www.AUSTRALIANACADEMICPRESS.com.au

ARTICLE AVAILABLE ONLINE
Twin Research and Human Genetics
Volume 14 ▮ Number 3 ▮ pp. 228–232

SNP Sets and Reading Ability: Testing Confirmation of a 10-SNP Set in a Population Sample

Michelle Luciano,[1] Grant W. Montgomery,[2] Nicholas G. Martin,[2] Margaret J. Wright[2] and Timothy C. Bates[1,2]
[1] *Centre for Cognitive Aging and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, United Kingdom*
[2] *Queensland Institute of Medical Research, Brisbane, Australia*

A set of 10 SNPs associated with reading ability in 7-year-olds was reported based on initial pooled analyses of 100K SNP chip data, with follow-up testing stages using pooling and individual testing. Here we examine this association in an adolescent population sample of Australian twins and siblings (N = 1177) aged 12 to 25 years. One (rs1842129) of the 10 SNPs approached significance (P = .05) but no support was found for the remaining 9 SNPs or the SNP set itself. Results indicate that these SNPs are not associated with reading ability in an Australian population. The results are interpreted as supporting use of much larger SNP sets in common disorders where effects are small.

■ **Keywords:** dyslexia, association, GWAS, normal reading, genes

Reading disability is a disorder for which genetic linkage has proven powerful (Cardon et al., 1994). Moreover, association studies focused within these linkage regions (Cope et al., 2005), clinical pedigree studies (Nopola-Hemmi et al., 2001), and translocations (Hannula-Jouppi et al., 2005) have led to strong evidence for candidate genes, including in normal samples (Bates et al., 2010a; 2010b; Lind et al., 2009; Luciano et al., 2007). Though valuable, gene variants discovered to date are far from sufficient to account for the heritable variance in dyslexia. Research is therefore turning to both pathway-based and hypothesis-free genome-wide association testing. To this end, Meaburn and colleagues (2008) conducted a DNA pooling-based association study for reading disability, finding support for association at 10 SNPs. An exciting prospect underpinning a significant portion of the future value of clinical genetics lies in aggregating small risk factors into diagnostic and prognostic tests (Rutter & Plomin, 2009). It is unclear in the absence of data how many SNPs will be required for utility, and utility itself will depend on purpose: Mendelian randomization may be aided greatly by even a relatively small poly-SNP set (Davey Smith, 2010), while accurate diagnosis of genetic risk may require tens of thousands of SNPs derived from large cohorts (Purcell et al., 2009). In this brief report, we examine confirmability of a small SNP set for a complex trait in a large, representative sample using individual-based testing.

Meaburn et al., (2008) implemented a three-stage design using the Twins Early Development Study sample. The phenotype used was age-7 scores on the Test of Word Reading Efficiency — which measures rate of reading aloud from lists of words and non-words — combined with teacher-ratings of reading ability. At stage 1, DNA-pools were formed for the top and bottom 25% of individuals based on reading scores in 3,043 twins (one from each twin pair selected at random); 302 SNPs differed in frequency by more than 10% between the pools. In the second stage, the top and bottom 10% of subjects from an increased sample of 4,258 (including co-twins of those in the pooling stage) were genotyped for 75 SNPs from the top 302; nine were significant. Stage 3 involved

RECEIVED 11 November, 2010; ACCEPTED 10 January, 2011.

ADDRESS FOR CORRESPONDENCE: Professor Timothy Bates/Dr Michelle Luciano, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, United Kingdom. E-mail: tim.bates@ed.ac.uk or michelle.luciano@ed.ac.uk

individual genotyping of these nine SNPs, plus 14 approaching significance in 3408 members of the stage 2 cohort (the middle 80% of scorers). Ten of these were associated with reading by one-tailed test, accounting for 0.15% of variance on average, with a total variance of 1% for the SNP set.

Background information on the selected SNPs has been reported (Meaburn et al., 2008). Five of the target SNPs were on the SNP chip used in the present study, the remaining five were substituted for the following proxy SNPs: rs1323381 (proxy rs1556876, $r^2 = 1.0$), rs1320490 (rs10495260, $r^2 = 0.93$), rs2192595 (rs2192594, $r^2 = 1.0$), rs2409411 (rs2833444, $r^2 = 0.96$) and rs4754752 (proxy rs6590849, $r^2 = 1.0$). With the exception of two of these proxy markers — which showed extremely strong r^2 — the remaining proxies were perfectly correlated with the original SNP and therefore will give the identical result to the original (ungenotyped) SNP. These SNPs were combined to form a SNP set to test for association with reading ability in our population sample.

Materials and Methods

Sample

Twins and their non-twin siblings were recruited from ongoing studies of melanoma risk factors and cognition (Wright et al., 2001): 1,177 individuals from 538 families (136 monozygotic, 343 dizygotic, 11 triplets) had phenotype and genotyping data. Their age ranged from 12.3 to 25.1 years (mean = 17.9, *SD* = 2.9), 54.5% were female, and 98% reported Caucasian ancestry, predominantly Anglo-Celtic (~ 82%). Ethical approval for this study was received from the Human Research Ethics Committee, Queensland Institute of Medical Research. Written informed consent was obtained from each participant and their parent/guardian (if younger than 18 years).

Measures

A quantitative measure of reading ability was formed as the principal component of the irregular-word, regular word and non-word scales for reading assessed using the CORE (Bates et al., 2004). This measure is an extended version of the Castles and Coltheart (1993) test, with additional items included to increase the difficulty level for an older sample. The three 40-item reading scales were administered untimed and assessed over the telephone by a trained researcher. For monozygotic twin pairs the mean for the two participants was used.

Genotyping

DNA was extracted from blood samples and genotyped with the Illumina 610K chip. Data-checking procedures were based on exclusion of unreliable samples and SNPs, as described in Benyamin et al., (2009). These included deviation from Hardy-Weinberg equilibrium at $p < 10^{-6}$, minor allele frequency < .01, and Mendelian errors. Subjects found to be of non-European ancestry by principal components analysis of the genotyping data were also excluded.

Analysis

Where proxy SNPs were used, we ran a linear regression of the number of minor alleles (0, 1, 2) at locus 1 (original SNP) on the number of minor alleles (0, 1, 2) at locus 2 (proxy SNP) using the HapMap CEPH data. This was to determine whether the linkage disequilibrium (LD) was positive or negative so that the allele could be coded in the same direction as the original study for composition of the SNP set score. To construct the SNP set score, the genotypes for each of the SNPs was recoded so that the decreaser allele homozygote was assigned a value of 0, the heterozygote assumed a value of 1, and the increaser allele homozygote was assigned a value of 2. These values were then summed across SNPs to derive a SNP set score.

For association with individual SNPs we considered additive models, including adjustment for the effects of age (and age squared), and sex and tester in MERLIN

TABLE 1

SNP names, Locations, Minor Allele Frequency (MAF), and Association Significance with Reading Ability.

	SNP	Location	Gene	MAF	One-tailed *p* value in Stage III Meaburn et al., (2008)	*p* value
1	rs10507218	12q23.3	—	0.36	.04	.17
2	rs1323381	9q31.3	—	0.13	.03	.87
3	rs1556876 (rs10485609)	20q13.13	*CSE1L*	0.25	.02	.17
4	rs10505938	20q13.13	*ARFGEF2*	0.21	.03	.07
5	rs1160219	11p15.1	*IGSF22*	0.23	.01	.25
6	rs10495260 (rs1320490)	1q42.11	*CDC42BPA*	0.18	.03	.52
7	rs1842129	6q22.31	*NKAIN2*	0.45	.02	.05
8	rs2192594 (rs2192595)	14q24.2	*DPF3*	0.16	< .01	.70
9	rs2833444 (rs2409411)	21q22	*TIAM1*	0.35	< .01	.58
10	rs6590849 (rs4754752)	11q22	—	0.21	.03	.70

Note: Where a proxy SNP was used the original Meaburn et al., (2008) SNP is shown in brackets.

https://doi.org/10.1375/twin.14.3.228 Published online by Cambridge University Press

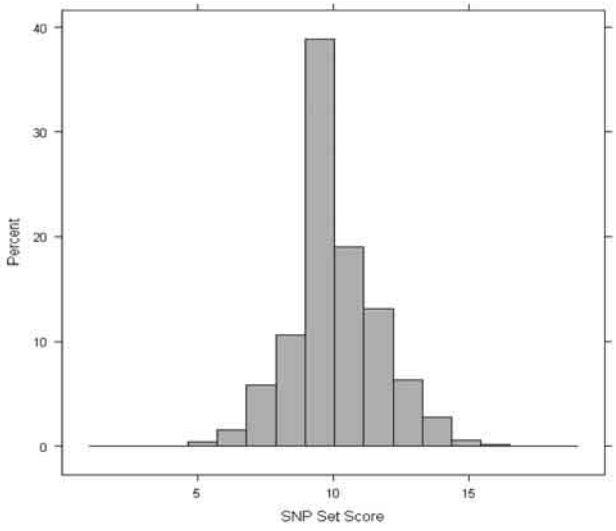

FIGURE 1
Frequency distribution of SNP set scores for individuals with complete data for all 10 SNPs.

(Chen & Abecasis, 2007). Age was correlated 0.26 with the reading principal component, and female participants scored higher than male participants. For a SNP explaining 1% of variance in our traits, under an additive model and against a background sibling correlation of ~0.30, we have > 95% power (α = .05) to detect association for a SNP with minor allele frequency above 0.05 (Purcell et al., 2003). A Bonferroni correction for 10 independent tests gave a new significance level of .005. The SNP set association was analyzed by linear regression in R (R Development Core Team, 2010).

Results

There was good variation in our reading scores, with adjusted standardized scores ranging from -5.67 to 2.49 (but note that scores below -4 were excluded as outliers). The results for individual SNP associations are shown in Table 1; none reached significance at an uncorrected level. Nevertheless, we analyzed the SNP set because of the gain in power that a combined measure can give. The distribution of the SNP set was normal, with scores ranging between 3 and 16 (see Figure 1). There was no association between the SNP set score and general reading ability (b = -0.03, p = .20). The distribution of reading scores for each of the observed SNP set scores is shown in Figure 2.

Discussion

This was the first test in an independent sample of the 10-SNP set identified by Meaburn et al., (2008) as influencing reading disability/ability. In our sample, this SNP set was not associated with general reading ability and nor were any individual SNPs. This lack of replication might be due to Type 1 error in the original study. A previous 5-SNP set (identified by DNA pooling) associated with general cognitive ability (Harlaar et al., 2005) also failed replication in independent samples (Luciano et al., 2008), suggesting that the small SNP set approach may not be better than prediction at an individual SNP level.

The SNPs in the set were originally identified through DNA pooling comparing low and high reading ability groups (Meaburn et al., 2008). But because co-twins of those in the pooling analysis were included in the second stage there was potential genetic non-independence and therefore bias in the second stage results. Furthermore, most SNPs (14/23) included at stage 3 had not passed stage 2, rendering the design weaker for these SNPs. Other biases in the DNA pooling method used, such as inefficient SNP selection, may also be relevant (Macgregor, 2010).

Sample discrepancies seem unlikely to explain the failure to replicate in this case as both samples drew on a population of very similar genetic background, that is, Caucasians in Britain and Australia (most of whom reported British ancestry). While our sample was older than the discovery sample and a different phenotype was used, we have previously replicated other SNPs originally identified in younger samples, also using different measures — for example, dyslexia diagnoses (Lind et al., 2010; Luciano et al., 2007). The use by Meaburn et al., (2008) of combined timed-TOWRE and teacher ratings may increase the loading on comprehension and/or timed aspects of reading. Raskind et al. (2005) reported linkage support at 2q that was specific for speed of non-word reading ('phonological decoding'). We found support for linkage in this region for our accuracy-based measures of regular word reading (Bates et al., 2007). This is consistent with the double-deficit hypothesis of reading in which lexical access speed is the basis for sight vocabulary (Wolf et al., 2000). Comprehension is closely linked to IQ, and

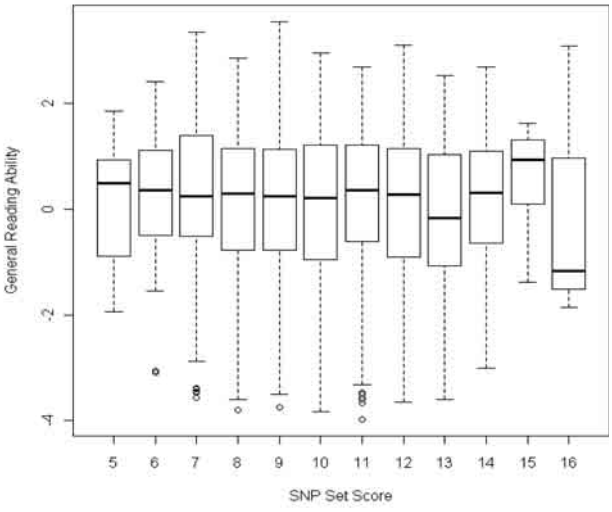

FIGURE 2
Standardized general reading ability scores for SNP set genotypic scores.

https://doi.org/10.1375/twin.14.3.228 Published online by Cambridge University Press

this may be a factor in the two results. Both speed and comprehension warrant further study.

While it is tempting to conclude that the report by Meaburn et al., (2008) suggests an upper limit of 0.15% for SNPs affecting reading, more optimistic conclusions are also compatible with the result. It is interesting that none of the SNPs reported were located in dyslexia candidate genes; for example, *KIAA0319* (Cope et al., 2005; Luciano et al., 2007) and *DYX1C1* (Bates et al., 2010a; Dahdouh et al., 2009). This suggests that genome coverage of the 100K SNP chip may be inadequate to detect signals known to be present, and is compatible with the view that SNPs of larger effect may well lie in regions not in LD with SNPs on this older chip. Work by Wray et al. (2009), indicates that even current chips are insensitive to well-known functional variants with psychiatric relevance, such as the *5HTLPR* polymorphism. There are also numerous other sources of genomic variance, including rare variants, indels, and copy number variants, which are not well covered by current SNP chips (Cooper et al., 2008).

While we found no evidence for a 10-SNP set predicting reading ability, SNP sets (on much larger scales) have been shown to predict psychiatric disease. For instance, the Schizophrenia consortium showed that as many as 38,000 SNPs taken from a discovery sample genome-wide association analysis could predict a third of risk for schizophrenia and even bipolar disorder in independent cohorts (Purcell et al., 2009). This study explored weighting allele scores by their effect size in the discovery sample. This and other weighting schemes might also improve prediction from small SNP sets. While the 10-SNP set did not generalize to our phenotype and sample, SNP sets clearly predict genetic risk in populations, and developing these for additional disorders, such as dyslexia, must be a priority, both for risk assessment and for their subsequent utility in identifying biological pathways of interest, affording targeted analysis of genes in such pathways.

Acknowledgments

We thank the twins and their parents for participating; Anjali Henders and Megan Campbell for managing sample processing and DNA extraction; Alison Mackenzie for coordinating the reading project; Marlene Grace, Ann Eldridge and the research interviewers for data collection. We thank the Office of the Chief Scientist of Scotland, the ARC and NHMRC for present and past support of this research. ML is a Royal Society of Edinburgh/Lloyds TSB Foundation for Scotland Personal Research Fellow.

References

Bates, T. C., Castles, A., Coltheart, M., Gillespie, N., Wright, M., & Martin, N. G. (2004). Behaviour genetic analyses of reading and spelling: a component processes approach. *Australian Journal of Psychology, 56*, 115–126.

Bates, T. C., Lind, P. A., Luciano, M., Montgomery, G. W., Martin, N. G., & Wright, M. J. (2010a). Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. *Molecular Psychiatry, 15*, 1190–1196.

Bates, T. C., Luciano, M., Castles, A., Coltheart, M., Wright, M. J., & Martin, N. G. (2007). Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *European Journal of Human Genetics, 15*, 194–203.

Bates, T. C., Luciano, M., Medland, S. E., Montgomery, G. W., Wright, M. J., & Martin, N. G. (2010b). Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. *Behavior Genetics*.

Benyamin, B., McRae, A. F., Zhu, G., Gordon, S., Henders, A. K., Palotie, A., Peltonen, L., Martin, N. G., Montgomery, G. W., Whitfield, J. B., & Visscher, P. M. (2009). Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. *American Journal of Human Genetics, 84*, 60–65.

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., & DeFries, J. C. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science, 266*, 276–279.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180.

Chen, W. M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *American Journal of Human Genetics, 81*, 913–926.

Cooper, G. M., Zerr, T., Kidd, J. M., Eichler, E. E., & Nickerson, D. A. (2008). Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nature Genetics, 40*, 1199–1203.

Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., Owen, M. J., O'Donovan, M.C., & Williams, J. (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *American Journal of Human Genetics, 76*, 581–591.

Dahdouh, F., Anthoni, H., Tapia-Paez, I., Peyrard-Janvid, M., Schulte-Korne, G., Warnke, A., Remschmidt, H., Ziegler, A., Kere, J., Müller-Myhsok, B., Nöthen, M. M., Schumacher, J., & Zucchelli, M. (2009). Further evidence for DYX1C1 as a susceptibility factor for dyslexia. *Psychiatric Genetics, 19*, 59–63.

Davey Smith, G. (2010). Mendelian randomization for strengthening causal inference in observational studies: Application to gene x environment interactions. *Perspectives on Psychological Science*, 527–545.

Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H., & Kere, J. (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet, 1*, e50.

Harlaar, N., Butcher, L. M., Meaburn, E., Sham, P., Craig, I. W., & Plomin, R. (2005). A behavioural genomic analysis of DNA markers associated with general cognitive ability in 7-year-olds. *Journal of Child Psychology and Psychiatry, 46*, 1097–1107.

Lind, P., Luciano, M., Duffy, D., Castles, A., Wright, M. J., Martin, N. G., & Bates, T. C. (2009). Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *European Journal of Human Genetics, 18*, 668–673.

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

Luciano, M., Lind, P. A., Deary, I. J., Payton, A., Posthuma, D., Butcher, L. M., Bochdanovits, Z., Whalley, L. J., Visscher, P. M., Harris, S. E., Polderman, T. J., Davis, O. S., Wright, M. J., Starr, J. M., de Geus, E. J., Bates, T. C., Montgomery, G. W., Boomsma, D. I., Martin, N. G., & Plomin, R. (2008). Testing replication of a 5-SNP set for general cognitive ability in six population samples. *European Journal of Human Genetics, 16*, 1388–1395.

Luciano, M., Lind, P. A., Duffy, D. L., Castles, A., Wright, M. J., Montgomery, G. W., Martin, N. G., & Bates, T. C. (2007). A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biological Psychiatry, 62*, 811–817.

Macgregor, S. (2010). Optimal selection of markers from DNA pooling experiments. *Behavior Genetics, 40*, 46–47; discussion 48.

Meaburn, E. L., Harlaar, N., Craig, I. W., Schalkwyk, L. C., & Plomin, R. (2008). Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. *Molecular Psychiatry, 13*, 729–740.

Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., Voutilainen, A., Kere, J., & Widén, E. (2001). A dominant gene for developmental dyslexia on chromosome 3. *Journal of Medical Genetics, 38*, 658–664.

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics, 19*, 149–150.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., & the International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*, 748–752.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raskind, W. H., Igo, R. P., Chapman, N. H., Berninger, V. W., Thomson, J. B., Matsushita, M., Brkanac, Z., Holzman, T., Brown, M., & Wijsman, E. M. (2005). A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2q that contributes to phonological decoding efficiency. *Molecular Psychiatry, 10*, 699–711.

Rutter, M., & Plomin, R. (2009). Pathways from science findings to health benefits. *Psychological Medicine, 39*, 529–542.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33*, 387–407.

Wray, N. R., James, M. R., Gordon, S. D., Dumenil, T., Ryan, L., Coventry, W. L., Statham, D. J., Pergadia, M. L., Madden, P. A., Heath, A. C., Montgomery, G. W., & Martin, N. G. (2009). Accurate, large-scale genotyping of 5HTTLPR and flanking single nucleotide polymorphisms in an association study of depression, anxiety, and personality measures. *Biological Psychiatry, 66*, 468–476.

Wright, M. J., De Geus, E., Ando, J., Luciano, M., Posthuma, D., Ono, Y., Hansell, N., Van Baal, C., Hiraishi, K., Hasegawa, T., Smith, G., Geffen, G., Geffen, L., Kanba, S., Miyake, A., Martin, N., & Boomsma, D. (2001). Genetics of cognition: Outline of a collaborative twin study. *Twin Research, 4*, 48–56.

https://doi.org/10.1375/twin.14.3.228 Published online by Cambridge University Press