Language Gene Network Patterns May Facilitate Relationship Setting-up between Language Genotypes and Students' Class-Performance

Wei Xia
School of Languages and Literature, Harbin Institute of Technology, Weihai, China
Email: xiawei2015@hitwh.edu.cn

Zhizhou Zhang
School of Marine Science and Technology, Harbin Institute of Technology, Weihai, China
Email: zhangzzbiox@hitwh.edu.cn

***Abstract*—How individual biological phenotypes are encoded by genome sequences will be elucidated more and more in the post-genomic era. Especially, the relationship between language abilities and language genes is to be decoded inevitably. In this article, it is conceptualized that different language ability-related class-performance of students is largely encoded by different combinations of a cluster of language genes. Any two persons have the same set of language genes, but each language gene holds different variations or mutations in its DNA sequence in the human population, and these variations brings up differential influence on the gene's function. The combinations of such variations in different language genes set up the molecular basis of the fact that almost every person is different from each other in the context of language abilities and performances. Some mutations in the key language genes (such as FOXP1 and FOXP2) are found to lead to severe language disorders, but for most students, only mild mutations or variations exist in their language genes, thus demonstrating normal language ability but differential levels of class-performance. Biological technology will gradually help to finish DNA sequences of every student, pinpoint his defects in some language genes, figure out his advantage and shortcoming, and thus promote a series of individualized approach for teaching and education.**

***Index Terms*—language gene, language ability, individualized, teaching, education**

I. Introduction

Speech is one of the most complex and refined motor skills of human being. Since the finding of FOX2 [1], more and more language genes have been characterized. About 7% 5-7 years old children develop speech and language disorders and such diseases or phenotypes are known to be highly heritable. Because multiple genes are involved in most cases, the inheritance patterns are usually complex. Besides, some types of disease, like autism, are apparently associated with speech and language disorders at personalized content. So, it is often concerned that we may need a quantitative regime to describe the defects of those children in order to set up personalized teaching approach for their education. The similar consideration is also obvious for those college students that possess apparently distinct language capacity and skills.

Functional study and category of known language genes is a prerequisite. In the past twenty years, about 15-20 language genes [2] were gradually distinguished in different language disorder-associated studies. This paper described several selected potential language genes one by one, and some potential implications in teaching or the general education are discussed.

II. Some Known Language Genes

A. FOXP1

Mutations in Foxp1 normally lead to neurodevelopmental disorders that sometimes include pronounced impairment in language and speech skills. Horn *et al* [3] found three children of 5-7 years old with moderate mental retardation but with sequence deletions in forkhead box P1 (FOXP1) gene and significant language and speech deficits. Considering the experiment scale of 1523 patients with mental retardation and 4104 ancestrally matched controls, the linkage between FOXP1 gene mutations and language and speech deficits is thought solid and causal. Hamdan *et al* [4] found a FOXP1 mutation in two nonsyndromic intellectual disability patients with autism. The patients also show severe language impairment, mood lability with physical aggressiveness, and specific obsessions and compulsions, but their oral expression seems normal. Song *et al* [5] discovered a FOXP1 de novo mutation that associates with severe speech delay in an individual belonging to a non-Caucasian population. She was 22 years old with a short stature (141 cm, body weight 44.3 kg) and delayed speech (unable to speak), but receptive language abilities were relatively well developed as indicated by her understanding of relational concepts.

Manuscript received February 1, 2017; revised May 1, 2017.

doi: 10.18178/ijlt.3.4.259-263

B. FOXP2

FOXP2 is the first characterized language gene [1] that encodes a protein associated with intriguing aspects of cognitive function in humans, non-human mammals, and song-learning birds. Mutations of the human FOXP2 gene cause a monogenic speech and language disorder. Single nucleotide polymorphism (SNP) in FOXP2 gene is a valuable consideration because many sequence variations or SNPs can be easily scanned with moderate cost in many students and then a molecular linkage can be set up between students' different language abilities and gene variation patterns.

C. CNTNAP2

Vernes *et al.* [6] measured SNPs in FOXP2 and CNTNAP2 in human samples from 184 families with specific language impairment (SLI). They found that almost all children with nonsense-word-repetition language defect possess a mutation in CNTNAP2 gene, and the mutation position is highly associated with autism in other studies.

D. FLNC/RBFOX2

Gialluisi *et al.* [7] performed a genome-wide association scan (GWAS) meta-analysis using three datasets comprising individuals with histories of reading or language problems, and their siblings. Language and reading abilities are heritable traits that share some genetic influences with each other. They identified novel associations at two SNPs located respectively at the FLNC and RBFOX2 genes. FLNC encodes a structural protein for cellular cytoskeleton re-modeling, and RBFOX2 regulates alternative splicing in neurons. Besides, RBFOX2 is a downstream target of FOXP2 gene, because a FOXP2-binding site was found 5kb from the RBFOX2 SNP position.

E. TM4SF20

In a genomic study of 15,493 children (all shared a diagnosis of communication disorder, ranging from early language delay to autism spectrum disorder) referred to the Medical Genetics Laboratories at Baylor College of Medicine, by using 180,000 oligonucleotide-based whole-genome microarray, Wiszniewski *et al*. [8] described a complex 4 kb deletion in TM4SF20 gene that segregates with early childhood communication disorders in 15 unrelated families mainly from Southeast Asia. The deletion removes the penultimate exon 3 of TM4SF20, a gene encoding a transmembrane protein of unknown function. Functional studies indicated that the deletion leads to a truncated form of the protein that is missing two of its four transmembrane domains and, although stable, fails to target to the plasma membrane and accumulates in the cytoplasm. Interestingly, most above children with the 4 kb deletion came from Southeast Asia or the Far East, including Thailand, Indonesia, Burma, Micronesia, Vietnam, and Philippines.

F. DCDC2

Davis *et al.* [9] demonstrated that there is a substantial genetic component to children's ability in reading and mathematics. They found evidence that reading ability is associated with a position in DCDC2 gene, which has been implicated in neuronal development as a susceptibility gene for dyslexia [10], [11]. Another study [12] consolidated the importance of DCDC2 with one of its SNP highly associated with dyslexia.

G. KIAA0319

Dyslexia is a disorder in the acquisition of reading and writing. Müller *et al.* [12] investigated SNPs previously linked to spelling or reading ability in a German case-control cohort. They characterized 16 SNPs within five genes for functional relevance and meta-analysed them with previous studies. Three SNPs were apparently associated with dyslexia: one within DCDC2, and two within KIAA0319. In the future, other less severe SNPs in the two genes will be of interest as potential detection targets to evaluate students' language abilities.

H. CNVs

Vernes Copy number variation (CNVs) is defined as a genomics phenomenon in which some fragments of a genome are repeated and the number of repeats in the genome varies between individuals. Copy number variation is a type of deletion or duplication event that affects various lengths of DNA. Genome research indicates that approximately two thirds of the entire human genome is composed of repeats and 4.8-9.5% of the human genome can be classified as CNVs [13]. A significant proportion of children with pronounced language difficulties cannot be explained by obvious neurological and medical causes, while CNVs have not been fully established to what extent they might contribute to language disorders. Pettigrew *et al.* [14] conducted a CNVs screen in 85 young children with language-related difficulties. They detected a de novo deletion on a genome position that is near by another locus disrupted in neurodevelopmental Prader-Willi and Angelman syndromes. That was the first report of a deletion being linked to language impairment. Interestingly, CNVs restricted to the close region have been associated with reading and mathematical difficulties and general cognitive functioning [15]. Simpson *et al*. [16] performed an exploratory genome-wide CNVs study in 127 independent cases with specific language impairment (SLI), their first-degree relatives (385 individuals) and 269 population controls. They found that children with SLI and their first-degree relatives have an increased burden of moderate size CNVs (both deletions and duplications) than population controls, suggesting that CNVs may contribute to SLI risk. Bioinformatics analysis of the genes present within the CNVs identified significant overrepresentation of acetylcholine binding, cyclic-nucleotide phosphodiesterase activity and MHC proteins as compared with controls. These genes may be good targets to develop detection methods for CNVs-mediated language phenotypes.

III. Language Gene Interaction Network

Language abilities are determined by language genes and other genes that interact with them. Two or more interacting genes form a gene-combination. Students' differential language-based class-performances can be regarded as multiple-gene relied phenotypes in which one or several gene-combinations (or patterns), not a single gene, determine a specific language ability.

Worthey *et al*. [17] performed whole genome sequencing on ten randomly collected samples of CAS (childhood apraxia of speech) children and found several genes mutations, especially in gene KIAA0319 and CNTNAP2, but none mutations in FOXP2. One of the important values of the report is that some language problems are not directly connected with FOXP2, but with FOXP2-based gene interaction network.

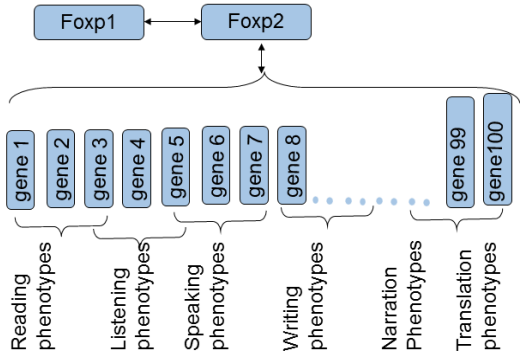

Figure 1. Foxp2 interacts with many genes that conceptually determine language-related phenotypes through different gene-combinations. Only 100 genes with strongest interaction with FOXP2 were illustrated [18].

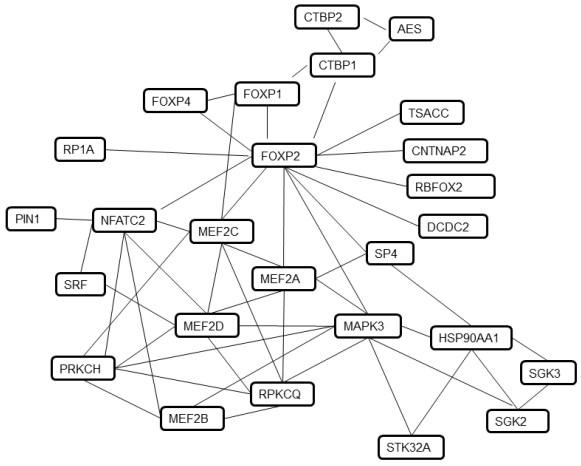

Figure 2. Physical interaction map of Foxp2 and other genes. Most data are collected from GeneCards database. All gene names and their functions can be checked out in GeneCards. Many interacting genes are not language genes but they may be involved in language ability development. Note, the functions of FOXP2 are not limited to language ability determination.

Vernes *et al.* [18] employed chromatin immunoprecipitation coupled with promoter microarrays (ChIP-chip) and successfully identified genomic sites directly bound by FOXP2 protein. They found that the promoter regions of about 303 genes have interaction with FOXP2, and 100 of them have very strong interactions. Presumably, different gene combinations among these 100 genes can contribute to different language abilities (Fig. 1), and these interactions may work as part of a large language-related molecular network (Fig. 2). In the language gene interaction network, some modules (combinations or patterns) may be more responsible for spoken and some other for written skills. Remarkably, almost every one of these genes has multiple SNPs and sequence variations, and one can imagine the potential number of the combinations among these genes is extremely large. This is the molecular basis that almost any two persons possess totally different language abilities.

In the above molecular interaction networks, the relationship between FOXP1 and FOXP2 is of special significance. FOXP1 and FOXP2 form heterodimers for transcriptional regulation on many other genes, they co-operate in common neurodevelopmental pathways through the co-regulation of common targets. Disruptions in FOXP1 have been reported in bringing autism spectrum disorder, gross motor delay and intellectual disability, while mutations in FOXP2 bring about orofacial dyspraxia, abnormalities in cortex and basal ganglia and receptive language impairment. The common phenotypes between FOXP1 and FOXP2 mutation consequences are different types of expressive language impairment [19], multiple cases of cognitive dysfunction, including intellectual disability and autism spectrum disorder, together with language impairment. The phenotypic spectra of FOXP1 and FOXP2 disruptions strongly indicate that these two interacting genes are involved in both shared and distinct neurodevelopmental pathways underlying cognitive diseases through the regulation of common and exclusive targets. So many cognitive deficits, deficiencies or disorders have more chance to originate from DNA variations of downstream interacting genes of FOXP1 and FOXP2, and direct disruptions in FOXP1 and FOXP2 are rare, since mutations in these two genes are likely linked with severe biological consequences.

TABLE I: Genes as Potential Measurement Targets

	Gene	Compromised ability (example)	Reference
1	FOXP1	Expressive language	[19]
2	FOXP2	Speech	[1]
	TPK1	Syntactic and lexical ability	[20], [21]
	ROBO1	Phonological buffer	[22], [23]
	KIAA0319	Reading, dyslexia	[24]-[27]
3	CNTNAP2	Early language development	[25], [28]-[29]
4	RBFOX2	Reading, language	[7]
	CMIP	Reading, memory	[25], [26], [30]
7	NFXL1	Speech	[31]
	ROBO2	Expressive vocabulary	[32]
	ATP2C2	Memory	[30]
	DCDC2	Reading, dyslexia	[26], [33]-[34]
8	TM4SF20	Language delay; communication disorder	[8]
9	FLNC	Reading, language	[7]
14	DYX1C1	Reading, dyslexia	[35], [36]
16	CNVs	Language	[14]-[16]

IV. Development of Techniques Applicable in Classrooms

There is a heavy task to do as characterizing language gene variations in different populations, especially different groups of students with differential language ability performance. Some known genes are listed in Table I as potential detection targets. It may take 20-30 years to fulfill the above task, and after that, every categorized language ability has its own DNA sequences as a marker. Different makers provide quantitative or semi-quantitative measurement for language ability classification. Most such measurements can be then developed as rapid, convenient and cost-effective techniques applicable in many places, including classrooms.

V. Conclusion

In this article, it is conceptualized that different language ability-related class-performance of students is largely encoded by different combinations of a cluster of language genes. Any language ability can be quantitatively or semi-quantitatively described with a group of genes, namely, the combination pattern(s) of DNA variations in a group of genes. Except for some rare disruptive mutations including deletions in some language genes, most gene variations are mild or nonsense. But aggregation of many such mild variations could lead to apparent difference in the general language ability and its performance. Simmons *et al*. [37] performed epistasis analysis using a functional coding variant in the brain-derived neurotrophic factor (BDNF) gene previously associated with reduced performance on memory tasks. Their analysis suggested that, when BDNF variation and another genomic position 13q21 susceptibility variation(s) happen together, the risk for SLI gets much higher, indicating that BDNF and 13q21 susceptibility variation(s) may be jointly part of the genetic architecture of SLI. Their analyses provide valuable insights for further cognitive neuroscience studies based on the models developed in their studies.

Acknowledgment

This work was supported in part by the National Science Foundation (No.31071170), GujingGong fund (2016) and HIT fund (ITBA10002010).

References

[1] C. S. Lai, *et al*., "A fork-head domain gene is mutated in a severe speech and language disorder," *Nature*, vol. 413, no. 6855, pp. 519-523, October 2001.
[2] A. Mozzi1, *et al*., "The evolutionary history of genes involved in spoken and written language: Beyond FOXP2, " *Sci. Rep.*, vol. 6, pp. 22157, February 2016.
[3] D. Horn, *et al*., "Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits," *Hum Mutat.*, vol. 31, no. 11, pp. 1851-1860, November 2010.
[4] F. F. Hamdan, *et al*. "De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment," *Am J. Hum Genet.*, vol. 87, no. 5, pp. 671-678, November 2010.
[5] H. Song, *et al*., "A case report of de novo missense FOXP1 mutation in a non-Caucasian patient with global developmental delay and severe speech impairment," *Clin Case Rep.*, vol. 3, no. 2, pp. 110-113, February 2015.
[6] S. C. Vernes, *et al*., "A functional genetic link between distinct developmental language disorders," *N Engl. J. Med*., vol. 359, no. 22, pp. 2337-2345, November 2008.
[7] A. Gialluisi, *et al*., "Genome-wide screening for DNA variants associated with reading and language traits," *Genes Brain Behav.*, vol. 13, no. 7, pp. 686-701, September 2014.
[8] W. Wiszniewski, *et al*., "TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities," *Am J. Hum Genet.*, vol. 93, no. 2, pp. 197-210, August 2013.
[9] O. Davis, *et al*., "The correlation between reading and mathematics ability at age twelve has a substantial genetic component," *Nat Commun*., vol. 5, pp. 4204, July 2014.
[10] J. Schumacher, *et al*., "Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia," *Am J. Hum Genet.*, vol. 78, no. 1, pp. 52-62, January 2006.
[11] T. S. Scerri, *et al*., "DCDC2, KIAA0319 and CMIP are associated with reading-related traits," *Biol Psychiatry.*, vol. 70, pp. 237-245, August 2001.
[12] B. Müller, A. Wilcke, I. Czepezauer, P. Ahnert, J. Boltze, H. Kirsten, and L. Consortium, "Association, characterisation and meta-analysis of SNPs linked to general reading ability in a German dyslexia case-control cohort," *Sci. Rep.*, vol. 6, pp. 27901, June 2016.
[13] M. Zarrei, *et al*., "A copy number variation map of the human genome," *Nat Rev Genet.*, vol. 16, no. 3, pp. 172-183, March 2015.
[14] K. A. Pettigrew, *et al*., "Copy number variation screen identifies a rare de novo deletion at chromosome 15q13.1-13.3 in a child with language impairment," *PLoS One*, vol. 10, no. 8, pp. e0134997, August 2015.
[15] H. Stefansson, *et al*., "CNVs conferring risk of autism or schizophrenia affect cognition in controls," *Nature*, vol. 505, no. 7483, pp. 361-366, January 2014.
[16] N. H. Simpson, *et al*., "Genome-wide analysis identifies a role for common copy number variants in specific language impairment," *Eur J Hum Genet.*, vol. 23, no. 10, pp. 1370-1377, October 2015.
[17] E. A. Worthey, *et al*., "Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech," *J. Neurodev Disord.*, vol. 5, no. 1, p. 29, October 2013.
[18] S. C. Vernes, *et al*., "High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders," *Am J. Hum Genet.*, vol. 81, no. 6, pp. 1232-1250, December 2007.
[19] C. Bacon and G. A. Rappold, "The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders," *Hum Genet.*, vol. 131, no. 11, pp. 1687-9168, November 2012.
[20] P. Villanueva, *et al*., "Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population," *Eur. J. Hum Genet.*, vol. 19, no. 6, pp. 687-695, June 2011.
[21] I. Fattal, *et al*., "The crucial role of thiamine in the development of syntax and lexical retrieval: A study of infantile thiamine deficiency," *Brain*, vol. 134, no. 6, pp. 1720-1739, June 2011.
[22] K. Hannula-Jouppi, *et al*., "The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia," *PLoS Genet.*, vol. p. e50, October 2005.
[23] T. C. Bates, *et al*., "Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits," *Behav Genet.*, vol. 41, pp. 50-57, January 2011.
[24] S. Paracchini, *et al*., "The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration," *Hum Mol Genet.*, vol. 15, pp. 1659-1666, May 2006.
[25] D. F. Newbury, *et al*., "Investigation of dyslexia and SLI risk variants in reading-and language-impaired subjects," *Behav Genet.*, vol. 41, no. 1, pp. 90-104, January 2011.
[26] T. S. Scerri, *et al*., "DCDC2, KIAA0319 and CMIP are associated with reading-related traits," *Biol. Psychiatry*, vol. 70, no. 3, pp. 237-245, August 2011.
[27] C. Francks, *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," *Am J. Hum Genet*., vol. 75, no. 6, pp. 1046-1058, December 2004.
[28] S. C. Vernes, *et al*., "A functional genetic link between distinct developmental language disorders," *N Engl J Med*., vol. 359, no. 22, pp. 2337-2345, November 2008.
[29] A. J. Whitehouse, *et al*., "CNTNAP2 variants affect early language development in the general population," *Genes Brain Behav.*, vol. 10, no. 4, pp. 451-456, June 2011.
[30] D. F. Newbury, *et al*., "CMIP and ATP2C2 modulate phonological short-term memory in language impairment," *Am J. Hum Genet*., vol. 85, no. 2, pp. 264-272, August 2009.
[31] P. Villanueva, *et al*., "Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment," *PLoS Genet*., vol. 11, no. 3, pp. e1004925, March 2015.
[32] B. St Pourcain, *et al*., "Common variation near ROBO2 is associated with expressive vocabulary in infancy," *Nat Commun*., vol. 5, p. 4831, September 2014.
[33] K. E. Deffenbacher, *et al*., "Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: Linkage and association analyses," *Hum. Genet*., vol. 115, no. 2, pp. 128-138, July 2004.
[34] J. Schumacher, *et al*., "Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia," *Am J. Hum Genet*., vol. 78, no. 1, pp. 52-62, January 2006.
[35] M. Taipale, *et al*., "A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain," *Proc Natl Acad. Sci. USA*, vol. 100, no. 20, pp. 11553-11558, September 2003.
[36] S. Paracchini, *et al*., "Analysis of dyslexia candidate genes in the Raine cohort representing the general Australian population," *Genes Brain Behav*., vol. 10, no. 2, pp. 158-165, March 2011.
[37] T. R. Simmons, *et al*., "Increasing genotype-phenotype model determinism: Application to bivariate reading/language traits and epistatic interactions in language-impaired families," *Hum Hered*., vol. 70, no. 4, pp. 232-244, 2010.

Wei Xia is interested in ESL learning and teaching theory/practice, language gene sequence information and language ability, language gene pattern recognition in student populations, and language gene-based education regime.

Zhizhou Zhang is a professor of Molecular Biology at Harbin Institute of Technology, China. He received his B.S. degree in Molecular Biology from the University of Science and Technology of China, and his Ph.D. degree in Biochemistry and Molecular Biology from Medical College of Ohio (now part of University of Toledo), USA. His research interests include gene manipulation technology, nanotechnology, language gene characterization and Marine science.

