Published online 10 December 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D877–D881***
doi: 10.1093/nar/gkv1340

HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease

Lucas D. Ward[1,2] and Manolis Kellis[1,2,*]

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [2]The Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

Received September 28, 2015; Revised November 14, 2015; Accepted November 16, 2015

ABSTRACT

More than 90% of common variants associated with complex traits do not affect proteins directly, but instead the circuits that control gene expression. This has increased the urgency of understanding the regulatory genome as a key component for translating genetic results into mechanistic insights and ultimately therapeutics. To address this challenge, we developed HaploReg (http://compbio.mit.edu/HaploReg) to aid the functional dissection of genome-wide association study (GWAS) results, the prediction of putative causal variants in haplotype blocks, the prediction of likely cell types of action, and the prediction of candidate target genes by systematic mining of comparative, epigenomic and regulatory annotations. Since first launching the website in 2011, we have greatly expanded HaploReg, increasing the number of chromatin state maps to 127 reference epigenomes from ENCODE 2012 and Roadmap Epigenomics, incorporating regulator binding data, expanding regulatory motif disruption annotations, and integrating expression quantitative trait locus (eQTL) variants and their tissue-specific target genes from GTEx, Geuvadis, and other recent studies. We present these updates as HaploReg v4, and illustrate a use case of HaploReg for attention deficit hyperactivity disorder (ADHD)-associated SNPs with putative brain regulatory mechanisms.

INTRODUCTION

Phenotype-associated loci from genome-wide association studies (GWAS) are usually non-coding, and functionally interpreting them is a challenge due to linkage disequilibrium (LD) and our almost complete inability to predict regulatory function directly from non-coding sequence. Therefore, regulatory genomic data such as maps of enhancers and transcription factor binding sites are essential to interpreting GWAS, developing mechanistic hypotheses, and ultimately understanding the genetic architecture of complex traits and disease (1–3). For human geneticists, these regulatory data can be unwieldy to translate from a genome browser to insights about a set of genomically-dispersed disease variants. HaploReg (4) integrates regulatory genomic maps together in the context of haplotype blocks, allowing researchers to intersect regulatory elements with genetic variants to quickly formulate functional hypotheses, both through dissection of multiple variants within a haplotype block and through global enrichment analysis of a set of associated loci. HaploReg annotation of GWAS has successfully been applied for haplotype fine-mapping (5–9) and enrichment analysis (7,10,11).

DATA AND INTERFACE UPDATES

HaploReg has been expanded substantially since it first launched in 2011. Here we describe the updates that have been incorporated in Haploreg v4 in response to new research in regulatory genomics and feedback from users.

Catalog of variants

HaploReg v4 defines a core set of 52 054 804 variants, consisting primarily of single-nucleotide polymorphisms (SNPs) using all refSNP IDs, hg19 positions and alleles from dbSNP release b137 (12). Corresponding hg38 coordinates for these variants were obtained from dbSNP release b141. This core set of dbSNP variants was integrated with other data sets either by rsID (for GWAS, eQTL and 1000 Genomes data) or by intersecting intervals by coordinate using the BEDTools software package (13) (for all other functional tracks.)

Linkage disequilibrium was calculated using phased low-coverage whole-genome autosomal sequences for four ancestral super-populations (AFR, AMR, ASN and EUR)

*To whom correspondence should be addressed. Tel: +1 617 253 2419; Fax: +1 617 452 5034; Email: manoli@mit.edu
Present address: Lucas D. Ward, Amgen, Inc., Cambridge, MA 02142, USA.

Query SNP: rs864643 and variants with r^2 >= 0.8

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
3	39494916	0.81	0.95	rs561543	G	A	0.52	0.23	0.23	0.19			4 tissues	VAS	HNF4A	HNF4		4 hits	5 hits	MOBP	intronic
3	39495310	0.83	0.93	rs72410685	ATGAAT	A	0.49	0.23	0.22	0.19			BLD			Pax-6,Pou3f2,Sox			3 hits	MOBP	intronic
3	39495518	0.9	0.95	rs4359752	A	G	0.51	0.25	0.26	0.20			BLD			E2A,TBX5,ZEB1			2 hits	MOBP	intronic
3	39495699	0.85	0.95	rs4113192	A	G	0.49	0.23	0.22	0.19			BLD			Hdx,RXRA,Zbtb3			2 hits	MOBP	intronic
3	39495779	0.85	0.95	rs4113193	G	A	0.48	0.23	0.22	0.19						15 altered motifs			2 hits	MOBP	intronic
3	39496043	0.85	0.95	rs6762335	T	C	0.49	0.23	0.22	0.19						5 altered motifs			2 hits	MOBP	intronic
3	39496111	0.86	0.97	rs6762416	T	C	0.49	0.23	0.22	0.19						HNF1,Pou2f2,STAT			2 hits	MOBP	intronic
3	39496489	0.85	0.96	rs6808636	A	G	0.51	0.23	0.22	0.19						EWSR1-FLI1,Pax-4,Sin3Ak-20			2 hits	MOBP	intronic
3	39496599	0.84	0.94	rs55780606	C	T	0.43	0.22	0.22	0.19						GR,Gfi1b			1 hit	MOBP	intronic
3	39496803	0.9	0.97	rs1768233	A	T	0.49	0.23	0.22	0.19						4 altered motifs			2 hits	MOBP	intronic
3	39496851	0.81	0.91	rs1708009	T	C	0.68	0.27	0.25	0.21										MOBP	intronic
3	39496891	0.81	0.92	rs1708015	A	G	0.49	0.24	0.23	0.20						ERalpha-a,Roaz,p300			1 hit	MOBP	intronic
3	39496978	0.89	0.97	rs1708018	C	T	0.49	0.23	0.22	0.19						7 altered motifs			2 hits	MOBP	intronic
3	39497049	0.89	0.97	rs533463	T	C	0.49	0.23	0.22	0.19						6 altered motifs			1 hit	MOBP	intronic
3	39497234	0.9	0.97	rs535220	T	C	0.49	0.23	0.22	0.19			BRN			CCNT2,Nr2e3,PLZF			2 hits	MOBP	intronic
3	39497603	0.9	0.97	rs538214	T	C	0.49	0.23	0.22	0.19			BRN			TATA,YY1			2 hits	MOBP	intronic
3	39497642	0.96	0.98	rs1708032	T	C	0.51	0.24	0.26	0.20			BRN			Crx,Foxo,Hoxb8		4 hits	1 hit	MOBP	intronic
3	39497652	0.89	0.96	rs538972	A	T	0.49	0.23	0.22	0.20			BRN					4 hits	3 hits	MOBP	intronic
3	39498035	0.91	0.98	rs614359	G	A	0.49	0.23	0.22	0.19		BRN	IPSC, BRN, GI			DMRT2,Ets		4 hits	3 hits	MOBP	intronic
3	39498075	0.98	0.99	rs563767	T	C	0.65	0.25	0.26	0.20		BRN	IPSC, BRN, GI			5 altered motifs			2 hits	MOBP	intronic

Figure 1. Example of haplotype summary view.

from the 1000 Genomes Project Phase 1 release (14), using a search space of all variants within 250 kilobases of each other. Allele frequencies were also obtained for each population.

Location of variants relative to genes was calculated using BEDTools and both GENCODE (15) and RefSeq (16).

Genome-wide association studies

GWAS were obtained from the EBI-NHGRI GWAS Catalog (17) (downloaded 30 October 2015). When there were multiple GWAS for the same trait, a trait-wide pruning was performed to retain only the strongest (lowest *P*-value) GWAS result from all studies on that trait, when two results from different studies were overlapping or within one megabase of each other.

Sequence conservation

Mammalian evolutionarily constrained elements are defined as originally reported, using both SiPhy elements (18) and GERP elements (19). Both of these comparative genomics studies report both base-level conservation scores as well as discretized elements; we chose to report discretized elements resulting from the authors' algorithms for the sake of simplicity and interpretability. A colored cell represents that that the element is conserved according to the algorithm.

Regulatory protein binding

Protein-binding sites from a variety of cell types and experimental conditions was obtained from the ENCODE Project ChIP-Seq data (20), processed by the narrowPeak algorithm.

Reference epigenomes

Epigenomic data from the Roadmap Epigenomics project (11) for the following data sets were included: ChromHMM states corresponding to enhancer or promoter elements, from the 15-state core model and 25-state model incorporating imputed data (21); histone modification ChIP-seq peaks using the gappedPeak algorithm for H3K27ac, H3K9ac, H3K9me1 and H3K9me3; and DNase hypersensitivity data peaks using the narrowPeak algorithm.

Expression quantitative trait loci

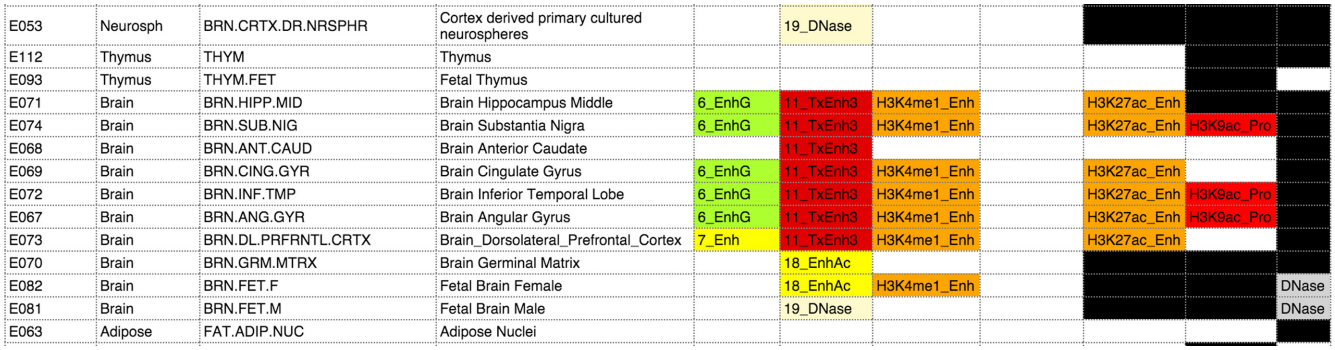
Expression QTL (eQTL) results were obtained from the GTEx pilot analysis v6 (22), the GEUVADIS project (23) and 12 other studies (10,24–34) in order to annotate variants with their putative regulatory target genes and the tissue(s) in which genotype has been associated with gene expression level. A wide range of QTLs, including eQTLs and other molecular QTLs such as metabolite QTLs, were also extracted from the GRASP database, build 2.0.0.0 (35,36).

Regulatory motifs

A library of position weight matrices from commercial, literature and motif-finding analysis of the ENCODE project (37) was used to score the effect of variants on regulatory motifs using the position weight matrix (PWM)-scanning process described previously (4).

Enrichment analysis

For a given set of lead SNPs from a GWAS or user-input SNPs, the overlap of SNPs with predicted enhancers in each reference epigenome is assessed. Users have four different options for defining enhancers, available in the option panel: using the 15-state core model, using the 25-state model incorporating imputed epigenomes, using H3K4me1 peaks and using H3K27ac peaks. The overlap with enhancers in each cell type is compared to two background models to assess enrichment: all 1000 Genomes variants with a frequency above 5% in any population and all independent GWAS catalog SNPs. The enrichment relative to these background frequencies is performed using a binomial test and uncorrected *P*-values are reported in an enrichment table underneath the haplotype views.



E053	Neurosph	BRN.CRTX.DR.NRSPHR	Cortex derived primary cultured neurospheres		19_DNase					
E112	Thymus	THYM	Thymus							
E093	Thymus	THYM.FET	Fetal Thymus							
E071	Brain	BRN.HIPP.MID	Brain Hippocampus Middle	6_EnhG	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh		
E074	Brain	BRN.SUB.NIG	Brain Substantia Nigra	6_EnhG	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
E068	Brain	BRN.ANT.CAUD	Brain Anterior Caudate		11_TxEnh3					
E069	Brain	BRN.CING.GYR	Brain Cingulate Gyrus	6_EnhG	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh		
E072	Brain	BRN.INF.TMP	Brain Inferior Temporal Lobe	6_EnhG	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
E067	Brain	BRN.ANG.GYR	Brain Angular Gyrus	6_EnhG	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh	H3K9ac_Pro	
E073	Brain	BRN.DL.PRFRNTL.CRTX	Brain_Dorsolateral_Prefrontal_Cortex	7_Enh	11_TxEnh3	H3K4me1_Enh		H3K27ac_Enh		
E070	Brain	BRN.GRM.MTRX	Brain Germinal Matrix		18_EnhAc					
E082	Brain	BRN.FET.F	Fetal Brain Female		18_EnhAc	H3K4me1_Enh				DNase
E081	Brain	BRN.FET.M	Fetal Brain Male		19_DNase					DNase
E063	Adipose	FAT.ADIP.NUC	Adipose Nuclei							

Figure 2. Example of epigenome details in a SNP detail view.

NHGRI-EBI GWAS hits

Trait	p-value	PMID
Attention deficit hyperactivity disorder	1E-8	18839057

GRASP QTL hits

Trait	p-value	PMID
Gene expression of MRPL15 in blood	7.3E-06	21829388
Serum ratio of (allantoin)/(quinate)	2.80E-04	21886157
Gene expression of MOBP (probeID ILMN_2298464) in cerebellum in Alzheimer's disease cases and controls	5.639E-33	22685416
Gene expression of MOBP (probeID ILMN_2298464) in cerebellum in Alzheimer's disease cases	1.398E-14	22685416
Gene expression of MOBP (probeID ILMN_2298464) in cerebellum in non-Alzheimer's disease samples	7.608E-18	22685416
Gene expression of MOBP (probeID ILMN_2298464) in temporal cortex in Alzheimer's disease cases and controls	1.262E-39	22685416
Gene expression of MOBP (probeID ILMN_2298464) in temporal cortex in Alzheimer's disease cases	1.471E-19	22685416
Gene expression of MOBP (probeID ILMN_2298464) in temporal cortex in non-Alzheimer's disease samples	1.4E-20	22685416
Gene expression of MOBP (probeID ILMN_2414962) in cerebellum in Alzheimer's disease cases and controls	0.000001177	22685416
Gene expression of MOBP (probeID ILMN_2414962) in temporal cortex in Alzheimer's disease cases and controls	5.381E-09	22685416
Gene expression of MOBP (probeID ILMN_2414962) in temporal cortex in non-Alzheimer's disease samples	0.00001133	22685416

Hits from selected eQTL studies

Study ID	Paper Title	PMID	Tissue	Correlated gene	p-value
Lappalainen2013	Transcriptome and genome sequencing uncovers functional variation in humans	24037378	Lymphoblastoid_EUR_exonlevel	ENSG00000168028.8_39449094_39449277	1.23112587621021e-05

Regulatory motifs altered

Position Weight Matrix ID (Library from Kheradpour and Kellis, 2013)	Strand	Ref	Alt	Match on:
				Ref: ATCCATGTGTCAGATGTAGCCAACGAATTATGTCAGAAGCAGAGAGAAAAGGCCTGAAA
				Alt: ATCCATGTGTCAGATGTAGCCAACGAATTGTGTCAGAAGCAGAGAGAAAAGGCCTGAAA
p300_disc6	+	13	1	ATTAYRWCA

Figure 3. Example of eQTL and motif alteration details from a SNP detail view.

USE EXAMPLE

To become acquainted with HaploReg, use the GWAS drop-down menu to select 'Attention deficit hyperactivity disorder (Lesch KP, 2008, 26 SNPs)' and select 'Submit'. Notice that the first two haplotype blocks from this study (38) are driven by lead SNPs with the same *P*-value = 1×10^{-8}. Go to the second haplotype result, for lead SNP rs864643 (Figure 1). Note that the top row in the haplotype block shows the SNP rs561543, and that it has LD of $r^2 = 0.81$ and $D' = 0.95$ with the lead variant rs864643. It overlaps with an HMM-predicted enhancer in four major tissue types; hover over '4 tissues' in that row to see a variety of enhancer tissues, including brain. Note that there is also an experiment with HNF4 protein bound by ChIP-Seq, 9 QTL results and an HNF4 motif disruption.

Notice the enrichment results at the bottom of the page below the haplotype results. Note that the strongest enrichment for enhancers (as defined by the 15-state core ChromHMM model) is in the angular gyrus sample from brain, with binomial $P = 2.0 \times 10^{-6}$ relative to all common SNPs.

Then go to the entry on the block for the lead SNP itself, rs864643. Click on the rsid, which is colored red because it is the lead SNP. Note that in the full table of epigenomic information from Roadmap Epigenomics (11), there is a cluster of enhancer activity in brain, and that it is classified as a genic enhancer by the 15-state core model and transcribed 3′ enhancer by the 25-state model (Figure 2). Note that H3K4me1, H3K27ac and H3K9ac all contribute to the chromatin state assignment at this locus. Black cells on the

right hand of this part of the table indicate that DNase was not assayed by Roadmap in these tissues.

Go to the bottom of the detail page for rs864643. Note that the SNP has been correlated with MOBP expression in two brain tissues (29), MPRL15 expression in blood (39) and serum ratio of allantoin to quinate (40); all three of these studies were curated by GRASP and found by cross-referencing this SNP to its database (35,36) (Figure 3). Looking at studies individually curated by HaploReg, notice that the SNP has been associated with differential expression of a single exon of RPSA in lymphoblastoid cells by the GEUVADIS study (23). In the motif table, note that the SNP changes the match to the p300 PWM, ATTAYRWCA, with the alternate allele changing a match to the fourth A to a G. Hover over the 'p300_disc' ID to see that the motif was discovered using the Trawler algorithm on a p300 ChIP-Seq experiment in HeLa cells from the ENCODE dataset (37).

These lines of evidence suggest regulatory mechanisms by which the SNPs from this GWAS may affect the complex phenotype of ADHD. While individually each piece of evidence is relatively weak, they offer ways in which molecular biologists could proceed with further experiments that would more definitively establish mechanisms. For example, the GWAS-wide enrichment suggests global differential gene regulation in angular gyrus, which has been associated with hyperactivation in ADHD by fMRI (41) and suggests a tissue to study gene expression directly in animal models. ChIP-seq and motif data suggest specifically testing HNF4 binding differentially to the alleles of rs561543, and the strong motif coupled with eQTL data suggest looking at whether p300 binds differentially to rs864643 in a brain tissue model. Finally, MOBP eQTL evidence suggests experiments to dissect the mechanism of MOBP differential expression, perhaps modulated by p300 at rs864643 and suggests that it may be useful to perform ADHD-relevant behavioral assays of MOBP-deficient mice, which do not show an overt behavioral phenotype (42).

ACKNOWLEDGEMENT

We thank HaploReg users and the reviewers of this manuscript for helpful suggestions and feedback.

FUNDING

National Institutes of Health (NIH) [R01-HG004037, RC1-HG005334, R01-HG008155]. Funding for open access charge: NIH [R01 HG004037].
Conflict of interest statement. None declared.

REFERENCES

1. Ward,L.D. and Kellis,M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.*, **30**, 1095–1106.
2. Paul,D.S., Soranzo,N. and Beck,S. (2014) Functional interpretation of non-coding sequence variation: concepts and challenges. *Bioessays*, **36**, 191–199.
3. Civelek,M. and Lusis,A.J. (2014) Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.*, **15**, 34–48.
4. Ward,L.D. and Kellis,M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
5. Verhoeven,V.J., Hysi,P.G., Wojciechowski,R., Fan,Q., Guggenheim,J.A., Hohn,R., MacGregor,S., Hewitt,A.W., Nag,A., Cheng,C.Y. *et al.* (2013) Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat. Genet.*, **45**, 314–318.
6. Chung,C.C., Kanetsky,P.A., Wang,Z., Hildebrandt,M.A., Koster,R., Skotheim,R.I., Kratz,C.P., Turnbull,C., Cortessis,V.K., Bakken,A.C. *et al.* (2013) Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.*, **45**, 680–685.
7. Lee,M.N., Ye,C., Villani,A.C., Raj,T., Li,W., Eisenhaure,T.M., Imboywa,S.H., Chipendo,P.I., Ran,F.A., Slowikowski,K. *et al.* (2014) Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*, **343**, 1246980.
8. Ruark,E., Seal,S., McDonald,H., Zhang,F., Elliot,A., Lau,K., Perdeaux,E., Rapley,E., Eeles,R., Peto,J. *et al.* (2013) Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat. Genet.*, **45**, 686–689.
9. Franceschini,N., Fox,E., Zhang,Z., Edwards,T.L., Nalls,M.A., Sung,Y.J., Tayo,B.O., Sun,Y.V., Gottesman,O., Adeyemo,A. *et al.* (2013) Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am. J. Hum. Genet.*, **93**, 545–554.
10. Westra,H.J., Peters,M.J., Esko,T., Yaghootkar,H., Schurmann,C., Kettunen,J., Christiansen,M.W., Fairfax,B.P., Schramm,K., Powell,J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–1243.
11. Roadmap Epigenomics Consortium, Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
12. Sherry,S.T., Ward,M.-H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
13. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
14. 1000 Genomes Project Consortium, Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M.A., Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T. and McVean,G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
15. Harrow,J., Denoeud,F., Frankish,A., Reymond,A., Chen,C.K., Chrast,J., Lagarde,J., Gilbert,J.G., Storey,R., Swarbreck,D. *et al.* (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol.*, **7**(Suppl. 1), S4.1–S4.9.
16. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
17. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
18. Lindblad-Toh,K., Garber,M., Zuk,O., Lin,M.F., Parker,B.J., Washietl,S., Kheradpour,P., Ernst,J., Jordan,G., Mauceli,E. *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476–482.
19. Davydov,E.V., Goode,D.L., Sirota,M., Cooper,G.M., Sidow,A. and Batzoglou,S. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.*, **6**, e1001025.
20. Encode Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
21. Ernst,J. and Kellis,M. (2015) Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.*, **33**, 364–376.
22. GTEx Consortium. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
23. Lappalainen,T., Sammeth,M., Friedlander,M.R., t Hoen,P.A., Monlong,J., Rivas,M.A., Gonzalez-Porta,M., Kurbatova,N., Griebel,T., Ferreira,P.G. *et al.* (2013) Transcriptome and genome

sequencing uncovers functional variation in humans. *Nature*, **501**, 506–511.
24. Montgomery,S.B., Sammeth,M., Gutierrez-Arcelus,M., Lach,R.P., Ingle,C., Nisbett,J., Guigo,R. and Dermitzakis,E.T. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, **464**, 773–777.
25. Schadt,E.E., Molony,C., Chudin,E., Hao,K., Yang,X., Lum,P.Y., Kasarskis,A., Zhang,B., Wang,S., Suver,C. *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.
26. Gibbs,J.R., van der Brug,M.P., Hernandez,D.G., Traynor,B.J., Nalls,M.A., Lai,S.L., Arepalli,S., Dillman,A., Rafferty,I.P., Troncoso,J. *et al.* (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.*, **6**, e1000952.
27. Stranger,B.E., Nica,A.C., Forrest,M.S., Dimas,A., Bird,C.P., Beazley,C., Ingle,C.E., Dunning,M., Flicek,P., Koller,D. *et al.* (2007) Population genomics of human gene expression. *Nat. Genet.*, **39**, 1217–1224.
28. Li,Q., Stram,A., Chen,C., Kar,S., Gayther,S., Pharoah,P., Haiman,C., Stranger,B., Kraft,P. and Freedman,M.L. (2014) Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum. Mol. Genet.*, **23**, 5294–5302.
29. Zou,F., Chai,H.S., Younkin,C.S., Allen,M., Crook,J., Pankratz,V.S., Carrasquillo,M.M., Rowley,C.N., Nair,A.A., Middha,S. *et al.* (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.*, **8**, e1002707.
30. Koopmann,T.T., Adriaens,M.E., Moerland,P.D., Marsman,R.F., Westerveld,M.L., Lal,S., Zhang,T., Simmons,C.Q., Baczko,I., dos Remedios,C. *et al.* (2014) Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One*, **9**, e97380.
31. Ramasamy,A., Trabzuni,D., Guelfi,S., Varghese,V., Smith,C., Walker,R., De,T. and U. K. Brain Expression Consortium, North American Brain Expression ConsortiumU. K. Brain Expression Consortium, North American Brain Expression Consortium and Coin,L. *et al.* (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.*, **17**, 1418–1428.
32. Fairfax,B.P., Humburg,P., Makino,S., Naranbhai,V., Wong,D., Lau,E., Jostins,L., Plant,K., Andrews,R., McGee,C. *et al.* (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, **343**, 1246949.
33. Grundberg,E., Adoue,V., Kwan,T., Ge,B., Duan,Q.L., Lam,K.C., Koka,V., Kindmark,A., Weiss,S.T., Tantisira,K. *et al.* (2011) Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet.*, **7**, e1001279.
34. Hao,K., Bosse,Y., Nickle,D.C., Pare,P.D., Postma,D.S., Laviolette,M., Sandford,A., Hackett,T.L., Daley,D., Hogg,J.C. *et al.* (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.*, **8**, e1003029.
35. Eicher,J.D., Landowski,C., Stackhouse,B., Sloan,A., Chen,W., Jensen,N., Lien,J.P., Leslie,R. and Johnson,A.D. (2015) GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
36. Leslie,R., O'Donnell,C.J. and Johnson,A.D. (2014) GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics*, **30**, i185–i194.
37. Kheradpour,P. and Kellis,M. (2014) Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.*, **42**, 2976–2987.
38. Lesch,K.P., Timmesfeld,N., Renner,T.J., Halperin,R., Roser,C., Nguyen,T.T., Craig,D.W., Romanos,J., Heine,M., Meyer,J. *et al.* (2008) Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.*, **115**, 1573–1585.
39. Fehrmann,R.S., Jansen,R.C., Veldink,J.H., Westra,H.J., Arends,D., Bonder,M.J., Fu,J., Deelen,P., Groen,H.J., Smolonska,A. *et al.* (2011) Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.*, **7**, e1002197.
40. Suhre,K., Shin,S.Y., Petersen,A.K., Mohney,R.P., Meredith,D., Wagele,B., Altmaier,E., CardioGram, Deloukas,P., Erdmann,J. *et al.* (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, **477**, 54–60.
41. Cortese,S., Kelly,C., Chabernaud,C., Proal,E., Martino,A., Milham,M.P. and Castellanos,F.X. (2012) Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. *Am. J. Psychiatry*, **169**, 1038–1055.
42. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E. and Mouse Genome Database, G. (2015) The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.*, **43**, D726–D736.