend of association of rs1005678 and rs12494414 with English spelling at the within-twin level (p = .057, p = .060) (**Supplementary table 4**).

Discussion

This study showed that word reading, vocabulary knowledge, and spelling had different degrees of genetic overlap across Chinese and English. The greatest genetic overlap was observed across Chinese and English vocabulary knowledge while the least was across Chinese and English spelling. We also replicated the association between rs2624839 and dyslexia by Doust for the first time in an unselected population of Chinese bilingual twins using quantitative vocabulary knowledge, word reading, and spelling. Investigation into

the proxies for rs2624839 (rs1005678, rs12632110, and rs12494414) proved the hypothesis that the region near rs2624839 affects language and literacy. The allele C of rs2624839, the allele G of rs1005678, the allele A of rs12632110, and the allele T of rs12494414 were the risk alleles. The allele direction was consistent with that reported (Doust et al., 2022). Besides, we showed an almost complete genetic overlap in vocabulary knowledge across Chinese and English and suggested that the region near rs2624839 may be one of the common genetic factors. The region near rs2624839 may also be a genetic factor of vocabulary knowledge and spelling confined to English but not in Chinese.

Our bivariate ACE model results found substantial genetic correlations between Chinese language and parallel English phenotypes. The genetic correlation reflects the extent to which the genetic factors underlying Chinese language overlap with the genetic factors that influence English in the model. In our study a near perfect genetic correlation (.995) was observed between Chinese and English vocabulary knowledge, indicating that the genetic factors that influence both traits are identical. We also found high genetic correlations for Chinese and English Word readings (.846), indicating that more than 80% of genetic factors influencing Chinese reading also have an influence on reading English word reading. These results are very similar to another twin study conducted in Hong Kong (Wong et al., 2014), which reported a perfect genetic correlation of 1.0 for receptive vocabulary and .90 for word reading. However, our study also additionally computed the genetic correlations between Chinese and English spelling (.687), which has never been reported before. These results suggested that though there is a high genetic correlation across Chinese and English, the degree of overlap in the genetic mechanisms underlying different phenotypes are distinct.

Genetic analysis on dyslexia or reading-related phenotypes was usually conducted in the European population, including the large cohort UK Avon Longitudinal Study of Parents and their Children (Luciano et al., 2013; Paracchini et al., 2008; Scerri et al., 2011), Australian Brisbane Adolescent Twin Study (Lind et al., 2010), and German population (Müller et al., 2016; Wilcke et al., 2009). There are three challenges to these kinds of studies: [1] Some of the significant associations cannot be replicated in other cohorts (Venkatesh et al., 2013) or survive after multiple testing correction (Waye et al., 2017). These inconsistent findings call for more replication studies across different populations. [2] Most studies typically used a qualitative design which may have potential sampling bias and population stratification. It could generate a bias of allele frequency in the sub-group of the extremes with z-score ±2, producing false-positive results (Tam et al., 2019). [3] The cultural context may be an essential influential factor. It is unclear whether the findings observed in the European countries are replicable in cultures where people use a non-alphabetic system and adopt different educational policies and curriculum guidelines. Therefore, more cohorts of non-European ancestry are required to replicate the genome-wide significant SNP association, together with the need for multiple replication cohorts that use different languages to validate the association (Hatzikotoulas et al., 2014).

Our Hong Kong Chinese bilingual twin cohort here acted as a unique replication cohort to validate the association, providing interesting and important results in three ways. [1] It is a Chinese population rather than a European population. Where most genetic association studies on dyslexia and language ability were on European populations, a Hong Kong Chinese population would avoid the problem of population stratification and aid the replication of significant association. Moreover, our Hong Kong Chinese cohort is Southern Han Chinese, which differs from the Northern Chinese

replication cohort reported by Doust (Doust et al., 2022) and Northern Chinese cohort reported by Wang (Wang et al., 2023) and provides additional validation. [2] The participants in our study are bilingual twin children in Hong Kong. We can explore the genetic overlap and non-overlap across languages. It is also advantageous to examine the reported significant association with parallel Chinese and English parameters in the same cohort, compared with that Doust did not test both English skills and Chinese skills in the same cohort. Our replication study in bilingual Chinese twins is the first to extend the previous findings in dyslexia (Doust et al., 2022) to word reading skills, as well as Chinese and English vocabulary knowledge and English spelling, in typically developing children. Moreover, quantitative measurement instead of qualitative measurement was used to facilitate replication in an unselected population. [3] Our twin samples allow the analysis at the within-twin level. The co-twin causality model takes environmental factors that are known to highly affect reading ability into account. It maximumly avoids the potential effects of environmental factors and demonstrates the direct effects of genetic factors in a more stringent way. More importantly, using a similar sample size, the within-twin-level analysis can detect significant associations which the population-level analysis cannot.

rs2624839 and its three proxy SNPs located from *SEMA3F* to *LSMEM2* revealed that the region near rs2624839 is associated with language and literacy. rs2624839 and rs1005678 are located at intron 2 of *SEMA3F* (NM_004186.5); rs12632110 is located at intron 18 of *SEMA3F* (NM_004186.5); rs12494414 is located at intron 1 of *LSMEM2* (NM_153215.3). The SpliceAI prediction score is .00 to .02 for any type of damaging consequence (Jaganathan et al., 2019). As the impact of the intronic variants is predicted to be low, these SNPs may not functionally affect the coded protein Sema3F or leucine-rich single-pass membrane protein 2. Nevertheless, the regulation of *SEMA3F* may explain the significant association with language and literacy. rs2624839 and rs1005678 are in the promoter region of *SEMA3F* (Abascal et al., 2020; Fishilevich et al., 2017). Single-cell RNA sequencing data shows that *SEMA3F* is highly expressed in the human brain cortex, specifically in neural progenitor cells, excitatory neurons, and inhibitory neurons (Zhou et al., 2023). The expression pattern is coherent with the function of Sema3f, which is synaptic scaling (Duncan et al., 2021; Wang et al., 2017). Polymorphisms at the locus near rs2624839 may lead to variable expression of *SEMA3F* and impact the interaction of Sema3f with Npn2 and PlexA3 which results in synaptic plasticity. Further study may focus on the neighboring SNPs or genes of rs2624839, such as *SEMA3F*, to investigate their potential role in dyslexia, and languages other than Chinese and English.

Our study provided evidence of the genetic overlap in vocabulary knowledge across Chinese and English and the association of locus near dyslexia-associated rs2624839 with vocabulary knowledge in Chinese and English among Hong Kong children. We found that the genetic contribution by the region near rs2624839 is common to spoken language (i.e. oral vocabulary knowledge) across Chinese and English and unique to English written language (i.e. word reading and spelling). Besides, we suggested that the variable regulatory effect of polymorphism in locus near rs2624839 leads to the variation in language and literacy. Our result highlighted the importance of large-scale GWAS to produce replicable results from a different population cohort and the value of the within-twin study design to provide a robust estimation of direct genetic factors. However, a larger sample size is still warranted to validate the significant association. Furthermore, genetic studies on language and literacy ability are subject to the differences in age and neurodevelopment of participants. Children aged from 5.4 to

10.8 years were recruited for this study; therefore, a replication study of adults is encouraged to confirm the significant genetic association.

Reference

Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O.Al, Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H. A., Barozzi, I., …Zimmerman, J. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature 2020 583:7818*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

Andreola, C., Mascheretti, S., Belotti, R., Ogliari, A., Marino, C., Battaglia, M., &Scaini, S. (2021). The heritability of reading and reading-related neurocognitive components: A multi-level meta-analysis. *Neuroscience & Biobehavioral Reviews*, *121*, 175–200. https://doi.org/10.1016/J.NEUBIOREV.2020.11.016

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., &Fox, J. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317. https://doi.org/10.1007/s11336-010-9200-6

Chen, Q., Kong, Y., Gao, W., &Mo, L. (2018). Effects of socioeconomic status, parent-child relationship, and learning motivation on reading ability. *Frontiers in Psychology*, *9*(JUL), 1297. https://doi.org/10.3389/FPSYG.2018.01297/BIBTEX

Chow, B. W. Y., Ho, C. S. H., Wong, S. W. L., Waye, M. M. Y., &Bishop, D. V. M. (2011). Genetic and Environmental Influences on Chinese Language and Reading Abilities. *PLOS ONE*, *6*(2), e16640. https://doi.org/10.1371/JOURNAL.PONE.0016640

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., DaRin Fioretto, L., …Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/NAR/GKAB1049

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaars, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A.V., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., Derosse, P., …Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, *9*(1), 1–16. https://doi.org/10.1038/s41467-018-04362-x

dbSNP. (n.d.). Retrieved June15, 2023, from https://www.ncbi.nlm.nih.gov/snp/

Doust, C., Fontanillas, P., Eising, E., Gordon, S. D., Wang, Z., Alagöz, G., Molz, B., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Coker, D., Partida, G. C., Dhamija, D., Das, S., Elson, S. L., …Luciano, M. (2022). Discovery of 42 genome-wide significant loci associated with dyslexia. *Nature Genetics*, 1–9. https://doi.org/10.1038/s41588-022-01192-y

Duncan, B. W., Mohan, V., Wade, S. D., Truong, Y., Kampov-Polevoi, A., Temple, B. R., &Maness, P. F. (2021). Semaphorin3F Drives Dendritic Spine Pruning through Rho-GTPase Signaling. *Molecular Neurobiology*, *58*(8), 3817. https://doi.org/10.1007/S12035-021-02373-2

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T. I., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., &Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database: The*

Journal of Biological Databases and Curation, *2017*, 1–17. https://doi.org/10.1093/DATABASE/BAX028

Friedman, N. P., Banich, M. T., &Keller, M. C. (2021). Twin studies to GWAS: there and back again. *Trends in Cognitive Sciences*, *25*(10), 855–869. https://doi.org/10.1016/J.TICS.2021.06.007

Hatzikotoulas, K., Gilly, A., &Zeggini, E. (2014). Using population isolates in genetic association studies. *Briefings in Functional Genomics*, *13*(5), 371. https://doi.org/10.1093/BFGP/ELU022

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., Gale, C. R., Davies, G., &Deary, I. J. (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, *24*(2), 169–181. https://doi.org/10.1038/s41380-017-0001-5

Ho, C. S. H., Chan, D. W. O., Lee, S. H., Tsang, S. M., &Luan, V. H. (2004). Cognitive profiling and preliminary subtyping in Chinese developmental dyslexia. *Cognition*, *91*(1), 43–75. https://doi.org/10.1016/S0010-0277(03)00163-X

Ho, S., Chan, D., Chung, K., Tsang, S., Lee, S., &Cheng, W. Y. R. (2007). *Hong Kong Test of Specific Learning Difficulties in Reading and Writing for Primary School Students-Second Edition*.

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., van derZee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., …Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, 1–12. https://doi.org/10.1038/s41588-022-01062-7

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., &Farh, K. K. H. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, *176*(3), 535-548.e24. https://doi.org/10.1016/J.CELL.2018.12.015

Kim, Y.-S. G. (2020). *Interactive Dynamic Literacy Model: An Integrative Theoretical Framework for Reading-Writing Relations*. 11–34. https://doi.org/10.1007/978-3-030-38811-9_2

Lind, P. A., Luciano, M., Wright, M. J., Montgomery, G. W., Martin, N. G., &Bates, T. C. (2010). Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *European Journal of Human Genetics*, *18*(6), 668–673. https://doi.org/10.1038/ejhg.2009.237

Lo, J. C. M., Ye, Y., Tong, X., McBride, C., Ho, C. S. H., &Waye, M. M. Y. (2018). Delayed copying is uniquely related to dictation in bilingual Cantonese–English-speaking children in Hong Kong. *Writing Systems Research*, *10*(1), 26–42. https://doi.org/10.1080/17586801.2018.1481902

Luciano, M., Evans, D. M., Hansell, N. K., Medland, S. E., Montgomery, G. W., Martin, N. G., Wright, M. J., &Bates, T. C. (2013). A genome-wide association study for reading and language abilities in two population cohorts. *Genes, Brain and Behavior*, *12*(6), 645–652. https://doi.org/10.1111/gbb.12053

Maughan, B., Messer, J., Collishaw, S., Pickles, A., Snowling, M., Yule, W., &Rutter, M. (2009). Persistence of literacy problems: spelling in adolescence and at mid-life. *Journal of Child Psychology and Psychiatry*, *50*(8), 893–901. https://doi.org/10.1111/J.1469-7610.2009.02079.X

McBride-Chang, C., Chung, K. K. H., &Tong, X. (2011). Copying skills in relation to word reading and writing in Chinese children with and without dyslexia. *Journal of Experimental Child Psychology*, *110*(3), 422–433. https://doi.org/10.1016/J.JECP.2011.04.014

Müller, B., Wilcke, A., Czepezauer, I., Ahnert, P., Boltze, J., Kirsten, H., Friederici, A. D., Emmrich, F., Brauer, J., Neef, N., Skeide, M., Schaadt, G., Kraft, I., &Dörr, L. (2016). Association, characterisation and meta-analysis of SNPs linked to general reading ability in a German dyslexia case-control cohort. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP27901

Neale, M, Eaves, L., Bartels, M., Boomsma, D., Posthuma, D., &Bates, T. (2017). The 2018 International Workshop on Statistical Genetic Methods for Human Complex Traits. *Behavior Genetics*, *47*, 729–730.

Neale, M., &Cardon, L. (1992). *Methodology for Genetic Studies of Twins and Families*. Springer Science & Business Media.

Paracchini, S., Steer, C. D., Buckingham, L.-L., Morris, A. P., Ring, S., Scerri, T., Stein, J., Pembrey, M. E., Ragoussis, J., Golding, J., &Monaco, A. P. (2008). Association of the KIAA0319 Dyslexia Susceptibility Gene With Reading Skills in the General Population. *American Journal of Psychiatry*, *165*(12), 1576–1584. https://doi.org/10.1176/APPI.AJP.2008.07121872

Perfetti, C. (2010). *Decoding, Vocabulary, and Comprehension. The Golden Triangle of Reading Skill. In Bringing reading research to life* (M. G.McKeown &L.Kucan (Eds.)). New York: Guilford Press.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., DeBakker, P. I. W., Daly, M. J., &Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, *81*(3), 559. https://doi.org/10.1086/519795

Sahay, A., Molliver, M. E., Ginty, D. D., &Kolodkin, A. L. (2003). Semaphorin 3F Is Critical for Development of Limbic System Circuitry and Is Required in Neurons for Selective CNS Axon Guidance Events. *The Journal of Neuroscience*, *23*(17), 6671. https://doi.org/10.1523/JNEUROSCI.23-17-06671.2003

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., DeLeeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., …Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., Bishop, D. V. M., &Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP Are Associated with Reading-Related Traits. *Biological Psychiatry*, *70*(3), 237–245. https://doi.org/10.1016/J.BIOPSYCH.2011.02.005

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., &Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics 2019 20:8*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1

Tong, X., McBride, C., Ho, C. S. H., Waye, M. M. Y., Chung, K. K. H., Wong, S. W. L., &Chow, B. W. Y. (2018). Within- and cross-language contributions of morphological awareness to word reading and vocabulary in Chinese–English bilingual learners. *Reading and Writing*, *31*(8), 1765–1786. https://doi.org/10.1007/S11145-017-9771-Z/TABLES/5

Tong, X., &McBride-Chang, C. (2010). Developmental models of learning to read Chinese

words. *Developmental Psychology*, *46*(6), 1662–1676. https://doi.org/10.1037/A0020611

Tong, X., Mcbride-Chang, C., Shu, H., &Wong, A. M. Y. (2009). Morphological Awareness, Orthographic Knowledge, and Spelling Errors: Keys to Understanding Early Chinese Literacy Acquisition. *Scientific Studies of Reading*, *13*(5), 426–452. https://doi.org/10.1080/10888430903162910

Tosto, M. G., Hayiou-Thomas, M. E., Harlaar, N., Prom-Wormley, E., Dale, P. S., &Plomin, R. (2017). The Genetic Architecture of Oral Language, Reading Fluency, and Reading Comprehension: A Twin Study From 7 to 16 Years. *Developmental Psychology*, *53*(6), 1115. https://doi.org/10.1037/DEV0000297

Venkatesh, S. K., Siddaiah, A., Padakannaya, P., &Ramachandra, N. B. (2013). Analysis of genetic variants of dyslexia candidate genes KIAA0319 and DCDC2 in Indian population. *Journal of Human Genetics*, *58*(8), 531–538. https://doi.org/10.1038/JHG.2013.46

Wang, Q., Chiu, S. L., Koropouli, E., Hong, I., Mitchell, S., Easwaran, T. P., Hamilton, N. R., Gustina, A. S., Zhu, Q., Ginty, D. D., Huganir, R. L., &Kolodkin, A. L. (2017). Neuropilin-2/PlexinA3 Receptors Associate with GluA1 and Mediate Sema3F-Dependent Homeostatic Scaling in Cortical Neurons. *Neuron*, *96*(5), 1084-1098.e7. https://doi.org/10.1016/J.NEURON.2017.10.029

Wang, Z., Zhao, S., Zhang, L., Yang, Q., Cheng, C., Ding, N., Zhu, Z., Shu, H., Liu, C., &Zhao, J. (2023). A genome-wide association study identifies a new variant associated with word reading fluency in Chinese children. *Genes, Brain and Behavior*, *22*(1), e12833. https://doi.org/10.1111/GBB.12833

Waye, M. M. Y., Poo, L. K., &Ho, C. S.-H. (2017). Study of Genetic Association With DCDC2 and Developmental Dyslexia in Hong Kong Chinese Children. *Clinical Practice and Epidemiology in Mental Health : CP & EMH*, *13*(1), 104. https://doi.org/10.2174/1745017901713010104

Wilcke, A., Weissfuss, J., Kirsten, H., Wolfram, G., Boltze, J., &Ahnert, P. (2009). The role of gene DCDC2 in German dyslexics. *Annals of Dyslexia*, *59*(1), 1–11. https://doi.org/10.1007/S11881-008-0020-7/TABLES/3

Wong, P. C. M., Kang, X., So, H. C., &Choy, K. W. (2022). Contributions of common genetic variants to specific languages and to when a language is learned. *Scientific Reports 2022 12:1*, *12*(1), 1–13. https://doi.org/10.1038/s41598-021-04163-1

Wong, S. W. L., Chow, B. W. Y., Ho, C. S. H., Waye, M. M. Y., &Bishop, D. V. M. (2014). Genetic and Environmental Overlap Between Chinese and English Reading-Related Skills in Chinese Children. *Developmental Psychology*, *50*(11), 2539. https://doi.org/10.1037/A0037836

Zhou, Z., Tan, C., Chau, M. H. K., Jiang, X., Ke, Z., Chen, X., Cao, Y., Kwok, Y. K., Bellgard, M., Leung, T. Y., Choy, K. W., &Dong, Z. (2023). TEDD: a database of temporal gene expression patterns during multiple developmental periods in human and model organisms. *Nucleic Acids Research*, *51*(D1), D1168–D1178. https://doi.org/10.1093/NAR/GKAC978

Zou, L., Zhu, K., Jiang, Q., Xiao, P., Wu, X., Zhu, B., &Song, R. (2022). Quality of life in Chinese children with developmental dyslexia: a cross-sectional study. *BMJ Open*, *12*(1), e052278. https://doi.org/10.1136/BMJOPEN-2021-052278

Declarations

Acknowledgements

We thank the participants for their participation.

Funding

This study was partially funded by the Research Grants Council of the Hong Kong Special Administration Region (C4054-17WF) and the Theme-based Research Scheme from the Hong Kong Special Administrative Region Research Grants Council (T44-410/21-N).

Competing Interests

The authors declare no competing interests.

Author Contributions

Cheuk Yan Chung and Kwong Wai Choy contributed and composed the study design. Cheuk Yan Chung, Dora Jue Pan, Mo Zheng drafted the manuscript, collected the genotype data, and interpreted the data. Kwong Wai Choy, Silvia Paracchini, Hon-Cheong So, Urs Maurer, Wenxuan Jiang, and Catherine McBride interpreted the data. The authors approved the final version of manuscript.

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Approval

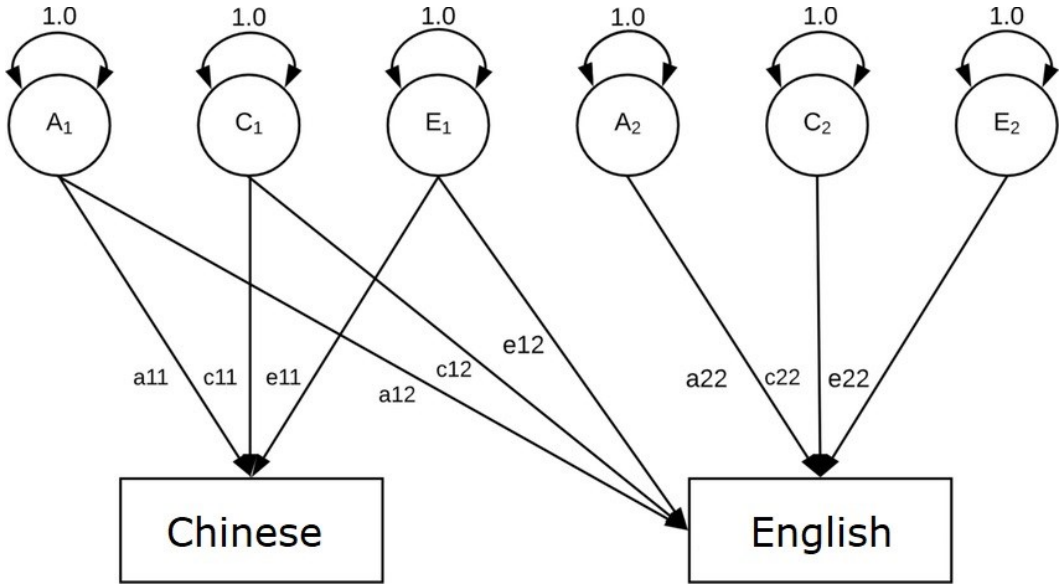
All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Chinese University of Hong Kong (CREC Ref. No.: 2021.237).

Consent to participate

Informed consent was obtained from all individual participants and their parents included in the study.

Figures

Figure 1 Bivariate Cholesky decomposition of variance and covariance of Chinese and English phenotype



A : additive genetic; C: shared environmental; E: non-shared environmental influence

Tables

Table 1 Descriptive statistics of phenotypes

Phenotypes	Number of participants	Mean (Standard deviation)	Range	Skewness	*p*-value of gender differences
Chinese word reading (Ho et al., 2007)	610	0.01 (0.91)	-4.22 to 1.93	-0.771	.101
Chinese vocabulary knowledge (Tong et al., 2018)	610	-0.04 (0.87)	-3.87 to 2.37	-0.295	.238
Chinese spelling (Lo et al., 2018)	616	0.04 (0.80)	-2.74 to 1.64	-0.742	**.005**
English word reading (Tong &McBride-Chang, 2010)	610	-0.03 (0.93)	-2.66 to 2.10	0.046	.511
English vocabulary knowledge (Tong et al., 2018)	610	-0.04 (0.96)	-2.31 to 2.13	0.101	.940
English spelling (Tong et al., 2009)	616	-0.04 (0.92)	-2.51 to 2.22	0.150	.762

Note: Significant results ($p < .05$) are in bold.

Table 2 Correlations between Chinese and English word reading, vocabulary knowledge, and spelling

	Chinese word reading	Chinese vocabulary knowledge	Chinese spelling	English word reading	English vocabulary knowledge	English spelling
Chinese word reading	1.000	**0.595**	**0.678**	**0.261**	0.025	**0.262**
Chinese vocabulary knowledge		1.000	**0.375**	**0.245**	**0.250**	**0.281**
Chinese spelling			1.000	**0.300**	**0.117**	**0.353**
English word reading				1.000	**0.848**	**0.942**
English vocabulary knowledge					1.000	**0.825**
English spelling						1.000

Note: Significant correlations ($p < .01$) are in bold.

Table 3 Standardized parameter estimates and calculated heritability for the three phenotypes in Chinese and English

Parameter estimate and heritability	Word Reading	Vocabulary Knowledge	Spelling
Bivariate Cholesky decomposition of variance and covariance of Chinese and English phenotypes			
a11	.7236	.6519	.4401
a12	.4487	.3203	.3581
a22	.2827	.0316	.3790
c11	.6200	.5385	.7353
c12	-.1763	-.0009	.1903
c22	.7855	.9069	.7780
e11	.3033	.5339	.5154
e12	.1447	.0812	.0966
e22	.2230	.2596	.2780
Heritability			
Chinese heritability	.5236	.4250	.1937
English heritability	.2812	.1036	.2718
Genetic correlation			
Genetic correlation between Chinese and English	.8461	.9952	.6868

Note: a: Additive genetic variable; c: Shared environmental variable; e: Non-shared environmental variable

Table 4 Bivariate ACE model fitting results for the three phenotypes in Chines and English and comparisons of nested models

Phenotype	Full and restricted models	*-2LL*	*Df*	*ΔLL*	*Δdf*	*p*
Word reading	Full ACE Model	2297.85	1209			
	Restricted Model 1 (a12=0)	2349.96	1210	52.11	1	**<.001**
	Restricted Model 2 (a22=0)	2309.55	1210	11.70	1	**<.001**
Vocabulary knowledge	Full ACE Model	2487.04	1209			
	Restricted Model 1 (a12=0)	2505.41	1210	16.78	1	**<.001**
	Restricted Model 2 (a22=0)	2487.04	1210	0.00	1	.976
Spelling	Full ACE Model	2425.82	1221			
	Restricted Model 1 (a12=0)	2434.96	1222	9.14	1	.002
	Restricted Model 2 (a22=0)	2427.31	1222	1.49	1	.222

Note: Significant results ($p < .001$) are in bold.

Table 5 Association analysis of rs1005678, rs12632110 and rs12494414 with Chinese and English vocabulary knowledge

			Chinese vocabulary knowledge				English vocabulary knowledge			
SNP	Level	Risk allele	Count	Coefficient [95% CI]		*p* value	Count	Coefficient [95% CI]		*p* value
rs1005678	Population	G	305	-0.145	[-0.291 – 0.002]	0.052	305	-0.131	[-0.304 – 0.041]	0.134
	Within-twin	G	342	-0.178	[-0.335 - -0.022]	**0.025**	342	-0.242	[-0.412 - -0.073]	**0.005***
rs12632110	Population	A	305	-0.148	[-0.290 - -0.007]	**0.040**	305	-0.105	[-0.272 – 0.062]	0.216
	Within-twin	A	342	-0.188	[-0.345 - -0.030]	**0.020**	342	-0.191	[-0.362 - -0.019]	**0.029**
rs12494414	Population	T	305	-0.095	[-0.239 – 0.049]	0.196	305	-0.135	[-0.304 – 0.034]	0.117
	Within-twin	T	342	-0.114	[-0.270 – 0.042]	0.153	342	-0.250	[-0.419 - -0.082]	**0.004***

Note: Significant results ($p < .05$) are in bold. Significant results after multiple testing correction ($p < .0125$) are with an asterisk (*).