OPEN ACCESS

Original research

Linkage analysis identifies an isolated strabismus locus at 14q12 overlapping with FOXG1 syndrome region

Xin (Cynthia) Ye,[1,2,3] Nicole M Roslin,[4] Andrew D Paterson,[4,5,6] Christopher J Lyons,[3,7] Victor Pegado,[7] Phillip Richmond,[1,3] Casper Shyr,[1,3] Oriol Fornes,[1,3] XiaoHua Han,[1] Michelle Higginson,[1] Colin J Ross,[8] Deborah Giaschi,[3,7] Cheryl Gregory-Evans,[7] Millan S Patel,[2,3] Wyeth W Wasserman [1,2,3]

For numbered affiliations see end of article.

Correspondence to
Professor Wyeth W Wasserman, Medical Genetics, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada; wyeth@cmmt.ubc.ca
Dr Millan S Patel; mpatel@cw.bc.ca

Received 2 June 2020
Revised 5 September 2020
Accepted 29 September 2020

ABSTRACT

Strabismus is a common condition, affecting 1%–4% of individuals. Isolated strabismus has been studied in families with Mendelian inheritance patterns. Despite the identification of multiple loci via linkage analyses, no specific genes have been identified from these studies. The current study is based on a seven-generation family with isolated strabismus inherited in an autosomal dominant manner. A total of 13 individuals from a common ancestor have been included for linkage analysis. Among these, nine are affected and four are unaffected. A single linkage signal has been identified at an 8.5 Mb region of chromosome 14q12 with a multipoint LOD (logarithm of the odds) score of 4.69. Disruption of this locus is known to cause FOXG1 syndrome (or congenital Rett syndrome; OMIM #613454 and *164874), in which 84% of affected individuals present with strabismus. With the incorporation of next-generation sequencing and in-depth bioinformatic analyses, a 4 bp non-coding deletion was prioritised as the top candidate for the observed strabismus phenotype. The deletion is predicted to disrupt regulation of *FOXG1*, which encodes a transcription factor of the Forkhead family. Suggestive of an autoregulation effect, the disrupted sequence matches the consensus FOXG1 and Forkhead family transcription factor binding site and has been observed in previous ChIP-seq studies to be bound by Foxg1 in early mouse brain development. Future study of this specific deletion may shed light on the regulation of *FOXG1* expression and may enhance our understanding of the mechanisms contributing to strabismus and FOXG1 syndrome.

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Ye X(Cynthia), Roslin NM, Paterson AD, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2020-107226

INTRODUCTION

Strabismus, also known as crossed eyes or squint, affects 1%–4% of individuals. Diagnosis and treatments for strabismus are well established, but the pathophysiology for most isolated strabismus remains largely unknown. Disturbances anywhere along the visual sensory or the oculomotor pathways can be postulated to lead to eye deviation.[1]

As early as Hippocrates' time, strabismus was recognised as a genetic disorder based on an observation of its tendency to cluster within families.[2] During the last century, twin and family studies have demonstrated a substantial genetic contribution to strabismus, and both autosomal dominant and autosomal recessive transmission patterns have been reported.[3] Recently, a genome-wide association study reported two variants with small effect sizes.[4] Strabismus occurs commonly as one phenotype among many in syndromes, such as congenital Rett syndrome (FOXG1 syndrome) and Joubert syndrome, in which 84% and 75% of individuals display a strabismus phenotype, respectively.[5 6] On the other hand, families displaying isolated strabismus transmitting in simple Mendelian patterns are uncommon. In a few such families, genetic loci on chromosomes 4, 6, 7, 12, 16 and 19 have been identified for isolated strabismus, but no causal gene has been identified in these regions.[3] Not overlapping the reported loci, 11 genes (*PHOX2A*, *ROBO3*, *KIF21A*, *SALL1*, *TUBB3*, *HOXB1*, *SALL4*, *CHN1*, *HOXA1*, *TUBB2B*, *MAFB*),[7] of which six encode transcription factors (underlined), have been identified for a subgroup of strabismus associated with congenital cranial dysinnervation disorders,[7] but the genetic aetiology of other strabismus subtypes remains elusive. Identification of a locus with high confidence, determination of a causal gene and detailed mechanistic insights at the nucleotide level would provide new insights into the molecular mechanisms of certain forms of strabismus.

We compiled a large, seven-generation, non-consanguineous pedigree with 22 individuals exhibiting isolated strabismus, consistent with an autosomal dominant inheritance pattern. Through genome-wide linkage analysis of eight affected and four unaffected individuals from one branch and one affected individual from a separate branch, we mapped this familial strabismus to chromosome 14q12, which overlaps with the FOXG1 syndrome locus. Next-generation sequencing and in-depth analysis identified a strong candidate deletion variant within this locus. The data support that the deletion localises to a FOXG1 transcription factor binding site (TFBS), suggesting an autoregulatory loop may be disrupted that controls *FOXG1* expression in early brain development.

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

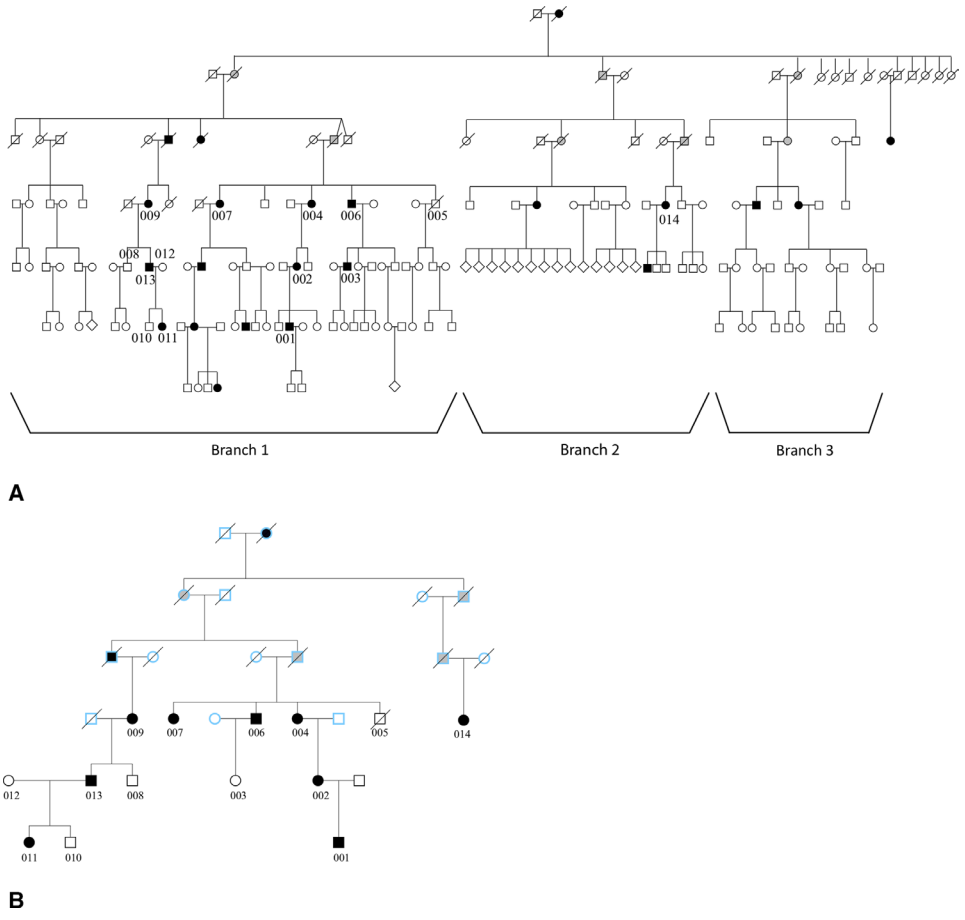


Figure 1 Pedigree for the subject family with isolated strabismus. (A) The pedigree represents a seven-generation family with 176 individuals, including deceased individuals. Three major branches are identified: 12 study participants come from branch 1 and one (014) comes from branch 2. Individual 012 was not included in the linkage analysis. Black represents affected individuals, white represents unaffected individuals and grey represents obligate carriers. (B) Simplified branch 1 of the subject family showing the genotyped individuals (with study ID) and ancestors required to link them. Individual 014 represents branch 2, and all the other individuals come from branch 1. Individuals whose status was not confirmed clinically were coded as unknown for the linkage analysis. These individuals were indicated by blue outlines.

RESULTS

Pedigree and participant profile

A seven-generation pedigree of European origin, with 176 individuals, including those who were deceased, was compiled. Individuals were labelled as strabismic either through medical records or strong family anecdotes. A roughly even distribution of strabismus cases was observed between females (13 individuals) and males (9 individuals) and the disorder was transmitted both maternally and paternally. Among the three extensively traced branches (figure 1A), branch 1 was most well documented and contained most of the participants in this study. In branch 1, strabismus was reported across four consecutive generations. An autosomal dominant inheritance model with high penetrance best matched the qualitative observation.

To characterise the strabismus in this family, eight of the nine affected descendants indicated in figure 1B were seen by one of three participating ophthalmologists specialised in strabismus. All individuals underwent complete ophthalmic examination with attention to ocular motility both before and after pharmacological cycloplegia. The specific characteristics of strabismus were not uniform across the descendants in the family (table 1). The strabismus phenotypes could be grouped into two broad directional categories: esotropia (an eye turns in) and hypertropia (an eye turns up). Both esotropia and hypertropia were noted in individual 014, but this individual had undergone multiple corrective surgeries and childhood medical records were not available. On the other hand, individuals 011, 013 and 009 are in consecutive generations and had no history of extraocular muscle surgery. Individual 011 presented with esotropia while the other two displayed hypertropia (009, 013). The unaffected status was confirmed by medical history provided by the participating family members. In the clinical ophthalmology examinations and in oral reports from subjects, there were no other phenotypes observed broadly in the individuals with strabismus.

Linkage analysis and haplotype analysis

Initial linkage analysis was performed with 12 family members; 8 affected and 4 unaffected individuals (excluding 012 and 014 from figure 1B). Simulations under the alternative hypothesis (linkage) generated a maximum LOD score of 3.56, under an autosomal dominant model with disease allele frequency q=0.005, 99% penetrance and 0.2% phenocopy rate. The LOD score curves did not change significantly across a range of disease allele frequency settings (results not shown).

With the same linkage analysis parameters as in the simulation, the largest LOD score based on the observed genotypes was 3.55, close to the simulated maximum. This chromosome 14 locus was the only region with a LOD score higher than 3, and exceeds the standard threshold for genome-wide significance

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

Table 1 Ophthalmological characteristics of strabismic participants in the subject family

Identifier	Age at examination (years)	Type of reported strabismus	Reported age of onset	Eye movement full	Prism cover test measurements	Visual acuity and refractive status	Strabismus surgery	Other ocular conditions
001*	24	Accommodative left esotropia with an A pattern	9 months	Y	Without correction, there was a 25 PD left esotropia. With correction, there was a 20 PD left esotropia in the primary position. This increased to 30 PD in upgaze and decreased to 10 PD in downgaze	Visual acuity was 20/20 in the right eye and 20/20–3 in the left eye with correction. Wearing: +2.50+3.25×20° OD, +2.25+3.00×135° OS	Bilateral strabismus surgery at age 2 and at age 3; botulinum toxin injection at the age of 4	
002*	60	Hypotropia and left exotropia. Consecutive exodeviation after initial esotropia surgery	Unclear	Y	20 PD left exotropia. 15 PD of left hypotropia in the primary position	Visual acuity was 20/20–2 OD, 20/20–1 OS with correction. Wearing: −3.50+0.50×50° OD, −4.50+0.75×90° OS	Strabismus surgery at ages 3, 8 and 16	
004*	82	Right esotropia	3 years	Y	12 PD comitant esotropia	Visual acuity was 20/200 right, 20/20 left with correction: −1.00+1.75×55° OD, −1.00+2.00×80° OS	Strabismus surgery at ages 7 and 10	Bilateral senile cataract, right dense amblyopia
007	83	Hypertropia	Unclear	Y	Poor fixation from reduced visual acuity in the left eye due to age-related macular degeneration precluded cover testing	Visual acuity was 20/30 OD, 20/800 OS with correction. Wearing: −1.50+2.00×100° OD, −0.50 OS	Yes. But no time specific information provided.	Macular degeneration left eye, macular drusen right eye. Bilateral senile cataract
009*	76	Right superior oblique muscle palsy; excyclotorsion	After age of 2, exact onset time unclear, aware of ocular misalignment at the age of 9	Y	8 PD right hypertropia which increased on right head tilt and decreased on left head tilt	Visual acuity: 20/20 right and left corrected: −0.50+1.00×170° OD, −1.25+1.25×35° OS	No	Presbyopia
011*	14	Left esotropia	Before the age of 2	Y	15 PD left esotropia in the primary position while wearing her glasses. Looking through the top part of glasses at near, her deviation increased to 30 PD. However, when used the near add, reducing the need to accommodate, left esotropia decreased to 15 PD	Visual acuity was 20/20 in each eye with correction. Wearing: −3.25+0.25×0° OD, −3.25 OS. She had an add of +1.50 sphere in each spectacle lens	No	Myopia
013*	45	Superior oblique palsy	Unclear, aware of ocular misalignment at the age of 6	Y	15 PD right hypertropia in the primary position, increasing to 20 PD on right head tilt, decreasing to 12 PD on left head tilt	Visual acuity: 20/20–2 in each eye with soft contact lens correction	No	Myopia
014*	72	Esotropia, right hypertropia; excyclotorsion in both eyes (a history suggests infantile esotropia)	Unclear, patient was informed the strabismus was present since birth	Y	25 PD esotropia SC at distance/8 PD right hypertropia SC at distance/20 PD esotropia SC at near/4 PD right hypertropia SC at near/ Titmus testing results were –/1, –/3 and –/9	Visual acuity was 20/25 OD and 20/25 OS	2 strabismus surgeries during 50 s	Senile cataract, latent nystagmus, left eye suppression (self-reported diplopia, but able to ignore it)

006 did not participate in the clinical study, and no supporting clinical data were available.
*The type of reported strabismus in individuals was confirmed by examination.
PD, prism diopter.

of 3.3.[8] This linked region spanned approximately 10 Mb and was bounded by rs7146411 and rs1951187, corresponding to chr14:22,779,843–32,908,192 (hg19). Chromosome 14q12 is therefore a novel locus for isolated strabismus.

Individual 014 representing branch 2 in the pedigree was recruited to the study after the genome-wide genotyping was performed. An expanded linkage analysis with corresponding SNPs extracted from 014's whole genome sequencing (WGS) data further supported the linkage to chromosome 14. Simulations under the alternative hypothesis (linkage) generated a maximum LOD score of 4.69 for the same region, under an autosomal dominant model with disease allele frequency q=0.005, 99% penetrance and 0.2% phenocopy rate. We observed a LOD score of 4.69, and the region remained as the sole candidate (figure 2A). In addition, we performed non-parametric linkage analyses and identified the same linkage region (figure 2B).

Haplotype analysis was used to complement linkage analysis by providing visual confirmation of the statistical testing. An approximately 8.5 Mb region (chr14:22,779,843–31,289,720) was shared between the nine affected descendants of the common ancestor. An unaffected descendant (005, subsequently deceased) shared a 5.5 Mb region within the linked region (figure 3A,B). Thus, an approximately 3 Mb region was shared exclusively by nine affected descendants, corresponding to chr14:28,467,136–31,289,720. The LOD score was above 4.60 for chr14:28,467,136–30,045,978 (online supplemental figure 1). The 3 Mb region situates within a gene poor region (figure 3B).

No impactful coding variant in the 10 Mb locus identified through WES and WGS

Whole exome sequencing (WES) showed that two affected third-degree cousins (001 and 011) shared 119 heterozygous non-synonymous variants across the entire exome. A subset of 60 variants among the 119 had a frequency lower than 1% in an

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

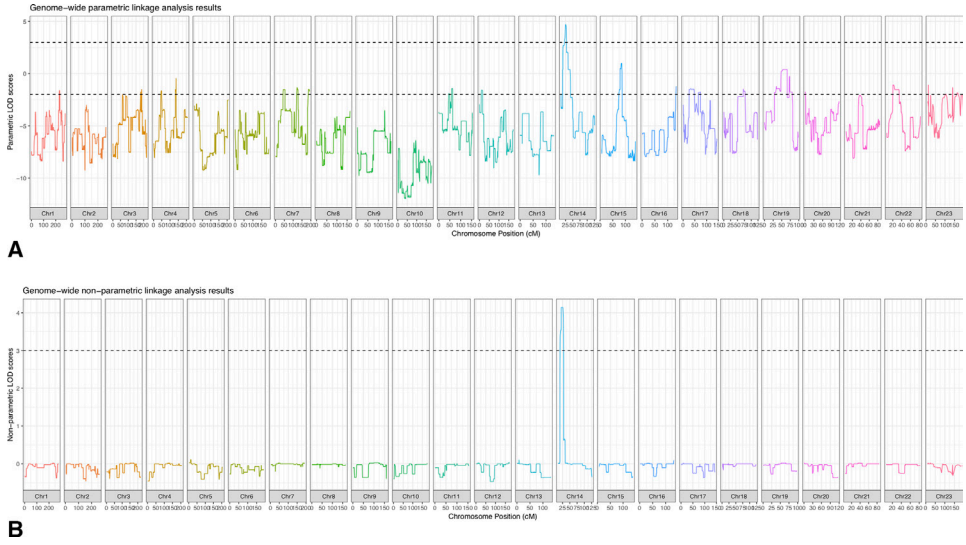


Figure 2 Linkage analysis for subject family. (A) Parametric analysis. An expanded linkage analysis was performed in all 13 individuals who shared the common ancestor. We observed a LOD score of 4.69 for the linkage region in chromosome 14. The top dashed line indicates a LOD of 3 and the bottom dashed line indicates a LOD of −2. (B) Non-parametric analysis. We performed non-parametric analyses and obtained the same linkage region on chromosome 14. The dashed line indicates a LOD of 3.

in-house database of a rare-disease-WES project.[9] Only one of the variants (chr14:31,061,628 A>G, rs145527124) was located within the 10 Mb locus, falling within exon 5 of *G2E3* (G2/M-phase specific E3 ubiquitin protein ligase) (NM_017769.4).

This variant leads to Ile113Val alteration in ENST00000206595. In gnomAD V.2.1.1 European (non-Finnish) population, there are two homozygotes for this variant and the overall allele frequency is 0.17% and two homozygotes while gnomAD V.3 has an overall allele frequency of 0.15% and another homozygote.[10] In addition, in the 1000 Genomes project,[11] the allele frequency in Punjabi from Lahore, Pakistan (PJL) is 1.6%. This *G2E3* variant was not supported as a candidate by computational analysis (predicted to be 'tolerated' with SIFT[12] and 'benign' with Polyphen[13]). Qualitative review of the literature did not suggest a potential role for *G2E3* in a strabismus phenotype.

The lack of candidate variants from WES motivated the generation of WGS for individuals 001, 013 and 014, who were selected to represent two sub-branches of branch 1 and branch 2 of the pedigree. The only coding variant detected with a frequency of ≤1% in the 10 Mb region was the aforementioned *G2E3* variant.

WGS and bioinformatics analyses highlight a heterozygous non-coding variant in a regulatory region of *FOXG1*

Our analyses showed that the WGS on 001, 013 and 014 did not capture some low complexity regions, raising concern that the protocol used at the time of generation might fail to detect repetitive sequences and small genomic structural alterations. As current WGS protocols could better detect such properties, we generated WGS for an affected parent–child trio (011, 012 and 013). For simplicity, we report variants from the trio WGS set for comprehensiveness in the following sections.

Within the 3 Mb region, a total of 664 variants were shared by both affected individuals (based on a dominant model of transmission). No CNVs were detected. We focused on variants with a frequency ≤1% and which have been reported in fewer than 10 homozygotes in gnomAD V.2.0, criteria met by 24 of the 664 variants. The only coding variant of these 24 was the *G2E3* coding variant reported previously. (These potentially identifying variants are available from the corresponding authors on request and under an appropriate data handling agreement.)

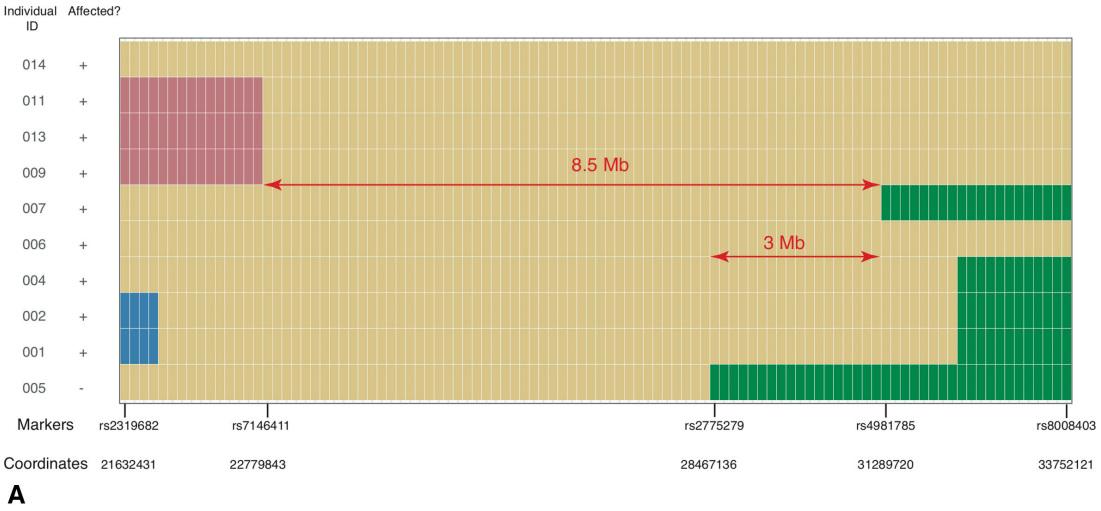
As 23 of 24 candidate variants prioritised in the WGS analysis were non-coding, we used diverse methods to annotate non-coding variants with regulatory information. There is no standard practice to annotate non-coding variants, so we used a variety of approaches to identify those within potential regulatory elements. One variant was noted recurrently using a variety of bioinformatic predictions. Among the 23 candidate variants, only chr14:29247628 TAAAC >T (online supplemental figure 2) has been assigned a CADD score over 20[14] and is ~10 kb 3′ from *FOXG1*. This variant is situated within a potential regulatory region, as suggested by the presence of DNase-seq peaks and the histone marks H3K4me1, H3K4me2, H3K4me3, H3K9ac and H3K27ac, which are associated with promoters/enhancers.[15] This deletion was absent in gnomAD V.2.1.1 and V.3 and dbSNP build 152 and 153.[16] This deletion was confirmed through Sanger sequencing to be present in all nine affected subjects and none of the unaffected individuals.

Since the upper limit of the reported prevalence of strabismus is 4%, additional variants with frequency >1% and ≤4% in gnomAD were also obtained for examination. A total of 54 additional variants were identified in the candidate region. However, none of these variants situated on protein coding regions or were assigned a CADD score over 20.

We examined topologically associating domains (TADs) for the 3 Mb region (figure 4) to suggest potential regulatory relationships between identified variants and nearby genes. The top candidate variant was located within the same TAD as *FOXG1*, and hereafter this TAD will be referred to as the FOXG1-TAD. Both *FOXG1* and the sequence surrounding the candidate variant are highly conserved across vertebrates, with the candidate sequence retained from fish to humans (figure 5).

In human genome annotations, the variant chr14:29247628 TAAAC >T was located within an alternative exon of a long non-coding RNA gene (*LINC01551*). Within the mouse, chicken and zebrafish annotation and supporting data, there were no RNA transcripts containing the variant.[17] As the variant position is conserved back to fish, and the transcript evidence is

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.



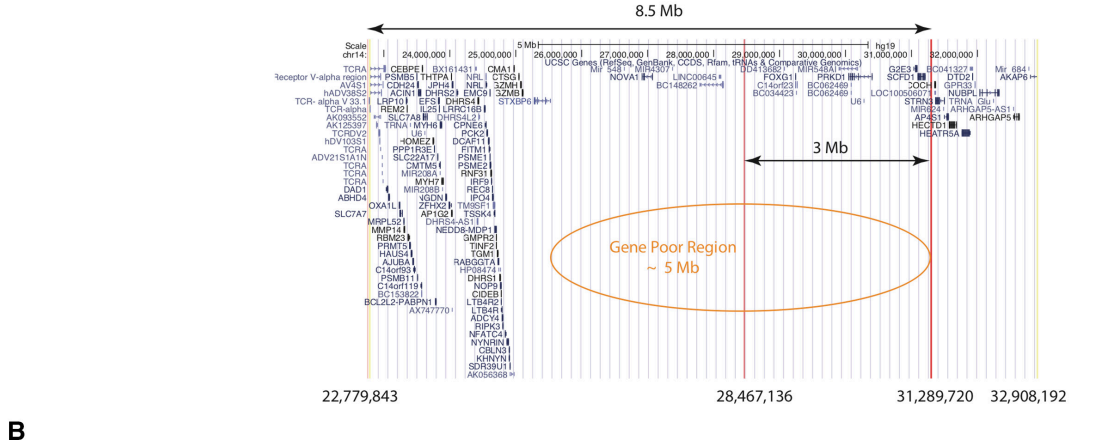


Figure 3 Linkage region. (A) Haplotype analysis for subject family. Each row represents an individual, and each column represents a marker used in the linkage analysis. The markers displayed span chr14:22,779,843–31,289,720 (~8.5 Mb) shared between nine affected individuals. In addition to the nine affected, an unaffected individual is included who shares a 5.5 Mb portion of the region. Thus, an approximately 3 Mb region was shared exclusively by 9 affected descendants, corresponding to chr14:28,467,136–31,289,720 (hg19). Yellow indicates the haplotype inherited from the common ancestor. Each of the other colours indicates a different haplotype from a different ancestor. The 8.5 Mb region and the core shared ~3 Mb region are indicated. For clarity, unaffected individuals not sharing a portion of the region with those affected are not displayed. (B) Genes across the linkage region. The genes reported in the UCSC gene track from the UCSC Genome Browser are displayed for the linkage region. The ~8.5 Mb region and the core ~3 Mb region are indicated. The core ~3 Mb region lies within a gene-poor region.

not supportive of transcription of the highly conserved region in other species, we considered whether the variant might be situated within a *cis*-regulatory region. We examined predicted transcription factor binding motifs overlapping the deletion and observed a match to Forkhead transcription factor binding sites, including JASPAR profile matches for FOXC1, FOXI1 and FOXG1 (figure 6). Proteins in the evolutionarily conserved superfamily of the Forkhead transcription factors share the presence of a DNA binding domain and a transactivation or transrepression effector region, and play a central role during development as well as in the adult.[18] The binding motif for both FOXC1 and FOXI1 profiles would be obliterated by the deletion. Due to two consecutive AAAC repeats, the FOXG1 binding motif is present twice in the reference sequence, with one copy remaining after the deletion. Publicly available mouse ChIP-seq data (GSE96070) showed that Foxg1 binds to this site in cortex tissue from E14–15 brain.[19] Thus, it appears that the deletion is situated within a Forkhead TF binding site, in a highly conserved region with conservation patterns consistent with a functional role in the *cis*-regulation of *FOXG1*.

DISCUSSION

We identified a new locus for isolated strabismus in a family, and this locus overlapped with the locus for FOXG1 syndrome, which has a high prevalence of strabismus. This finding suggests that the strabismus phenotype within FOXG1 syndrome may represent an independent disruptive mechanism from the other phenotypes. We recognised the limitation of the available clinical data from eight individuals, of whom five have a history of strabismus surgery. The nature of the surgery was unknown and since they were studied at different ages, the effect of prolonged adaptation to incomitance would have affected the strabismus at the time of examination, potentially masking the initial pathology. The interpretation of the phenotypic diversity should be re-evaluated if additional information becomes available. In-depth phenotyping of individuals without strabismus

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

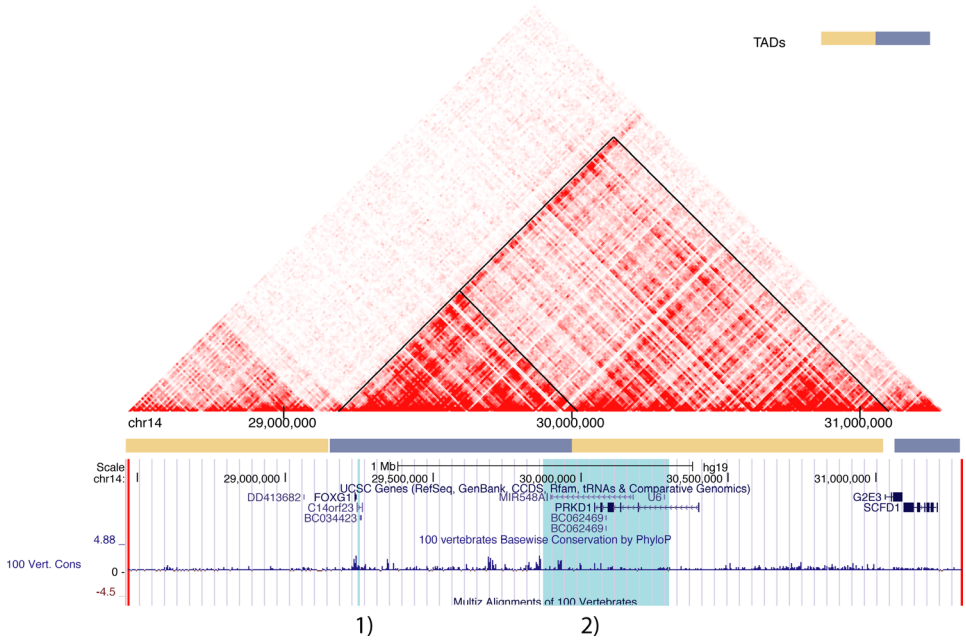


Figure 4 Topologically associating domains within the 3 Mb core region. This heatmap illustrates the chromatin interaction based on Hi-C data.[42] The deeper the red colour, the stronger the intra-chromosomal interaction between corresponding segments of the DNA. FOXG1-TADs are indicated by the black triangle shapes. The blue highlights from left to right correspond to the putative regulatory region within the FOXG1-TAD: (1) chr14:29,247,628 TAAAC >T; (2) the SRO (smallest region of deletion overlap) regulation region affecting FOXG1 expression.[22]

surgery illustrated clinical heterogeneity of strabismus within the family. This suggests that while a specific molecular lesion may lead to strabismus, the specific clinical type is determined by other factors. We examined both coding and non-coding variants, which led to identification of a potential strabismus causing sequence alteration of a Forkhead TFBS within the FOXG1-TAD, suggesting disruption of *cis*-regulation.

To the best of our knowledge, our report contains the largest isolated strabismus pedigree in the literature with the highest LOD score. A LOD score of 4.69 was obtained for a single linkage peak on chromosome 14. Moreover, the segregation of the disease in this family is consistent with autosomal dominant inheritance. A 3 Mb haplotype was shared by all affected participants and was absent from unaffected participants.

In this family, the strabismus types were not uniform. The clinical presentation in the majority of patients at the time of examination did not adhere to a classical neurological pattern, such as cranial nerve IV palsy. The various types of strabismus, however, appear to be caused by the same genetic factor since the linked haplotype is shared by all strabismic individuals. Our observation of strabismus variability in the subject family, review of the literature[20 21] and personal communication with other research groups suggests the current strabismus classification scheme is unhelpful for genetic studies: existing strabismus classification systems may inappropriately be separating individuals sharing a common underlying genetic cause and thus weaken study power. This may explain the paucity of studies detecting and confirming strabismus loci in the literature.[3]

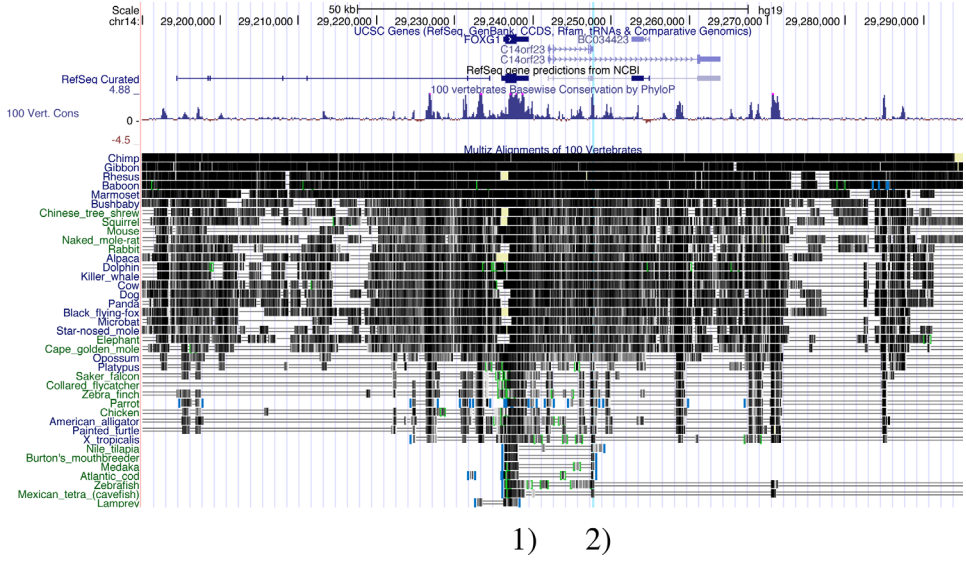


Figure 5 Ultra-conservation regions. (1) FOXG1; (2) the 4 bp deleted region.

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

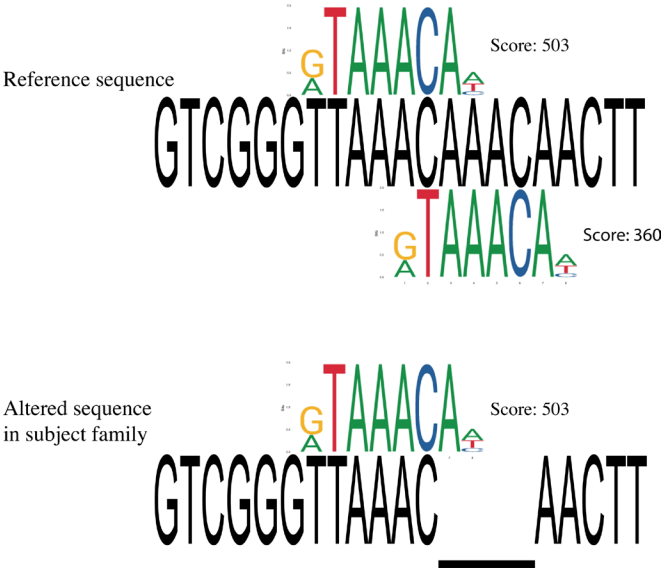


Figure 6 FOXG1 transcription factor binding site matching to reference and alternative sequence. Two FOXG1 TFBS are identified in reference sequence with scores of 503 and 360, respectively. Only one FOXG1 TFBS is identified in sequence with the 4 bp deletion. Scores are based on PWMScan with 'JASPAR CORE 2018 vertebrates' library (Ambrosini G., PWMTools, http://ccg.vital-it.ch/pwmtools).

This 3 Mb region overlaps with microdeletions/microduplications known to cause FOXG1 syndrome in which a high prevalence of strabismus is observed. FOXG1 syndrome, which is also known as congenital Rett syndrome, is a neurological disorder characterised by impaired development and structural brain abnormalities. Strikingly, 84% of affected individuals display strabismus.[5] Distal microdeletions that disrupt the topologically associating domains can lead to FOXG1 syndrome while *FOXG1* remains intact, indicating that mis-regulation of *FOXG1* can cause phenotypic change.[22] Due to the close proximity and shared TAD with *FOXG1*, the 4 bp deletion is speculated to alter *FOXG1* expression (figure 4). A spectrum of partially overlapping phenotypes have been reported in patients with FOXG1 syndrome.[5] The separability of the strabismus phenotype from intellectual disability and other severe disabling phenotypes therefore represents an important insight.

Close examination of diverse data provided important insights into the potential regulatory impact of the 4 bp deletion. First, the sequences surrounding the deletion are highly conserved in the genomes of vertebrates, suggesting that it is under evolutionary selection and that a change may be more likely to have a functional impact. Indeed, the sequence containing the deletion and the coding region of *FOXG1* were the only two highly conserved elements in a 180 kb neighbourhood (figure 5). In addition, this conserved sequence was not supported as being part of a long non-coding RNA in other species (eg, mouse, chicken, frog), implying a *cis*-regulatory effect. Second, the conserved sequence disrupted by the deletion was predicted to be a TFBS for Forkhead transcription factors, including FOXG1, according to the binding site profiles from JASPAR.[23] Third, Foxg1 ChIP-seq data from E14–15 mouse brain (GSE96070) showed that Foxg1 bound to this sequence.

The binding of Foxg1 to this sequence in mouse provides the basis for the hypothesis of disrupted FOXG1 autoregulation leading to strabismus in the subject family. The proposed autoregulatory model is illustrated in figure 7. *FOXG1* is transcribed

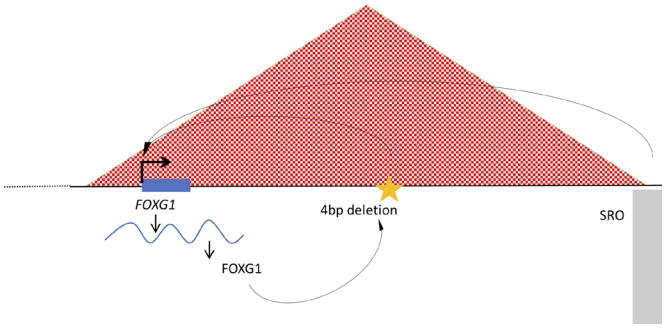


Figure 7 *Cis*-regulatory mechanism within FOXG1-TAD. (1) 4 bp deletion chr14:29,247,628 TAAAC >T; (2) the SRO (smallest region of deletion overlap) regulation region affecting FOXG1 expression.[22]

and translated, the transcription factor binds to the target sequence, helping to maintain the appropriate expression of *FOXG1* during critical developmental period. The disruption of the FOXG1 binding site leads to dysregulation of *FOXG1* expression in a highly specific developmental context that results in the isolated strabismus phenotype.

Autoregulation for key developmental transcription factors in vision is not new to the field. The SIMO regulatory sequence controlling expression of the *PAX6* transcription factor gene is an example of such a distal autoregulatory element.[24] While Pax6 is a crucial transcription factor for delineating the dorsal forebrain in mouse E10.0, Foxg1 is a critical transcription factor for delineating the ventral forebrain in mouse E9.0.[25] Thus, they may share similar sensitivity to regulatory disruption.

FOXG1 expression is strongest in fetal brain and its dysregulation leads to unbalanced development of excitatory and inhibitory synapses in iPSC-derived neurons and in mice.[26] In combination with other transcription factors, Foxg1 in pyramidal neurons is crucial for establishing cortical layers and axon trajectory of callosal projection neurons. Moreover, some Foxg1-directed processes are more vulnerable to dosage changes than others.[27] These observations suggest that Foxg1 has a dosage-sensitive and time-sensitive role in different brain structures. This implies that an alteration in Foxg1 expression pattern can have a very specific impact, and the specific phenotype can be separable from the rest.

In summary, we identified a 3 Mb region on chromosome 14 that is linked to autosomal dominant transmission of isolated strabismus. The region contains *FOXG1*, which has been previously associated with strabismus in 84% of patients with syndromic disruptions. Within the 3 Mb region, the top candidate variant is situated within a FOXG1 transcription factor binding motif, suggesting that disrupted autoregulation could be the mechanism underlying the observed strabismus phenotype. As the causal functional alteration remains to be proven, additional studies will be required to identify other families with genetic forms of strabismus mapping to the locus and to conclusively prove the causal sequence alteration and its pathophysiological mechanism.

MATERIALS AND METHODS

Patient ascertainment

A seven-generation pedigree was constructed based on family records, including photos displaying eye alignment. Thirteen participants were descendants of a common ancestor; nine of them reported early-onset isolated strabismus, and the other four

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

reported no strabismus. The common ancestor was reported to be of European origin.

Except for 006, the other eight affected descendants were examined by one or more of three ophthalmologists (Drs J Horton, VP and CL). All participants were asked about the age of onset (if applicable), ocular history and medical history. Examination included visual acuity, pupil observation, eye movements, ocular alignment, stereopsis, slit-lamp examination, fundus examination and intraocular pressure. Individuals 009, 011 and 013 did not have a history of extraocular muscle surgery and therefore underwent full orthoptic examinations. Subjects were asked about other medical or physical characteristics, with no reports spanning beyond immediate nuclear family members.

DNA isolation

Genomic DNA of participants was isolated from either saliva or blood. At least 4 mL blood samples or 6 mL saliva samples were collected for one round of next-generation sequencing, and at least a 2 mL saliva sample was collected from participants for array genotyping. Blood samples were collected in a clinical setting while saliva samples were collected using Oragene-DNA (OG-500) saliva kits. DNA was extracted from blood samples using the Qiagen QIAsymphony SP instrument and the QIAsymphony DNA Midi Kit and from saliva samples with DNA Genotek prepIT-L2P sample preparation kit following protocol no. PD-PR-015. Approximately 7–10 µg DNA per sample at a concentration no less than 70 ng/µL was sent for sequencing. A 500 ng DNA per sample at a concentration of at least 50 ng/µL was sent for array genotyping.

Genotyping: statistical linkage analysis and haplotype analysis

Genotyping was performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. The assay was performed on HumanOmni2.5–8 v1_C, using the Infinium LCG assay (Illumina). Standard quality control steps were performed on the genotypes, including sex check, call rate, autosomal heterozygosity and verification of the pedigree structure. Simulations were performed to determine the maximum possible LOD (logarithm of the odds) score for different model parameters under the alternative hypothesis (linkage). SLINK 3.02[28] was used to simulate pedigrees under dominant and recessive models with a range of disease allele frequencies and penetrance. For a particular model, the maximum LOD score from the analysis of 1000 simulated pedigrees was declared the maximum LOD score.

Multiple filters were applied to select a set of markers suitable for linkage analysis. Only markers with alleles unambiguous for strand information on the autosomes and X chromosomes were kept. Genotype data from HapMap3 European populations[29] were used to estimate marker allele frequencies, and a minor allele frequency >0.45 and pairwise r^2 <0.1 were selected. A set of 17 779 SNPs was obtained after a SNP filtering step. Merlin V.1.1.2[30] was used to perform multipoint linkage analysis under the same model as in the SLINK simulation. Analysis of the X chromosome was performed within Merlin using the standard procedures.[30]

As individual 014 was recruited at a later time point, we extracted SNPs from WGS data. We then used the combined SNP genotypes for linkage analysis on the family and performed genome-wide parametric and non-parametric linkage analyses using Merlin V.1.1.2. To refine the boundaries of the linked region, we examined the SNPs from both edges (rs714641 and rs1951187) manually between descendants and identified the minimum shared haplotype region.

Whole-exome sequencing

We performed WES on 001 and 011, affected third cousins. WES was performed via the Agilent SureSelect Human All Exon 38 Mb kit and Illumina HiSEQ 2000 platform (performed by Perkin Elmer) with an average coverage of 27X. The genomic aligners, Bowtie (V.0.12.9) and BWA (V.0.6.1), were used to map the paired-end reads to the hg19 reference genome.[31 32] The Genome Analysis Toolkit (GATK) (V.1.0) performed local re-alignment, which allowed for correcting misalignment at the extremity of reads.[33] SAMtools (V.0.1.18) was applied to call variants from aligned WES reads.[34] In-house scripts were used to filter variants according to the following criteria: under an autosomal dominant model, with a frequency not higher than 1% in dbSNP build 135, non-synonymous coding variants, and predicted by SIFT[12] to be 'damaging' or indeterminate.

Whole-genome sequencing

Two rounds of WGS were conducted as the project progressed. WGS was first performed on 001, 013 and 014, who were three distantly related affected cousins, on an Illumina HiSEQ 2000 platform (BGI America) generating paired-end reads of 125 bp and average coverage of 37X. An informatics pipeline (similar to the WES pipeline, but with newer versions of software) was applied to this set of WGS data: Bowtie (V.1.0.0) and BWA (V.0.7.5a) for mapping the paired-end reads to the hg19 reference genome,[31 32] GATK (V.2.8) for local re-alignment[33] and SAMtools (V.0.1.19) for variant calling.[34]

Variants located within the linkage region were selected for further analysis. Allele frequency was assessed using dbSNP build 137 and Exome Variant Server (http://evs.gs.washington.edu/EVS/) and variants with a frequency higher than 1% were excluded. Heterozygous variants shared across the three samples were selected, and SnpEff[35] (with hg19 database) was applied to annotate those variants.

Later, in order to obtain PCR-free results suitable for analysis of short tandem repeats, WGS was performed on 011, 012 and 013, a trio, on an Illumina NovaSeq platform (Macrogen) with an average coverage of 45X. A different informatic pipeline was applied to this set of WGS data: BWA mem (V.0.7.12) for mapping the paired-end reads to the GRCh37 reference (http://www.bcgsc.ca/downloads/genomes/9606/hg19/1000genomes/bwa_ind/genome/), SAMtools (V.1.2) for file format conversion, duplicate marking with Picard (V.1.139), GATK for local re-alignment (V.3.4–46) and GATK HaplotypeCaller for variant calling (V.3.4–46). Variants were soft filtered using BCFTools (V.1.8) keeping variants with at least 10 reads supporting the alternate allele and a maximum depth of 300. Filtered variants were then annotated and normalised using SnpEff (V.4.11; gene version GRCh37.75), VT (V.0.5772) and VCFAnno (V.0.2.8).[36] Filtered, annotated variants are then converted into a GEMINI database (V.0.19.1)[37] using VCF2DB (https://github.com/quinlan-lab/vcf2db). Specific GEMINI queries were performed for variants under the autosomal dominant model, with details below. Scripts for processing the data, and details regarding databases annotated against using VCFAnno and CNV analyses can be found online (https://github.com/Phillip-a-richmond/AnnotateVariants/tree/master/Strabismus).[38]

Variants under the autosomal dominant model (shared between 011 and 013) located within the linkage region were selected for further investigation. Reflecting the upper end of

J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

reported strabismus population frequency, all variants with allele frequency <4% were manually reviewed. However, it is expected that dominant transmission of isolated strabismus is a rare event, and therefore the results reported in this manuscript pertain to those variants with allele frequency <1% and <10 homozygous individuals for the minor allele in gnomAD V.2.0.[10] Candidate rare variants within the region were then assigned to gene and then manually examined for potential molecular impact of the gene by database searches of tissue expression patterns and previously published data.

Non-coding variant annotation and interpretation

To enable analysis of non-coding variants, we used multiple databases and corresponding bioinformatic tools to annotate such variants, including functional annotation of the mammalian genome 5 (FANTOM5) database,[39] JASPAR,[23] Segway,[40] RegulomeDB[41] and Combined Annotation Dependent Depletion (CADD).[14] Data related to the candidate variants were visualised in the UCSC Genome Browser. The 100 Vertebrates Basewise Conservation PhyloP track was used for examining the conservation status. Due to the incomplete annotation of non-coding RNAs across different species, Emsembl BLAST/BLAT was used to find potential orthologs in other vertebrates, and RefSeq genes and ESTs were considered as evidence for RNA transcripts.[17] Based on the qualitative assessment, the top prioritised variant was confirmed to be present by Sanger sequencing in all of the affected subjects and absent from unaffected subjects.

SUMMARY

Eye misalignment, or strabismus, can affect up to 4% of individuals. When strabismus is detected early, intervention in young children based on eye patching and/or corrective lenses can be beneficial. In some cases, corrective surgeries are used to align the eyes, with many individuals requiring multiple surgeries over a lifetime. A better understanding of the causes of strabismus may lead to earlier detection as well as improved treatment options. Hippocrates observed that strabismus runs in families over 2400 years ago, an early recognition of what we now recognise as a portion of cases arising from genetic causes. We describe a large family affected by strabismus and identify a single region on chromosome 14 that may be responsible. The region contains *FOXG1*, in which mutations are known to cause a severe neurodevelopmental syndrome, with 84% of affected individuals also having strabismus. We identify a 4 bp deletion in the region that appears to autoregulate when *FOXG1* is active. Future study of this genetic alteration may enhance our understanding of the mechanisms of strabismus.

Author affiliations
[1]Centre for Molecular Medicine and Therapeutics, The University of British Columbia, Vancouver, British Columbia, Canada
[2]Department of Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada
[3]BC Children's Hospital Research Institute, The University of British Columbia, Vancouver, British Columbia, Canada
[4]The Centre for Applied Genomics, Hospital for Sick Children Research Institute, University of Toronto, Toronto, Ontario, Canada
[5]Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada
[6]Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[7]Department of Ophthalmology and Visual Sciences, The University of British Columbia, Vancouver, British Columbia, Canada
[8]Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

Twitter Oriol Fornes @ofornes

Acknowledgements We thank Dr Jonathan C Horton (University of California San Francisco) for generously offering his time to examine multiple participants in his clinic. We thank Care4Rare (Dr. Kym Boycott, PI, Children's Hospital of Eastern Ontario ResearchInstitute, University of Ottawa) for support of the linkage analysis, and recognize that authors NMR and ADPparticipated as members of the Care4Rare project. We thank Ms Dora Pak for research management support.

Contributors XCY: substantial contributions to the conception and design of the work, the acquisition, analysis and interpretation of data. Drafting the work and revising it critically for important intellectual content. NMR, ADP (collaborated through Care4Rare Canada); PR, CS, OF, XH, MH, CJR: substantial contributions to analysis and interpretation of data. Revising the work critically for important intellectual content. CJL, VP: substantial contributions to the acquisition, analysis and interpretation of data. Revising the work critically for important intellectual content. DG, CG-E: substantial contributions to the conception of the work. Revising the work critically for important intellectual content. MSP, WWW: substantial contributions to the conception and design of the work. Drafting the work and revising it critically for important intellectual content. All the listed authors: final approval of the version published. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding Funding for XCY is provided by a Frederick Banting and Charles Best Canada Graduate Scholarship (GSD-146285) from the Canadian Institutes of Health Research, BCCHRI-CIHR-UBC MD/PhD Studentship, and NSERC Discovery Grant (RGPIN-355532-10, RGPIN-2017-06824). The SNP genotyping was performed under the Care4Rare Canada Consortium funded by Genome Canada and the Ontario Genomics Institute (OGI-147), the Canadian Institutes of Health Research, Ontario Research Fund, Genome Alberta, Genome British Columbia, Genome Quebec, and Children's Hospital of Eastern Ontario Foundation. CJR is supported by a Michael Smith Foundation for Health Research Scholar Award.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the University of British Columbia Children's & Women's Research Ethics Board (approval number CW10-0317/H10-03215), and written consent forms were obtained from all participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The genome-wide sequencing data cannot be shared publicly because of privacy concern. The identifying variants are available from the corresponding authors on request and under an appropriate data handling agreement.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Wyeth W Wasserman http://orcid.org/0000-0001-6098-6412

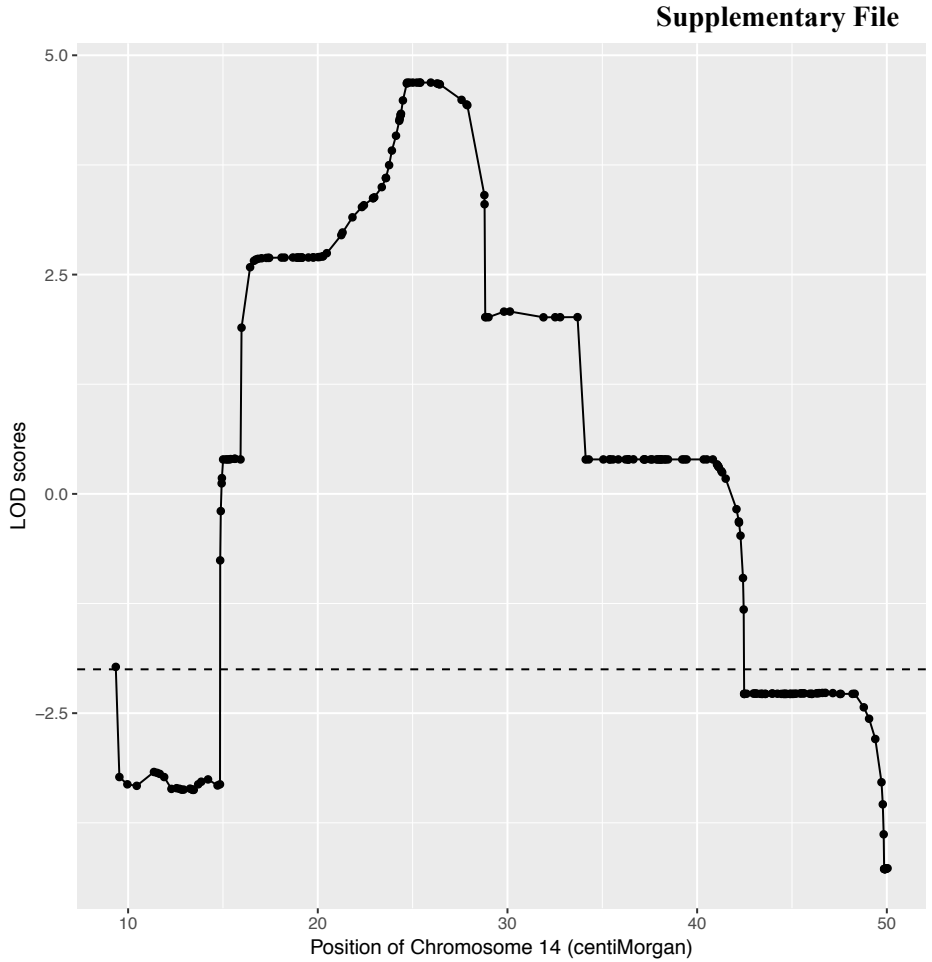
REFERENCES

1 Demer JL. Neuroanatomical strabismus. In: *Pediatric ophthalmology, neuro-ophthalmology, genetics*. Springer Berlin Heidelberg, 2010: 59–75.
2 Lorenz B. Genetics of isolated and syndromic strabismus: facts and perspectives. *Strabismus* 2002;10:147–56.
3 Ye XC, Pegado V, Patel MS, Wasserman WW. Strabismus genetics across a spectrum of eye misalignment disorders. *Clin Genet* 2014;86:103–11.
4 Shaaban S, MacKinnon S, Andrews C, Staffieri SE, Maconachie GDE, Chan W-M, Whitman MC, Morton SU, Yazar S, MacGregor S, Elder JE, Traboulsi EI, Gottlob I, Hewitt AW, Hunter DG, Mackey DA, Engle EC, Strabismus Genetics Research Consortium. Genome-wide association study identifies a susceptibility locus for comitant esotropia and suggests a parent-of-origin effect. *Invest Ophthalmol Vis Sci* 2018;59:4054–64.
5 Mitter D, Pringsheim M, Kaulisch M, Plümacher KS, Schröder S, Warthemann R, Abou Jamra R, Baethmann M, Bast T, Büttel H-M, Cohen JS, Conover E, Courage C, Eger

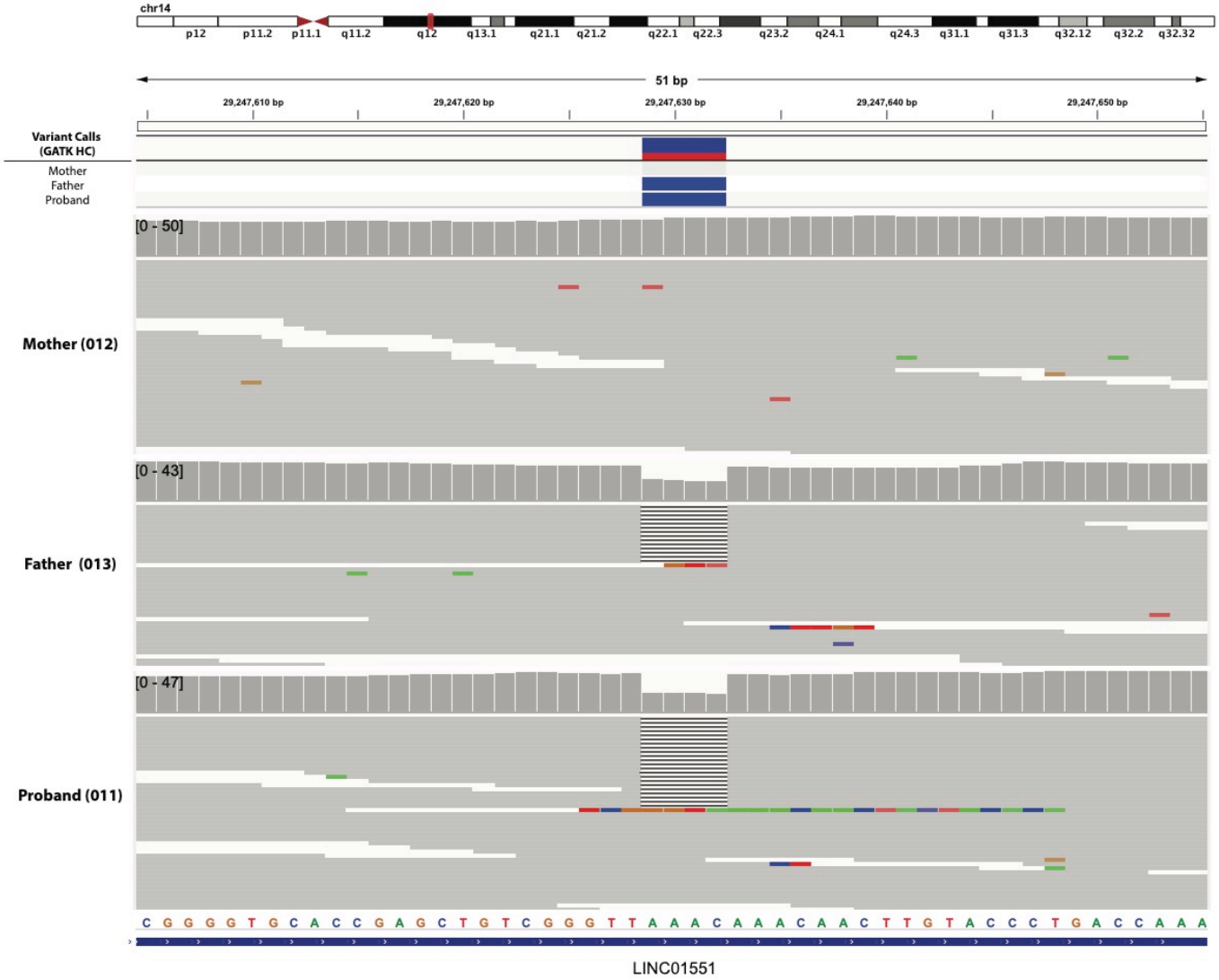
J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

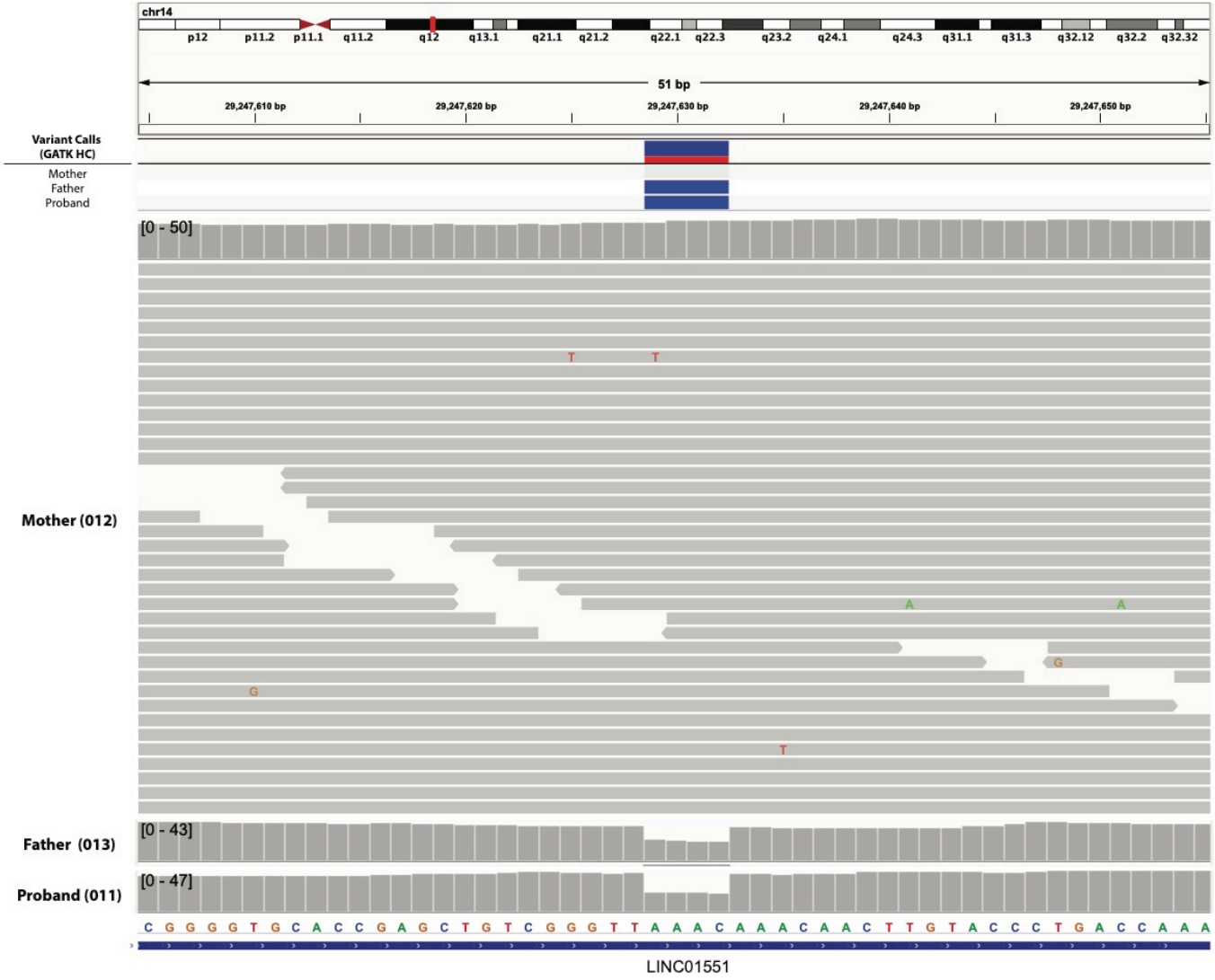
A, Fatemi A, Grebe TA, Hauser NS, Heinritz W, Helbig KL, Heruth M, Huhle D, Höft K, Karch S, Kluger G, Korenke GC, Lemke JR, Lutz RE, Patzer S, Prehl I, Hoertnagel K, Ramsey K, Rating T, Rieß A, Rohena L, Schimmel M, Westman R, Zech F-M, Zoll B, Malzahn D, Zirn B, Brockmann K. FOXG1 syndrome: genotype–phenotype association in 83 patients with *FOXG1* variants. *Genet Med* 2018;20:98–108.
6 Brooks BP, Zein WM, Thompson AH, Mokhtarzadeh M, Doherty DA, Parisi M, Glass IA, Malicdan MC, Vilboux T, Vemulapalli M, Mullikin JC, Gahl WA, Gunay-Aygun M. Joubert syndrome: ophthalmological findings in correlation with genotype and hepatorenal disease in 99 patients prospectively evaluated at a single center. *Ophthalmology* 2018;125:1937–52.
7 Chilton JK, Guthrie S. Axons get ahead: insights into axon guidance and congenital cranial dysinnervation disorders. *Dev Neurobiol* 2017;77:861–75.
8 Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;11:241–7.
9 Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, Zhang L-H, Bhavsar AP, Lee JJY, Drögemöller BI, Abdelsayed M, Alfadhel M, Armstrong L, Baumgartner MR, Burda P, Connolly MB, Cameron J, Demos M, Dewan T, Dionne J, Evans AM, Friedman JM, Garber I, Lewis S, Ling J, Mandal R, Mattman A, McKinnon M, Michoulas A, Metzger D, Ogunbayo OA, Rakic B, Rozmus J, Ruben P, Sayson B, Santra S, Schultz KR, Selby K, Shekel P, Sirrs S, Skrypnyk C, Superti-Furga A, Turvey SE, Van Allen MI, Wishart D, Wu J, Wu J, Zafeiriou D, Kluijtmans L, Wevers RA, Eydoux P, Lehman AM, Vallance H, Stockler-Ipsiroglu S, Sinclair G, Wasserman WW, van Karnebeek CD. Exome sequencing and the management of neurometabolic disorders. *N Engl J Med* 2016;374:2246–55.
10 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
11 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015;526:68–74.
12 Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40:W452–7.
13 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
14 Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94.
15 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
16 Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
17 Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. Ensembl 2018. *Nucleic Acids Res* 2018;46:D754–61.
18 Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: key players in health and disease. *Trends Genet* 2011;27:224–32.
19 Godbole G, Shetty AS, Roy A, D'Souza L, Chen B, Miyoshi G, Fishell G, Tole S. Hierarchical genetic interactions between FOXG1 and LHX2 regulate the formation of the cortical hem in the developing telencephalon. *Development* 2018;145. doi:10.1242/dev.154583. [Epub ahead of print: 09 01 2018].
20 Ferreira RdaC, Oelrich F, Bateman B. Genetic aspects of strabismus. *Arq Bras Oftalmol* 2002;65:171–5.
21 Khan AO, Shinwari J, Abu Dhaim N, Khalil D, Al Sharif L, Al Tassan N. Potential linkage of different phenotypic forms of childhood strabismus to a recessive susceptibility locus (16p13.12-p12.3). *Mol Vis* 2011;17:971–6.
22 Mehrjouy MM, Fonseca ACS, Ehmke N, Paskulin G, Novelli A, Benedicenti F, Mencarelli MA, Renieri A, Busa T, Missirian C, Hansen C, Abe KT, Speck-Martins CE, Vianna-Morgante AM, Bak M, Tommerup N. Regulatory variants of FOXG1 in the context of its topological domain organisation. *Eur J Hum Genet* 2018;26:186–96.
23 Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, JASPAR MA. Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 2018;2018:D260–6.
24 Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, Damante G, Grainger R, van Heyningen V, Kleinjan DA. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved Pax6 enhancer causes aniridia. *Am J Hum Genet* 2013;93:1126–34.
25 Hébert JM, Fishell G. The genetics of early telencephalon patterning: some assembly required. *Nat Rev Neurosci* 2008;9:678–85.
26 Patriarchi T, Amabile S, Frullanti E, Landucci E, Lo Rizzo C, Ariani F, Costa M, Olimpico F, W Hell J, M Vaccarino F, Renieri A, Meloni I. Imbalance of excitatory/inhibitory synaptic protein expression in iPSC-derived neurons from FOXG1$^{(+/-)}$ patients and in foxg1$^{(+/-)}$ mice. *Eur J Hum Genet* 2016;24:871–80.
27 Cargnin F, Kwon J-S, Katzman S, Chen B, Lee JW, Lee S-K. FOXG1 orchestrates neocortical organization and cortico-cortical connections. *Neuron* 2018;100:1083–96.
28 Schäffer AA, Lemire M, Ott J, Lathrop GM, Weeks DE. Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees. *Hum Hered* 2011;71:126–34.
29 , Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJR, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE, International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–8.
30 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101.
31 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
32 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
33 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
34 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genomes Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
35 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
36 Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol* 2016;17:118.
37 Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol* 2013;9:e1003153.
38 van Kuilenburg ABP, Tarailo-Graovac M, Richmond PA, Drögemöller BI, Pouladi MA, Leen R, Brand-Arzamendi K, Dobritzsch D, Dolzhenko E, Eberle MA, Hayward B, Jones MJ, Karbassi F, Kobor MS, Koster J, Kumari D, Li M, MacIsaac J, McDonald C, Meijer J, Nguyen C, Rajan-Babu I-S, Scherer SW, Sim B, Trost B, Tseng LA, Turkenburg M, van Vugt JJFA, Veldink JH, Walia JS, Wang Y, van Weeghel M, Wright GEB, Xu X, Yuen RKC, Zhang J, Ross CJ, Wasserman WW, Geraghty MT, Santra S, Wanders RJA, Wen X-Y, Waterham HR, Usdin K, van Karnebeek CDM. Glutaminase deficiency caused by short tandem repeat expansion in *GLS*. *N Engl J Med* 2019;380:1433–41.
39 Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, Mungall CJ, Arner E, Baillie JK, Bertin N, Bono H, de Hoon M, Diehl AD, Dimont E, Freeman TC, Fujieda K, Hide W, Kaliyaperumal R, Katayama T, Lassmann T, Meehan TF, Nishikata K, Ono H, Rehli M, Sandelin A, Schultes EA, 't Hoen PAC, Tatum Z, Thompson M, Toyoda T, Wright DW, Daub CO, Itoh M, Carninci P, Hayashizaki Y, Forrest ARR, Kawaji H, FANTOM consortium. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 2015;16:22.
40 Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* 2012;9:473–6.
41 Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7.
42 Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, Hardison R, Wang T, Yue F. The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:151.

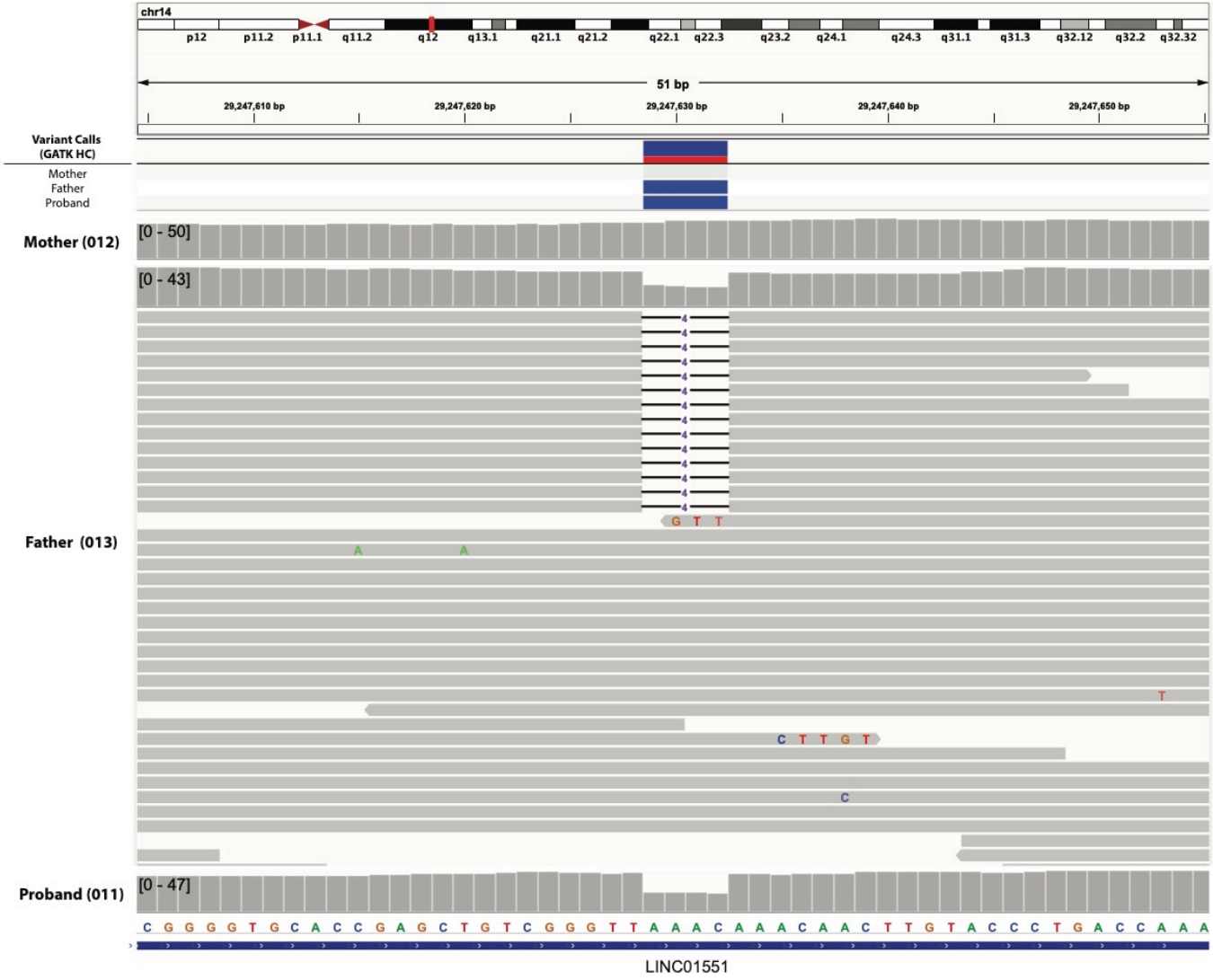
J Med Genet: first published as 10.1136/jmedgenet-2020-107226 on 30 November 2020. Downloaded from http://jmg.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

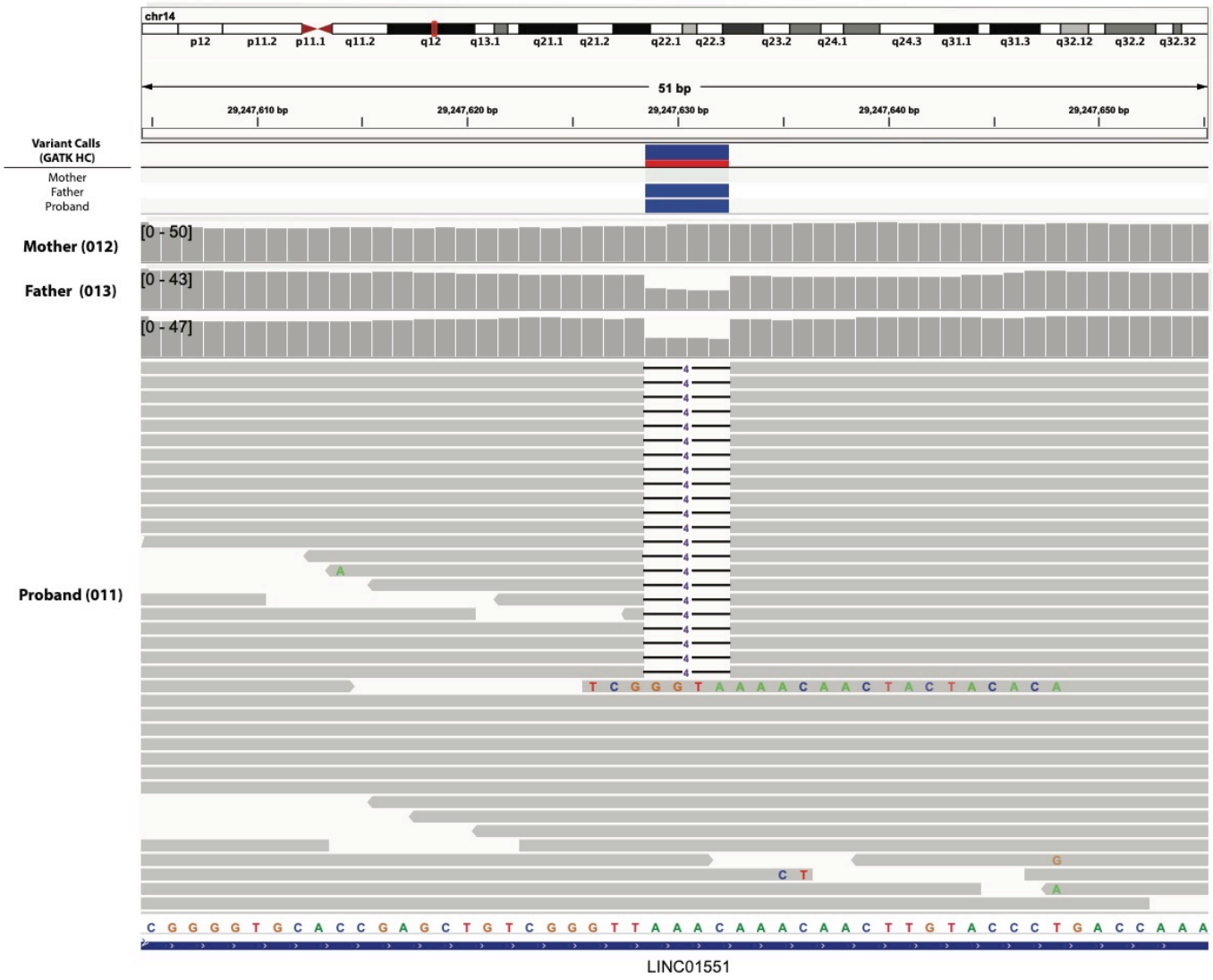


Supplementary Figure 1. LOD score plot for the observed linked region on chromosome 14. The horizontal axis represents the location in terms of centiMorgans. The displayed region spans the physical positions of chr14:28,467,136 – 30,045,978 (hg19/GRCh37). The vertical axis represents LOD score. The dash line indicates the threshold of LOD <-2.


chr14
p12
p11.2
p11.1
q11.2
q12
q13.1
q21.1
q21.2
q22.1
q22.3
q23.2
q24.1
q24.3
q31.1
q31.3
q32.12
q32.2
q32.32
51 bp
29,247,610 bp
29,247,620 bp
29,247,630 bp
29,247,640 bp
29,247,650 bp
Variant Calls (GATK HC)
Mother
Father
Proband
[0 - 50]
Mother (012)
[0 - 43]
Father (013)
[0 - 47]
Proband (011)
C G G G G T G C A C C G A G C T G T C G G G T T A A A C A A A C A A C T T G T A C C C T G A C C A A A
LINC01551


chr14
p12 p11.2 p11.1 q11.2 q12 q13.1 q21.1 q21.2 q22.1 q22.3 q23.2 q24.1 q24.3 q31.1 q31.3 q32.12 q32.2 q32.32
51 bp
29,247,610 bp
29,247,620 bp
29,247,630 bp
29,247,640 bp
29,247,650 bp
Variant Calls (GATK HC)
Mother
Father
Proband
[0 - 50]
Mother (012)
[0 - 43]
Father (013)
[0 - 47]
Proband (011)
C G G G G T G C A C C G A G C T G T C G G G T T A A A C A A A C A A C T T G T A C C C T G A C C A A A
LINC01551


chr14
p12 p11.2 p11.1 q11.2 q12 q13.1 q21.1 q21.2 q22.1 q22.3 q23.2 q24.1 q24.3 q31.1 q31.3 q32.12 q32.2 q32.32
51 bp
29,247,610 bp
29,247,620 bp
29,247,630 bp
29,247,640 bp
29,247,650 bp
Variant Calls (GATK HC)
Mother
Father
Proband
Mother (012)
[0 - 50]
[0 - 43]
Father (013)
Proband (011)
[0 - 47]
C G G G G T G C A C C G A G C T G T C G G G T T A A A C A A A C A A C T T G T A C C C T G A C C A A A
LINC01551


chr14
p12 p11.2 p11.1 q11.2 q12 q13.1 q21.1 q21.2 q22.1 q22.3 q23.2 q24.1 q24.3 q31.1 q31.3 q32.12 q32.2 q32.32
51 bp
29,247,610 bp
29,247,620 bp
29,247,630 bp
29,247,640 bp
29,247,650 bp
Variant Calls (GATK HC)
Mother
Father
Proband
Mother (012)
[0 - 50]
Father (013)
[0 - 43]
[0 - 47]
Proband (011)
T C G G G T A A A A C A A C T A C T A C A C A
C T
G
A
C G G G G T G C A C C G A G C T G T C G G G T T A A A C A A A C A A C T T G T A C C C T G A C C A A A
LINC01551

Supplementary Figure 2.

a) Integrative Genomic Viewer (IGV) screenshot of the WGS for the parent-child trio. (Affected: 011 and 013; Unaffected: 012). The horizontal axis represents the sequence of the reference genome. A small portion of the aligned reads for each of the three individuals is shown as indicated by the labels along the vertical axis.

b) Extended view of IGV screenshot for 012 (Mother).

c) Extended view of IGV screenshot for 013 (Father).

d) Extended view of IGV screenshot for 011 (Proband).

(Detailed information about the IGV display can be obtained on https://software.broadinstitute.org/software/igv/)