



HAL
open science

Genome wide association study of incomplete hippocampal inversion in adolescents

Claire Cury, Marzia Antonella Scelsi, Roberto Toro, Vincent Frouin, Eric Artiges, Andreas Heinz, Herve Lemaitre, Jean-Luc Martinot, Jean-Baptiste Poline, Michael N Smolka, et al.

► **To cite this version:**

Claire Cury, Marzia Antonella Scelsi, Roberto Toro, Vincent Frouin, Eric Artiges, et al.. Genome wide association study of incomplete hippocampal inversion in adolescents. PLoS ONE, Public Library of Science, 2020, 15 (1), 10.1371/journal.pone.0227355 . hal-02416654

HAL Id: hal-02416654

<https://hal.inria.fr/hal-02416654>

Submitted on 17 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genome wide association study of incomplete hippocampal inversion in adolescents

Claire Cury^{1,*}, Marzia Antonella Scelsi¹, Roberto Toro^{2,3}, Vincent Frouin⁴, Eric Artiges⁵, Antoine Grigis⁴, Andreas Heinz⁶, Hervé Lemaître⁷, Jean-Luc Martinot⁸, Jean-Baptiste Poline^{9,10}, Michael N. Smolka¹¹, Henrik Walter⁶, Gunter Schumann¹², Andre Altmann^{1,‡}, Olivier Colliot^{13,14,15,16,17,‡} and the IMAGEN Consortium[^]

1 Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK

2 Centre National de la Recherche Scientifique, Genes, Synapses and Cognition, URA 2182, Institut Pasteur, Paris, France

3 Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France.

4 NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

5 Inserm U 1000 “Neuroimaging & Psychiatry”, University Paris Sud, University Paris Descartes - Sorbonne Paris Cité; and Psychiatry Department 91G16, Orsay Hospital, France

6 Charité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charitéplatz 1, Berlin, Germany

7 Inserm U 1000 “Neuroimaging & Psychiatry”, Faculté de médecine, Université Paris-Sud, Le Kremlin-Bicêtre; and Université Paris Descartes, Sorbonne Paris Cité, Paris, France

8 Inserm U 1000 “Neuroimaging & Psychiatry”, University Paris Sud, University Paris Descartes - Sorbonne Paris Cité; and Maison de Solenn, Paris, France

9 McGill University, Faculty of Medicine, Montreal Neurological Institute and Hospital, McConnell Brain Imaging Center, Ludmer Centre for Neuroinformatics and Mental Health, Canada.

10 Henry H. Wheeler Jr. Brain Imaging Center, Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA

11 Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

31 12 Medical Research Council - Social, Genetic and Developmental Psychiatry Centre, Institute of
32 Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
33 13 Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France
34 14 Inserm, U 1127, F-75013, Paris, France
35 15 CNRS, UMR 7225, F-75013, Paris, France
36 16 Sorbonne Université, F-75013, Paris, France
37 17 Inria, Aramis project-team, F-75013, Paris, France

38

39 ‡ Shared senior authors

40

41 ^ Whole author list is given in the supporting information file S1 File and in the acknowledgment
42 section.

43

44 * **Corresponding author information:**

45 Email: claire.cury.pro@gmail.com (CC)

46 Current affiliation:

47 Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn (ex VISAGES) – ERL U 1228, F-
48 35000 Rennes, France

49

50 **Abstract**

51 Incomplete hippocampal inversion (IHI), also called hippocampal malrotation, is an atypical
52 presentation of the hippocampus present in about 20% of healthy individuals. Here we conducted
53 the first genome-wide association study (GWAS) in IHI to elucidate the genetic underpinnings that
54 may contribute to the incomplete inversion during brain development. A total of 1381 subjects
55 contributed to the discovery cohort obtained from the IMAGEN database. The incidence rate of
56 IHI was 26.1%. Loci with $P < 1e-5$ were followed up in a validation cohort comprising 161 subjects
57 from the PING study. Summary statistics from the discovery cohort were used to compute IHI
58 heritability as well as genetic correlations with other traits. A locus on 18q11.2 (rs9952569;
59 $OR=1.999$; $Z=5.502$; $P=3.755e-8$) showed a significant association with the presence of IHI. A
60 functional annotation of the locus implicated genes *AQP4* and *KCTD1*. However, neither this locus
61 nor the other 16 suggestive loci reached a significant p-value in the validation cohort. The h^2
62 estimate was 0.54 (sd: 0.30) and was significant ($Z=1.8$; $P=0.036$). The top three genetic
63 correlations of IHI were with traits representing either intelligence or education attainment and
64 reached nominal $P \leq 0.013$.

65

66 Introduction

67

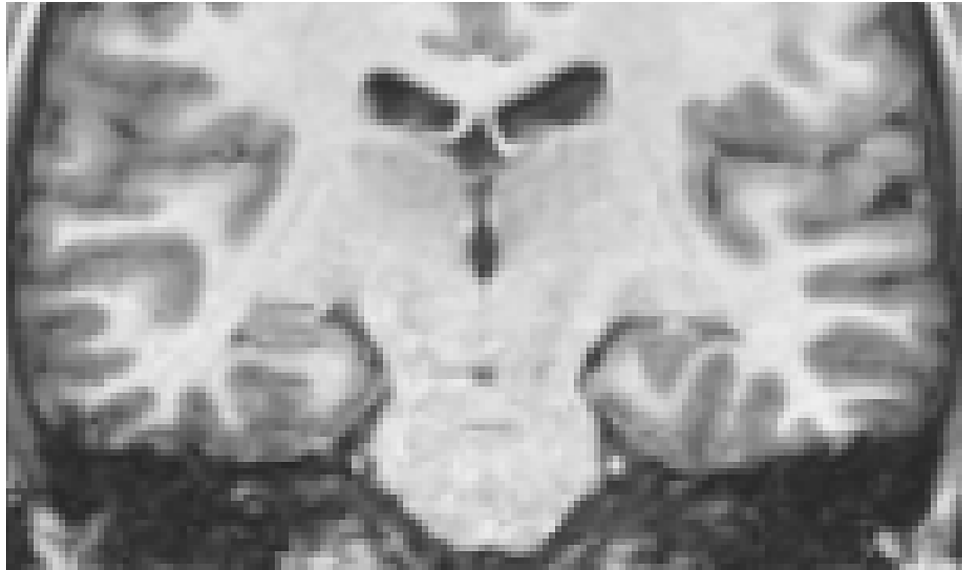
68 Human hippocampi are small structures, one in each temporal lobe that belongs to the brain's
69 limbic system and is known to be mainly involved in memory processes such as long term
70 memorisation and spatial navigation [1]. The limbic system and the hippocampus influence the
71 activity of the hypothalamic Pituitary Adrenocortical (HPA) axis, a major neuroendocrine mediator
72 of stress, playing a role in emotional stress responses [2]. Thus, the hippocampus is implicated,
73 with evidence of morphological changes, in a variety of neurological pathologies and psychiatric
74 disorders, such as Alzheimer's disease where hippocampal atrophy increases with the pathology
75 [3]; major depressive disorder where hippocampal volume can predict the response to
76 antidepressants [4,5], is related to suicide attempts [6], and is linked to cortisol disruption
77 (highlighting the implication of the hippocampus in the HPA axis) [7]; Schizophrenia, where
78 patients have smaller hippocampi [8]; or temporal lobe epilepsy, the most frequent form of chronic
79 focal epilepsy in adults, linked to hippocampal sclerosis [9]. Furthermore, during brain
80 development, the growth of the left and the right hippocampi shows distinct responses to postnatal
81 maternal stress [10]. Anatomically, there is a variation to the typical presentation of the
82 hippocampi in normal subjects: the incomplete hippocampal inversion (IHI) also referred to as
83 hippocampal malrotation (Fig 1). This anatomical variant has been initially observed in healthy
84 subjects by [11] and then mostly observed in patients with epilepsy [12,13]. IHIs are mainly left-
85 sided and characterized by a rounded or vertical shape, a medial positioning and a deep collateral
86 sulcus [13–15] and are present in around 20% of the normal population [15]. It has been reported
87 that IHI impacts the hippocampal volume: subjects with incomplete inversions appear to have
88 smaller hippocampi [16], and more specifically, the hippocampal subfield CA1 seems to be related
89 to the IHI severity [17]. Also it has been suggested that IHI might interfere with the quality of
90 hippocampal segmentation for volumetric analysis [16,18], which may be clinically relevant, since

91 the hippocampal volume can predict the response to antidepressant in patients without IHI [4,5].
92 Additionally, a sulcal morphometry analysis suggested that morphological changes associated
93 with IHI are not confined to the hippocampus [15]; significant differences in cortical sulci located
94 along the limbic system are shown between participants with and without complete inversion.
95 Several studies suggest that IHI have their origin in developmental processes [19,20]. For
96 example, [21] observed that during the rotational growth of the hemispheres, the major portion of
97 the hippocampus is carried dorso-laterally and then ventrally to lie in the medial part of the
98 temporal lobe. As the neocortex expands and evolves, the allocortex (the 3 layers cortex) is
99 displaced inferiorly, medially and internally into the temporal horn. This rotational growth of the
100 cortex implies an inversion of the hippocampus during normal development, which in some cases
101 may remain incomplete. Following this hypothesis, [22] conducted a study using foetal MRI and
102 found a correlation between the degree of in-folding and the number of gestational weeks. In a
103 recent study [15] described detailed criteria to evaluate IHI, ultimately making the IHI evaluation
104 more reproducible. In the same study, the introduced criteria had been applied to assess the IHI
105 status of 2000 adolescents without neurological disorders. Results showed a prevalence of about
106 20% of IHI among this normal population. The majority of the IHI cases were left-sided (17% on
107 left side). The lateral preference of left-sided over right-sided IHI may be rooted in the observation
108 of asymmetrical hippocampus development in neonates with the right hippocampus developing
109 faster than the left one [23]. In addition to these developmental observations, IHI has been
110 reported to be associated with genetic changes. For instance, IHI was observed at higher
111 prevalence in subjects with chromosome 22q11.2 microdeletion [24], which leads to DiGeorge
112 syndrome.

113

114

Fig 1: T1 weighted MRI in a coronal view. The left hippocampus (right side in the image) presents an incomplete hippocampal inversion (IHI). The right hippocampus (left side in the image) is an example of a normal or properly inverted hippocampus.



115

116

117 Given that recent evidence implicates developmental processes in the aetiology of IHI and the
118 observation that the structure and shape of subcortical structures, including the hippocampus, are
119 under genetic control [25], we aimed at elucidating specific genetic variants contributing to IHI. To
120 this end we conducted the first genome-wide association study on the genetics of incomplete
121 hippocampal inversion.

122 **Methods**

123

124 **Subjects**

125 Subjects were investigated from two cohorts: IMAGEN [26] and PING [27]. The IMAGEN cohort
126 comprises >2000 subjects collected at eight sites across Europe [26], and local ethics committee
127 approved the study (see at the end of the paper for details and study [26]). At the time of baseline

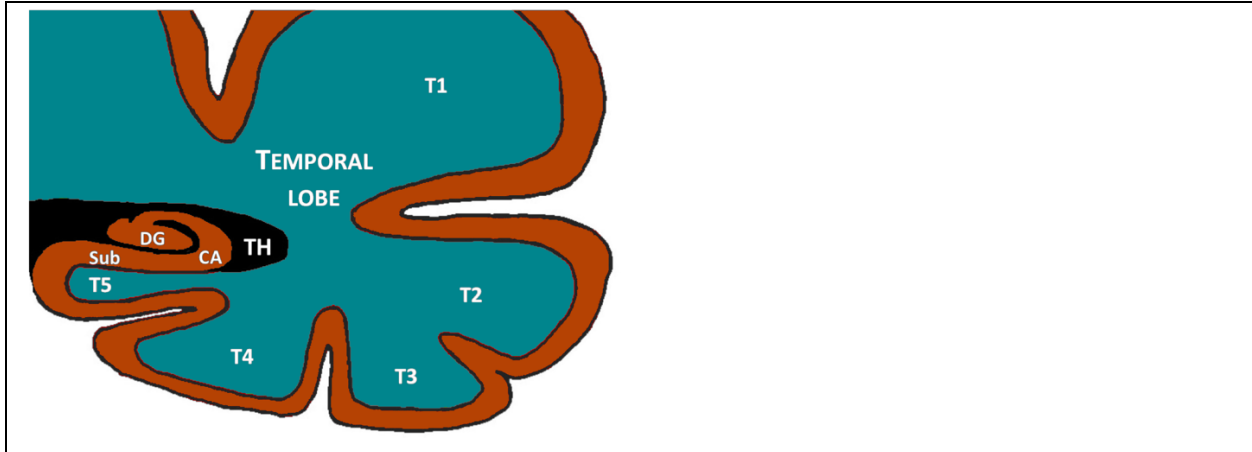
128 data collection and study inclusion all participants were 14 years of age. The second cohort was
129 obtained from the Pediatric Imaging Neurocognition and Genetics (PING) Study database
130 (<http://ping.chd.ucsd.edu/>). PING was launched in 2009 by the National Institute on Drug Abuse
131 (NIDA) and the Eunice Kennedy Shriver National Institute Of Child Health & Human Development
132 (NICHD) as a 2-year project of the American Recovery and Reinvestment Act. The primary goal
133 of PING has been to create a data resource of highly standardized and carefully curated magnetic
134 resonance imaging (MRI) data, comprehensive genotyping data, and developmental and
135 neuropsychological assessments for a large cohort of developing children aged 3 to 20 years.
136 The scientific aim of the project is, by openly sharing these data, to amplify the power and
137 productivity of investigations of healthy and disordered development in children, and to increase
138 understanding of the origins of variation in neurobehavioral phenotypes. Access to the dataset
139 was granted through a Federal Wide Assurance (FWA). For up-to-date information,
140 see <http://ping.chd.ucsd.edu/> and [27]. All methods were performed in accordance with relevant
141 guidelines and regulations.

142 **Image data processing and IHI scoring**

143 The procedure for scoring IHI [15], which has been previously described in detail and shown a
144 good intra- and inter-reproducibility [15], was applied to the subjects used in this study (from
145 IMAGEN and PING). Inter- and intra-rater variability were assessed in a previous publication [15].
146 This was studied on 42 participants from the discovery cohort using the kappa statistic. In all
147 cases, intra- and inter-rater agreements were beyond substantial ($\kappa \geq 0.64$). Very strong
148 agreements ($\kappa \geq 0.8$) were observed in the majority of comparisons (14/20). Rating on the
149 validation cohort was conducted after by a single rater (CC), thus the rater was not blinded to
150 whether subjects were from the discovery or validation cohort. In brief, the IHI score is composed
151 of four different criteria: (1) assessing the roundness of the hippocampal body; (2) evaluating the
152 verticality of the collateral sulcus which is located between the 4th and the 5th temporal lobe

153 convolution (Fig 2); (3) the mediality of the hippocampal body; and (4) the depth of the fusiform
154 gyrus, separating the 4th and the 3rd convolution of the temporal lobe (Fig 2). Each criterion is
155 assessed from a coronal point of view after registering the subjects' T1 weighted MRI into the
156 standard MNI space using the FSL's affine transformation FLIRT [28,29]. Evaluation was carried
157 out using an inhouse Java interface (https://github.com/cclairec/viewerIHI_java). During scoring
158 each criterion received a score between 0.0 and 2.0. The first three criteria have a step size of
159 0.5, the fourth criterion is binary (0 or 2), and the 5th criterion, assessed between 0 and 2, has a
160 step size of 1.0. The sum of those criteria forms the overall IHI score ranging from 0.0 to 10.0.
161 This is a semi-continuous score (with a step of 0.5), where an IHI score of 0.0 indicates the total
162 absence of IHI, and a score of 8.0 represents a very pronounced presentation of IHI. In their
163 previous study [15] established an optimal cut-off (at 3.75) of the overall IHI score to indicate
164 presence or absence of IHI, by maximising the accuracy of the classification of a global criterion
165 (blind to individual criteria or IHI scores), indicating if a given hippocampus presents or not an IHI
166 (an intermediate score for partial IHI were present but not used in the estimation of the optimal
167 cut-off): hippocampi without IHI correspond to IHI score < 4.0 and hippocampi with IHI correspond
168 to IHI scores ≥ 4 .

Fig 2: Hippocampal anatomy in coronal view. The hippocampus comprises the Dentate gyrus (DG), the cornus ammonis (CA) and the subiculum (sub). The temporal lobe is composed of five convolutions: T1, T2, T3, T4 and T5. The collateral sulcus divides T5 from T4 and the sulci of the fusiform gyrus separates T4 from T3. TH indicates the location of the temporal horn of the lateral ventricles.



169
170

171 For this genetic study, the phenotype was IHI in either left or right hippocampus. To
172 determine IHI, we applied the same cut-off of 4.0 for left and right hippocampi and used, for the
173 IMAGEN cohort, the previously processed data from the IMAGEN study [15].

174

175 **SNP genotyping and pre-processing**

176 IMAGEN subjects were genotyped from blood samples on 610-Quad SNP and 660-Quad SNP
177 arrays from Illumina. Genetic data was available for 1,841 subjects. In a first round of quality
178 control (QC) we performed subject-level QC by removing subjects with mismatching self-reported
179 sex and genotype inferred sex (N=10) or where more than 10% of SNPs were missing (N=0).
180 Next, we performed ancestry matching based on the HapMap3 data [30]. Population outliers were
181 defined as subjects exhibiting more than five standard deviations distance from the CEU and TSI
182 population in any of the first five principal components. Based on these criteria, 220 subjects were
183 excluded from further analysis (S1 Fig). For the remaining subjects the genetic relationship matrix
184 (GRM) was computed on common SNPs (minor allele frequency [MAF] >5%) after LD pruning
185 using GCTA [31]. Another 18 subjects were removed due to relatedness (i.e., PIHAT > 0.05)
186 leaving a total of 1593 subjects for the analysis. The raw genotyping data were prepared for

187 imputation using a series of scripts (<http://www.well.ox.ac.uk/~wrayner/tools/>). Haplotype
188 reference consortium (HRC) v1.1 [32] SNPs were imputed on the Sanger imputation server
189 (<https://imputation.sanger.ac.uk>) using EAGLE2 [33] for pre-phasing and PBWT [34] for
190 imputation. Data from the two different genotyping chips were imputed independently. Genotypes
191 were hard called based on the maximal genotype posterior probability with a threshold of 0.9.
192 That is, if none of the three genotypes reached a posterior probability of at least 0.9, then the SNP
193 was set to missing in the corresponding subject. Finally, an additional round of QC was conducted
194 on SNP level based on imputation quality (INFO score > 0.3), missingness (< 5%), minor allele
195 frequency (MAF>1%) and deviation from Hardy-Weinberg-Equilibrium ($p < 1e-6$) leaving 6,742,645
196 SNPs across the autosomes for the association analysis.

197 PING subjects were genotyped from saliva samples on Human660W-Quad arrays from
198 Illumina. After QC, genetic data for 1,391 participants was suitable for analysis. Individual SNPs
199 of the PING dataset were accessed through the PING data portal (ping-dataportal.ucsd.edu).
200 Ancestry and admixture proportions in the PING participants were based on the ADMIXTURE
201 software [35] and downloaded through the data portal (for details see [27]). We restricted the
202 validation cohort to participants of at least 12 years of age and of European ancestry (minimum
203 90% European ancestry as per ADMIXTURE; N=197).

204

205 **Genome wide association study**

206 The genome wide association study was carried out with Plink v1.9 [36] assuming an additive
207 genetic model and computing for every SNP a logistic regression while correcting for sex, age at
208 imaging (in days) and five principal components for population structure. Phenotype or covariate
209 information was missing for 212 participants. Thus, the discovery GWAS comprised 1,381
210 unrelated subjects. The genome-wide statistical significance threshold was set to the standard
211 threshold of $p < 5e-8$ and regional association plots were generated with LocusZoom [37]. SNPs

212 exceeding the threshold for suggestive association with IHI ($p < 1e-5$) were followed up in an
213 independent cohort of adolescents (PING). In case the top SNP was not genotyped in PING,
214 LDlink [38](<https://analysistools.nci.nih.gov/LDlink/>) was used to identify a proxy in LD (r^2) within
215 +/- 50kb of the top SNP's location. Association with single SNPs was tested in R using the glm
216 function; the logistic model was corrected for age and sex.

217

218 **Functional annotation of GWA summary statistics**

219 The GWA summary statistics were annotated using the web-based version of the FUncional
220 Mapping and Annotation (FUMA) tool [39] (<http://fuma.ctglab.nl/>). In order to elucidate the
221 functional consequences of genetic risk loci, FUMA approaches the mapping in two separate
222 steps: first, lead SNPs are identified and mapped to relevant genes on the basis of strand
223 proximity, expression quantitative trait loci (eQTL) and chromatin interaction; second, the
224 reprioritized genes returned by the first step are annotated with respect to expression levels and
225 overrepresentation in differentially expressed gene sets among a wide range of human tissues.

226 For the purposes of this study, SNP-to-gene mapping was performed according to the
227 following parameters: SNPs with $p < 5e-8$ were identified as lead SNPs, and genomic risk loci were
228 constructed by including SNPs in linkage disequilibrium with independent lead SNPs ($LD\ r^2 > 0.6$
229 in the 1000 Genomes Phase 3 EUR panel) and with a minimum MAF of 1%. Positional mapping
230 was performed by linking lead SNPs to genes in a 50kb window. Mapping based on eQTL was
231 performed by using only SNP-gene pairs significant at $FDR < 0.05$ in all tissues/cell types from 4
232 data repositories (GTEx [40], the Westra blood eQTL dataset [41], the BIOS QTL browser [42]
233 and BRAINEAC [43]); the available data only covers *cis*-eQTLs with up to 1 Mb distance between
234 SNP and gene. Chromatin interaction mapping was also performed to take into account potential
235 long-range interactions between risk loci and genes due to chromatin folding. We based mapping
236 on interactions significant at $FDR < 1e-6$ in 14 tissue types and seven cell lines from [44]. We also

237 based mapping on tissue/cell type specific enhancer or promoter regions annotated in 111
238 epigenomes from the Roadmap Epigenomics Project [45]. The Major Histocompatibility Complex
239 (MHC) was excluded from annotations, and mapping to all functional gene classes (protein-
240 coding, non-coding RNA, long intergenic ncRNA, processed transcripts, pseudogenes) was
241 enabled.

242 After mapping lead SNPs to relevant genes, we performed annotation of the prioritized
243 genes in biological context, mainly with respect to tissue-specific expression levels. Average
244 expression levels (\log_2 Read Per Kilobase per Million (RPKM+1)) of protein-coding genes in 53
245 tissues from GTEx v6 were visualized through heat maps, allowing for comparison of expression
246 level across genes and tissue types. Candidate genes were tested for overrepresentation in sets
247 of differentially expressed genes (DEG), as well as sets of genes up- and down-regulated, across
248 53 specific tissue types from GTEx v6 using hypergeometric tests. The same gene-set enrichment
249 analysis strategy was applied to test for overrepresentation of biological functions among the
250 prioritized genes, using gene sets from the Molecular Signatures Database version 5.2 [46],
251 WikiPathways [47] and the GWAS catalog [48], and applying the Benjamini-Hochberg multiple
252 testing correction procedure.

253

254 **Heritability analysis and genetic correlation**

255 We used LD score regression [49] in order to estimate IHI heritability from the GWAS summary
256 statistics data. Next, we computed partitioned heritability estimates using the LD score method
257 described in [50] and [51]. Heritability estimates were partitioned into 53 overlapping functional
258 categories, derived from 24 main annotations, from [50]. Stratified LD score regression was also
259 used to test for heritability enrichment in genes specifically expressed in a number of tissues of
260 cell types. For this analysis, we used the specifically expressed gene lists compiled by [51] for the
261 following datasets: expression levels from RNA-seq experiments in the 53 GTEx tissues and cell

262 types, as well as only the 13 GTEx brain regions; the Cahoy dataset, comprising microarray
263 expression data from three cell types (astrocyte, neuron, oligodendrocyte) in the mouse brain
264 [52]; the Franke dataset, comprising microarray expression data in 152 tissues and cell types
265 from human, mouse and rat [53]; and the Immunological Genome Project Consortium dataset,
266 comprising microarray expression data for 292 immune cell types in the mouse [54]. Enrichment
267 p-values were corrected for multiple comparisons using the Benjamini-Hochberg procedure.

268 Given the reported higher prevalence of IHI in patients with epilepsy, we used LD score
269 regression to compute the genetic correlation [55] between IHI and epilepsy susceptibility based
270 on a recent GWAS [56]. Finally, we conducted an exploratory analysis of genetic correlation
271 between IHI and traits from 832 GWASs using the LD hub [57] (<http://ldsc.broadinstitute.org/>).

272

273 **Gene expression in the developing human brain**

274 In order to explore the transcription pattern of two candidate genes, we downloaded their
275 expression values from BrainSpan through the web interface
276 (<http://www.brainspan.org/rnaseq/search/index.html>). The data comprises post-mortem gene
277 expression data of 42 subjects at ages spanning from prenatal development (eight post
278 conception weeks) till adulthood (40 years). Brains were sampled across 26 brain structures.
279 Gene expression was measured using RNA-sequencing and expression levels for each gene
280 were provided as reads per kilobase of exon model per million mapped reads (RPKM). We
281 analysed this data using a linear mixed effects model implemented in the lme4 package in R. In
282 these analyses, the gene expression level was the target variable, subject ID and structure ID
283 were random effects, while an indicator variable for age less than 25 weeks post conception was
284 the fixed effect. We tested for the significant effect of age<25 post conception weeks (pcw) on
285 gene expression. This threshold was selected based on the estimated occurrence of hippocampal
286 inversion between pcw 20 and 30 [20].

287 Results

288 Subjects: cohort statistics

289 In IMAGEN 1381 subjects had genotyping and all phenotype and confounding information
290 available. Incidence rate of IHI was 26.1%. In PING, for the 197 European subjects aged 12 years
291 or older, we could successfully access and score 161 T1 weighted MR images for analysis, and
292 IHI incidence rate was 23.6%; both at the 4.0 cutoff. There was a higher incidence rate of IHI in
293 the left hemisphere in both cohorts. Summary statistics for both cohorts can be found in Table 1.

294 **Table 1:** Cohort summary statistics of participants used for the genetic analyses.

Cohort	IMAGEN	PING
Participants (N)	1381	161
Female (%)	687 (49.7)	78 (48.4%)
Age (SD)	14.5 (0.41)	16.06 (2.54)
IHI (%)	360 (26.1)	38 (23.6)
Left/Right/Bilateral	251/46/63	24/7/7

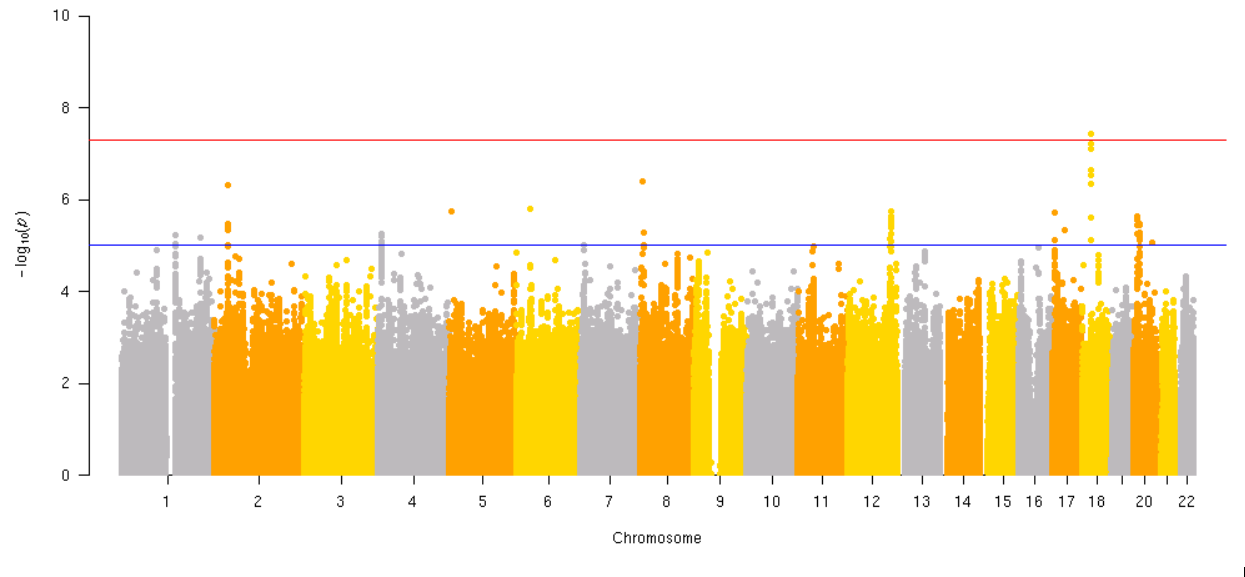
295

296 Genome wide association study and functional annotation

297 We tested each of the 6.7mio SNPs for an association with the presence of IHI. In the discovery
298 dataset comprising subjects from the IMAGEN study, 17 loci passed the threshold for suggestive
299 association (Fig 3; Table 2). One locus on chromosome 18 reached genome-wide significance
300 (top SNP: rs9952569; OR=1.999; Z=5.502; P=3.755e-8; Fig 4). There was no inflation in p-values
301 ($\lambda=1.017$; S2 Fig). The top SNP shows a consistently strong association with the continuous IHI
302 score (i.e., not applying the cutoff at 4.0) and the global criterion (C0), but misses the genome-
303 wide significance threshold in both cases (S3 Fig; beta=0.5021, Z= 5.129, P=3.332e-07 for the
304 continuous score and beta= 0.2542, Z= 5.299, P=1.354e-07 for C0). Functional annotation of the

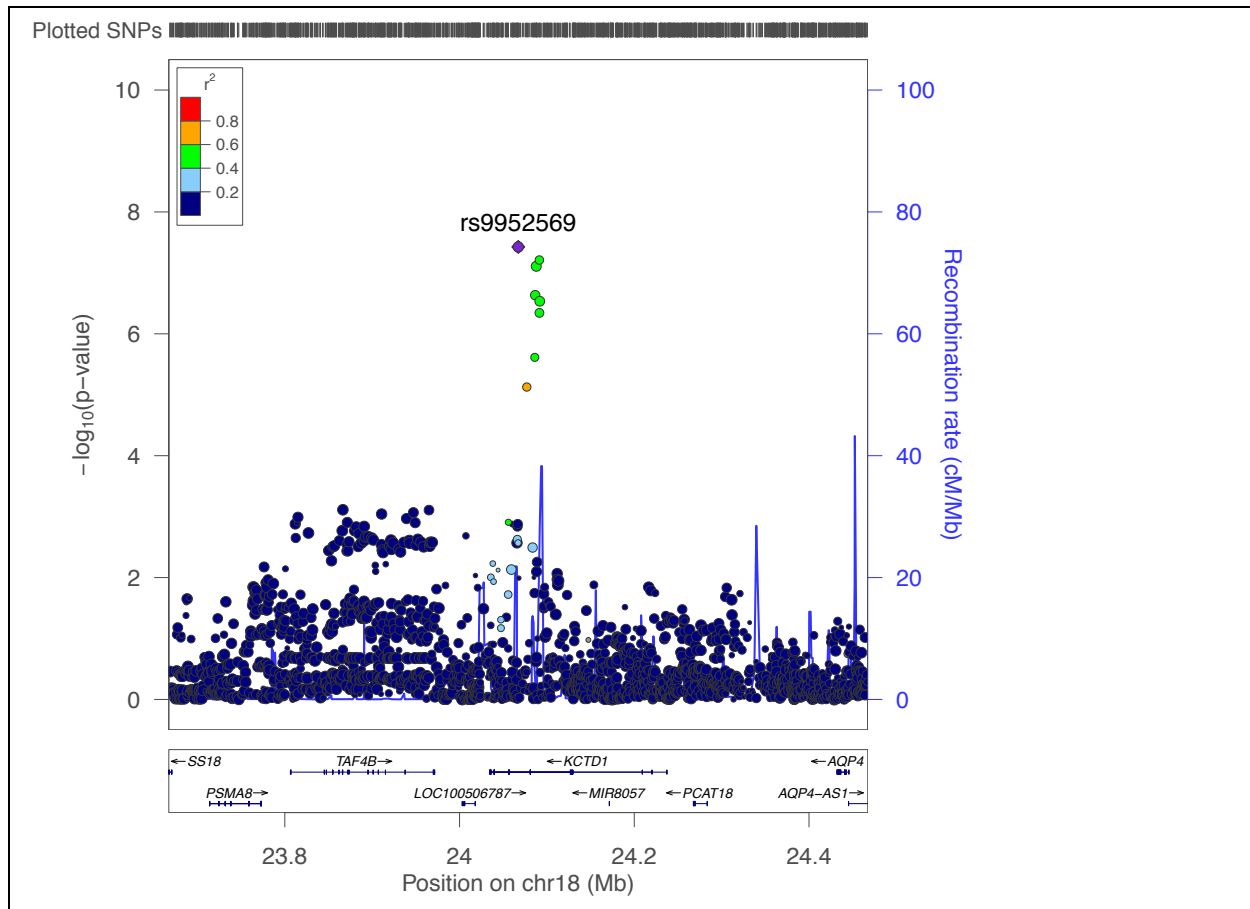
305 GWAS result was carried out using FUMA and linked the significant locus to six genes: *AQP4*,
306 *AQP4-AS*, *CIAPIN1P*, *KCTD1*, *RNU6-1289P*, and *U3* (S4 Fig. left). In fact the top associated SNP
307 is located in an intron of *KCTD1*. Brain gene expression based on GTEx shows high and brain-
308 specific expression for *AQP4* and moderate to high expression levels for *KCTD1* (S4 Fig, right).

Fig 3: Manhattan plot. The y-axis depicts the $-\log_{10}(\text{p-value})$ of the association between SNP and presence of IHI assuming an additive model in the discovery cohort. The SNPs tested in the study are ordered along their chromosomal position on the x-axis. The red horizontal line denotes genome wide significance at the Bonferroni threshold ($P=5e-8$), while the blue horizontal line marks the threshold for suggestive association ($P=1e-5$).



309

Fig 4: Regional association plot for rs9952569.



310

311 The gene set enrichment analysis showed that the four protein-coding genes prioritized
 312 by FUMA (*AQP4*, *AQP4-AS*, *KCTD1* and *U3*) were statistically significantly enriched
 313 ($p < 0.000314$; Bonferroni corrected for $3 * 53$ tests) in overexpressed genes in hippocampal and
 314 caudate tissue from GTEx v6 (S5 Fig).

315 We tested the 17 top SNPs within each suggestive loci in the PING cohort for an
 316 association with IHI. In PING, there were either genotyped or SNPs in near perfect LD ($r^2 > 0.9$)
 317 for seven loci, intermediate ($0.25 < r^2 < 0.9$) proxies for five loci and weak proxies for the
 318 remaining five loci ($r^2 < 0.25$). None of the selected candidate SNPs showed a nominal significant
 319 association (uncorrected $p < 0.05$) with IHI in this cohort. The top SNP from the discovery cohort,
 320 rs9952569, was not significant in the validation sample and showed an effect in the opposite
 321 direction (OR=0.597; $P=0.29$; Table 2).

322 **Table 2: Suggestive loci from discovery GWAS and results in validation cohort (below).** Left side: The table displays the leads
 323 SNPs in the discovery cohort (IMAGEN) with $p < 1e-5$ along with their chromosome, base pair (BP), alleles (A1, A2), subjects (N), odds
 324 ratio (OR), statistics (STAT) and P-value. Right side: The second table depicts the p-values for the SNPs (VALSNP) in the validation
 325 cohort (PING), along with their base pair, correlation to the discovery SNP (R2), alleles (A1, A2), OR and P-value.
 326

Discovery								
CHR	SNP	BP	A1	A2	NMISS	OR	STAT	P
1	rs79030318	147233364	A	G	1354	1.773	4.534	5.79E-06
1	rs3009985	212191991	G	A	1369	1.513	4.501	6.77E-06
2	rs10490445	37617484	G	T	1379	1.812	5.027	4.97E-07
4	rs6852934	7760243	A	G	1379	1.948	4.538	5.68E-06
5	rs77126180	6478922	A	G	1337	3.31	4.78	1.75E-06
6	rs35806781	37081139	A	G	1344	4.889	4.798	1.60E-06
7	rs11764012	8207615	A	G	1348	0.6388	-4.427	9.56E-06
8	rs1870396	9261456	G	C	1333	1.741	5.065	4.08E-07
8	rs2256087	11496193	C	A	1380	0.5757	-4.552	5.32E-06
12	rs491825	113048587	T	C	1344	1.522	4.487	7.21E-06
12	rs471231	118187337	A	G	1379	0.5948	-4.77	1.84E-06
17	rs75997523	6730523	A	G	1350	2.895	4.756	1.97E-06
17	rs11656431	32392629	A	G	1327	5.12	4.576	4.74E-06
18	rs9952569	24067227	C	T	1374	1.999	5.502	3.76E-08
20	rs6056644	9467164	C	A	1380	1.816	4.722	2.33E-06
20	rs6044984	17718017	A	G	1374	0.5503	-4.645	3.40E-06
20	rs117452506	50185618	A	G	1336	2.385	4.448	8.68E-06

Validation								
CHR	VALSNP	BP	R2	A1	A2	OR	STAT	P
1	rs12061877	147264252	0.755	C	T	1.5584465	1.0434	0.2331419
1	rs3738449	212240063	0.899	A	G	0.8060804	-0.7252	0.4743359
2	rs10490445	37617484	1	G	T	1.5846445	1.0391	0.1837016
4	rs2285769	7761286	1	A	G	1.4751031	0.5908	0.3685725
5	rs272456	6484012	0.042	A	G	0.7735566	-0.2829	0.3407685
6	rs2395670	37082809	0.011	A	G	1.3437002	1.1122	0.2280719
7	rs2058519	8214574	0.478	G	A	1.2305445	-0.0447	0.4622293
8	rs6601306	9264711	0.224	T	C	0.8644291	-1.1714	0.7694562
8	rs2256087	11496193	1	C	A	1.5806599	0.9401	0.116553
12	rs17824050	113036817	0.553	A	G	0.8728475	0.4514	0.6146593
12	rs471231	118187337	1	A	G	1.2764438	1.6962	0.3977656
17	rs12601392	6699944	0.537	G	A	1.2769892	0.4180	0.6042291
17	rs11658185	32394763	0.158	G	A	0.8852353	0.1711	0.7424022
18	rs9952569	24067227	1	C	T	0.5970534	-1.0367	0.2872871
20	rs6056647	9476953	0.976	A	G	1.4052022	0.8625	0.3847928
20	rs13042529	17698510	0.493	A	G	0.8046117	-1.4447	0.5881554
20	rs8121883	50197080	0.106	G	A	1.4445344	0.8726	0.2055499

327

328 Heritability analysis and genetic correlation

329 Heritability of IHI was estimated from the GWAS summary statistics using LD score regression.
 330 The h^2 estimate was 0.54 (0.30) and statistically significant using a one-sided test ($Z=1.8$;
 331 $P=0.036$). We next sought to identify genomics regions or cell type marker genes that show
 332 enriched heritability. However, none of the tested genomic regions or gene sets showed
 333 statistically significant enrichment after FDR correction (S6 Fig).

334 Motivated by reported increased prevalence of IHI in persons with epilepsy we computed
 335 the genetic correlation (r_g) between IHI and epilepsy susceptibility. The estimate of r_g was -0.0854
 336 (0.2612) which did not reach statistical significance ($Z=-0.3269$; $P=0.7437$).

337 LDhub was used to compute r_g between IHI and 832 GWAS summary statistics; the
 338 computation was successful for 749 GWAS (Supporting S1 Table). None of the traits survived the
 339 FDR corrected p-value threshold ($P_{FDR}<0.05$). A total of 20 traits reached nominal significance
 340 ($p<0.05$, Table 3). The top three positively correlated traits were: intelligence [58], College or
 341 University degree based on a UK BioBank (UKBB) GWAS and Years of Schooling [59]. Among
 342 the 20 nominal significant genetic correlations was also Fluid Intelligence Score (UKBB).

343
 344 **Table 3: Genetic correlations between IHI and other GWASs with nominal significance**
 345 **($p<0.05$).**

Trait	PMID	Category	r_g	se	z	p
Intelligence	28530673	cognitive	0.3479	0.139	2.5037	0.0123
Qualifications: College or University degree	0	ukbb	0.3472	0.1392	2.494	0.0126
Years of schooling 2016	27225129	education	0.3227	0.13	2.4828	0.013
HDL cholesterol	20686565	lipids	0.3992	0.1742	2.2923	0.0219
Ulcerative colitis	26192919	autoimmune	0.4354	0.1978	2.2015	0.0277
Started insulin within one year diagnosis of diabetes	0	ukbb	0.7574	0.3488	2.1716	0.0299
Mothers age at death	0	ukbb	0.6082	0.2801	2.1717	0.0299

Medication for cholesterol_ blood pressure or diabetes: Insulin	0	ukbb	1.2001	0.5571	2.154	0.0312
Creatinine	27005778	metabolites	0.7254	0.3513	2.0652	0.0389
Pain type(s) experienced in last month: Back pain	0	ukbb	-0.3682	0.1826	-2.0165	0.0438
Illnesses of father: Chronic bronchitis/emphysema	0	ukbb	-0.4744	0.2357	-2.0132	0.0441
Triglycerides in small VLDL	27005778	metabolites	0.643	0.3196	2.0123	0.0442
Age at Menopause	26414677	reproductive	0.4687	0.2331	2.0103	0.0444
Fluid intelligence score	0	ukbb	0.2584	0.1286	2.0094	0.0445
Smoking/smokers in household	0	ukbb	-0.5759	0.2881	-1.9987	0.0456
Tinnitus: Yes_ now some of the time	0	ukbb	-1.3776	0.6904	-1.9954	0.046
Concentration of small VLDL particles	27005778	metabolites	0.6411	0.3221	1.9904	0.0465
Time spent using computer	0	ukbb	0.2483	0.1249	1.9883	0.0468
Pack years adult smoking as proportion of life span exposed to smoking PREVIEW ONLY	0	ukbb	-0.3127	0.1587	-1.9702	0.0488
Serum total triglycerides	27005778	metabolites	0.6408	0.3258	1.9671	0.0492

346

347

348 **Gene expression in the developing human brain**

349 We extracted the gene expression levels in the developing brain for *AQP4* and *KCTD1*. The
350 summary of the data up to post-conception week (pcw) 100 are depicted in S7 Fig. Expression of
351 *KCTD1* remains rather stable across the entire time frame, while the expression of *AQP4* starts
352 very low and increases with the progression of brain maturation and reaches its peak around the
353 time of term birth (pcw 37-40). Gene expression was significantly different before and after pcw
354 25 for both *KCTD1* ($p=3.115e-04$) and *AQP4* ($p=2.424e-07$) when limited to data acquired before
355 pcw 40.

356

357 Discussion

358

359 Incidence rate of IHI was consistently around 25% in both, the discovery and the validation cohort.
360 This was comparable to previous reports of 18-19% [60,61], especially considering that the IHI
361 score at the 4.0 cutoff includes not only strong IHI (as in the cited studies), but also lighter IHI
362 [15], therefore increasing the IHI rate. In both cohorts there was a higher incidence rate of IHI in
363 the left hippocampus, which agrees well with the observations that the right hippocampus matures
364 faster and thereby inverts correctly. The GWAS highlighted one genome-wide significant locus on
365 chromosome 18, which is linked through chromatin interaction maps (Fig. S4, left) to six genes,
366 two of which show substantial expression in brain tissue: *KCTD1* and *AQP4*. Of note, the locus
367 showed consistently strong association with continuous scales of the IHI phenotype. Furthermore,
368 a genome-wide screen of those continuous scales revealed two genome-wide significant loci
369 (Fig. S3), one of which exceeded the suggestive threshold in the original GWAS (Table 2;
370 rs35806781, OR=4.889, Z=4.798, P=1.603e-06). The other SNP (rs186025034) had a low minor
371 allele frequency (about 1%) and missed the suggestive threshold in the original GWAS
372 (OR=5.159, Z=4.34, P=1.423e-05). Overall, we observed consistency across IHI definitions and
373 their genetic associations.

374 The Potassium Channel Tetramerization Domain Containing 1 (*KCTD1*) gene negatively
375 regulates the AP-2 family of transcription factors and the Wnt signalling pathway, which controls
376 normal embryonic development, cellular proliferation and growth [62]. Interestingly, mutations in
377 *KCTD1* have been linked to Scalp-Ear-Nipple syndrome [63], which is a rare, autosomal-dominant
378 disorder characterized by cutis aplasia of the scalp as well as minor anomalies of the external
379 ears, digits, nails, and malformations of the breast. Clearly, *KCTD1* has the ability to influence
380 developmental processes. Thus, it is conceivable that more benign variation in *KCTD1* may play
381 a role in the generation of IHI. Furthermore, *KCTD1* is a potassium channel gene and various

382 members of the potassium channel gene family have been linked as causes of epilepsy [64–66].
383 In a recent GWAS for epilepsy susceptibility SNPs in the *KCTD1* gene reached p-values as low
384 as 0.0003758 (S8 Fig) [56].

385 Aquaporin-4 (*AQP4*) is a bidirectional water channel that is found on astrocytes throughout
386 the central nervous system (S4 Fig). However, while *AQP4* expression in brain tissue is in general
387 high in children and adults, its expression is quite low before post-conception week 20 (S7 Fig).
388 MRI studies of IHI during brain development [20,23] show lack of hippocampal inversion during
389 the early phases of development <25 post-conception weeks, which coincides with the time point
390 of increased *AQP4* expression. Furthermore, *AQP4* has been linked through various lines of
391 evidence to epilepsy, e.g., the lack of aquaporin-4 water channels increased seizure threshold
392 and seizure duration in mice [67,68] and *AQP4* expression among chronic temporal lobe epilepsy
393 patients is increased almost twofold in the hippocampus of the affected hemisphere compared to
394 the contralateral hemisphere [69]. Taken together, astroglial *AQP4* may modulate neuronal
395 excitability by regulating the extraneuronal and extrasynaptic environments and thereby affect the
396 epileptogenesis. This may explain the observed increased rates of IHI in persons with epilepsy.
397 Interestingly, the four protein-coding genes prioritized by FUMA were also enriched in genes
398 overexpressed in hippocampal and caudate tissue (S5 Fig). However, enrichment results tend to
399 be unstable when only a small gene set is tested for enrichment, thus, this result should be
400 considered with caution regarding its interpretation.

401 We attempted to validate the genome-wide significant locus in a second independent
402 cohort of adolescents. However, neither the genome-wide significant locus, nor any of the
403 suggestive loci reached nominal significance in the validation cohort. One major contributor to this
404 lack of replication was the limited sample size of the validation cohort. Despite the equally large
405 set of participants in both studies, age and ethnicity restrictions severely limited the available
406 sample size for the validation cohort (N=161) and drastically lowered the statistical power to detect
407 differences (power of 35% for just the top variant). Larger validation cohorts are needed to confirm

408 the association of the identified locus with IHI: e.g., to validate the top variant with 80% at least
409 N=500 subjects are required. Although there are growing imaging and genetic datasets, e.g., the
410 UKBB that aims at 100,000 participants with genetics and brain imaging data, few studies focus
411 on healthy younger subjects (children, adolescents or young adults), which is beneficial for the
412 validation in order to exclude confounding by disease processes or age-related atrophy. One such
413 option is the Philadelphia Neurodevelopmental Cohort (PNC) [70]. However, it is important to
414 keep in mind that the evaluation of IHI is not restricted to adolescents. IHI can be observed in
415 children and adults too, without extra difficulties. The study here focus on adolescents, since the
416 discovery cohort is the dataset used for the reference study [15]. Still, in older patients, even
417 though IHI still exist, their detection may be more difficult and less reliable due to the confounding
418 effect of hippocampal atrophy to ageing, hippocampal sclerosis or neurodegeneration. Also
419 scoring bigger databases (such as UKBB) will be feasible only after automatic methods for IHI
420 scoring have been developed.

421 We estimated heritability of IHI based on the state-of-the-art LD score regression method
422 that operates on the GWAS summary statistics. The inferred heritability was substantial with
423 $h^2=0.54$; the estimate was subject to high uncertainty as reflected by the high standard deviation
424 of 0.3, which is likely a direct reflection of the low sample size of the discovery cohort. Analysis
425 on twin data, such as the healthy young adult twins participating in the Queensland Twin IMaging
426 (QTIM) study [71], can be used to confirm this preliminary heritability estimate, but would require
427 a significant effort in manually scoring IHI in these large cohorts. In addition, the magnitude of h^2
428 is comparable to recently published estimates on the heritability of hippocampal volume and
429 shape from more than 3600 subjects [25]; h^2 ranges from 0.08 to 0.337 depending on hemisphere
430 and structural measure, with heritability of volume being generally the lowest. Given the impact
431 of IHI on hippocampal shape and appearance together with the high prevalence of IHI in nearly
432 25% of healthy subjects, it is likely that the observed heritability of hippocampal shape reported
433 by [25] was in part due to IHI.

434 We used the generated genome-wide summary statistics for two additional explorations.
435 First, we sought to investigate if heritability was enriched in any particular region of the genome,
436 characterized by its function or by marker genes for specific cell types. None of the investigated
437 categories achieved statistical significance after FDR correction. Second, we computed genetic
438 correlations with other traits. We hypothesized that there may be genetic link with epilepsy,
439 however the resulting correlation was non-significant and negative, i.e., people with IHI were less
440 likely to be affected by epilepsy, thereby contradicting earlier reports. The exploratory analysis
441 with 832 additional traits highlighted a positive genetic correlation between IHI and intelligence
442 and education attainment. There are various reports highlighting the contribution of the
443 hippocampus and its subregions to various mental aspects that collectively are referred to as
444 intelligence, e.g., spatial processing [72] and working memory [73]. Moreover, one recent study
445 linked hippocampal shape to cognitive performance [74]. In particular, in males the radial distance
446 of the hippocampus correlated with better test scores (e.g., general factor of intelligence, abstract-
447 fluid intelligence, and the rotation of solid figures). In females, the effect was reversed. Therefore,
448 a genetic correlation of IHI with intelligence in the broad sense is conceivable.

449 In conclusion, we presented the first genome-wide association study of IHI, where we
450 identified a genome-wide significant locus. Additional exploration of the resulting summary
451 statistics revealed a high heritability and suggested positive genetic correlation of IHI with traits
452 linked to intelligence and education attainment.

453 **Acknowledgements**

454 MAS acknowledges financial support by the EPSRC-funded UCL Centre for Doctoral Training in
455 Medical Imaging (EP/L016478/1). AA holds an MRC eMedLab Medical Bioinformatics Career
456 Development Fellowship. This work was supported by the Medical Research Council [grant
457 number MR/L016311/1]. The research leading to these results has received funding
458 from the program “Investissements d’avenir” ANR-10-IAIHU-06 (Agence Nationale de la
459 Recherche-10-IA Institut Hospitalo-Universitaire-6) and from the “Contrat d’Interface Local”
460 program (to OC) from Assistance Publique-Hôpitaux de Paris (AP-HP).

461
462 Data and/or research tools used in the preparation of this manuscript were obtained and analyzed
463 from the controlled access datasets distributed from the NIMH-supported research Domain
464 Criteria Database (RDoCdb). RDoCdb is a collaborative informatics system created by the
465 National Institute of Mental Health to store and share data resulting from grants funded through
466 the Research Domain Criteria (RDoC) project. Dataset identifier(s): #2607.

467 Data collection and sharing for this project was funded by the Pediatric Imaging, Neurocognition
468 and Genetics Study (PING) (National Institutes of Health Grant RC2DA029475).

469
470 IMAGEN was supported by the European Union-funded FP6 (LSHM-CT-2007-037286).

471 **IMAGEN Consortium author list:**

472 Contact: Gunter Schumann (gunter.schumann@kcl.ac.uk)

473 Tobias Banaschewski M.D., Ph.D.¹⁸; Gareth J. Barker* Ph.D.¹⁹; Arun L.W. Bokde Ph.D.²⁰; Uli
474 Bromberg Ph.D.²¹; Christian Büchel M.D.²¹; Erin Burke Quinlan, PhD¹⁶; Sylvane Desrivières
475 Ph.D.¹⁶; Herta Flor Ph.D.^{22,23}; Hugh Garavan Ph.D.²⁴; Penny Gowland Ph.D.²⁵; Bernd Ittermann
476 Ph.D.²⁶; Marie-Laure Paillère Martinot M.D., Ph.D.²⁷; Eric Artiges M.D., Ph.D.²⁸; Frauke Nees
477 Ph.D.^{18,22}; Dimitri Papadopoulos Orfanos Ph.D.⁹; Tomáš Paus M.D., Ph.D.²⁹; Luise Poustka
478 M.D.³⁰; Sarah Hohmann M.D.¹⁸; Sabina Millenet Dipl.-Psych.¹⁸; Juliane H. Fröhner Dipl.-Psych.¹⁵;
479 Robert Whelan Ph.D.³¹

480 ¹⁸Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental
481 Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim,
482 Germany; ¹⁹Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience,

483 King's College London, United Kingdom; ²⁰Discipline of Psychiatry, School of Medicine and Trinity
484 College Institute of Neuroscience, Trinity College Dublin; ²¹University Medical Centre Hamburg-
485 Eppendorf, House W34, 3.OG, Martinstr. 52, 20246, Hamburg, Germany; ²²Department of
486 Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty
487 Mannheim, Heidelberg University, Square J5, Mannheim, Germany; ²³ Department of
488 Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany;
489 ²⁴Departments of Psychiatry and Psychology, University of Vermont, 05405 Burlington, Vermont,
490 USA; ²⁵Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of
491 Nottingham, University Park, Nottingham, United Kingdom; ²⁶Physikalisch-Technische
492 Bundesanstalt (PTB), Braunschweig and Berlin, Germany [or depending on journal requirements
493 can be: Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2 – 12, Berlin, Germany;
494 ²⁷Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 “Neuroimaging &
495 Psychiatry”, University Paris Sud, University Paris Descartes; Sorbonne Université; and AP-HP,
496 Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France; ²⁸Institut
497 National de la Santé et de la Recherche Médicale, INSERM Unit 1000 “Neuroimaging &
498 Psychiatry”, University Paris Sud, University Paris Descartes – Sorbonne Paris Cité; and
499 Psychiatry Department 91G16, Orsay Hospital, France; ²⁹Bloorview Research Institute, Holland
500 Bloorview Kids Rehabilitation Hospital and Departments of Psychology and Psychiatry, University
501 of Toronto, Toronto, Ontario, M6A 2E1, Canada; ³⁰Department of Child and Adolescent
502 Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075,
503 Göttingen, Germany; ³¹School of Psychology and Global Brain Health Institute, Trinity College
504 Dublin, Ireland;
505

506 **Author contributions**

507 CC and AA are guarantor of integrity of the entire study. All authors contributed to manuscript
508 drafting or manuscript revision for important intellectual content. All authors gave their approval
509 of the final version of submitted manuscript. CC, MAS, RT, AA and OC did literature research,
510 designed the study, performed experiments and analysed data. CC, MAS, AA and OC wrote the
511 article. VF contributed to genotyping data, JBP contributed to the statistical processing of data.
512 EA, AG, AH, HW, HL, JLM, MS and GS contributed to the clinical study and data acquisition.
513

514 **Ethics committee:**

515 London: Psychiatry, Nursing and Midwifery (PNM) Research Ethics Subcommittee (RESC),
516 Waterloo Campus, King's College London. Nottingham: University of Nottingham Medical School
517 Ethics Committee. Mannheim: Medizinische Fakultät Mannheim, Ruprecht Karl Universität
518 Heidelberg and Ethik-Kommission II an der Fakultät für Klinische Medizin Mannheim.
519 Dresden: Ethikkommission der Medizinischen Fakultät Carl Gustav Carus, TU Dresden
520 Medizinische Fakultät. Hamburg: Ethics board, Hamburg Chamber of Physicians. Paris: CPP
521 IDF VII (Comité de protection des personnes Ile de France), ID RCB: 2007-A00778-45 September
522 24th 2007. Dublin: TCD School of Psychology REC. Berlin: ethics committee of the Faculty of
523 Psychology. And Mannheim's ethics committee approved the whole study. All informed consents
524 have been obtained from a parent and/or legal guardian. Written informed assent and consent
525 were obtained, respectively, from all adolescents and their parents after complete description of
526 the study. For the PING dataset, written parental informed consent was obtained for all PING
527 subjects below the age of 18, and child assent was also obtained for all participants between the
528 ages of 7 and 17. Written informed consent was obtained directly from all participants aged 18
529 years or older.

530

531 **References**

532

- 533 1. Burgess N, Maguire EA, O'Keefe J. The Human Hippocampus and Spatial and Episodic
534 Memory. *Neuron*. 2002 Aug;35[4]:625–41.
- 535 2. Jankord R, Herman JP. Limbic Regulation of Hypothalamo-Pituitary-Adrenocortical Function
536 during Acute and Chronic Stress. *Annals of the New York Academy of Sciences*. 2008
537 Dec;1148[1]:64–73.
- 538 3. Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, et al. A meta-analysis of
539 hippocampal atrophy rates in Alzheimer's disease. *Neurobiology of Aging*. 2009
540 Nov;30[11]:1711–23.
- 541 4. Colle R, Cury C, Chupin M, Deflesselle E, Hardy P, Nasser G, et al. Hippocampal volume
542 predicts antidepressant efficacy in depressed patients without incomplete hippocampal
543 inversion. *NeuroImage: Clinical*. 2016 Feb 1;12:949–55.
- 544 5. Sämann PG, Höhn D, Chechko N, Kloiber S, Lucae S, Ising M, et al. Prediction of
545 antidepressant treatment response from gray matter volume across diagnostic categories.
546 *European Neuropsychopharmacology*. 2013 Nov;23[11]:1503–15.

- 547 6. Colle R, Chupin M, Cury C, Vandendrie C, Gressier F, Hardy P, et al. Depressed suicide
548 attempters have smaller hippocampus than depressed patients without suicide attempts.
549 *Journal of Psychiatric Research*. 2015 Feb;61:13–8.
- 550 7. Travis SG, Coupland NJ, Hegadoren K, Silverstone PH, Huang Y, Carter R, et al. Effects of
551 cortisol on hippocampal subfields volumes and memory performance in healthy control
552 subjects and patients with major depressive disorder. *Journal of Affective Disorders*. 2016
553 Sep;201:34–41.
- 554 8. van Erp TGM, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA, et al.
555 Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540
556 healthy controls via the ENIGMA consortium. *Mol Psychiatry*. 2016 Apr;21[4]:547–53.
- 557 9. Helmstaedter C, Elger CE. Chronic temporal lobe epilepsy: a neurodevelopmental or
558 progressively dementing disease? *Brain*. 2009 Oct 1;132[10]:2822–30.
- 559 10. Qiu A, Rifkin-Graboi A, Chen H, Chong Y-S, Kwek K, Gluckman PD, et al. Maternal anxiety
560 and infants' hippocampal development: timing matters. *Transl Psychiatry*. 2013
561 Sep;3[9]:e306–e306.
- 562 11. Bronen RA, Cheung G. MRI of the normal hippocampus. *Magnetic Resonance Imaging*.
563 1991;9[4]:497–500.
- 564 12. Lehericy S, Dormont D, Sémah F, Clémenceau S, Granat O, Marsault C, et al. Developmental
565 abnormalities of the medial temporal lobe in patients with temporal lobe epilepsy. *AJNR*
566 *American journal of neuroradiology*. 1995;16[4]:617–26.
- 567 13. Baulac M, De Grissac N, Hasboun D, Oppenheim C, Adam C, Arzimanoglou A, et al.
568 Hippocampal developmental changes in patients with partial epilepsy: magnetic resonance
569 imaging and clinical aspects. *Annals of neurology*. 1998;44:223–33.
- 570 14. Bernasconi N, Kinay D, Andermann F, Antel S, Bernasconi A. Analysis of shape and
571 positioning of the hippocampal formation: an MRI study in patients with partial epilepsy and
572 healthy controls. *Brain : a journal of neurology*. 2005;128:2442–2452.
- 573 15. Cury C, Toro R, Cohen F, Fischer C, Mhaya A, Samper-González J, et al. Incomplete
574 Hippocampal Inversion: A Comprehensive MRI Study of Over 2000 Subjects. *Frontiers in*
575 *neuroanatomy*. 2015;9:160.
- 576 16. Cury C. Statistical shape analysis of the anatomical variability of the human hippocampus in
577 large populations. Sorbonne Université; 2015.
- 578 17. Colenutt J, McCann B, Knight MJ, Coulthard E, Kauppinen RA. Incomplete Hippocampal
579 Inversion and Its Relationship to Hippocampal Subfield Volumes and Aging: Incomplete
580 Hippocampal Inversion and Aging. *Journal of Neuroimaging*. 2018 Jul;28[4]:422–8.
- 581 18. Kim H, Chupin M, Colliot O, Bernhardt BC, Bernasconi N, Bernasconi A. Automatic
582 hippocampal segmentation in temporal lobe epilepsy: Impact of developmental
583 abnormalities. *NeuroImage*. 2012 Feb 15;59[4]:3178–86.

- 584 19. Okada Y, Kato T, Iwai K, Iwasaki N, Ohto T, Matsui A. Evaluation of hippocampal infolding
585 using magnetic resonance imaging. *NeuroReport*. 2003;14[10]:1405–9.
- 586 20. Bajic D, Ewald U, Raininko R. Hippocampal development at gestation weeks 23 to 36. An
587 ultrasound study on preterm neonates. *Neuroradiology*. 2010;52[6]:489–94.
- 588 21. Baker LL, Barkovich AJ. The large temporal horn: MR analysis in developmental brain
589 anomalies versus hydrocephalus. *American Journal of Neuroradiology*. 1992;13[1]:115–22.
- 590 22. Righini A, Zirpoli S, Parazzini C, Bianchini E, Scifo P, Sala C, et al. Hippocampal infolding
591 angle changes during brain development assessed by prenatal MR imaging. *American*
592 *Journal of Neuroradiology*. 2006;27[10]:2093–7.
- 593 23. Bajic D, Canto Moreira N, Wikström J, Raininko R. Asymmetric development of the
594 hippocampal region is common: A fetal MR imaging study. *American Journal of*
595 *Neuroradiology*. 2012;33[3]:513–8.
- 596 24. Andrade D, Krings T, Chow EWC, Kiehl TR, Bassett AS. Hippocampal malrotation is
597 associated with chromosome 22q11.2 microdeletion. *Canadian Journal of Neurological*
598 *Sciences*. 2013;40[5]:652–6.
- 599 25. Roshchupkin G V., Gutman BA, Vernooij MW, Jahanshad N, Martin NG, Hofman A, et al.
600 Heritability of the shape of subcortical brain structures in the general population. *Nature*
601 *Communications*. 2016;7:13738.
- 602 26. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN
603 study: reinforcement-related behaviour in normal brain function and psychopathology.
604 *Molecular psychiatry*. 2010;15[12]:1128–39.
- 605 27. Jernigan TL, Brown TT, Hagler DJ, Akshoomoff N, Bartsch H, Newman E, et al. The Pediatric
606 Imaging, Neurocognition, and Genetics (PING) Data Repository. *NeuroImage*. 2016;124[Pt
607 B]:1149–54.
- 608 28. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain
609 images. *Medical Image Analysis*. 2001;5[2]:143–56.
- 610 29. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and
611 accurate linear registration and motion correction of brain images. *NeuroImage*.
612 2002;17[2]:825–41.
- 613 30. The International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E,
614 Schaffner SF, et al. Integrating common and rare genetic variation in diverse human
615 populations. *Nature*. 2010;467:52–58.
- 616 31. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait
617 analysis. *American journal of human genetics*. 2011;88[1]:76–82.
- 618 32. McCarthy S, Das S, Kretschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference
619 panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*. 2016;48:1279–1283.

- 620 33. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al.
621 Reference-based phasing using the Haplotype Reference Consortium panel. *Nature*
622 *Genetics*. 2016;48[11]:1443–8.
- 623 34. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler
624 transform (PBWT). *Bioinformatics*. 2014;30[9]:1266–72.
- 625 35. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
626 individuals. *Genome Research*. 2009;19[9]:1655–1664.
- 627 36. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation
628 PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4[1].
- 629 37. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom:
630 regional visualization of genome-wide association scan results. *Bioinformatics*.
631 2010;26[18]:2336–7.
- 632 38. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific
633 haplotype structure and linking correlated alleles of possible functional variants.
634 *Bioinformatics*. 2015;31[21]:3555–7.
- 635 39. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation
636 of genetic associations with FUMA. *Nature communications*. 2017;8[1]:1826.
- 637 40. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot
638 analysis: multitissue gene regulation in humans. *Science (New York, NY)*.
639 2015;348[6235]:648–60.
- 640 41. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic
641 identification of trans eQTLs as putative drivers of known disease associations. *Nature*
642 *Genetics*. 2013;45:1238–1243.
- 643 42. Zhernakova D V., Deelen P, Vermaat M, Van Iterson M, Van Galen M, Arindrarto W, et al.
644 Identification of context-dependent expression quantitative trait loci in whole blood. *Nature*
645 *Genetics*. 2017;49:139–145.
- 646 43. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability
647 in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience*.
648 2014;17[10]:1418–28.
- 649 44. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact
650 Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*. 2016;17:2042–
651 2059.
- 652 45. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et
653 al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–330.
- 654 46. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP.
655 Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739–1740.

- 656 47. Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen EL, Bohler A, et al. WikiPathways:
657 Capturing the full diversity of pathway knowledge. *Nucleic Acids Research*. 2016;44:D488–
658 D494.
- 659 48. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS
660 Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*.
661 2014;42:D1001–D1006.
- 662 49. Bulik-Sullivan B, Loh P, Finucane H, Ripke S. LD Score regression distinguishes confounding
663 from polygenicity in genome-wide association studies. *Nature*. 2015;
- 664 50. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning
665 heritability by functional annotation using genome-wide association summary statistics.
666 *Nature genetics*. 2015;47[11]:1228–35.
- 667 51. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability
668 enrichment of specifically expressed genes identifies disease-relevant tissues and cell
669 types. *Nature genetics*. 2018;50[4]:621–9.
- 670 52. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, et al. A
671 transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for
672 understanding brain development and function. *The Journal of neuroscience : the official*
673 *journal of the Society for Neuroscience*. 2008;28[1]:264–78.
- 674 53. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological
675 interpretation of genome-wide association studies using predicted gene functions. *Nature*
676 *communications*. 2015;6:5890.
- 677 54. Heng TSP, Painter MW, Consortium IGP. The Immunological Genome Project: networks of
678 gene expression in immune cells. *Nature immunology*. 2008;9[10]:1091–4.
- 679 55. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic
680 correlations across human diseases and traits. *Nature genetics*. 2015;47[11]:1236–41.
- 681 56. The International League Against Epilepsy Consortium on Complex Epilepsies. Genetic
682 determinants of common epilepsies: a meta-analysis of genome-wide association studies.
683 *Lancet neurology*. 2014;13[9]:893–903.
- 684 57. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a
685 centralized database and web interface to perform LD score regression that maximizes the
686 potential of summary level GWAS data for SNP heritability and genetic correlation analysis.
687 *Bioinformatics*. 2017;33[2]:272–9.
- 688 58. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, et al. Genome-
689 wide association meta-analysis of 78,308 individuals identifies new loci and genes
690 influencing human intelligence. *Nature Genetics*. 2017;49[7]:1107–12.
- 691 59. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide
692 association study identifies 74 loci associated with educational attainment. *Nature*.
693 2016;533[7604]:539–42.

- 694 60. Bajic D, Wang C, Kumlien E, Mattsson P, Lundberg S, Eeg-Olofsson O, et al. Incomplete
695 inversion of the hippocampus - A common developmental anomaly. *European Radiology*.
696 2008;18[1]:138–42.
- 697 61. Bajic D, Kumlien E, Mattsson P, Lundberg S, Wang C, Raininko R. Incomplete hippocampal
698 inversion - Is there a relation to epilepsy? *European Radiology*. 2009;19[10]:2544–50.
- 699 62. Li X, Chen C, Wang F, Huang W, Liang Z, Xiao Y, et al. KCTD1 Suppresses Canonical Wnt
700 Signaling Pathway by Enhancing β -catenin Degradation. *PLoS ONE*. 2014;9[4]:e94343.
- 701 63. Marneros AG, Beck AE, Turner EH, McMillin MJ, Edwards MJ, Field M, et al. Mutations in
702 KCTD1 cause scalp-ear-nipple syndrome. *American Journal of Human Genetics*.
703 2013;92[4]:621–6.
- 704 64. Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, Melis R, et al. A novel potassium
705 channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nature Genetics*.
706 1998;18:25–29.
- 707 65. Jentsch TJ. Neuronal KCNQ potassium channels: physiology and role in disease. *Nature*
708 *reviews Neuroscience*. 2000;1:21–30.
- 709 66. Liu Z, Xiang Y, Sun G. The KCTD family of proteins: structure, function, disease relevance.
710 *Cell & bioscience*. 2013;3[45].
- 711 67. Binder DK, Yao X, Verkman AS, Manley GT. Increased seizure duration in mice lacking
712 aquaporin-4 water channels. *Acta Neurochirurgica, Supplementum*. 2006;96.
- 713 68. Binder DK, Oshio K, Ma T, Verkman AS, Manley GT. Increased seizure threshold in mice
714 lacking aquaporin-4 water channels. *NeuroReport*. 2004;15[2]:259–62.
- 715 69. Das A, Wallace GC, Holmes C, McDowell ML, Smith JA, Marshall JD, et al. Hippocampal
716 tissue of patients with refractory temporal lobe epilepsy is associated with astrocyte
717 activation, inflammation, and altered expression of channels and receptors. *Neuroscience*.
718 2012;220:237–46.
- 719 70. Satterthwaite TD, Elliott MA, Ruparel K, Loughhead J, Prabhakaran K, Calkins ME, et al.
720 Neuroimaging of the Philadelphia neurodevelopmental cohort. *NeuroImage*. 2014;86:544–
721 53.
- 722 71. de Zubicaray GI, Chiang M-C, McMahon KL, Shattuck DW, Toga AW, Martin NG, et al.
723 Meeting the Challenges of Neuroimaging Genetics. 2008;2[4]:258–63.
- 724 72. Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RSJ, et al.
725 Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the*
726 *National Academy of Sciences*. 2000;97[8]:4398–403.
- 727 73. Axmacher N, Henseler MM, Jensen O, Weinreich I, Elger CE, Fell J. Cross-frequency
728 coupling supports multi-item working memory in the human hippocampus. *Proceedings of*
729 *the National Academy of Sciences*. 2010;107[7]:3228–33.

730 74. Colom R, Stein JL, Rajagopalan P, Martínez K, Hermel D, Wang Y, et al. Hippocampal
731 structure and human cognition: Key role of spatial processing and evidence supporting the
732 efficiency hypothesis in females. *Intelligence*. 2013;41[2]:129–140.

733

734 **Supporting information**

735 **S1 Fig. Ancestry matching of the IMAGEN cohort.** The panels show pairwise scatter plots for
736 the first five principal components of the genetic relationship matrix. The HapMap3 subjects are
737 represented as filled circles and color coded by population (see legend in the top left panel).
738 IMAGEN participants are represented by black open circles; red open circles indicate subjects
739 that were excluded based on distance to the CEU+TSI European ancestry.

740

741 **S2 Fig. Quantile-quantile plot.** QQ plot for the genome-wide association study. The x-axis
742 shows the expected $-\log_{10}(p\text{-value})$ while the y-axis shows the study $-\log_{10}(p\text{-value})$.
743 There was no evidence of p-value inflation $\lambda=1.017$.

744

745 **S3 Fig. Manhattan plots for continuous IHI scores.** The y-axis depicts the $-\log_{10}(p\text{-value})$ of
746 the association between SNP and presence of IHI assuming an additive model in the discovery
747 cohort. The SNPs tested in the study are ordered along their chromosomal position on the x-axis.
748 The red horizontal line denotes genome wide significance at the Bonferroni threshold ($P=5e-8$),
749 while the blue horizontal line marks the threshold for suggestive association ($P=1e-5$). Upper plot:
750 Result for the sum of the five criteria and then maxed over left and right hippocampus. Two loci
751 exceed the genome wide significant threshold: on chromosome 6 rs35806781 ($\beta=1.478$,
752 $Z=5.766$, $P=1.006e-08$) and on chromosome 9 rs186025034 ($\beta=1.867$, $Z=6.202$, $P=7.408e-$
753 10). Lower plot: Result for the global criterion presenting 3 classes: 0 = non IHI, 2 = IHI, 1 = partial
754 IHI, and taking the max over left and right hippocampus. One locus exceeds genome-wide
755 significance on chromosome 9: rs186025034 ($\beta=0.8251$, $Z=5.575$, $P=2.98e-08$).

756

757

758 **S4 Fig. FUMA results.** Left: chromatin interaction plot, mapping the genome-wide significant
759 locus to six genes on chromosome 18. Right: expression heatmap (average $\log_2(\text{RPKM})$ in 53
760 GTEx tissues) for the four mapped protein-coding genes.

761

762 **S5 Fig. Enrichment analysis of prioritized protein-coding genes in GTEx.** The four prioritized
763 genes are tested for enrichment in tissue-specific gene lists for 53 tissue from the GTEx dataset.
764 Each list of differentially expressed genes is split into overexpressed and underexpressed genes.
765 P-value threshold for significance was the Bonferroni corrected threshold for $3 * 53$ tests
766 ($p < 0.000314$). Tissues showing significant enrichment are indicated by red bars.

767

768 **S6 Fig. Heritability partitioning analysis:** panels 1-5, among cell/tissue-specific differentially
769 expressed genes; panel 6, among genomic functional annotation categories.

770

771 **S7 Fig. Gene expression in the developing human brain for AQP4 and KCTD1.** The x-axis
772 shows the subjects' age in weeks post conception (pcw). The y-axis depicts the mean of the log₂
773 RPKM (reads per kilobase per million) provided by BrainSpan. AQP4 and KCTD1 are indicated
774 by different colors.

775

776 **S8 Fig. Local association plot for KCTD1 in ILAE epilepsy GWAS.**

777

778 **S1 Table. Genetic correlation between IHI and 832 GWASs obtained from LD hub.**

779

780 **S1 File. Imagen Consortium author list.**