Association of the DYX1C1 Gene with Chinese Literacy in a Healthy Chinese Population by Latent Class and LASSO Analyses

Mary Miu Yee Waye
The Nethersole School of Nursing,
The Chinese University of Hong Kong
Mary-waye@cuhk.edu.hk

Cynthia O. Siu
COS and Associates Ltd.
csiu2004@gmail.com

Catherine McBride
Dept. of Psychology,
The Chinese University of Hong Kong
cmcbride@psy.cuhk.edu.hk

Connie Suk-han Ho
Dept. of Psychology,
The University of Hong Kong
shhoc@hku.hk

Cheuk Wa Wong
School of Biomedical Sciences,
The Chinese University of HK
allanwalker1@yahoo.com.hk

Abstract

DYX1C1, the first dyslexia candidate gene, has been associated with developmental dyslexia in different populations, but its influence on reading abilities in the general population is less well known. Copy number variants (CNVs) have been implicated in neurodevelopmental and childhood-onset disorders involving cognitive development in previous studies. In this report, we investigated the extent to which genomic CNVs for the SNP previously linked to dyslexia, -3G/A (rs3743205) in the gene DYX1C1, contribute to Chinese and English literacy in the general population in a Chinese cohort, and whether these processes, in turn, are influenced by environmental factors, such as family income, parents' education, and IQ. Our findings suggest that the logR ratio (which is a way to detect CNVs) of a previously reported dyslexia-related SNP, -3G/A (rs3743205) is significantly associated with Chinese literacy in a cohort of Chinese children with normal reading abilities.

1. Introduction

Reading abilities could be affected in children with normal intellectual abilities. The reason behind developmental dyslexia has perplexed neuroscientists for decades. With the discovery of the first dyslexia candidate gene DYX1C1, it was hoped that this could help us better understand reading difficulties and serve as a bridge to enhance reading abilities in the general population [1], [2]. The original report implicating DYX1C1 (including SNP rs3743205) in the etiology of dyslexia was found in a study of Finnish dyslexia family with a child with dyslexia by breakpoint mapping [3], with the gene located in a loci on chromosome 15q21.3 [4]. Another study confirmed that DYX1C1 influences reading and spelling ability with additional effects on short-term information storage or rehearsal, but rs3743205 was not associated with dyslexia as reported by Taipale et al; the reason could be that this SNP has a very low minor allele frequency in that sample (0.002) [5]. Since then, many more candidate genes have been reported and there are genes that seems to be affecting different populations, e.g. association of DYX1C1 with German [6], Chinese [7]–[9], Australian [5], [10], and Swiss children (in the NeuroDys Consortium including children from Austria, France, Germany, The Netherlands, Switzerland, Finland, Hungary, and the United Kingdom [11], while others are not so clearly universal [12].

DYX1C1 was shown to be one of the dyslexia candidate genes that was associated with white matter

URI: http://hdl.handle.net/10125/50243
ISBN: 978-0-9981331-1-9

volume in the left temporoparietal region which affects reading ability [13]. However this gene's effect on brain volume is unclear since the knockdown of Dyx1c1 (through RNA interference, RNAi, in rats) did not affect the cortical volume, hippocampal volume, or midsagittal area of the corpus callosum in rats, even though knockdown rats has long-term deficits in their visual attention abilities, in addition to rapid and complex auditory processing abilities [14].

The mechanism of DYX1C1 as a dyslexia susceptibility gene was postulated to be due to its role in neuronal migration mediated by its ciliary function as a dynein axonemal assembly factor [15]–[18], as well as its possible interactions with the U-box protein CHIP (carboxy terminus of Hsc70-interacting protein) which is involved in the degradation of estrogen receptors alpha (ERalpha) and beta (ERbeta). These protein interaction complexes have been demonstrated along the neurites of primary rat hippocampal neurons, and is thought to affect brain development and regulate cognitive functions [19]. Supporting this idea, neuroimaging results also revealed that there was both a main effect of the gene and a gene × age interaction, possibly not only reflecting the participation of DYX1C1 in early brain development but also in pathways that could affect later myelination at both time points of measurement, separated by 2 years [20]. The interaction of DYX1C1 with estrogen receptors (and thus implicating its role in the sex hormone signaling pathway) is also supported by the fact that two of the polymorphisms studied by Darki's group, rs3743204 and rs3743205, are part of a haplotype of three SNPs associated specifically with female dyslexic subjects in a sample of 366 German trios [6].

Although DYX1C1 has been shown to be associated with developmental dyslexia, the role of this gene in reading development in the general population without dyslexia is not clear. There was a study showing that in a group 248 children in Beijing, there was marginal association of certain SNPs of DYX1C1 (rs11629841) with children's orthographic judgements for children at age 7 and 8 but less so at age 6, and association of this SNP was found with Chinese character dictation at ages 9, 10 and 11 years, but no significant association with other SNPs of DYX1C1 they have studied (rs3743205 or rs57809907) [8]. The role of environmental factors in literacy skills has also been studied previously [21], [22].

2. Materials and Method

2.1. Participants

Informed consent was given to the participant's parent prior to testing. Depending upon the parent's preference, the children were tested either in a laboratory setting, their school, or their home by trained experimenters. The test took approximately 2 hours; five-minute breaks were given every 30 minutes of testing. A performance report and $720 HKD incentive were given for participation.

2.2. Inclusion Criteria

Ninety-six Hong Kong Chinese-English bilingual children (Age M=7.6 years, SD=1.17, Grades 1-4; 65 females) were included in the analysis. They were selected from a fellow classmate of pairs of twins in Hong Kong involved in another study: a twin study with the following criteria: They were not diagnosed as having any special educational needs, including developmental dyslexia, hearing or visual impairment, physical or intellectual disability, speech and language impairment, attention deficit/hyperactivity disorder, or autistic spectrum disorders.

2.3. Phenotypes

Cognitive-linguistic and literacy measures, including non-verbal intelligence, rapid digit naming, phonological awareness, lexical decision, morphological construction, vocabulary knowledge, pure copying, delayed copying, and spelling were given to the participants as described previously [22]. For rapid digit naming, the child will be asked to name 8 rows of 5 digits given to them rapidly. Two trials were given and the time averaged. For phonological awareness, children were required to: 1). take away one syllable from three-syllable words or pseudowords and 2). say a one to three syllable word or pseudoword

aloud, but without saying the onset. For lexical decision, the children were asked to distinguish real words from pseudowords. For discourse skills, the children were asked to arrange the sequence of several sentences related to a story correctly. For morphological construction, children were asked to construct new compounds/words for the presented concepts based on previously acquired morphemes. For vocabulary knowledge, children were tested on their receptive vocabulary, expressive vocabulary and vocabulary definitions. For reading comprehension, the children were asked to respond to some questions orally after reading 3 passages. For one minute word reading, the children were asked to read as many of the 90 words given in one minute. For Chinese word reading, they were asked to read as many of the words until they fail 15 words in a row. For English word reading, they were asked to read until they fail 4 words in a row. For pure copying, children were asked to copy some unfamiliar objects. For delayed copying, the children were asked to copy a word that disappears after 2 seconds using powerpoint slides. For spelling, children were asked to spell some pre-recorded words.

2.4. Genotypes

Saliva samples were collected from each participant using the Oragene DNA collection kits from DNA Genotek (DNA Genotek, 2016) and DNA was extracted according to the manufacturer's instructions. Microarray analysis of the DNA was carried out at the Li Ka Shing Institute of Health Sciences, Chinese University of Hong Kong, located in the Prince of Wales Hospital in Hong Kong, using Illumina's Infinium OmniZhongHua-8 v1.1, according to the manufacturer's protocol (Illumina Genotyping, 2016).

2.5. Statistical Analyses

Data were analyzed using Infinium OmniZhongHua-8 v1.3 (Illumina Inc., San Diego, CA). CNV detection was based on two metrics: B allele frequency and log R ratio. The log R ratio, a measure of normalized total signal intensity, was defined as the log (base 2) ratio of the observed normalized R value for the SNP, divided by the expected normalized R value for the SNP, where expected normalized R was computed from a linear interpolation of canonical genotype clusters [23]. R is the intensity of dyelabeled molecules that have hybridized to the beads on the array and with potentially different patterns of CNV influences on literacy. For each SNP, raw signal intensity values were measured for the A and B alleles and subject to a five-step normalization procedure using the signal intensity of all SNPs (see Illumina white paper at https://icom.illumina.com/icom/software.ilmn). This procedure produces the X and Y values for each SNP. The total signal intensity R = X + Y, where the X and Y values for each SNP represent the experiment-wide normalized signal intensity on the A and B alleles, respectively. Composite literacy measures were defined by combined "word reading" and "dictation" skills in Chinese and English languages.

In the LASSO (Least Absolute Shrinkage and Selection Operator) network model, the nodes represent cognitive-linguistic and literacy measures (phenotypes) and edges (lines connecting phenotype nodes) represent significant pairwise partial correlations (conditional independence relationships between 2 nodes) estimated by LASSO regression method [24], [25], with FDR (false discovery rate) adjusted P of 0.05. LASSO performs both Extended Bayesian Information Criterion (EBIC) model selection and regularization. In order to enhance interpretability of the network model and avoid overfitting, we applied LASSO in statistical programming language R package qgraph [26] to shrink all edges in the network of connected nodes and assign 0 to small edges so that the covariance structure of all nodes can be explained by the smallest number of edges in the final parsimonious network model.

Factor Mixture Model (FMM) was applied to model a mixture of subgroups (latent classes LC) within the study sample. The relationships between CNVs and language skills were evaluated using regression models. The objective of the analysis was to investigate whether the latent classes which represented the sociodemographic factors moderated the relationships between CNV and language measures.

The potential impact of latent class (LC) on the relationships between CNVs and language skills was assessed by including LC and interaction term for LC x language skills in the model. An exploratory analysis of the relationships in the sociodemographic characteristics, neonatal and infancy history, behavior

and language skills was conducted using a LASSO network model.

3. Results

3.1. Phenotype analyses

Figure 1 showed the network clustering patterns of the sociodemographic characteristics, neonatal and infantry history, the behavior and language skills. Chinese language skills were closely associated with neonatal and infancy history with short network path (proportional to the strength of correlation between variables), while English language skills had direct significant correlations with family socioeconomic status. Based on the LASSO optimal model, there are significant relationships between the Chinese and English language skills and their relationships with the sociodemographic features of the participants as well as some environmental factors. Environmental factors such as parents' education levels and income had a significantly greater influence on the English language skills (literacy, morphological awareness and vocabulary) compared to Chinese language skills.

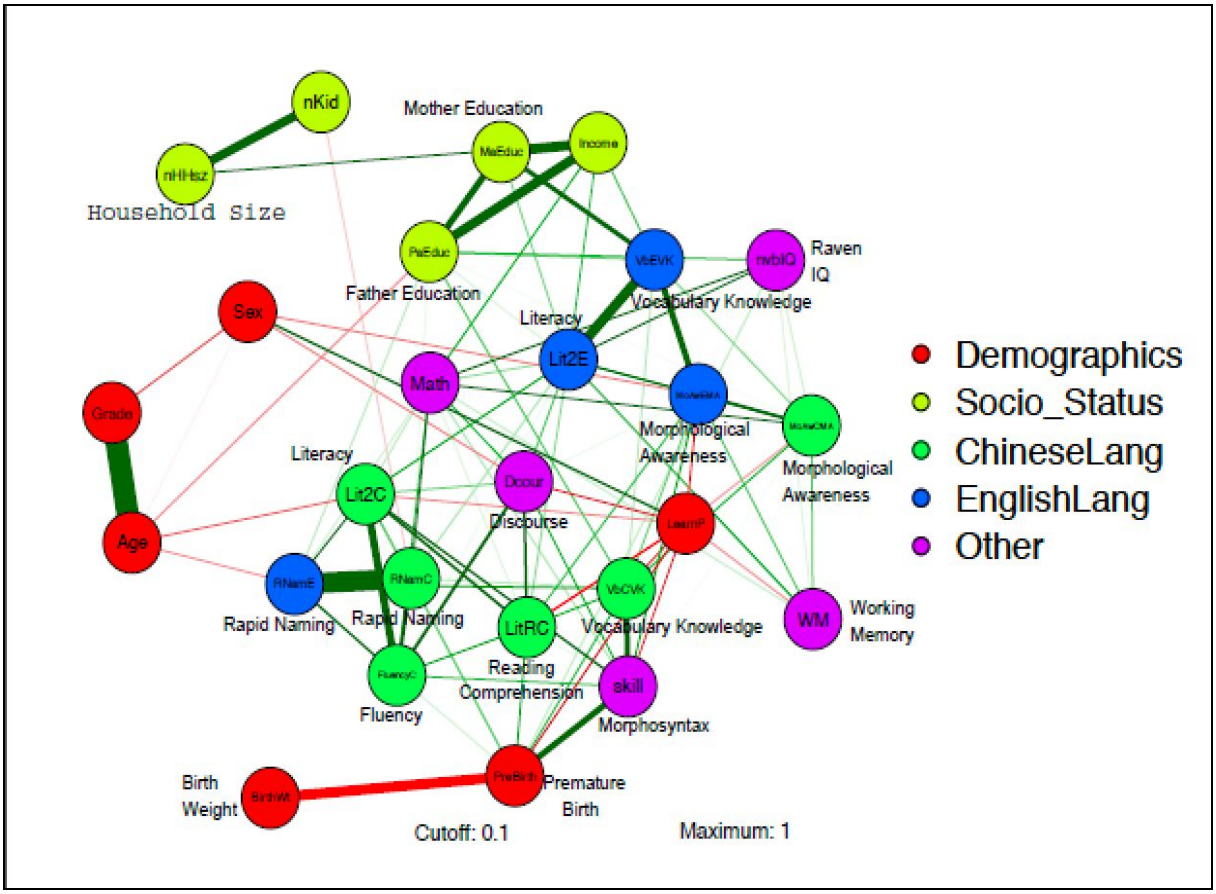

Figure 1. Network Clustering Patterns of Language Skills and Sociodemographic Characteristics

3.2. Latent Class Analysis

We performed a latent class analysis using FMM approach to assess the heterogeneity of the study sample. FMM analysis identified 2 latent classes (LC) which were comparable in age, gender, and grade, but LC1 (N=71) had

significantly higher family socio-economic status (family income, father education, mother education, and IQ) (all P's < 0.05), better Chinese and English literacy and language skill measure (all P's < 0.05) compared to LC2 (N=24).

	Latent Class 1 (n=71)	Latent Class 2 (n=24)	P-value
Age (months)	96.9 (SD=1.7)	101.3 (SD=3.1)	0.189
Gender Male Female	47.9% 52.1%	54.2% 45.8%	0.642
Grade 1 2 3 4	21 (29.6%) 22 (31.0%) 14 (19.7%) 14 (19.7%)	6 (25.0%) 8 (33.3%) 5 (20.8%) 5 (20.8%)	0.761
Income <= 25000 (HK$) > 25000	23 (32.4%) 48 (67.6%)	17 (70.8%) 7 (29.2%)	<0.001
Father Education Form 5 or below Post Secondary University	 19 (26.8%) 22 (31.0 %) 30 (42.2%)	 14 (58.3%) 7 (29.2 %) 3 (12.5%)	0.002
Mother Education Form 5 or below Post Secondary University	 20 (28.2 %) 25 (35.2 %) 26 (36.6 %)	 16 (66.7 %) 6 (25 %) 2 (8.3 %)	<0.001
IQ Raven <=100 101 – 115 >115	9 (12.7%) 28 (39.4%) 34 (47.9%)	10 (41.7%) 10 (41.7%) 4 (16.7%)	0.003

Table 1. Comparison of Latent Classes. † One subject had genotype data. but missing data on phenotypes

3.3. Quantitative trait association study

The SNP rs3743205 has low minor allele frequency (~0 MAF) and extreme homozygosity (AA 0, AB 2.1%, BB 97.9%). Therefore LogR ratio was used as a proxy for copy number measurement. Table 2 summarizes the results for the LC-by-Language interaction effects which test the significance of differences between the 2 latent classes (LC 1 and LC 2) in the associations between CNV (log R) and language measures. We also provide the adjusted effect size (regression slope of language measure on log R within each latent class), showing the magnitude of association between CNV (log R) and language measure within the latent classes.

Associations of CNVs with Chinese literacy measures (CWR and CDICT) were statistically significant (P<0.05), while associations of CNVs with English literacy measures (EWR and EDICT) were statistically non-significant (P>0.05). No significant associations were found in all other measures for rs3743205. There was a significant latent class effect on moderating the relationship of CNV with Chinese literacy measure (P<0.05), with log R ratio increased towards 0 as Chinese literacy improved in Latent Class 1. The association between log R ratio and Chinese Literacy was, however, non-significant in Latent Class 2. A significant adjusted association was found between Log R ratio of the SNP rs3743205 of DYX1C1 and 2 of the 3 Chinese literacy measures (NS in RC) in a

subpopulation of subjects (Latent class 1) which differs from Latent class 2 in almost all sociodemographic and language measures. Latent class 2 has lower performance in all language ability measures compared to Latent class 1.

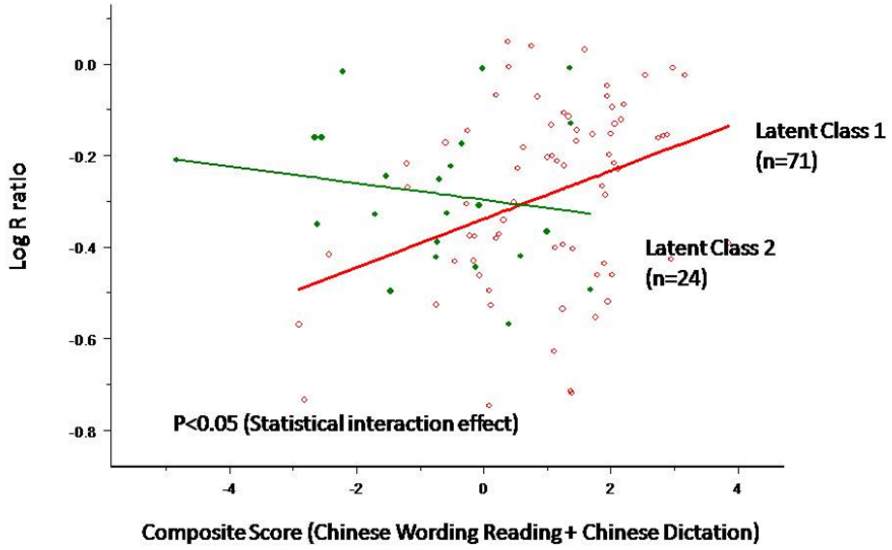

Figure 2. Associations of a Composite Score of Language Measures with Log R Ratio (indirect measure of copy number of SNP) for SNP rs3743205 (AA 0; AB 2%; BB 97.8%; C>T)

			Adjusted Associations		Adjusted Associations	
			Latent Class 1 (n=71)		Latent Class 2 (n= 24)	
		Interaction P-value	Effect Size	Nominal P-value	Effect Size	Nominal P-value
Chinese Literacy	Chinese Word Reading (CWR Norm Score)	0.190	0.40	0.022	-0.07	0.708
	Chinese Dictation (CDICT)	0.130	0.54	0.0009	0.115	0.610
	Chinese Reading Comprehension (RC)	0.494	0.013	NS	0.046	NS
	Chinese Word Reading + Dictation Composite Score	0.038	0.29	0.0018	<0.01	0.991
English Literacy	English Word Reading (EWR)	0.207	0.044	NS	-0.024	NS

	English Dictation (EDICT)	**0.677**	**0.021**	**NS**	**-0.002**	**NS**
	English Word Reading + Dictation composite Score	**0.382**	**0.018**	**NS**	**-0.007**	**NS**
Fluency	**Chinese One Minute Word Reading (COM Norm Score)**	**0.806**	**0.014**	**NS**	**0.002**	**NS**
Vocabulary	**Chinese Vocabulary Knowledge**	**0.899**	**0.021**	**NS**	**0.028**	**NS**
	English Vocabulary Knowledge	**0.802**	**0.017**	**NS**	**0.003**	**MS**

Table 2: Associations of 3 Language Measures with Log R Ratio for SNP rs3743205
† One subject had genotype data but missing data on phenotypes. NS: P > 0.05.

4. Discussions

Our results shown that there are interactive relationships between the Chinese and English language skills and the sociodemographic features of the participants. Our results are consistent with the latent class regression analyses done earlier with a group of kindergarten children, and the findings that there was a higher frequency of phonological awareness (PA) and rapid automatized naming (RAN) deficits in children from lower socioeconomic status backgrounds [27]. The relationships between the language skills with environmental factors is consistent with previous reports that the presence of Filipino helpers in the households of Hong Kong children tend to associate with better English skills of the children [21]. Our results carry important implications for improvement of genomic analyses in studies of dyslexia or normal reading and writing skills, based on stratification of latent classes.

Here we found that a polymorphism in DYX1C1 previously reported to be associated with developmental dyslexia is also correlated with two of the three Chinese literacy measures we studied in a normal cohort of Chinese children. The results will require replication in an independent population since our sample size is small for a genetic study.

Log-R-ratio has been previously used to identify mosaic autosomal chromosomal abnormalities including copy number variation and loss of heterozygosity. For example, a study of patients with Log-R-ratio ad electronic medical records shown that a third of those with chromosomal abnormalities had hematologic abnormalities while a much lower rate of hematologic malignancies (2% vs 24%, P<0.001) was observed for the normal age- and sex-matched controls [28]. Using tumour samples versus blood samples from breast cancer patients, log-R ratio was used to compare the disparity from heterozygous toward either the A/B allele homozygous (allelic disparity) and association with or B allele homozygous (allelic disparity) states, and there was an association between levels of disparity in SNPs and clinical and tumour-related parameters [29]. Detection of 18 *de novo* rare CNV deletions (rDELs) was possible by using the Log-R ratio intensity data from a dense single nucleotide polymorphism (SNP) array of 6808 type 1 diabetes patients, and 9,954 controls. These rDELs were detected in affected offspring but none for the unaffected siblings (P=0.03) [30]. Log R ratio was also used to demonstrate that EPHA3 deletion was a potential protective factor for amyotrophic lateral sclerosis (ALS) [31]. Log R ratio

was also found to be an advantageous method (rather than relying on calling algorithms) in Illumina 1 M data as a continuous measure to detect their association of CNVs with bladder cancer using simple regression models [32]. Software such as VegaMC can also convert allele signal intensities in log R ratio and B allele frequency, to enable the detection of loss of heterozygosity and was used to demonstrate the ability of detecting aberrant genes for colon adenocarcinoma and glioblastoma multiforme from the Cancer Genome Atlas (TCGA) project [33].

In our study, we used the Log-R ratio intensity data from a dense single nucleotide polymorphism (SNP) array to detect rare CNV deletions in children with detailed assessment of their reading and writing skills. A linear regression was used to associate the Log R ratio for a SNP from the DYX1C1 gene that is known to be associated with developmental dyslexia with Chinese literacy measurements. Chinese dictation and Chinese word reading scores showed significant association with the Log-R ratio whereas Chinese Reading Comprehension scores did not have significant association. Thus we also performed association with a composite score for both Chinese dictation and Chinese word reading scores and found a significant association ($p<0.05$).

It appears that DYX1C1 may be associated with decoding skills but not text-level processing. Latent class 2 may be a group of weaker readers with other genes playing a more prominent role affecting their literacy skills, while DYX1C may be associated with decoding only and have no problem with text-level processing or comprehension.

Reading comprehension is an intentional, active, interactive process during reading which involves a complex array of cognitive processes including understanding phonemes, phonics and the ability to comprehend or construct meaning from the text. The reason that reading comprehension is not significantly associated with Log-R ratio of DYX1C1 in our study, but only Chinese word reading and dictation are associated, (not English word reading and diction) may be due to the fact that they involve different cognitive processes in different parts of the brain. This is reflected from previous fMRI data: Reading comprehension was associated with more left-lateralized activation and with left inferior occipital cortex (including fusiform gyrus) activation [34]. On the other hand, Chinese character recognition relies more on ventral visual-spatial processing than English word recognition [35]. Second language (i.e. English) word reading may cause more right hemisphere activation than native language (e.g. Chinese) [36], [37], and the left caudate and fusiform regions are more important in reading skills in a second language [38].

Although the integrity of DNA could vary due to sample storage and processing issues and such variation could lead to variance of log-R ratio as shown by pathology archives of clinical material in the form of formalin-fixed paraffin-embedded (FFPE) tissue samples [39], we do not think the log-R association observed is due to sample quality issues due to our vigilance in sample storage using high quality collection kits (Oragene) that is well tested, and our prompt processing of the saliva samples as soon as they are collected by a research assistant dedicated to this project, and also in light of the fact that the association is only observed in latent class 2 and not in latent class 1 samples. From the value of log-R ratio we observed (lower than 0), this region of DYX1C1 appeared to be altered by deletion rather than by amplification. The extent of these small deletions is not well understood but up to 4,989 small scale copy number variations of the Korean HapMap samples have been described in 79 samples, suggesting that the human genome structure is a lot more complicated than previously thought [40].

5. Conclusion

We found association between the log-R ratio of DYX1C1 and two measures of Chinese literacy skills in a cohort of Hong Kong Chinese children with normal reading abilities. Whether there is selective pressure for this loci of DYX1C1 to be preferentially deleted remains to be studied. Nevertheless, early identification of genetic factors and environmental factors should be useful for smart decision in implementation of early interventions. The methodology described in this paper for visualization of factors affecting language abilities of healthy children can be also used for other big data sets related to other social or health care needs.

6. Acknowledgements

This work was supported by the Collaborative Research Funds of the Research Grants Council of Hong Kong (CUHK8/CRF/13G) and LCWIIM-BRAIN grant (#8303402). We thank all the participants and the central core facility of the Faculty of Medicine of CUHK for the genotyping studies. The authors report no financial interests or potential conflicts of interest.

7. References:

[1] S. E. Fisher and C. Francks, "Genes, cognition and dyslexia: learning to read the genome.," *Trends Cogn Sci (Regul Ed)*, vol. 10, no. 6, pp. 250–257, Jun. 2006.

[2] S. Paracchini, T. Scerri, and A. P. Monaco, "The genetic lexicon of dyslexia.," *Annu Rev Genomics Hum Genet*, vol. 8, pp. 57–79, 2007.

[3] M. Taipale, N. Kaminen, J. Nopola-Hemmi, T. Haltia, B. Myllyluoma, H. Lyytinen, K. Muller, M. Kaaranen, P. J. Lindsberg, K. Hannula-Jouppi, and J. Kere, "A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain.," *Proc Natl Acad Sci U S A*, vol. 100, no. 20, pp. 11553–11558, Sep. 2003.

[4] J. Nopola-Hemmi, M. Taipale, T. Haltia, A. E. Lehesjoki, A. Voutilainen, and J. Kere, "Two translocations of chromosome 15q associated with dyslexia.," *J Med Genet*, vol. 37, no. 10, pp. 771–775, Oct. 2000.

[5] T. C. Bates, P. A. Lind, M. Luciano, G. W. Montgomery, N. G. Martin, and M. J. Wright, "Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation.," *Mol Psychiatry*, vol. 15, no. 12, pp. 1190–1196, Dec. 2010.

[6] F. Dahdouh, H. Anthoni, I. Tapia-Páez, M. Peyrard-Janvid, G. Schulte-Körne, A. Warnke, H. Remschmidt, A. Ziegler, J. Kere, B. Müller-Myhsok, M. M. Nöthen, J. Schumacher, and M. Zucchelli, "Further evidence for DYX1C1 as a susceptibility factor for dyslexia.," *Psychiatr Genet*, vol. 19, no. 2, pp. 59–63, Apr. 2009.

[7] C. K. P. Lim, C. S. H. Ho, C. H. N. Chou, and M. M. Y. Waye, "Association of the rs3743205 variant of DYX1C1 with dyslexia in Chinese children.," *Behav Brain Funct*, vol. 7, p. 16, May 2011.

[8] Y. Zhang, J. Li, T. Tardif, M. Burmeister, S. M. Villafuerte, C. McBride-Chang, H. Li, B. Shi, W. Liang, Z. Zhang, and H. Shu, "Association of the DYX1C1 dyslexia susceptibility gene with orthography in the Chinese population.," *PLoS ONE*, vol. 7, no. 9, p. e42969, Sep. 2012.

[9] Z. Wang, G. Cui, A. Zhao, P. Li, D. Liu, L. Shen, and D. Li, "[Genome-wide association studies on the developmental dyslexia children].," *Wei Sheng Yan Jiu*, vol. 44, no. 5, pp. 767–70, 779, Sep. 2015.

[10] S. Paracchini, Q. W. Ang, F. J. Stanley, A. P. Monaco, C. E. Pennell, and A. J. O. Whitehouse, "Analysis of dyslexia candidate genes in the Raine cohort representing the general Australian population.," *Genes Brain Behav*, vol. 10, no. 2, pp. 158–165, Mar. 2011.

[11] J. Becker, D. Czamara, T. S. Scerri, F. Ramus, V. Csépe, J. B. Talcott, J. Stein, A. Morris, K. U. Ludwig, P. Hoffmann, F. Honbolygó, D. Tóth, F. Fauchereau, C. Bogliotti, S. Iannuzzi, Y. Chaix, S. Valdois, C. Billard, F. George, I. Soares-Boucaud, and J. Schumacher, "Genetic analysis of dyslexia candidate genes in the European cross-linguistic NeuroDys cohort.," *Eur J Hum Genet*, vol. 22, no. 5, pp. 675–680, May 2014.

[12] A. Carrion-Castillo, B. Maassen, B. Franke, A. Heister, M. Naber, A. van der Leij, C. Francks, and S. E. Fisher, "Association analysis of dyslexia candidate genes in a Dutch longitudinal sample.," *Eur J Hum Genet*, vol. 25, no. 4, pp. 452–460, Jan. 2017.

[13] F. Darki, M. Peyrard-Janvid, H. Matsson, J. Kere, and T. Klingberg, "Three dyslexia susceptibility genes, DYX1C1, DCDC2, and KIAA0319, affect temporo-parietal white matter structure.," *Biol Psychiatry*, vol. 72, no. 8, pp. 671–676, Oct. 2012.

[14] C. E. Szalkowski, A. B. Booker, D. T. Truong, S. W. Threlkeld, G. D. Rosen, and R. H. Fitch, "Knockdown of the candidate dyslexia susceptibility gene homolog dyx1c1 in rodents: effects on auditory processing, visual attention, and cortical and thalamic anatomy.," *Dev Neurosci*, vol. 35, no. 1, pp. 50–68, Apr. 2013.

[15] G. D. Rosen, J. Bai, Y. Wang, C. G. Fiondella, S. W. Threlkeld, J. J. LoTurco, and A. M. Galaburda, "Disruption of neuronal migration by RNAi of Dyx1c1 results in neocortical and hippocampal malformations.," *Cereb Cortex*, vol. 17, no. 11, pp. 2562–2572, Nov. 2007.

[16] A. Tarkar, N. T. Loges, C. E. Slagle, R. Francis, G. W. Dougherty, J. V. Tamayo, B. Shook, M. Cantino, D. Schwartz, C. Jahnke, H. Olbrich, C. Werner, J. Raidt, P. Pennekamp, M. Abouhamed, R. Hjeij, G. Köhler, M. Griese, Y. Li, K. Lemke, and H. Omran, "DYX1C1 is required for axonemal dynein assembly and ciliary motility.," *Nat Genet*, vol. 45, no. 9, pp. 995–1003, Sep. 2013.

[17] Y. Wang, M. Paramasivam, A. Thomas, J. Bai, N. Kaminen-Ahola, J. Kere, J. Voskuil, G. D. Rosen, A. M. Galaburda, and J. J. Loturco, "DYX1C1 functions in neuronal migration in developing neocortex.," *Neuroscience*, vol. 143, no. 2, pp. 515–522, Dec. 2006.

[18] L. A. Gabel, C. J. Gibson, J. R. Gruen, and J. J. LoTurco, "Progress towards a cellular neurobiology of reading disability.," *Neurobiol Dis*, vol. 38, no. 2, pp. 173–180, May 2010.

[19] S. Massinen, K. Tammimies, I. Tapia-Páez, H. Matsson, M.-E. Hokkanen, O. Söderberg, U. Landegren, E. Castrén, J.-A. Gustafsson, E. Treuter,

and J. Kere, “Functional interaction of DYX1C1 with estrogen receptors suggests involvement of hormonal pathways in dyslexia.,” *Hum Mol Genet*, vol. 18, no. 15, pp. 2802–2812, Aug. 2009.

[20] F. Darki, F. Nemmi, A. Möller, R. Sitnikov, and T. Klingberg, “Quantitative susceptibility mapping of striatum in children and adults, and its association with working memory performance.,” *Neuroimage*, vol. 136, pp. 208–214, Aug. 2016.

[21] E. C. Crebo, “Filipina domestic helpers in Hong Kong and their role in English language learning,” *Thesis*, 2003.

[22] X. Tong, J. C. M. Lo, C. McBride, C. S.-H. Ho, M. M. Y. Waye, K. K. H. Chung, S. W. L. Wong, and B. W.-Y. Chow, “Coarse and fine N1 tuning for print in younger and older Chinese children: Orthography, phonology, or semantics driven?,” *Neuropsychologia*, vol. 91, pp. 109–119, Oct. 2016.

[23] D. A. Peiffer, J. M. Le, F. J. Steemers, W. Chang, T. Jenniges, F. Garcia, K. Haden, J. Li, C. A. Shaw, J. Belmont, S. W. Cheung, R. M. Shen, D. L. Barker, and K. L. Gunderson, “High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping.,” *Genome Res*, vol. 16, no. 9, pp. 1136–1148, Sep. 2006.

[24] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (* ..., 1996.

[25] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso.,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008.

[26] S. Epskamp, A. O. J. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom, “qgraph : network visualizations of relationships in psychometric data,” *J Stat Softw*, vol. 48, no. 4, 2012.

[27] O. Ozernov-Palchik, E. S. Norton, G. Sideridis, S. D. Beach, M. Wolf, J. D. E. Gabrieli, and N. Gaab, “Longitudinal stability of pre-reading skill profiles of kindergarten children: implications for early screening and theories of reading.,” *Dev Sci*, Oct. 2016.

[28] H. Jouni, K. Shameer, Y. W. Asmann, R. Hazin, M. de Andrade, and I. J. Kullo, “Clinical Correlates of Autosomal Chromosomal Abnormalities in an Electronic Medical Record-Linked Genome-Wide Association Study: A Case Series.,” *J Investig Med*, vol. 1, no. 4, 2013.

[29] F. Kaveh, H. Edvardsen, A.-L. Børresen-Dale, V. N Kristensen, and H. K. Solvang, “Allele-specific disparity in breast cancer.,” *BMC Med Genomics*, vol. 4, p. 85, Dec. 2011.

[30] N. J. Cooper, C. J. Shtir, D. J. Smyth, H. Guo, A. D. Swafford, M. Zanda, M. E. Hurles, N. M. Walker, V. Plagnol, J. D. Cooper, J. M. M. Howson, O. S. Burren, S. Onengut-Gumuscu, S. S. Rich, and J. A. Todd, “Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes.,” *Hum Mol Genet*, vol. 24, no. 6, pp. 1774–1790, Mar. 2015.

[31] Ö. Uyan, Ö. Ömür, Z. S. Ağım, A. Özoğuz, H. Li, Y. Parman, F. Deymeer, P. Oflazer, F. Koç, E. Tan, H. Özçelik, and A. N. Başak, “Genome-wide copy number variation in sporadic amyotrophic lateral sclerosis in the Turkish population: deletion of EPHA3 is a possible protective factor.,” *PLoS ONE*, vol. 8, no. 8, p. e72381, Aug. 2013.

[32] G. Marenne, F. X. Real, N. Rothman, B. Rodríguez-Santiago, L. Pérez-Jurado, M. Kogevinas, M. García-Closas, D. T. Silverman, S. J. Chanock, E. Génin, and N. Malats, “Genome-wide CNV analysis replicates the association between GSTM1 deletion and bladder cancer: a support for using continuous measurement from SNP-array data.,” *BMC Genomics*, vol. 13, p. 326, Jul. 2012.

[33] S. Morganella and M. Ceccarelli, “VegaMC: a R/bioconductor package for fast downstream analysis of large array comparative genomic hybridization datasets.,” *Bioinformatics*, vol. 28, no. 19, pp. 2512–2514, Oct. 2012.

[34] A. Buchweitz, R. A. Mason, L. M. B. Tomitch, and M. A. Just, “Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension.,” *Psychol Neurosci*, vol. 2, no. 2, pp. 111–123, 2009.

[35] Y. Sun, Y. Yang, A. S. Desroches, L. Liu, and D. Peng, “The role of the ventral and dorsal pathways in reading Chinese characters and English words.,” *Brain Lang*, vol. 119, no. 2, pp. 80–88, Nov. 2011.

[36] G. Ding, C. Perry, D. Peng, L. Ma, D. Li, S. Xu, Q. Luo, D. Xu, and J. Yang, “Neural mechanisms underlying semantic and orthographic processing in Chinese-English bilinguals.,” *Neuroreport*, vol. 14, no. 12, pp. 1557–1562, Aug. 2003.

[37] L. H. Tan, J. A. Spinks, C.-M. Feng, W. T. Siok, C. A. Perfetti, J. Xiong, P. T. Fox, and J.-H. Gao, “Neural systems of second language reading are shaped by native language.,” *Hum Brain Mapp*, vol. 18, no. 3, pp. 158–166, Mar. 2003.

[38] L. H. Tan, L. Chen, V. Yip, A. H. D. Chan, J. Yang, J.-H. Gao, and W. T. Siok, “Activity levels in the left hemisphere caudate-fusiform circuit predict how well a second language will be learned.,” *Proc Natl Acad Sci U S A*, vol. 108, no. 6, pp. 2540–2544, Feb. 2011.

[39] A. N. Hosein, S. Song, A. E. McCart Reed, J. Jayanthan, L. E. Reid, J. R. Kutasovic, M. C. Cummings, N. Waddell, S. R. Lakhani, G. Chenevix-Trench, and P. T. Simpson, “Evaluating the repair of DNA derived from formalin-fixed paraffin-embedded tissues prior to genomic profiling by SNP-CGH analysis.,” *Lab Invest*, vol. 93, no. 6, pp. 701–710, Jun. 2013.

[40] J. S. Bae, H. S. Cheong, B. L. Park, L. H. Kim, C. S. Han, T. J. Park, J. Y. Kim, C. F. A. Pasaje, J. S. Lee, and H. D. Shin, “Genome-wide profiling of structural genomic variations in Korean HapMap individuals.,” *PLoS ONE*, vol. 5, no. 7, p. e11417, Jul. 2010.