

## ARTICLE

# Intelligence: shared genetic basis between Mendelian disorders and a polygenic trait

Sanja Franić<sup>\*1</sup>, Maria M Groen-Blokhuis<sup>1</sup>, Conor V Dolan<sup>1,2</sup>, Mathijs V Kattenberg<sup>1</sup>, René Pool<sup>1</sup>, Xiangjun Xiao<sup>3</sup>, Paul A Scheet<sup>3</sup>, Erik A Ehli<sup>4</sup>, Gareth E Davies<sup>4</sup>, Sophie van der Sluis<sup>5</sup>, Abdel Abdellaoui<sup>1</sup>, Narelle K Hansell<sup>6</sup>, Nicholas G Martin<sup>6</sup>, James J Hudziak<sup>7</sup>, Catherina E M van Beijsterveldt<sup>1</sup>, Suzanne C Swagerman<sup>1</sup>, Hilleke E Hulshoff Pol<sup>8</sup>, Eco J C de Geus<sup>1</sup>, Meike Bartels<sup>1</sup>, H Hilger Ropers<sup>9</sup>, Jouke-Jan Hottenga<sup>1</sup> and Dorret I Boomsma<sup>1</sup>

Multiple inquiries into the genetic etiology of human traits indicated an overlap between genes underlying monogenic disorders (eg, skeletal growth defects) and those affecting continuous variability of related quantitative traits (eg, height). Extending the idea of a shared genetic basis between a Mendelian disorder and a classic polygenic trait, we performed an association study to examine the effect of 43 genes implicated in autosomal recessive cognitive disorders on intelligence in an unselected Dutch population ( $N=1316$ ). Using both single-nucleotide polymorphism (SNP)- and gene-based association testing, we detected an association between intelligence and the genes of interest, with genes *ELP2*, *TMEM135*, *PRMT10*, and *RGS7* showing the strongest associations. This is a demonstration of the relevance of genes implicated in monogenic disorders of intelligence to normal-range intelligence, and a corroboration of the utility of employing knowledge on monogenic disorders in identifying the genetic variability underlying complex traits.

*European Journal of Human Genetics* (2015) 23, 1378–1383; doi:10.1038/ejhg.2015.3; published online 25 February 2015

## INTRODUCTION

Multiple inquiries into the genetic etiology of complex human traits have indicated that, for a number of phenotypes, the genetic variants affecting continuous, polygenic phenotypic variation may be concentrated in the same genes as those giving rise to monogenic (ie, Mendelian) disorders. For instance, 180 loci associated with normal variation in the classic polygenic trait of adult height were shown to be enriched in genes underlying skeletal growth disorders.<sup>1</sup> Many rare genetic variants in three candidate genes (*ABCA1*, *APOA1*, and *LCAT*), which give rise to pathologically low levels of HDL cholesterol in plasma, are also found in individuals with the common, complex version of the low-HDL-cholesterol trait.<sup>2–4</sup> Genes underlying Mendelian disorders of lipid levels, and those affecting their normal concentration overlap almost entirely.<sup>5</sup> Other examples include hemoglobin F levels,<sup>5</sup> fat mass,<sup>6</sup> type 2 diabetes,<sup>5,7,8</sup> and Parkinson's disease.<sup>9,10</sup>

Genes underlying Mendelian disorders, in which protein functioning is severely altered, may therefore provide an opportunity to localize and understand the genetic variability that underlies susceptibility to a similar common polygenic phenotype.<sup>2</sup> In the present study, we utilize this idea to examine the effects of 43 genes implicated in autosomal recessive cognitive disorders<sup>11</sup> on intelligence in a Dutch sample from the general population ( $N=1316$ ; see Materials and Methods, and Supplementary Figures S1 and S2). Despite its being one of the most heritable human traits (with heritability estimates ranging from 0.6 to

0.8 in adolescence and adulthood<sup>12,13</sup>), no loci consistently associated with normal-range variation in intelligence have thus far been reported.<sup>13–16</sup> The two largest genome-wide association studies (GWAS) to date failed to find replicable genome-wide association in SNP-based analyses in adults and children, respectively.<sup>15,17</sup> The 43 genes considered in the present study are a subset of the genes identified in a recent study that used homozygosity mapping, exon enrichment, and next-generation sequencing in consanguineous families with autosomal-recessive intellectual disabilities to identify single, presumably disease-causing variants in 50 novel candidate genes.<sup>11</sup> The genome-wide data set of the Netherlands Twin Register (NTR),<sup>18</sup> used in the present study, contains SNP data on 43 of these 50 genes (Table 1), including 1227 genotyped SNPs in total (Supplementary Table S1).

## MATERIALS AND METHODS

### Sample

The data were obtained from the NTR.<sup>18,19</sup> The NTR is a population-based register of Dutch twins recruited at birth and measured longitudinally at ages 1 through 18. The sample consisted of 1316 individuals from 662 families (978 twins, 231 siblings, and 107 of their parents). To keep the genetic within-family covariance matrix approximately compound symmetric (ie, to keep the genetic covariances between each type of relatives approximately equal), the data were selected so as to contain no complete MZ twin pairs and no more than one parent per family. Thus, each family consisted of individuals who were

<sup>1</sup>Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; <sup>2</sup>Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands; <sup>3</sup>Division of OVP, Cancer Prevention and Population Sciences, Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>4</sup>Avera Institute for Human Genetics, Avera McKennan Hospital, University Health Center, Sioux Falls, SD, USA; <sup>5</sup>Section Functional Genomics, Department of Clinical Genetics, VU Medical Center Amsterdam, Amsterdam, The Netherlands; <sup>6</sup>Genetic Epidemiology, Molecular Epidemiology and Neurogenetics Laboratories, Queensland Institute of Medical Research, Brisbane, Australia; <sup>7</sup>Department of Psychiatry and Medicine, University of Vermont, Burlington, VT, USA; <sup>8</sup>Neuroimaging Research Group, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>9</sup>Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany  
\*Correspondence: Dr S Franić, Department of Biological Psychology, VU University Amsterdam, Faculty of Psychology and Education, Van der Boechorststraat 1, Amsterdam, 1081 BT The Netherlands. Tel: +31 20 598 4027; E-mail: s.franic@vu.nl

Received 12 February 2014; revised 16 December 2014; accepted 25 December 2014; published online 25 February 2015

**Table 1 Chromosomal position (hg19), length, and number of genotyped SNPs for the 43 genes**

Gene	Chromosome	Start (bp)	End (bp)	Length (kb)	N of SNPs
PARP1	1	224097741	226600780	2503.039	385
RGS7	1	240926554	241525530	598.976	177
TMEM135	11	86743886	87039756	295.870	83
LAMA1	18	6936743	7122813	186.070	68
FRY	13	32600437	32875794	275.357	67
ADK	10	75905960	76474061	568.101	48
SCAPER	15	76635526	77202785	567.259	48
ASCC3	6	100951070	101334248	383.178	31
PECR	2	216856052	216952678	96.626	27
POLR3B	12	106746436	106908976	162.540	24
ENTPD1	10	97449774	97642023	192.249	23
ACBD6	1	180239515	180477089	237.574	23
NDST1	5	149860381	149942773	82.392	18
ZBTB40	1	22773344	22862650	89.306	17
INPP4A	2	99056317	99212496	156.179	17
ELP2	18	33704407	33762909	58.502	16
TAF2	8	120738015	120850103	112.088	14
LINS	15	101094574	101148435	53.861	14
KDM5A	12	384223	503620	119.397	13
CACNA1G	17	48633429	48709835	76.406	13
SLC31A1	9	115978842	116034217	55.375	11
CAPN10	2	241521133	241562122	40.989	9
KIF7	15	90147020	90203682	56.662	8
RALGDS	9	135968107	136044301	76.194	8
WDR45L	17	80567438	80611411	43.973	8
GON4L	1	155714508	155834191	11.9683	8
C9orf86	9	139689818	139740639	50.821	6
TTI2	8	33325904	33376119	50.215	5
UBR7	14	93668401	93700561	32.160	4
ZCCHC8	12	122952417	122990518	38.101	4
CCNA2	4	122732599	122750087	17.488	3
C11orf46	11	30339598	30364774	25.176	3
FASN	17	80031215	80061106	29.891	3
PRMT10	4	148553936	148610381	56.445	3
MAN1B1	9	139976379	140008635	32.256	3
CNKSRI	1	26498894	26521377	22.483	3
HIST1H4B	6	26022124	26032480	10.356	2
HIST3H3	1	228607546	228618026	10.480	2
EEF1B2	2	207019309	207032652	13.343	2
CASP2	7	142980308	143009789	29.481	2
ASCL1	12	103346464	103359294	12.830	2
KDM6B	17	7738222	7763106	24.884	1
ERLIN2	8	37589117	37621619	32.502	1
Total					1227

genetically either siblings or parent-offspring, that is, the expected genetic correlation between any given pair of family members was 0.5. The observed intraclass correlation between the family members was 0.57 (SE = 0.025). In all, 45.8% of the sample were males. The mean ages of children and parents were 12.7 (SD = 4.1) and 43.9 (SD = 4.1), respectively. The age distribution (showing each participant's mean age across measurement occasions) is given in Supplementary Figure S1.

**Phenotype data**

Intelligence was assessed longitudinally using the Revised Amsterdam Children Intelligence Test (RAKIT<sup>20</sup>), Wechsler Intelligence Scale for Children (WISC<sup>21–23</sup>), Raven's Standard and Advanced Progressive Matrices (SPM, APM<sup>24,25</sup>), and the Wechsler Adult Intelligence Scale (WAIS<sup>26,27</sup>), the choice of test being largely dependent on participants' age. A previous study employing

the same data set demonstrated a high genetic stability of intelligence scores as assessed by the different tests (the autoregressive coefficients between the additive genetic factors at consecutive measurement occasions ranging from 0.8 to 1).<sup>28</sup> Therefore, the individuals' mean scores across the different ages were used as measure of the phenotype. The IQ scores were derived based on the age- and sex-appropriate norms for the RAKIT, WISC, or WAIS, and subsequently converted to z-scale within each measurement occasion and averaged over measurement occasions. For the 154 participants for whom only the Raven's Matrices scores were available, we used z-transformed scores on Raven's Matrices. The distribution of intelligence scores is given in Supplementary Figure S2.

**Genotype data**

Blood and/or buccal samples for DNA extraction were collected as part of several projects within the NTR. Genotyping was performed using the Affymetrix Human SNP Array 6.0. Genotypes were called using the BIRDSEED V2 algorithm. SNPs in Hardy-Weinberg equilibrium ( $P > 0.00001$ ) with a minor allele frequency exceeding 0.01 and a missingness rate below 5% were included in the analyses. Samples were selected if their call rate exceeded 95% and were checked for Mendelian errors, excessive heterozygosity ( $-0.1 < F < 0.1$ ), and discrepancies in relatedness.<sup>29</sup> Genotypes displaying Mendelian inheritance errors were excluded from the analyses.

For the present study, we selected all genotyped SNPs from the 50 genes of interest, including a 5-kb border around each gene. In all, 7 out of the 50 genes contained no genotyped SNPs. The distribution of the SNPs (1227 in total) over the remaining 43 genes is shown in Table 1. The full list of SNPs is given in Supplementary Table S1. In all, 0.85% of the SNPs were in exonic regions, and out of that, 61.3% were non-synonymous.

**Analyses**

**SNP-based analyses.** As a first step, we tested for an association between the phenotype and each of the 1227 SNPs. As the observations were clustered in families, the analyses were performed using a multilevel regression model with random intercepts to account for the within-family covariance structure. Specifically, the model for phenotype of person  $i$  in family  $j$  was  $ph_{ij} = b_{0j} + b_1 * SNP_{ij} + res_{ij}$ , where  $ph$  denotes phenotype,  $b_{0j}$  is intercept in family  $j$ ,  $b_1$  is a (fixed) slope parameter, and  $res_{ij}$  denotes an individual-specific residual term. The intercept term can be further decomposed as  $b_{0j} = g_0 + k_{0j}$ , where  $g_0$  is a fixed component and  $k_{0j}$  is a component that is random over families. Using random intercepts prevents the inflation of type I error associated with applying a standard (fixed-effects) regression model to family-clustered data. The within-family genetic covariance structure was approximately compound symmetric (ie, the expected genetic correlation between any given type of relatives was 0.5). The analyses were implemented using the 'nlme' package in R.<sup>30</sup> The code used to carry out the analyses is given in Script S1.

Additionally, we performed association testing using the Plink software package.<sup>31</sup> Here the association between the phenotype and each of the 1227 SNPs was examined using the Huber-White sandwich variance estimator to account for the family structure in the data. The results were compared with those obtained using the multilevel regression model in R. A high degree of correspondence between the results obtained using the multilevel regression model (which effectively assumes an AE background covariance structure among first-degree relatives) and those obtained using the Huber-White sandwich estimator (which corrects for relatedness without assuming a background model) would imply that any background misspecification in the random effects model has not affected the conclusions. A high degree of correspondence is expected, because the test of a fixed effect in the multilevel regression model is fairly robust to possible background misspecification.<sup>32</sup>

To empirically evaluate the results obtained for the 1227 SNPs, we drew a number of random samples of: (a) 1227 SNPs from the entire genome, (b) 1227 SNPs from intragenic regions of the genome, and (c) 43 genes (including all SNPs on those genes) from the autosomal genome. All samples excluded the 1227 SNPs of interest. Each of the random samples was subjected to the analyses described above. The resulting QQ plots and genomic inflation factors ( $\lambda$ ) were compared to those obtained for the 1227 SNPs of interest.

As additional verification of the results, permutation was employed to generate an empirical distribution of  $\lambda$  values under the null hypothesis of no association. The genotypes (ie, the 1227 SNPs) were randomly reallocated over the phenotypes 1000 times, and each of the 1000 permuted data sets was analyzed using the random intercept multilevel regression model described above. To account for the background covariance structure arising from the clustering of data in families, family data were relocated jointly: the genotypes of any two-member family were reassigned to phenotypes of another randomly selected two-member family, and the same was done for three- and four-member families. Thus, the family structure in the permuted data sets remained intact. As in the original analyses, the family structure was subsequently corrected for using a multilevel model. The null distribution of  $\lambda$  values generated using the permuted data sets was compared to the  $\lambda$  obtained for the 1227 SNPs of interest.

Finally, a genome-wide association study was performed. Here, the phenotype was regressed on each of the available genotyped SNPs (538652 SNPs) using the Plink software package.<sup>31</sup>

All analyses were performed using an additive model and included eight genomic principal components<sup>33</sup> as covariates to account for any possible effects of population stratification. All  $\lambda$  values were estimated as regression coefficients of the observed on the expected  $-\log_{10}$  of the  $P$ -values, using the GenABEL package in R.<sup>30</sup>

**Gene-based analyses.** In the next step, the SNP-based  $P$ -values obtained using the multilevel regression model were used as input for gene-based analysis. A gene-based association test that employs the extended Simes procedure (GATES) was used.<sup>34</sup> GATES involves jointly analyzing all available SNPs in a gene to obtain a single  $P$ -value associated with the gene. The method assumes that an association test between the phenotype and all available SNPs on the gene has been carried out, and that the resulting  $P$ -values and pair-wise allelic correlation coefficients  $r$  for all SNPs are available. In the present case, we used the  $P$ -values obtained in the SNP-based multilevel regression analysis, and pair-wise allelic correlation coefficients obtained using the  $-r$  option in Plink. Given  $m$  SNPs on a gene, a gene-based  $P$ -value is obtained through an iterative procedure by combining the ascendingly ordered  $m$   $P$ -values in the following way:  $P_G = \min(m_e p_{(j)} / m_{e(j)})$ , where  $m_e$  is the effective number of independent  $P$ -values among the  $m$  SNPs,  $m_{e(j)}$  is the effective number of independent  $P$ -values among the top  $j$  SNPs ( $j = 1, \dots, m$ ), and  $p_j$  is the  $j$ -th lowest  $P$ -value (ie, the  $P$ -value associated with the  $j$ -th top SNP). The null hypothesis of this gene-based test is that none of the SNPs are associated with the phenotype; the alternative is that at least one SNP is associated. The effective number of independent  $P$ -values among the  $m$  SNPs,  $m_e$ , is estimated as  $m_e = m - \sum [I(\lambda_i > 1)(\lambda_i - 1)] \lambda_i > 0$ , where  $I(x)$  is an indicator function and  $\lambda_i$  is the  $i$ -th eigenvalue of the  $m \times m$  correlation matrix ( $\rho$ ) of the  $P$ -values obtained in the SNP-based association test. The pair-wise  $P$ -value correlation coefficient,  $\rho_{ij}$ , can be approximated by a 6th order polynomial function of the

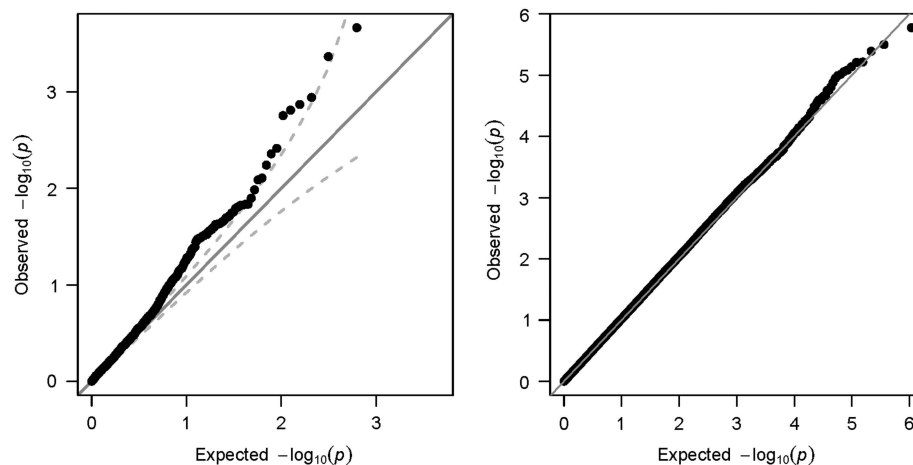
allelic correlation coefficient  $r_{ij}$ :  $\rho_{ij} = 0.2982r_{ij}^6 - 0.0127r_{ij}^5 + 0.0588r_{ij}^4 + 0.0099r_{ij}^3 + 0.6281r_{ij}^2 - 0.0009r_{ij}$ , where  $\rho_{ij}$  and  $r_{ij}$  are the  $ij$ -th elements of the SNP  $P$ -value correlation matrix  $\rho$ , and of the allelic correlation matrix  $r$ , respectively. For a full overview of the method, we refer the reader to the original publication<sup>34</sup> and to Script S2, which contains our implementation GATES in R. The R script performs the test  $k$  times given  $k$  genes in the input file.

Additionally, we performed a gene-based association test using the Versatile Gene-Based Test for Genome-wide Association Studies (VEGAS),<sup>35</sup> and compared the results to those obtained using GATES. VEGAS is a simulation-based method that uses information from the full set of SNPs within a gene and accounts for linkage disequilibrium (LD) by using simulations from the multivariate normal distribution. The analyses were performed using the VEGAS web-interface.<sup>35</sup>

## RESULTS

### SNP-based analyses

Association between intelligence scores and each of the 1227 SNPs (see Materials and Methods) was examined using an additive model and eight principal components<sup>33</sup> to account for the possible effects of population stratification (Script S1). The left panel of Figure 1 shows a quantile-quantile (QQ)-plot, including 95% confidence intervals (CIs), of the association  $P$ -values (also see Supplementary Figures S3 and S4, and Supplementary Table S2). The CI estimates were obtained while taking into account the LD structure between the SNPs: instead of  $N = 1227$ , we used an estimate of the effective number of independent  $P$ -values ( $N = 625$ ). This approach produces relatively broader CIs; we thus adopt a more stringent approach to evaluate the significance of the difference between the expected and the observed distributions. As evident from the figure, the distribution of the observed  $P$ -values differs markedly from that expected under the null hypothesis of no effect, indicating an enrichment of the 43 candidate genes for polymorphisms associated with intelligence. Note that the significant inflation along nearly the entire length of the QQ plot (genomic inflation factor  $\lambda = 1.26$ ) is not necessarily indicative of population stratification, particularly in the context of a candidate SNP study. Here, the observed inflation is expected under the alternative hypothesis of (polygenic) effects of a relatively large number of the candidate SNPs tested.<sup>36</sup> As the analyses were performed while adjusting for eight principal components (seven of which were correlated with geographic latitude and longitude in the present sample,<sup>33</sup> thereby feasibly representing differences in ancestry),



**Figure 1** Left: QQ-plot based on the 1227 candidate SNPs. Right: Genome-wide QQ-plot based on 538652 SNPs. Dashed lines: 95% confidence intervals (CIs).

population stratification does not appear to be a likely cause of the inflation.

To empirically verify the finding and confirm the absence of population stratification, we performed SNP-based association testing on samples of SNPs drawn randomly from the genome. In particular, we drew 1000 random samples of: (a) 1227 SNPs from the entire genome, (b) 1227 SNPs from intragenic regions of the genome, and (c) 43 genes (including all SNPs on those genes) approximately matched for size with the 43 candidate genes and sampled from the entire autosomal genome. All random samples excluded the 1227 SNPs of interest. The distributions of the  $\lambda$  values obtained for each set of random samples, along with the  $\lambda$  obtained for the 1227 SNPs of interest (marked by a horizontal line), are depicted in Figure 2. As evident from the figure (panels a and b), the effect found for the SNPs of interest did not replicate in any of the 2000 random samples obtained by sampling SNPs from the entire genome or from the intragenic regions of the genome. For SNPs residing on randomly sampled sets of 43 genes (panel c, Figure 2), only 3.6% of  $\lambda$  values exceed the  $\lambda$  obtained for the candidate SNPs. Note that the higher variance in panel c of Figure 2 relative to that in Panels a and b is expected given that the degree of non-independence of SNPs (ie, LD) is considerably higher in SNPs sampled from the same gene relative to those sampled from the entire genome. A reduced effective number of independent SNPs is expected to result in a less precise estimate of  $\lambda$ , that is, in a higher dispersion around the mean  $\lambda$  value.

As further empirical verification, we performed permutation testing to obtain an empirical distribution of  $\lambda$  values under the null hypothesis of no association: the genotypes (ie, the 1227 SNPs of interest) were randomly reallocated over the phenotypes 1000 times,

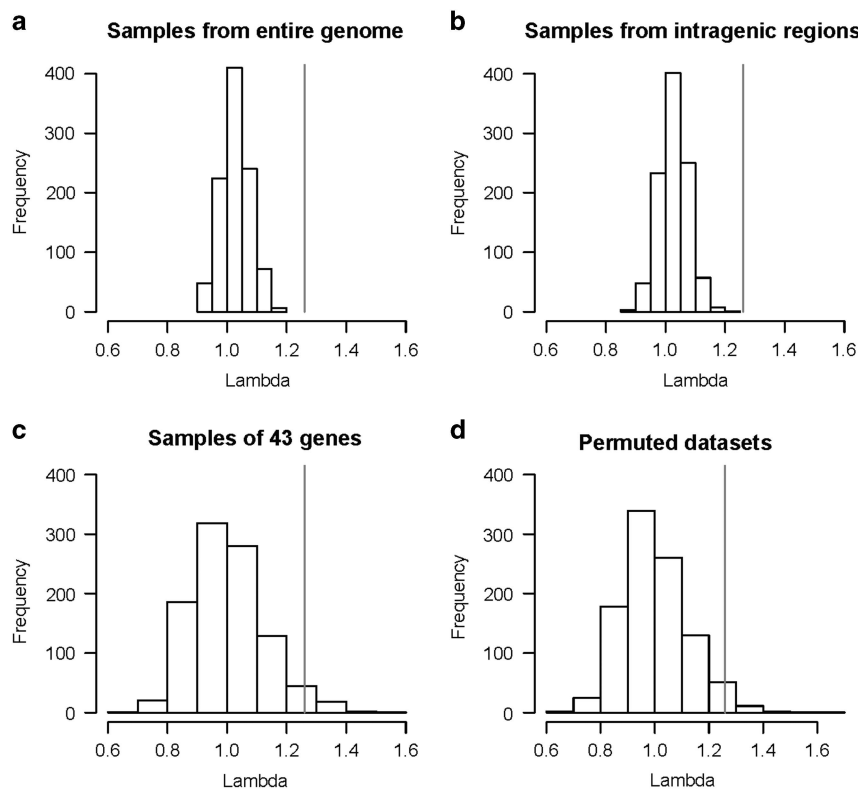
and each of the 1000 permuted datasets was analyzed using SNP-based association testing. The resulting distribution of  $\lambda$  values and the  $\lambda$  obtained for the non-permuted data set ( $\lambda=1.26$ ) are shown in panel d of Figure 2. Here, only 2.9% of the  $\lambda$  values exceed the  $\lambda$  value of interest; an empirical  $P$ -value consistent with that obtained from random sampling.

Finally, a genome-wide association analysis was performed. Here, the phenotype was regressed on each of the available genotyped SNPs (538652 SNPs). The resulting QQ plot is depicted in the right panel of Figure 1. As evident from the figure, the genome-wide  $P$ -values in the right panel show no notable inflation ( $\lambda = 1.03$ ), in contrast to the left panel ( $\lambda = 1.26$ ).

The present results thus consistently indicate an enrichment of the candidate set of genes for polymorphisms associated with intelligence, while plausibly ruling out population structure as the cause of the observed effect. The former is demonstrated by the significant inflation of the association  $P$ -values for the candidate set of SNPs as compared with random subsets of SNPs (empirical  $P=0.036$ ) and to a permutation-based null distribution (empirical  $P=0.029$ ). The latter is established by (a) the inclusion of genetic principal components into the association study, (b) the near absence of comparable  $P$ -value inflation in randomly selected sets of SNPs, and (c) the absence of genome-wide  $P$ -value inflation.

#### Gene-based analyses

Next, gene-based testing was carried out (see Materials and Methods and Script S2). The full list of gene-based results is given in Supplementary Table S3. Genes *ELP2* ( $P=0.007$ ), *TMEM135* ( $P=0.007$ ), *PRMT10* ( $P=0.019$ ), and *RGS7* ( $P=0.044$ ) displayed



**Figure 2** Distribution of genomic inflation factors ( $\lambda$ ) obtained for 1000 (a) random samples of 1227 SNPs from the entire genome, (b) random samples of 1227 SNPs from intragenic regions of the genome, (c) random samples of 43 genes from the entire genome, and (d) permuted data sets. Horizontal line:  $\lambda$  obtained for in the non-permuted data set for the 1227 SNPs of interest.

the strongest associations, although no association was significant after correction for multiple testing (with  $\alpha=0.05/43$ ). Notably, 2 out of the 50 genes from the Najmabadi *et al*<sup>11</sup> study harbor more than one mutation associated with cognitive disabilities; one of those genes is *ELP2*, which, in the present study, shows the strongest evidence of association.

Focusing on the four nominally significant genes, we examined the positions of the most strongly associated SNPs in these genes relative to the mutations in Najmabadi *et al*<sup>11</sup> (Supplementary Figure S5). As evident from the figure, both mutations in *ELP2*, as well as the mutations in *TMEME135* and *PRMT10*, are relatively close to our top SNP for their respective genes; the distances range from 4.8 kb to 31.4 kb. On *RGS7*, the distance between the mutation and the top SNP is relatively large (535.7 kb). Note that any distance between the disease-causing mutation and our top SNP is consistent with the logic of the present study however, as the gene is viewed as a functional unit with regard to its etiological relevance to intelligence, regardless of the distribution of the functionally relevant polymorphisms along the gene.

For validation, both the gene-based analyses and the SNP-based analyses were performed using several different methods (see Materials and Methods). The results obtained using the different methods converged highly: the  $\log_{10}$  of the *P*-values obtained using two methods of SNP-based testing correlated 0.88, and the *P*-values obtained using two different gene-based tests correlated 0.89 (Supplementary Table S3).

## DISCUSSION

The present study focused on 43 genes implicated in autosomal recessive cognitive disorders in consanguineous Iranian families,<sup>11</sup> and found these to be enriched for polymorphisms associated with normal-range intelligence in a Dutch population-based sample. This is a demonstration of the relevance of genes implicated in monogenic disorders of cognitive ability to continuous variability in intelligence. Despite the high heritability of intelligence,<sup>12,28,37,38</sup> the progress in the identification of loci consistently associated with variation in its normal range has thus far been limited.<sup>15,17,38–42</sup> Exceptions are the apolipoprotein E (*APOE*) gene at older ages<sup>43</sup> and formin binding protein 1-like (*FNBPL1*), the latter having recently been shown to be associated with both childhood and adulthood intelligence.<sup>15,17</sup> The present approach utilizes the idea that the differentially sized effects of individual mutations located within a gene functionally relevant to the phenotype may range from severe disruptions of protein functioning (resulting in a Mendelian disorder) to smaller effects underlying polygenic variation. Utilizing prior knowledge on genetics of Mendelian disorders may therefore prove a valuable approach to the identification of genetic variability underlying polygenic traits, with the advantage of requiring considerably smaller sample sizes than GWAS to achieve adequate power. This may prove especially useful in the study of phenotypes for which large samples are difficult to obtain, for instance because the phenotype is difficult or costly to measure (eg, neuropsychological or fMRI measures), and/or in detection of genetic variants characterized by small effect sizes. For instance, in the present study we clearly demonstrate enrichment, although none of the *P*-values for individual SNPs fall below the Bonferroni-corrected significance threshold ( $\alpha=0.05/1227=0.00004$ , or  $\alpha=0.05/625=0.00008$  if one corrects by the number of independent SNPs<sup>34</sup>), indicating that the magnitudes of individual SNP effects are too small to be detected in regular GWAS.

Although larger sample sizes are needed to identify the exact genes and genetic variants driving the association in the present study, we

focus on the top four genes that reach nominal significance. The most strongly associated gene, *ELP2* (elongator complex protein 2), encodes a subunit of the RNA polymerase II elongator complex,<sup>44</sup> involved in acetylation of histones H3 and probably H4 and possibly in chromatin remodeling. *TMEME135* (transmembrane protein 135) is involved in fat metabolism and energy expenditure.<sup>45</sup> *PRMT10* (protein arginine methyltransferase 10) affects chromatin remodeling leading to transcriptional regulation, RNA processing, DNA repair, and cell signaling.<sup>46</sup> *RGS7* (regulator of G-protein signaling 7) interacts with 14-3-3 protein, tau, and snapin (a component of the SNARE complex required for synaptic vesicle docking and fusion).<sup>47</sup>

The utilization of knowledge on monogenic disorders to identify polymorphisms that affect the variability of continuous phenotypes is a cost-efficient approach to understand the genetic variability underlying polygenic traits. At present, the causal variants for a large number of monogenic disorders have been identified (over 3000 disorders; Online Mendelian Inheritance in Man (OMIM): <http://www.ncbi.nlm.nih.gov/omim>), and recent developments in sequencing technologies have made it possible to employ exome sequencing or whole-genome sequencing, possibly in combination with homozygosity mapping, as an efficient approach to identifying novel causal variants underlying Mendelian disorders.<sup>48–50</sup> The National Human Genome Research Institute has opened Centers for Mendelian Genomics (NHGRI Genome Sequencing Program, <http://www.genome.gov/>), whose primary goal is the discovery of as yet unknown variation underlying Mendelian disorders. Thus, at present, the utilization of existing and impending knowledge on variants underlying Mendelian disorders to identify the variation underlying polygenic traits may prove a viable, efficient, and cost-effective complement to standard approaches such as GWAS. The present finding highlights the importance of continuing the efforts directed at studying monogenic diseases<sup>50,51</sup> at a time when focus has shifted away from them, as they can advance our understanding of multifactorial traits.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

Funding was obtained from the European Science Council (ERC Advanced, 230374), the Netherlands Organization for Scientific Research (NWO): grants 480-04-004, 051.02.060, SPI56-464-14192, NWO/MaGW: VIDI-452-12-014), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI NL, 184.021.007), the VU University's Institute for Health and Care Research (EMGO+), Neuroscience Campus Amsterdam (NCA), Avera Institute, Sioux Falls, South Dakota (USA), and the National Institutes of Health: Grand Opportunity grant 1RC2MH089951-01. Statistical analyses were partly carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI Posthuma) and the Dutch Brain Foundation and the VU University Amsterdam. This work is part of the programme of BiG Grid, the Dutch e-Science Grid, which is financially supported by the NWO.

- Allen HL, Estrada K, Lettre G *et al*: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832–838.
- Antonarakis SE, Beckmann JS: Opinion - Mendelian disorders deserve more attention. *Nat Rev Genet* 2006; **7**: 277–282.
- Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; **305**: 869–872.

- 4 Frikke-Schmidt R, Nordestgaard BG, Jensen GB, Tybjaerg-Hansen A: Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* 2004; **114**: 1343–1353.
- 5 Hirschhorn JN, Gajdos ZK: Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annu Rev Med* 2011; **62**: 11–24.
- 6 Loos RJ, Lindgren CM, Li S *et al*: Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; **40**: 768–775.
- 7 Sandhu MS, Weedon MN, Fawcett KA *et al*: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 2007; **39**: 951–953.
- 8 Winckler W, Weedon MN, Graham RR *et al*: Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* 2007; **56**: 685–693.
- 9 Gasser T: Mendelian forms of Parkinson's disease. *Biochim Biophys Acta* 2009; **1792**: 587–596.
- 10 Lesage S, Brice A: Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* 2009; **18**: R48–R59.
- 11 Najmabadi H, Hu H, Garshasbi M *et al*: Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011; **478**: 57–63.
- 12 Plomin R, DeFries JC, McClearn GE, McGuffin P: *Behavioral Genetics*. W.H. Freeman & Co Ltd: New York, NY. 2008.
- 13 Deary IJ, Johnson W, Houlihan LM: Genetic foundations of human intelligence. *Hum Genet* 2009; **126**: 215–232.
- 14 Butcher LM, Davis OSP, Craig IW, Plomin R: Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500 K single nucleotide polymorphism microarrays. *Genes Brain Behav* 2008; **7**: 435–446.
- 15 Davies G, Tenesa A, Payton A *et al*: Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 2011; **16**: 996–1005.
- 16 Chabris CF, Hebert BM, Benjamin DJ *et al*: Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci* 2011; **23**: 1314–1323.
- 17 Benyamin B, Pourcain B, Davis O *et al*: Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol Psychiatry* 2013; **19**: 253–258.
- 18 van Beijsterveldt CE, Groen-Blokhuis M, Hottenga JJ *et al*: The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Res Human Genet* 2013; **16**: 252.
- 19 Boomsma DI, de Geus EJC, Vink JM *et al*: Netherlands Twin Register: From twins to twin families. *Twin Res Human Genet* 2006; **9**: 849–857.
- 20 Bleichrodt N, Drenth P, Zaai J, Resing W: *Revisie Amsterdamse Kinder Intelligentie Test, RAKIT*. Swets and Zeitlinger: Lisse. 1984.
- 21 Wechsler D, Kort W, Compaan EL, Bleichrodt N, Resing WCM, Schittkatte M: *Wechsler Intelligence Scale for Children-Third Edition. Dutch version*. Swets and Zeitlinger: Lisse, The Netherlands. 2002.
- 22 Sattler JM: *Assessment of Children: WISC-III and WPPSI-R Supplemental*. San Diego, CA. 1992.
- 23 Van Haasen P, De Bruyn E, Pijl Y *et al*: *Wechsler Intelligence Scale for Children-Revised, Dutch Version*. Swets & Zeitlinger BV: Lisse, The Netherlands. 1986.
- 24 Raven JC: *Guide to the Standard Progressive Matrices: Sets A, B, C, D and E*. London: HK Lewis, 1960.
- 25 Raven J, Raven JC, Court JH: *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4, The Advanced Progressive Matrices*. Oxford Psychologists Press: Oxford, England. 1998.
- 26 Stinissen J, Willems P, Coetsier P, Hulsman W: *Manual for the Dutch Translated and Adapted Version of the Wechsler Adult Intelligence Scale (WAIS)*. Swets and Zeitlinger: Lisse. 1970.
- 27 Wechsler D: *Wechsler Adult Intelligence Scale-Third Edition, Dutch Version*. Swets and Zeitlinger: Lisse, The Netherlands. 1997.
- 28 Franic S, Dolan CV, Borsboom D, van Beijsterveldt CEM, Boomsma DI: Multivariate genetic analysis of longitudinally measured cognitive abilities. *Behav Genet* 2010; **40**: 792–792.
- 29 Scheet P, Ehli EA, Xiao X *et al*: Twins, tissue, and time: an Assessment of SNPs and CNVs. *Twin Res Hum Genet* 2012; **15**: 737.
- 30 R\_Development\_Core\_Team R: *A Language and Environment for Statistical Computing*. 2.13.0 edn R Foundation for Statistical Computing: Vienna, Austria. 2011.
- 31 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 32 Minica CC, Dolan CV, Hottenga J-J, Willemsen G, Vink JM, Boomsma DI: The use of imputed sibling genotypes in sibship-based association analysis: on modeling alternatives, power and model misspecification. *Behav Genet* 2013; 1–13.
- 33 Abdellaoui A, Hottenga J-J, de Knijff P *et al*: Population structure, migration, and diversifying selection in the Netherlands. *Eur J Hum Genet* 2013; **21**: 1277–1285.
- 34 Li M-X, Gui H-S, Kwan JSH, Sham PC: GATES: a rapid and powerful gene-based association test using extended simes procedure. *Am J Hum Genet* 2011; **88**: 283–293.
- 35 Liu JZ, McRae AF, Nyholt DR *et al*: A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–145.
- 36 Yang J, Weedon MN, Purcell S *et al*: Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 2011; **19**: 807–812.
- 37 Haworth C, Wright M, Luciano M *et al*: The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatry* 2009; **15**: 1112–1120.
- 38 Deary IJ, Johnson W, Houlihan L: Genetic foundations of human intelligence. *Hum Genet* 2009; **126**: 215–232.
- 39 Chabris CF, Hebert BM, Benjamin DJ *et al*: Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci* 2012; **23**: 1314–1323.
- 40 Need AC, Attix DK, McEvoy JM *et al*: A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet* 2009; **18**: 4650–4661.
- 41 Butcher LM, Davis OS, Craig IW, Plomin R: Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500 K single nucleotide polymorphism microarrays. *Genes Brain Behav* 2008; **7**: 435–446.
- 42 Davis OS, Butcher LM, Docherty SJ *et al*: A three-stage genome-wide association study of general cognitive ability: hunting the small effects. *Behav Genet* 2010; **40**: 759–767.
- 43 Deary IJ, Whiteman MC, Pattie A *et al*: Ageing: cognitive change and the APOE  $\epsilon$  4 allele. *Nature* 2002; **418**: 932–932.
- 44 Fellows J, Erdjument-Bromage H, Tempst P, Svejstrup JQ: The E1p2 subunit of elongator and elongating RNA polymerase II holoenzyme is a WD40 repeat protein. *J Biol Chem* 2000; **275**: 12896–12899.
- 45 Exil VJ, Avila DS, Benedetto A *et al*: Stressed-induced TMEM135 protein is part of a conserved genetic network involved in fat storage and longevity regulation in *Caenorhabditis elegans*. *PLoS ONE* 2010; **5**: e14228.
- 46 Fisk JC, Zurita-Lopez C, Sayegh J, Tomasello DL, Clarke SG, Read LK: TbPRMT6 is a type I protein arginine methyltransferase that contributes to cytokinesis in *Trypanosoma brucei*. *Eukaryot Cell* 2010; **9**: 866–877.
- 47 Ilardi JM, Mochida S, Sheng Z-H: Snapin: a SNARE-associated protein implicated in synaptic transmission. *Nat Neurosci* 1999; **2**: 119–124.
- 48 Najmabadi H, Motazacker MM, Garshasbi M *et al*: Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum Genet* 2007; **121**: 43–48.
- 49 Inlow JK, Restifo LL: Molecular and comparative genetics of mental retardation. *Genetics* 2004; **166**: 835–881.
- 50 Ku C-S, Naidoo N, Pawitan Y: Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011; **129**: 351–370.
- 51 Ropers H-H: Single gene disorders come into focus-again. *Dialogues Clin Neurosci* 2010; **12**: 95.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)