Correlation analysis among single nucleotide polymorphisms in thirteen language genes and culture/education parameters from twenty-six countries

Bo Sun, Changlu Guo and Zhizhou Zhang*

BIOX Biotechnology Center, Harbin Institute of Technology, Weihai, China

*Corresponding: zhangzzbiox@hitwh.edu.cn

Abstract

Language is a vital feature of any human culture, but whether language gene polymorphisms have meaningful correlations with some cultural characteristics during the long-run evolution of human languages largely remains obscure (uninvestigated). This study would be an endeavor example to find evidences for the above question's answer. In this study, the collected basic data include 13 language genes and their randomly selected 111 single nucleotide polymorphisms (SNPs), SNP profiles, 29 culture/education parameters, and estimated cultural context values for 26 representative countries. In order to undertake principal component analysis (PCA) for correlation search, SNP genotypes, cultural context and all other culture/education parameters have to be quantitatively represented into numerical values. Based on the above conditions, this study obtained its preliminary results, the main points of which contain: (1) The 111 SNPs contain several clusters of correlational groups with positive and negative correlations with each other; (2) Low cultural context level significantly influences the correlational patterns among 111 SNPs in the principal component analysis diagram; and (3) Among 29 culture/education parameters, several basic characteristics of a language (the numbers of alphabet, vowel, consonant and dialect) demonstrate least correlations with 111 SNPs of 13 language genes.

Key words: language gene; single nucleotide polymorphism; culture/education; cultural context; correlation

Introduction

Numerous studies are undertaken on food and human living environment, especially biological and physiological consequences they bring to human body, including effects on gene expression in different tissues. Meanwhile, much less studies are taken on similar effects resulted from human cultural and educational activities (CEAs). CEAs do not directly bring in and out of any eye-visible substance from human body (as food and human living environment do), but only induce psychological consequences plus neurological recognition-like output. In fact, more and more evidences have been gathered in that ECAs are

capable of reshaping gene expression profiles of human body[1-6]; ECAs-induced elevation of brain recognition level will influence people's daily life and directly bring changes in behaviors for food intake and many other human activities, thus indirectly bring in and out of eye-visible substance from human body. In this way, CEAs shall be associated with human gene evolution and revolution. Indeed, ECAs still belong to human phenotypic behavior, while any phenotype has innate interaction with genotype and influences each other in a long run.

Language genes are a group of specific genes in charge of unique human language functions. In the past twenty years, some progress has been seen on language gene polymorphisms from different ethnics[7-12]. These polymorphisms are primarily derived from long-term interactions between human and surrounding geographical environments to get adapted. Most of these polymorphisms are even likely produced and optimized in the early stage of human emergence. In a specific profile of geographical conditions, there is normally some type of ECAs for a certain human community that gets adapted to the specific geographical conditions. To what extent these ECAs can participate in the evolution of language gene polymorphism, is an interesting and not yet investigated a question.

As geographical factors, cultural and educational parameters also occur and evolve in a large scale space (at least in a scale as small as a village), and are hard to be modeled in a laboratory way in which variables can be changed one by one to observe relevant consequences in other variables. Even if we can employ a small number of people in a village-like space to perform culture-education activity-based research, it is still a big challenge to keep the observation for enough long a time (such as 20-50 years). As we know, human gene mutation happens all the time, but a stable polymorphism point in a genome is through selection in many generations before it can stably exist in a population. This means heritable SNPs with a stable pattern are normally accumulated in a long time as several hundred or thousand years, even longer. So what we can do is to employ correlation analysis method(s) to tackle questions as to what extent culture/education parameters play a role in reshaping the profile of language gene polymorphisms.

This study focused on correlation analysis among a total 111 SNPs from these 13 genes and a series of culture/education factors collected from 26 countries, and the author did find some interesting correlational parameters, including several strongest positive and negative correlations.

Methods

Language genes and their SNPs

Language is an emergent complicated function of human being, though many other animals also have their own 'languages'. If a gene mutation is statistically or experimentally associated with a certain language function loss, it would be called language gene. Language gene SNP data were all randomly selected for each gene in the dbSNP database: https://www.ncbi.nlm.nih.gov/snp/; Table 1 listed 13 language genes,

and a total 111 SNPs from these 13 genes were randomly selected for this study (Table 2, Supplementary file-1). Quantification of SNPs was described in Supplementary file-2.

Table 1 Thirteen language genes

	Gene name	Compromised ability when mutated (example)	Reference
1	FOXP1	Expressive language	13
2	FOXP2	Speech	14
3	CNTNAP2	Early language development	15,16,17
4	TPK1	Syntactic and lexical ability	18,19
5	DCDC2	Reading, dyslexia	20,21,22
6	KIAA0319	Reading, dyslexia	23,15,20,24
7	TM4SF20	Language delay; communication disorder	25
8	FLNC	Reading, language	26
9	ATP2C2	Memory	27
10	ROBO1	Phonological buffer	28,29
11	ROBO2	Expressive vocabulary	30
12	CMIP	Reading, memory	15,20,27
13	NFXL1	Speech	31

Table 2 111 SNPs of thirteen language genes

	Variables	Abbreviation		Variables	Abbreviation
1	ROBO1 rs34841026	ROBO1	57	DCDC2 rs33943110	DCD-4
2	ROBO2 rs11127602	ROBO-1	58	DCDC2 rs33914824	DCD-5
3	ROBO2 rs10865561	ROBO-2	59	DCDC2 rs9467075	DCD-6
4	ROBO2 rs5788280	ROBO-3	60	DCDC2 rs9460973	DCD-7
5	ROBO2 rs3923745	ROBO-4	61	DCDC2 rs3846827	DCD-8
6	ROBO2 rs3923744	ROBO-5	62	DCDC2 rs3789219	DCD-9
7	ROBO2 rs1163750	ROBO-6	63	CNTNAP2 rs1637842	CNTN-1
8	ROBO2 rs1163749	ROBO-7	64	CNTNAP2 rs1637841	CNTN-2
9	ROBO2 rs1163748	ROBO-8	65	CNTNAP2 rs1479837	CNTN-3
10	ROBO2 rs1031377	ROBO-9	66	CNTNAP2 rs1468370	CNTN-4
11	TM4SF20 rs6724955	TM1	67	CNTNAP2 rs1062072	CNTN-5
12	TM4SF20 rs44675173	TM2	68	CNTNAP2 rs1062071	CNTN-6
13	TM4SF20 rs4675172	TM3	69	CNTNAP2 rs987456	CNTN-7
14	TM4SF20 rs4673192	TM4	70	CNTNAP2 rs700309	CNTN-8
15	TM4SF20 rs4438464	TM5	71	CNTNAP2 rs700308	CNTN-9
16	TM4SF20 rs4428010	TM6	72	CNTNAP2 rs3194	CNTN-10
17	TM4SF20 rs4408717	TM7	73	CMIP rs201316817	CMI-1
18	TPK1 rs113536847	TPK-1	74	CMIP rs183876152	CMI-2
19	TPK1 rs79464600	TPK-2	75	CMIP rs183075361	CMI-3
20	TPK1 rs77358162	TPK-3	76	CMIP rs114894868	CMI-4
21	TPK1 rs28380423	TPK-4	77	CMIP rs79979027	CMI-5
22	TPK1 rs17170295	TPK-5	78	CMIP rs74031247	CMI-6

	Variables	Abbreviation		Variables	Abbreviation
23	TPK1 rs12333969	TPK-6	79	CMIP rs60152409	CMI-7
24	TPK1 rs6953807	TPK-7	80	CMIP rs57603843	CMI-8
25	NEXI 1 rs1964425	NEX-1	81	CMIP rs35429777	CMI-9
26	NEXI 1 rs1822030	NEX-2	82	CMIP rs34119643	CMI-10
27	NEXI 1 rs1822029	NEX-3	83	ATP2C2 rs78371901	ATP-1
28	NEXI 1 rs1812964	NEX-4	84	ATP2C2 rs74038217	ATP-2
29	NEXI 1 rs1545200	NEX-5	85	ATP2C2 rs62640935	ATP-3
30	NEXI 1 rs1440228	NEX-6	86	ATP2C2 rs62640932	ATP-4
31	NEXI 1 rs1371730	NEX-7	87	ATP2C2 rs62640931	ATP-5
32	NEXI 1 rs1036681	NEX-8	88	ATP2C2 rs62050917	ATP-6
33	NEXI 1 rs978094	NEX-9	89	ATP2C2 rs16973859	ATP-7
34	NEXI 1 rs920462	NEX-10	90	ATP2C2 rs13334642	ATP-8
35	EXP1 rs7638391	EXP1	91	ATP2C2 rs4782970	ATP-9
36	FOXP1 rs76145927	FOXP1-1	92	ATP2C2 rs4782948	ATP-10
37	FOXP1 rs75214049	FOXP1-2	93	KIAA0319 rs138160539	KIA-1
38	FOXP1 rs17008544	FOXP1-3	94	KIAA0319 rs117692893	KIA-2
39	FOXP1 rs17008063	FOXP1-4	95	KIAA0319 rs114195393	KIA-3
40	FOXP1 rs11914627	FOXP1-5	96	KIAA0319 rs699461	KIA-4
41	FOXP1 rs7639736	FOXP1-6	97	KIAA0319 rs699462	KIA-5
42	FOXP1 rs1499893	FOXP1-7	98	KIAA0319 rs699463	KIA-6
43	FOXP1 rs1053797	FOXP1-8	99	KIAA0319 rs730860	KIA-7
44	FLNC rs2291569	FLN-1	100	KIAA0319 rs10946705	KIA-8
45	FLNC rs2291568	FLN-2	101	KIAA0319 rs75674723	KIA-9
46	FLNC rs2291566	FLN-3	102	KIAA0319 rs75720688	KIA-10
47	FLNC rs2291565	FLN-4	103	FOXP2 rs10227893	FOXP2-1
48	FLNC rs2291563	FLN-5	104	FOXP2 rs10244649	FOXP2-2
49	FLNC rs2291562	FLN-6	105	FOXP2 rs12705977	FOXP2-3
50	FLNC rs2291561	FLN-7	106	FOXP2 rs61732741	FOXP2-4
51	FLNC rs2291560	FLN-8	107	FOXP2 rs61758964	FOXP2-5
52	FLNC rs2291558	FLN-9	108	FOXP2 rs62640396	FOXP2-6
53	FLNC rs2249128	FLN-10	109	FOXP2 rs73210755	FOXP2-7
54	DCDC2 rs35029429	DCD-1	110	FOXP2 rs1058335	FOXP2-8
55	DCDC2 rs2274305	DCD-2	111	FOXP2 rs61753357	FOXP2-9
56	DCDC2 rs34584835	DCD-3			

Collection of culture/ducation parameters

Most education/culture parameters were collected from three websites, including Baidu (https://baike.baidu.com), Bing (https://cn.bing.com), omniglot (https://omniglot.com/writing/languages.htm) and United Nations databases (UND) (http://data.un.org/Default.aspx). (Supplementary file-3) Detailed data are not provided in this manuscript because of page limitation, but can be requested from the corresponding author. This study is a part of our larger scale one that has to collect much more parameters (language gene single nucleotide polymorphism or SNPs, geographical factors, societal factors, etc.) except for education/culture parameters. Besides 29 education/culture parameters (Table 3), we are collecting 111 (SNPs) + 61 (Education/ Culture/ Geography etc.) = 172 parameters in total for around 150 countries, and only 26 countries in Table 4 have been collected for all 172 parameters by the time this manuscript is written.

Table 3 Education/culture parameters

	Variables (in number)	Abbreviation	Data source
1	Dialect types in the sample country	dial	Baidu, Bing
2	Alphabet of the country's main language	alph	Baidu, Bing, omniglot
3	Vowels of the country's main language	vowe	Baidu, Bing, omniglot
4	Consonants of the country's main language	cnso	Baidu, Bing, omniglot
5	Technology literature	tech	PubMed
6	Engineering literature	engi	PubMed
7	College/institute	inst	Baidu, Bing
8	News agency	news	Baidu, Bing
9	Newspaper/periodicals	newp	Baidu, Bing
10	Broadcast language types	brod	Baidu, Bing
11	Teachers in pre-primary education, both sexes	tpp	UND
12	Teachers in primary education, both sexes	tpe	UND
13	Teachers in secondary education, both sexes	tse	UND
14	Exports of cultural goods	ecgd	UND
15	Imports of cultural goods	icgd	UND
16	Exports of books and press goods	ebpd	UND
17	Exports of performance and celebration goods	epcd	UND
18	Exports of visual arts and crafts goods	evad	UND
19	Natural relics	nrel	UND
20	Cultural relics	crel	UND
21	Natural/cultural double relics	ncrl	UND
22	Total world cultural relics	twcr	UND
23	Power distance Ref.	PDI	Ref.[32-34]
24	Uncertainty avoidance	UAI	Ref.[32-34]
25	Individualism vs collectivism	InCo	Ref.[32-34]
26	Masculinity and Femininity	Masc	Ref.[32-34]
27	Long and short term orientation	Long	Ref.[32-34]
28	Indulgence and restraint	Rest	Ref.[32-34]
29	Cultural context	cuco	Ref.[35]

Selected countries with decent representation

The countries were selected based on a range of cultural context vale (Table 4). In order to give cultural context parameter quantitative estimation (thus can be conveniently used in statistical analysis), the relative cultural context levels for known countries/areas in Table 3 are estimated from 2 to 42 as shown in the column 2. Then other countries were provided similar or same estimated curo level (ECL) value by comparison[36-38]. For example, seven African countries (Benin, Cameroon, Gambia, Kenya, Senegal, Tunisia, and Tanzania) were given the same ECL value as 28. Nepal, Myanmar, Bhutan, Sri Lanka, Bangladesh were given the same ECL value as Vietnam because they all belong to south-western Asia.

Table 4 Estimated curo level (ECL) values of selected 26 countries in this study

Nation/People/Region	ECL	Nations used in this study
Japanese	42	Japan
Korea	40	
Chinese	38	China, Mongolia
Vietnam	36	Vietnam, Nepal, Myanmar, Bhutan, Sri Lanka, Bangladesh
Arab, Middle East, West Asia	34	Iran
Indian	32	
Greek, South European	30	Slovenia, Greece, Bosnia and Herzegovina
Africa	28	Benin, Cameroon, Gambia, Kenya, Senegal, Tunisia, Tanzania
Mexican, Middle America	26	Barbados, Dominican, Peru
Spanish	24	
Italian	22	
Russian	20	
French	18	France
French Canadian	16	
English	14	UK
English Canadian	12	
Australian	10	
North American	8	
Scandinavian	6	
Swiss	4	
German	2	Germany

Correlation analysis

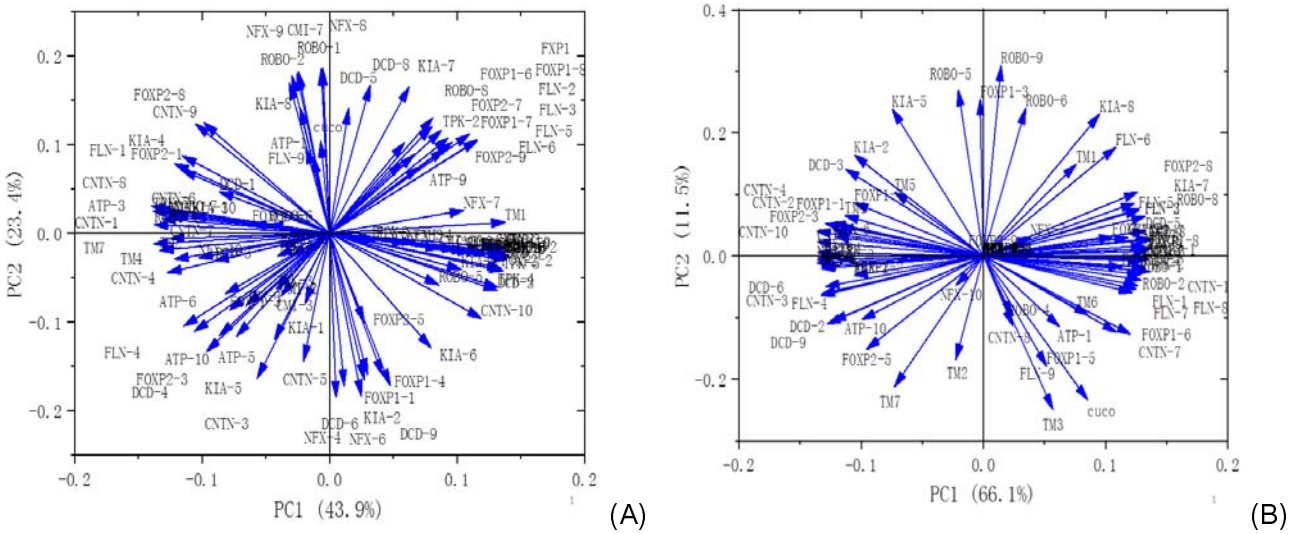
The software Origin was employed to analyze the potential correlation relationship between all parameters in this study. By using Origin2019, principal components analysis (PCA) was performed to extract potential

correlation patterns among multiple numerical variables. PCA is a feasible technique to emphasize variation and visualize strong patterns in a large dataset. In a typical PCA diagram, correlated variables are drawn as short or long arrows in which long arrow represents strong correlation and short arrow represents weak correlation; plus, two arrows can form a right, obtuse or acute angle, representing no-correlation, negative correlation or positive correlation, respectively. Quantitative correlation between any two factors was undertaken as follow. The basic variables to measure in a PCA result plot are arrow length and the angle between two arrows. The correlation score = angle score ×arrow score; the angle score ranges from -5 to +5. The angles (0-15], (15-30], (30-45], (45-60], (60-75], (75-105], (105-120], (120-135], (135-150], (150-165], (165-180] are scored 5, 4, 3, 2,1, 0, -1, -2, -3, -4, -5, respectively. The arrow score = the length of arrow-1× the length of arrow-2. The angle value and the arrow length value can be conveniently obtained with ImageJ software (Supplementary file-4, Supplementary file-5).

Results and Discussion

General correlations among all selected SNPs

Figure 1(A) demonstrated a general landscape of correlational relationship among 111 SNPs. Here are several apparent findings. First, no all SNPs from a single gene stay aggregated in a corner of the PCA map, as expected; because all genes passed through million years of mutual adaptation; if all SNPs from a single gene stay aggregated, that would means that many other SNPs from other genes counteract with them. Such a single gene would likely be lost during evolution. Second, we still can find that a couple of (not all) SNPs from a single gene can aggregate at a specific position of the PCA map. Such examples include FLN-2~FLN-3~FLN-5, FOXP1-6~FOXP1-5~FOXP1-7, FOXP1-1~FOXP1-4, ATP-5~ATP-10, CNTN-1~CNTN-6, etc. Third, there are about four big clusters of SNPs, any one of them having a totally negative correlational group and two other groups with much less level of correlations.



(A) (B)

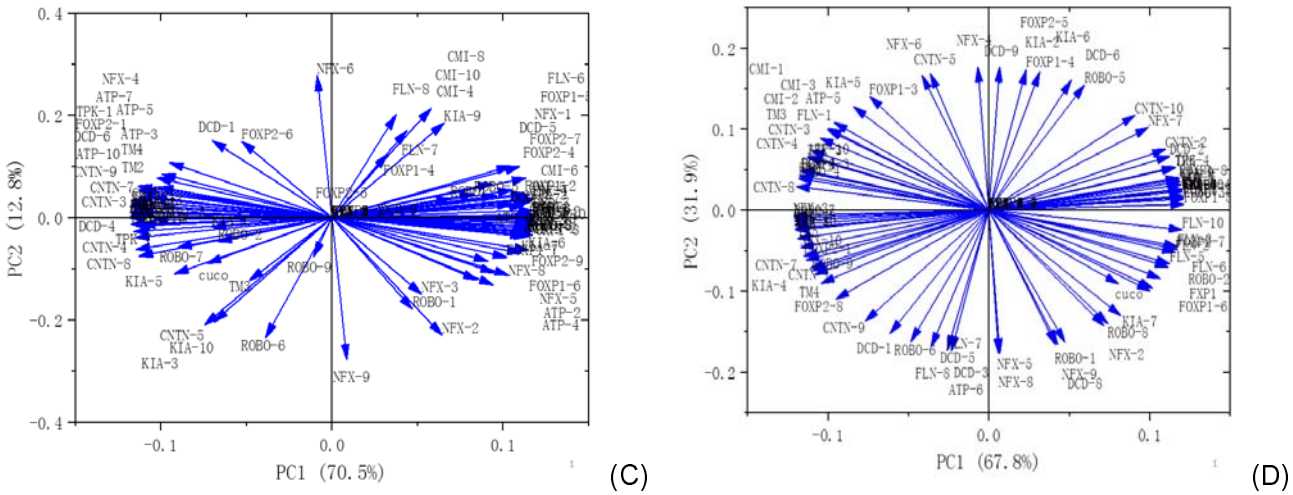


Figure 1 PCA for 111 language gene SNPs. (A) All SNPs in all 26 countries; (B) All SNPs in 10 high cultural context countries; (C) All SNPs in 11 intermediate cultural context countries; (D) All SNPs in 6 low cultural context countries.

Effect of cultural context (cuco) on correlational patterns among language gene SNPs

Cultural context is an important parameter that can distinguish individual country/region from each other. Countries can be classified as high, intermediate and low cultural context types. In table 4, countries were divided as three groups by cuco values: [2,26], [28,36) , and ≥36. Correlation analysis among 111 SNPs was performed for three groups of country, and it was found that SNP correlation pattern changed significantly. In the low cuco group, correlations among language gene SNPs became evenly distributed (Fig.1D), which is not found in high (Fig.1B) and intermediate (Fig.1C) cuco groups.

Table 5 provided parameters that have strongest correlations with cuco in three groups of country. The SNP FOXP1-5 is positively correlated with cuco under high cultural context, but became negatively correlated under intermediate cultural context. The other three SNPs, KIA-5, FLN-6 and CNTN-5, demonstrated similar conversions. Though very limited number of countries was employed for PCA analysis, the above results were still very intriguing, in that this study may provide a good case in which a culture/education factor highly influences genotype interaction pattern(s).

Table 5 Parameters that have strongest correlations with cuco in three groups of country

cuco level	Positive correlation with cuco parameter	Negative correlation with cuco parameter
High[a]	TM-3, Foxp1-5, FLN-9, ATP-1	KIA-5, ROBO-5, TM5, KIA-2, DCD-3, Foxp1-4
Intermediate[b]	KIA-5, ROBO-7, CNTN-8, TM-3,	FLN-6, Foxp1-5, NFX-1, FLN-8, KIA-9, FLN-7,

	CNTN-5, KIA-10, KIA-3	Foxp1-4, CMI-4, CMI-8, CMI-10
Low[c]	Foxp1-6, FXP1, KIA-7, ROBO-8, ROBO-2, NFX-2, FLN-6	ATP-5, KIA-5, Foxp1-3, CMI-1, CMI-2, CMI-3, CNTN-3, CNTN-4, FLN-1
all	ROBO-1, ROBO-2, CMI-7, KIA-8, NFX-9, DCD-5	DCD-6, DCD-9, NFX-4, Foxp1-1, Foxp1-4, KIA-2, NFX-6, CNTN-5

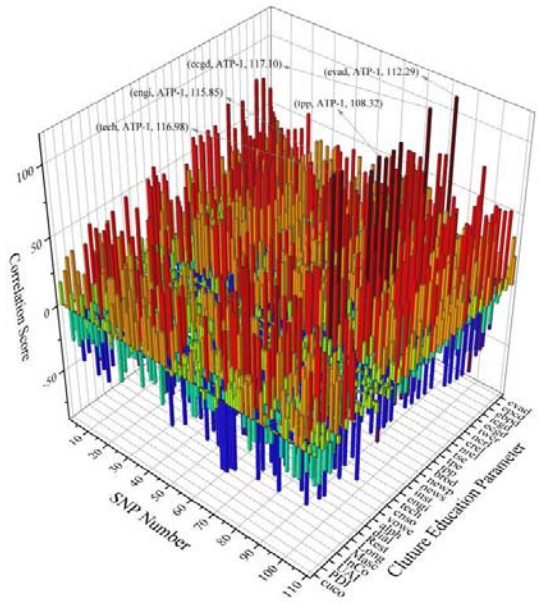
Note: [a]cuco value≥36; [b]cuco value between 26~36; [c]cuco value ≦26 ;

Correlation between a specific gene SNP and culture/education parameters

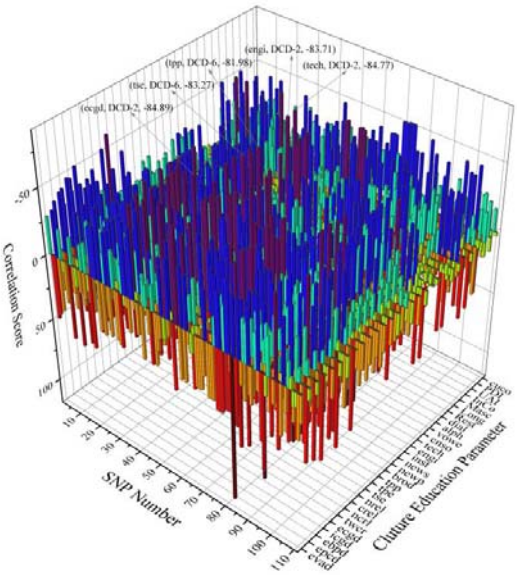
Correlation analysis among all variables in Table 2 and Table 3 is of interest. In theory, culture/educational activities still belong to human phenotypes, while any phenotype will automatically bring feedback to genotypes that determine it. So the long-run human culture/education factors shall somehow interact with gene polymorphisms and form related (weak or strong) functional correlations.

In Figure 2(A), strongest positive correlations were seen at ATP-1-tech, ATP-1-ecgd, ATP-1-engi, ATP-1-tpp and ATP-1-evad. ATP-1 is one of the SNPs of language gene ATP2C2, which encodes the ATPase secretory pathway Ca2+ transporting-2 protein. Diseases associated with ATP2C2 include specific language impairment and speech/ communication disorders. The five parameters, tech (Technology literature), engi (Engineering literature), ecgd (Exports of cultural goods), tpp (Teachers in pre-primary education, both sexes) and evad (Exports of visual arts and crafts goods), have similar highest correlation values. It is surprising that the five strongest correlations were all formed by ATP-1 SNP. How these five parameters mechanically correlate with the language gene SNP ATP-1 will be worth tackling in the future.

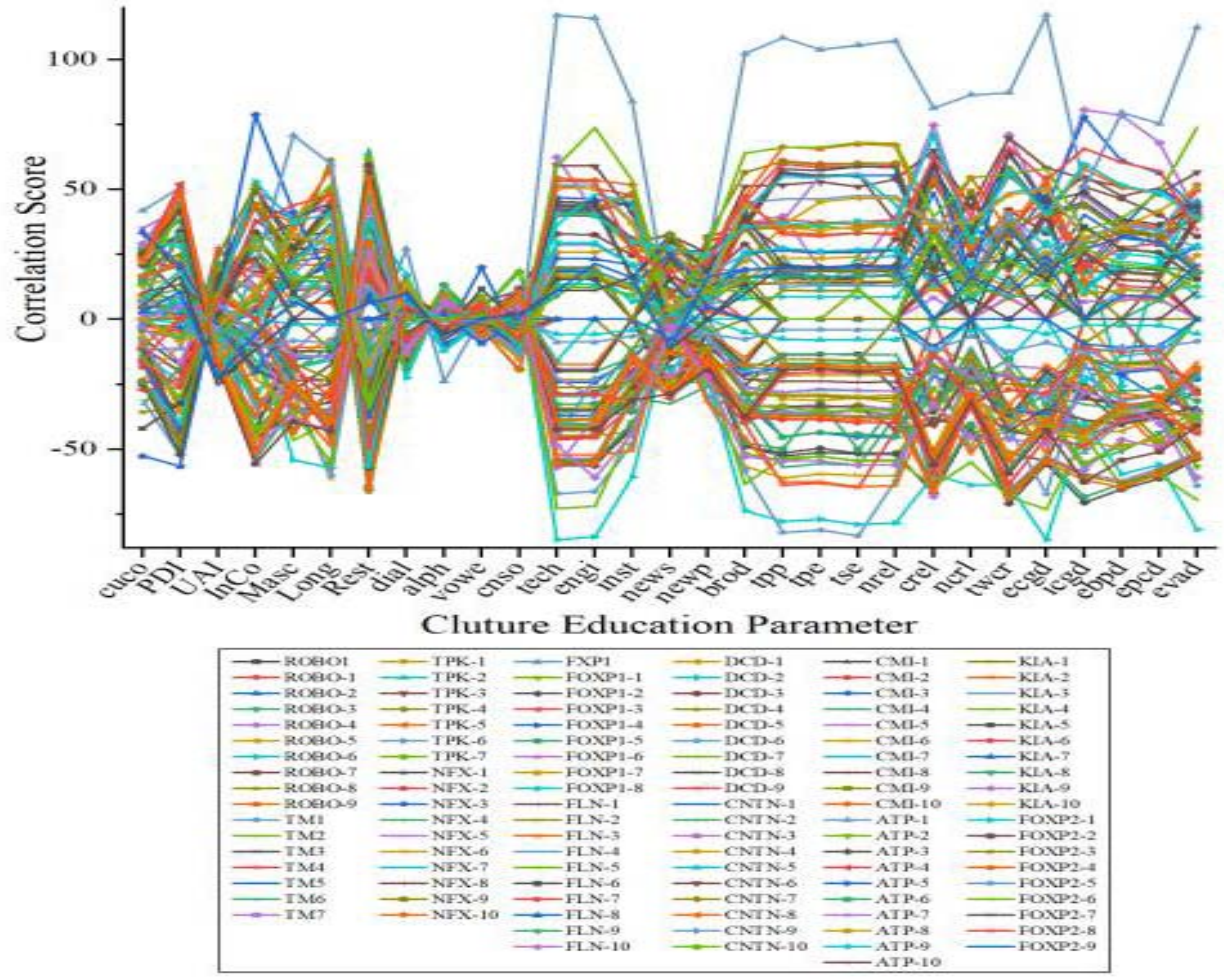
In Figure 2(B), strongest negative correlation were seen at DCD-2-tech, DCD-2-engi, DCD-6-tse (Teachers in secondary education, both sexes), DCD-2-ecgd, and DCD-2-evad. DCD-2 and DCD-6 are two SNPs of language gene DCDC2.This gene encodes a doublecortin domain-containing family member, which is thought to function in neuronal migration where it may affect the signaling of primary cilia. Mutations in this gene are associated with reading disability (RD) type 2 (developmental dyslexia)[40]. Robert Plomin[39] claimed that a person's achievement in English, mathematics, science, literacy and arts is significantly heritable and influenced by similar groups of genes. They compared 12,500 pairs of twins' genome sequences and exam available scores of all school tests. One SNP locus (rs807701, not the same DCDC2 SNP as in this study) on DCDC2 gene seemed highly associated with reading ability, and previous studies already found that DCDC2 mutations will lead to reading difficulty[40]. The large scale sampling of the above study enabled the authors to abstract relatively confident conclusions on how genes affect a person's academic performance, including language abilities and general output in the liberal education[41]. Such performance is supposed to have a good chance to influence people's behavior in cultural/educational activities.



(A)



(B)



(C)

Figure 2 Correlation among language gene SNPs and 30 culture/education parameters. (A) Five strongest positive correlations were shown; (B) Five strongest negative correlations were shown; (C) Smallest correlations were exposed clearly at five culture/education parameters.

Figure 2(C) indicates that little correlation was found for UAI (Uncertainty avoidance), dial (Dialect types in the sample country), vome (Vowels of the country's main language), alph (Alphabet of the country's main language) and cnso (Consonants of the country's main language); According to Hofstede[32-34], uncertainty avoidance describes the extent to which people attempt to cope with anxiety by minimizing uncertainty. Cultures with high score in uncertainty avoidance prefer rules and structured disciplines, and employees tend to keep longer with their current employer. UAI may be a universal parameter in any country/region though its content differs a lot in different country/region, and that may be why it has a very small correlation value with all SNPs; The other four parameters, dial, vome, alph and cnso, are all basic language factors. It is interesting why they also have very small correlation values with all language gene SNPs.

Conclusions

In the past several years the authors have been collecting multiple parameters (history /geography /religion/ genetics/ education/ culture/ society) for many countries, and started to supply with human language gene polymorphism data plus appropriate correlation analysis[38,42-43]. Such multi-layer data, when in a scale large enough, shall be useful for investigations on interdisciplinary questions on the boundary of natural science and social science. Recently, researchers found interesting interplay of genetics and culture in Ethiopia[7], highlighting the importance to employ data from both natural science and social science to deepen the understanding of cultural questions.

In this study, the basic data include 13 language genes and their randomly selected 111 single nucleotide polymorphisms (SNPs), SNP profiles in 26 countries, 29 culture/education parameters in 26 countries, and estimated cultural context values for 26 countries. In order to undertake principal component analysis (PCA), SNP genotypes, cultural context and all other culture/education parameters have to be quantitatively represented into numerical values.

In this study, only 26 countries were used for all analysis. However, the namelist (Table 4) contains descent representation, including one eastern (Asia) developed country, three western developed countries, two developing countries in eastern Asia, three ordinary European countries, seven African countries, six ordinary countries in south Asia, and several typical ones in west Asia (one) and Middle America (three). So the results in this study would be preliminary but valuable.

Based on the above conditions, this study obtained its preliminary results, the main points of which contain: (1) The 111 SNPs form several clusters of correlational groups with positive and negative correlations with each other; (2) Cultural context level apparently influences the correlational patterns among 111 SNPs in the principal component analysis diagram; and (3) Among 29 culture/education parameters, several basic characteristics of a language (the numbers of alphabet, vowel, consonant and dialect) surprisingly demonstrate least correlations with 111 SNPs of 13 language genes.

ACKNOWLEDGMENTS

This study was supported by National Research Center for Foreign Language Education Grant (ZGWYJYJJ10A042) and State Language Commission Research Grant (YB135-117). The authors thank Nie Qi, Pan Yusheng for their help in collecting sample data.

References

1. David, Bueno. Genetics and Learning: How the Genes Influence Educational Attainment. Frontiers in psychology. 2019;10: 1622.

2. Rezapour S, Shiravand M, Mardani M. Epigenetic changes due to physical activity. Biotechnology and Applied Biochemistry. 2018;65(6): 761-767.

3. Zhang M, Yan S, Pan W, et al. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. Nature. 2019;569(7754): 112–115.

4. Hernandez LM, Blazer DG, Medicine IO. Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate. National Academies Press. 2006.

5. Heinrichs M, Baumgartner T, Kirschbaum C, et al. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biological Psychiatry. 2003;54(12): 1389-1398.

6. Kosfeld M, Heinrichs M, Zak PJ, et al. Oxytocin Increases Trust in Humans. Nature. 2005;435(7042): 673–676.

7. S López, Tarekegn A, Band G, et al. Evidence of the interplay of genetics and culture in Ethiopia. Nature Communications. 2021;12(1).

8. Lang X, Zhang W, Song X, et al. FOXP2 contributes to the cognitive impairment in chronic patients with schizophrenia. Aging (Albany NY). 2019;11(16): 6440-6448.

9. Rao W, Du X, Zhang Y, et al. Association between forkhead-box P2 gene polymorphism and clinical symptoms in chronic schizophrenia in a Chinese population. Journal of Neural Transmission. 2017;124(7): 1-7.

10. Jin Y, Birlea SA, Fain PR, et al. Common variants in FOXP1 are associated with generalized vitiligo. Nature Genetics. 2010;42(7): 576–578.

11. Zare S, F Mashayekhi, Bi Dab Adi E. The association of CNTNAP2 rs7794745 gene polymorphism and autism in Iranian population. Journal of Clinical Neuroence. 2017;39: 189-192.

12. Karaca I, Yilmaz SG, Palamar M, et al. Evaluation of CNTNAP2 gene rs2107856 polymorphism in Turkish population with pseudoexfoliation syndrome. International Ophthalmology. 2019;39(1): 167-173.

13. Bacon C, Rappold GA. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Human Genetics. 2012;131(11): 1687-1698.

14. Lai C, Fisher SE, Hurst JA, et al. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001;413(6855): 519-523.

15. Newbury DF, Paracchini S, Scerri TS, et al. Investigation of Dyslexia and SLI Risk Variants in Reading- and Language-Impaired Subjects. Behavior Genetics. 2011;41(1): 90-104.

16. Vernes SC, Newbury DF, Abrahams BS, et al. A functional genetic link between distinct developmental language disorders. New England Journal of Medicine. 2008;359(22): 2337-2345.

17. Whitehouse A, Bishop D, Ang QW, et al. CNTNAP2 variants affect early language development in the general population. Genes, Brain and Behavior. 2012;11(4): 501.

18. Villanueva P, Newbury DF, Jara L, et al. Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. European Journal of Human Genetics. 2011;9(6): 687-695.

19. Fattal I, Friedmann N, Fattal-Valevski A. The crucial role of thiamine in the development of syntax and lexical retrieval: A study of infantile thiamine deficiency. Brain. 2011;134(6): 1720-1739.

20. Scerri TS, Morris AP, Buckingham LL, et al. DCDC2, KIAA0319 and CMIP Are Associated with Reading-Related Traits. Biological psychiatry. 2011;70(3): 237-245.

21. Deffenbacher KE , Kenyon JB , Hoover DM , et al. Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. Human Genetics. 2004;115(2): 128-138.

22. Schumacher J, Anthoni H, Dahdouh F, et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. The American Journal of Human Genetics. 2006;78(1): 52-62.

23. Silvia P, Ankur T, Sandra C, et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. Human Molecular Genetics. 2006;15(10): 1659-1666.

24. Francks C, Paracchini S, Smith SD, et al. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. The American Journal of Human Genetics. 2004;75(6): 1046-1058.

25. Wiszniewski W, Hunter JV, Hanchard NA, et al. TM4SF20 Ancestral Deletion and Susceptibility to a Pediatric Disorder of Early Language Delay and Cerebral White Matter Hyperintensities. American Journal of Human Genetics. 2013;93(2): 405.

26. Gialluisi A, Newbury DF, Wilcutt EG, et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain and Behavior. 2014;13(7): 686-701.

27. Newbury FD, Winchester L, Addis Paracchini S, et al. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. American Journal of Human Genetics. 2009;85(2): 264–272.

28. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, et al. The Axon Guidance Receptor Gene ROBO1 Is a Candidate Gene for Developmental Dyslexia. Plos Genetics. 2005;1(4): e50.

29. Bates TC, Luciano M, Medland SE, et al. Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. Behavior Genetics. 2011;41(1): 50-57.

30. Pourcain BS, Cents R, Whitehouse A, et al. Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nature Communications. 2014;5: 4831.

31. Villanueva P, Nudel R, Hoischen A, et al. Exome Sequencing in an Admixed Isolated Population Indicates NFXL1 Variants Confer a Risk for Specific Language Impairment. PLOS Genetics. 2015;11(3): e1004925.

32. Hofstede GJ. Culture's Consequences: International Differences In Work-Related Values. Sage Publications. 1984.

33. Yasemin H, Ülk ü han BD, Ş ü kran SY. Relationship between uncertainty avoidance culture, entrepreneurial activity and economic development. Procedia-Social and Behavioral Sciences. 2014;150: 908-916.

34. Hofstede G, Hofstede GJ, Minkov M. Cultures and Organizations, Software of the mind. Intercultural Cooperation and Its Importance for survival. Southern Medical Journal. 2010;13(3): S219–S222.

35. Xia W, Sun B, Zhang ZZ.Correlation analysis between cultural context level and

education/culture/geography/society-related parameters in twenty-six countries. International Conference on Digital Technology in Education. 2021.

36. Morain G, Hall ET, Hall MR. Understanding Cultural Differences: Germans, French and Americans. Modern Language Journal. 1990;75(1): 135.
37. Hall ET. Beyond Culture. New York: Doubleday. 1976.
38. Hall ET. The Sileng Language. New York: Doubleday. 1959.
39. Davis OS, Band G, Spencer CC, et al. The correlation between reading and mathematics ability at age twelve has a substantial genetic component. Nature Communications. 2014;5: 4204-4211.
40. Meng HY, Shelley DS, Karl H, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. PNAS. 2005;102(47): 17053-17058.
41. Xia W, Qin H, Li XW, et al. Language gene basis for precision personalized education and liberal education.International conference on E-society, E-education and E-technology. 2018;112-116.
42. Xia W, Zhang ZZ, Guo CL. Correlation Analysis between English-Chinese Translation-Based Writing Error Types and Language Gene Polymorphisms for Chinese Graduate Students. International Journal of Learning and Teaching. 2019;333-338.
43. Xia W, Zhang ZZ, Guo CL. Novel Education Technology May Derive from Personal Genome Data: A Language Gene Polymorphism Site Potentially Associated with Translation-writing Errors in A Bilingual Classroom of Chinese Students. International Conference on Modern Educational Technology. 2019;45-48.

Supplementary files

Supplementary file-1 Selected 111 SNPs from 13 language genes
Supplementary file-2 SNP quantification method and SNPs data in 26 countries
Supplementary file-3 Culture education parameters data
Supplementary file-4 PCA for language gene and culture education parameters
Supplementary file-5 SNP and culture education correlation value data