Genome sequencing for rightward hemispheric language dominance

Running title: Genetics of language dominance

Authors

Amaia Carrion-Castillo[1], Lise Van der Haegen[2], Nathalie Tzourio-Mazoyer[3], Tulya Kavaklioglu[1], Solveig Badillo[3,4], Marie Chavent[4], Jérôme Saracco[4], Marc Brysbaert[2], Simon E Fisher[1,5], Bernard Mazoyer[3] and Clyde Francks[1,5]

[1] Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[2] Department of Experimental Psychology, Ghent Institute for Functional and Metabolic Imaging, Ghent University, Belgium

[3] Groupe d'Imagerie Neurofonctionnelle, Institut des Maladies Neurodégénératives, Centre National de la Recherche Scientifique, Commissariat à l'Energie Atomique, et Université de Bordeaux, Bordeaux, France

[4] Institut de Mathématiques de Bordeaux, Centre National de la Recherche Scientifique, Institut National de la Recherche en Informatique et Automatique, et Université de Bordeaux, Bordeaux, France

[5] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Corresponding author

Clyde Francks

Max Planck Institute for Psycholinguistics

PO BOX 310, 6500 AH Nijmegen

The Netherlands

Phone: +31-24-3521929

Fax: +31-24-3521213

E-mail: clyde.francks@mpi.nl

Date of submission: 18.03.2019

Abstract: 209/250 words

Introduction: 880 words

Discussion: 1430 words

Abstract

Most people have left-hemisphere dominance for various aspects of language processing, but only roughly 1% of the adult population has atypically reversed, rightward hemispheric language dominance (RHLD). The genetic-developmental program that underlies leftward language laterality is unknown, as are the causes of atypical variation. We performed an exploratory whole-genome-sequencing study, with the hypothesis that strongly penetrant, rare genetic mutations might sometimes be involved in RHLD. This was by analogy with *situs inversus* of the visceral organs (left-right mirror reversal of the heart, lungs etc.), which is sometimes due to monogenic mutations. The genomes of 33 subjects with RHLD were sequenced, and analysed with reference to large population-genetic datasets, as well as thirty-four subjects (14 left-handed) with typical language laterality. The sample was powered to detect rare, highly penetrant, monogenic effects if they would be present in at least 10 of the 33 RHLD cases and no controls, but no individual genes had mutations in more than 5 RHLD cases while being un-mutated in controls. A hypothesis derived from invertebrate mechanisms of left-right axis formation led to the detection of an increased mutation load, in RHLD subjects, within genes involved with the actin cytoskeleton. The latter finding offers a first, tentative insight into molecular genetic influences on hemispheric language dominance.

Introduction

Non-invasive imaging methods such as functional magnetic resonance imaging (fMRI) have shown that roughly 85% of people have left-hemisphere language dominance, while most remaining people are ambilateral for language, and only a small minority of around 1% show rightward hemisphere language dominance[1-3]. The degree of laterality assessed with fMRI varies with the type of language task used, and is usually more pronounced for language production than perception tasks[4]. Roughly 90% of people are right-handed, 10% left-handed, and a small remainder ambidextrous[5]. Although more than 70% of left-handers have left-hemisphere language dominance[3], over 90% of people with RHLD are also left-handed[3]. Therefore RHLD usually involves a broader re-organization of left-right laterality than purely for language functions, but may represent an etiological group that is distinct from the bulk of left-handers.

Gene expression and *in utero* ultrasound studies of human embryos have indicated that lateralized development is already underway in the human central nervous system by five to eight weeks post-conception[6-8], which indicates a genetic-developmental program underlying the typical form of functional brain laterality. One study reported a non-significant heritability (<1%) for the laterality of speech sound perception, based on the dichotic listening method, and considering the full range of trait variation from left- to right-ear-advantage[9]. However, atypical functional language dominance, i.e. a categorical trait defined to include both rightward hemisphere language dominance (RHLD) and ambilateral dominance, has been shown to have a heritability of roughly 30%, measured with functional transcranial Doppler sonography during language production[10-12]. There have been no twin or family-based studies of RHLD heritability itself, likely due to the rarity of the trait. Twin and family studies have reported moderate heritability estimates for left-handedness (24-39%)[11,13], although heritability estimates based on genomic similarity between unrelated people in the general population are much lower for left-handedness (heritability=1-3%)[14,15].

Regardless, molecular mechanisms for the initial 'symmetry breaking' process in the mammalian brain, i.e. for establishing a left-right axis in the very early embryo, remain unknown[16]. In contrast, much is known about the developmental origins of asymmetry of the visceral organs (i.e. heart, lungs etc.). Increased activation of the nodal signalling cascade on the left side of an early embryonic structure, called the node, ultimately results in asymmetric organogenesis[17]. Motile cilia within the node are important for this process, because their unidirectional rotation, arising from the chirality of their protein constituents, produces a right-to-left fluid flow that triggers left-sided nodal expression[17,18]. Monogenic mutations in genes that encode components of motile cilia, or otherwise affect ciliary functions, can cause the disorder primary ciliary dyskinesia (PCD) together with *situs inversus totalis* (SIT), a condition affecting roughly 1/6000 to 1/8000 people, in which the visceral organs are placed as the mirror image of the usual arrangement[18,19]. PCD with SIT is a genetically heterogeneous condition, which can be caused by mutations in at least 37 different genes[20], although one gene accounts for 15-28% of cases (*DNAH5*)[21,22].

Intriguingly, people with PCD and SIT do not show an increased rate of RHLD or left-handedness, which suggests a fundamental dissociation between nodal-ciliary mechanisms of visceral axis formation and the brain functional lateralities for language and hand dominance[23-25]. Thus, the typical form of human brain functional laterality may instead originate from a genetic-developmental mechanism that is brain-intrinsic. Recent studies in *Drosophila* have revealed that cellular chirality induces left-right asymmetry of individual organs in an organ-intrinsic manner, without being induced by the ciliary-nodal pathway[26-29]. In these mechanisms, chirality is a transient property of whole cell morphology at key points in embryonic development[26]. A role of actin-related genes in establishing cellular chirality has been observed in both invertebrate (*Drosophila,* snail)[26-29] and vertebrate models (cultured cells, frog,

zebrafish) [26,30,31], suggesting that this mechanism is important to establish left-right organ asymmetry across bilaterian groups. Apart from the cilia-related nodal signalling pathway, cellular chirality is the only biological mechanism that has been shown to give rise to organ asymmetry in multicellular animals, of which we are aware.

Recent analyses using the UK biobank dataset, based on more than 300,000 participants, have reported that alleles of the microtubule-associated gene *MAP2* have very small effects on the probability of becoming left-handed, as well as some other loci which did not clearly implicate individual genes[32,33]. However, the rarer trait of RHLD, found in only roughly 10% of left-handers and less than 1% of right-handers, has not been subject to any previous molecular genetic studies. By analogy with SIT, here we investigated whether RHLD might sometimes arise due to high-penetrance genetic mutations. We sequenced the genomes of 33 people with RHLD as assessed using fMRI, as well as 34 typically lateralized subjects (20 right-handed, 14 left-handed), and interrogated the data with reference to large population genetic databases (Figure 1).

As this was an exploratory study, we performed separate analyses under recessive and dominant models, allowing for allelic heterogeneity (different causative mutations within a given gene) or genetic heterogeneity (causative mutations in different genes). We also tested for an increased rate of rare mutations in RHLD within specific candidate gene sets, in case an increased load of mutations affecting specific biological processes might increase the chance of having RHLD. The candidate sets included genes involved in visceral laterality or the actin cytoskeleton, as well as a set of 18 genes which have been tentatively associated with human brain laterality in previous studies[16,32,34].

Methods

Datasets and functional laterality measurement

A total of 67 participants (33 with RHLD) were included in the present study, all of whom gave written informed consent. All RHLD subjects except one were left-handed (Edinburgh Handedness Inventory (EHI) median = -87.50), while the controls included 14 left-handed and 20 right-handed participants (EHI median=76.39; this composition allowed us to perform post hoc analysis using control groups of different handedness, see below). Summary statistics for language laterality measures and handedness are provided in Table 1, Figure 2 and Figure S2.

The subjects in this study were recruited from two separate sources, i.e. the BIL&GIN dataset (France) and the GOAL dataset (Belgium).

BIL&GIN

17 RHLD subjects and 22 controls were drawn from a larger dataset of healthy, young adults, balanced for handedness (N=297, of which 153 left-handers)[3]. Informed consent was obtained from all participants and the study was approved by the Basse-Normandie local ethics committee (reference: CPP-2006-16).

We studied hemispheric lateralization for three language tasks, namely production, reading, and listening, using fMRI to calculate Global Hemispheric Functional Laterality Indexes (HFLIs), as described previously[35]. Each participant underwent a slow event-related functional MRI protocol including three runs, one for each language task, presented in a random order. The three runs followed the same structure, alternating execution of the task at the sentence level and at the word list level. Word lists used in the tasks consisted of ordered lists of the months of the year or days of the week. Functional magnetic resonance imaging was performed on a Philips Achieva 3Tesla MRI scanner. For each run,

functional volumes were acquired with a T_2^*-weighted echo planar imaging acquisition (192 volumes; TR = 2s; TE = 35ms; flip angle = 80°; 31 axial slices; 3.75mm^3 isotropic voxel size).

fMRI data analysis was performed using the SPM5 software (www.fil.ion.ucl.ac.uk/spm/). Scans of each participant and each run were normalized to our site-specific template, corrected for motion during the run, and then warped into the standard MNI space using a tri-linear interpolation, with subsequent smoothing using a 6-mm FWHM Gaussian filtering. We then computed for each participant the BOLD signal difference maps and associated t-maps corresponding to the "sentence versus word-list" contrast for the production, reading and listening runs. For each individual and each language task, we computed a Hemispheric Functional Lateralization Index (HFLI) using the LI-toolbox applied to the individual contrast t-map of the considered language task[36].

A two-step procedure was then implemented to select RHLD subjects and typically lateralized controls. We first selected the ten individuals previously identified as strongly right-lateralized in this dataset using a stringent criterion based only on language production (HFLI for language production < -50)[3]. Then, in order to identify individuals exhibiting a right-lateralized profile in all three language conditions, but who may have been overlooked in the first-step, we modelled the joint distribution of the 3 HFLI using a mixture of 3D Gaussian functions and applied a robust consensus clustering approach[37]. This second step uncovered 14 individuals having HFLI values < -15 for each of the three language conditions, including 7 of those already selected in the first step. In total, 17 subjects were thus identified as having RHLD on the basis of their HFLIs for production, reading and listening. These 17, plus another 22 control subjects with typical left-hemisphere language dominance, comprised the 39 BIL&GIN participants of the present study. HFLI distributions for RHLD and controls subjects are shown in Figure 2. The median age of RHLD subjects was 23 years, range 19-38 years, and for controls the median age was also 23 years, range 19-38 years. Information on sex is given in Table 1. We deliberately over-represented left-handedness in our selection of control subjects (14 left-handed out of 22) in order to carry out post hoc analysis with respect to handedness (see Results). Handedness was assessed based on the Edinburgh inventory[38].

GOAL

16 RHLD participants were selected from a larger dataset of healthy left-handers (N=250) [39] that was first evaluated using the behavioural visual half field (VHF) task to identify likely RHLD subjects, and then confirmed using fMRI to calculate Global Hemispheric Functional Laterality Indexes (HFLIs) based on a language production task[2]. Participants were asked to covertly think of as many words as possible beginning with a letter presented in the middle of the screen for 15 seconds. Ten different letters were presented in randomized order. The baseline condition consisted of ten 15-second blocks with silent repetition of the non-word baba. Experimental and baseline blocks were alternated with 20 rest periods of again 15 seconds, during which a horizontal line was displayed at the screen centre. Images were acquired on a 3-Tesla Siemens Trio MRI scanner (Siemens Medical Systems, Erlangen, Germany) with an 8-channel radiofrequency head coil. First, a high resolution anatomical image was collected using a T1-weighted 3D MPRAGE sequence (TR = 1550 ms, TE = 2.39 ms, image matrix = 256 ×256, FOV = 220 mm, flip angle = 9◦, voxel size = 0.9 mm × 0.9 mm × 0.9 mm). Functional images were then obtained using a T_2^*-weighted gradient-echo EPI sequence. Forty axial slices covering the whole brain were acquired (TR = 2630 ms; TE = 35 ms; flip angle = 80°; image matrix = 64 × 64, FOV = 224 mm, slice thickness = 3.0 cm, distance factor = 17%, and voxel size = 3.5 mm ×3.5 mm ×3 mm).

The sixteen strongly right-lateralized individuals all met a stringent criterion for RHLD (HFLI for language production < -50)[2]. Twelve controls were collected separately but their language lateralization was assessed using the same fMRI paradigm. The twelve controls each had a strongly leftward HFLI score

(>50). HFLI distributions for RHLD and controls subjects are shown in Figure 2, and information on sex is given in Table 1. The median age of RHLD subjects was 24.5 years, range 20-29 years, and for controls the median age was 19 years, range 18-24 years. All control subjects were right-handed in the GOAL dataset as assessed by the Edinburgh inventory[38].

Informed consent was obtained from all participants and ethical approval for the study was obtained from the Ethics Committee of the Ghent University Hospital.

Whole genome sequencing, pre-processing and variant calling

BIL&GIN

Whole genome sequencing of the 39 BIL&GIN subjects was performed using Illumina's HiSeq technology by the genomics research organization and service company BGI (HongKong/Shenzhen) (https://emea.illumina.com/systems.html). Thirteen additional subjects of European descent, who were not part of the present study, were also sequenced at the same time, and their data processed together with the 39 through pre-processing and variant calling stages (as some of the processing steps below benefit from being run on the greatest sample size available; a minimum of 30 is recommended[40]).

Sequencing was done at 20 times average coverage depth, with 90 base pair (bp) paired-end reads for 11 of the RHLD subjects and 14 controls, and 150 bp paired-end reads for 6 RHLD subjects and 8 controls. Raw reads were cleaned by excluding adapter sequences, reads with low-quality bases for more than 50% of their lengths, and reads with unknown bases for more than 10% of their lengths. Clean reads were mapped onto the human reference genome (hg19) using the software Burrows-Wheeler Aligner[41]. Bam files were sorted using SAMtools v1.2 [42] and PCR duplicate reads were marked using Picard v1.134. Re-alignment around indels (insertion/deletions) and base quality control recalibration was performed using the Genome analysis toolkit software (GATK v3.5)[43,44]. Genetic variants were called using the HaplotypeCaller (HC) tool of GATK (v3.5). HC was run separately per sample using the '-ERC GVCF' mode, and then merged together using the GenotypeGVCFs tool, as recommended in the GATK best practices. We performed Variant Quality Score Recalibration (VQSR) to exclude low quality variants (phred-scaled Qscore < 30) and to flag the rest into the sensitivity tier they fell into (90, 99, 99.9 and 100).

These variants were then normalized and variants belonging to any VQSR sensitivity tier over 99 % were excluded. For the 39 BIL&GIN subjects of this study, the variant calling of SNPs and indels identified on average 4,165,806 variants per subject for the 90bp protocol (range: 4,079,049-4,330,101), and 4,484,638 per subject for the 150bp protocol (range: 4,354,345-4,657,333).

GOAL

The genomics company Novogene (Hong Kong/Shenzhen) performed WGS on the 28 samples of the GOAL dataset using Illumina's HiSeq Xten technology, and paired-end sequencing with reads of 150 base pairs and 30x sequence depth. The same pipeline as that applied to the BIL&GIN data was used for alignment (build 37), variant calling, annotation and filtering (but updated to SAMtoolsv1.3.1, Picard v2.0.1, GATK v4.0.1.1 and Gemini v20.0.1, as sequencing of the GOAL subjects was done later). The variant calling and VQSR steps were done together with data from 34 European-descent subjects who were not part of the present study, again because these steps benefit from a larger number of subjects. These variants were then normalized using the software tool vt normalize (v0.5772-60f436c3) [45]and variants belonging to any VQSR sensitivity tier over 99 % were excluded. This process resulted in an average of 4,518,323 SNPs and indels per subject (range: 4,318,448-4,701,297).

Stratification and inbreeding

Within the BIL&GIN and GOAL datasets separately, population structure was assessed by calling genotypes from the sequence data for selected sets of common variants (BIL&GIN: 77,553 variants, GOAL: 41,273 variants) spanning the autosomes. These were high-confidence single nucleotide polymorphism (SNP) sites identified by the 1000 Genomes Project, 1000G_phase1.snps.high_confidence.hg19.vcf.gz with minor allele frequencies (MAF) > 10% in each dataset[40], and had been pruned to be in low linkage disequilibrium (LD) with one another using the program PLINK (v1.9) (maximum LD r-square 0.2)[46,47]. Multidimensional scaling (MDS) was used to visualize the major dimensions of genome-wide variability (Figure S1). None of the first five dimensions was associated with the RHLD versus control distinction in either of the datasets (all $|T|<1$, $P>0.33$). Inbreeding was assessed with the F coefficient estimate within each dataset using PLINK (v1.9)[47]. The measure was not associated with the RHLD versus control distinction in either dataset (both $|T|<1$, $P>0.39$).

Note that common genetic variants were only used for the purposes of assessing population stratification and inbreeding within the datasets, whereas the rest of the study was focused on rare genetic variation, which has the potential to involve highly penetrant effects.

Annotation of SNPs and indels

SNPs and indels were annotated using Annovar [48] and Variant Effect Predictor (VEP v88)[49]. In the genome, non-synonymous protein-coding variants, and variants which affect splice donor and acceptor sites, are *a priori* the most likely to grossly alter gene function. Accordingly, Gemini (v.20.0) [50]was used to select protein coding variants with 'MEDIUM' or 'HIGH' impact severity annotations, as well as non-coding variants with 'HIGH' impact severity annotations (in practice those altering splice donor or acceptor sites). Additional filtering was done in R and comprised the removal of 'MEDIUM' variants with a PolyPhen [51] prediction score of "benign". Minor allele frequency (MAF) information was assigned as the maximum MAF across the GNOMAD (v1), ExAC (v3), 1KG, and ESP datasets (i.e. 'max_aaf_all' in Gemini), which together comprise whole exome or whole genome data from more than 120,000 people from various population datasets[52] (http://evs.gs.washington.edu/EVS/, http://www.internationalgenome.org/home). Within the BIL&GIN and GOAL datasets separately, any variants present in at least 19 participants (case or control) were excluded as they are likely to be platform-specific errors or else common variants not previously detected by other sequencing platforms or protocols, and would necessarily be present in at least two control subjects in BIL&GIN or three controls in GOAL (hence unlikely to be high-penetrance mutations for RHLD).

Monogenic mutation models

Recessive: Here we considered only homozygous or compound heterozygous mutations as potentially trait-causal. For screening purposes, compound heterozygosity was assigned when a given gene had at least two different mutations, although allelic phase information was not usually available due to the limited sequence read lengths. Variants were excluded when they had MAF >=10% on the basis of on-line population databases (see above). At 10% MAF, assuming Hardy-Weinberg equilibrium, the variant would be present in homozygous form at 1% in the population, i.e. roughly equal to the RHLD frequency in the population. In the case that 50% penetrance might arise from L-R randomization, as has been observed for mutations which cause *situs inversus* with primary ciliary dyskinesia[53], it is theoretically possible that a single causal variant in a gene could have up to 14% population frequency under a recessive model and Hardy-Weinberg equilibrium, and still be consistent with a trait frequency of 1%, if it was the only variant involved and caused all cases of the trait. However, allelic and genetic heterogeneity are typical for monogenic traits. Therefore a MAF threshold of 10% under a recessive model is an inclusive rather than strict filter. Variants not present or with no MAF information in the

population databases were retained. There were on average 43 recessively mutated genes per subject for the BIL&GIN-90bp protocol (range: 31-61), 64 per subject for the BIL&GIN-150bp protocol (range: 55-77), and 45 per subject for the GOAL dataset (range: 33-64). Integrative Genome Viewer (IGV v2.3.55) was used to visualize the possible compound heterozygous mutations, and genes carrying these were discarded when both mutations were definitely present on the same allele (i.e. "in phase") on a given sequence read.

Dominant: Here we considered heterozygous or homozygous mutations as potentially trait-causative. Variants were excluded as potentially causative when they had population MAF >=1% in the population databases, on a similar logic as for the recessive model above, but appropriate for allelic dominance and the frequency of RHLD in the population (roughly 1%). Variants not present or with no MAF information in the population databases were retained. There were on average 196 genes per subject for the BIL&GIN-90bp protocol (range: 154-215), 240 per subject for the BIL&GIN-150bp protocol (range: 208-268), and 262 per subject for the GOAL dataset (range: 229-300).

Gene-level testing

The BIL&GIN and GOAL datasets were combined for subsequent analysis.

We first verified that the total number of mutated genes per subject did not differ significantly between RHLD and control subjects, under either the dominant or recessive model (T-tests, all $p>0.10$). Significance for single-gene analysis was then assessed separately for individual genes and models (recessive or dominant), using the one-tailed Fisher's exact test for a 2x2 contingency table, for the categories 'mutated' and 'not mutated' in 33 RHLD subjects and 34 controls. The minimum number of mutated RHLD subjects to achieve a nominally significant P value (i.e. less than 0.05) was 5, i.e. if a gene would be mutated in 5 out of the 33 RHLD subjects and none of the 34 controls, that gene would show a nominally significant P value of association with RHLD, as a putative major-genetic effect (P value= 0.0267). This approach allows for allelic heterogeneity, i.e. the unit of testing is the gene, within which a variety of different mutations can be present. Note that the power and sample size considerations when modelling highly penetrant effects are different to typical genome-wide association studies of common traits, in which large samples are screened for common variants of small effect. Here we focus only on rare variants and interrogate the data with respect to the possibility of high penetrance. Note also that the Fisher's exact test is robust for the sample size, since the significance is assessed with respect to all of the actual possibilities that might have arisen in the contingency table in this set of subjects.

We calculated that for an individual gene to be significant at $P<0.05$ after Bonferroni multiple testing correction, it would have to be mutated in at least 11 (dominant) or 10 (recessive) of the 33 RHLD subjects, and no controls, leading to nominal P=0.000186 (dominant) or P=0.000373 (recessive) in the Fisher's exact test, i.e. the gene would need to be a monogenic cause for roughly one third of the instances of RHLD. For these calculations, we counted how many individual genes, y, have mutations in at least x subjects, for every value of x from 1 to 67 subjects. For each value of x, we then calculated the minimum number of RHLD subjects with mutations in a given gene that would be required to produce a P value less than $0.05/y$ in the Fisher's exact test.

We performed a post hoc filtering step in which we further excluded from consideration, as potentially monogenic effects, all genes which were mutated in at least one control subject, as these genes were unlikely to be causal monogenically for RHLD. Note that this filter was only applied after the statistical analysis, in order not to bias the multiple testing correction.

Mutational load in gene sets

We tested whether the RHLD cases had an increased mutational load in specific candidate gene-sets (see the Introduction for the rationale). These candidate sets, based on the gene ontology (GO) as defined within AmiGO's direct annotation[54,55] (http://geneontology.org/gene-associations/goa_human.gaf.gz downloaded 16-Nov-2017), were 'cilium' (GO:0005929), 'left-right axis specification' (GO:0070986), 'actin cytoskeleton' (GO:0015629), plus two sets defined on the basis of visceral laterality phenotypes or disorders: 58 genes related to primary ciliary dyskinesia and asymmetry disorders[20]; 62 genes either implicated in visceral asymmetry disorders or known to be involved in the visceral left-right developmental pathway[22], as well as a final set of 18 candidate genes which have been tentatively associated with human brain laterality in previous studies[16,32,34].

The GO terms were defined within AmiGO's[54,55] direct annotation (http://geneontology.org/gene-associations/goa_human.gaf.gz downloaded 16-Nov-2017). Additional sets were investigated *post hoc* as child sets of the actin cytoskeleton set (Table S2). Only gene sets comprising at least ten genes were considered.

To test for an increased mutational load within a given gene-set in RHLD, the sum of the number of mutated genes (as defined above) per subject within the set was compared between RHLD subjects and controls by means of the one-tailed exact binomial test, i.e. considering the sum of mutated genes per subject in RHLD subjects only, the total sum across RHLD and controls combined, and the proportion of all subjects who were RHLD (33/67). Again, as an exact test, the binomial is robust for the subject sample size, and does not require assumptions on the number of mutations per individual.

Association with handedness within the UK Biobank

Since the large majority of people with RHLD are left-handed, any monogenic contributions to RHLD would likely also be strongly penetrant for left-handedness. We checked whether a specific mutation of interest in the gene *TCTN1,* rs188817098, which we initially considered a potential candidate for causing RHLD in some subjects (see Results), is also associated with handedness the UK Biobank cohort data. There were 330,474 subjects (32,367 left-handed) available for this analysis. In this dataset, rs188817098 had been directly genotyped and was in Hardy Weinberg equilibrium (p=1), and the minor allele C had a frequency of 0.001305. Handedness (UK biobank field ID: 1707.0.0) was self-reported and coded for the present purposes as 'left-handed' or 'right-handed', as described elsewhere[56]. We performed association analysis of rs188817098 with handedness using the program BOLT-LMM (v2.3) which uses linear mixed effects regression under an additive genetic model [57]. The top 40 principal components capturing genetic diversity in the genome-wide genotype data, calculated using fastPCA[58] and provided by the UK biobank[59], were included as covariates to control for population structure, as well as sex, age, genotyping array, and assessment centre. The UK Biobank data were obtained as part of research application 16066, with Clyde Francks as the principal applicant. The data collection for the UK Biobank has been described elsewhere[60]. Informed consent was obtained by the UK Biobank for all participants.

Results

Monogenic mutational models

We focused on mutations in the 33 RHLD cases which are known to be relatively rare in the general population on the basis of large-scale genetic databases, and predicted to disruptively affect protein sequence, while not being mutated in a set of 34 control subjects (see methods). As noted above, a given gene would need to be a monogenic cause for at least 10 or 11 of the 33 RHLD cases in this study, and not mutated in controls, in order to be detected at a significant level after multiple testing

correction. There were no genes which met this threshold, under either the dominant or recessive models.

Under a recessive model, no gene was even nominally significant (i.e. showed unadjusted P<0.05), which could have arisen from being mutated in as few as five RHLD cases and no controls.

In the dominant model, *TCTN1* was the only nominally significant gene (p=0.0267 before multiple testing correction), with five RHLD cases and no controls having heterozygous mutations (Table 2). *TCTN1* encodes a member of a family of secreted and transmembrane proteins, and is a component of the tectonic-like complex, which forms a barrier between the ciliary axoneme and the basal body[61]. This gene tolerates missense and loss of function variation well (as reflected by the ExAC missense Z-score[52]: z = 0.20). Recessive mutations in *TCTN1* cause Joubert syndrome (JBTS, MIM #614173), a ciliopathy characterized by cerebellar and brainstem malformations[61,62].

Three of the five RHLD cases shared the same *TCTN1* missense variant (chr12:111080154 G/C, rs188817098), which has a maximum population frequency of 0.001199 (in ExAC non-Finnish Europeans). This variant is present in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) as a variant of uncertain significance with potential relevance to Joubert syndrome/Meckel-Gruber syndrome patients (SCV000634600.1). The other two *TCTN1* mutations were a missense variant (chr12:111078865 G/C, rs201990420) and an in-frame deletion (chr12:111070349 GATA/G), each present in one RHLD case only, and with maximum population frequencies of 0.0008 and 0.0033, respectively.

Rs188817098 was also associated with handedness in the UK biobank dataset (p=0.034), with the minor allele C (frequency 0.001305) associated with left-handedness (odd ratio = 1.24). However, this modest effect does not seem compatible with a role of this variant as a highly penetrant cause of RHLD and left-handedness.

Gene-set analysis

We analyzed a small number of candidate gene sets involved either in visceral laterality or else the actin cytoskeleton (see Introduction for the rationale). We observed an enrichment of mutations within the 'actin cytoskeleton' (GO:0015629) gene-set (Table 3). This gene set comprises 205 human genes (Table S1) which contribute to the actin cytoskeleton, i.e. the internal framework of the cell, composed of actin and associated proteins. Within the genomes of the 67 participants of this study, there were 171 different mutations present in 92 genes belonging to this set. 59.6% of the instances of mutated genes (102 out of 171) were in the subjects with RHLD, whereas the null probability of a mutated gene falling in a subject with RHLD was 49.25% (i.e. 33/67), exact binomial test P=0.0040 (Table 3 and Figure S3). This suggests that individuals with RHLD have a significant enrichment of rare, disruptive mutations in genes involved in actin cytoskeleton structure and function.

In contrast, no differences were found between participants with RHLD and controls for the gene ontology sets 'cilium' (GO:0005929), 'left-right axis specification' (GO:0070986), or sets defined on the basis of visceral laterality phenotypes or disorders[20,22], as well as the set of 18 candidate genes which have been tentatively associated with human brain laterality in previous studies (Table 3), consistent with language dominance being largely or wholly independent of these pathways/sets.

We investigated subsets of genes defined as belonging to specific components of the actin cytoskeleton, which included “actin filament” (GO:0005884), “myosin complex” (GO:0016459), and “cortical actin cytoskeleton” (GO:0030864), but saw no significant increase in mutation rates in RHLD in these sets (Table S2). This may indicate that subsets of actin cytoskeleton genes that are more specifically relevant to lateralized brain development have not been defined within the gene ontology.

Post hoc analysis of mutational load within the actin cytoskeleton gene set was further performed in different subsets of subjects according to handedness: RHLD versus right-handed controls only (P=0.04), RHLD versus left-handed controls only (P=0.004), right-handed controls versus left-handed controls (P=0.88) (Table S3). This pattern indicates that left-handedness without RHLD is not linked to an increased rate of mutations in actin cytoskeleton genes, and that the tentative increase was a specific property of the RHLD subjects.

Per dataset analysis showed that the increased mutational load in the actin cytoskeleton gene set was mostly driven by the BIL&GIN dataset (P=0.0006), while the effect was not significant in the GOAL dataset (P=0.4) despite having a similar trend of increased mutational load in RHLD cases (Table S4, Figure S3).

Discussion

Laterality is an important feature of the human brain's structural and functional organization[16,63,64]. Despite this, very little is known of the genetic contributions to typical brain laterality and its variation. In the present study, we performed the first molecular genetic investigation of RHLD, a trait which is present in only roughly 1% of the population. We focused on relatively rare coding variants that are predicted to disrupt protein functions. A highly penetrant mutated gene in roughly one third of the RHLD cases, and no controls, could have been detected at a significant level after adjusting for multiple testing in this study. This is a similar level of genetic heterogeneity as found in *situs inversus* of the visceral organs when it occurs together with primary ciliary dyskinesia, for which up to roughly one quarter of cases are due to mutations in a single gene, *DNAH5* [21].

However, we found no individual genes mutated in RHLD at this level, in the present study. It remains possible that some monogenic causes of RHLD were present in our dataset, but we could not distinguish them with the present sample size. Note that the sample size precluded an investigation of common genetic effects with low penetrance, i.e. the kinds of effects that are tested in typical genome-wide association studies of common traits. The approach here was necessarily focused only on rare variants, which might have sometimes acted as highly penetrant mutations. Nonetheless, it appears on the basis of our data that substantial genetic heterogeneity is likely to be involved in any heritable contribution to RHLD, even if some individual effects might be strongly penetrant. As noted in the introduction, non-leftward language dominance has previously been shown to have a heritability of roughly 30%, although the trait definition in that study included ambilateral individuals in addition to RHLD[11].

As RHLD is mostly found in left-handed people[3], and comprises roughly 10% of the left-handed population, then any highly penetrant genetic effects on RHLD would presumably also be strongly associated with left-handedness. One individual gene, *TCTN1*, carried rare, protein-altering mutations in five RHLD cases and no controls. Three of these cases carried the same rare variant, and the very large UK Biobank dataset, comprising hundreds of thousands of participants, allowed us to test this rare variant for association with left-handedness. (No functional imaging measures of language laterality were available in the UK Biobank to study RHLD in that dataset.) Although the *TCTN1* variant showed a significant association with left-handedness, in the expected direction (i.e. the minor allele associated with left-handedness), the effect size was not compatible with a highly penetrant effect. Therefore, this finding remains ambiguous.

In the present study, candidate genes which have been tentatively associated with human brain laterality in previous studies showed no evidence for an increase in mutation load in RHLD. The only gene among these that had more mutations in RHLD cases than in controls was *AR* (8 in RHLD cases, 6 in

controls). For most of these genes, there is no clear mechanism that might link them to left-right axis determination through chiral properties.

We also found no evidence that candidate gene sets involved in visceral laterality or primary ciliary dyskinesia have an enrichment of rare, protein-altering mutations in RHLD. This finding is consistent with the fact that people with *situs inversus* of the viscera, when it occurs together with primary ciliary dyskinesia, have shown normal population rates of left-handedness and left hemisphere language dominance[23-25]. Therefore, there appears to be a developmental disconnect between nodal-ciliary-induced visceral laterality and the functional brain lateralities for hand dominance and language. This suggests that at least some aspects of human functional brain laterality arise from an independent and unknown mechanism, which may be brain-intrinsic. A molecular-developmental pathway for laterality in the zebrafish brain has been relatively well described, but this appears to take its original cues from the nodal-visceral pathway, and thus the relevance for human functional brain laterality is not clear[65,66]. A relatively small-scale genome-wide association study in humans reported that genes involved in visceral laterality showed an enrichment of association signals with left-versus-right hand motor skill[67], but a much larger study of binary-trait handedness in the UK Biobank dataset, based on roughly 350,000 subjects, found no genetic link of handedness to visceral asymmetry genes[32]. Early life factors can also influence handedness, including birthweight, twinning, and breastfeeding, but to an extent which is not remotely predictive at the individual level[56].

Intriguingly, it may be that *situs inversus* of the visceral organs does associate with left-handedness when not due to mutations affecting the nodal ciliary pathway[25], although no causal genes were identified in a recent study which investigated the trait combination of *situs inversus* and left-handedness without primary ciliary dyskinesia[68]. Here we found initial evidence that people with RHLD have an elevated rate of rare, protein-altering mutations in genes involved in the structure and function of the actin cytoskeleton. This effect was robust to the use of either left or right-handed control groups, and thus was a specific property of RHLD subjects in this dataset, rather than left-handedness in general. We speculate that functional language laterality may be grounded in an evolutionarily ancient mechanism of inducing organ-intrinsic left-right morphogenesis, which can be traced back to the ancestral bilateria, and which arises from fundamental aspects of cellular biology and mechanics[26,29,69]. Developmental studies will be needed to assess whether cellular chirality is transiently present prior to asymmetric embryonic development of the mammalian brain. An understanding of how mutations of actin cytoskeleton genes might affect such a process will depend on detailed analysis of cellular models. An increased load of heterozygous mutations in genes affecting the actin cytoskeleton might affect brain laterality, while being otherwise well tolerated during development, due to compensation by non-mutated alleles at most of the genes involved. Given that common variants of the microtubule-associated gene *MAP2* have recently been associated with left-handedness by large-scale GWAS[32,33], our findings here in relation to RHLD may be broadly concordant, insofar as they also implicate the cytoskeleton in the developmental origins of human brain laterality.

The possible link of RHLD to actin cytoskeleton genes will need to be replicated in larger independent datasets. Within the present study, we combined the BIL&GIN and GOAL datasets to maximize the power to detect genetic effects on RHLD, although the functional tasks used to define RHLD differed between these two datasets: hemispheric dominance was defined using a contrast at the sentence level in BIL&GIN, and a word-level contrast in GOAL (see Methods). However, we are not aware of a large-scale data collection in existence, or currently underway, in which a harmonized phenotypic measure of RHLD will become available and which would be well-powered for GWAS.

Given the sample size for the present study, we focused on rare, protein-altering mutations which had the potential to be highly penetrant effects. Whole genome sequence data, of the type produced in the

present study, also contain information on noncoding variation. Rare noncoding variation has recently been implicated in neurodevelopmental disorders such as autism[70,71], and a significant fraction of this variation is potentially important for gene function and regulation[72]. The noncoding genome comprises 98% of the genome, and interpreting the variation within these regions is challenging. Several attempts have been made to rank potentially causative variants across the genome based on scores that integrate different types of information, including conservation of DNA sequence, regulatory information[73], and population genomic data. These ranking approaches include CADD[74], DANN[75], GWAVA[76], M-CAP[77], MetaSVM [78] or REVEL[79]. However, these ranking approaches are not very concordant with each other[72]. Moreover, the methods rely on assumptions about the deleteriousness/pathogenicity of variants, so that the overall approach is not an obvious fit for a non-pathogenic trait such as RHLD. Thus we did not pursue investigation of non-coding variation, which must await larger sample sizes and an improved understanding of the role of rare, non-coding variation in non-disease phenotypic variation.

Datasets based on hundreds of thousands of participants, such as the UK biobank[80], permit the estimation of how much of the variance in brain traits can be explained by common genetic variants, and the detection of genetic loci with very small effect sizes. However, the use of such large datasets is usually at the expense of detailed and accurate phenotypic characterization. Correlated structural [81] or resting-state derived indices [82] may offer alternative ways to study RHLD in large datasets, but these approaches will always be indirect. Hence, the approach taken in the present study is complementary to large-scale studies. We expect that convergent evidence arising from different strategies will help us better understand the biological underpinnings of language lateralization.

References

1. Knecht S, Drager B, Deppe M, et al. Handedness and hemispheric language dominance in healthy humans. *Brain.* Dec 2000;123 Pt 12:2512-2518.

2. Van der Haegen L, Cai Q, Seurinck R, Brysbaert M. Further fMRI validation of the visual half field technique as an indicator of language laterality: a large-group analysis. *Neuropsychologia.* Aug 2011;49(10):2879-2888.

3. Mazoyer B, Zago L, Jobard G, et al. Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLoS ONE.* 2014;9(6):e101165.

4. Tzourio-Mazoyer N, Perrone-Bertolotti M, Jobard G, Mazoyer B, Baciu M. Multi-factorial modulation of hemispheric specialization and plasticity for language in healthy and pathological conditions: A review. *Cortex.* Jan 2017;86:314-339.

5. Gilbert AN, Wysocki CJ. Hand preference and age in the United States. *Neuropsychologia.* Jul 1992;30(7):601-608.

6. Hepper PG, Shahidullah S, White R. Handedness in the human fetus. *Neuropsychologia.* 1991;29(11):1107-1111.

7. de Kovel CGF, Lisgo S, Karlebach G, et al. Left-Right Asymmetry of Maturation Rates in Human Embryonic Neural Development. *Biol. Psychiatry.* Aug 2017;82(3):204-212.

8. de Kovel CGF, Lisgo SN, Fisher SE, Francks C. Subtle left-right asymmetry of gene expression profiles in embryonic and foetal human brains. *Sci Rep.* Sep 2018;8(1):12606.

9. Ocklenburg S, Strockens F, Bless JJ, Hugdahl K, Westerhausen R, Manns M. Investigating heritability of laterality and cognitive control in speech perception. *Brain Cogn.* Nov 2016;109:34-39.

10. Somers M, Ophoff RA, Aukes MF, et al. Linkage analysis in a Dutch population isolate shows no major gene for left-handedness or atypical language lateralization. *The Journal of neuroscience : the official journal of the Society for Neuroscience.* 2015;35(23):8730-8736.

11. Somers M, Ophoff RA, Aukes MF, et al. Linkage Analysis in a Dutch Population Isolate Shows No Major Gene for Left-Handedness or Atypical Language Lateralization. *Journal of Neuroscience.* 2015;35(23):8730-8736.

12. Ocklenburg S, Strockens F, Bless JJ, Hugdahl K, Westerhausen R, Manns M. Investigating heritability of laterality and cognitive control in speech perception. *Brain Cogn.* Nov 2016;109:34-39.

13. Medland SE, Duffy DL, Wright MJ, et al. Genetic influences on handedness: data from 25,732 Australian and Dutch twin families. *Neuropsychologia.* Jan 2009;47(2):330-337.

14. Ge TaC, Chia-Yen and Neale, Benjamin M. and Sabuncu, Mert R. and Smoller, Jordan W. Phenome-wide heritability analysis of the UK Biobank. *PLOS Genetics.* 04 2017;13(4):1-21.

15. de Kovel CGF, Carrion-Castillo A, Francks C. A large-scale population study of early life factors influencing left-handedness. *Scientific reports.* Jan 24 2019;9(1):584.

16. Francks C. Exploring human brain lateralization with molecular genetics and genomics. *Ann. N. Y. Acad. Sci.* Nov 2015;1359:1-13.

17. Norris DP. Cilia, calcium and the basis of left-right asymmetry. *BMC Biol.* Dec 2012;10:102.

18. Fliegauf M, Benzing T, Omran H. When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell Biol.* Nov 2007;8(11):880-893.

19. Peeters H, Devriendt K. Human laterality disorders. *Eur J Med Genet.* 2006;49(5):349-362.

20. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nat. Rev. Mol. Cell Biol.* Sep 2017;18(9):533-547.

21. Hornef N, Olbrich H, Horvath J, et al. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am. J. Respir. Crit. Care Med.* Jul 2006;174(2):120-126.

22. Deng H, Xia H, Deng S. Genetic basis of human left-right asymmetry disorders. *Expert Rev Mol Med.* Jan 2015;16:e19.

23. McManus IC, Martin N, Stubbings GF, Chung EM, Mitchison HM. Handedness and situs inversus in primary ciliary dyskinesia. *Proc. Biol. Sci.* Dec 2004;271(1557):2579-2582.

24. Tanaka S, Kanzaki R, Yoshibayashi M, Kamiya T, Sugishita M. Dichotic listening in patients with situs inversus: brain asymmetry and situs asymmetry. *Neuropsychologia.* Jun 1999;37(7):869-874.

25. Vingerhoets G, Li X, Hou L, et al. Brain structural and functional asymmetry in human situs inversus totalis. *Brain Struct Funct.* Jan 4 2018.

26. Inaki M, Liu J, Matsuno K. Cell chirality: its origin and roles in left-right asymmetric development. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* Dec 2016;371(1710).

27. McDowell G, Rajadurai S, Levin M. From cytoskeletal dynamics to organ asymmetry: a nonlinear, regulative pathway underlies left-right patterning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences.* 2016;371(1710).

28. Okumura T, Utsuno H, Kuroda J, Gittenberger E, Asami T, Matsuno K. The development and evolution of left-right asymmetry in invertebrates: lessons from Drosophila and snails. *Dev. Dyn.* Dec 2008;237(12):3497-3515.

29. Lebreton G, Geminard C, Lapraz F, et al. Molecular to organismal chirality is induced by the conserved myosin 1D. *Science.* 11 2018;362(6417):949-952.

30. Davison A, McDowell GS, Holden JM, et al. Formin Is Associated with Left-Right Asymmetry in the Pond Snail and the Frog. *Curr. Biol.* Mar 2016;26(5):654-660.

31. Tee YH, Shemesh T, Thiagarajan V, et al. Cellular chirality arising from the self-organization of the actin cytoskeleton. *Nat. Cell Biol.* Apr 2015;17(4):445-457.

32. de Kovel CGF, Francks C. The molecular genetics of hand preference revisited. *bioRxiv.* 2018.

33. Wiberg A, Douaud G, Ng M, et al. Human Handedness: Genetics, Microtubules, Neuropsychiatric Diseases and Brain Language Areas. *bioRxiv.* 2018.

34. Gunturkun O, Ocklenburg S. Ontogenesis of Lateralization. *Neuron.* Apr 19 2017;94(2):249-263.

35. Labache L, Joliot M, Saracco J, et al. A SENtence Supramodal Areas AtlaS (SENSAAS) based on multiple task-induced activation mapping and graph analysis of intrinsic connectivity in 144 healthy right-handers. *Brain structure & function.* Dec 7 2018.

36. Wilke M, Lidzba K. LI-tool: a new toolbox to assess lateralization in functional MR-data. *J. Neurosci. Methods.* Jun 2007;163(1):128-136.

37. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning.* Jul 2003;52:91-118.

38. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* Mar 1971;9(1):97-113.

39. Van der Haegen L, Brysbaert M. Behavioral language and face lateralization2018.

40. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* Oct 2013;11(1110):1-11.

41. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* Jul 2009;25(14):1754-1760.

42. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* Aug 2009;25(16):2078-2079.

43. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* Sep 2010;20(9):1297-1303.

44. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* May 2011;43(5):491-498.

45. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics.* Jul 2015;31(13):2202-2204.

46. Purcell S, Chang C. PLINK [1.90]. *PLINK [1.90]*.

47. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.

48. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* Sep 2010;38(16):e164.

49. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* Aug 2010;26(16):2069-2070.

50. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* 2013;9(7):e1003153.

51. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat. Methods.* Apr 2010;7(4):248-249.

52. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* Aug 2016;536(7616):285-291.

53. Olbrich H, Haffner K, Kispert A, et al. Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat. Genet.* Feb 2002;30(2):143-144.

54. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* May 2000;25(1):25-29.

55. Consortium G. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* Jan 2017;45(D1):D331-D338.

56. de Kovel CGF, Carrion-Castillo A, Francks C. A large-scale population study of early life factors influencing left-handedness. *bioRxiv.* 2018.

57. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics.* Mar 2015;47(3):284-290.

58. Galinsky KJ, Bhatia G, Loh PR, et al. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *American journal of human genetics.* Mar 3 2016;98(3):456-472.

59. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* 2017.

60. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* Mar 2015;12(3):e1001779.

61. Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, et al. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat. Genet.* Jul 2011;43(8):776-784.

62. Srour M, Hamdan FF, McKnight D, et al. Joubert Syndrome in French Canadians and Identification of Mutations in CEP104. *Am. J. Hum. Genet.* Nov 2015;97(5):744-753.

63. Willems RM, Van der Haegen L, Fisher SE, Francks C. On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat. Rev. Neurosci.* 03 2014;15(3):193-201.

64. Toga AW, Thompson PM. Mapping brain asymmetry. *Nat. Rev. Neurosci.* Jan 2003;4(1):37-48.

65. Concha ML, Bianco IH, Wilson SW. Encoding asymmetry within neural circuits. *Nat. Rev. Neurosci.* Dec 2012;13(12):832-843.

66. Duboc V, Dufourcq P, Blader P, Roussigne M. Asymmetry of the Brain: Development and Implications. *Annu. Rev. Genet.* 2015;49:647-672.

67. Brandler WM, Morris AP, Evans DM, et al. Common variants in left/right asymmetry genes and pathways are associated with relative hand skill. *PLoS Genet.* 2013;9(9):e1003751.

68. Postema MC, Carrion-Castillo A, Fisher SE, Vingerhoets G, Francks C. The genetics of situs inversus totalis without primary ciliary dyskinesia. *bioRxiv.* 2018.

69. McDowell G, Rajadurai S, Levin M. From cytoskeletal dynamics to organ asymmetry: a nonlinear, regulative pathway underlies left\textendashright patterning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences.* 2016;371(1710).

70. Turner TN, Hormozdiari F, Duyzend MH, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. *Am. J. Hum. Genet.* Jan 2016;98(1):58-74.

71. Yuen RK, Merico D, Cao H, et al. Genome-wide characteristics of de novo mutations in autism. *NPJ Genom Med.* Aug 2016;1:160271-1602710.

72. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.* Aug 2017.

73. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* Sep 2012;489(7414):57-74.

74. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* Mar 2014;46(3):310-315.

75. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics.* Mar 2015;31(5):761-763.

76. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat. Methods.* Mar 2014;11(3):294-296.

77. Jagadeesh KA, Wenger AM, Berger MJ, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* Dec 2016;48(12):1581-1586.

678 **78.** Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods
679 for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* Apr
680 2015;24(8):2125-2137.
681 **79.** Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the
682 Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* Oct 2016;99(4):877-885.
683 **80.** Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK
684 Biobank prospective epidemiological study. *Nat. Neurosci.* Sep 2016.
685 **81.** Tzourio-Mazoyer N, Marie D, Zago L, et al. Heschl's gyrification pattern is related to speech-
686 listening hemispheric lateralization: FMRI investigation in 281 healthy volunteers. *Brain Struct*
687 *Funct.* 2015;220(3):1585-1599.
688 **82.** Joliot M, Tzourio-Mazoyer N, Mazoyer B. Intra-hemispheric intrinsic connectivity asymmetry and
689 its relationships with handedness and language Lateralization. *Neuropsychologia.* Mar 2016.

690

691

692 Acknowledgements

693 Thanks to all of the study participants. This research was funded by the Netherlands Organization for
694 Scientific Research (NWO, grant No. 054-15-101) and French National Research Agency (ANR, grant No.
695 15-HBPR-0001-03) as part of the FLAG-ERA consortium project 'MULTI-LATERAL', a Partner Project to the
696 European Union's Flagship Human Brain Project. Additional support was from the Max Planck Society
697 (Germany) and an InterLabex 2013 grant from the Bordeaux Initiative of Excellence (ANR grant No.–10–
698 IDEX-03-02). The GOAL data were collected with an Odysseus Grant awarded by the Government of
699 Flanders to M. Brysbaert. Some of the data in this study were obtained from the UK Biobank cohort, as
700 part of research application 16066, with Clyde Francks as the principal applicant. We are grateful to UK
701 Biobank for making data available.

702 Competing Financial Interests Statement

703 The authors declare no competing interests.

704 Figure legends

705 **Figure 1: Schematic figure showing the study design.** Images are shown from an example subject with
706 typical left-hemisphere language dominance, and an example subject with atypical Rightward
707 Hemispheric Language Dominance (RHLD), as assessed by functional Magnetic Resonance Imaging
708 (fMRI). Genomic analysis was focused on rare, protein-altering variants within genes and candidate
709 gene-sets.

710 **Figure 2: Hemispheric Functional Laterality Index (HFLI) distributions for the language task contrasts**
711 **within RHLD and control subjects.** Negative HFLIs indicate rightward functional laterality. Note that
712 GOAL samples were only assessed using Production HFLI.

Tables

Table 1: Summary statistics for language laterality measures and handedness, within the 67 participants of this study. EHI: Edinburgh Handedness Inventory score: median [min-max]. Median [min; max] values are shown for the three HFLI indexes. PROD, production; READ, reading; LIST, listening. See also Figure 2.

Dataset	Group	N	Sex (M/F)	Handedness (LH/RH)	EHI	$HFLI_{PROD}$	$HFLI_{READ}$	$HFLI_{LIST}$
BIL&GIN	RHLD	17	8/9	16/1	-22.92 [-100;100]	-58 [-72;-15]	-61 [-84;24]	-59 [-72;52]
	Controls	22	10/12	14/8	-77.78 [-100;100]	61 [29;83]	59 [16;84]	57 [25;79]
GOAL	RHLD	16	4/12	16/0	-100 [-100;-16]	-77 [-94;-45]	-	-
	Controls	12	0/12	0/12	90.5 [67; 100]	83 [49;90]	-	-

Table 2: All putative mutations within *TCTN1*. Chr: chromosome. Ref: Reference allele. Alt: Alternative allele. MAF: maximum minor allele frequency across 1KG, ExAC, gnomAD populations. RS ID refers the variant identity in dbSNP. AA: amino acid. PolyPhen prediction: PosD: possibly damaging. Sift prediction D: deleterious. PFAM: protein domain. CADD: CADD score v1. The RHLD and Ctrl columns show the numbers of these mutations in cases and controls (all were heterozygous).

Chr	Position	Ref	Alt	MAF	RS ID	Impact	AA change	Gemini severity	Poly Phen	Sift	PFAM	CADD	RHLD	Ctrl
12	111070349	GA TA	G	3.3E-3	rs529269328	inframe del	p.N235del	MED	-	-	DUF1619	-	1	0
12	111078865	G	C	8.0E-4	rs201990420	missense	p.V339L	MED	PosD	D	DUF1619	16.3	1	0
12	111080154	G	C	0.0014	rs188817098	missense	p.V431L	MED	PosD	D	-	26.1	3	0

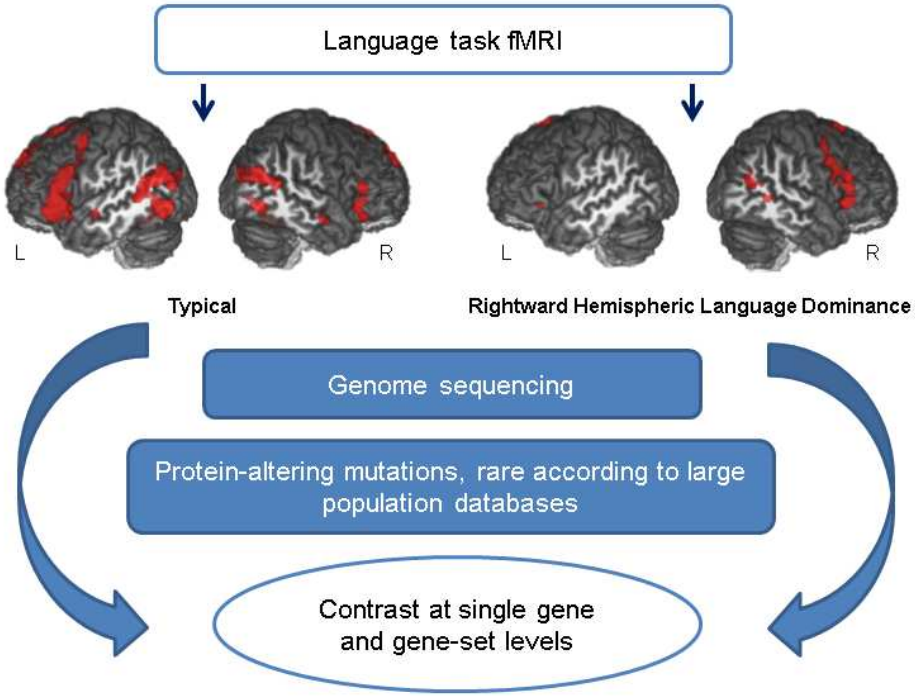
Table 3: Mutation load analysis of candidate gene sets. Set size: number of genes within set. RHLD: instances of genes carrying mutations within RHLD cases; Total: instances of genes carrying mutations in RHLD cases and controls combined. The P-value is shown from the exact binomial test, where the null probability was 0.493 (33/67 participants being RHLD) and alternative hypothesis = "greater". Reiter & Leroux (2017): 58 genes related to primary ciliary dyskinesia and asymmetry disorders. Deng et al. (2015): 62 genes either implicated in visceral asymmetry disorders or known to be involved in the visceral left-right developmental pathway. Francks (2015), Gunturkun & Okclenburg (2017), de Kovel & Francks (2018): 18 genes previously associated with brain/behavioural laterality phenotypes in humans.

Gene set	Set size	GO ID	RHLD	Total	P
Actin cytoskeleton	205	GO:0015629	102	171	0.004048
Cilium	173	GO:0005929	86	177	0.60
Left/right axis specification	13	GO:0070986	6	13	0.69
Reiter & Leroux (2017)	58	-	25	49	0.46
Deng et al. (2015)	63	-	29	60	0.61
Francks (2015) Gunturkun & Ocklenburg (2017) de Kovel & Francks (2018)	18	-	21	41	0.46

738 # Figures

739

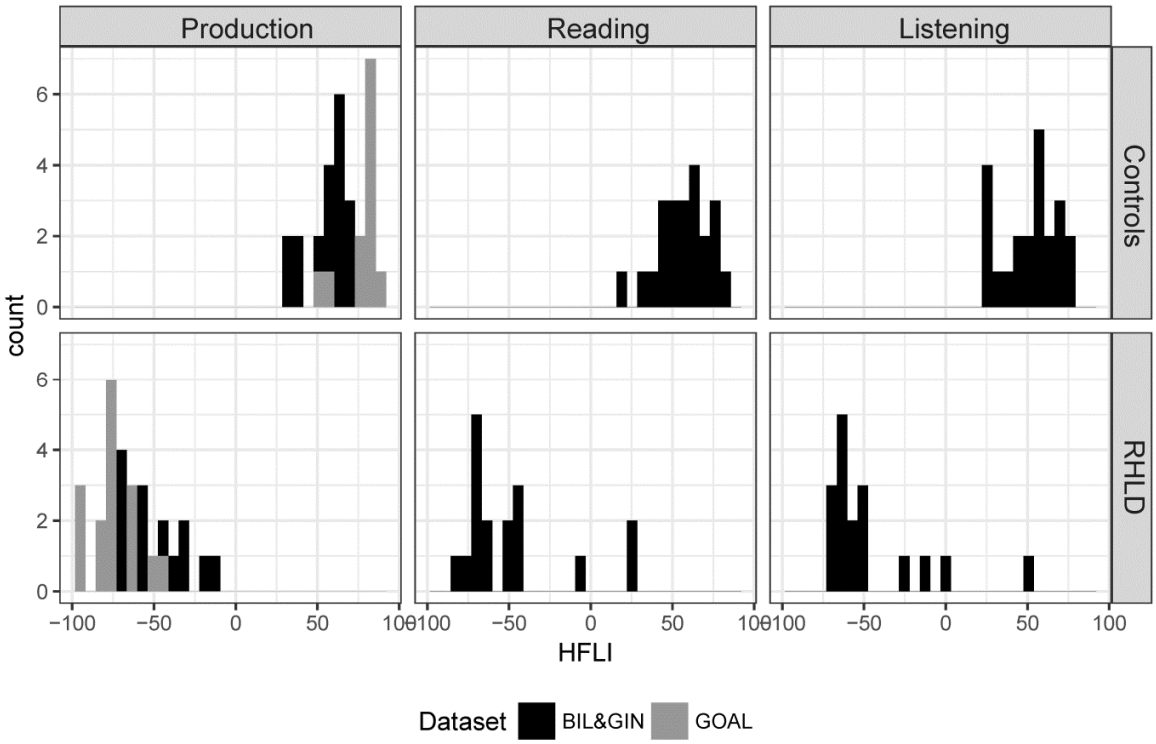
740 **Figure 1**



741

742

743 **Figure 2**



744