Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence

Suzanne Sniekers1, Sven Stringer1, Kyoko Watanabe1, Philip R Jansen1,2, Jonathan R I Coleman3,4, Eva Krapohl3, Erdogan Taskesen1,5, Anke R Hammerschlag1, Aysu Okbay1,6, Delilah Zabaneh3, Najaf Amin7, Gerome Breen3,4, David Cesarini8, Christopher F Chabris9,21, William G Iacono10, M Arfan Ikram11, Magnus Johannesson12, Philipp Koellinger1,9, James J Lee10,13, Patrik E Magnusson14, Matt McGue10, Mike B Miller10, William E R Ollie15, Antony Payton15, Neil Pendleton16, Robert Plomin3, Cornelius A Rietveld6,17, Henning Tiemeier2,11,18, Cornelia M van Duijn7,19 & Danielle Posthuma1,20

Intelligence is associated with important economic and health-related life outcomes1. Despite intelligence having substantial heritability (0.54) and a confirmed polygenic nature, initial genetic studies were mostly underpowered3–5. Here we report a meta-analysis for intelligence of 78,308 individuals. We identify 336 associated SNPs (METAL P < 5 × 10−8) in 18 genomic loci, of which 15 are new. Around half of the SNPs are located inside a gene, implicating 22 genes, of which 11 are new findings. Gene-based analyses identified an additional 30 genes (MAGMA P < 2.73 × 10−6), of which all but one had not been implicated previously. We show that the identified genes are predominantly expressed in brain tissue, and pathway analysis indicates the involvement of genes regulating cell development (MAGMA competitive P = 3.5 × 10−6). Despite the well-known difference in twin-based heritability for intelligence in childhood (0.45) and adulthood (0.80), we show substantial genetic correlation (rG = 0.89, LD score regression P = 5.4 × 10−29). These findings provide new insight into the genetic architecture of intelligence.

We combined genome-wide association study (GWAS) data for intelligence in 78,308 unrelated individuals from 13 cohorts (Online Methods). Of these individuals, full GWAS results for intelligence on n = 48,698 have been published in two different studies5,6 (n = 12,441 and n = 36,257, respectively), while GWAS results for the remaining 29,610 individuals have not been published previously. Across the different cohorts, various tests to measure intelligence were used. Therefore—following previous publications on combining intelligence phenotypes across different cohorts5,7—the cohorts either calculated Spearman’s g or used a primary measure of fluid intelligence (Supplementary Table 1), which is known to correlate highly with g. Previous research has shown that many different aspects of intelligence are highly correlated to each other and that Spearman’s g captures the latent general intelligence trait, irrespective of the specific tests used to construct it8,10.

All association studies were performed on individuals of European descent; standard quality control procedures included correcting for population stratification and filtering on minor allele frequency (MAF) and imputation quality (Online Methods). As 8 of the 13 cohorts consisted of children (aged <18 years; total n = 19,509) and 5 consisted of adults (n = 58,799; aged 18–78 years), we first performed meta-analysis of the children- and adult-based cohorts separately using METAL software11 and subsequently calculated rG using LD score regression12. The estimated rG was 0.89 (s.e.m. = 0.08, P = 5.4 × 10−29), indicating substantial overlap between the genetic variants influencing
Figure 1  Regional association and linkage disequilibrium plots for 18 genome-wide significant loci. The y axis represents the negative logarithm (base 10) of the SNP $P$ value and the x axis represents the position on the chromosome, with the name and location of genes in the UCSC Genome Browser shown in the bottom panel. The SNP with the lowest $P$ value in the region is marked by a purple diamond. The colors of the other SNPs indicate the $r^2$ of these SNPs with the lead SNP. Plots were generated with LocusZoom. 

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
intelligence in childhood and adulthood, and warranting a combined meta-analysis. The genetic correlations between all individual cohorts were generally larger than 0.80 except for those involving some of the smaller-sized cohorts (n < 4,000), which, given the large standard errors of the r-values, is likely due to the relatively low sample sizes in some of the individual cohorts (Supplementary Table 2). The full meta-analysis of all 13 cohorts (maximum n = 78,308) included 12,104,294 SNPs. The quantile–quantile plot of all SNPs exhibited some inflation (λ_m = 1.21; Supplementary Fig. 1 and Supplementary Table 3), which is within the expected range for a polygenic trait at the current sample size and heritability (13). We performed LD score regression to quantify the proportion of inflation in the mean (\lambda_m) for the 336 genome-wide significant SNPs. The lower the score, the more likely it is that a SNP has a regulatory function. For b–d, the numbers in parentheses in the legends refer to the number of lead SNPs for that category. ncRNA, noncoding RNA; TSS, transcription start site; TF, transcription factor.

Results of SNP-based meta-analysis for intelligence based on 78,308 individuals. Association results from the GWAS meta-analysis pertaining to individuals of European descent. (a) Negative log10-transformed P values for each SNP (y axis) are plotted by chromosomal position (x axis). The red and blue lines represent the thresholds for genome-wide statistically significant associations (P = 5 × 10^{-8}) and suggestive associations (P = 1 × 10^{-5}), respectively. Green dots represent the independent hits. (b) Functional categories for the 336 genome-wide significant SNPs. (c) The minimum (most active) chromatin state across 127 tissues for the 336 genome-wide significant SNPs. (d) The Regulome database score for the 336 genome-wide significant SNPs. The lower the score, the more likely it is that a SNP has a regulatory function.
To examine the robustness of the 336 SNPs and 47 genes that reached genome-wide significance in the primary analyses, we sought replication. Because there are no reasonably large GWAS for intelligence available and given the high genetic correlation with educational attainment, which has been used previously as a proxy for intelligence\(^7\), we used the summary statistics from the latest GWAS for educational attainment\(^8\) for proxy-replication (Online Methods). We first deleted overlapping samples, resulting in a sample of 196,931 individuals for educational attainment. Of the 336 top SNPs for intelligence, 306 were available for look-up in educational attainment, including 16 of the independent lead SNPs. We found that the effects of 305 of the 306 available SNPs in educational attainment were significant in the replication, with 150 of these being the independent SNPs, representing 47 genes that were significantly associated with intelligence in the GWAS\(^1\). Of the 47 genes that were significantly associated with intelligence in the GWAS, 15 were also significantly associated with educational attainment \((P < 0.05/47; \text{Supplementary Table 15})\). Given the high (0.70) but not perfect genetic correlation between educational attainment and intelligence, these results strongly support the involvement of the proxy-replicated SNPs and genes in intelligence.

The strongest emerging association with intelligence is with rs2490272 (6q21) in an intronic region of FOXO3 and neighboring SNPs in the promoter of the same gene. This gene is part of the insulin/insulin-like growth factor 1 signaling pathway and is believed to be involved in long-term memory formation\(^2\). The gene is also reported to be involved in the insulin/insulin-like growth factor 1 signaling pathway and is believed to be involved in long-term memory formation\(^2\). The gene is also reported to be involved in the insulin/insulin-like growth factor 1 signaling pathway and is believed to be involved in long-term memory formation\(^2\).
Figure 3 Gene-based genome-wide analysis for intelligence and genetic overlap with other traits. (a) Negative log_{10}-transformed P values for each gene are plotted. Green dots represent significantly associated genes from GWGAS. The threshold for gene-wide statistical significant associations was set at the Bonferroni threshold of \( P = 2.73 \times 10^{-6} \); the suggestive threshold was set at \( P = 2.73 \times 10^{-5} \). (b) Heat map of gene expression levels of genes for intelligence in 45 tissue types (see Supplementary Table 1 for n values per tissue). A value above zero (red) depicts a relatively high expression level with a value below zero (blue) depicts a relatively low expression level. (c) Epigenetic states of genes. The bars denote the proportions of epigenetic states across 127 tissue types. (d) Genetic correlations between intelligence and 32 health-related outcomes. Error bars show 95% confidence intervals for estimates of \( r_g \). Red bars represent the traits that showed a significant genetic correlation after correction for multiple testing (\( P < 1.56 \times 10^{-5} \)), pink bars represent the traits that showed a nominally significant correlation (\( P < 0.05 \)) and blue bars represent the traits that did not show a genetic correlation significantly different from zero. Note, as Alzheimer’s disease is an age-related disorder, we calculated the \( r_g \) with this phenotype across three age groups and found no difference in the \( r_g \) values (Supplementary Note). TSS, transcription start site.
genetic overlap with several neuropsychiatric and metabolic disorders. These findings provide starting points for understanding the molecular neurobiological mechanisms underlying intelligence, one of the most investigated traits in humans.

URLs. UK Biobank, http://www.ukbiobank.ac.uk; genotyping and quality control of UK Biobank, http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580; CHIC summary statistics http://ssgac.https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; MAGMA, http://ctg.cncr.nl/software/magma; MSigDB, http://software.broadinstitute.org/gsea/msigdb/collections.jsp; METAL, http://genome.sph.umich.edu/wiki/METAL_Program; LD score regression (LDSC), https://github.com/bulik/lodsc.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was funded by the Netherlands Organization for Scientific Research (NWO VICI 453-14-005). The analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003), by VU University, Amsterdam, the Netherlands, and by the Dutch Brain Foundation and is hosted by the Dutch National Computing and Networking Services SurfSARA. This research has been conducted using the UK Biobank resource under application number 16406. We thank the participants and researchers who collected and contributed to the data.

Summary statistics have been made available for download from http://ctg.cncr.nl/software/summary_statistics.

AUTHOR CONTRIBUTIONS

S. Sneekes performed the analyses. D.P. conceived the study. S. Stringer performed quality control on the UK Biobank data. K.W. and E.T. conducted in silico follow-up analyses. P.R.J., E.K. and J.R.L.C. conducted polygenic risk score analyses. P.K., C.A.R., D.Z., H.T., C.M.v.D., N.A., P.M., D.C., M.J., M.M., M.B.M., W.G.I., J.L., G.B., R.P., N.P., A.P., W.E.R.O., M.A.I. and C.F.C. contributed data. A.R.H. provided scripts for the pathway analyses. A.O. performed the educational attainment meta-analysis. S. Sneekes and D.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Deary, I.J. Intelligence. Annu. Rev. Psychol. 63, 453–482 (2012).
2. Polderman, T.J.C. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. 47, 702–709 (2015).
3. Chabris, C.F. et al. Most reported genetic associations with general intelligence are probably false positives. Psychol. Sci. 23, 1314–1323 (2012).
4. Davies, G. et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol. Psychiatry 16, 996–1005 (2011).
5. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with FYNBP1L. Mol. Psychiatry 19, 253–258 (2014).
6. Davies, G. et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Mol. Psychiatry 21, 758–767 (2016).
7. Rietveld, C.A. et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proc. Natl. Acad. Sci. USA 111, 13790–13794 (2014).
8. Deary, I.J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. Nat. Rev. Neurosci. 11, 201–211 (2010).
9. Johnson, W., Bouchard, T.J., Krueger, R.F., McGue, M. & Gottesman, I.I. Just one thing: consistent results from three test batteries. Intelligence 32, 95–107 (2004).
10. Ree, M.J. & Eales, J.A. The stability of g across different methods of estimation. Intelligence 15, 271–278 (1991).
11. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
12. Bulik-Sullivan, B.K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
13. Yang, J. et al. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19, 807–812 (2011).
14. Davies, G. et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). Mol. Psychiatry 20, 183–192 (2015).
15. Boyle, A.P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1792 (2012).
16. Vilhjálmsson, B.J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 97, 576–592 (2015).
17. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).
18. Ashburner, M. et al. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29 (2000).
19. Fabregat, A. et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 44, D481–D487 (2016).
20. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550 (2005).
21. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533, 539–542 (2016).
22. Newman, A.B. & Murabito, J.M. The epidemiology of longevity and exceptional survival. Epidemiol. Rev. 35, 181–197 (2013).
23. Willcox, B.J. et al. FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. USA 105, 13987–13992 (2008).
24. Flachsbart, F. et al. Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc. Natl. Acad. Sci. USA 106, 2700–2705 (2009).
25. Behrens, P., Brinkmann, U. & Wellmann, A. CSE1L/CAS: its role in proliferation and apoptosis. Apoptosis 8, 39–44 (2003).
26. Velez Edwards, D.R. et al. Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women's Health Initiative SHARE Study. Hum. Genet. 132, 323–336 (2013).
27. Spellotes, E.K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 42, 937–948 (2010).
28. Willer, C.J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat. Genet. 41, 25–34 (2009).
29. Locke, A.E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197–206 (2015).
30. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).
31. Comuzzie, A.G. et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS One 7, e51954 (2012).
32. Berndt, S.I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat. Genet. 45, 501–512 (2013).
33. Wheeler, E. et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nat. Genet. 45, 513–517 (2013).
34. Pruim, R.J. et al. LocusZoo: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
Discovery sample. The current study was based on 78,308 individuals. The origin of the samples is as follows:

1. UK Biobank web-based measure (UKB-wb; \( n = 17,862 \)). GWAS results have not yet been published; raw genotypic data were available for the present study.

2. UK Biobank touchscreen measure (UKB-ts; \( n = 36,257 \)), non-overlapping with UKB-wb. Results have been published before\(^6\), raw genotypic data were available for the present study.

3. CHIC consortium\(^7\) (\( n = 12,441 \)). Results have been published before; meta-analysis summary statistics were available for the present study.

4. Five additional cohorts (\( n = 11,748 \)). For these, 69 SNP associations with IQ have previously been published as part of a lookup effort\(^8\), but full GWAS results have not been published previously. Per-cohort full GWAS summary statistics were available for the present study.

We describe these data sets in more detail below.

UK Biobank samples (UKB-wb, UKB-ts). We used the data provided by the UK Biobank Study\(^9\) resource (see URLs), which is a major national health resource including >500,000 participants. All participants provided written informed consent; the UK Biobank received ethical approval from the National Research Ethics Service Committee North West–Haydock (reference 11/NW/0382), and all study procedures were performed in accordance with the World Medical Association Declaration of Helsinki ethical principles for medical research. The current study was conducted under UK Biobank application number 16406.

The study design of the UK Biobank has been described in detail elsewhere\(10,11\). Briefly, invitation letters were sent out in 2006–2010 to ~9.2 million individuals, including all people aged 40–69 years who were registered with the National Health Service and living up to ~25 miles from one of the 22 study assessment centers. A total of 503,325 participants were subsequently recruited into the study\(12\). Apart from registry-based phenotypic information, extensive self-reported baseline data have been collected by questionnaire, in addition to anthropometric assessments and DNA collection. For the present study, we used imputed data obtained from UK Biobank (May 2015 release) including ~73 million genetic variants in 152,249 individuals. Details on the data are provided elsewhere (see URLs). In summary, the first ~50,000 samples were genotyped on the UK BiLEVE Axiom array, and the remaining ~100,000 samples were genotyped on the UK Biobank Axiom array. After standard quality control of the SNPs and samples, which was centrally performed by UK Biobank, the data set comprised 641,018 autosomal SNPs in 152,256 samples for phasing and imputation. Imputation was performed with a reference panel that included the UK10K haplotype panel and the 1000 Genomes Project Phase 3 reference panel.

We used two fluid intelligence phenotypes from the Biobank data set. These are based on questionnaires that were taken either in the assessment center at the initial intake (‘touchscreen’, field 20016) or at a later moment at home (‘web-based’, field 20191). The measures indicate the number of correct answers out of 13 fluid intelligence questions. The data distribution roughly approximates a normal distribution.

For the analyses in our study, we only included individuals of European descent. After removal of related individuals and those with discordant sex, who withdrew consent or had missing phenotype data, 36,257 individuals remained for analysis for the fluid intelligence touchscreen measure and 28,846 remained for the web-based version. As 10,984 individuals had taken both the touchscreen and web-based test, we only included the data from the touchscreen test for these individuals. This resulted in 54,119 individuals with a score on either the fluid intelligence web-based (UKB-wb) or touchscreen (UKB-ts) version (Supplementary Table 1). At the time of taking the test, the age of the participants ranged between 40 and 78 years. Half of the participants were between 40 and 60 years old. 44% were between 60 and 70 years old and 6% were older than 70 years. The mean age was 58.98 years with a standard deviation of 8.19.

Summary statistics from the CHIC consortium. We downloaded the publicly available combined GWAS results from the meta-analyses as reported by CHIC\(^3\) (see URLs). Details on the included cohorts and performed analyses are reported in the original publication\(^5\). Briefly, CHIC includes six cohorts totaling 12,441 individuals: the Avon Longitudinal Study of Parents and Children (ALSPAC, \( n = 5,517 \)), the Lothian Birth Cohorts of 1921 and 1936 (LBC1921, \( n = 464 \); LBC1936, \( n = 947 \)), the Brisbane Adolescent Twin Study subsample of the Queensland Institute of Medical Research (QIMR, \( n = 1,752 \)), the Western Australian Pregnancy Cohort Study (Raine, \( n = 936 \)) and the Twins Early Development Study (TEDS, \( n = 2,825 \)). All individuals are children aged from 6–18 years. Within each cohort, the cognitive performance measure was adjusted for sex and age and principal components were included to adjust for population stratification. See also Supplementary Table 1.

SNP analysis in the UK Biobank sample. Association tests were performed in SNPTEST\(27\) (see URLs), using linear regression. Both phenotypes were corrected for a number of covariates, including age, sex and a minimum of five genetically determined principal components, depending on how many were associated with the phenotype (5 for the web-based test and 15 for the touchscreen version, tested by linear regression). Additionally, we included the Townsend deprivation index as a covariate, which is based on postal code and measures material deprivation. The touchscreen version of the phenotype was also corrected for assessment center and genotyping array. SNPs with imputation quality score <0.8 and MAF <0.01 (based on all Europeans present in the total sample) were excluded after the association analysis, resulting in 12,573,858 and 12,595,966 SNPs for the touchscreen and web-based test, respectively.

Gene analysis. The SNP-based \( P \) values from the meta-analysis were used as input for the gene-based analysis. We used all 19,427 protein-coding genes from the NCBI 37.3 gene definitions as the basis for a genome-wide gene association analysis (GWGAS) in MAGMA (see URLs). After SNP annotation, values from the meta-analysis were used as input for the gene-based analysis. We used all 19,427 protein-coding genes in a gene set is significantly larger than the combined effect of all other genes, whereas self-contained \( P \) values are not interpreted and not reported by us. Competitive \( P \) values were corrected for multiple testing using MAGMAs built-in empirical multiple-testing correction with 10,000 permutations.

Pathway analysis. We used MAGMA to test for association of predefined gene sets with intelligence. A total of 6,166 GO and 674 Reactome gene sets were obtained (see URLs). We computed competitive \( P \) values, which are less likely to be below the threshold of significance than self-contained \( P \) values. Competitive \( P \) values are the outcomes of the test that the combined effect of genes in a gene set is significantly larger than the combined effect of all other genes, whereas self-contained \( P \) values are informative when testing against the null hypothesis of no association. Self-contained \( P \) values are not interpreted and not reported by us. Competitive \( P \) values were corrected for multiple testing using MAGMAs built-in empirical multiple-testing correction with 10,000 permutations.

Meta-analysis. Meta-analysis of the results of the 13 cohorts was performed in METAL\(13\) (see URLs). We did not include SNPs that were not present in the UK Biobank sample. The analysis was based on \( P \) values, taking sample size and direction of effect into account using the sample size scheme.
Genetic correlations. Genetic correlations ($r_g$) were calculated between intelligence and 32 other traits for which summary statistics from GWAS were publicly available, using LD score regression (see URLs). This method corrects for sample overlap, by estimating the intercept of the bivariate regression. A conservative Bonferroni-corrected threshold of 1.56 × 10^{-3} was used to determine significant correlations.

Functional annotation. We identified all SNPs that had an $r^2$ value of 0.1 or higher with the 18 independent lead SNPs and were included in the METAL output. We used the 1000 Genomes Project Phase 3 reference panel to calculate $r^2$. We further filtered on SNPs with $P < 0.05$. In addition, we only annotated SNPs with MAF >0.01.

Positional annotations for all lead SNPs and SNPs in LD with the lead SNPs were obtained by performing ANNOVAR gene-based annotation using RefSeq genes. In addition, CADD scores38 and RegulomeDB15 scores were annotated to SNPs by matching chromosome, position, reference and alternative alleles. For each SNP, eQTLs were extracted from GTEx (44 tissue types)39, the Blood eQTL browser40 and BIOS gene-level eQTLs41. The eQTLs obtained from GTEx were filtered on gene $P < 0.05$, and eQTLs obtained from the other two databases were filtered on FDR < 0.05. The FDR values were provided by GTEx, BIOS and the Blood eQTL browser. For GTEx eQTLs, there is one FDR value available per gene–tissue pair. As such, the FDR is identical for all eQTLs belonging to the same gene–tissue pair. For BIOS and the Blood eQTL browser, an FDR value was computed for each SNP.

To test whether the SNPs were functionally active by means of histone modifications, we obtained epigenetic data from the NIH Roadmap Epigenomics Mapping Consortium42 and ENCODE43. For every 200 bp of the genome, a 15-core chromatin state was predicted by a hidden Markov model based on five histone marks (H3K4me3, H3K4me1, H3K27me3, H3K9me3 and H3K36me3) for 127 tissue and cell types44. We annotated chromatin states (15 states in total) to SNPs by matching chromosome and position for every tissue or cell type. We computed the minimum state (1, the most active state) and the consensus state (majority of states) across 127 tissue and cell types for each SNP.

Chromatin states were also determined for the 52 genes (47 from the gene-based test + 5 additional genes implicated by single-SNP GWAS). For each gene and tissue, the chromatin state was obtained per 200-bp interval in the gene. We then annotated the genes by means of a consensus decision when multiple states were present for a single gene; that is, the state of the gene was defined as the modus of all states present in the gene.

Tissue expression of genes. RNA sequencing data from 1,641 tissue samples with 45 unique tissue labels were derived from the GTEx consortium39. This set includes 313 brain samples over 13 unique brain regions (see Supplementary Table 18 for sample size per tissue). Of the 52 genes implicated by either the GWAS or the GWWGAS, 44 were included in the GTEx data. Normalization of the data was performed as described previously45. Briefly, genes with RPKM value smaller than 0.1 in at least 80% of the samples were removed. The remaining genes were log2 transformed (after using a pseudocount of 1), and finally a zero-mean normalization was applied.

Proxy replication in educational attainment. For the replication analysis, we used a subset of the data from ref. 21. In particular, we excluded the Erasmus Rucphen Family Study, the Minnesota Center for Twin and Family Research Study, the Swedish Twin Registry Study, the 23andMe data and all individuals from UK Biobank, to make sure that there was no sample overlap with our IQ data set. Genetic correlation between intelligence and educational attainment in this non-overlapping subsample was $r_g = 0.73$, s.e.m. = 0.03, $P = 1.4 \times 10^{-14}$. The replication analysis was based on the phenotype EduYears, which measures the number of years of schooling completed. A total of 306 of our 336 top SNPs (and 16 of 18 independent lead SNPs) were available in the educational attainment sample. We performed a sign concordance analysis for the 16 independent lead SNPs, using the exact binomial test. For each independent signal we determined whether either the lead SNP had a P value smaller than 0.05/16 in the educational attainment analysis or another (correlated) top SNP in the same locus had such a P value, if this was not the case for the lead SNP. All 47 genes implicated in the GWWGAS for intelligence were available for lookup in the educational attainment sample. For each gene, we determined whether it had a P value smaller than 0.05/16 in the educational attainment analysis.

Polygenic risk score analysis. We used LDpred16 to calculate the variance explained in intelligence in independent samples by a polygenic risk score based on our discovery analysis, as well as two previous GWAS for intelligence5,6. LDpred adjusts GWAS summary statistics for the effects of LD by using an approximate Gibbs sampler that calculates the posterior means of effects, conditional on LD information, when calculating polygenic risk scores. We used varying priors for the fraction of SNPs with nonzero effects (priors: 0.01, 0.05, 0.1, 0.5, 1 and an infinitesimal prior). Independent data sets available for polygenic risk score analyses are described in the Supplementary Note.

Data availability. Summary statistics have been made available for download from http://cogeneration.co.nz/software/summary_statistics. Genotype data that underlie the findings of this study are available from UK Biobank but restrictions apply to the availability of these data, which were used under license for the current study (application number 16406) and so are not publicly available. Summary statistics from the CHIC consortium are available from http://ssgac.org/documents/CHIC_Summary_Benyamin2014.txt.gz. Additional supporting data are provided in the supplementary material.

35. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779 (2015). 36. Allen, N.E., Sudlow, C., Peakman, T. & Collins, R. UK Biobank data: come and get it. Sci. Transl. Med. 6, 224ed4 (2014). 37. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007). 38. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315 (2014). 39. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015). 40. Westra, H.-J. et al. Systematic identification of trans-eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013). 41. Bonder, M.J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat. Genet. 49, 131–138 (2016). 42. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015). 43. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012). 44. Ernst, J. & Kellis, M. ChromatinMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012). 45. Taskesen, E. & Reinders, M.J.T. 2D representation of transcriptomes by t-SNE exposres relatedness between human tissues. PLoS One 11, e0149853 (2016).
Erratum: Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence

Suzanne Sniekers, Sven Stringer, Kyoko Watanabe, Philip R Jansen, Jonathan R I Coleman, Eva Krapohl, Erdogan Taskesen, Anke R Hammerschlag, Aysu Okbay, Delilah Zabaneh, Najaf Amin, Gerome Breen, David Cesarini, Christopher F Chabris, William G Iacono, M Arfan Ikram, Magnus Johannesson, Philipp Koellinger, James J Lee, Patrik K E Magnusson, Matt McGue, Mike B Miller, William E R Ollier, Antony Payton, Neil Pendleton, Robert Plomin, Cornelius A Rietveld, Henning Tiemeier, Cornelia M van Duijn & Danielle Posthuma

Nat. Genet.; doi:10.1038/ng.3869; corrected online 31 May 2017

In the version of this article initially published online, heritability was misspelled in the penultimate sentence of the abstract. The error has been corrected in the print, PDF and HTML versions of this article.