

OPEN

# The genetics of situs inversus without primary ciliary dyskinesia

Merel C. Postema<sup>1</sup>, Amaia Carrion-Castillo<sup>1</sup>, Simon E. Fisher<sup>1,2</sup>, Guy Vingerhoets<sup>1,3</sup> & Clyde Francks<sup>1,2\*</sup>

*Situs inversus* (SI), a left-right mirror reversal of the visceral organs, can occur with recessive Primary Ciliary Dyskinesia (PCD). However, most people with SI do not have PCD, and the etiology of their condition remains poorly studied. We sequenced the genomes of 15 people with SI, of which six had PCD, as well as 15 controls. Subjects with non-PCD SI in this sample had an elevated rate of left-handedness (five out of nine), which suggested possible developmental mechanisms linking brain and body laterality. The six SI subjects with PCD all had likely recessive mutations in genes already known to cause PCD. Two non-PCD SI cases also had recessive mutations in known PCD genes, suggesting reduced penetrance for PCD in some SI cases. One non-PCD SI case had recessive mutations in *PKD1L1*, and another in *CFAP52* (also known as *WDR16*). Both of these genes have previously been linked to SI without PCD. However, five of the nine non-PCD SI cases, including three of the left-handers in this dataset, had no obvious monogenic basis for their condition. Environmental influences, or possible random effects in early development, must be considered.

Roughly 1:6,000–8,000 people have *situs inversus* (SI), a mirror reversal of the normal asymmetrical arrangement of the viscera<sup>1,2</sup>. SI can occur in combination with Primary Ciliary Dyskinesia (PCD), a recessive genetic disorder which involves mutations that disrupt motile cilia<sup>3</sup>. Cilia are hair-like organelles that protrude from the cell surface into the extracellular space<sup>3</sup>. They are expressed in various tissues including the respiratory epithelium<sup>4</sup>, so that a disruption of ciliary motility can cause symptoms such as chronic bronchitis, inflamed or infected sinuses<sup>5</sup>, which are often present in PCD.

Motile cilia are also expressed early in development, within an embryonic structure called the ‘node’, where they generate a leftward fluid flow that helps to create the left-right body axis<sup>4</sup>. The leftward direction of the nodal flow may be explained by a posterior tilt of the cilia together with their clockwise rotation, arising ultimately from molecular chirality of their component proteins<sup>6–8</sup>. When leftward nodal flow is absent due to recessive PCD-causing mutations, many of which affect component proteins of the ciliary cytoskeleton, it becomes a matter of chance whether the viscera will take up the normal or mirror-reversed positioning<sup>9</sup>. The same principle of randomization appears in mice, for example through recessive mutations in the *Iv* gene<sup>10,11</sup>. Randomization can also explain why some human monozygotic twins are discordant for SI<sup>12</sup>.

However, roughly three quarters of people with SI do not have PCD<sup>1</sup>, and the causes of their SI remain largely unknown. A sibling recurrence risk of 100 times the population prevalence has been classically documented for SI<sup>13</sup>, although it is not clear whether SI individuals with PCD were distinguished from those without PCD when making this calculation, and we have been unable to obtain the primary reference<sup>14</sup>. A few genes have been reported to be involved in SI without PCD, and some of these can also cause partial disruptions of visceral laterality, known as heterotaxy or *situs ambiguus* (SA), including *ZIC3* [MIM:300265]<sup>15,16</sup>, *CCDC11* [MIM:614759]<sup>17</sup>, *CFAP52* [MIM:609804, also known as *WDR16*]<sup>18</sup>, *NME7* [MIM:613465]<sup>19</sup>, and *PKD1L1* [MIM:609721]<sup>20</sup>. The mechanisms by which these non-PCD genes influence visceral laterality are not well understood, but some code for cilia-associated proteins, rather than coding directly for cytoskeletal components of cilia.

A recent exome sequencing study of 323 people with laterality defects included 11 that had SI, while the rest had SA or cardiac malformations such as dextrocardia<sup>16</sup>. Of the 11 people with SI, only one had a clear candidate recessive mutation, in the known SA gene *MMP21*. PCD was not an inclusion criterion for that study, and none of the people with SI had recessive mutations in genes known to cause PCD. One additional participant with SI had a dominant mutation in *ROCK2*, and another had an X-linked mutation in *RAI2* – both of these genes are

<sup>1</sup>Department of Language & Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

<sup>2</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.

<sup>3</sup>Department of Experimental Psychology, Ghent University, Ghent, Belgium. \*email: [clyde.francks@mpi.nl](mailto:clyde.francks@mpi.nl)

important for heart development, but it is not clear whether mutations in either were causative for SI<sup>16</sup>. That was the largest previous study, of which we are aware, that screened the exome for mutations in people with non-PCD SI.

Approximately 90% of people are right-handed and have left-hemisphere dominance for language, among other lateralized functions, but the developmental basis for brain asymmetry remains unknown<sup>21–23</sup>. One possibility is that early embryonic mechanisms that give rise to asymmetries of the visceral organs also impact on brain asymmetries. Indeed, such a link has already been established in the *frequent-situs-inversus (fsi)* line of zebrafish<sup>24</sup>. One study of humans reported a possible genetic link between a continuous measure of left-versus-right hand motor skill and genes involved in visceral laterality, based on analyzing genetic variants which are common in the population<sup>25</sup>. However, the sample size of under 3000 subjects was low for complex-trait genome-wide association analysis using common genetic variants. A much larger study of over 300,000 subjects from the UK Biobank found no support for a link of left-handedness to genes involved in visceral asymmetry<sup>26</sup>. The same large study identified an association of a common variant at the microtubule-associated gene *MAP2* with left-handedness, with a small effect<sup>26</sup>.

Intriguingly, the general population rate of 10–15% left-handedness is not altered in SI with PCD<sup>27,28</sup>. This implies a developmental dissociation between brain laterality for hand motor control and nodal-ciliary visceral patterning. However, in the only previous study of handedness in SI to specifically target non-PCD individuals, Vingerhoets *et al.*<sup>29</sup> recently reported that five of nine non-PCD SI subjects were left-handed. Assuming a population rate for left-handedness of 10%, an exact, one-tailed binomial test gives  $P = 0.0009$  for this elevated rate of left-handedness in non-PCD SI. A further six SI subjects studied by Vingerhoets *et al.*<sup>29</sup> had PCD – but, as expected, they showed no elevated rate of left-handedness (1 out of 6).

The data of Vingerhoets *et al.*<sup>29</sup> suggest the existence of developmental mechanisms that link brain asymmetry and visceral laterality, independently of genes involved in PCD, although the small sample size cannot provide proof of this. In fact, a study from 1950 included 270 people with SI that were identified based on mass X-ray photography of the Norwegian population, and left-handedness was not over-represented compared to the general population<sup>12</sup>. The ascertainment method in that study suggests that most SI individuals did not necessarily have PCD, although this was not investigated formally<sup>12</sup>. If most SI individuals in the 1950 study did not have PCD, it would indicate that there is no elevation of left/handedness in non/PCD SI, and therefore that the elevated rate of left-handedness among the nine non-PCD SI participants of Vingerhoets *et al.*<sup>29</sup> was a chance finding in a small sample, or else an unusual feature of this particular set of subjects. Like the 1950 study, an even older study, from 1938, also reported a normal population rate of left-handedness in SI, although here ‘most of those included were examined on account of illness or a congenital malformation’<sup>30</sup>. In the 1938 study, therefore, it seems possible that many of the SI individuals had PCD, although further information was not given<sup>30</sup>. Left-handedness has a heritability of roughly 25% based on twin and family data<sup>31</sup>, but only around 2% based on genome-wide genotype data for common polymorphisms, within the UK Biobank dataset<sup>32</sup>. However, it has been proposed that left-handedness may sometimes occur due to genetic mutations that are rare in the population, but with substantial effects on brain laterality when present<sup>33,34</sup>. Rare genetic effects are not well captured in genome-wide association or SNP-heritability studies, which are typically based on common variation in the genome<sup>26</sup>.

Here we performed an exploratory whole genome sequencing study in the same set of 15 SI subjects studied by Vingerhoets *et al.*<sup>29</sup> (six with PCD and nine without), as well as 15 healthy controls matched for age, sex, education, and handedness (Table 1). The primary goal was to identify rare, highly penetrant mutations in the nine people with non-PCD SI, as relatively little is known about the genetic contributions to non-PCD SI. As these nine subjects showed an elevated rate of left-handedness, our study also had the potential to yield novel insights into the developmental origins of left-right patterning of both brain and body. The six SI subjects with PCD were included as positive controls, in which we expected to find recessive mutations in known PCD-causing genes.

## Methods

**Participants.** Table 1 gives an overview of the participants and their characteristics. Fifteen people with radiologically documented SI (mean age = 33 years, 8 males), and 15 controls with normal *situs* matched for age, sex, education and handedness (mean age = 33 years, 8 males), were recruited from Ghent University Hospital and Middelheim Hospital Antwerp. All participants were tested and scanned at Ghent University Hospital following ethical approval by the “Commissie voor Medische Ethisiek, Universitair Ziekenhuis Gent”, study approval number 2012/073. All participants gave informed consent for DNA sample collection and genomic analysis in relation to body, brain and behavioural laterality. All methods were performed in accordance with the relevant guidelines and regulations.

Details about the recruitment, diagnosis and selection/exclusion procedure have been described previously<sup>29</sup>. Radiological information (RX or CT) of thorax and complete abdomen was available for eight SI participants, and of thorax and upper abdomen in seven SI participants. The medical reports confirmed that all 15 participants presented with radiologically documented SI.

In five participants with SI, a formal diagnosis of PCD or Kartagener syndrome was found in their medical records. Kartagener syndrome refers to the clinical triad of SI, chronic sinusitis, and bronchiectasis<sup>9</sup>. A sixth SI-participant was identified on account of a radiological consultation regarding infertility. The participant also presented with chronic sinusitis and mild chronic bronchitis. Although no formal diagnosis of PCD was obtained in this case, the presence of chronic upper and lower respiratory infection and infertility in an individual with SI warrants suspicion of Kartagener syndrome. Consequently, we included the participant within the PCD group. In addition, all six members of the PCD group reported having a daily wet cough, and had PICADAR scores of between 8 and 12<sup>35</sup>, and thus predictive probabilities of having PCD of between 66% and 99%, based on this recently-developed, questionnaire-based tool (note that five of these subjects anyway had formal medical diagnoses).

Subj	SI group	Sex/Age	EHI	NH	CHD	Daily wet cough	Type	Gene	Clinvar	rs ID	Start position	ref	alt	MAF	AAC	impact
SI02	non-PCD	M/50	0.9	R	0	?	hom	<i>PKDIL1</i>	SA   SI	rs528302390	47870810	T	TCA	0.00109	—	splice donor
SI03	non-PCD	F/26	-0.8	L	0	yes		unsolved								
SI04	non-PCD	M/23	-1	L	1	no		unsolved								
SI05	non-PCD	M/27	0.9	R	0	no		unsolved								
SI07	non-PCD	F/35	0.9	R	1	?		unsolved								
SI09	non-PCD	F/36	0.7	L <sup>§</sup>	0	no		unsolved								
SI12	non-PCD	F/40	0.9	R	0	no	chet	<i>DNAH5</i>	Kartagener	None	13900415	A	C	—	E/*	stop gained
							chet	<i>DNAH5</i>	Kartagener	None	13769691	G	GC	0.00002	A/X	frameshift
SI14	non-PCD	M/18	-0.8	L	1	no	chet	<i>CFAP52/WDR16</i>	SI	rs148905544	9489207	T	A	0.00066	H/Q	Missense variant
							chet	<i>CFAP52/WDR16</i>	SI	rs374035006	9536224	C	T	0.00012	R/*	Stop gained
SI16	non-PCD	M/21	-1	L	0	no	hom	<i>CCDC151</i>	Kartagener	rs765121016	11533429	T	TCTC	0.00011	E/-	inframe deletion
SI06	PCD	M/46	1	R	0	yes	hom	<i>LRRC6</i>	Kartagener	rs767624733	133687728	T	C	0.00017	—	splice donor
SI08	PCD	F/23	0.9	R	0	yes	hom	<i>DNAH11</i>	PCD	rs373706559	21659620	A	C	0.00006	Y/*	stop gained
SI11	PCD	F/32	0.9	R	0	yes	chet	<i>DNAAF1</i>	Kartagener	rs569633512	84203963	C	T	—	—	splice donor
							chet	<i>DNAAF1</i>	Kartagener	rs373103805	84193302	G	C	0.00023	N/K	missense
SI13	PCD	M/48	0.6	L <sup>§</sup>	0	yes	hom	<i>CCDC114</i>	Kartagener	rs779459076	48814907	CACG	C	0.00093	-/R	Inframe insert
SI15	PCD	F/31	0.7	R	0	yes	chet	<i>DNAH5</i>	Kartagener	None	13786289	A	T	—	K/*	stop gained
							chet	<i>DNAH5</i>	Kartagener	rs397515540	13753397	G	GA	0.00034	D/X	frameshift
SI17	PCD	M/39	0.5	R	0	yes	chet	<i>DNAH5</i>	Kartagener	rs548521732	13839638	C	T	0.00021	—	splice acceptor
							chet	<i>DNAH5</i>	Kartagener	rs769458738	13753597	T	C	0.00004	R/H	missense

**Table 1.** The most likely causal recessive mutations in the fifteen SI subjects. Subjects shown with two mutations have possible compound heterozygous mutations, otherwise they have homozygous mutations. The Genome Reference Consortium (GRC) build 37 decoy version was used as reference sequence. EHI: Edinburgh Handedness Inventory score; NH: natural handedness; CHD: Congenital Heart Disease. Type: type of genetic mutation, i.e., homozygous (hom) or compound heterozygous (chet); MAF: minor allele frequency in population databases, if known; AAC: amino acid change. <sup>§</sup>Self-identified natural lefthander made to convert to right-handedness.

A seventh SI subject (SI03) had no medical record pertaining to PCD but did report daily wet cough. This subject had a PICADAR score of 8. In the study of Vingerhoets *et al.* this SI subject was classified as non-PCD, before the PICADAR assessment was available<sup>29</sup>. For the purposes of the present study, we also treated this subject as non-PCD given the lack of formal medical history of PCD and the ambiguous PICADAR score, as well as the fact that we found no likely recessive mutations in PCD-causing genes in this subject (see Results).

Eight other SI subjects had no medical record of PCD or PCD-like symptoms, and were classified as non-PCD SI cases. Six of these reported no daily wet cough, and two did not answer in this regard. PICADAR scores can only be calculated in the presence of daily wet cough, so that none of these eight cases received PICADAR scores. Three of these cases had been previously diagnosed with congenital heart disease that required surgical treatment, and their radiological files all referred to their cardiac condition. Congenital heart disease is a frequent comorbidity of SI, as the cardiac circulation appears particularly sensitive to perturbation in normal left-right positional information<sup>36</sup>.

Handedness was assessed using the Edinburgh Handedness Inventory (EHI)<sup>37</sup>. All participants had EHI scores  $>|0.5|$  (Table 1), so that none were ambidextrous<sup>38</sup>. Note that one non-PCD SI subject reported being forced to switch from left- to right-handedness in childhood, in which case five out of nine non-PCD SI cases were naturally left-handed. One of the six cases with PCD also reported forced left-to-right switching, otherwise the rest were right-handed (Table 1).

**Whole genome sequencing (WGS) and Pre-processing.** DNA was extracted from saliva samples using the Oragene kit (Oragene). WGS was performed by Novogene (Hong Kong) using Illumina high throughput sequencing (HiSeq-PE150), creating paired end reads with a length of 150 base pairs (bp). Raw reads, stored in BAM files, were aligned to the human reference genome (the extended “decoy” version of b37) using Burrows-Wheeler Aligner (BWA) software<sup>39</sup>, and sorted and reordered using SAMtools (v1.3.1)<sup>40</sup>. PCR duplicates, which could arise during cluster amplification, were marked using Picard (v2.9.0). Genome Analysis Toolkit (GATK v3.7)<sup>41</sup> software was used to realign reads around insertions/deletions (indels) and to recalibrate base quality scores per sample.

**Variant calling and quality control.** Indels (insertion/deletions) and single nucleotide polymorphisms (SNPs) were called as recommended by the GATK Best Practices workflow. Variant Quality Score Recalibration (VQSR) was performed, and variants with a high probability of being false positive were flagged according to their sensitivity tranche (90, 99, 99.9 and 100). All SNPs and indels within a VQSR tranche of 99.9% or higher were discarded. Variants with a quality depth  $\leq 9$  or a call rate  $\leq 0.8$  were also removed. VcfTools v1.13 was used to create

a summary table in the Variant Call Format (vcf). A total of 13,989,941 SNPs and indels were identified across the 30 subjects, with an average number of 5,186,055 (min = 5,053,188, max = 5,272,561) alternative alleles per subject (i.e. different to the reference genome, build37 decoy version).

**Genomic-level evaluation.** For overall descriptive analysis of the participant genomes, a subset of 40,387 variants distributed genome-wide was used that had known Minor Allele Frequencies (MAFs)  $>0.1$ , and were the result of pruning to be in low linkage disequilibrium with one another. For this, the flag–indep-pairwise function in Plink (v1.90b3w)<sup>42</sup> was applied with a pairwise linkage disequilibrium  $r^2$  greater than 0.2, based on a SNP-SNP correlation matrix of 1500 by 150 in window size. The resulting data were then used as the basis for inferring pairwise identity by descent (IBD) sharing between subjects (i.e. genetic relatedness), inbreeding, and inconsistencies with reported sex, using the Plink operations–genome,–ibc/het, and–check-sex.

A different subset of 61,467 independent variants distributed genome-wide was used for visualizing population structure through Multi-Dimensional Scaling (MDS) analysis with respect to the 1000 genomes reference dataset (v37)<sup>43</sup> (downloaded from: [ftp://climb.genomics.cn/pub/10.5524/100001\\_101000/100116/1kg\\_phase1\\_all.tar.gz](ftp://climb.genomics.cn/pub/10.5524/100001_101000/100116/1kg_phase1_all.tar.gz)) using the Plink operation –mds-plot.

**Annotation and filtering of single nucleotide variants affecting protein sequences.** Variants were normalized using the software tool *vt normalize* (v0.5772–60f436c3)<sup>44</sup>, which ensures consistent representation of variants across the genome. Normalized SNVs were annotated using Annovar<sup>45</sup> and Variant Effect Predictor (v88)<sup>46</sup>. Gemini (v 0.20.0) 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 synonymous variants, and of ‘MEDIUM’ variants with a PolyPhen<sup>47</sup> prediction of ‘benign’ unless annotated as ‘deleterious’ by Sift<sup>48</sup>, and those annotated as ‘tolerated’ by Sift unless predicted to be ‘probably damaging’ by PolyPhen. Data were then processed and analyzed separately under recessive and dominant models:

**Recessive model.** For the recessive model we further excluded variants with a known population frequency  $\geq 0.005$  in any of the following databases: GNOMAD<sup>49</sup>, ESP<sup>50</sup>, 1000 Genomes<sup>43</sup> and ExAC databases<sup>49</sup>. The remaining low-frequency variants were considered as putative mutations. Gene-level mutation counts per subject were then made, with a given gene being assigned as recessively mutated when it carried two copies of the same mutation (homozygous) or two different mutations (possible compound heterozygous). Integrative Genome Viewer (IGV v2.3.97)<sup>51</sup> was used to visualize the possible compound heterozygous mutations, and genes carrying these were discarded as recessive candidates when both mutations were definitely present on the same allele (i.e. “in phase”) on a given sequence read. Finally, genes recessively mutated according to these criteria in any of the fifteen unaffected control subjects were excluded as being potentially causative in cases, and also removed for the purposes of Gene Set Enrichment Analysis (GSEA; see below): this step had the advantage of removing false variants from potential technical artifacts, or variants which are common in the population but not annotated as such in on-line databases. In total, 17 genes were excluded based on overlap with the unaffected control subjects. Genes on chromosome X were processed as part of the recessive pipeline, such that females would need to carry two mutations in a given gene, and males one mutation.

**Dominant model.** A maximum population frequency threshold of  $5 \times 10^{-5}$  was applied in this case, and genes carrying at least one rare variant according to this criterion were considered as potentially causative. Again, genes mutated according to this criterion in any of the fifteen unaffected control subjects ( $N = 47$  genes) were excluded as being potentially dominantly causative in cases, and removed for the purposes of GSEA analysis (below).

**Gene set enrichment analysis.** To test whether a list of mutated genes in a given set of subjects contained functionally related genes across subjects, we performed Gene Set Enrichment Analysis (GSEA) using the g:Profiler R package (version 0.6.1)<sup>52</sup>. A gene classification scheme derived from the Gene Ontology (GO) database<sup>53,54</sup> was used. This specified a total of 6380 functionally defined gene-sets, based on a background of 19,327 genes with functional annotations. Many genes are not represented in the GO, particularly mRNA transcripts of no known function or homology, so that the numbers of mutated genes in a given set of subjects was always higher than the subset used as input for GSEA.

Mutated genes on chromosome X were combined with recessively mutated autosomal genes for the purposes of GSEA. GSEAs were performed separately for the mutated gene lists in SI subjects with PCD (54 genes, of which 40 are in the GO), non-PCD SI subjects (60 genes, 38 in the GO), left handed non-PCD SI subjects (42 genes, 26 in the GO), and non-PCD SI cases that were unsolved under the recessive model (27 genes, 13 in the GO).

For the dominant model, GSEA was performed separately for mutated genes in the non-PCD SI subjects (330 genes, 271 in the GO), the subset of non-PCD SI subjects that remained unsolved under the recessive model (183 genes, 148 in the GO), and the left handed subset of non-PCD subjects (201 genes, 163 in the GO). PCD subjects were not tested under a dominant model, as PCD is known to be a recessive phenotype.

In order to confirm that unaffected controls showed no significant pathway enrichment among their mutated genes, a mirrored exclusion was performed whereby any genes mutated in cases were excluded from the control gene lists. This resulted in 56 genes exclusively mutated in controls for the recessive model (of which 34 genes are in the GO), and 533 genes exclusively mutated in controls under the dominant model (440 in the GO).

The identities of genes were first converted using the g:Convert tool<sup>52</sup> to ensure recognition by the GO schema. The following settings were then used for GSEA: minimum set size = 15, maximum set size = 500, minimum intersect number = 2, hierarchical filtering = moderate. P-values were corrected for multiple testing across gene

sets, based on the gSCS correction method in gProfiler, separately for each input list of mutated genes corresponding to a given set of subjects. This method of multiple testing correction takes into account the hierarchical structure of the sets<sup>52</sup>. We applied a cut-off P-value of adjusted 0.01.

**Candidate gene lists.** We queried the genetic data in R (v3.3), based on searching the Clinvar<sup>55</sup> column of our annotated variant data for the terms: “*situs inversus*”, “*heterotaxy | heterotaxia | situs ambiguus*”, “*PCD | ciliary dyskinesia | Kartagener*”, “*left-right*”, and “*asymmetry | laterality*”.

A candidate gene list was also created which included genes from the literature that were suggested to be associated with human laterality phenotypes and/or ciliopathies, but not (yet) present in our initial list of candidate genes (Supplementary Table S1). Specifically, a list of ciliopathy genes from a review of this topic<sup>5</sup> was searched for the same terms as mentioned above plus ‘congenital heart disease’, yielding 57 candidate genes. Additionally, a list of genes potentially associated with human laterality disorders, as compiled in a 2015 review<sup>1</sup>, resulted in the addition of 38 candidate genes. Five more genes were included that were considered as possibly causal of laterality defects in a recent exome sequencing study<sup>16</sup> (see Introduction). Eleven genes containing variants that were previously reported to associate with measures of hand preference<sup>26</sup> or *planum temporale* asymmetry<sup>56</sup> were also added to the list of candidate genes. Finally, 47 more genes – arising from a search for the ‘*situs inversus*’ phenotype – were retrieved from the Mouse Genome Database (MGD)<sup>57</sup> at the Mouse Genome Informatics website, the Jackson laboratory, Bar Harbor, Maine. (URL: <http://www.informatics.jax.org>) [Oct, 2017].

**Structural variants.** For all participant genomes, structural Variants (SVs) (>50 kilobases in length) were called using a combination of two SV calling algorithms: CNVnator (v0.3.3)<sup>58</sup> and Lumpy (v 0.2.13)<sup>59</sup>. These algorithms complement each other regarding the kinds of signals in WGS data that they can detect, as CNVnator is based on read-depth, whereas Lumpy is based on paired-end mapping<sup>60</sup>. For CNVnator the bin size was set to 100 base pairs for all genomes except for two, where it was 150 base pairs (we determined a roughly optimal bin size for each subject’s genome, such that that the average read depth for that genome would be at least 4 standard deviations from zero). Lumpy was run using default parameters in “lumpyexpress”.

SVs were then annotated using SV2<sup>61</sup>. Variants that were present in any of the 15 healthy controls were removed from consideration as potentially causative for SI. Only variants that passed the default SV2 filtering criteria for quality were included<sup>61</sup>.

## Results

**Protein-altering single nucleotide variants.** *Recessive mutations.* Our variant calling, filtering and annotation pipeline produced between 5 and 15 recessively mutated genes per SI subject. We included six SI subjects with PCD as positive controls, in order to ensure that the variant calling and mutation definition criteria were well calibrated. As PCD is known to be a recessive phenotype for which at least 30 different genes have already been identified<sup>1</sup>, we expected most, or all, of these six subjects to have identifiable mutations in known PCD-causing genes. As expected, each of these six cases had just one recessively mutated gene which was annotated ‘Kartagener’ or ‘PCD’ in the Clinvar database<sup>55</sup>, and was therefore the most likely monogenic cause for their condition (Table 1). These genes were *LRRC6* [MIM:614930], *DNAH11* [MIM:603339], *DNAAF1* [MIM:613190], *CCDC114* [MIM:615038], and *DNAH5* [MIM: 603335] (the latter mutated in two SI cases with PCD, consistent with *DNAH5* being the most common cause of PCD in European-ancestry populations<sup>62</sup>) (Table 1). The PCD subject with a homozygous *LRRC6* mutation (subject SI06) was the only individual to show an elevated inbreeding coefficient and non-European ancestry (Supplementary Table S2, Supplementary Fig. S1).

None of the fifteen unaffected control subjects had any recessively mutated genes annotated ‘Kartagener’, ‘PCD’, ‘SA’ or ‘SI’ in Clinvar.

Surprisingly, two of the nine non-PCD SI cases had recessive mutations in genes annotated ‘Kartagener’ in the Clinvar database<sup>55</sup>. These were subject SI12 (again involving mutations in *DNAH5*), and subject SI16 (in *CCDC151* [MIM:615956]) (Table 1). As these subjects had no medical records pertaining to PCD-like symptoms, and no daily wet cough, then the findings suggest reduced penetrance for PCD.

Two of the nine non-PCD SI subjects had recessive mutations in genes annotated in Clinvar with ‘*heterotaxy*’ and/or ‘*situs inversus totalis*’, but not annotated as PCD-causing (Table 1). These genes were *PKD1L1* [MIM: 617205] in subject SI02, and *CFAP52* [MIM: 609804] in subject SI14. A homozygous missense mutation in *PKD1L1* was previously reported in an individual with SI and congenital heart disease but no PCD, as well as recessive splicing mutations in two individuals with heterotaxy<sup>20</sup>. Our subject SI02 had no diagnosis of congenital heart disease (CHD), and was right-handed (Table 1). As *PKD1L1* is a known recessive cause of non-PCD SI, we consider this gene to be the most likely cause of the non-PCD SI in subject SI02. As regards *CFAP52* (also known as *WDR16*), a homozygous deleterious deletion was previously reported to segregate with *situs* anomalies, including heterotaxy and *situs inversus totalis*, in a consanguineous family<sup>18</sup>. SI14 carried possible compound heterozygous mutations in this gene (Table 1), which are therefore the most likely causes of non-PCD SI in this individual. SI14 is left-handed and has CHD (Table 1).

Subject SI03, who had no formal medical history of PCD, but had reported an intermediate PICADAR score, showed no recessive mutations in known PCD genes, which tends to support non-PCD status for this subject. This meant that five non-PCD SI cases did not have recessive mutations in genes known to cause human laterality disorders, as annotated in Clinvar, and therefore remained ‘unsolved’ under a recessive model (Table 1). Among these five non-PCD SI cases, three were left-handed, and were therefore of primary interest for the present study. Three of these same subjects also had CHD (Table 1).

We constructed a list of known or suspected laterality genes with reference to the literature and mouse laterality phenotypes (Methods; Supplementary Table S1), but none of these genes had recessive mutations in the five unsolved non-PCD SI subjects.

We note possible compound heterozygous mutations in *PKD1* [MIM:601313], as a paralogue of *PKD1L1* [MIM:609721], in subject SI14 (Supplementary Table S3). However, mutations in this gene would be expected to cause autosomal dominant Polycystic Kidney Disease<sup>63,64</sup>, and since there was no such diagnostic record for this subject, one or both of these specific mutations probably has limited functional impact and is therefore unlikely to be a monogenic cause for SI either. As noted above, SI14 anyway had likely causative mutations in *CFAP52*.

*KIF13B* [MIM:607350] was putatively recessively mutated in subjects SI12 (likely solved through *DNAH5*) and SI14 (likely solved through *CFAP52*) (Supplementary Table S3). Although no literature has linked *KIF13B* to PCD or laterality phenotypes, *KIF3A* [MIM:604683], another kinesin encoding gene, is known to be involved in LR determination<sup>65</sup>. Moreover, *KIF3B* [MIM: 603754] is known to affect motility of nodal cilia and LR determination<sup>66</sup>. Together with dyneins, kinesins allow ciliary proteins to enter the organelle via the transition zone by transporting them as cargo<sup>67,68</sup>, and thereby play an important role in ciliary construction and maintenance<sup>67,68</sup>. The mutations in *KIF13B* might therefore conceivably have affected phenotypic outcomes in subjects SI12 and SI14, but we cannot confidently assign roles to *KIF13B* on the basis of this evidence.

**Dominant mutations.** For the five non-PCD SI cases who remained unsolved under a recessive model, we also considered a dominant model using a maximum known mutation frequency of  $5 \times 10^{-5}$ , and again cross-referenced the mutated genes against Clinvar and our candidate gene list (Supplementary Table S1), but no likely causative genes emerged (Supplementary Table S3) (see Discussion).

Subject SI05 showed a heterozygous mutation in *LRRC6* [MIM:614930] (Supplementary Table S3), which was included among our candidate genes. However, since recessive – but not dominant - mutations in this gene have been associated with PCD<sup>69</sup>, it is unlikely that this mutation is causal for non-PCD SI in this subject.

We also note a heterozygous mutation in *WDR62* [MIM:613583] in the unsolved case SI09 (Supplementary Table S3). Although mice with mutations in this gene have shown dextrocardia and right pulmonary isomerism (MGI:5437081)<sup>57</sup>, humans with recessive *WDR62* mutations do not show altered laterality. Instead, they have shown infantile spasm, microcephaly and intellectual disability<sup>70</sup>. It is therefore unlikely that *WDR62* is a dominant cause of altered laterality in humans.

The same subject SI09 also had a heterozygous mutation in *PLXND1* [MIM:604282] (Supplementary Table S3), a gene which appears among search results for the phenotype ‘*situs inversus*’ within the Mouse Genome Database<sup>57</sup>. However, while *PLXND1* is annotated as a cause of aortic arch and atrial abnormalities in this database, it is not annotated with *situs inversus* or heterotaxia, so that the basis for the search result is unclear. We did not find evidence for this gene’s involvement in visceral laterality in a further literature search.

Subject SI03 had a heterozygous mutation in *SPEF2* [MIM:610172] (Supplementary Table S3). Loss of *SPEF2* function in mice results in spermatogenesis defects and primary ciliary dyskinesia<sup>71</sup>, but this is through recessive mutations, so that we do not consider the heterozygous mutation in SI03 as a strong causal candidate for non-PCD SI.

**Gene set enrichment analysis.** We first performed gene set enrichment analysis using the positive control set of six SI subjects with PCD. As noted above, the six PCD subjects had likely recessive monogenic causes in five different genes (two subjects had mutations in the same gene, *DNAH5*). As expected, from the total combined list of recessively or X-linked mutated genes in these six subjects, of which 40 genes were included in the GO schema, we obtained significant results for various cilia-related pathways, such as ‘axoneme’ ( $p = 0.004$ ), ‘outer dynein arm assembly’ ( $p = 3.85 \times 10^{-5}$ ), ‘dynein complex’ ( $p = 4.89 \times 10^{-5}$ ), and ‘microtubule based movement’ ( $p = 1.41 \times 10^{-4}$ ) (Table 2) (all P-values adjusted for multiple testing across gene sets, see Methods). When the single most likely causative gene for each PCD subject was removed from the gene list, which left 36 recessively or X-linked mutated genes in these subjects that are present in the GO schema, there were no longer any significant enrichment terms, which further supports that the monogenic causes had been correctly identified in these subjects. The gene set enrichment analysis in the PCD subjects confirmed that, despite a relatively small number of subjects (i.e. six), the analysis was well powered to identify affected biological processes, even when most individual subjects had monogenic causes in different genes, and each subject had other, non-causative mutated genes.

With this in mind, we then performed gene set enrichment analysis in the non-PCD SI subjects, who were of primary interest for the present study. However, no significant enrichments were found when testing the list of recessively or X-linked mutated genes in the nine subjects with non-PCD SI, of which 38 genes were included in the GO schema. There was also no significant functional enrichment when testing the recessively mutated or X-linked genes in the five left-handed subjects with non-PCD SI, of which 26 genes were in the GO schema, and nor when testing the list of genes in the five unsolved non-PCD SI subjects, of which 12 genes were present in the GO schema (Table 2). In addition, gene-set enrichment analysis with dominantly mutated genes as input did not produce significant results in the non-PCD SI cases, nor the left-handed or unsolved subsets.

As expected, the lists of recessively/X-linked and dominantly mutated genes in the fifteen unaffected control subjects did not produce any significant gene set enrichment terms (Table 2).

**Structural variant analysis.** We additionally screened the genomes of all subjects for structural variants (such as larger-scale deletions and duplications), using two complementary algorithms (see Methods). SI subjects had SVs affecting an average of 9.7 genes per subject (min = 2 SVs, max = 16 SVs), and controls had SVs affecting an average of 9.6 genes per subject (min = 5 SVs, max = 16 SVs). None of the SI subjects, regardless of PCD status, had SVs affecting genes that were annotated SI, PCD, Kartagener, *situs ambiguus* (SA), or Heterotaxy (HTX) in Clinvar, nor affecting genes from our list of candidate laterality genes.

Group	p-value <sup>a</sup>	term size	query size <sup>b</sup>	overlap size	term ID	term name	intersection	samples	Pval Fisher <sup>c</sup>
Non-PCD SI (n=9) <sup>d</sup>	NS		38						
Non-PCD SI LH (n=5) <sup>d</sup>	NS		26						
Unsolved cases (n=5 <sup>d</sup> )	NS		22						
SI with PCD (n=6)	$3.85 \times 10^{-5}$	17	40	4	GO:0036158	outer dynein arm assembly	DNAH5, CCDC114, LRRC6, DNAAF1	SI06, SI11, SI12, SI13, SI15, SI17	$7.71 \times 10^{-13}$
SI with PCD (n=6)	$1.41 \times 10^{-4}$	296	40	8	GO:0007018	microtubule-based movement	DNAH5, CCDC114, DNAH11, WDR60, LRRC6, DNAAF1, FMN2, DNAH12	SI06, SI08, SI11, SI12, SI13, SI15, SI17	$9.74 \times 10^{-15}$
SI with PCD (n=6)	$4.89 \times 10^{-5}$	48	40	5	GO:0030286	dynein complex	DNAH5, CCDC114, DNAH11, WDR60, DNAH12	SI08, SI11, SI12, SI13, SI15, SI17	$1.68 \times 10^{-13}$
SI with PCD (n=6)	0.004	113	40	5	GO:0005930	axoneme	DNAH5, CCDC114, DNAH11, DNAAF1, RP1L1	SI03, SI08, SI11, SI12, SI13, SI15, SI17	$1.38 \times 10^{-11}$
SI with PCD (n=6)	0.002	427	40	8	GO:0099568	cytoplasmic region	DNAH5, CCDC114, DNAH11, SHROOM2, DNAAF1, FMN2, RP1L1, PCLO	SI03, SI06, SI08, SI11, SI12, SI13, SI15, SI17	$1.88 \times 10^{-13}$
Unaffected controls (n=15)	NS		34						

**Table 2.** Gene set enrichment analysis under a recessive mutation model. LH: left-handed. <sup>a</sup>P-values are corrected for multiple testing using the gSCS method. NS: no significant gene sets. <sup>b</sup>The number of mutated genes present in the GO schema. <sup>c</sup>Uncorrected P-values were calculated using Fisher's exact test. <sup>d</sup>Results were the same after excluding subject SI03.

## Discussion

In this study we aimed to identify rare, highly penetrant genetic mutations that cause SI without PCD, by analysing whole genome sequence data from nine SI subjects without medical histories of PCD, five of whom were left-handed. We additionally included six SI subjects with PCD as positive controls, and fifteen unaffected subjects as negative technical controls.

Likely monogenic causes were identified in all positive controls, i.e. each of the six PCD subjects had one recessively mutated gene (usually a different gene in each subject) that is already known to cause PCD when mutated. The six PCD subjects also confirmed that significant pathway enrichment could be detected on the basis of their mutated gene lists, as multiple gene sets related to ciliary functions were detected. The fifteen unaffected control subjects showed no recessive mutations in genes known to cause PCD or laterality phenotypes, as expected.

Two of the nine non-PCD SI subjects had likely recessive monogenic causes in known PCD genes. This may indicate reduced penetrance of these mutations for PCD-like symptoms, such as lung symptoms, although they can apparently affect *situs* in early development. One non-PCD SI subject, who was right-handed, had a homozygous mutation in a gene already known to cause SI without PCD when mutated, i.e. *PKD1L1*<sup>20</sup>. This gene encodes a component of a calcium channel which is associated with non-motile cilia<sup>72</sup>. Another non-PCD SI subject, who was left-handed and had CHD, had compound heterozygous mutations in *CFAP52*, which is also known as a cause of either heterotaxy or SI without PCD, when mutated<sup>18</sup>. *CFAP52* protein has been proposed to play a role in cilia-related signal transduction processes<sup>73</sup>.

However, the five remaining non-PCD SI subjects had no obvious monogenic basis for their condition, i.e., they did not have likely causative mutations in genes known to cause human laterality disorders as annotated in the Clinvar database, or within a list of known or suspected laterality genes based on literature searching and mouse phenotypes. Our findings are similar to the findings from Li *et al.*<sup>16</sup>, whose exome sequencing study of laterality defects included 11 subjects with SI, of which 8 had no likely candidate mutations (see Introduction).

Furthermore, gene set enrichment analysis of the mutated genes in non-PCD SI subjects did not identify significant biological pathways, in either the whole set, or the left-handed subset, or the subset that was 'unsolved' by recessive monogenic causes. In other words, the biology of their non-PCD SI could not be linked via the mutations that they carried. Finally, we also considered larger genomic rearrangements known as Copy Number Variants (CNVs), but no obvious candidate genes emerged. A monogenic model is still possible for the majority of the non-PCD SI cases, but would have to involve genetic heterogeneity across a set of genes which are not currently linked in terms of their known biology in the Gene Ontology, at least to an extent that was discernible in this dataset, as it was for the PCD subjects.

Among the five 'unsolved' non-PCD SI subjects, three were left-handed, and therefore comprised the majority of left-handed non-PCD SI subjects in this study. This meant that we did not identify a genetic-developmental pathway that links handedness and visceral asymmetry in this study. The question of whether, and how, handedness is linked developmentally to visceral laterality in humans remains unanswered. As noted in the Introduction, the existing literature remains ambiguous, and on balance does not support overt links<sup>12,22,25–30</sup>. It is possible that continuous measures of lateralized hand performance, such as pegboard measures or hand grip strength, might reveal stronger links with genes involved in visceral laterality<sup>25</sup>. However, we had no such measures available in the present study.

Some studies based on small samples have suggested that aspects of brain structural asymmetry can be altered in SI<sup>74–77</sup>. Indeed, this was previously assessed in the same SI subjects and controls as analyzed in the present

study<sup>29,78</sup> - briefly, *planum temporale* and arcuate fasciculus asymmetries were not altered in SI, but the overall cerebral torque showed evidence of left-right reversal in non-PCD SI. At that time there was no genetic information available for these individuals. In light of the genetic data in the present study, the group of nine non-PCD SI individuals has now fragmented into two with likely low-penetrance PCD, two with likely causative mutations in non-PCD-causing genes, and five genetically unsolved. We do not attempt here to do group-based statistics with quantitative measures of brain structural asymmetry in these very small groups. However, this should be a line of research as larger datasets become available.

In the present study we focused on rare, protein-coding mutations, but genetic contributions to non-PCD SI might also involve non-coding variation, or rare combinations of multiple common variants. The noncoding genome comprises 98% of the genome, but 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<sup>79</sup>, and population genomic data<sup>80–85</sup>. However, these ranking approaches are currently not very concordant with each other<sup>86</sup>. For the present study we did not address these possibilities, which must await larger sample sizes and an improved understanding of the role of rare, non-coding mutations in phenotypic variation.

*In utero* environmental effects such as prenatal drug exposure might also affect left-right determination<sup>87</sup>. As regards handedness, this is known to associate with various early life factors including birthweight and breastfeeding, although not to a degree which is remotely predictive at the individual level<sup>88</sup>. As noted in the introduction, left-handedness has a heritability of roughly 25% in family and twin studies, and lower in SNP-based heritability studies. Regardless, the primary causes of this trait remain unknown.

In this study there was a degree of diagnostic uncertainty as regards the PCD status of some SI subjects. However, it was not the purpose of the present study to achieve a clinical diagnosis of the presence or absence of PCD, nor to confirm already-known PCD genes. In this context we did not, for example, sequence the mutations in the PCD subjects by another technique for validation, nor confirm allelic phase in the compound heterozygotes. Rather, we were concerned to identify potentially causative mutations in the nine SI subjects without medical histories of PCD, who also showed an elevated rate of left-handedness, with the potential to yield important basic insights into body and brain laterality. If we had found clear candidate mutations in genes not previously linked to PCD or SI, then further validation and functional characterization of those mutations would have been appropriate, but this did not arise.

Regardless of the PCD status of any individual SI subject in this study, the overall pattern of results was clearly different between the PCD and non-PCD groups, where all six positive control subjects in the PCD group had mutations in known PCD-causing genes, while five out of nine in the non-PCD group had no obvious monogenic mutations. Also, the pathway analyses in various different subsets of the non-PCD SI group showed a consistent lack of significant enrichment, whereas clear signals emerged from the PCD group, which further supports an overall distinction of the groups. Nonetheless, further detailed investigation of subjects SI12 and SI16 with PCD diagnostic tools might reveal lung and other ciliary symptoms to an extent, as these two subjects had likely causative mutations for their SI in known PCD-causing genes<sup>89</sup>.

Although ciliary-induced nodal flow plays a crucial role in symmetry breaking, some organisms, such as chicks and fruit flies (*Drosophila melanogaster*), do not have motile nodal cilia, yet they do show asymmetrical organs<sup>90</sup>. For example, left-right asymmetry in chicks involves rearranging the relative orientations of cells expressing critical genes at the node, in a non-ciliary-dependent manner<sup>91</sup>. Furthermore, left-right asymmetry of the *Drosophila* gut is established by intracellular cytoskeletal organization that may give rise to cellular shape chirality, by means of unidirectional tilting of radial fibers, and anti-clockwise swirling of transverse fibers<sup>92</sup>. Similarly, the actin-related *Lsdia1* gene affects shell coiling direction in the water snail<sup>93</sup>. Whether such mechanisms also influence left-right organ asymmetry in mammals is unclear. In the present study we saw no mutations in the orthologues of two genes implicated in cellular chirality in *Drosophila*, *MYO1D* or *MYO1C*<sup>94</sup>, nor in the orthologues of two genes thought to affect laterality in chicks and frogs, *SLC6A4* and *SLC18A2*<sup>95</sup>. We were not able to identify a human homologue of *Lsdia1*.

Future studies in larger human cohorts may help to identify more genetic contributions to non-PCD SI, and clarify the extent to which developmental mechanisms might be shared with brain asymmetry.

## Data availability

Requests to access the genomic datasets generated for the current study will be considered in relation to the consents, relevant rules and regulations, and can be made via the corresponding author.

Received: 11 February 2019; Accepted: 13 February 2020;

Published online: 28 February 2020

## References

1. Deng, H., Xia, H. & Deng, S. Genetic basis of human left-right asymmetry disorders. *Expert. Rev. Mol. Med.* **16**, e19 (2015).
2. Arnold, G. L., Bixler, D. & Girod, D. Probable autosomal recessive inheritance of polysplenia, situs inversus and cardiac defects in an Amish family. *Am. J. Med. Genet.* **16**, 35–42 (1983).
3. Mirvis, M., Stearns, T. & James Nelson, W. Cilium structure, assembly, and disassembly regulated by the cytoskeleton. *Biochemical J.* **475**, 2329–2353 (2018).
4. Fliegauf, M., Benzing, T. & Omran, H. When cilia go bad: cilia defects and ciliopathies. *Nat. reviews. Mol. Cell Biol.* **8**, 880–893 (2007).
5. Reiter, J. F. & Leroux, M. R. Genes and molecular pathways underpinning ciliopathies. *Nat. reviews. Mol. Cell Biol.* **18**, 533–547 (2017).
6. Okada, Y., Takeda, S., Tanaka, Y., Belmonte, J. I. & Hirokawa, N. Mechanism of nodal flow: a conserved symmetry breaking event in left-right axis determination. *Cell* **121**, 633–644 (2005).

7. Cartwright, J. H., Piro, O. & Tuval, I. Fluid-dynamical basis of the embryonic development of left-right asymmetry in vertebrates. *Proc. Natl Acad. Sci. USA* **101**, 7234–7239 (2004).
8. Hilfinger, A. & Julicher, F. The chirality of ciliary beats. *Phys. Biol.* **5**, 016003 (2008).
9. Leigh, M. W. *et al.* Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. *Genet. Med.* **11**, 473–487 (2009).
10. Layton, W. M. Jr. Random determination of a developmental process: reversal of normal visceral asymmetry in the mouse. *J. heredity* **67**, 336–338 (1976).
11. Brueckner, M., D'Eustachio, P. & Horwitz, A. L. Linkage mapping of a mouse gene, iv, that controls left-right asymmetry of the heart and viscera. *Proc. Natl Acad. Sci. USA* **86**, 5035–5038 (1989).
12. Torgersen. Situs inversus, Asymmetry, and Twinning. (1950).
13. Cavalli-Sforza, L. L. & Bodmer, W. F. *The genetics of human populations*. (W. H. Freeman 1971).
14. Newcombe, H. B. In M. Fishbein, ed. *Papers and Discussions of the Second International Conference on Congenital Malformations*. New York: International Medical Congress. (1964).
15. Ware, S. M. *et al.* Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *Am. J. Hum. Genet.* **74**, 93–105 (2004).
16. Li, A. H. *et al.* Genetic architecture of laterality defects revealed by whole exome sequencing. *European journal of human genetics: EJHG* (2019).
17. Gur, M. *et al.* Roles of the cilium-associated gene CCDC11 in left-right patterning and in laterality disorders in humans. *Int. J. developmental Biol.* **61**, 267–276 (2017).
18. Ta-Shma, A. *et al.* A human laterality disorder associated with a homozygous WDR16 deletion. *Eur. J. Hum. Genet.* **23**, 1262–1265 (2015).
19. Reish, O. *et al.* A Homozygous Nme7 Mutation Is Associated with Situs Inversus Totalis. *Hum. Mutat.* **37**, 727–731 (2016).
20. Vetrini, F. *et al.* Bi-allelic Mutations in PKDIL1 Are Associated with Laterality Defects in Humans. *Am. J. Hum. Genet.* **99**, 886–893 (2016).
21. Francks, C. Exploring human brain lateralization with molecular genetics and genomics. *Ann. N. Y. Acad. Sci.* **1359**, 1–13 (2015).
22. de Kovel, C. G. F., Lisgo, S. N., Fisher, S. E. & Francks, C. Subtle left-right asymmetry of gene expression profiles in embryonic and foetal human brains. *Sci. Rep.* **8**, 12606 (2018).
23. de Kovel, C. G. F. *et al.* Left-Right Asymmetry of Maturation Rates in Human Embryonic Neural Development. *Biol. Psychiatry* **82**, 204–212 (2017).
24. Barth, K. A. *et al.* fsi zebrafish show concordant reversal of laterality of viscera, neuroanatomy, and a subset of behavioral responses. *Curr. Biol.* **15**, 844–850 (2005).
25. Bandler, W. M. *et al.* Common variants in left/right asymmetry genes and pathways are associated with relative hand skill. *PLoS Genet.* **9**, e1003751 (2013).
26. de Kovel, C. G. F. & Francks, C. The molecular genetics of hand preference revisited. *Sci. Rep.* **9**, 5986 (2019).
27. McManus, I. C., Martin, N., Stublings, G. F., Chung, E. M. & Mitchison, H. M. Handedness and situs inversus in primary ciliary dyskinesia. *Proc. Biol. Sci.* **271**, 2579–2582 (2004).
28. Afzelius, B. A. & Stenram, U. Prevalence and genetics of immotile-cilia syndrome and left-handedness. *Int. J. developmental Biol.* **50**, 571–573 (2006).
29. Vingerhoets, G. *et al.* Brain structural and functional asymmetry in human situs inversus totalis. *Brain Struct Funct* (2018).
30. Cockayne. the genetics of transposition of the viscera. (1938).
31. Medland, S. E. *et al.* Genetic influences on handedness: data from 25,732 Australian and Dutch twin families. *Neuropsychologia* **47**, 330–337 (2009).
32. Neale, B. Heritability of > 2,000 traits and disorders in the UK Biobank. (2017).
33. Armour, J. A., Davison, A. & McManus, I. C. Genome-wide association study of handedness excludes simple genetic models. *Heredity* **112**, 221–225 (2014).
34. Kavaklıoglu, T., Ajmal, M., Hameed, A. & Francks, C. Whole exome sequencing for handedness in a large and highly consanguineous family. *Neuropsychologia* **93**, 342–349 (2016).
35. Behan, L. *et al.* PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. *Eur. respiratory J.* **47**, 1103–1112 (2016).
36. Kosaki, K. & Casey, B. Genetics of human left-right axis malformations. *Semin. Cell developmental Biol.* **9**, 89–99 (1998).
37. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
38. Robinson, J. Edinburgh Handedness Inventory. In: Volkmar F.R. (eds) *Encyclopedia of Autism Spectrum Disorders*, Springer, New York, NY, (2013).
39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma.* **25**, 1754–1760 (2009).
40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma.* **25**, 2078–2079 (2009).
41. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
42. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
43. Auton, A. *et al.* A global reference for human genetic variation. *Nat.* **526**, 68–74 (2015).
44. Tan, A., Abecasis, G. R. & Kang, H. M. Unified representation of genetic variants. *Bioinforma.* **31**, 2202–2204 (2015).
45. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids Res.* **38**, e164 (2010).
46. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
47. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. methods* **7**, 248–249 (2010).
48. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
49. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nat.* **536**, 285–291 (2016).
50. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nat.* **493**, 216–220 (2013).
51. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
52. Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic acids Res.* **44**, W83–89 (2016).
53. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
54. Ashburner, M. *et al.* Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids Res.* **45**, D331–d338 (2017).
55. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids Res.* **46**, D1062–d1067 (2018).
56. Carrion-Castillo, A. *et al.* Genetic effects on planum temporale asymmetry and their limited relevance to neurodevelopmental disorders, intelligence or educational attainment. *Cortex* **124**, 137–153, <https://doi.org/10.1016/j.cortex.2019.11.006> (2019).
57. Smith, C. L., Blake, J. A., Kadın, J. A., Richardson, J. E. & Bult, C. J. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic acids Res.* **46**, D836–d842 (2018).
58. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).

59. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
60. Pirooznia, M., Goes, F. S. & Zandi, P. P. Whole-genome CNV analysis: advances in computational approaches. *Front. Genet.* **6**, 138 (2015).
61. Antaki, D., Brandler, W. M. & Sebat, J. SV2: Accurate Structural Variation Genotyping and De Novo Mutation Detection from Whole Genomes. *Bioinformatics* (2017).
62. Hornef, N. et al. DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am. J. respiratory Crit. care Med.* **174**, 120–126 (2006).
63. Ranjzad, F. et al. Identification of Three Novel Frameshift Mutations in the PKD1 Gene in Iranian Families with Autosomal Dominant Polycystic Kidney Disease Using Efficient Targeted Next-Generation Sequencing. *Kidney blood Press. Res.* **43**, 471–478 (2018).
64. Xu, D. et al. Novel Mutations in the PKD1 and PKD2 Genes of Chinese Patients with Autosomal Dominant Polycystic Kidney Disease. *Kidney blood Press. Res.* **43**, 297–309 (2018).
65. Hirokawa, N., Tanaka, Y. & Okada, Y. Left-right determination: involvement of molecular motor KIF3, cilia, and nodal flow. *Cold Spring Harb. Perspect. Biol.* **1**, a000802 (2009).
66. Nonaka, S. et al. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**, 829–837 (1998).
67. Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nat. Rev. Genet.* **11**, 331–344 (2010).
68. Verhey, K. J., Dishinger, J. & Kee, H. L. Kinesin motors and primary cilia. *Biochemical Soc. Trans.* **39**, 1120–1125 (2011).
69. Horani, A. et al. LRRC6 mutation causes primary ciliary dyskinesia with dynein arm defects. *PLoS One* **8**, e59436 (2013).
70. Nardello, R. et al. A novel mutation of WDR62 gene associated with severe phenotype including infantile spasm, microcephaly, and intellectual disability. *Brain Dev.* **40**, 58–643 (2018).
71. Sironen, A. et al. Loss of SPEF2 function in mice results in spermatogenesis defects and primary ciliary dyskinesia. *Biol. Reprod.* **85**, 690–701 (2011).
72. Masyuk, A. I., Gradilone, S. A. & LaRusso, N. F. Calcium signaling in cilia and ciliary-mediated intracellular calcium signaling: are they independent or coordinated molecular events. *Hepatology* **60**, 1783–1785 (2014).
73. Hirschner, W. et al. Biosynthesis of Wdr16, a marker protein for kinocilia-bearing cells, starts at the time of kinocilia formation in rat, and wdr16 gene knockdown causes hydrocephalus in zebrafish. *J. Neurochem.* **101**, 274–288 (2007).
74. Tubbs, R. S., Wellons, J. C., 3rd, Salter, G., Blount, J. P. & Oakes, W. J. Intracranial anatomic asymmetry in situ inversus totalis. *Anat. Embryol.* **206**, 199–202 (2003).
75. Kennedy, D. N. et al. Structural and functional brain asymmetries in human situs inversus totalis. *Neurology* **53**, 1260–1265 (1999).
76. Ihara, A. et al. Neuroimaging study on brain asymmetries in situ inversus totalis. *J. Neurol. Sci.* **288**, 72–78 (2010).
77. Schuler, A. L. et al. Mens inversus in corpore inverso? Language lateralization in a boy with situs inversus totalis. *Brain Lang.* **174**, 9–15 (2017).
78. Mannaert, L., Verhelst, H., Gerrits, R., Bogaert, S. & Vingerhoets, G. White matter asymmetries in human situs inversus totalis. *Brain Struct. Funct.* **224**, 2559–2565 (2019).
79. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nat.* **489**, 57–74 (2012).
80. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
81. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinforma.* **31**, 761–763 (2015).
82. Ritchie, G. R., Dunham, I., Zeggini, E. & Flück, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **11**, 294–296 (2014).
83. Jagadeesh, K. A. et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
84. Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
85. Ioannidis, N. M. et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
86. Eilbeck, K., Quinlan, A. & Yandell, M. Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.* (2017).
87. van Veenendaal, N. R., Kusters, C. D., Oostra, R. J., Bergman, J. E. & Cobben, J. M. When the right (Drug) should be left: Prenatal drug exposure and heterotaxy syndrome. *Birth defects research. Part A, Clin. Mol. teratology* **106**, 573–579 (2016).
88. de Kovel, C. G. F., Carrion-Castillo, A. & Francks, C. A large-scale population study of early life factors influencing left-handedness. *Sci. Rep.* **9**, 584 (2019).
89. Shapiro, A. J. et al. Diagnosis of Primary Ciliary Dyskinesia. An Official American Thoracic Society Clinical Practice Guideline. *Am. J. respiratory Crit. care Med.* **197**, e24–e39 (2018).
90. Vandenberg, L. N. & Levin, M. A unified model for left-right asymmetry? Comparison and synthesis of molecular models of embryonic laterality. *Developmental Biol.* **379**, 1–15 (2013).
91. Gros, J., Feistel, K., Viebahn, C., Blum, M. & Tabin, C. J. Cell movements at Hensen's node establish left/right asymmetric gene expression in the chick. *Sci.* **324**, 941–944 (2009).
92. Tee, Y. H. et al. Cellular chirality arising from the self-organization of the actin cytoskeleton. *Nat. Cell Biol.* **17**, 445–457 (2015).
93. Abe, M. & Kuroda, R. The development of CRISPR for a mollusc establishes the formin Lsdia1 as the long-sought gene for snail dextral/sinistral coiling. *Development (Cambridge, England)* **146** (2019).
94. Lebreton, G. et al. Molecular to organismal chirality is induced by the conserved myosin 1D. *Sci.* **362**, 949–952 (2018).
95. Fukumoto, T., Blakely, R. & Levin, M. Serotonin transporter function is an early step in left-right patterning in chick and frog embryos. *Developmental Neurosci.* **27**, 349–363 (2005).

## Acknowledgements

Thanks to all of the study participants. This research was funded by the Max Planck Society (Germany) and the Netherlands Organization for Scientific Research (NWO, Grant No. 054-15-101) as part of the FLAG-ERA consortium project ‘MULTI-LATERAL’, a Partner Project to the European Union’s Flagship Human Brain Project. The study was also funded by the Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO-Grant No. G.0114.16 N) assigned to GV.

## Author contributions

G.V. recruited and characterized the subjects. G.V., S.E.F. and C.F. conceived the study. M.C.P performed the genetic data analysis with support from A.C.C. M.C.P. and C.F. wrote the manuscript. G.V., A.C.C. and S.E.F. edited the manuscript. C.F. directed the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-60589-z>.

**Correspondence** and requests for materials should be addressed to C.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020