Title: Functional consequences of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals

Jonathan R. I. Coleman[1,2], Julien Bryois[3], Héléna A. Gaspar[1], Philip R. Jansen[4,5], Jeanne Savage[4], Nathan Skene[6], Robert Plomin[1], Ana B. Muñoz-Manchado[6], Sten Linnarsson[6], Greg Crawford[7,8], Jens Hjerling-Leffler[6], Patrick F. Sullivan[3,9], Danielle Posthuma[4,10,¶,*], Gerome Breen[1,2,¶,*]

Author Affiliations:

[1] MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK.
[2] NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London SE5 8AF, UK.
[3] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden.
[4] Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands.
[5] Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
[6] Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden.
[7] Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA
[8] Department of Pediatrics, Medical Genetics Division, Duke University, Durham, NC 27708, USA
[9] Departments of Genetics, University of North Carolina, Chapel Hill, NC, 27599-7264, USA.
[10] Department of Clinical Genetics, VU University Medical Center (VUMC), Neuroscience Campus Amsterdam, Amsterdam, 1081 HV, The Netherlands.
[¶] These authors contributed equally to the work
* Correspondence to: gerome.breen@kcl.ac.uk or danielle.posthuma@vu.nl

Running title: Secondary analysis of IQ GWAS

Conflict of Interest Statement

PF Sullivan reports the following potentially competing financial interests: Lundbeck (advisory committee), Pfizer (Scientific Advisory Board member), and Roche (grant recipient, speaker reimbursement). G Breen reports consultancy and speaker fees

from Eli Lilly and Illumina and grant funding from Eli Lilly. All other authors declare no financial interests or potential conflicts of interest.

Abstract

Variance in IQ is associated with a wide range of health outcomes, and 1% of the population are affected by intellectual disability. Despite a century of research, the fundamental neural underpinnings of intelligence remain unclear. We integrate results from genome-wide association studies (GWAS) of intelligence with brain tissue and single cell gene expression data to identify tissues and cell types associated with intelligence. GWAS data for IQ (N = 78,308) were meta-analyzed with an extreme-trait cohort of 1,247 individuals with mean IQ ~170 and 8,185 controls. Genes associated with intelligence implicate pyramidal neurons of the somatosensory cortex and CA1 region of the hippocampus, and midbrain embryonic GABAergic neurons. Tissue-specific analyses find the most significant enrichment for frontal cortex brain expressed genes. These results suggest specific neuronal cell types and genes may be involved in intelligence and provide new hypotheses for neuroscience experiments using model systems.

Genome-wide association studies (GWAS) have successfully identified statistical associations with a wide range of behavioral phenotypes and neuropsychiatric disorders [1–3]. Increasing sample sizes has begun to yield findings for intelligence [4–6]. The largest and most recent study of intelligence reported 18 loci significantly associated with intelligence [4]. Significant genetic correlations were observed between intelligence and a variety of behavioral (educational attainment, smoking behaviors), anthropometric (cranial morphology, height, body composition), and psychiatric phenotypes (schizophrenia, autism, depressive symptoms), mirroring epidemiological evidence for correlations between intelligence and a broad range of health-related outcomes [4,7].

Considered alone, not all associations identified by GWAS precisely localize biological mechanisms amenable to subsequent experimentation. For instance, the most associated variant in a significant locus may not be the causal variant [8,9], there may be multiple causal variants in a locus [10], loci may act through altering the expression of distant genes [11], and the associations identified by GWAS of complex traits are often spread across the genome [12–14]. To extract meaningful biological inferences from GWAS results, it is necessary to integrate data from other sources, such as studies of gene expression [15]. Results from gene-wise analyses in Sniekers et al. (2017) identified expression predominant in the brain for 14 of the 44 genes with significant association, although some transcription was inferred for most genes across most tissues [4]. Integration of genomic results with data on biological pathways suggested a prominent role for nervous system development in intelligence.

In silico functional annotation of GWAS results is dependent on high-quality biological reference data. Recently, data from the Karolinska Institutet (KI) mouse

superset of single-cell RNA sequencing (scRNAseq) of ~10,000 single cells from multiple brain regions was used to map schizophrenia GWAS results to brain cell types [16]. Genes that previously showed association with schizophrenia were expressed with higher specificity in pyramidal cells, medium spiny neurons, and interneurons than in 20 other brain cell types. This demonstrates the potential of cell-type specific annotation to enable the construction of new functional hypotheses for complex traits.

We sought to develop a better understanding of the neurobiological underpinnings of intelligence through combining GWAS results with a number of data sources concerned with tissue and cell-type specific gene expression. To this end, we meta-analyzed the most recent GWAS of intelligence [4] with an extreme-trait GWAS that compared individuals of very high intelligence to a group from the general population (HiQ) [17]. We then analyzed the enrichment of associations with intelligence in single cell expression data from the KI mouse superset [16], and in brain genomic and transcriptomic data from the GTEx project [18]. Finally, we combined genomic and tissue-specific expression data with information from biological, disease-relevant and drug-target pathway databases to further assess the potential impact of biological mechanisms explaining variance in intelligence.

Results

Meta-analysis

The genetic correlation between Sniekers et al. (2017) and HiQ was 0.92 (SE: 0.07) and was sufficiently high to justify meta-analysis. 25 loci met genome-wide significance with intelligence (i.e., $p < 5 \times 10^{-8}$; Table 1, Figure 1a). Of these, 13 were genome-wide significant in Sniekers et al. (2017) and 12 were novel (Table 1; Supplementary Table 1). The single locus previously identified in the HiQ sample was not genome-wide significant in this analysis ($p = 0.00595$; [17]).

Assessment of genome-wide inflation yielded a median λ_{GC} of 1.24. The LD score regression intercept from was 1.004 (SE: 0.01), suggesting that this inflation is caused by polygenicity rather than confounding (Figure 1b) [19]. Annotation of specific genomic loci to databases of interest suggested overlapping loci between intelligence and a variety of phenotypes (Supplementary Table 2). The largest overlap was observed with educational attainment (14/25 loci), but overlap at multiple loci was widespread, including with age at menarche, height, body mass index, autoimmune disease, and schizophrenia. Genes previously implicated in intellectual disability or developmental delay (ID/DD) were present within 9/25 loci.

Heritability and partitioned heritability

LD Score regression yielded a SNP-heritability estimate of 0.221 (SE: 0.01) in line with the previously reported SNP-heritability in Sniekers et al (2017). Partitioning this heritability across 58 functional SNP annotations identifies conserved regions as significantly enriched contributors to the heritability of intelligence (proportion of heritability = 0.340, enrichment = 13.3 fold, $p = 3.26 \times 10^{-8}$), consistent with previous

Locus	Chr	Locus BP	Index SNP	Pos	A1	A2	Dir	FreqA1	Zscore	p
1	1	22354246-22470115	rs7526484	22451845	T	C	++	0.257	5.46	4.82×10^{-8}
2	**1**	**41748897-41850182**	**rs12744310**	**41773502**	**T**	**C**	**--**	**0.208**	**-5.51**	**3.63×10^{-8}**
3	1	98149785-98660622	rs72737821	98581337	A	T	++	0.712	5.48	4.37×10^{-8}
4	1	103548734-103703521	rs6577362	103632361	A	G	--	0.666	-5.77	8.05×10^{-9}
5	**2**	**71434829-71701353**	**rs6745907**	**71609073**	**A**	**G**	**--**	**0.433**	**-5.82**	**5.78×10^{-9}**
6	**2**	**100576304-101031561**	**rs13010010**	**100852734**	**T**	**C**	**++**	**0.390**	**6.57**	**5.09×10^{-11}**
7	**2**	**144152539-144272229**	**rs13428598**	**144250487**	**T**	**C**	**++**	**0.375**	**5.85**	**4.95×10^{-9}**
8	2	161903399-162071338	rs10930013	162070325	A	G	--	0.475	-5.55	2.87×10^{-8}
9	2	217286343-217397419	rs13411858	217306698	T	G	--	0.843	-5.57	2.60×10^{-8}
10	**3**	**23839884-24207650**	**rs7646501**	**24079291**	**A**	**G**	**++**	**0.749**	**6.67**	**2.62×10^{-11}**
11	3	49381898-50247824	rs1987628	49399259	A	G	++	0.319	5.49	4.01×10^{-8}
12	4	106066982-106316427	rs2647257	106199505	A	T	++	0.636	5.54	2.98×10^{-8}
13	5	176855627-176898619	rs2545797	176864202	T	C	++	0.455	5.53	3.12×10^{-8}
14	**6**	**98310291-98591810**	**rs12206087**	**98582900**	**A**	**G**	**++**	**0.504**	**6.8**	**1.05×10^{-11}**
15	**6**	**108856378-109020032**	**rs768023**	**108876002**	**A**	**G**	**++**	**0.564**	**7.2**	**5.84×10^{-13}**
16	**7**	**32209459-32399658**	**rs10236197**	**32291761**	**T**	**C**	**++**	**0.654**	**6.03**	**1.61×10^{-9}**
17	**7**	**32472099-32831680**	**rs6949851**	**32493807**	**T**	**G**	**++**	**0.675**	**5.45**	**4.97×10^{-8}**
18	**7**	**133130940-133727497**	**rs4728302**	**133630463**	**T**	**C**	**--**	**0.582**	**-5.66**	**1.48×10^{-8}**
19	9	23344737-23407386	rs11794152	23345347	A	G	--	0.589	-6.23	4.72×10^{-10}
20	**9**	**72014104-72213799**	**rs11138902**	**72103314**	**A**	**G**	**++**	**0.537**	**6.02**	**1.77×10^{-9}**
21	**13**	**106580722-106688603**	**rs2251499**	**106639856**	**T**	**C**	**++**	**0.250**	**6.85**	**7.60×10^{-12}**
22	14	103987140-104018455	rs10149470	104017953	A	G	--	0.486	-5.49	4.04×10^{-8}
23	**16**	**28490517-28917746**	**rs12928404**	**28847246**	**T**	**C**	**++**	**0.663**	**5.59**	**2.26×10^{-8}**
24	**20**	**47413079-47938626**	**rs6095360**	**47532536**	**A**	**G**	**--**	**0.693**	**-6.16**	**7.26×10^{-10}**
25	22	51089716-51151631	rs5770820	51150473	A	G	--	0.209	-5.63	1.82×10^{-8}

Table 1: Genome-wide significant variants from single-variant association analyses.
Loci significant in Sniekers et al. (2017) are in bold.
FreqA1 = A1 frequency in non-Finnish samples from the 1000 Genomes project.
Dir = direction of effect in Sniekers and in TIP.

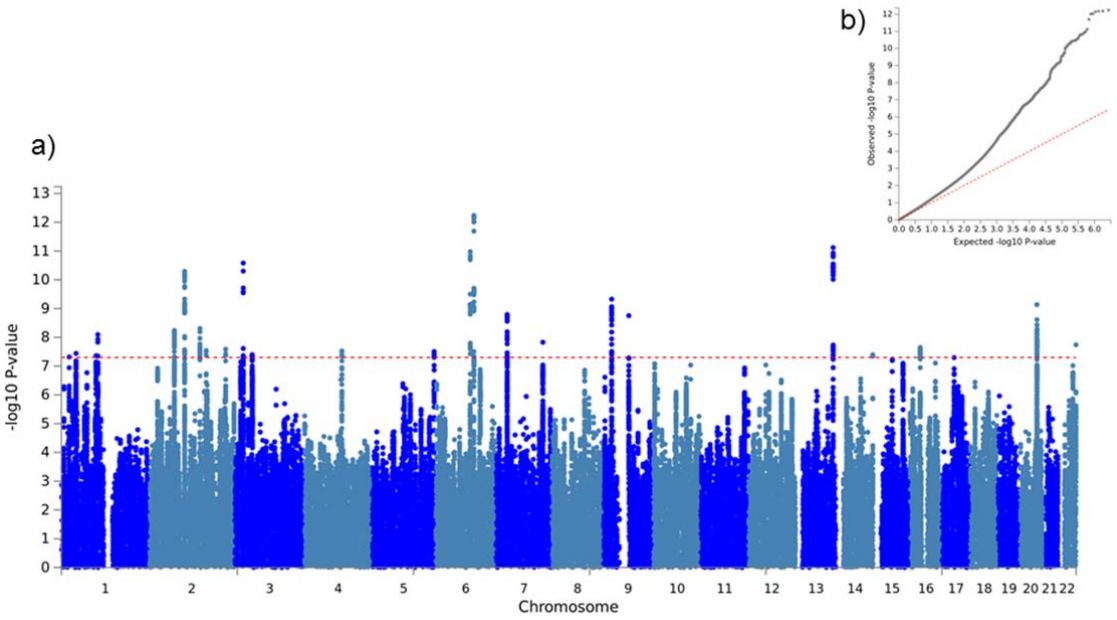

Figure 1: a) Manhattan plot and b) QQ plot of meta-analysis results

reports in a subset of our meta-analyzed cohorts [20]. Four extended annotations were also significantly enriched ($p < 8.62 \times 10^{-4}$; Figure 2), suggesting that genetic variation located in the vicinity of conserved regions, enhancers (specifically H3K4me1 elements) and open chromatin in brain dorsolateral prefrontal cortex [21] and fetal cells) are enriched in heritability for intelligence.

Gene-wise analyses

93 genes at 41 loci were identified at genome-wide significance ($p < 2.65 \times 10^{-6}$, Bonferroni threshold for 18,839 genes; Table 2, Supplementary Figure 1). 28 of these genes (20 loci) were also identified in Sniekers et al. (2017). 11 of the 93 genes were previously implicated in intellectual disability or developmental delay, although this overlap does not significantly from what is expected under the null hypothesis of no enrichment ($p = 0.073$, hypergeometric test; Table 2).

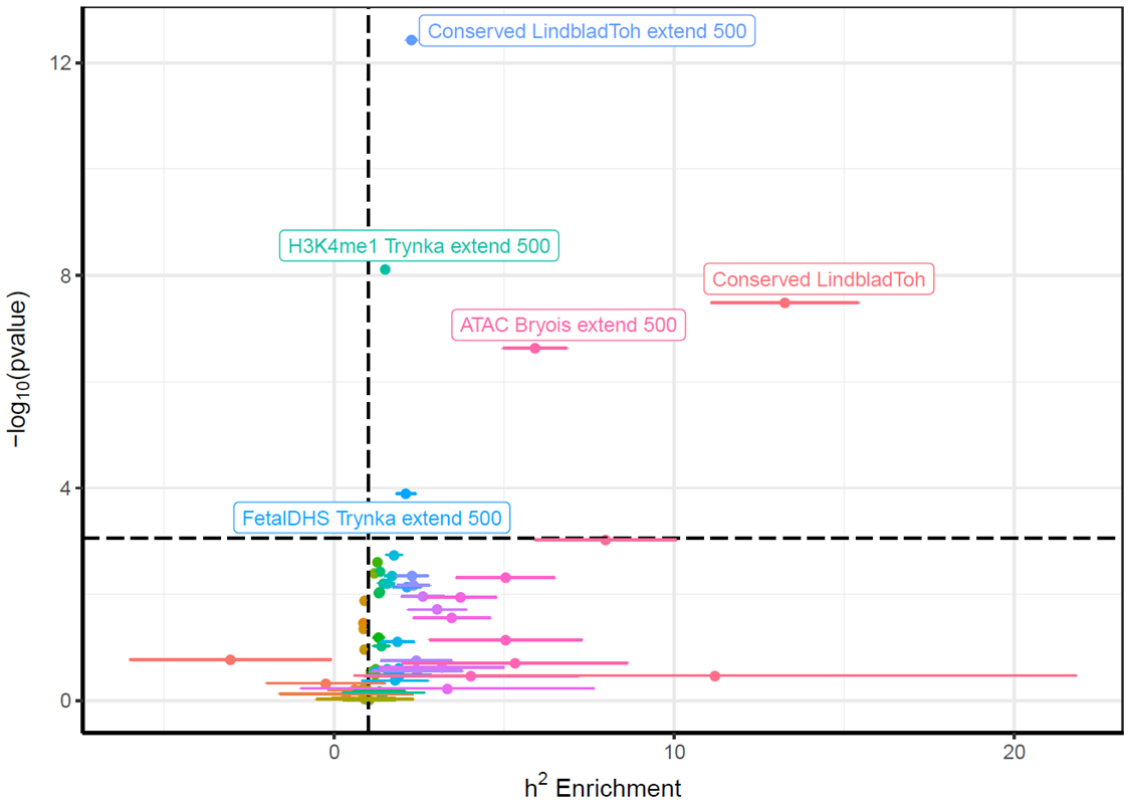

Figure 2: Heritability enrichment of genomic annotations. Horizontal line is p = 8.62×10^{-4}, Bonferroni-corrected threshold for 58 annotations. Vertical line is enrichment = 1 (that is, no enrichment)

Locus	Gene Names	Chr	Pos	SNPS	P
1	CDC42 ***WNT4***	**1**	**22347138-2479067**	188 **197**	5.32×10^{-7} $\mathbf{6.33\times10^{-7}}$
2	FOXO6	1	41431939-41830086	130	1.86×10^{-8}
Chr1:71.8-72.8 Mb	NEGR1	1	71861623-72748417	1424	1.09×10^{-6}
4	*COL11A1*	2	103403088-103773269	1027	4.25×10^{-7}
5	**ZNF638** DYSF	**2**	**71381822-71702582**	**492** 621	$\mathbf{1.41\times10^{-8}}$ 3.48×10^{-7}
7	**ARHGAP15**	**2**	**144151161-144356509**	**1088**	$\mathbf{2.10\times10^{-7}}$
8	DPP4	2	161868668-162319725	100	1.25×10^{-6}
Chr3:16.8-17.2 Mb	PLCL2	3	16844159-17132086	712	2.30×10^{-9}
10	*THRB* **RPL15**	**3**	**23961762-24190184**	818 **86**	1.12×10^{-8} $\mathbf{3.04\times10^{-7}}$
11	TCTA *AMT* RHOA *DAG1* NICN1 TRAIP BSN GPX1 CTD-2330K9.3 AMIGO3 *GMPPB* **IP6K1** MST1 MON1A CAMKV **RNF123** FAM212A CDHR4 MST1R UBA7 RBM6 GNAT1 APEH USP4 SEMA3F	**3**	**48725014-50125108**	68 60 127 132 60 70 173 51 57 53 53 **116** 50 77 52 **72** 59 60 54 52 256 66 52 104 93	2.03×10^{-9} 8.39×10^{-9} 1.88×10^{-8} 3.26×10^{-8} 3.78×10^{-8} 3.97×10^{-8} 5.25×10^{-8} 5.86×10^{-8} 7.08×10^{-8} 1.50×10^{-7} 1.50×10^{-7} $\mathbf{1.58\times10^{-7}}$ 1.65×10^{-7} 1.66×10^{-7} 1.91×10^{-7} $\mathbf{2.13\times10^{-7}}$ 2.25×10^{-7} 2.52×10^{-7} 2.81×10^{-7} 3.69×10^{-7} 4.33×10^{-7} 9.27×10^{-7} 9.76×10^{-7} 1.25×10^{-6} 1.46×10^{-6}
12	TET2	4	106044468-106428563	289	2.22×10^{-6}
Chr5:139.5-139.6 Mb	CYSTM1	5	139554227-139661637	214	1.45×10^{-6}

13	DBN1 **GRK6** F12 **PRR7**	**5**	**176848019-176909800**	35 **45** 23 **44**	6.60×10^{-9} $\mathbf{1.19\times10^{-8}}$ 1.31×10^{-7} $\mathbf{2.07\times10^{-7}}$
Chr6:26.5-26.6Mb	BTN1A1	6	26501449-26510650	139	1.74×10^{-6}
15	**FOXO3**	**6**	**108856378-109020332**	**211**	$\mathbf{1.35\times10^{-12}}$
17	**PDE1C**	**7**	**32259273-32858489**	**1634**	$\mathbf{1.75\times10^{-9}}$
Chr7:86.2-86.5 Mb	GRM3	7	86273230-86494200	424	1.13×10^{-6}
18	**EXOC4**	**7**	**132950174-133734597**	**1171**	$\mathbf{9.61\times10^{-8}}$
20	**APBA1**	**9**	**72014104-72213075**	**365**	$\mathbf{7.88\times10^{-9}}$
Chr10:64.9-65.3 Mb	**JMJD1C**	**10**	**64926981-65225722**	**525**	$\mathbf{1.99\times10^{-6}}$
Chr10:94.0-94.2 Mb	MARCH5	10	94050920-94113721	110	1.55×10^{-7}
Chr11:133.7-133.9 Mb	IGSF9B	11	133778459-133826880	170	1.50×10^{-6}
Chr11:123.9-124.1 Mb	OR10D3 VWA5A	11	123986069-124056952	136 184	1.92×10^{-6} 2.24×10^{-6}
Chr11:124.1-124.2 Mb	OR8D2	11	124189134-124190184	113	2.16×10^{-6}
Chr12:49.3-49.5 Mb	PRKAG1 *KMT2D* **DDN** RHEBL1	**12**	**49388932-49463808**	36 60 **30** 42	4.33×10^{-7} 4.43×10^{-7} $\mathbf{5.76\times10^{-7}}$ 2.48×10^{-6}
Chr12:93.1-93.4 Mb	**EEA1** PLEKHG7	**12**	**93115281-93323107**	**351** 155	$\mathbf{5.69\times10^{-8}}$ 1.54×10^{-6}
Chr12:123.4-123.7 Mb	PITPNM2	12	123468027-123634562	174	2.46×10^{-6}
Chr14:69.5-69.7 Mb	**DCAF5**	**14**	**69517598-69619867**	**190**	$\mathbf{1.37\times10^{-6}}$
22	CKB BAG5	14	103833065-104186052	75 73	2.47×10^{-6} 2.64×10^{-6}
Chr15:82.4-82.6 Mb	**EFTUD1** FAM154B	**15**	**82422571-82577271**	**414** 154	$\mathbf{4.08\times10^{-8}}$ 4.09×10^{-8}
Chr16:12.9-13.4 Mb	SHISA9	16	12995477-13334272	952	5.33×10^{-7}
23	**ATXN2L** ***TUFM*** **SH2B1** **ATP2A1**	**16**	**28298418-29008079**	**47** **57** **92** **66**	$\mathbf{2.27\times10^{-8}}$ $\mathbf{7.72\times10^{-8}}$ $\mathbf{1.87\times10^{-7}}$ $\mathbf{5.37\times10^{-7}}$
Chr16:72.0-72.1 Mb	*DHODH*	16	72042487-72058954	202	9.60×10^{-8}

Chr17:34.5-35.0 Mb	GGNBP2 DHRS11 MYO19 TBC1D3H TBC1D3G	17	34581086-34957235	85 28 112 4 4	$7.38x10^{-7}$ $8.31x10^{-7}$ $1.57x10^{-6}$ $2.28x10^{-6}$ $2.28x10^{-6}$
Chr18:12.9-13.0 Mb	SEH1L	18	12947132-12987535	181	$2.35x10^{-6}$
Chr18:49.8-51.1 Mb	**DCC**	**18**	**49866542-51057784**	**3737**	**$5.86x10^{-8}$**
Chr19:31.7-31.9 Mb	TSHZ3	19	31765851-31840453	166	$3.64x10^{-7}$
24	**CSE1L** **STAU1** ***ARFGEF2*** PREX1 DDX27	**20**	**47422577-47933479**	**206** **213** **364** 658 156	**$1.14x10^{-10}$** **$5.86x10^{-10}$** **$7.58x10^{-10}$** $8.50x10^{-8}$ $4.65x10^{-7}$
Chr22:31.7-31.8 Mb	LINC01521	22	31742875-31744670	72	$1.69x10^{-6}$
Chr22:42.3-42.5 Mb	**WBP2NL** **SEPT3** FAM109B SMDT1	**22**	**42372276-42480288**	**157** **143** 50 45	**$1.78x10^{-7}$** **$6.95x10^{-7}$** $1.56x10^{-6}$ $1.78x10^{-6}$
25	***SHANK3*** ACR	**22**	**51083118-51155826**	**162** 66	**$2.22x10^{-8}$** $7.08x10^{-7}$

Table 2: Genome-wide significant genes from gene-wise association analyses. Genetic locus is that identified in single-variant analyses, or the genomic region otherwise. P value from competitive test in MAGMA. Loci in bold were significantly associated in Sniekers et al. (2017). Genes in italics have previously been implicated in developmental delay or intellectual disability.

Tissue- and cell-specific gene expression

Tissue-specific enrichment analysis identified an enrichment of genes with high brain-specific expression associated with intelligence ($p_{MAGMA} = 4.43 \times 10^{-9}$, $p_{LDSC} = 4.23 \times 10^{-6}$; Figure 3a, Supplementary Table 3a). Across 10 brain regions in GTEx[18], stronger gene associations with intelligence were associated with increased specificity of gene expression to the frontal cortex ($p_{MAGMA} = 0.00305$, $p_{LDSC} = 2.66 \times 10^{-4}$; Figure 3b, Supplementary Table 3b).

In the KI level 1 (broad cell groups) cell-type specific analyses, both linear regression (MAGMA) [22] and heritability enrichment-based analyses (LD Score) [23] supported enrichment of genes with high specificity to pyramidal neurons in the somatosensory neocortex ($p_{MAGMA} = 1.41 \times 10^{-6}$, $p_{LDSC} = 5.81 \times 10^{-4}$) and in the CA1 region of the hippocampus ($p_{MAGMA} = 9.08 \times 10^{-5}$, $p_{LDSC} = 1.12 \times 10^{-3}$), as well as to midbrain embryonic GABAergic neurons ($p_{MAGMA} = 9.47 \times 10^{-5}$, $p_{LDSC} = 1.61 \times 10^{-3}$; Figure 3c, Supplementary Table 3c). Level 2 analyses (narrowly-defined cell types) suggested significant enrichment of genes with high specificity to type pyramidal cells of the CA1 region ($p_{MAGMA} = 2.16 \times 10^{-4}$, $p_{LDSC} = 1.19 \times 10^{-4}$), although considerable variability was observed between methods at this level of granularity (Figure 3d, Supplementary Table 3d).

Cell-type specific analyses were repeated conditioning on each enriched cell-type in turn. When controlling for gene expression in pyramidal neurons of the somatosensory neocortex, the previously observed enrichment in CA1 pyramidal cells is lost. In contrast, the patterns of enrichment observed when conditioning on expression in CA1 pyramidal cells or in midbrain embryonic GABAergic neurons were consistent with those observed without conditioning (Figure 4).

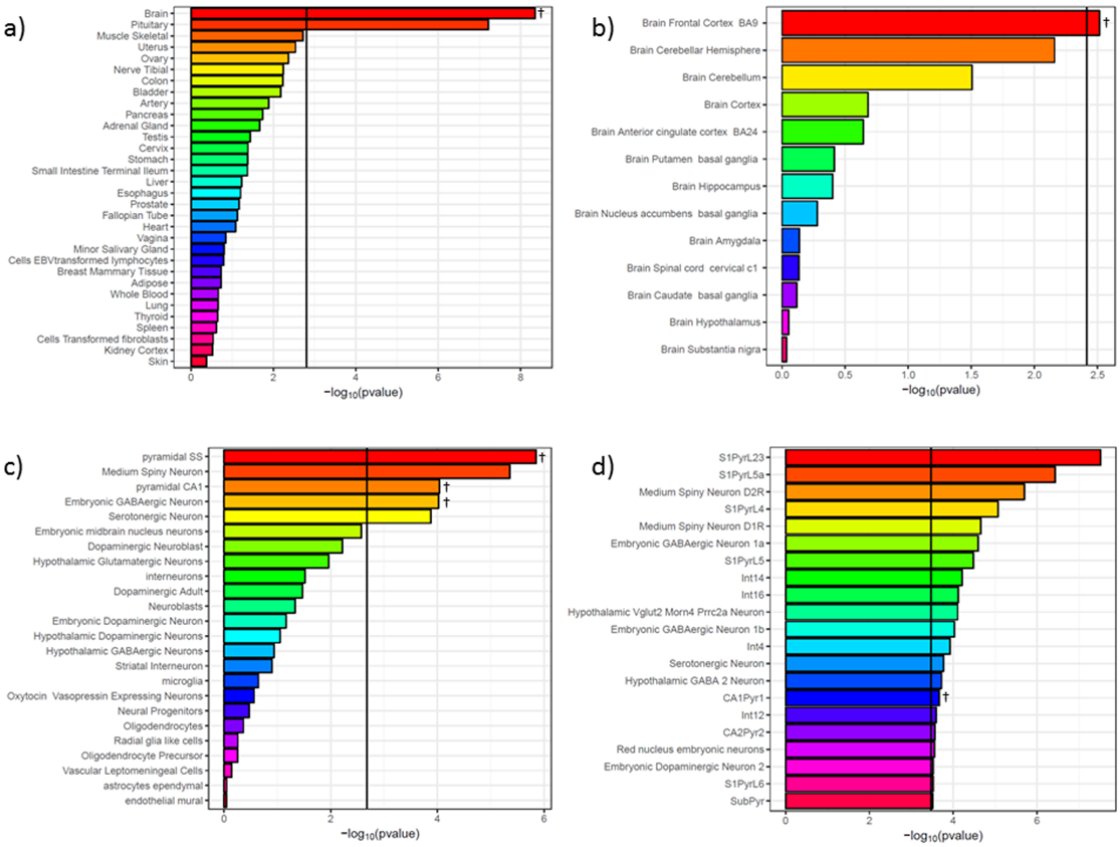

Figure 3: Results of tissue- and cell-specific analyses. a) Whole-body analyses; b) Brain tissue analyses; c) KI level 1 cell analyses; d) KI level 2 cell analyses (only those significant in MAGMA competitive analyses after correction for multiple testing shown). Vertical lines are the Bonferroni threshold for each analysis. Results shown are from MAGMA analyses - tissues and cell-types also significant in LDSC analyses are indicated with †. Full results are shown in Supplementary Table 3.

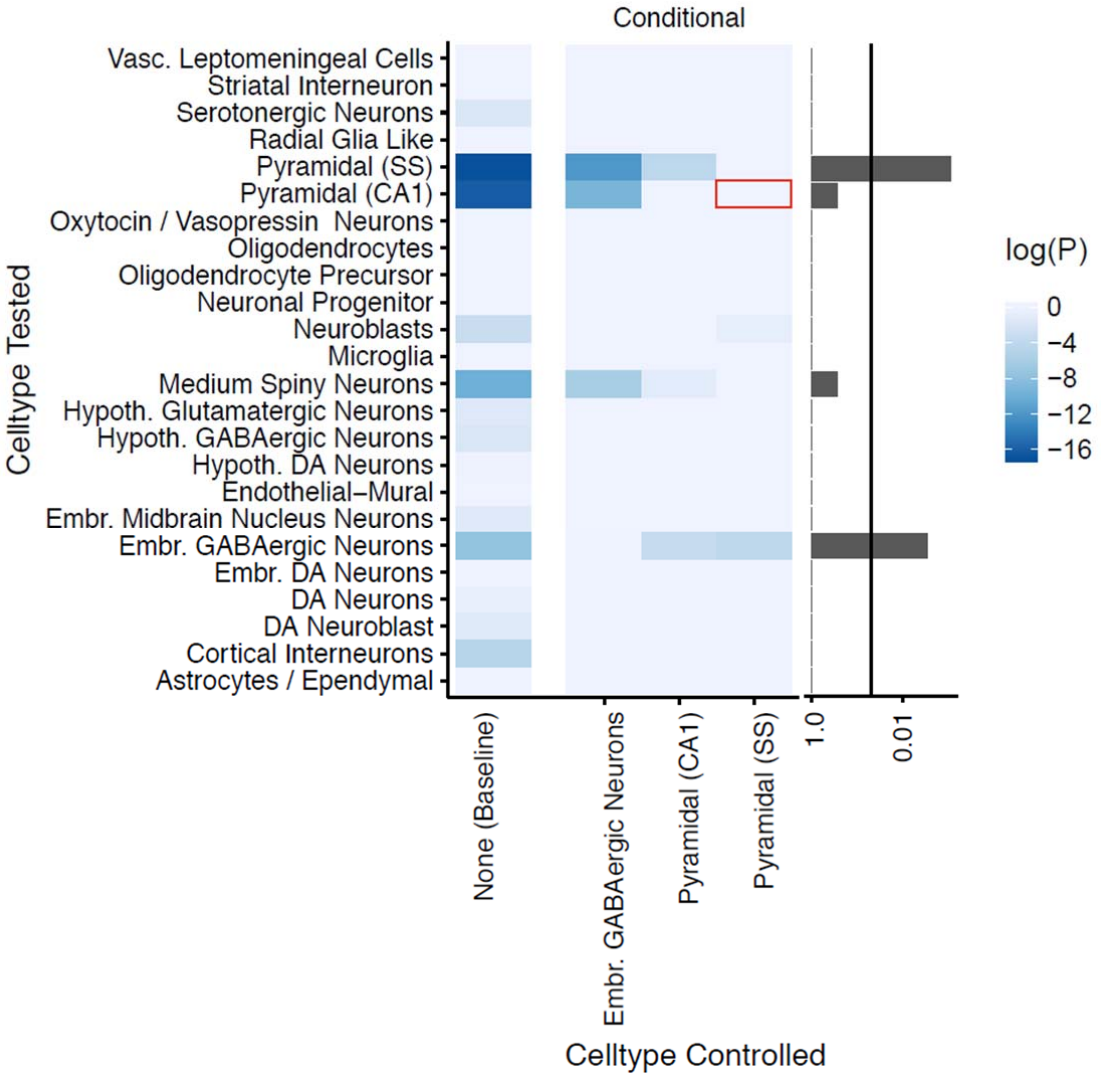

Figure 4: Conditional cell-type enrichment analysis. X-axis lists target cell-types. Y-axis lists other cell-types. Colors correspond to the enrichment probability of the other cell type after conditioning (P). Values of log(P) approaching zero indicate no enrichment after conditioning. Barplot on the right shows the minimum value of P for each cell-type across all conditional analyses (excluding analyses of the target cell-type with itself); the vertical line marks p=0.05. Red box highlights the loss of significant enrichment in CA1 pyramidal neurons when conditioning on somatosensory (SS) neurons.

Predicted tissue-specific gene expression

Across the 10 GTEx brain tissues, results from MetaXcan suggested significant effects on the expression of 16 genes ($p < 1.60 \times 10^{-6}$; Table 3). Eight of these genes are situated at locus 11 (Chr3, 48.7-50.2 Mb). Genetic variation was predicted to upregulate 6 genes and downregulate 10 genes, with both single-region (9 genes) and multiple-region (7 genes) patterns of altered expression implied. Three genes (*NAGA, TUFM* and *GMPPB*) have previously been implicated in ID.

Tissue-specific pathway analysis

Tissue-specific pathway analyses identified 32 nested pathways with $p \leq 5.34 \times 10^{-6}$ (Bonferroni correction for 9,361 effectively independent pathways), of which 29 contained genes that were significant in the gene-wise analysis (Supplementary Table 4; Supplementary Figure 2). 7 pathways remained significant after further correcting for all tissue-specific tests: "self-reported educational attainment", "modulation of synaptic transmission", "neurodegenerative disease", "neuron spine", "schizophrenia", "rare genetic neurological disorder", and "potentially synaptic genes" (Table 4).

226

Locus	Gene	Tissues	Direction	P	Genewise?
11	RNF123	Putamen basal ganglia Nucleus accumbens basal ganglia Frontal cortex Cerebellar Hemisphere	- - - -	5.77×10^{-7} 3.98×10^{-10} 1.84×10^{-7} 6.37×10^{-7}	No
11	NPRL2	Cortex	+	1.06×10^{-7}	No
11	MST1	Hypothalamus	+	1.33×10^{-7}	Yes
11	MST1R	Cerebellum Cerebellar Hemisphere	- -	2.25×10^{-7} 8.98×10^{-7}	Yes
11	RBM6	Frontal cortex	+	2.28×10^{-7}	Yes
11	FAM212A	Cerebellar Hemisphere	-	4.40×10^{-7}	Yes
11	AMIGO3	Cortex	-	1.31×10^{-6}	Yes
11	*GMPPB*	Cortex	-	1.38×10^{-6}	Yes
Chr12: 58.3-58.4 Mb	XRCC6BP1	Nucleus accumbens basal ganglia Hypothalamus Cortex Caudate basal ganglia	+ + + +	1.15×10^{-6} 1.50×10^{-6} 9.27×10^{-7} 9.33×10^{-7}	No
15	FOXO3	Nucleus accumbens basal ganglia	-	4.93×10^{-10}	Yes
23	*TUFM*	Putamen basal ganglia Nucleus accumbens basal ganglia	- -	4.89×10^{-7} 1.38×10^{-7}	Yes
23	NPIPB6	Caudate basal ganglia	-	1.33×10^{-6}	No
Chr17:34.9-35.0 Mb	DHRS11	Putamen basal ganglia Nucleus accumbens basal ganglia Caudate basal ganglia	- - -	3.75×10^{-7} 1.39×10^{-6} 1.09×10^{-6}	No
Chr18:12.9-13.2Mb	CEP192	Cortex	+	9.41×10^{-8}	No
Chr22:42.4-42.5 Mb	*NAGA*	Putamen basal ganglia Cerebellum Cerebellar Hemisphere	- - -	1.29×10^{-7} 2.67×10^{-7} 2.19×10^{-7}	No
Chr22:42.5-42.6 Mb	CYP2D6	Hypothalamus	+	5.70×10^{-7}	No

227

228 Table 3: Genes predicted to be upregulated (+) or downregulated (-) in specific
229 tissues. Genetic locus is that identified in single-variant analyses, or the genomic
230 region otherwise. Genewise = genome-wide significant in gene-wise analyses.
231 Genes in italics have previously been implicated in developmental delay or
232 intellectual disability.

233

Pathway Name	Tissues significant	Significant genes	Min p [Tissue]
Self Reported Educational Attainment	Adipose, Artery, Bladder, Brain, Breast Mammary Tissue, Cells Transformed fibroblasts, Cervix, Colon, Fallopian Tube, Liver, Lung, Nerve Tibial, Ovary, Unweighted, Small Intestine Terminal Ileum, Spleen, Uterus, Whole Blood, Amygdala, Anterior cingulate cortex (BA24), Caudate (basal ganglia), Cortex, Frontal Cortex (BA9), Hippocampus, Hypothalamus, Nucleus accumbens (basal ganglia), Putamen (basal ganglia), Spinal cord (cervical c1)	ATXN2L, RNF123, NEGR1, DPP4, BTN1A1, JMJD1C	1.03×10^{-9} [Frontal Cortex (BA9)]
GO:Neuron Spine	Artery, Bladder, Cells EBVtransformed lymphocytes, Cervix, Esophagus, Lung, Minor Salivary Gland, Unweighted, Uterus, Vagina, Cerebellar Hemisphere, Cerebellum, Hypothalamus	*ARFGEF2, SHANK3,* EEA1, GRM3	5.29×10^{-8} [Cerebellum]
GO:Modulation Of Synaptic Transmission	Artery, Bladder, Cells EBVtransformed lymphocytes, Cervix, Esophagus, Ovary, Thyroid, Uterus, Vagina, Hippocampus	STAU1, PLCL2, DBN1, *SHANK3*, SHISA9, GRM3, IGSF9B	2.77×10^{-8} [Uterus]
Schizophrenia	Brain, Pituitary, Testis, Cortex, Frontal Cortex .BA9.	FOXO3, *ARFGEF2*, PLCL2, DBN1, APBA1, *AMT*, *THRB*, GRK6, FOXO6, *SHANK3*, GPX1, DCC, EXOC4, F12, CDC42, SHISA9, NEGR1, GRM3, DPP4, IGSF9B	5.57×10^{-8} [Pituitary]
Rare Genetic Neurological Disorder	Brain, Nerve Tibial, Pituitary	FOXO3, CSE1L, *ARFGEF2*, PDE1C, DBN1, *AMT*, *THRB*, GRK6, ZNF638, FOXO6, RHOA, *SHANK3*, ATXN2L, *DAG1*, TRAIP, BSN, EEA1, GPX1, DCC, *TUFM*, *DHODH*, EXOC4, F12, *GMPPB*, MARCH5, SH2B1, MST1R, DYSF, TSHZ3, *COL11A1*, PRKAG1, *KMT2D*, CDC42, ATP2A1, DDN, *WNT4*, ACR, APEH, GRM3, DPP4, CYSTM1, SEMA3F, JMJD1C, TET2, CKB	9.09×10^{-8} [Brain]

Neurodegenerative Disease	Testis	FOXO3, CSE1L, *ARFGEF2*, PDE1C, DBN1, APBA1, *THRB*, GRK6, ZNF638, *SHANK3*, ATXN2L, *DAG1*, BSN, EEA1, GPX1, DCC, *TUFM*, EXOC4, F12, MARCH5, DYSF, *COL11A1*, *KMT2D*, CDC42, ATP2A1, ACR, APEH, NEGR1, GRM3, DPP4, CYSTM1, SEMA3F, JMJD1C, VWA5A, SEH1L	4.75×10^{-8} [Testis]
Potentially Synaptic All	Brain	FOXO3, CSE1L, *ARFGEF2*, PDE1C, PLCL2, DBN1, APBA1, *THRB*, ZNF638, FOXO6, RHOA, *SHANK3*, ATXN2L, EFTUD1, BSN, EEA1, GPX1, DCC, *TUFM*, PREX1, EXOC4, IP6K1, SH2B1, CAMKV, PRR7, ARHGAP15, TSHZ3, RBM6, *KMT2D*, CDC42, SHISA9, ATP2A1, DDN, SEPT3, NEGR1, GRM3, USP4, DPP4, DCAF5, IGSF9B, JMJD1C, TET2, VWA5A, SEH1L, PITPNM2, CKB	1.16×10^{-7} [Brain]

Table 4: Pathways significantly enriched for genes associated with intelligence after correction for all tissue-effective pathways tests. Genes in italics have previously been implicated in developmental delay or intellectual disability.

240 ***Conditional gene set enrichment***

241 Results from tissue-specific and gene-set analyses identified a number of
242 gene sets associated with intelligence. Of specific interest were synaptic genes
243 (post-synaptic density proteome list; [24]), RBFOX family binding partners [25], CELF4
244 binding partners, and previously reported intellectual disability genes. Additional
245 analyses were performed to assess whether the association of these gene sets was
246 independent of the enrichment for gene expression in pyramidal cells of the
247 somatosensory cortex. Bootstrapped analyses confirmed the association of each
248 gene set prior to conditional analysis (empirical $p < 0.05$; Supplementary Table 5).
249 Conditioning on expression in gene expression in pyramidal cells of the
250 somatosensory cortex, the enrichment for synaptic genes is no longer significant,
251 indicating that this enrichment is not independent of that observed in the pyramidal
252 neurons. This effect was not observed for RBFOX or CELF4 targets. Intellectual
253 disability genes were weakly enriched prior to conditioning, and conditioning had little
254 effect on enrichment. However, subdividing ID/DD genes into those associated with
255 severe and those associated with moderate ID/DD indicated that the enrichment
256 stemmed predominantly from moderate ID/DD genes, and this was not altered by
257 conditioning on somatosensory pyramidal neuron gene expression (Supplementary
258 Table 5).
259
260

Discussion

The overall power of our GWAS meta-analysis was equivalent to a sample size of ~99,000 individuals due to the inclusion of the extreme trait HiQ sample, which contributes equivalently to a population cohort of ~21,000 individuals, and is likely to be enriched for alleles associated with intelligence in the normal range. We mapped results from our GWAS to tissue and cell-type specific gene expression data, identifying enrichment of specificity at multiple levels: in the brain, the frontal cortex, midbrain embryonic GABAergic neurons and pyramidal neurons, especially those in the somatosensory cortex. A number of genes previously implicated in intellectual disability or developmental delay are predicted from the GWAS results to show differential gene expression for normal range IQ in different brain regions, and are associated with variation in intelligence in the normal range from gene-wise analyses.

RNA sequencing data suggest that genes more strongly associated with intelligence are enriched for brain-specific expression in general. While the dominance of brain-specificity over other body tissue specificity is pronounced, assessing differences within the brain is more difficult. Genes more strongly associated with intelligence showed higher specificity for frontal cortical expression, but the differences in cell composition between brain tissues means that cell-type analyses may be more informative. This can be seen within our results, in that pyramidal neurons of the somatosensory cortex were significant in the cell-type specific analysis, but the cortex as a whole is not significant in the GTEx brain-tissue analysis. This is perhaps due to the fact that the cortex is a highly heterogeneous mixture of cell-types. Our results suggest expression in pyramidal neurons in one area of the cortex is relevant in intelligence, but expression in the other cell-types

and areas of the cortex may not be [26]. A further caveat to this interpretation is that the full cellular composition of the cortex (and the brain overall) is not reflected in the KI mouse superset, and as such our conclusions about wider effects must be constrained.

Genes with higher specificity to pyramidal neurons were enriched for associations with intelligence. However, the location of the most interesting neuron population is not yet clear. Observed enrichment in the pyramidal cells of the CA1 region of the hippocampus was lost when accounting for gene expression in pyramidal cells of the somatosensory cortex. An uncaptured population of pyramidal neurons (for example, in the frontal cortex) may exist that similarly overlaps in expression with the somatosensory pyramidal neurons, and accounts for the enrichment observed in the latter population. While the KI superset is the largest brain scRNAseq resource to date, it covers a limited set of regions and developmental stages [16].

Tissue and cell-specific analyses were performed using related but distinct approaches, namely linear regression of binned specificity and heritability enrichment analysis of the top 10% of specific genes. Linear regression tended to give smaller p-values than heritability enrichment analysis (Supplementary Table 3). While this may be a statistical artefact generated by the differing assumptions of the methods, this could also result from the pattern of enrichment. For example, a small set of genes associated with intelligence, all with very high specificity to a given tissue would result in a lower p-value from heritability enrichment than if the association with intelligence was spread more broadly across genes with generally enriched tissue-specificity.

The results of analyses in the KI mouse superset for intelligence can be contrasted with those for schizophrenia [16]. Both phenotypes initially demonstrate enrichment for the same hippocampal and somatosensory pyramidal neuron populations. However, conditional analyses in schizophrenia demonstrated significant differences in the results between schizophrenia and intelligence - the enrichment observed in the somatosensory cortex in schizophrenia could be fully explained by that in the hippocampus, while the opposite pattern was observed in intelligence. Additionally, unlike in schizophrenia, genes implicated in intelligence also showed specificity in midbrain embryonic cells. Analyses in schizophrenia implicated more level two cell types than were significant in our analyses. Such differences may reflect differences in the common-variant contribution to variance in intelligence and schizophrenia; despite a similar effective sample size to the schizophrenia GWAS (40,675 cases, 64,643 controls, N_{eff} ~ 100K), this GWAS identified only 25 loci compared with the 140 associated with schizophrenia [27] highlights the potentially more polygenic nature of the observed heritability of intelligence, which is approximately equal to that of schizophrenia.

The KI mouse superset has a number of strengths; it is the largest and broadest scRNAseq dataset to date, captures extra-nuclear as well as nuclear transcripts, and was generated using identical methods [16]. However, there are also limitations to the use of this dataset. One issue is that the expression data used is derived from mice rather than humans. Gene expression in the brain is conserved across mammals, such that the principal axes of variation in comparative studies of gene expression capture inter-tissue, rather than inter-species, variance [28]. Furthermore, there is a high degree of conservation of gene expression between mouse and human brains specifically [16,29,30]. Previous analyses using the KI mouse

superset have made extensive comparison between mouse and human gene expression and found high concordance in mapping mouse to human genes [16].

Nevertheless, cell types that are not enriched for genes associated with intelligence in this study should not be prematurely excluded, as some cell-types are not present in the dataset and others will have dissimilar functions or have been exposed to different evolutionary pressures in mouse and in human. Intelligence is a major characteristic that differentiates humans from other mammals [31]. As such, it may be the case that genes with higher specificity to regions dissimilar between humans and mice could be enriched for associations with intelligence, which would not be captured by this approach.

Our results highlight potential insights beyond those gleaned from tissue-specific expression patterns. Several genes previously implicated in ID/DD were present in loci associated in the GWAS. Perhaps the most interesting example of this is *GMPPB* (GDP-Mannose Pyrophosphorylase B), which is in locus 11 of the GWAS results, and was genome-wide significant in gene-wise analyses. It is a member of the "rare genetic neurological disorder" gene set (significantly enriched, specifically in neural tissues), and the expression of *GMPPB* was predicted to be downregulated in the cortex of individuals with higher intelligence. Rare loss of function mutations in the *GMPPB* gene have been identified as causal mutations in tens of individuals with muscular dystrophies and myasthenias, many of which present with mild to severe ID [32–36]. The product of this gene is important for the glycosylation of a-DG (alpha-dystroglycan - the dystroglycan gene *DAG* is also present in locus 11 and is a significantly associated gene in this analysis; [32]). Glycosylation is required for the interaction of a-DG with extracellular ligands, with a variety of consequences including the organization of axon guidance [37]. Our results, in the context of the

360 medical genetic literature, tentatively suggest the effects on axon guidance by
361 glycosylated a-DG may be an area worthy of further exploration to understand the
362 biology of intelligence.

363 However, the observed overall overlap of all ID/DD genes with loci from the
364 GWAS does not differ from that expected by chance. This lack of significant
365 enrichment is partly reflected in the pathway analysis - there are several overlapping
366 gene sets designed to capture ID/DD genes. Of these, only "rare genetic
367 neurological disorder" was significantly enriched following correction. Further insight
368 is obtained from conditional gene-set analyses - although the overall "intellectual
369 disability" gene set is only nominally associated with intelligence, stratifying this gene
370 set demonstrates considerable enrichment of mild intellectual disability genes,
371 independent of gene expression in somatosensory pyramidal neurons. The presence
372 of genes causative of ID/DD in loci associated with intelligence in the normal range,
373 and the enrichment of specific pathways and gene sets derived from the ID/DD
374 literature, may indicate shared biology between ID/DD and normal intelligence.
375 However, the lack of broad enrichment of all such genes suggests that there may
376 also be distinct pathways contributing to normal intelligence that are not commonly
377 affected in ID/DD. ID/DD is not a single condition, but a group of disorders with
378 differing etiologies - our results are still consistent with a two-group model of ID/DD
379 etiology [38–40].

380 The meta-analysis results presented herein extend previous findings [4]. The
381 results are largely consistent with those previously reported, which is unsurprising.
382 More genes were identified in our gene-wise analysis due to an analytical decision to
383 extend the boundaries by which each gene is defined 35kb upstream and 10kb
384 downstream of the coding region [41,42]. Defining genes using this boundary (as

385 opposed to no boundary extension) captures additional transcriptional elements -
386 these may be specific to the target gene, but may also capture elements with more
387 distal regulatory effects.

388 Deriving testable biological hypotheses from the statistical associations of
389 GWAS results is one of the central challenges for the immediate future of complex
390 genetics [3,15]. The provision of high-quality reference datasets encompassing genetic
391 information from variation to translation, and the integration of genomic data to such
392 reference data is invaluable to this aim [16,18]. We have demonstrated that some
393 insights into the biology of intelligence can be derived from GWAS, and have
394 suggested potential avenues for further exploration. Our results could indicate that
395 intelligence represents optimal pyramidal neuron functioning. Cognitive tests are
396 highly correlated with general intelligence (*g*), which may depend on pyramidal
397 neuron function [7].

398 Understanding how biology underlies variation in intelligence is an active area
399 of research that is beginning to yield results. Unifying these new genetic results with
400 data from multiple approaches, can increase the power of each approach, has the
401 potential to yield greater understanding of the biology of intelligence, which in turn
402 could inform the study of many health-related phenotypes with which intelligence is
403 correlated.

Online Methods

Cohort descriptions

Sniekers intelligence GWAS [4]

The cohort analyzed in Sniekers et al. (2017) was drawn from 7 cohorts, primarily consisting of data from the UK Biobank pilot genotyping release (N = 54,119) and the Childhood Intelligence Consortium (N = 12,441) as well as seven additional cohorts (N = 11,748). The phenotype for analysis was Spearman's *g*, or a primary measure of fluid intelligence that correlates highly with *g* [4,43]. Summary statistics from this analysis are available at https://ctg.cncr.nl/software/summary_statistics. Full details on cohort characteristics, genotyping and analysis are supplied in the Supplementary Material.

HiQ high-intelligence GWAS

The Duke University Talent Identification Program (TIP) cohort has been described previously [17]. In brief, TIP is a non-profit organization that recruits and nurtures academically gifted children of extremely high intelligence (top 3%) from the US population. For genomic study, 1,247 participants from the top 1% of TIP (top 0.03% of population) were selected as a high-intelligence cohort (HiQ). IQ was inferred from performance on the Scholastic Assessment Test (SAT) or American College Test (ACT) taken at age 12 rather than the usual age of 18 years. A population comparison cohort (N = 8,185) was obtained from the The University of Michigan Health and Retirement Study (HRS). Participants were assumed to be

428 drawn from the normal distribution of intelligence. Full details on genotyping and
429 analysis are supplied in the Supplementary Material.
430
431 ***Meta-analysis***
432 Summary statistics from Sniekers et al. (2017) and HiQ were meta-analyzed
433 using METAL [4,44]. To account for the increased discovery power afforded by the
434 extreme-sampling method used in TIP, analyses were weighted by their respective
435 non-centrality parameters (NCP), estimated using the Genetic Power Calculator [45–
436 47]. Specifically, NCPs were estimated assuming a causal variant of 20% frequency,
437 capturing 0.1% of variance in each phenotype, assuming HiQ controls are drawn
438 from the normal distribution (+/- 2 SD from the mean) and HiQ cases are sampled
439 from 4 SD above the mean, consistent with IQ > 160 in 99% of the cohort [17]. The
440 NCP of the Sniekers cohort (N = 78,308) was 78.4, while the NCP of the HiQ cohort
441 was 21.6, suggesting the HiQ cohort contributes equivalently to a population cohort
442 of ~21,000 individuals. Only variants present in both cohorts were retained for
443 analysis.
444 Following association analysis, genome-wide significant loci were defined via
445 clumping in PLINK2 [48]. Index variants ($p < 5x10^{-4}$) were merged into loci if in linkage
446 disequilibrium ($r^2 > 0.1$ within 500kb) with a variant with a lower p-value. Loci within
447 50kb of each other were merged. Manhattan and QQ plots were generated using
448 FUMA ([49]. Annotation of genomic results with: data from the EBI GWAS catalog;
449 OMIM; GENCODE genes; genes previously implicated in autism and in intellectual
450 disability; copy-number variants previously implicated in psychiatric disorders; and
451 mouse knockout phenotypes was performed with RegionAnnotator version 1.63
452 (https://github.com/ivankosmos/RegionAnnotator).

453

454 ***Heritability and partitioned heritability***

455 The heritability of intelligence accounted for by common variants was
456 estimated using LD Score, limited to the HapMap3 variants and pre-computed LD
457 scores provided with the package [19]. Heritability was then partitioned across the 53
458 genomic annotations provided with the package [23], with the addition of 5 annotations:
459 open chromatin regions (ATAC and ATAC Bryois extend 500, which increases the
460 window around the region by 500 bases in both directions), the intersection between
461 ATAC and conserved regions of the genome (ATAC-conserved) and regions present
462 in the Neanderthal genome (Neanderthal and Neanderthal extend 500; [21,50]. Regions
463 of open chromatin were identified in prefrontal cortical tissue from 135 schizophrenic
464 individuals and 137 controls using ATAC sequencing, which identifies stretches of
465 DNA free of nucleosomes and other DNA-binding proteins [21,51].

466

467 ***Gene-wise analyses***

468 Results from the meta-analysis were filtered to retain only single nucleotide
469 variants (SNPs) present in the European superpopulation of 1000 Genomes Phase 3
470 [52]. SNPs were annotated to a gene using MAGMA v1.06, assigning SNPs to genes if
471 they lay between 35kb upstream and 10kb downstream of the gene location (as
472 supplied on the MAGMA website, build 37; [22]. Gene-wise p-values were obtained
473 from MAGMA as the aggregate of the mean and smallest p-value across all SNPs
474 annotated to the gene. MAGMA accounts for possible confounders such as gene
475 size, gene density, linkage disequilibrium and minor allele count. The threshold for
476 genome-wide significance was defined as $p = 2.65 \times 10^{-6}$, the Bonferroni correction
477 for the 18,839 genes tested. Genes passing genome-wide significance were defined

as coming from the same locus if their locations were within 50kb of each other, or if they lay within clumped loci from the single variant analysis. Significant genes were cross-referenced to the intellectual disability (ID) gene list provided with RegionAnnotator. The significance of the observed overlap was quantified as a hypergeometric test in R [53], using as background 1,366 ID/DD genes in 18,839 autosomal genes.

Tissue- and cell-specific gene expression

Tissue-specific and cell-type specific proportions of gene expression were calculated following the method described in detail in [16]). Tissue expression data was drawn from the GTEx Consortium [18], and brain cell-type expression data was drawn from scRNAseq data from mouse brain [16]. For each gene, the value for each tissue (or cell-type) was calculated by dividing the mean Unique Molecular Identifier (UMI) counts for the given tissue by the summed mean UMI counts across all tissues [16].

Associations between gene-wise p-values from the meta-analysis and tissue-specific (cell-type specific) gene expression were calculated using two methods, implemented in MAGMA [22] and in LD Score [19]. In MAGMA, genes were grouped into 40 equal bins by specificity of expression, and bin membership was regressed on gene-wise association with intelligence in the meta-analysis (For these analyses, gene-wise association was defined as the mean p-value across all SNPs assigned to the gene.) In LD Score, the 10% of genes with the highest specificity within each tissue were used as a gene set for partitioned heritability analysis. Results were considered significant if the association p-values were smaller than the relevant Bonferroni threshold for both methods.

Conditional cell-specific analyses were performed as a secondary analysis to test whether each enriched cell-type observed was independent of all others. Full details of the method implemented are provided in Skene et al., 2017. In brief, for each enriched cell-type in turn (the target cell-type), z-scores from gene-wise association analyses with intelligence were randomly resampled without replacement. The mean z-score within each expression-specificity decile of the target cell-type was held constant, but the mean z-score of each specificity decile of other cell types was randomized. Empirical p-values were derived for each of the other cell-types, and this procedure was repeated 10,000 times. Expression in each of the other cell-types is considered to be associated with intelligence independently of expression in the target cell type if the observed p-value is lower than the 500th empirical p-value (i.e. 95% of the empirical distribution; [16]).

Predicted tissue-specific gene expression

Results from the variant-level meta-analysis were used to predict gene expression using MetaXcan and genomic and transcriptomic reference data from the brain regions assayed in the GTEx project [18,54]. Associations between predicted gene expression levels and intelligence were calculated. Significance was set at 1.60×10^{-6}, the Bonferroni correction for the 31303 gene-tissue pairs tested [54]. Significant genes were cross-referenced to the intellectual disability gene list provided with RegionAnnotator.

Pathway analysis

A pathway matrix **P** was generated with elements $P_{g,p} = 1$ if gene g was in pathway p and $P_{g,p} = 0$ otherwise. The elements in the matrix were multiplied by

binned gene expression weights obtained from GTEx data (as in the foregoing section on tissue-specific expression) for 13 brain regions and 32 tissues, generating 45 weighted gene/pathway matrices. Only genes with expression data were taken into account. Association between these tissue-weighted pathways and gene-wise associations with intelligence was computed using MAGMA. 13,564 pathways were drawn from OpenTargets (downloaded January 2017; [55]), GO ontologies, canonical pathways drawn from MSigSB v5.2 C2 and C5 datasets [56], and biological pathways related to psychiatric disorders found in various scientific publications (a link to each pathway source is provided in Supplementary Table 4). The pathways assessed by this approach are related to each other in a complex fashion; GO pathways are hierarchical and MSigDB and OpenTargets pathways capture related gene sets. Accordingly, in order to control for multiple testing, the effective number of pathways tested was established by computing the number of principal components accounting for 99.5% of explained variance in the pathway similarity matrix, obtained by computing the Tanimoto similarity between pathways. This results in a Bonferroni-corrected threshold of $p = 5.34 \times 10^{-6}$ for 9,361 effectively independent tests for each matrix. A more stringent threshold was applied secondarily, taking into account all tissue-specific pathway matrices for a total for 9,361 x 46 tests and threshold $p = 1.16 \times 10^{-7}$.

Conditional gene set enrichment

Gene sets of interest were drawn from the results of pathway analyses. Specifically, the human postsynaptic density proteome gene set [24] was used to capture the effects of synaptic and neuronal pathways. Gene sets of RBFOX and CELF4 targets were included as they showed enrichment that appears to be driven

552 by brain-specificity, and have been previously implicated in other brain-related traits
553 [16,25,57]. ID/DD gene sets were tested as they are of specific interest to the study of
554 intelligence. To assess whether the enrichment of these gene sets is independent of
555 gene expression in somatosensory pyramidal neurons, conditional analyses were
556 performed following the method described above for conditional cell-type analyses,
557 modified for the use of gene sets [16].

558 ***Data availability***

559 Summary statistics from the GWAS meta-analysis will be made available at
560 [link available upon publication].

561

562 Acknowledgements

563

564 We gratefully acknowledge the contribution of all of the researchers and participants
565 involved in the collection and analysis of the data included. This study includes data
566 from Sniekers et al., 2017, which made use of the UK Biobank resource under
567 application number 16406 (as previously acknowledged).

568

569 Analysis in this paper represents independent research funded by the National
570 Institute for Health Research (NIHR) Biomedical Research Centre at South London
571 and Maudsley NHS Foundation Trust and King's College London. The views
572 expressed are those of the authors and not necessarily those of the NHS, the NIHR
573 or the Department of Health. Analyses were performed using high performance
574 computing facilities funded with capital equipment grants from the GSTT Charity
575 (TR130505) and Maudsley Charity (980).

576

577 Analysis from Sniekers et al., 2017 was funded by the Netherlands Organization for
578 Scientific Research (NWO VICI 453-14-005), and carried out on the Genetic Cluster
579 Computer, which is financed by the Netherlands Scientific Organization (NWO:480-
580 05-003), by VU University, Amsterdam, the Netherlands, and by the Dutch Brain
581 Foundation and is hosted by the Dutch National Computing and Networking Services
582 SurfSARA. PRJ is supported by the 'Stichting Vrienden van Sophia' (grant nr: 14-27)
583 awarded to DP.

584

585 Research on the HiQ cohort was supported by a European Research Council
586 Advanced Investigator award (295366) to RP. Collecting DNA from the highest-

scoring TIP individuals was supported by an award from the John Templeton Foundation (13575) to RP.

Summary statistics from this analysis have been made available at [link available upon publication].

Author contributions

GB, DP and PFS conceived the study. JRIC, JB, HAG, PRJ, JS and NS performed statistical analyses. RP, ABM, SL, GC and JH acquired data. JRIC and GB wrote the manuscript. All authors reviewed the manuscript.

605 References

606 1. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated
607 with educational attainment. *Nature* **533,** 539–542 (2016).
608 2. Lo, M.-T. *et al.* Genome-wide analyses for personality traits identify six genomic
609 loci and show correlations with psychiatric disorders. *Nat. Genet.* (2016).
610 doi:10.1038/ng.3736
611 3. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *bioRxiv*
612 115600 (2017). doi:10.1101/115600
613 4. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals
614 identifies new loci and genes influencing human intelligence. *Nat. Genet.* (2017).
615 doi:10.1038/ng.3869
616 5. Davies, G. et al. Genome-wide association study of cognitive functions and
617 educational attainment in UK Biobank (N=112 151). Mol. Psychiatry 21, 758–
618 767 (2016).
619 6. Benyamin, B. *et al.* Childhood intelligence is heritable, highly polygenic and
620 associated with FNBP1L. *Mol. Psychiatry* **19,** 253–258 (2013).
621 7. Deary, I. J. Intelligence. *Annu. Rev. Psychol.* **63,** 453–482 (2012).
622 8. Wray, N. R. Allele frequencies and the r2 measure of linkage disequilibrium:
623 impact on design and interpretation of association studies. *Twin Res. Hum.*
624 *Genet.* **8,** 87–94 (2005).
625 9. Carlson, C. S. *et al.* Selecting a maximally informative set of single-nucleotide
626 polymorphisms for association analyses using linkage disequilibrium. *Am. J.*
627 *Hum. Genet.* **74,** 106–120 (2004).
628 10. Hormozdiari, F. *et al.* Widespread Allelic Heterogeneity in Complex Traits. *Am.*
629 *J. Hum. Genet.* **100,** 789–802 (2017).

11. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).
12. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169,** 1177–1186 (2017).
13. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *Am. J. Hum. Genet.* **99,** 139–153 (2016).
14. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47,** 1385–1392 (2015).
15. Breen, G. *et al.* Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19,** 1392–1396 (2016).
16. Skene, N. G. *et al.* Genetic Identification Of Brain Cell Types Underlying Schizophrenia. *bioRxiv* 145466 (2017). doi:10.1101/145466
17. Zabaneh, D. *et al.* A genome-wide association study for extremely high intelligence. *Mol. Psychiatry* (2017). doi:10.1038/mp.2017.121
18. Consortium, G. TEx. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
19. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
20. Hill, W. D. *et al.* Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl. Psychiatry* **6,** e980 (2016).
21. Bryois, J. *et al.* Evaluation Of Chromatin Accessibility In Prefrontal Cortex Of

Schizophrenia Cases And Controls. *bioRxiv* 141986 (2017). doi:10.1101/141986

22. 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).
23. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).
24. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14,** 19–21 (2011).
25. Lee, J. A. *et al.* Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron* **89,** 113–128 (2016).
26. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–1142 (2015).
27. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *bioRxiv* 068593 (2016). doi:10.1101/068593
28. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478,** 343–348 (2011).
29. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167,** 566–580.e19 (2016).
30. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149,** 483–496 (2012).
31. Flinn, M. V., Geary, D. C. & Ward, C. V. Ecological dominance, social competition, and coalitionary arms races. *Evol. Hum. Behav.* **26,** 10–46 (2005).
32. Carss, K. J. *et al.* Mutations in GDP-mannose pyrophosphorylase B cause congenital and limb-girdle muscular dystrophies associated with

hypoglycosylation of α-dystroglycan. *Am. J. Hum. Genet.* **93,** 29–41 (2013).

33. Jensen, B. S. *et al.* GMPPB-Associated Dystroglycanopathy: Emerging Common Variants with Phenotype Correlation. *Hum. Mutat.* **36,** 1159–1163 (2015).

34. Rodríguez Cruz, P. M. *et al.* Clinical features of the myasthenic syndrome arising from mutations in GMPPB. *J. Neurol. Neurosurg. Psychiatry* **87,** 802–809 (2016).

35. Belaya, K. *et al.* Mutations in GMPPB cause congenital myasthenic syndrome and bridge myasthenic disorders with dystroglycanopathies. *Brain* **138,** 2493–2504 (2015).

36. Cabrera-Serrano, M. *et al.* Expanding the phenotype of GMPPB mutations. *Brain* **138,** 836–844 (2015).

37. Wright, K. M. *et al.* Dystroglycan organizes axon guidance cue localization and axonal pathfinding. *Neuron* **76,** 931–944 (2012).

38. Roberts, J. A. The genetics of mental deficiency. *Eugen. Rev.* **44,** 71–83 (1952).

39. McGue, M. & Gottesman, I. I. Classical and Molecular Genetic Research on General Cognitive Ability. *Hastings Cent. Rep.* **45,** S25–31 (2015).

40. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* **17,** 9–18 (2016).

41. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18,** 199–209 (2015).

42. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7,** 29–59 (2006).

43. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence

differences. *Nat. Rev. Neurosci.* **11,** 201–211 (2010).

44. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
45. Purcell, S., Cherny, S. S. & Sham, P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19,** 149–150 (2003).
46. Peloso, G. M. *et al.* Phenotypic extremes in rare variant study designs. *Eur. J. Hum. Genet.* **24,** 924–930 (2016).
47. de Bakker, P. I. W. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17,** R122–R128 (2008).
48. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
49. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: Functional mapping and annotation of genetic associations. *bioRxiv* 110023 (2017). doi:10.1101/110023
50. Vernot, B. *et al.* Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352,** 235–239 (2016).
51. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
52. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526,** 75–81 (2015).
53. Team, R. C. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. *Document freely*

available on the internet at: http://www. r-project. org (2015).

54. Barbeira, A. *et al.* MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results. *bioRxiv* 045260 (2016). doi:10.1101/045260

55. Koscielny, G. *et al.* Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw1055

56. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1,** 417–425 (2015).

57. Wagnon, J. L. *et al.* CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet.* **8,** e1003067 (2012).