DOI:
10.1515/2157-5665.1053

Entrepreneurship Research Journal

Manuscript 1053

A Study of Gene-Environment Interactions In Entrepreneurship

Lydia Quaye, *King's College London*
Nicos Nicolaou, *University of Cyprus & Cass Business School, City University*
Scott Shane, *Case Western Reserve University*
Juliette Harris, *King's College London*

DOI: 10.1515/2157-5665.1053

A Study of Gene-Environment Interactions In Entrepreneurship

Lydia Quaye, Nicos Nicolaou, Scott Shane, and Juliette Harris

Abstract

We examined the interactions between four genes associated with dyslexia (*ROBO1, KIAA0319, DCDC2, DYX1C1*) and education on the tendency to become an entrepreneur. We used a two-staged design consisting of a discovery sample of 692 individuals, and a replication sample of 797 participants from the TwinsUK cohort. Associations were identified between entrepreneurship and interactions of education and *ROBO1* rs654867 and *KIAA0319* rs6902039 with the stage 1 samples. However these were not independently replicated and the associations were no longer significant when the samples from the 2 stages were combined. A tagging SNP approach was used to investigate the effect of the interactions between education and 191 tagging SNPs from the candidate genes on entrepreneurship. While we found several significant interactions (*DCDC2, KIAA0319* and *ROBO1*), none passed the stringent threshold for significance of a Bonferroni correction. Similar to the case with other behavioural genetics phenotypes, large sample sizes will be required to identify significant gene-environment interactions in entrepreneurship after making Bonferroni corrections.

KEYWORDS: entrepreneurship, genetics, gene-environment interactions

Author Notes: We would like to thank the twins that participated in our study. We are most grateful for funding from Philip Zepter grant no. ucy 31028.

INTRODUCTION

Dyslexia is an inherited disability that makes spelling, reading and writing a challenge (DeFries et al., 1987; Stevenson et al., 1987; Hohnen and Stevenson, 1999; Scerri and Schulte-Korne, 2010). This disorder influences occupational choice because dyslexics disproportionately select jobs that demand less reading (Taylor and Walter, 2003).

One occupation with relatively low demands for reading is business ownership, which focuses more on interpersonal communication (Taylor and Walter, 2003). Entrepreneurs are more than twice as likely as non-entrepreneurs to be dyslexic (Pinker, 2008), with one study reporting that 35% of U.S. entrepreneurs are dyslexic, as compared to only 15% of the overall U.S. population (Logan, 2009). Notable examples of dyslexics who have become entrepreneurs include Richard Branson (Virgin Group), Charles Schwab (Charles Schwab and Company), Ted Turner (Turner Broadcasting Systems), John Chambers (CISCO Systems), Henry Ford (Ford Motor Company), Paul Orfalea (Kinko's) and William Hewlett (Hewlett Packard) (Smith, 2008).

Dyslexia has a genetic component. Familial clustering for dyslexia has been reported since the early 1900s (Thomas, 1905; Stephenson, 1907) while twin studies using monozygotic and same-sex dizygotic twins found higher concordances for dyslexia in monozygotic than in dizygotic twins (Bakwin, 1973; DeFries and Alarcon, 1996). The heritability of dyslexia has been estimated between 40 and 70% (Olson, 2002; Scerri and Schulte-Korne, 2010).

Molecular genetics studies of dyslexia have also been conducted. Hannula-Jouppi et al. (2005) found that the axon guidance receptor gene *ROBO1,* which influences the production of proteins involved in axon[1] growth and neuronal migration[2], is a candidate for dyslexia. Specifically, Hannula-Jouppi et al. (2005) found that a haplotype[3] spanning *ROBO1* was absent or attenuated in dyslexic individuals. Several studies have also implicated the *KIAA0319, DCDC2* and *DYX1C1* genes in dyslexia susceptibility (Cope et al., 2005; Dennis et al, 2009; Galaburda et al., 2006; Harold et al., 2006; Luciano et al., 2007; Meng et al., 2005; Paracchini et al., 2008; Scerri and Schulte-Korne, 2010; Taipale et al., 2003).

Because entrepreneurship is also highly heritable (Nicolaou et al, 2008), one theory is that the observed preference of dyslexics for entrepreneurship might result through gene-environment correlations. There are a number of ways

[1] The axon is part of a neuron (a neuron is a type of cell that makes up the nervous system) that "carries information to other neurons and transmits an action potential (Ward, 2010, p.18).
[2] Neuronal migration is "the movement of new neurons to their final and appropriate location in their developing brain" (Gazzaniga et al., 2009, p. G-8).
[3] A haplotype is a group of alleles that are inherited together.

through which this may take place. First, people who have more trouble processing written words are more likely to choose human-interaction-oriented activities and less likely to choose reading or mathematics intensive ones. These choices cumulate over time, leading people with dyslexia to be more likely to select training for occupations that demand less reading.

Second, dyslexics are better in visuo-spatial processing (Von Karolyi et al., 2003) and have greater spatial ability (Hornsby, 1984) than non-dyslexics. These characteristics may make them more likely than others to recognise entrepreneurial opportunities and become entrepreneurs because “this increased level of spatial advantage can manifest itself as a skill or gift at recognizing patterns” (Smith, 2008, p. 306). Coppola (2007) quotes John Chambers, the founder of CISCO Systems who said that “dyslexia helps him step back and see the big picture". Specifically, Chambers argues that “dyslexia forces you to look at things in totality and not just as a single chess move” (Coppola, 2007).

Third, dyslexics are more creative in tasks needing novelty and insight and are more likely to view and solve a problem from an unusual perspective (Everatt et al., 1999). Because creativity and novelty are very important in entrepreneurship (Shane, 2003) dyslexics may be more likely than others to become entrepreneurs.

However, the effect of a genetic propensity to dyslexia on occupational choice may interact with the environmental condition of education. For people with low levels of education, the effect of a genetic predisposition to dyslexia on the tendency to be an entrepreneur should be relatively small because few low education occupations require much reading. However, at high levels of education, where different occupations lead to jobs with far ranging requirements for reading, those individuals with a genetic predisposition to dyslexia are more likely to select occupations that demand less reading than those without the genetic predisposition.

This theory was tested in a 2-stage design with ~1,500 Caucasians from the TwinsUK population cohort. This study reports the findings from the evaluation of the effects of common tagging single nucleotide polymorphisms (SNPs) from genes which have been associated with dyslexia (*ROBO1*, *KIAA0319*, *DCDC2* and *DYX1C1*) on entrepreneurship and the effect of an interaction between the common variants and education on entrepreneurship.[4]

[4] Six of these SNPs within or flanking the *ROBO1* and *KIAA0319* genes have previously been found to be associated with entrepreneurship (unpublished results – refer to Table 1).

METHODS

The study involved a two-stage design using data from the TwinsUK Adult twin registry (www.twins.ac.uk; Spector and Williams, 2006). The sample is made up of volunteers not enriched for any particular disease or trait (Spector and Williams, 2006). The first (discovery) stage of the study consisted of 692 Caucasian individuals and the second (validation) stage was made up of 797 participants, also of Caucasian ethnicity. Volunteers provided informed consent and were administered a protocol approved by the St. Thomas Hospital ethics committee. The volunteers completed a written survey and provided a blood sample from which DNA was extracted.

The mean age of the discovery cohort (stage 1) was 58 years (standard deviation [SD] =11 years) and 91% were female. The mean age of the validation cohort (stage 2) was also 58 with the same standard deviation. Approximately 86% of the validation cohort were female. The large prevalence of females in the TwinsUK adult twin registry is because the twins were initially recruited to study osteoporosis, a disease that occurs mostly in females. Descriptive statistics of the data are presented in Table 2.

Genotyping and tagging SNP selection

The TwinsUK samples were genotyped with the Infinium 610k assay (Illumina, San Diego, USA). The positions of the gene coding regions for the candidate genes were identified.

Six SNPs within or flanking genes associated with dyslexia (*ROBO1* and *KIAA0319*) were selected for analysis. These SNPs (rs6548678; rs6902039; rs331199; rs9827022; rs10511114 and rs13095041) were chosen because of their nominal association with entrepreneurship from a previous, unpublished study. The *ROBO1* rs1509278 SNP was omitted from analysis because it is in complete linkage equilibrium with rs6548678 (r^2=1). Linkage disequilibrium (LD) is the non-random association of alleles at two or more loci. Neighbouring SNPs tend to be in LD – the SNPs are correlated with each other. SNPs which are in strong LD are inherited together, but their polymorphisms may have different MAF. Sometimes, the term “tag” is used to describe the correlation between SNPs with the same or similar minor allele frequencies (MAF). The correlation between neighbouring SNPs makes it unnecessary to genotype all the SNPs within a gene in order to test for association with a trait. The correlation between two SNPs is measured by r^2. An r^2 threshold of 0.8 is usually chosen to in genetic association studies, which means there is at least 80% correlation between the tagging SNP and all the SNPs it tags.

Table 1. Nominally significant SNPs associated with entrepreneurship

Chromosome	Gene	Marker	P-value
3	*ROBO1**	rs6548678	0.003
3	*ROBO1**	rs1509278	0.0038
6	*KIAA0319*	rs6902039	0.0091
3	*ROBO1*	rs331199	0.02
3	*ROBO1**	rs9827022	0.025
3	*ROBO1**	rs10511114	0.046
3	*ROBO1**	rs13095041	0.047

* flanking region

Table 2. Descriptive Statistics

(a) Continuous variables

Variable	Mean	Standard Deviation	Min	Max	Descriptives
Age (years)	58.1	10.88	26	83	80% ≥ 50 years (N=1234)
Education (years)	12.5	3.05	7	30	N=1489

(b) Dichotomous variables

Variable	Descriptions	
	Stage 1	Stage 2
Entrepreneurship	146 entrepreneurs 546 non-entrepreneurs	169 entrepreneurs 628 non-entrepreneurs
Gender	58 males 634 females	111 males 686 females
Zygosity	231 monozygotic twins 461 dizygotic twins	434 monozygotic twins 361 dizygotic twins

(c) Genotype distribution of rs6548678

Genotype	Entrepreneurs		Non-entrepreneurs		Total
	Stage 1	Stage 2	Stage 1	Stage 2	
TT	50 (34%)	45 (22%)	151 (28%)	171 (25%)	417 (28%)
TC	66 (45%)	81 (52%)	253 (46%)	307 (49%)	707 (47%)
CC	30 (21%)	43 (26%)	142 (26%)	150 (26%)	365 (25%)
Total	146 (100%)	169 (100%)	546 (100%)	628 (100%)	1489 (100%)

In addition to *ROBO1* and *KIAA0319, DCDC2* and *DYX1C1*, which have also been associated with dyslexia, were selected for a candidate gene association study in order to ascertain whether the interaction between education and polymorphisms within and flanking these genes correlated with entrepreneurship. The tagging SNP approach was used to select SNPs from the candidates. In order to select tagging SNPs, genotyping data from the CEPH reference panel were downloaded for the four candidate genes from The International HapMap Project (www.hapmap.org), and Haploview 4.2 (Barrett et al, 2005) were used to tag the SNPs in order to prevent the unnecessary analysis of highly correlated SNPs ($r^2 \geq 0.8$). The PLINK whole genome data analysis toolset software (http://pngu.mgh.harvard.edu/purcell/plink/; Purcell et al., 2007) was used to extract genotyping data within and flanking *ROBO1*, *DCDC2*, *KIAA0319* and *DYX1C1* of TwinsUK participants. Tagging SNPs from the candidate genes were selected if there was imputed or genotyping data available from the TwinsUK cohort. These included 130 tagging SNPs in *ROBO1*, 44 tagging SNPs in *KIAA0319*, 7 tagging SNPs in *DCDC2* and 10 tagging SNPs in *DYX1C1*, amounting to a total of 191 tagging SNPs.

Quality control

Only participants with entrepreneurship and education data were included in this research. SNPs with genotype call rates <90% were excluded from analysis. Eleven *ROBO1* tagging SNPs (rs1447834, rs6790493, rs7642059, rs10511124, rs12633213, rs11713043, rs9682532, rs11714102, rs1350366, rs11925452 and rs1913250) were excluded from analysis due to genotype call rates below 90%. However, 167 (87%) of the tagging SNPs had call rates >95%. Deviations from Hardy-Weinberg equilibrium were assessed with the χ^2 test. The Hardy-Weinberg measure evaluation was used to estimate whether the population is in equilibrium. This was used as a method of quality control, to ensure that there were no errors in the genotyping data, and the observed genotype frequencies do not defer significantly from the expected.

Measures

Entrepreneurship was measured by a dichotomous variable indicating a positive response to a question commonly used for this purpose in the social science literature (Mesch and Czamanzki, 1997; Delmar and Davidsson, 2000): “In your working life, have you ever started a new business?”

We used an additive genetic model (for all SNPs) whereby common homozygotes were coded as 0, heterozygotes were coded as 1, and rare homozygotes were coded as 2. Education was measured as the number of years of full-time education from primary school onwards. The effects of age (in years), gender and zygosity on entrepreneurship were assessed with logistic regression. Age and zygosity were not associated with entrepreneurship; however gender was significantly associated, with a greater proportion of males being entrepreneurs, compared with females (P=0.003).

Statistical Analysis

To examine the effect of the interactions between each gene and education on the tendency to be an entrepreneur, we ran hierarchical logistic regressions using STATA version 11 (Cohen and Cohen, 1983; Jaccard, 2001). We employed clustered effects to deal with the non-independence of the twin sample (Hosmer and Lemeshow, 2001). We used the Wald χ^2 test to assess the overall significance of each model.

RESULTS

Previous unpublished analysis of four candidate genes associated with dyslexia (*ROBO1*, *DCDC2*, *KIAA0319*, and *DYX1C1)* showed 7 polymorphisms to be nominally significant for entrepreneurship (see Table 1)[5]. As rs6548678 and rs1509278, which flank *ROBO1*, were in linkage disequilibrium, we conducted the analyses for the first (discovery) and second (validation) stages of the study using six of these SNPs (the results for rs6548678 and rs1509278 would be identical). The 6 single nucleotide polymorphisms (rs6548678, rs6902039, rs331199, rs9827022, rs10511114 and rs13095041) were assessed for association with entrepreneurship and interaction with education.

Logistic regression modelling showed that age and zygosity were not significantly associated with entrepreneurship (P>0.05), therefore, these variables were excluded from further analyses. Males were approximately twice as likely as females to be entrepreneurs, thus this covariate was adjusted for in all further analyses. Table 3 shows the results of the hierarchical logistic regression analyses to examine the effect of the interaction between the subject's level of education and each of the SNPs[6].

When the effects of an interaction between the rs6548678 polymorphism and education were examined, the interaction was significant (odds ratio [OR]=0.88 with corresponding 95% confidence intervals [0.81-0.96], P=0.003) as

[5] Six out of the seven SNPs were in the flanking regions of the genes.
[6] We report the odds ratios rather than the estimated coefficients as this is the standard practice in behavioral genetics research. We note, however, that the odds ratio interpretation of interaction terms is not as straight forward (Ai and Norton, 2003; Norton, Wang and Ai, 2004).

was the Wald χ^2 statistic for the model (χ^2=21.79, P=0.001)[7]. The results suggest that the environmental stimulus of high education increases the odds that a carrier of two copies of allele T will be an entrepreneur. The interaction of this genetic predisposition and education increases the amount of variance in the likelihood of being an entrepreneur by 1%.

We further examined the significant interaction effect for *ROBO1* rs6548678. As illustrated in Figure 1, which shows the genotype distribution between entrepreneurs and non-entrepreneurs across the TT, TC and CC genotypes for three different levels of education (less than 10 years; between 10 and 14 years and greater than or equal to 15 years). The 2-bar histogram shows that individuals who had two copies of the T allele of rs6548678 and were highly educated were more likely to be entrepreneurs than people who were highly educated and had TC or CC genotypes.

We also plotted the predicted log odds for entrepreneurship for the combinations of education listed below and *ROBO1* rs6548678 values for a series of simple regression equations (Jaccard et al., 1990), using three values for the rs6548678 polymorphism: TT, TC and CC genotypes. Inputting the values corresponding to three different levels of education (1 standard deviation below the mean [9.8 years], the mean value [12.5 years] (approximately equal to A-Levels) and 1 SD above the mean [15.2 years]), the graph shows that people who were homozygous for the T allele were more likely to become entrepreneurs if they have high levels of education than if they have low levels of education.

The results of the interaction between the *KIAA0319* rs6902039 polymorphism and education are also shown in Table 3. The interaction (OR=1.13 [1.03-1.24], P=0.013) and the model (Wald χ^2 =20.21, P=0.0005) were also statistically significant at the 5% level. This indicates that education increases the odds that a carrier of two copies of allele C will be an entrepreneur. The interaction of this genetic predisposition and education increases the amount of variance in the likelihood of being an entrepreneur by 1%. A statistically significant association was also identified between the interaction between education and *ROBO1* rs9827022 and entrepreneurship with the stage 1 samples.

No statistically significant association was found with an interactive term between the *ROBO1* rs331199, rs10511114 and rs13095041 polymorphisms and education with the stage 1 samples (P>0.05).

In the validation stage of the study we attempted to replicate the significant findings above using an additional sample of 797 individuals from the TwinsUK cohort. The results from the validation stage are shown in Table 4. Two

[7] There is a significant difference between models 2 and 5 in the effect of rs6548678 on the likelihood of engaging in entrepreneurship. However, as model 5 includes the interaction term, the coefficient of rs6548678 does not represent a "main effect", but a conditional effect based on the level of education (Aiken and West, 1991).

statistically significant associations were found with an interactive term between the *ROBO1* rs10511114 and rs331199 polymorphisms and education in the stage 2 sample ($P<0.05$). However, none of the interaction effects that were significant in the stage 1 sample were significant with the stage 2 samples.

The effects of the interactions between other tagging SNPs from *DCDC2*, *DYX1C1*, *KIAA0319* and *ROBO1* and education on entrepreneurship were also evaluated.

The results from the SNPs which passed QC with P-values <0.1 are presented in Table 5. There was no evidence of association between the interaction of tagging SNPs from the *DYX1C1* and education on an individual's tendency to be an entrepreneur. Interactive associations were identified with tagging SNPs from or flanking *DCDC2*, *KIAA0319* and *ROBO1* ($P<0.05$, see Table 5). The 3 most significant interactions were all within *KIAA0319*. The strongest association with entrepreneurship was the interaction between the *KIAA0319* rs12201003 polymorphism and education OR=0.81 (0.71-0.92), P=0.001. The second most significant interaction was between the *KIAA0319* rs3181245 polymorphism and education OR=1.1 (1.03-1.17), P=0.004. The third most significant interaction was between the *KIAA0319* rs699461 polymorphism and education OR= 1.08 (1.02-1.14), P=0.011. There was little correlation between these *KIAA0319* tagging SNPs - the strongest correlation was between rs699461 and rs12201003 ($r^2=0.203$). The interaction between education and 6 *ROBO1* tagging SNPs (rs9864412, rs2608018, rs11713043, rs2043661, rs1447833, rs4443127) were also associated with entrepreneurship ($P<0.05$, see Table 5).

In order to adjust for multiple testing we used Bonferroni correction. The threshold for significance after this adjustment is 2.6×10^{-4}. None of the interaction terms passed this threshold required for significance. We note that the Bonferroni method is very stringent in this case because it assumes independent tests, while many of the SNPs analysed were correlated (e.g. rs2660739 and rs1995693 were correlated at $r^2=0.78$, as were rs13325521 and rs9845223).

Table 3. Logistic Regression Results of variables on entrepreneurship with Cluster Effects (Caucasians – stage 1 samples)

Variables	Model 1		Model 2		Model 3		Model 4		Model 5	
	Odds ratio (95% CI)	**P-value**	**Odds ratio (95% CI)**	**P-value**	**Odds ratio (95% CI)**	**P-value**	**Odds ratio (95% CI)**	**P-value**	**Odds ratio (95% CI)**	**P-value**
Gender	**0.43 (0.24-0.77)**	**0.004**	**0.43 (0.24-0.76)**	**0.004**	**0.46 (0.25-0.84)**	**0.010**	**0.46 (0.25-0.83)**	**0.01**	**0.44 (0.24-0.80)**	**0.008**
Education			1.04 (0.98-1.11)	0.210			1.04 (0.97-1.11)	0.267	**1.15 (1.05-1.26)**	**0.003**
rs6548678					0.80 (0.61-1.04)	0.093	0 .80 (0.62-1.04)	0.101	**4.10 (1.34-12.5)**	**0.013**
rs6548678xEdu									**0.88 (0.81-0.96)**	**0.003**
Wald χ^2 (DF)	**9.03 (1)**	**0.012**	**12.19 (2)**	**0.016**	**11.05 (2)**	**0.026**	**13.43 (3)**	**0.020**	**21.79 (4)**	**0.001**
rs6902039					**0.73 (0.56-0.96)**	**0.026**	**0.73 (0.55-0.96)**	**0.024**	**0.16 (0.04-0.54)**	**0.003**
rs6902039xEdu									**1.13 (1.03-1.24)**	**0.013**
Wald χ^2 (DF)	**9.03 (1)**	**0.012**	**12.19 (2)**	**0.016**	**12.46 (2)**	**0.0002**	**15.44 (3)**	**0.0015**	**20.21 (4)**	**0.0005**

N=692, DF: degrees of freedom, CI: confidence intervals.

Model 1 includes the control variables. Model 2 includes the control variables and education. Model 3 includes the control variables and genotype. Model 4 includes the control variables, genotype and education. Model 5 includes the control variables, genotype, education and the interaction term between genotype and education.

Table 4. Candidate SNPs from previous study – effect of SNPs and their interaction with education on entrepreneurship

Gene	SNP	Stage 1	Stage 2	Combined
		OR (95% CI), P-value	**OR (95% CI), P-value**	**OR (95% CI), P-value**
		SNP effects		
ROBO1	rs9827022	0.87 (0.67-1.15), P=0.333	1.04 (0.81-1.33), P=0.78	0.95 (0.79-1.15), P=0.607
ROBO1	rs10511114	0.92 (0.69-1.22), P=0.549	1.16 (0.88-1.52), P=0.285	1.03 (0.84-1.26), P=0.763
ROBO1	rs13095041	0.85 (0.69-1.12), P=0.242	1.06 (0.82-1.37), P=0.667	0.95 (0.79-1.15), P=0.601
ROBO1	rs6548678	0.8 (0.62-1.04), P=0.101	1.04 (0.82-1.31), P=0.749	0.92 (0.77-1.1), P=0.349
KIAA0319	rs6902039	**0.73 (0.55-0.96), P=0.024**	1.16 (0.91-1.48), P=0.22	0.94 (0.78-1.13), P=0.496
ROBO1	rs331199	0.89 (0.67-1.17), P=0.402	0.95 (0.74-1.21), P=0.66	0.92 (0.76-1.11), P=0.369
		SNP interactions with education		
ROBO1	rs9827022xEdu	**0.9 (0.82-0.98), P=0.021**	1.02 (0.97-1.06), P=0.448	0.99 (0.96-1.02), P=0.627
ROBO1	rs10511114xEdu	0.91 (0.83-1), P=0.054	**1.05 (1-1.1), P=0.047**	1.01 (0.97-1.04), P=0.687
ROBO1	rs13095041xEdu	0.94 (0.86-1.04), P=0.233	0.99 (0.94-1.03), P=0.567	0.98 (0.94-1.01), P=0.146
ROBO1	rs6548678xEdu	**0.88 (0.81-0.96), P=0.003**	1.02 (0.98-1.06), P=0.359	0.99 (0.96-1.02), P=0.375
KIAA0319	rs6902039xEdu	**1.13 (1.02-1.24), P=0.014**	1.02 (0.98-1.06), P=0.43	1 (0.97-1.03), P=0.987
ROBO1	rs331199xEdu	0.98 (0.89-1.07), P=0.61	**1.05 (1.01-1.09), P=0.025**	1.02 (0.99-1.06), P=0.148

OR: odds ratio. CI: confidence interval.

Table 5. Effect of interaction between education and tagging SNPs from dyslexia candidate genes (P<0.1) (combined stages 1 & 2)

Gene	SNP	MAF	N	Genotype Call rate	SNP	SNPxEdu interaction
					OR (95% CI), P-value	OR (95% CI), P-value
KIAA0319	**rs12201003**	0.22	1470	99%	**0.77 (0.64-0.91), P=0.003**	**0.81 (0.71-0.92), P=0.001**
KIAA0319	**rs3181245**	0.08	1373	92%	**1.1 (1.02-1.2), P=0.021**	**1.1 (1.03-1.17), P=0.004**
KIAA0319	**rs699461**	0.17	1474	99%	1.05 (0.97-1.13), P=0.209	**1.08 (1.02-1.14), P=0.011**
ROBO1	**rs9864412**	0.04	1477	99%	0.94 (0.88-1.01), P=0.077	**0.94 (0.89-0.99), P=0.023**
KIAA0319	**rs807509**	0.22	1433	96%	1.05 (0.97-1.13), P=0.206	**1.07 (1.01-1.14), P=0.025**
DCDC2	**rs1337736**	0.42	1400	94%	**1.15 (1.05-1.26), P=0.003**	**1.08 (1.01-1.16), P=0.026**
KIAA0319	**rs3756821**	0.17	1457	98%	**1.1 (1.02-1.19), P=0.019**	**1.07 (1.01-1.14), P=0.027**
ROBO1	**rs2608018**	0.48	1480	99%	**0.9 (0.82-0.98), P=0.015**	**0.93 (0.87-0.99), P=0.027**
ROBO1	**rs2043661**	0.42	1489	100%	0.93 (0.79-1.09), P=0.368	**0.92 (0.85-0.99), P=0.034**
ROBO1	**rs1447833**	0.20	1488	100%	**0.91 (0.84-1), P=0.048**	**0.91 (0.84-1), P=0.038**
ROBO1	**rs4443127**	0.01	1424	96%	0.97 (0.87-1.07), P=0.503	**1.21 (1.01-1.44), P=0.039**
KIAA0319	**rs2760157**	0.31	1462	98%	1.44 (0.85-2.44), P=0.18	**1.08 (1-1.16), P=0.043**
ROBO1	rs2872006	0.19	1481	99%	1.06 (0.97-1.15), P=0.221	0.95 (0.9-1), P=0.05
ROBO1	rs9874934	0.06	1481	99%	0.95 (0.89-1.01), P=0.118	0.95 (0.9-1), P=0.074
ROBO1	rs6809232	0.42	1489	100%	0.96 (0.89-1.02), P=0.204	0.93 (0.87-1.01), P=0.075
ROBO1	rs2120705	0.32	1488	100%	0.93 (0.85-1.01), P=0.087	1.05 (0.99-1.12), P=0.076
DCDC2	rs17631086	0.13	1488	100%	**1.09 (1.01-1.18), P=0.018**	0.9 (0.8-1.01), P=0.086
ROBO1	rs6807427	0.12	1401	94%	0.92 (0.82-1.04), P=0.204	0.95 (0.9-1.01), P=0.088
KIAA0319	rs7763790	0.40	1489	100%	0.94 (0.87-1.01), P=0.072	0.95 (0.89-1.01), P=0.093
KIAA0319	rs2760179	0.44	1459	98%	0.97 (0.9-1.05), P=0.406	0.95 (0.89-1.01), P=0.094
KIAA0319	rs16889523	0.40	1487	100%	**0.77 (0.64-0.91), P=0.003**	0.95 (0.89-1.01), P=0.094
ROBO1	rs7619949	0.34	1488	100%	**1.1 (1.02-1.2), P=0.021**	0.95 (0.9-1.01), P=0.098

MAF: minor allele frequency; N: total number of samples; CI: confidence interval

Figure 1. Two-bar histogram of genotype distribution and the level of education for entrepreneurs and non-entrepreneurs

	Low <10 years			10-14 years			High ≥15 years		
	TT	TC	CC	TT	TC	CC	TT	TC	CC
Entrepreneur	0	0	0	27.68	48.21	24.11	53.85	38.46	7.69
Non-entrepreneur	5	75	20	28.84	45.81	25.35	29.57	41.74	28.7

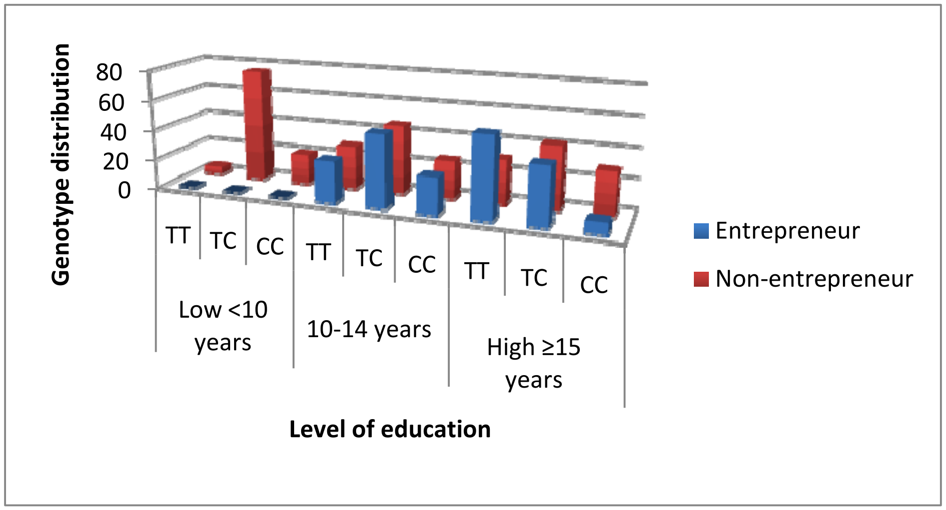
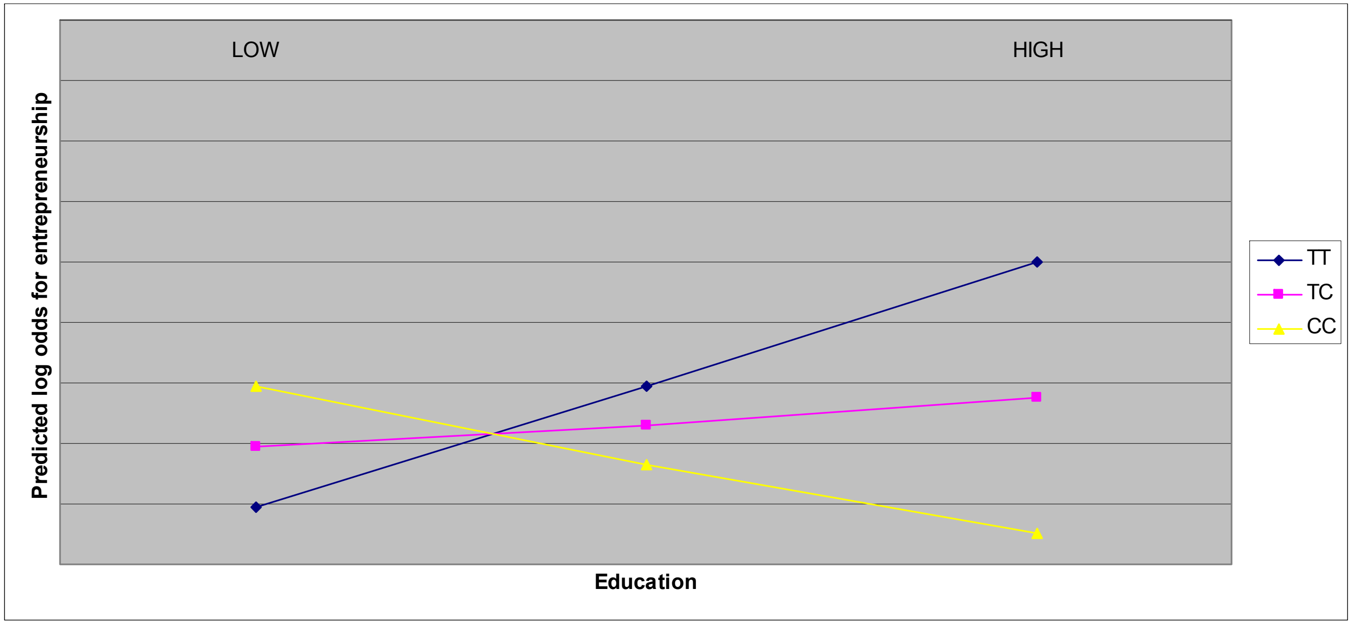


Figure 2. Interaction effects between rs6548678 and education



3 different levels of education (mean, one standard deviation above the mean, and one standard deviation below the mean) were used for each of the TT, TC and CC alleles.

DISCUSSION

This study used a 2 stage design to examine whether an interaction between a genetic polymorphism from 4 candidate dyslexia genes (*DCDC2*, *DYX1C1*, *KIAA0319* and *ROBO1*) and an aspect of the environment are associated with the tendency to be an entrepreneur. We found a significant interaction between education and rs6548678 on the tendency to be an entrepreneur. Highly educated individuals who carry two copies of allele T at the SNP rs6548678 were more likely to be entrepreneurs than individuals with the other alleles. However, when we tested for all tagging SNPs from the four genes with ~1,500 samples in the study, we did not find any tagging SNP which attained the required level of significance after correcting for multiple testing using a Bonferroni correction.

The rs6548678 locus flanks the *ROBO1* gene, which is involved in axon growth and neuronal migration, which, in turn, can influence sensorimotor, perceptual and cognitive processes indispensable for learning (Galaburda et al., 2006). *ROBO1* has a known function in the regulation of axon crossing across the midline between the hemispheres of the brain (Kidd et al., 1999; Seeger et al., 1993; Hannula-Jouppi et al. (2005).

The small amount of variance explained by the interaction of education and rs6548678 (1%) suggests that this gene-environment interaction alone does not predict the tendency to be an entrepreneur, but may be one of many mechanisms involved.

The lack of a significant interaction when the analysis is corrected for multiple testing underlines the difficulty in identifying gene-environment interactions for "complex" phenotypes like entrepreneurship (Young-Wolff et al., 2011; Field et al., 2011). Thousands of cases and controls are required to detect relative risks (Bookman et al., 2011). Future studies could focus on identifying gene-environment interactions for intermediate endophenotypes as a first stage in the process of identifying interactions for such phenotypes (Moffitt, 2005; Bookman et al., 2011). They could also aggregate polymorphisms into tagging SNP sets, which can reduce the multiple testing problem (Docherty et al., 2011). Moreover, they could increase the homogeneity in the operationalization of entrepreneurship and lower the measurement error - both of which would increase the power in detecting a gene-environment interaction.

We note that the use of a twin cohort is a potential limitation of the study. However, a comparison of lifestyle characteristics and health of the TwinUK cohort and a parallel singleton population found only a difference in birth weight between the groups, suggesting that findings using the twin cohort are generalizable to the general population (Andrew et al., 2001). Another limitation is the use of a self-reported single item measure of the tendency to be an entrepreneur. However, other research has shown the validity of this measure

(Delmar and Davidsson, 2000; Nicolaou et al., 2008). An additional limitation of the study is the relatively small sample size. Furthermore, education may be partly genetically influenced and therefore may not be an exogenous environmental variable (Behrman and Taubman, 1989). However, it is very difficult to obtain completely exogenous environmental measures when conducting gene-environment interaction studies as most "environmental" factors will have some genetic predisposition (Plomin et al., 2008).

Our study highlights some of the challenges involved in identifying gene-environment interactions in entrepreneurship. We have reported our null findings, following Bookman et al. (2011) who argued that study designs "should be informed by gene-environment hypotheses and negative results should be published to inform future hypotheses" (p.3).

Our lack of significant results after Bonferroni correction does not negate the importance of examining gene-environment interactions for entrepreneurship. These effects should be greater than the main effects of genetic and environmental factors (Moffitt, 2005).

REFERENCES

Ai, C., Norton E. C. (2003). "Interaction Term in Logit and Probit Models". *Economic Letters*, 80(1) 123-129.

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Andrew, T., Hart, D.J., Sneider, H., de Lange, M., Spector, T.D., MacGregor, A.J. (2001). "Are twins and singletons comparable? A study of disease-related and lifestyle and lifestyle characteristics in adult women". *Twin Research* 4: 464-77.

Bakwin, H. (1973). "Reading disability in twins". *Dev Med Child Neurol* 15(2):184–187.

Barrett JC, Fry B, Maller J, Daly MJ. (2005). "Haploview: analysis and visualization of LD and haplotype maps". *Bioinformatics.* 21(2): 263-5

Behrman, J. R., Taubman, P. (1989). "Is schooling mostly in the genes? Nature-nurture decomposition using data on relatives". *Journal of Political Economy*, 97(6): 1425-1446.

Bookman, E. B., McAllister, K., Gillanders, E., Wanke, K., Balshaw, D., et al. (2011). "Gene-environment interplay in common complex diseases: forging an integrative model-recommendations from an NIH workshop". *Genetic Epidemiology*. Feb 9. doi: 10.1002/gepi.20571. [Epub ahead of print]

Cohen, J., Cohen, P. (1983). *Applied Multiple Regression/Correlation Techniques for the Behavioral Sciences.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Cope N, Harold D, Hill G, *et al.* (2005). "Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia". *Am. J. Hum. Genet.* 76 (4): 581–91.

Coppola, G. (2007). "Why dyslexics make great entrepreneurs". Bloomberg Businessweek. December 12.

DeFries, J. C., Fulker, D. W., Labuda, M. C. (1987). "Reading disability in twins: evidence for a genetic aetiology". *Nature* 329:537–539.

DeFries, J.C., Alarcon, M. (1996). "Genetics of specific reading disability". *Ment Retard Dev Disabil Res Rev* 2:39–47.

Delmar, F., P. Davidsson. (2000). "Where do they come from? Prevalence and characteristics of nascent entrepreneurs". *Entrepreneurship Regional Development* 12 1–23.

Dennis, M.Y., Paracchini, S., Scerri, T.S., Prokunina-Olsson, L., Knight, J.C., et al. (2009). "A common variant associated with dyslexia reduces expression of the *KIAA0319* gene". *PLoS Genet.* 5:e1000436.

Docherty, S. J., Kovas, J., Plomin, R. (2011). "Gene-environment interaction in the etiology of mathematical ability using SNP Sets". *Behavior Genetics* 41 141-154.

Everatt, J., Steffert, B., & Smythe, I. (1999). "An Eye for the Unusual: Creative Thinking in Dyslexics". *Dyslexia*: *An International Journal of Research and Practice* 5: 28-46.

Field**,** A., Baxter, K., Terry, S.F. (2011). "Understanding gene-environment interactions". *Genet Test Mol Biomarkers* 15(6) 371-2.

Galaburda, A., LoTurk, J., Ramus F., Fitch, R., and Rosen, G. (2006). "From genes to behavior in developmental dyslexia". *Nature Neuroscience*, 9(10): 1213-1217.

Gazzaniga, M. S., Ivry, R. B., Mangun, G. R. (2008). *Cognitive neuroscience* (3rd edition). Norton and Company.

Hannula-Jouppi K, Kaminen-Ahola N, Taipale M et al. (2005). "The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia". *PLoS Genet* 1(4):e50.

Harold D, Paracchini S, Scerri T, Dennis M, Cope N, et al. (2006). "Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia". *Mol Psychiatry* 11: 1085–1091, 1061.

Hohnen, B., Stevenson, J. (1999). "The structure of genetic influences on general cognitive, language, phonological, and reading abilities". *Dev Psychol* 35:590–603.

Hornsby, B. (1984). *Overcoming Dyslexia: A straightforward guide for families and teachers*. Marin Dunitz Ltd, London.

Hosmer, D. W., S. Lemeshow. (2001). *Applied Logistic Regression.* Wiley, New York, NY.

Jaccard, J. (2001). *Interaction effects in logistic regression*. Newbury Park: Sage.

Jaccard, J., Turrisi, R., Wan, C. K. (1990). *Interaction Effects in Multiple Regression*. Sage, Thousand Oaks, CA.

Kidd**,** T., Bland, K. S., Goodman, C. S. (1999). "Slit is the midline repellent for the Robo receptor in Drosophila". *Cell*. 96:785-794.

Logan, J. (2009). "Dyslexic entrepreneurs: The incidence; their coping strategies and their business skills". *Dyslexia*. 15 328-346.

Luciano M, Lind PA, Duffy DL, Castles A, Wright MJ, et al. (2007). "A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability". *Biol Psychiatry* 62: 811–817.

Meng, H., Smith, S.D., Hager, K. et al. (2005). "DCDC2 is associated with reading disability and modulates neuronal development in the brain". *Proc Natl Acad Sci* 102(47):17053–17058.

Mesch,G.S., Czamanski, D. (1997). "Occupational Closure and Immigrant Entrepreneurship: Russian Jews in Israel". *The Journal of Socioeconomics*, 26(6), 597-610.

Moffitt, T. E. (2005). "The new Look of Behavioral Genetics in Developmental Psychopathology: Gene - Environment Interplay in Antisocial Behaviors". *Psychological Bulletin*, 131, 522-554.

Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J., Spector, T. (2008). "Is the tendency to engage in entrepreneurship genetic?" *Management Science,* 54(1), 167-179.

Norton, E.C., Ai., C. (2004). "Computing interaction effects and standard errors in logit and probit models". *The Stata Journal*, 4(2):103-116.

Olson, R. K. (2002). "Dyslexia: nature and nurture". *Dyslexia* 8 143-159.

Paracchini S, Steer CD, Buckingham LL, Morris AP, Ring S, et al. (2008). "Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population". *Am J Psychiatry* 12: 1576–1584.

Pinker, S. (2008). *The Sexual Paradox: Men, Women, and the Real Gender Gap*, New York: Scribner.

Plomin, R. DeFries, J. C., McClearn, G. E., McGufin, P. (2008). *Behavioral Genetics* (5th ed.). New York: Worth.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C. (2007). "PLINK: a toolset for whole-genome association and population-based linkage analysis". *American Journal of Human Genetics* 81.

Scerri, T. S., Schulte-Körne**,** G. (2010). "Genetics of developmental dyslexia. *Eur Child Adolesc Psychiatry*. 19(3) 179-197.

Seeger, M., Tear, G., Ferres-Marco, D., Goodman, C.S. (1993). "Mutations affecting growth cone guidance in Drosophila: genes necessary for guidance toward or away from the midline". *Neuron* 10 409–426.

Shane, S. (2003). *A General Theory of Entrepreneurship*. Edward Elgar.

Smith, R. (2008). Entrepreneurs and Dyslexia: Learning Lessons from Dyslexic Entrepreneurs. In Harrison, R.T. & Leitch, C.M. (Eds) *Entrepreneurial Learning: Conceptual Frameworks and Applications*. Routledge.

Spector, T. D., Williams, F. M. (2006). "The UK Adult Twin Registry (TwinsUK)". *Twin Research and Human Genetics* 9(6): 899-906.

Stephenson, S. (1907). "Six cases of congenital word-blindness affecting three generations of one family". *Ophthalmoscope* 5:482–484.

Stevenson, J, Graham, P, Fredman, G, McLoughlin, V. (1987). "A twin study of genetic influences on reading and spelling ability and disability". *J Child Psychol Psychiatry* 28:229–247.

Taipale, M., Kaminen, N., Nopola-Hemmi, J. et al. (2003). "A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain". *Proc Natl Acad Sci* 100(20):11553–11558.

Taylor, K., Walter, J. (2003). "Occupation choices of adults with and without Symptoms of dyslexia". *Dyslexia*, 9: 177-185.

Thomas, C. J. (1905). "Congenital ‘‘word-blindness’’ and its treatment". *Ophthalmoscope* 3:380–385.

von Karolyi, C., Winner, E., Gray, W. & Sherman, G. F. (2003). "Dyslexia linked to talent: Global visual-spatial ability". *Brain and Language* 85:427–31.

Ward, J. (2010). *The student’s guide to cognitive neuroscience*. Psychology Press, London.

Young-Wolff, K**.** C**.**, Enoch, M. A., Prescott, C. A. (2011). "The Influence of Gene-Environment Interactions on Alcohol Consumption and Alcohol Use Disorders: A Comprehensive Review". *Clinical Psychology Review* 31 800-816.