A *DACT1* enhancer modulates brain asymmetric temporal regions involved in language processing

Yann Le Guen[a,1], François Leroy[b], Cathy Philippe[a], IMAGEN consortium, Jean-François Mangin[a], Ghislaine Dehaene-Lambertz[b,2], Vincent Frouin[a,1,2]

[a]UNATI; [b]Cognitive Neuroimaging Unit, U992, INSERM; Neurospin, Institut Joliot, CEA, Université Paris-Saclay, Gif-sur-Yvette 91191, France
Author contributions: Y.L.G., F.L., G.D.L and V.F. designed the research; Y.L.G. performed the research; IMAGEN consortium contributed to replication data; Y.L.G., F.L., C.P., J.F.M., G.D.L and V.F. analyzed the results; and Y.L.G., C.P., G.D.L. and V.F. wrote the paper.
[1]To whom correspondence may be addressed. Email: yleguen@stanford.edu or vincent.frouin@cea.fr
[2]G.D.L and V.F. contributed equally to this work.

Abstract

Uncovering the genes that contribute to the variability in brain regions involved in language processing might shed light on the evolution of brain structures essential to the emergence of language in Homo sapiens. The superior temporal asymmetrical pit (STAP), which is not observed in chimpanzees, represents an ideal phenotype to investigate the genetic variations that support human communication. The depth of the left STAP was significantly associated with a *DACT1* enhancer region in the UK Biobank British discovery sample (N=16,515). This association was replicated in the IMAGEN cohort (N=1,726) and the UK Biobank non-British validation sample (N=2,161). This genomic region was also associated to a lesser extent with the right STAP depth and the formation of sulcal interruptions, *plis de passage*, in the bilateral STAP but not with other structural brain MRI phenotypes, highlighting its specific expression in the superior temporal regions. Diffusion MRI emphasized an association with the fractional anisotropy of the left auditory fibers of the corpus callosum and with networks involved in linguistic processing in resting-state functional MRI. Finally, *DACT1* is mainly expressed during the first two trimesters of pregnancy just before the superior temporal sulcus becomes visible and the first evidence of speech processing is observed in preterm neonates. Overall, this evidence demonstrates a significant, specific relationship between variants in this genomic region and the establishment of the superior temporal regions that support human communication.

Significance statement

The search for "language genes" has mainly focused on families with language disorders. Here, using a macroscopic marker of cerebral asymmetry observed only in humans, the STAP, we discovered a genetic variant that modulates many anatomical and functional features of the superior temporal sulcus, particularly in the left hemisphere, in the general population. Whole-brain analyses showed that this modulation is nearly limited to the superior temporal region. Finally, the gene effect is most significant during the first six months of gestation, decreasing just before the appearance of thalamocortical connectivity that enables external auditory inputs. Thus, we uncovered a genomic region that probably contributes to building the networks underlying human language.

Introduction

How did humans evolve a complex communication system? This question remains unanswered. The first approach has been to look for rare monogenic mutations in families with language and speech disorders, revealing the role of *FOXP2* (1) and *KIAA0319* (2) for example. Although these genes have sometimes been labeled "language genes", this complex faculty cannot be reduced to a few abruptly mutated genes. Language in modern humans results from multiple evolutionary adaptations at different neural levels, from sensory-motor capacities to produce and perceive speech to abstract abilities to maintain and combine symbolic representations in hierarchical sequences to convey conceptual content. Nevertheless, the discovery of these genes has generated productive research in humans and animals in the following directions: first, which brain areas do these genes modulate either structurally or functionally in humans; and second, how do animal variants and even the human variants in animals affect neural networks and animal abilities? This research may result in the proposal of neuro-cognitive mechanisms supporting language and communication. For example, *FOXP2* is currently seen as an important gene in vocal production in birds, mice and humans in agreement with the modulation of activity observed in the basal ganglia and inferior frontal region in human carriers of its mutation (3) and allelic variants (4).

Another approach, which is now possible due to genotyping and neuroimaging in large cohorts, is to investigate the relation between certain brain structures and allelic variations in the normal human population (5). Although all participants have language, this type of research can pinpoint particular genetic and neural architectures, opening new windows to understand language. For example, Dediu and Ladd (6) proposed that typological features of languages might depend on certain allele frequencies in human groups, and they explored the relation between tonal languages and the distribution of allelic variants of *ASPM* and *Microcephalin*. Functional and structural asymmetries might be one of the human brain phenotypes whose genetic study might shed light on this issue. Indeed, human communication is based on the retrieval of two crucial types of information in speech, who speaks and what is said. In most adults, the message and the messenger are processed in parallel in the superior temporal region, with the linguistic message (7) and voice identification (8) processed in the left and right hemispheres, respectively. This channeling is present early in human infants (9, 10). At six months of gestation, activations to syllables are differently lateralized when preterm infants discriminate a phonetic change (/ba/ vs /ga/) or a voice change (male vs female) (11). Thus, the left and right peri-sylvian regions are distinctly functionally organized from the onset of thalamocortical connectivity. Therefore, this organization occurs before sensory input can shape these regions, suggesting genetically determined protomaps.

These functional asymmetries should be discussed in relation to the notable asymmetries visible at the macroscopic level and reported near the Sylvian fissure. The Sylvian fissure is more dorsal and oblique in the right hemisphere, the *planum temporale* is larger in the left hemisphere (12) and the right superior temporal sulcus (STS) is deeper than the left (13). Some asymmetries, such as the shape and location of the Sylvian fissure and the left-right size difference in the *planum,* are observed in apes (14), but the STS depth asymmetry is the first macroscopic brain asymmetry that is observed only in humans. A 45 mm-long segment located under Heschl's gyrus, superior temporal asymmetrical pit (STAP), is deeper in the right hemisphere than in the left in 96% of human infants, children, adults but not in chimpanzees (13). During gestation, the right STS appears earlier than the left similar to many other sulci (15, 16), but other left sulci catch up, and the left STS remains shallower than the right and is more often interrupted by one or more *plis de passage* (PPs) (13, 17). The observation of these structural and functional differences before term

age suggests that genetically driven mechanisms guide the development of peri-sylvian regions to endorse the complexity of verbal and nonverbal communication in the human species. Because of the sharp distinction between chimpanzees and humans, Leroy et al. (13) suggested that the STAP might represent an important macroscopic marker of the recent cytoarchitectonic changes (18) in the primate lineage supporting the communication function in humans. We have previously examined the heritability (i.e., the genetic proportion in phenotypic variance across individuals) of STS characteristics. Because sulcal pits correspond to the deeper parts of sulci where the initial or primary folds would occur, they show less intersubject variability (19, 20). In the left STAP area, the two sulcal pit regions (STS b and c in (19)) are highly heritable in contrast to the right STS yielding the largest asymmetric estimate of heritability for all brain sulci. In addition, the PPs that interrupt some sulci are more frequent and heritable in the left STS ($h^2 = 0.53$) than in the right ($h^2 = 0.27$) (17). In most brain regions, this interhemispheric difference in heritability estimates was not detected, therefore suggesting a distinctive genetic modulation in the superior temporal lobes.

Therefore, to clarify the genetic influence on the development of the human communication system, we studied the genetic variants significantly associated with STAP phenotypes (i.e., the right and left sulcal depths, their asymmetry and the formation of PPs). We took advantage of the most recent release of the UK Biobank, which includes approximately 22,000 individuals genotyped and scanned with the same MRI scanner and imaging protocol (21). We used 16,515 genetically confirmed individuals of British ancestry as our discovery sample and the remaining 2,161 non-British ancestry subjects as a first replication sample. In addition, we used the IMAGEN cohort of 1,726 adolescents (14-15 years) as the second replication sample. We found that a region upstream of *DACT1* on 14q23.1 strongly modulates STS phenotypes. To further understand the functional role of this region, we first used databases collecting information on gene tissue expression and chromatin states in various brain tissues at different stages of development. Second, we investigated the effect of this region on brain phenotypes by taking advantage of the multimodal sequences acquired in the same UK Biobank subjects (T1w, diffusion tensor imaging (DTI) and resting state). We observed that this gene has a localized effect on the superior temporal regions, particularly on the left STS, which suggests a role in the establishment of the linguistic network.

Results

Heritability and genome-wide association. The first two columns in **Table 1** present the heritability estimates of STAP phenotypes in the following populations: the extended pedigree of the Human Connectome Project (HCP) using SOLAR (22) and unrelated individuals in the UK Biobank using GCTA (23). Although weaker due to the different cohort design, the higher heritability estimates in the left hemisphere of the depth potential function (PF) and presence of PP previously obtained in HCP (17) were reproduced in UK Biobank individuals (details in **Table S1**). The depth asymmetry index was also heritable. We computed the genome-wide association (GWA) with these phenotypes in the UK Biobank discovery sample (see **Fig. S1-S3** for the Manhattan and QQ plots of these association tests). A region upstream (26.5 kb) of the *DACT1* gene in the cytogenetic band 14q23.1 contains single nucleotide polymorphisms (SNPs) whose associated p-values passed the Bonferroni corrected genomic threshold ($p < 5 \cdot 10^{-8}/8 = 6.25 \cdot 10^{-9}$) (**Fig. 1A**). The strongest association in imputed SNPs was with rs160458 ($p = 2.0 \cdot 10^{-12}$, N = 16,203, base pair (bp) position = 59074878 (GRCh37)) and in genotyped SNPs with rs160459 ($p = 4.2 \cdot 10^{-12}$, N=16,515, base pair position = 59074136), which are both annotated in Ensembl (24) as "human-specific base". Strong

associations ($10^{-6} < p < 10^{-8}$) were observed for other left/right STAP phenotypes, while the asymmetry index of the geodesic depth and DPF were significant at $p < 0.05$ (uncorrected). In the following section, we focus on the genotyped SNP rs160459 because its chromatin state in the brain appears more consistent with a regulatory role (**Fig. 1B**), and both SNPs are in strong linkage disequilibrium ($r^2 = 0.936$ in British ancestry, distance in build GRCh37 = 742 bp).

Replication of the association in two other populations (**Table 1** for rs160459 and **Table S2** for rs160458). The association with the STAP left geodesic depth was replicated in the two samples above the nominal significance level ($p < 0.05$), with $p = 8.0 \cdot 10^{-5}$ in the IMAGEN cohort and $p = 0.041$ in the non-British UK Biobank sample. On average, the allelic repartition across the three population samples is 21.6% 'CC', 49.9% 'CA', and 28.5% 'AA'. **Fig. S4** illustrates the effect size of the rs160459 allelic configuration on the geodesic depth and depth PF in the three populations revealing shallower STS in the 'AA' allelic configuration. **Fig. 2A** shows the influence of rs160459 along the STAP geodesic depth profile. Its allelic configuration exerts a strong control on the posterior part of the left STAP notably at the average location of the posterior PP when it exists.

Role of rs160459 region in *DACT1* regulation. Fig. 1B presents the chromatin state functional annotation of the genomic region upstream of DACT1 (i.e, the region significantly associated with the left STAP geodesic depth). The subregion around rs160459 has its chromatin state in brain tissues annotated as enhancer (Enh) and/or flanking active transcription start site (TssAFlnk) (25, 26), thus revealing its active chromatin state in all tested brain areas, with the notable exception of the dorso-lateral prefrontal cortex. In contrast, this region has a quiescent chromatin state in most other body tissues except stem cells and epithelial cells suggesting a tropism for neuro-ectoderm derived cell lines (**Fig. S5**). **Fig. 1C** emphasizes that rs160459 is an expression quantitative trait locus (eQTL) of *DACT1* in temporal cortex tissues ($p = 0.022$, uncorrected, 35 postmortem samples 'CC', 62 'CA' and 37 'AA') only among all other tested brain tissues (**Fig. S6**) (27).

***DACT1* gene expression across lifespan.** We performed gene-level analysis using genes prioritized by FUMA (28) from our GWA results on the left STAP geodesic depth. Among these genes (see criteria in the Methods), *DACT1* and *DAAM1* are key components in the following gene ontology (GO) biological processes obtained from the molecular signature database (29): regulation of the establishment of planar polarity, regulation of organ morphogenesis, and Wnt-signaling pathway. *DACT1* and *DAAM1* expression varies across age in cerebral tissues (**Fig. 1D**, data from Brainspan (30)). Moreover, these genes are mainly expressed during early brain development with the following distinct periods: high expression from the early to late mid-prenatal period (8 to 24 weeks post conception (pcw), no data before 8 pcw), followed by a decrease from the late prenatal period (26 pcw) until the end of the first year, and minimal expression after early childhood (from 2 years of age).

Multimodality MRI phenotypes association with rs160459. To characterize the potential role of this enhancer genomic region in shaping the human brain, we performed a phenome-wide association (PheWA) of rs160459 with various brain phenotypes computed on structural, diffusion and functional MRI exams available from the UK Biobank. We extracted (i) the brain sulci features with Morphologist (31); (ii) cortical thickness, surface area and gray matter volumes with Freesurfer on the Desikan (32) and Destrieux (33) atlases; and (iii) the subcortical volumes with Freesurfer. From the UK Biobank, we obtained (iv) tract-based spatial statistics (TBSS) (34) on white matter tracts from the ICBM atlas (35) and (v) resting-state functional MRI (rfMRI) phenotypes (see Methods).

Table 2 summarizes the 20 phenotypes with the strongest associations (smallest p-values) with rs160459 (**Table S3** for rs160458, **Table S4** with all phenotypes for rs160459). The Bonferroni significance threshold is $p < 9.7 \cdot 10^{-6}$ (0.05/5179 phenotypes, strict correction not accounting for correlations between phenotypes). A majority of phenotypes are directly related to lateral and superior parts of the temporal cortex, notably the left STS (depth, gray matter volume, surface area and thickness of its banks).
Fig. S7 shows the mapping of the p-values of the association between rs160459 and brain sulci features (maximum depth, mean depth, sulcal width, surface, gray matter thickness along the sulcus, and sulcus length). This genetic variant clearly has a localized effect on the superior temporal region curvature, with a large effect on both the left and right STS as well as on the anterior inferior temporal sulcus. Similarly, regarding cortical features obtained from Freesurfer parcellations (gray volume, surface area, and cortical thickness), the effect is localized to the temporal regions, notably the left banks of the STS in the Desikan atlas (**Fig. S8**) and the left temporal pole and middle temporal gyrus in the Destrieux atlas (**Fig. S9**).
Two main associations were uncovered with diffusion MRI analysis. The largest association was with the fractional anisotropy (FA) measured in the left tapetum of the corpus callosum, and the second corresponded to the mode of anisotropy (MO) of the left cerebellar peduncle.
Finally, we analyzed resting-state activity. rs160459 modulates several components that comprise linguistic regions (**Fig. S10**, $p < 0.01$, uncorrected), such as the superior temporal regions, insula, inferior frontal region and basal ganglia (component 18 for ICA25; and components 9 and 38 for ICA100). Associations with ICA100 amplitudes included the left fusiform and associative visual areas (components 52 and 8 respectively). Correlations between resting-state networks revealed associations with components 29-41 and 8-31 for ICA100 and 5-18 for ICA25 (**Fig. S11**), which are implicated in language processing (**Fig. S12**). In particular, **Fig. 3A** presents the top associated pair, 29-41, with rs160459. Notably, component 29 bilaterally comprises the angular and supra-marginal gyri, STAP and posterior cingulate. Component 41 includes a string of areas from Broadman area 19 and fusiform gyri to the posterior STS.

Characterizing the local influence of rs160459 on brain anatomy. The 'CC' and 'AA' allelic configurations were observed in a roughly similar number of subjects. We compared brain anatomy (DTI FA values, sulcal depth and thickness values) in 'CC' vs 'AA' subjects. A voxelwise difference in FA values in MNI space was detected between the 4,742 British subjects with the 'AA' allele and the 3,470 subjects with the 'CC' allele (**Fig. 3C**, left). These regions are projected on the cortex in (**Fig. 3C**, right). The tapetum, which comprises the auditory callosal fibers, is delineated bilaterally, in addition to a region close to the hippocampus. We also noted a difference in FA in the left STS region at the average boundary between gray and white matter, which corresponds to the average location of the posterior PP in the STAP when present. Similarly, **Fig. 3B** highlights the vertexwise difference in sulcal depth between individuals 'AA' and 'CC' (other configuration differences in **Fig. S13**). The left STAP clearly appears as the region with the most substantial difference in depth between these two groups of individuals. Two other areas display reverse effects relative to the STAP; the first area is in the left intraparietal sulcus, and the second is in the left inferior temporal gyrus along the STAP. In the posterior part of the left STAP relative to 'CC' subjects, 'AA' subjects have a shallower sulcus surrounded by more pronounced gyri whereas the white matter is more isotropic under the sulcus and anisotropic under the *planum temporale* and middle temporal gyrus. The vertexwise gray matter thickness difference between 'AA' and 'CC' subjects did not produce remarkable differences although a small cluster in the left inferior temporal gyrus along the STAP is notable (**Fig. S14**).

Discussion

A genomic effect focused on the language comprehension network. Starting from a recent evolutionary brain feature, the asymmetry of the depth of the STS, we explored its genetic associations and found that the depth of the left STS was heritable in two different population samples (HCP, UK Biobank from British ancestors) and strongly associated with a genomic region upstream of *DACT1*. This genomic region modulates several other features of the left STS (the gray matter volume, thickness and surface areas of its banks) and of close brain parcels within the left auditory/linguistic network (i.e., Jensen sulcus at the border between the supra-marginal and angular gyri, Heschl's gyrus volume and thickness of the left middle temporal lobe), and to a lesser extent, their right homologues (i.e., STS depth and thickness). Notably, although we started from an asymmetric feature, rs160459 associations with brain phenotypes were often significant in both hemispheres. However, the most important effects were always observed in the left hemisphere. As summarized in **Fig. 3C**, this SNP mainly modulated the macro and micro structure of the temporal and supra-marginal components of the linguistic network with an effect limited to the left STS for some phenotypes such as sulcal depth (**Fig. S7-S11**). This tropism for the auditory/linguistic regions was confirmed by resting-state analyses. rs160459 modulated the amplitude of components comprising the classical language areas (superior temporal and inferior frontal regions and the basal ganglia) as indicated by the correlation between two resting-state components (ICA100: 29-41, $p \sim 10^{-6}$). The ICA recovers bilateral components, making it difficult to disentangle the contribution of each hemisphere. Nonetheless, in the left hemisphere, the correlation between components 29-41 is reminiscent of the ventral and dorsal pathway of the reading network. This interpretation related to reading would also be consistent with the significant modulation of the amplitude of the left fusiform gyrus (ICA100: component 52). The relation between this genomic region and language is also supported by two patients with a 14q23.1 deletion encompassing *DACT1* who had moderate to severe expressive language delay (36, 37).

An interesting structure was also highlighted in our analyses: the left tapetum, i.e. the auditory fibers of the corpus callosum. This result may be linked to the observation that patients with corpus callosum agenesis are also the only group with no significant STS depth asymmetry in (13). The authors speculated that the dense local connectivity useful for phonetic encoding (38) intertwined with many long-range bundles (39) (the callosal auditory fibers but also the arcuate fasciculus, inferior occipito-frontal fasciculus, and middle and inferior longitudinal fasciculi), which underlie the left STS, create different tensions and tissue stiffness compared to that in the right hemisphere explaining the depth asymmetry of this sulcus. Therefore, the elevation of the bottom of the left sulcus, particularly in its posterior part, could be related to a localized higher density of fibers, which may eventually result in a PP (**Fig. 2, 3B**). Among these tracts, the arcuate fasciculus, which is the most left-lateralized tract (40), was not directly extracted in our analyses, but the voxelwise differences between allelic variants below the left STAP could correspond to it.

A role in establishing the auditory/linguistic network. rs160459 positively regulates *DACT1*, a gene mainly active, in human brain tissues, during gestation until the age of two, a crucial period for language acquisition and social communication. The protein encoded by *DACT1* belongs to the dapper family and is involved in regulating intracellular signaling pathways during development. Among its related pathways is the noncanonical Wnt/planar cell polarity pathway, which is activated in the neuronal differentiation process (41) and synaptogenesis (42). *DACT1* is reportedly involved in excitatory synapse organization, dendrite and synapse formation and establishment of spines in hippocampal pyramidal neurons (43), as well as in the migration of cortical interneurons (44). Thus, *DACT1* plays a major role in the morphology of pyramidal

cells and interneurons. Although the function of *DACT1* is quite general, our study reveals that in the superior temporal region, its expression is specifically regulated by rs160459. The STS becomes visible at approximately 25-29 wGA (16, 45) - the right STS appears one or two weeks earlier than the left - when *DACT1* expression level significantly decreases (**Fig. 1D**). At this age, the majority of neurons remain in the subplate but the main tracts (e.g., corpus callosum and fronto-occipital fasciculus) become visible, and short-range connectivity emerges, bending around the bottom of the sulci. The brain begins to process external stimuli, such as sounds and voices, along distinctive pathways, but as maturation progresses and stimulation becomes richer, dendrite development and axonal arborization complexifies the gyrification pattern, accentuating the initial hemispheric biases.

Can the allelic variation in the normal population explain something about language? There is no reason to believe that 'AA' and 'CC' subjects are more or less linguistically efficient, especially when the variants are distributed equally among the population. However, any language uses only a subset of the possible linguistic features, such as tones, stress, and complexity of the phonetic repertoire. Following the relation between allelic variants of *ASPM* and *Microcephalin* and the use of linguistic tones in certain languages proposed by Dediu and Ladd (6), Wong et al. (46) observed a significant correlation between *ASPM* variants and the correct identification of rising and falling Mandarin tones in Europeans, non-speakers of Mandarin. Here, we can speculate that rs160459 might control the column organization and early connectivity in the posterior superior temporal cortex and notably the left-right hemispheric balance through the corpus callosum. The corpus callosum has asymmetric connections early on with earlier maturation of the right to left connections, reinforcing the left hemispheric lateralization for language (47) and may contribute to the integration of features better processed by the right hemisphere (e.g., prosody, and tones) into the linguistic network. For example, tones that are processed by the right hemisphere in subjects speaking atonal languages are processed by the left in those who learned tonal languages (48). The left lateralization for language is strengthened during the first semester of life (49), and infants with greater FA changes in the splenium from 6 to 24 months produced more words at 2 years (50).

Another population that might point in the same direction is premature infants, who are at higher risk of language and sociability deficits than full-term neonates (51). Language performances at 2 years of age have been correlated with the microstructure measured at term age of the left superior temporal gyrus (52) and arcuate fasciculus (53) but also with the volume of the auditory fibers in the splenium of the corpus callosum in adolescents (54). Furthermore, compared to full-term infants, preterm infants raised in a more silent environment had both poorer language performances and a decrease in the STS depth asymmetry (55). Although no studies have examined the relationship between language, STS depth asymmetry and the volume of the temporal callosal fibers, we postulate that interaction between these structures might favor early language performances. It is also possible that allelic variants may not affect daily linguistic performances in normal subjects but may be protective. For example, preterm neonates are exposed to an environment that can be harmful due to the loss of proximity to the maternal voice and to the rhythmicity of the heart beats in the fetal environment while being exposed to unexpected and not-muffled sounds that are sometimes stressful, such as alarms. It would be interesting to examine whether the frequency of language deficits in this population is also dependent on this gene.

In conclusion, the genomic region in linkage disequilibrium with rs160459 modulates the STS depth and other related MRI phenotypes in the temporal lobe. The concordance between *DACT1* gene expression and STS formation timeframes presumably makes *DACT1* a key gene in the formation of superior temporal

networks, especially those involved in language. Based on a convergence of observations, notably in the preterm population, we speculated regarding the role of the corpus callosum in early language acquisition, but further studies should specify the role of numerous tracts crossing in this region and relating the close and distant linguistic areas.

Material and Methods

Subjects. The UK Biobank is a health research resource that aims to improve the prevention, diagnosis and treatment of a wide range of diseases. Between 2006 and 2010, half a million participants aged 40 to 73 were recruited in Great Britain. In this work, we used the data released in October 2018, including 22,392 subjects with T1-weighted MRI who were aged 44 to 81 at the time of the imaging visit. The genetic data from the UK Biobank have been subject to rigorous quality control, which was carried out at the Wellcome Trust Centre for Human Genetics (56). Here, we limited our analysis to individuals identified by the UK Biobank as belonging to the main subset of white British ancestry (using the variable *in.white.British.ancestry.subset* in the file *ukb_sqc_v2.txt*). Subjects with high genotyped missingness, high heterozygosity or sex mismatches were excluded from our analysis. Additional information on MRI data acquisition and on the number of SNPs included in the analysis is provided in **Supplementary Methods**.

STAP depth extraction. The STAP section was extracted using the automated method described in (17), Fig. 1. Briefly, we first extracted the sulcal pits (primary cortical folds) using a watershed algorithm (57). Then, we summed all the individual pits densities on a symmetric template using surface-based interhemispheric registration as proposed in (58). We performed a watershed algorithm on this group pit density, which yielded a group parcellation of the sulcal pits. We projected these group areas onto the individual meshes and identified areas denoted STS b, c, d (19). Then, we extracted the geodesic path of the vertices that follow the sulcal fundus on the white mesh between the deepest pits STS b and border STS c/d, which approximately correspond to the STAP Talaraich coordinates defined by Leroy et al. (13). From this set of vertices, we obtained the individual profiles of the geodesic depth and depth PF, which considers both the sulcal depth and convexity. We averaged these values along the STAP in each subject to obtain our depth phenotypes. We also applied the PP detection method described in (17) and benchmarked on manually labelled data from (13) to identify subjects with at least one sulcal interruption in their depth profile. This process leads to the binary phenotype "presence or absence of PP". Of our British discovery sample that completed the pipeline (16,515 subjects), 9,501 subjects had a left PP (57.53%), and 2,587 had a right PP (15.66%).

Phenome-wide analysis. We processed the T1w and T2w images with Freesurfer v6 pipeline (59), and extracted individual measures from the following output files: ?h.aparc.stats (Desikan parcellation (32)), ?h.aparc.a2009s.stats (Destrieux parcellation (33)), and aseg.stats (subcortical volumes). We processed the T1w images with Brainvisa Morphologist pipeline (31) and obtained individual sulcus measures per subject. We obtained diffusion MRI and resting-state MRI phenotypes from the UK Biobank (described in 16, 53). Briefly, diffusion MRI measures were extracted using TBSS (34) from the ICBM 48 white matter tracts atlas (35). For resting-state fMRI (rfMRI), independent component analysis (ICA) was performed by the UK Biobank on the first 4,181 participants (21). A total of 21 and 55 good components were identified for 25 and 100 component analyses, respectively (60). These components were subsequently used for rfMRI acquired in the following subjects. The rfMRI phenotypes include the amplitude of the ICA weighted rfMRI

BOLD signal, referred to as ICA amplitude, the Pearson correlation and partial correlations between pairs of ICA time series, which are also referred to as functional connectivity.

Quantitative genetics. In our discovery sample, we used genome complex trait analysis (GCTA), (23) which provides an estimate of heritability in population studies with unrelated genotyped participants. We considered the genotyped SNP variants common to the UKBiobank and UKBileve arrays. To compute the kinship matrix of the population, we selected specific SNPs with PLINK v1.9 (61) using the following thresholds: missing genotype = 5% (70,737 variants excluded), Hardy-Weinberg equilibrium test (hwe) = 10^{-6} (18,462), and minor allele frequency (maf) = 1.0% (101,839). We kept the SNPs in moderate linkage disequilibrium with variation inflation factor 10 within a window of 50 SNPs (91,496 variants excluded). Then, we computed the genetic relationship matrix with GCTA using the 501,722 SNPs left. The amount of phenotypic variance captured by this matrix is estimated using a mixed-effects linear model. As covariates, we included gender, genotyping array type, age at the MRI session and the 10 genetic principal components (PCs) provided by the UK Biobank to account for population stratification.

The heritability estimates for the Human Connectome Project were published in our previous study (17). These estimates are computed using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) (22) on an extended pedigree of 820 Caucasian individuals, which includes twins and siblings.

Genome-wide associations. The analyses of genotype-phenotype association were performed with PLINKv1.9 (61) using the imputed genotype data. The same thresholds as above were applied (missing genotype: 5%, hwe: 10^{-6}, and maf: 0.01). We included the same covariates as in the quantitative genetic section.

Functional annotation. The functional annotations of loci were provided by FUMA (28), which notably annotates SNP location, expression quantitative trait effects (27) and chromatin state (26) in various brain tissues. FUMA prioritized genes from summary p-values of the GWAs based on base pair distance and linkage disequilibrium, as well as probability of being loss-of-function intolerant (pLI) scores and non-coding residual variation intolerance score (ncRVIS) from the ExAC database. FUMA also examines the Brainspan atlas (30) and Molecular signature database (29) to annotate prioritized genes from GWAs.

ACKNOWLEDGMENTS. The present analyses were conducted under UK Biobank data application number 25251. The acknowledgments of IMAGEN contributing consortium authors can be found in the Supplementary Notes.

References:

1. Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. *Nat Genet* 18(2):168–170.
2. Francks C, et al. (2004) A 77-Kilobase Region of Chromosome 6p22.2 Is Associated with Dyslexia in Families From the United Kingdom and From the United States. *Am J Hum Genet* 75(6):1046–1058.
3. Liégeois F, et al. (2003) Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat Neurosci* 6(11):1230–1237.
4. Pinel P, et al. (2012) Genetic Variants of FOXP2 and KIAA0319/TTRAP/THEM2 Locus Are Associated with Altered Brain Activation in Distinct Language-Related Regions. *J Neurosci* 32(3):817–825.
5. Cai DC, et al. (2014) A genome-wide search for quantitative trait loci affecting the cortical surface area and thickness of Heschl's gyrus. *Genes, Brain Behav* 13(7):675–685.
6. Dediu D, Ladd DR (2007) Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc Natl Acad Sci U S A* 104(26):10944–9.
7. Mazoyer BM, et al. (1993) The cortical representation of speech. *J Cogn Neurosci* 5(4):467–479.
8. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403(6767):309–312.
9. Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L (2002) Functional neuroimaging of speech perception in infants. *Science* 298(5600):2013–5.
10. Bristow D, et al. (2009) Hearing faces: How the infant brain matches the face it sees with the speech it hears. *J Cogn Neurosci* 21(5):905–921.
11. Mahmoudzadeh M, et al. (2013) Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proc Natl Acad Sci* 110(12):4846–4851.
12. Geschwind N, Levitsky W (1968) Human Brain: Left-Right Asymmetries in Temporal Speech Region. *Science (80-)* 161(3837):186–187.
13. Leroy F, et al. (2015) New human-specific brain landmark: The depth asymmetry of superior temporal sulcus. *Proc Natl Acad Sci* 112(4):1208–1213.
14. Hopkins WD, Pilcher DL, MacGregor L (2000) Sylvian fissure asymmetries in nonhuman primates revisited: A comparative MRI study. *Brain Behav Evol* 56(6):293–299.
15. Dubois J, et al. (2008) Mapping the early cortical folding process in the preterm newborn brain. *Cereb Cortex* 18(6):1444–1454.
16. Habas PA, et al. (2012) Early folding patterns and asymmetries of the normal human brain detected from in utero MRI. *Cereb Cortex* 22(1):13–25.
17. Le Guen Y, et al. (2018) The chaotic morphology of the left superior temporal sulcus is genetically constrained. *Neuroimage* 174:297–307.
18. Cantalupo C, Hopkins WD (2001) Asymmetric Broca's area in great apes. *Nature* 414(6863):505–505.
19. Le Guen Y, et al. (2018) Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds. *Cereb Cortex* 28(6):1922–1933.
20. Meng Y, Li G, Lin W, Gilmore JH, Shen D (2014) Spatial distribution and longitudinal development of deep cortical sulcal landmarks in infants. *Neuroimage* 100:206–218.
21. Miller KL, et al. (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 19(11):1523–1536.
22. Almasy L, Blangero J (1998) Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. *Am J Hum Genet* 62(5):1198–211.
23. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet* 88(1):76–82.
24. Zerbino DR, et al. (2018) Ensembl 2018. *Nucleic Acids Res* 46(D1):D754–D761.

25. Roadmap Epigenomics Consortium, et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–329.
26. Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9(3):215–216.
27. Ramasamy A, et al. (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* 17(10):1418–1428.
28. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8(1):1–10.
29. Liberzon A, et al. (2015) The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* 1(6):417–425.
30. Miller JA, et al. (2014) Transcriptional landscape of the prenatal human brain. *Nature* 508(7495):199–206.
31. Perrot M, Rivière D, Mangin J-F (2011) Cortical sulci recognition and spatial normalization. *Med Image Anal* 15(4):529–50.
32. Desikan RS, et al. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31(3):968–980.
33. Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53(1):1–15.
34. Smith SM, et al. (2006) Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31(4):1487–1505.
35. Mori S, et al. (2008) Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage* 40(2):570–582.
36. Jiang YH, et al. (2008) De novo and complex imbalanced chromosomal rearrangements revealed by array CGH in a patient with an abnormal phenotype and apparently "balanced" paracentric inversion of 14(q21q23). *Am J Med Genet Part A* 146(15):1986–1993.
37. Martínez-Frías ML, et al. (2014) Interstitial deletion 14q22.3-q23.2: Genotype-phenotype correlation. *Am J Med Genet Part A* 164(3):639–647.
38. DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci* 109(8):E505–E514.
39. Turken AU, Dronkers NF (2011) The Neural Architecture of the Language Comprehension Network: Converging Evidence from Lesion and Connectivity Analyses. *Front Syst Neurosci* 5(February):1–20.
40. Dubois J, et al. (2010) Structural asymmetries of perisylvian regions in the preterm newborn. *Neuroimage* 52(1):32–42.
41. Jiao S, Liu Y, Yao Y, Teng J (2018) miR-124 promotes proliferation and neural differentiation of neural stem cells through targeting DACT1 and activating Wnt/β-catenin pathways. *Mol Cell Biochem* 0(0):0.
42. Rosso SB, Inestrosa NC, Rosso SB (2013) WNT signaling in neuronal maturation and synaptogenesis. *Front Cell Neurosci* 7:103.
43. Okerlund ND, et al. (2010) Dact1 Is a Postsynaptic Protein Required for Dendrite, Spine, and Excitatory Synapse Development in the Mouse Forebrain. *J Neurosci* 30(12):4362–4368.
44. Arguello A, et al. (2013) Dapper Antagonist of Catenin-1 Cooperates with Dishevelled-1 during Postsynaptic Development in Mouse Forebrain GABAergic Interneurons. *PLoS One* 8(6):e67679.
45. Chi JG, Dooling EC, Gilles FH (1977) Gyral development of the human brain. *Ann Neurol* 1(1):86–93.
46. Wong PCM, Chandrasekaran B, Zheng J (2012) The Derived Allele of ASPM Is Associated with Lexical Tone Perception. *PLoS One* 7(4):e34243.
47. Adibpour P, Dubois J, Moutard M-L, Dehaene-Lambertz G (2018) Early asymmetric inter-hemispheric transfer in the auditory network: insights from infants with corpus callosum agenesis. *Brain Struct Funct* 223(6):2893–2905.
48. Pierce LJ, Klein D, Chen J-K, Delcenserie A, Genesee F (2014) Mapping the unconscious

maintenance of a lost first language. *Proc Natl Acad Sci U S A* 111(48):17314–9.
49. Shultz S, Vouloumanos A, Bennett RH, Pelphrey K (2014) Neural specialization for speech in the first months of life. *Dev Sci* 17(5):766–774.
50. Swanson MR, et al. (2017) Splenium development and early spoken language in human infants. *Dev Sci* 20(2):e12360.
51. Charkaluk M, Truffert P, Fily A, Ancel P, Pierrat V (2010) Neurodevelopment of children born very preterm and free of severe disabilities: the Nord-Pas de Calais Epipage cohort study. *Acta Paediatr* 99(5):684–689.
52. Aeby A, et al. (2013) Language development at 2 years is correlated to brain microstructure in the left superior temporal gyrus at term equivalent age: A diffusion tensor imaging study. *Neuroimage* 78:145–151.
53. Salvan P, et al. (2017) Language ability in preterm children is associated with arcuate fasciculi microstructure at term. *Hum Brain Mapp* 38(8):3836–3847.
54. Northam GB, et al. (2012) Interhemispheric temporal lobe connectivity predicts language impairment in adolescents born preterm. *Brain* 135(12):3781–3798.
55. Pineda RG, et al. (2014) Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments. *J Pediatr* 164(1):52–60.e2.
56. Bycroft C, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562(7726):203–209.
57. Auzias G, Brun L, Deruelle C, Coulon O (2015) Deep sulcal landmarks: Algorithmic and conceptual improvements in the definition and extraction of sulcal pits. *Neuroimage* 111:12–25.
58. Greve DN, et al. (2013) A surface-based analysis of language lateralization and cortical asymmetry. *J Cogn Neurosci* 25(9):1477–92.
59. Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781.
60. Elliott LT, et al. (2018) Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562(7726):210–216.
61. Purcell S, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81(3):559–575.

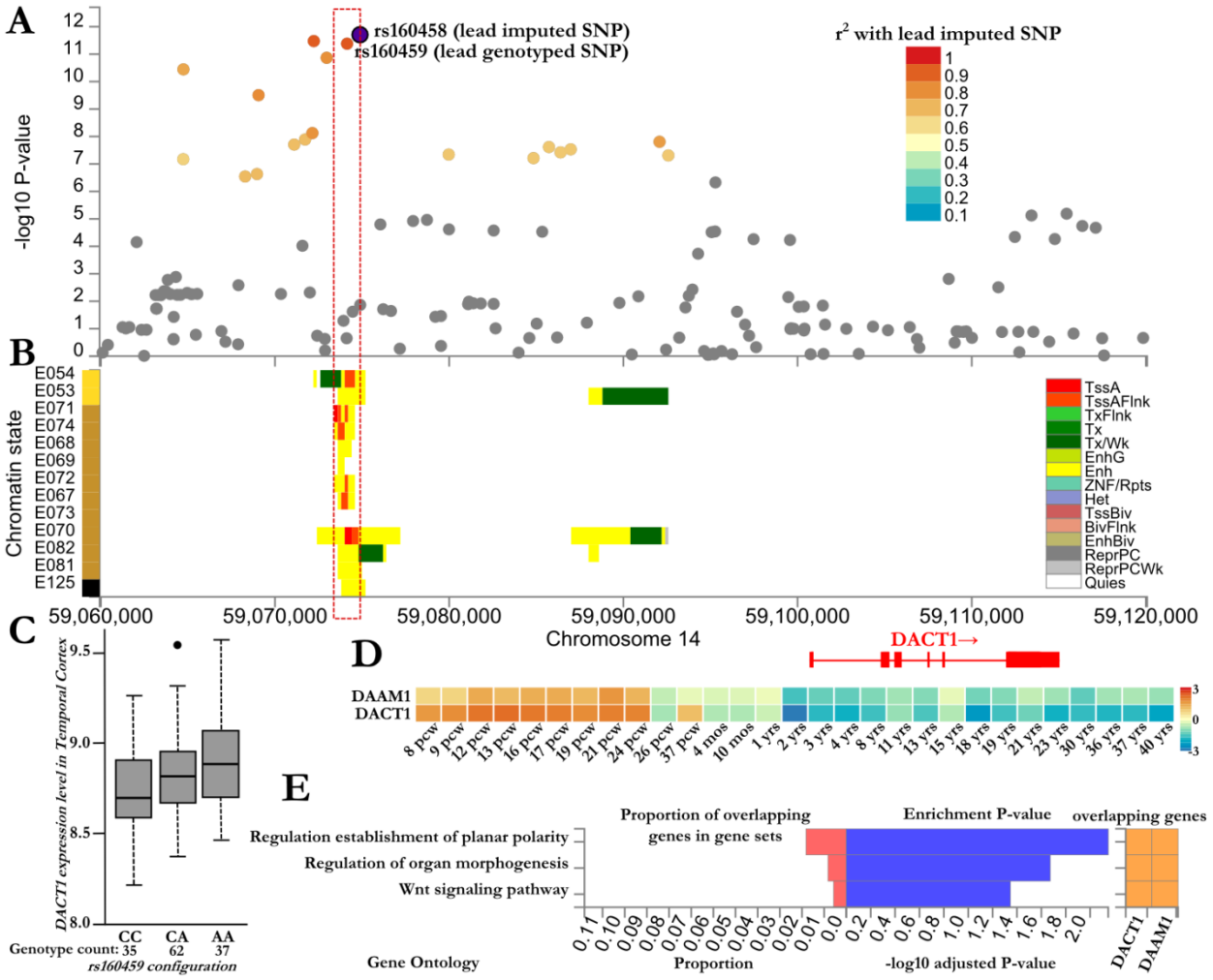

Fig. 1 (A) Manhattan plot of the geodesic depth in the left STAP zoomed in on the significant associations near the *DACT1* gene (14q23.1). (B) Chromatin state of the genomic region near lead SNPs rs160458 (imputed) and rs160459 (genotyped). *DACT1* position and exons are highlighted in red on the axis showing the base pair position (GRCh37) on chromosome 14. Yellow correspond to enhancer (Enh) state, red-orange to flanking active transcription start site (TssAFlnk), green to weak transcription (Tx/Wk) and white to quiescent (Quies) (see ref. (25, 26) for the complete legend). **(C) *DACT1* expression level in the temporal cortex stratified by rs16049 configuration from Braineac** (27)**. (D) Average normalized gene expression (zero mean across samples) of *DACT1* and *DAAM1* at 29 developmental ages from Brainspan** (30)**. (E) Gene ontologies with significant enrichment in the gene sets prioritized by FUMA** (28) **in our GWAs.** Brain tissue name abbreviations in **(B)**: E054: Ganglion Eminence-derived primary cultured neurospheres, E053: Cortex-derived primary cultured neurospheres, E071: Brain Hippocampus Middle, E074: Brain Substantia Nigra, E068: Brain Anterior Caudate, E069: Brain Cingulate Gyrus, E072: Brain Inferior Temporal Lobe, E067: Brain Angular Gyrus, E073: Brain Dorsolateral Prefrontal Cortex, E070: Brain Germinal Matrix, E082: Fetal Brain Female, E081: Fetal Brain Male, E125: Astrocyte Primary Cells.

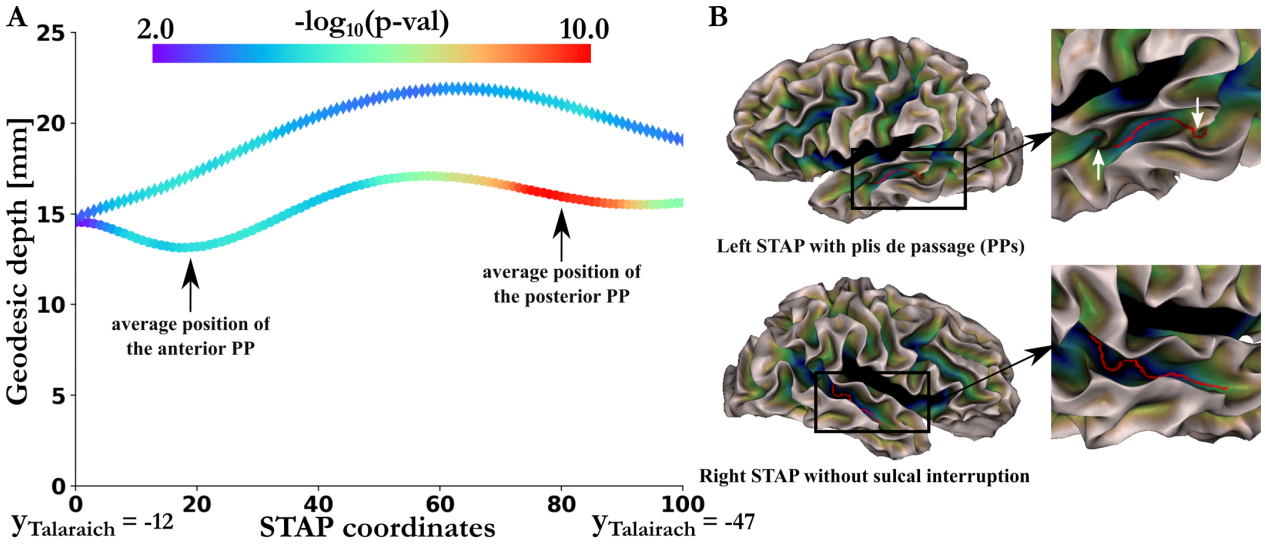

Fig. 2 (A) Mapping of the p-values along the STAP profile averaged over 16,515 British subjects of the association between rs160459 and geodesic depth. Left values are represented by disks and right values by diamonds (more details **Fig. S15**). **(B) Example of asymmetries in an individual subject with anterior and posterior PPs in the left STAP and no PP in the right STAP.** Note that most subjects have only one PP either anterior or posterior in the left STAP, and no sulcal interruption in the right STAP (17).

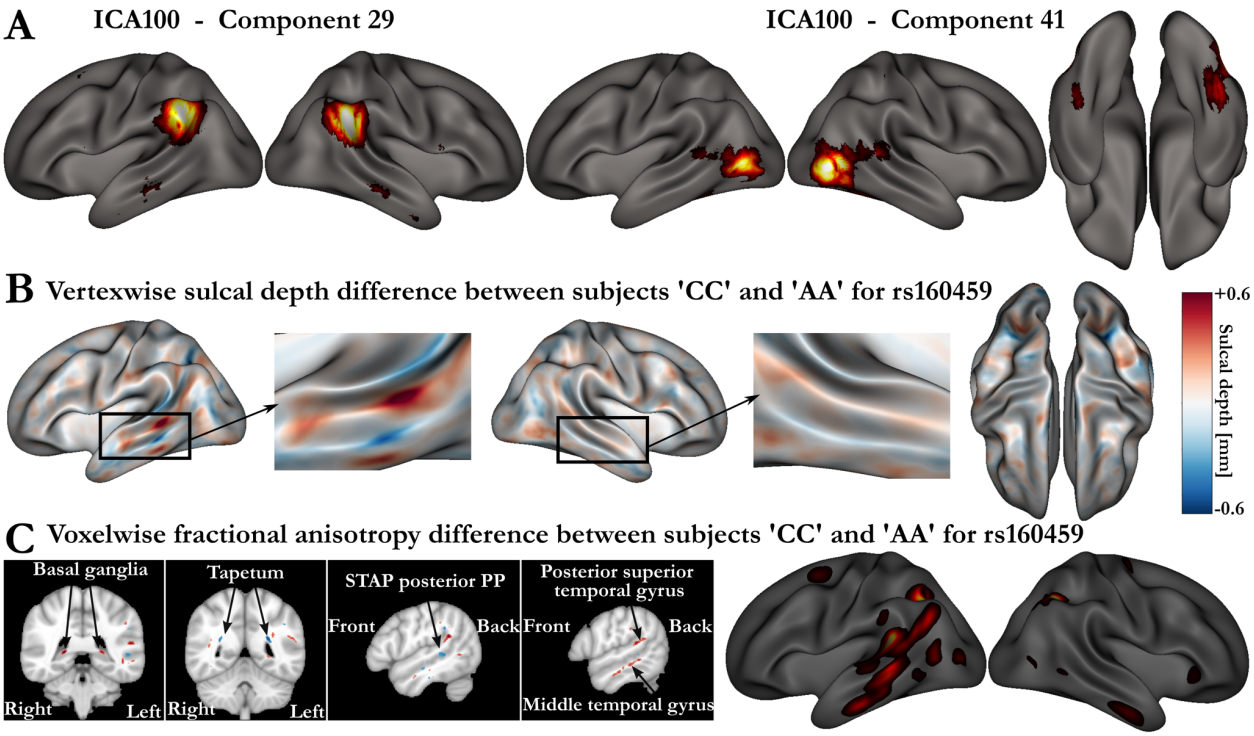

Fig. 3 (A) ICA100 pair (29-41) for which their full correlation has the highest association p-value with rs160459. (B) Vertexwise sulcal depth difference between subjects 'CC' and 'AA' for rs160459. Red values show higher depth values for CC subjects and blue for AA. **(C) Voxelwise fractional anisotropy difference in the MNI space between subjects 'CC' and 'AA' for rs160459**. These figures are thresholded at 0.005 [unitless fractional measure] to underline the main differences. In the right column, these differences in white matter are projected onto the cortex.

Table 1. P-values of rs160459 association with the STAP phenotypes in three populations. * significant heritability estimates ($p < 6.3 \cdot 10^{-3}$, details in Table S1.). AI: asymmetry index, PF: potential function, PP: *pli de passage*.

Phenotype/Cohort	Heritability		Association		
	HCP	UKB	UKB discovery N=16,515 British	IMAGEN replication N=1,726	UKB replication N=2,161 Non-British
Left STAP geodesic depth	48%*	21%*	**$p = 4.2 \cdot 10^{-12}$**	$p = 8.0 \cdot 10^{-5}$	p = 0.0133
Right STAP geodesic depth	46%*	15%*	$p = 1.3 \cdot 10^{-7}$	$p = 2.2 \cdot 10^{-4}$	p = 0.1563
AI of the STAP geodesic depth	30%*	9%*	p = 0.0017	p = 0.0406	p = 0.1675
Left STAP depth PF	25%*	18%*	$p = 3.8 \cdot 10^{-8}$	p = 0.0100	p = 0.1004
Right STAP depth PF	38%*	4%	$p = 4.1 \cdot 10^{-6}$	p = 0.0971	p = 0.0249
AI of the STAP depth PF	22%*	11%*	p = 0.0053	p = 0.0909	p = 0.5414
PP in the left STAP	53%*	16%*	$p = 2.7 \cdot 10^{-7}$	p = 0.0022	p = 0.2447
PP in the right STAP	27%	0%	$p = 2.6 \cdot 10^{-7}$	p = 0.0034	p = 0.2067

Table 2. Top 20 MRI phenotypes associated with rs160459 (significant after Bonferroni correction $p < 9 \cdot 10^{-6}$, 0.05/5179). (GM: gray matter, DKT: Desikan-Killiany atlas, DST: Destrieux atlas, rfMRI: resting-state functional MRI). Color code: orange: Brainvisa phenotypes, green: Freesurfer phenotypes, blue: diffusion MRI phenotypes, red: rfMRI phenotypes.

Phenotypes	Pvals	N subjects (British)
Mean depth of the left STS	$1.4 \cdot 10^{-13}$	18101
Surface area of the left pole temporal - DST	$5.8 \cdot 10^{-12}$	17127
GM volume of the left banks STS - DKT	$7.6 \cdot 10^{-8}$	17127
Surface area of the right cuneus gyrus - DST	$2.3 \cdot 10^{-7}$	17127
Surface area of the left banks STS - DKT	$2.8 \cdot 10^{-7}$	17127
Fractional anisotropy of the left tapetum of corpus callosum	$5.7 \cdot 10^{-7}$	16541
Surface area of the right cuneus - DKT	$6.6 \cdot 10^{-7}$	17127
GM thickness of the right middle temporal gyrus - DKT	$6.9 \cdot 10^{-7}$	17127
GM thickness of the left banks STS - DKT	$8.8 \cdot 10^{-7}$	17127
GM thickness of the left middle temporal - DKT	$1.1 \cdot 10^{-6}$	17127
Maximum depth of the left anterior inferior temporal sulcus	$1.5 \cdot 10^{-6}$	18101
GM thickness of the left middle temporal gyrus - DST	$1.9 \cdot 10^{-6}$	17127
rfMRI correlation between ICA100: 29 and 41 (edge 1146)	$2.1 \cdot 10^{-6}$	15859
Mean depth of the right STS	$2.3 \cdot 10^{-6}$	18100
Surface area of the right temporal pole - DST	$2.4 \cdot 10^{-6}$	17126
GM thickness of the right banks STS - DKT	$6.0 \cdot 10^{-6}$	17127
rfMRI correlation between ICA25: 5 and 18 (edge 87)	$7.6 \cdot 10^{-6}$	15864
GM volume of the left Jensen sulcus - DST	$7.6 \cdot 10^{-6}$	17108
GM volume of the left transverse temporal - DKT	$7.9 \cdot 10^{-6}$	17126
Surface area of the left anterior inferior temporal sulcus	$8.6 \cdot 10^{-6}$	18101