

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

Short Title:

New isolated strabismus locus in FOXG1 syndrome region

Authors: Xin (Cynthia) Ye^{1,2,3}, Nicole M. Roslin⁴, Andrew D. Paterson^{4,5,7}, Care4Rare Canada⁶, Christopher Lyons^{3,8}, Victor Pegado⁸, Phillip Richmond^{1,3}, Casper Shyr^{1,3}, Oriol Fornes^{1,3}, Xiaohua Han¹, Michelle Higginson¹, Colin J. Ross⁹, Deborah Giaschi^{3,8}, Cheryl Y. Gregory-Evans⁸, Millan Patel^{*2,3}, Wyeth W. Wasserman^{*1,2,3}

1. Centre for Molecular Medicine and Therapeutics, University of British Columbia, BC, Canada
2. Department of Medical Genetics, University of British Columbia, BC, Canada
3. BC Children's Hospital Research Institute, University of British Columbia, 950 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada
4. The Centre for Applied Genomics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
5. Genetics & Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada;
6. Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada, K1H 8L1
7. Divisions of Epidemiology and Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
8. Department of Ophthalmology and Visual Sciences, University of British Columbia, BC, Canada
9. Faculty of Pharmaceutical Science, University of British Columbia, BC, V6T, 1Z3, Canada

Corresponding Contact:

* Wyeth W. Wasserman (wyeth@cmmt.ubc.ca)

* Millan Patel (mPatel@cw.bc.ca)

CONFLICT OF INTEREST

The authors declare no conflict of interest.

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, and a single linkage signal has been identified at chromosome 14q12 with a multipoint LOD score of 4.69. Disruption of this locus is known to cause FOXP1 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 4bp non-coding deletion was prioritized as the top candidate for the observed strabismus phenotype. The deletion is predicted to disrupt regulation of *FOXP1*, which encodes a transcription factor of the Forkhead family. Suggestive of an auto-regulation effect, the disrupted sequence matches the consensus FOXP1 and Forkhead family transcription factor binding site and has been observed in previous ChIP-seq studies to be bound by Foxp1 in early mouse brain development. The findings of this study indicate that the strabismus phenotype commonly observed within FOXP1 syndrome is separable from the more severe syndromic characteristics. Future study of this specific deletion may shed light on the regulation of *FOXP1* expression and may enhance our understanding of the mechanisms contributing to strabismus and FOXP1 syndrome.

Author 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 2,400 years ago, an early recognition of what we now recognize 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 syndrome, with 84% of affected individuals also having strabismus. We identify a 4bp deletion in the region that appears to auto-regulate when *FOXG1* is active. Future study of this genetic alteration may enhance our understanding of the mechanisms of strabismus.

Keywords: Strabismus, Linkage Analysis, Whole Genome Sequencing, Bioinformatics, FOXG1 Syndrome

1 **Introduction**

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

6 As early as Hippocrates' time, strabismus was recognized as a genetic disorder based on
7 an observation of its tendency to cluster within families (2). During the last century, twin and
8 family studies have demonstrated a substantial genetic contribution to strabismus, and both
9 autosomal dominant and autosomal recessive transmission patterns have been reported (3).
10 Recently a genome wide association study reported two variants with small effect sizes (4).
11 Strabismus occurs commonly as one phenotype amongst many in syndromes, such as congenital
12 Rett syndrome (FOXP1 syndrome) and Joubert syndrome, in which 84% and 75% of individuals
13 display a strabismus phenotype respectively (5,6). On the other hand, families displaying isolated
14 strabismus transmitting in simple Mendelian patterns are uncommon. In a few such families,
15 genetic loci on chromosomes 4, 6, 7, 12, 16, 19 have been identified for isolated strabismus, but
16 no causal gene has been identified in these regions (3). Not overlapping the reported loci, eleven
17 genes (*PHOX2A*, *ROBO3*, *KIF21A*, *SALL1*, *TUBB3*, *HOXB1*, *SALL4*, *CHN1*, *HOXA1*, *TUBB2B*,
18 *MAFB*) (7), of which five encode transcription factors (underlined), have been identified for a
19 subgroup of strabismus associated with congenital cranial dysinnervation disorders (7), but the
20 genetic etiology of other strabismus subtypes remains elusive. Identification of a locus with high
21 confidence, determination of a causal gene and detailed mechanistic insights at the nucleotide
22 level would provide new insights into the molecular mechanisms of strabismus.

23 We compiled a large, seven-generation, non-consanguineous pedigree with 21
24 individuals exhibiting isolated strabismus, consistent with an autosomal dominant inheritance
25 pattern. Through genome-wide linkage analysis of 8 affected and 4 unaffected individuals from
26 one branch and one affected individual from a separate branch, we mapped this familial
27 strabismus to chromosome 14q12, which overlaps with the FOXP1 syndrome locus. Next
28 generation sequencing and in-depth analysis identified a strong candidate deletion variant within
29 this locus. The data supports that the deletion localizes to a FOXP1 transcription factor binding
30 site (TFBS), suggesting an auto-regulatory loop may be disrupted that controls *FOXP1*
31 expression in early brain development.

32

33 **Results**

34 **Pedigree and participant profile**

35 A seven-generation pedigree of European origin, with over 176 individuals, including those who
36 were deceased, was compiled. Individuals were labelled as strabismic either through medical
37 records or strong family anecdotes. A roughly even distribution of strabismus cases was
38 observed between females (12 individuals) and males (nine individuals) and the disorder was
39 transmitted both maternally and paternally. Among the three extensively traced branches (Figure
40 1a), branch 1 was most well-documented and contained most of the participants in this study. In
41 branch 1, strabismus was reported across four consecutive generations. An autosomal dominant
42 inheritance model with high penetrance best matched the qualitative observation.

43 To characterize the strabismus in this family, each of the nine affected descendants
44 indicated in Figure 1b was seen by one of three participating ophthalmologists specialized in
45 strabismus. All individuals underwent complete ophthalmic examination with attention to ocular

46 motility both before and after pharmacologic cycloplegia. The specific characteristics of
47 strabismus were not uniform across the descendants in the family (Table 1). The strabismus
48 phenotypes could be grouped into two broad directional categories: esotropia (an eye turns in)
49 and hypertropia (an eye turns up). Both esotropia and hypertropia were noted in individual 014,
50 but this individual had undergone multiple corrective surgeries and childhood medical records
51 were not available. On the other hand, individuals 011, 013 and 009 are in consecutive
52 generations and had no history of extraocular muscle surgery. Individual 011 presented with
53 esotropia while the other two displayed hypertropia (009, 013). The unaffected status was
54 confirmed by past medical history provided by the participating family members. In the clinical
55 ophthalmology exams and in oral reports from subjects, there were no other phenotypes observed
56 broadly in the individuals with strabismus.

57

58 **Linkage analysis and haplotype analysis**

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

65 With the same linkage analysis parameters as in the simulation, the largest LOD score
66 based on the observed genotypes was 3.55, close to the simulated maximum. This chromosome
67 14 locus was the only region with a LOD score higher than 3, and exceeds the standard threshold
68 for genome-wide significance of 3.3 (8). This linked region spanned approximately 10 Mb and

69 was bounded by rs7146411 and rs1951187, corresponding to chr14: 22,779,843 - 32,908,192
70 (hg19). Chromosome 14q12 is therefore a novel locus for isolated strabismus.

71 Individual 014 representing branch 2 in the pedigree was recruited to the study after the
72 genome-wide genotyping was performed. An expanded linkage analysis with corresponding
73 SNPs extracted from 014's whole genome sequencing (WGS) data further supported the linkage
74 to chromosome 14. Simulations under the alternative hypothesis (linkage) generated a maximum
75 LOD score of 4.69 for the same region, under an autosomal dominant model with disease allele
76 frequency $q = 0.005$, 99% penetrance, and 0.2% phenocopy rate. We observed a LOD score of
77 4.69, and the region remained as the sole candidate (Figure 2a). In addition, we performed non-
78 parametric linkage analyses and identified the same linkage region (Figure 2b).

79 Haplotype analysis was used to complement linkage analysis by providing visual
80 confirmation of the statistical testing. An approximately 8.5 Mb region (chr14:22,779,843 -
81 31,289,720) was shared between the nine affected descendants of the common ancestor. An
82 unaffected descendant (005, subsequently deceased) shared a 5.5 Mb region within the linked
83 region (Figure 3a and Figure 3b). Thus, an approximately 3 Mb region was shared exclusively by
84 nine affected descendants, corresponding to chr14: 28,467,136 - 31,289,720. The LOD score was
85 above 4.60 for chr14:28,467,136 – 30,045,978 (Supplementary Figure 1). The 3 Mb region
86 situates within a gene poor region (Figure 3b).

87

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

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

92 WES project (9). Only one of the variants (chr14: 31061628 A>G, rs145527124) was located
93 within the 10 Mb locus, falling within exon 5 of *G2E3* (G2/M-phase specific E3 ubiquitin
94 protein ligase) (NM_017769.4).

95 This variant leads to Ile113Val alteration in ENST00000206595. In gnomAD v 2.1.1
96 European (non-Finnish) population, there are two homozygotes for this variant and the overall
97 allele frequency is 0.17% and 2 homozygotes while gnomAD v3 has an overall allele frequency
98 of 0.15% and another homozygote (10). In addition, in the 1000 Genomes project (11) the allele
99 frequency in Punjabi from Lahore, Pakistan (PJL) is 1.6%. This *G2E3* variant was not supported
100 as a candidate by computational analysis (predicted to be “tolerated” with SIFT (12) and
101 “benign” with Polyphen (13)). Qualitative review of the literature did not suggest a potential role
102 for *G2E3* in a strabismus phenotype.

103 The lack of candidate variants from WES motivated the generation of WGS for
104 individuals 001, 013, and 014, who were selected to represent the three branches of the pedigree.
105 The only coding variant detected with a frequency of $\leq 1\%$ in the 10 Mb region was the
106 aforementioned *G2E3* variant.

107

108 **WGS and bioinformatics analyses highlight a heterozygous non-coding variant in a** 109 **regulatory region of *FOXG1***

110 Our analyses showed that the WGS on 001, 013, and 014 did not capture some low complexity
111 regions, raising concern that the protocol used at the time of generation might fail to detect
112 repetitive sequences and small genomic structural alterations. As current WGS protocols could
113 better detect such properties, we generated WGS for an affected parent-child trio (011, 012, and

114 013). For simplicity, we report variants from the trio WGS set for comprehensiveness in the
115 following sections.

116 Within the 3 Mb region, a total of 664 variants were shared by both affected individuals
117 (based on a dominant model of transmission). No copy number variations (CNV) were detected.
118 We focused on variants with a frequency $\leq 1\%$ and which have been reported in fewer than 10
119 homozygotes in gnomAD 2.0, criteria met by 24 of the 664 variants. The only coding variant of
120 these 24 was the *G2E3* coding variant reported above. (These potentially identifying variants are
121 available from the corresponding author upon request and under an appropriate data handling
122 agreement.)

123 As 23 of 24 candidate variants prioritized in the WGS analysis were non-coding, we
124 used diverse methods to annotate non-coding variants with regulatory information. There is no
125 standard practice to annotate non-coding variants, so we used a variety of approaches to identify
126 those within potential regulatory elements. One variant was noted recurrently as interesting using
127 a variety of bioinformatic predictions. Among the 23 candidate variants, only chr14:29247628
128 TAAAC>T (Supplementary Figure 2) has been assigned a CADD score over 20 (14) and is ~10
129 kb 3' from *FOXG1*. This variant is situated within a potential regulatory region, as suggested by
130 the presence of DNase-seq peaks and the histone marks H3K4me1, H3K4me2, H3K4me3,
131 H3K9ac and H3K27ac, which are associated with promoters/enhancers (15). This deletion was
132 absent in gnomAD v 2.1.1 and v 3 and dbSNP build 152 and 153 (16). This deletion was
133 confirmed through Sanger sequencing to be present in all nine affected subjects and none of the
134 unaffected individuals.

135 Since the upper limit of the reported prevalence of strabismus is 4%, additional variants
136 with frequency $>1\%$ and $\leq 4\%$ in gnomAD were also obtained for examination. A total of 54

137 additional variants were identified in the candidate region. However, none of these variants
138 situated on protein coding regions or were assigned a CADD score over 20.

139 We examined topologically associating domains (TADs) for the 3 Mb region (Figure 4)
140 to suggest potential regulatory relationships between identified variants and nearby genes. The
141 top candidate variant was located within the same TAD as *FOXG1*, and hereafter this TAD will
142 be referred to as the FOXG1-TAD. Both *FOXG1* and the sequence surrounding the candidate
143 variant are highly conserved across vertebrates, with the candidate sequence retained from fish to
144 humans (Figure 6).

145 In human genome annotations, the variant chr14:29247628 TAAAC>T was located
146 within an alternative exon of a long non-coding RNA gene (*LINC01551*). Within the mouse,
147 chicken, and zebrafish annotation and supporting data, there were no RNA transcripts containing
148 the variant (17). As the variant position is conserved back to fish, and the transcript evidence is
149 not supportive of transcription of the highly conserved region in other species, we considered
150 whether the variant might be situated within a *cis*-regulatory region. We examined predicted
151 transcription factor binding motifs overlapping the deletion and observed a match to Forkhead
152 transcription factor binding sites, including JASPAR profile matches for FOXC1, FOXI1, and
153 FOXG1 (Figure 5). Proteins in the evolutionarily conserved superfamily of the Forkhead
154 transcription factors share the presence of a DNA binding domain and a transactivation or
155 transrepression effector region, and play a central role during development as well as in the adult
156 (18). The binding motif for both FOXC1 and FOXI1 profiles would be obliterated by the
157 deletion. Due to two consecutive AAAC repeats, the FOXG1 binding motif is present twice in
158 the reference sequence, with one copy remaining after the deletion. Publicly available mouse
159 ChIP-seq data (GSE96070) showed that Foxg1 binds to this site in cortex tissue from E14-15

160 brain. Thus, it appears that the deletion is situated within a Forkhead TF binding site, in a highly
161 conserved region with conservation patterns consistent with a functional role in the *cis*-regulation
162 of *FOXG1*.

163

164 **Discussion**

165 We identified a new locus for isolated strabismus in a family, and this locus overlapped with the
166 locus for FOXG1 syndrome, which has a high prevalence of strabismus. This finding suggests
167 that the strabismus phenotype within FOXG1 syndrome can be isolated from the other
168 phenotypes. In-depth phenotyping of individuals without strabismus surgery illustrated clinical
169 heterogeneity of strabismus within the family, suggesting that while a specific molecular lesion
170 may lead to strabismus, the specific clinical type is determined by other factors. We examined
171 both coding and non-coding variants, which led to identification of a potential strabismus
172 causing sequence alteration of a Forkhead TFBS within the FOXG1-TAD, suggesting disruption
173 of *cis*-regulation.

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

179 In this family the strabismus types were not uniform. The various types of strabismus,
180 however, appear to be caused by the same genetic factor since the linked haplotype is shared by
181 all strabismic individuals. Our observation of strabismus variability in the subject family, review
182 of the literature (19,20), and personal communication with other research groups suggests the

183 current strabismus classification scheme is unhelpful for genetic studies: existing strabismus
184 classification systems may inappropriately be separating individuals sharing a common
185 underlying genetic cause and thus weaken study power. This may explain the paucity of studies
186 detecting and confirming strabismus loci in the literature (3).

187 Since different types of strabismus appear to arise from the same genetic variant, a more
188 reliable or nuanced classification system may be required. In phenotype ontologies, hierarchical
189 classification systems allow for groupings to be examined at multiple levels of resolution. This
190 important capacity is missing in our current approach to strabismus phenotyping. An expansion
191 of the classification system could lead to an improved clinical understanding of strabismus and
192 more efficient strategies to study the genetics of strabismus.

193 This 3 Mb region overlaps with microdeletions/microduplications known to cause
194 FOXP1 syndrome in which a high prevalence of strabismus is observed. FOXP1 syndrome,
195 which is also known as congenital Rett syndrome, is a neurological disorder characterized by
196 impaired development and structural brain abnormalities. Strikingly, 84% of affected individuals
197 display strabismus (5). Distal microdeletions that disrupt the topologically associating domains
198 can lead to FOXP1 syndrome while *FOXP1* remains intact, indicating that mis-regulation of
199 *FOXP1* can cause phenotypic change (21). Due to the close proximity and shared TAD with
200 *FOXP1*, the 4bp deletion is speculated to alter *FOXP1* expression (Figure 4). A spectrum of
201 partially overlapping phenotypes have been reported in patients with FOXP1 syndrome (5). The
202 separability of the strabismus phenotype from intellectual disability and other severe disabling
203 phenotypes therefore represents an important insight.

204 Close examination of diverse data provided important insights into the potential
205 regulatory impact of the 4bp deletion. First, the sequences surrounding the deletion are highly

206 conserved in the genomes of vertebrates, suggesting that it is under evolutionary selection and
207 that a change may be more likely to have a functional impact. Indeed, the sequence containing
208 the deletion and the coding region of *FOXG1* were the only two highly conserved elements in a
209 180 kb neighborhood (Figure 6). In addition, this conserved sequence was not supported as being
210 part of a long non-coding RNA in other species (e.g. mouse, chicken, frog), implying a *cis*-
211 regulatory effect. Second, the conserved sequence disrupted by the deletion was predicted to be a
212 TFBS for Forkhead transcription factors, including FOXG1, according to the binding site
213 profiles from JASPAR (22). Third, Foxg1 ChIP-seq data from E14-15 mouse brain (GSE96070)
214 showed that Foxg1 bound to this sequence.

215 The binding of Foxg1 to this sequence in mouse provides the basis for the hypothesis of
216 disrupted FOXG1 auto-regulation leading to strabismus in the subject family. The proposed
217 auto-regulatory model is illustrated in Figure 7. *FOXG1* is transcribed and translated, the
218 transcription factor binds to the target sequence, helping to maintain the appropriate expression
219 of *FOXG1* during critical developmental period. The disruption of the FOXG1 binding site leads
220 to dysregulation of *FOXG1* expression in a highly specific developmental context that results in
221 the isolated strabismus phenotype.

222 Auto-regulation for key developmental transcription factors in vision is not new to the
223 field. The SIMO regulatory sequence controlling expression of the *PAX6* transcription factor
224 gene is an example of such a distal auto-regulatory element (23). While Pax6 is a crucial
225 transcription factor for delineating the dorsal forebrain in mouse E10.0, Foxg1 is a critical
226 transcription factor for delineating the ventral forebrain in mouse E9.0 (24). Thus, they may
227 share similar sensitivity to regulatory disruption.

228 FOXG1 expression is strongest in fetal brain and its dysregulation leads to unbalanced
229 development of excitatory and inhibitory synapses in iPSC-derived neurons and mice (25). In
230 combination with other transcription factors, Foxg1 in pyramidal neurons is crucial for
231 establishing cortical layers and axon trajectory of callosal projection neurons. Moreover, some
232 Foxg1-directed processes are more vulnerable to dosage changes than others (26). These
233 observations suggest that Foxg1 has a dosage and time sensitive role in different brain structures.
234 This implies that an alteration in Foxg1 expression pattern can have a very specific impact, and
235 the specific phenotype can be separable from the rest.

236 In summary, we identified a 3 Mb region on chromosome 14 that is linked to autosomal
237 dominant transmission of isolated strabismus. The region contains *FOXG1*, which has been
238 previously associated with strabismus in 84% of patients with syndromic disruptions. Within the
239 3 Mb region, the top candidate variant is situated within a FOXG1 transcription factor binding
240 motif, suggesting that disrupted auto-regulation could be the mechanism underlying the observed
241 strabismus phenotype. As the causal functional alteration remains to be proven, additional
242 studies will be required to identify other families with genetic forms of strabismus mapping to
243 the locus and to conclusively prove the causal sequence alteration and its pathophysiological
244 mechanism.

245

246 **Materials and Methods**

247 **Patient ascertainment**

248 The study was approved by the University of British Columbia Children's & Women's Research
249 Ethics Board (approval number CW10-0317/H10-03215), and written consent forms were
250 obtained from all 14 participants. A seven-generation pedigree was constructed based on family

251 records, including photos displaying eye alignment. Thirteen participants were descendants of a
252 common ancestor; nine of them reported early onset isolated strabismus, and the other four
253 reported no strabismus. The common ancestor was reported to be of European origin.

254 Except for 006 and 005, the other eight affected descendants and four unaffected
255 descendants were examined by one or more of three ophthalmologists (Drs. J. Horton, V.
256 Pegado, and C. Lyons). All participants were asked about the age of onset (if applicable), ocular
257 history, and medical history. Examination included visual acuity, pupil observation, eye
258 movements, ocular alignment, stereopsis, slit lamp examination, fundus examination, and
259 intraocular pressure. Individuals 009, 011, and 013 did not have a history of extraocular muscle
260 surgery and therefore underwent full orthoptic exams. Subjects were asked about other medical
261 or physical characteristics, with no reports spanning beyond immediate nuclear family members.

262

263 **DNA isolation**

264 Genomic DNA of participants was isolated from either saliva or blood. At least 4 ml blood
265 samples or 6 ml saliva samples were collected for one round of next generation sequencing, and
266 at least a 2 ml saliva sample was collected from participants for array genotyping. Blood samples
267 were collected in a clinical setting while saliva samples were collected using Oragene-DNA
268 (OG-500) saliva kits. DNA was extracted from blood samples using the Qiagen QIASymphony
269 SP instrument and the QIASymphony DNA Midi Kit and from saliva samples with DNA
270 Genotek prepIT-L2P sample preparation kit following protocol # PD-PR-015. Approximately, 7-
271 10 μg DNA per sample at a concentration no less than 70 $\text{ng}/\mu\text{l}$ was sent for sequencing. A 500
272 ng DNA per sample at a concentration of at least 50 $\text{ng}/\mu\text{l}$ was sent for array genotyping.

273

274 **Genotyping: statistical linkage analysis and haplotype analysis**

275 Genotyping was performed by The Centre for Applied Genomics, The Hospital for Sick
276 Children, Toronto, Canada. The assay was performed on HumanOmni2.5-8v1_C, using the
277 Infinium LCG assay (Illumina). Standard quality control steps were performed on the genotypes,
278 including sex check, call rate, autosomal heterozygosity, and verification of the pedigree
279 structure. Simulations were performed to determine the maximum possible LOD (logarithm of
280 the odds) score for different model parameters under the alternative hypothesis (linkage). SLINK
281 3.02 (27) was used to simulate pedigrees under dominant and recessive models with a range of
282 disease allele frequencies and penetrance. For a particular model, the maximum LOD score from
283 the analysis of 1000 simulated pedigrees was declared the maximum LOD score.
284 Multiple filters were applied to select a set of markers suitable for linkage analysis. Only markers
285 with alleles unambiguous for strand information on the autosomes and X chromosomes were
286 kept. Genotype data from HapMap3 European populations (28) were used to estimate marker
287 allele frequencies, and a minor allele frequency >0.45 and pairwise $r^2 < 0.1$ were selected. A set
288 of 17,779 SNPs was obtained after a SNP filtering step. Merlin 1.1.2 (29) was used to perform
289 multipoint linkage analysis under the same model as in the SLINK simulation. Analysis of the X
290 chromosome was performed within Merlin using the standard procedures (29).

291

292 As individual 014 was recruited at a later time point, we extracted single nucleotide
293 polymorphisms (SNPs) from whole genome sequencing (WGS) data. We then used the
294 combined SNP genotypes for linkage analysis on the family and performed genome-wide
295 parametric and non-parametric linkage analyses using Merlin 1.1.2. To refine the boundaries of

296 the linked region, we examined the SNPs from both edges (rs714641 and rs1951187) manually
297 between descendants and identified the minimum shared haplotype region.

298

299 **Whole-exome sequencing**

300 We performed whole exome sequencing (WES) on 001 and 011, affected third cousins. WES
301 was performed via the Agilent SureSelect Human All Exon 38Mb kit and Illumina HiSEQ 2000
302 platform (performed by Perkin Elmer) with an average coverage of 27X. The genomic aligners,
303 Bowtie (version 0.12.9) and BWA (version 0.6.1), were used to map the paired-end reads to the
304 hg19 reference genome (30,31). The Genome Analysis Toolkit (GATK) (version 1.0) performed
305 local re-alignment, which allowed for correcting misalignment at the extremity of reads (32).
306 SAMtools (versions 0.1.18) was applied to call variants from aligned WES reads (33). In-house
307 scripts were used to filter variants according to the following criteria: under an autosomal
308 dominant model, with a frequency not higher than 1% in dbSNP build 135, non-synonymous
309 coding variants, and predicted by SIFT (12) to be ‘damaging’ or indeterminate.

310

311 **Whole-genome sequencing**

312 Two rounds of WGS were conducted as the project progressed. WGS was first performed on
313 001, 013, and 014, who were three distantly related affected cousins, on an Illumina HiSEQ 2000
314 platform (BGI America) generating paired-end reads of 125 bp and average coverage of 37X. An
315 informatics pipeline (similar to the WES pipeline, but with newer versions of software) was
316 applied to this set of WGS data: Bowtie (version 1.0.0) and BWA (version 0.7.5a) for mapping
317 the paired-end reads to the hg19 reference genome (30,31), GATK (version 2.8) for local re-
318 alignment (32), and SAMtools (version 0.1.19) for variant calling (33).

319 Variants located within the linkage region were selected for further analysis. Allele
320 frequency was assessed using dbSNP build 137 and Exome Variant Server (URL:
321 <http://evs.gs.washington.edu/EVS/>) and variants with a frequency higher than 1% were excluded.
322 Heterozygous variants shared across the three samples were selected, and SnpEff (34) (with hg19
323 database) was applied to annotate those variants.

324 Later, in order to obtain PCR-free results suitable for analysis of short tandem repeats,
325 WGS was performed on 011, 012, and 013, a trio, on an Illumina NovaSeq platform (Macrogen)
326 with an average coverage of 45X. A different informatic pipeline was applied to this set of WGS
327 data: BWA mem (version 0.7.12) for mapping the paired-end reads to the GRCh37 reference
328 (http://www.bcgsc.ca/downloads/genomes/9606/hg19/1000genomes/bwa_ind/genome/),
329 SAMtools (version 1.2) for file format conversion, duplicate marking with Picard (version
330 1.139), GATK for local re-alignment (version 3.4-46), and GATK HaplotypeCaller for variant
331 calling (version 3.4-46). Variants were soft-filtered using BCFTools (version 1.8) keeping
332 variants with at least 10 reads supporting the alternate allele and a max depth of 300. Filtered
333 variants were then annotated and normalized using SnpEff (version 4.11; gene version
334 GRCh37.75), VT (version 0.5772), and VCFAnno (version 0.2.8) (35). Filtered, annotated
335 variants are then converted into a GEMINI database (version 0.19.1) (36) using VCF2DB
336 (<https://github.com/quinlan-lab/vcf2db>). Specific GEMINI queries were performed for variants
337 under the autosomal dominant model, with details below. Scripts for processing the data, and
338 details regarding databases annotated against using VCFAnno and CNV analyses can be found
339 online (<https://github.com/Phillip-a-richmond/AnnotateVariants/tree/master/Strabismus>) (37).
340 Variants under the autosomal dominant model (shared between 011 and 013) located within the
341 linkage region were selected for further investigation. Reflecting the upper end of reported

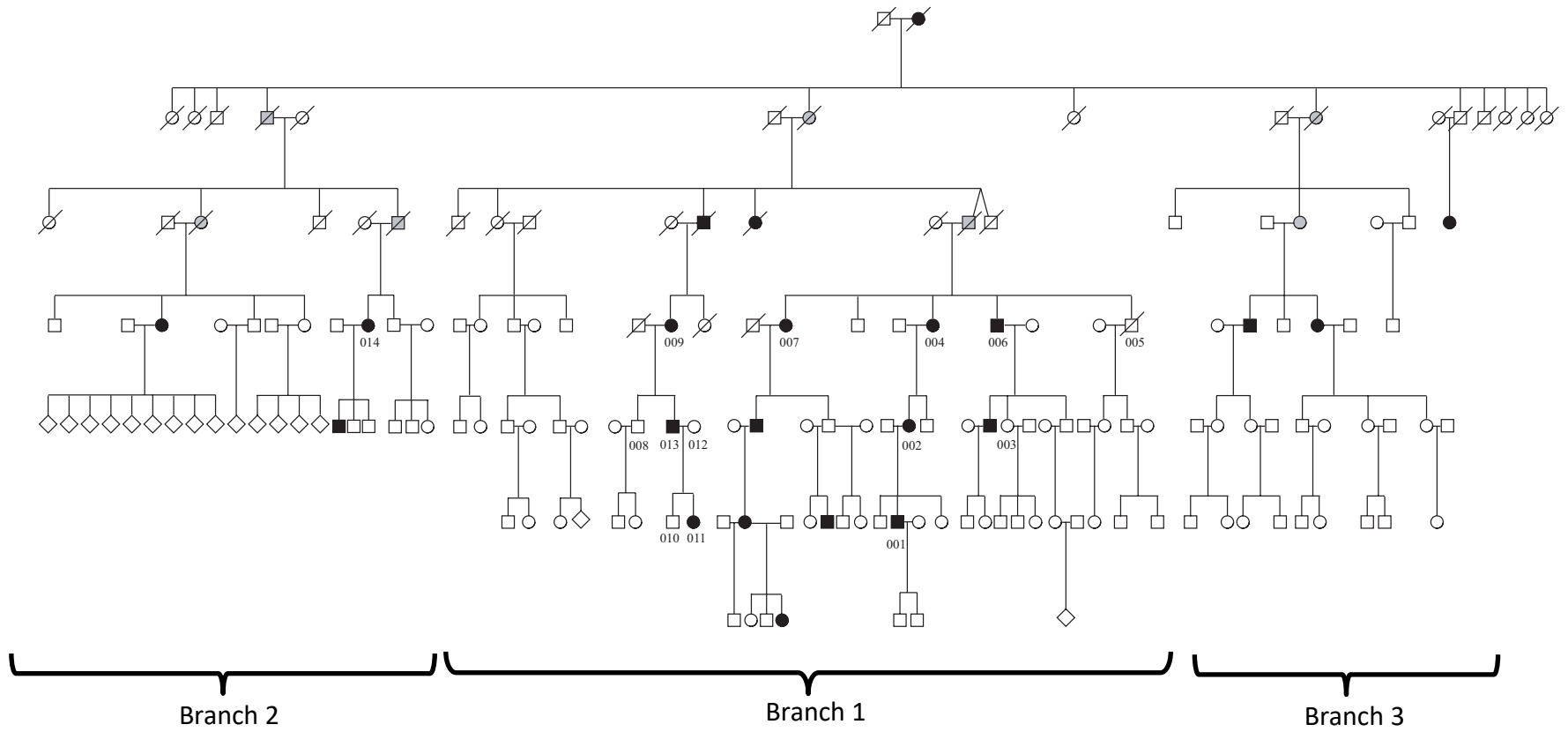
342 strabismus population frequency, all variants with allele frequency <4% were manually
343 reviewed. However, it is expected that dominant transmission of isolated strabismus is a rare
344 event, and therefore the results reported in this manuscript pertain to those variants with allele
345 frequency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

349 **Non-coding variant annotation and interpretation**

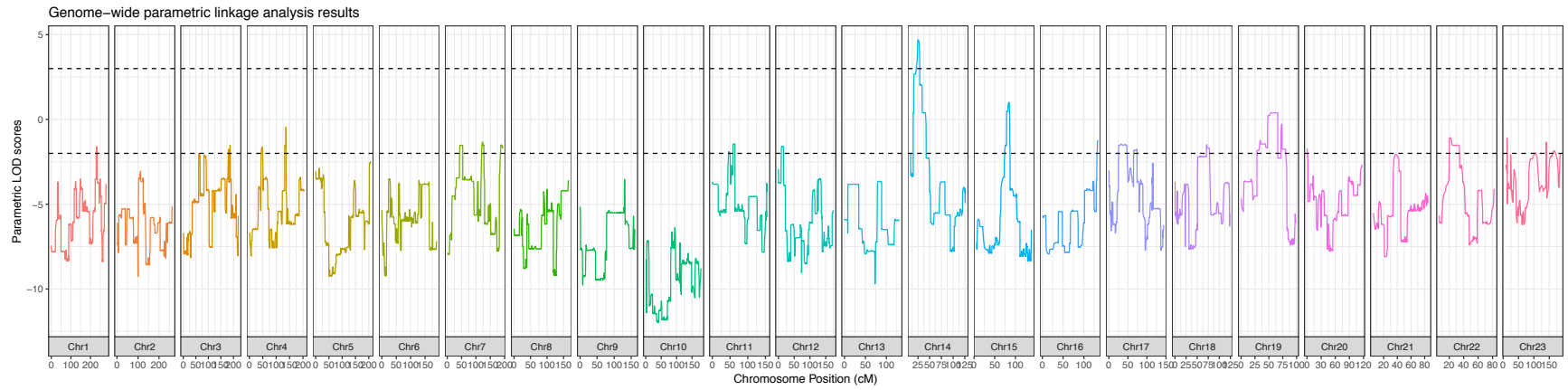
350 To enable analysis of non-coding variants, we used multiple databases and corresponding
351 bioinformatic tools to annotate such variants, including functional annotation of the mammalian
352 genome 5 (FANTOM5) database (38), JASPAR (22), Segway (39), RegulomeDB (40) and
353 Combined Annotation Dependent Depletion (CADD) (14). Data related to the candidate variants
354 were visualized in the UCSC Genome Browser. The 100 Vertebrates Basewise Conservation
355 PhyloP track was used for examining the conservation status. Due to the incomplete annotation
356 of noncoding RNAs across different species, Emsembl BLAST/BLAT was used to find potential
357 orthologs in other vertebrates, and RefSeq genes and ESTs were considered as evidence for RNA
358 transcripts (17). Based on the qualitative assessment, the top prioritized variant was confirmed to
359 be present by Sanger sequencing in all of the affected subjects and absent from unaffected
360 subjects.

361 **Acknowledgement**

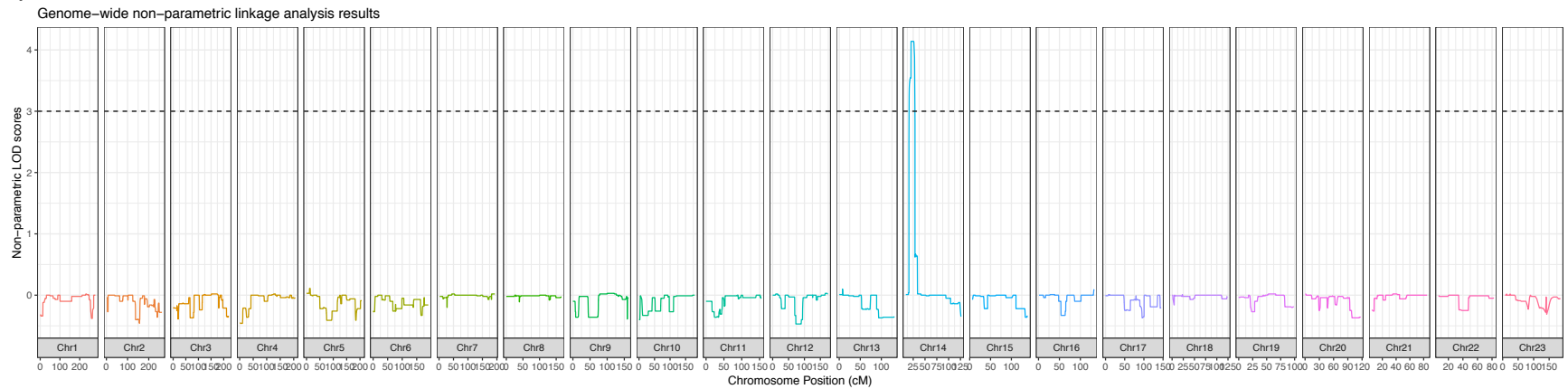
362 We thank Dr. Jonathan C. Horton (University of California San Francisco) for generously
363 offering his time to examine multiple participants in his clinic. We thank Ms. Dora Pak for
364 research management support. Funding for X.C.Y. is provided by a Frederick Banting and
365 Charles Best Canada Graduate Scholarship (GSD-146285) from the Canadian Institutes of
366 Health Research, BCCHRI-CIHR-UBC MD/PhD Studentship, and NSERC Discovery Grant
367 (RGPIN-355532-10, RGPIN-2017-06824). The SNP genotyping was performed under the
368 Care4Rare Canada Consortium funded by Genome Canada and the Ontario Genomics Institute
369 (OGI-147), the Canadian Institutes of Health Research, Ontario Research Fund, Genome
370 Alberta, Genome British Columbia, Genome Quebec, and Children's Hospital of Eastern Ontario
371 Foundation. C.J.R is supported by a Michael Smith Foundation for Health Research Scholar
372 Award.



a)



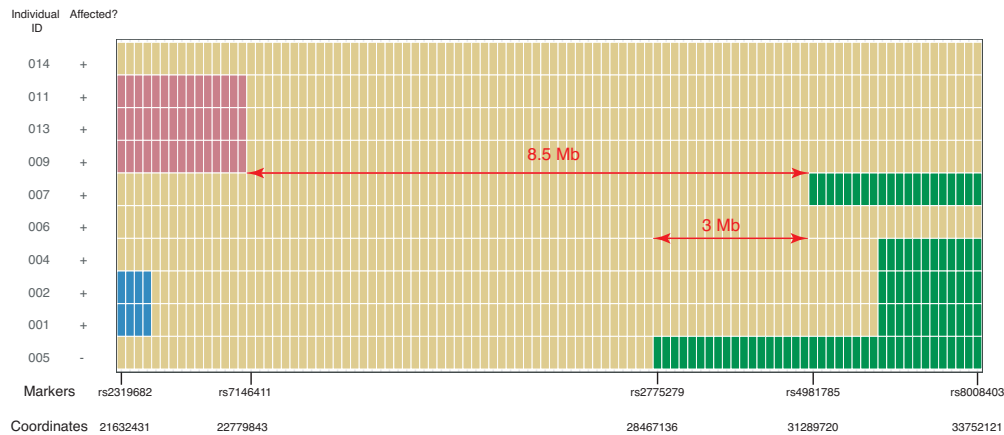
a)



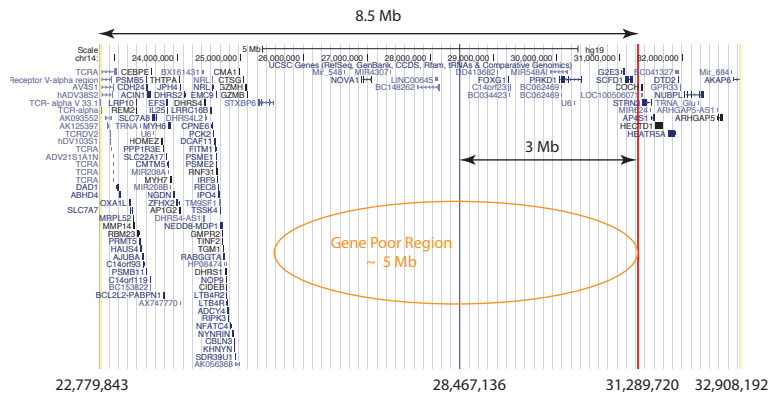
b)

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



a)



b)

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 colors indicate a different haplotype from a different ancestor. The 8.5 Mb and the core shared ~3 Mb 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.

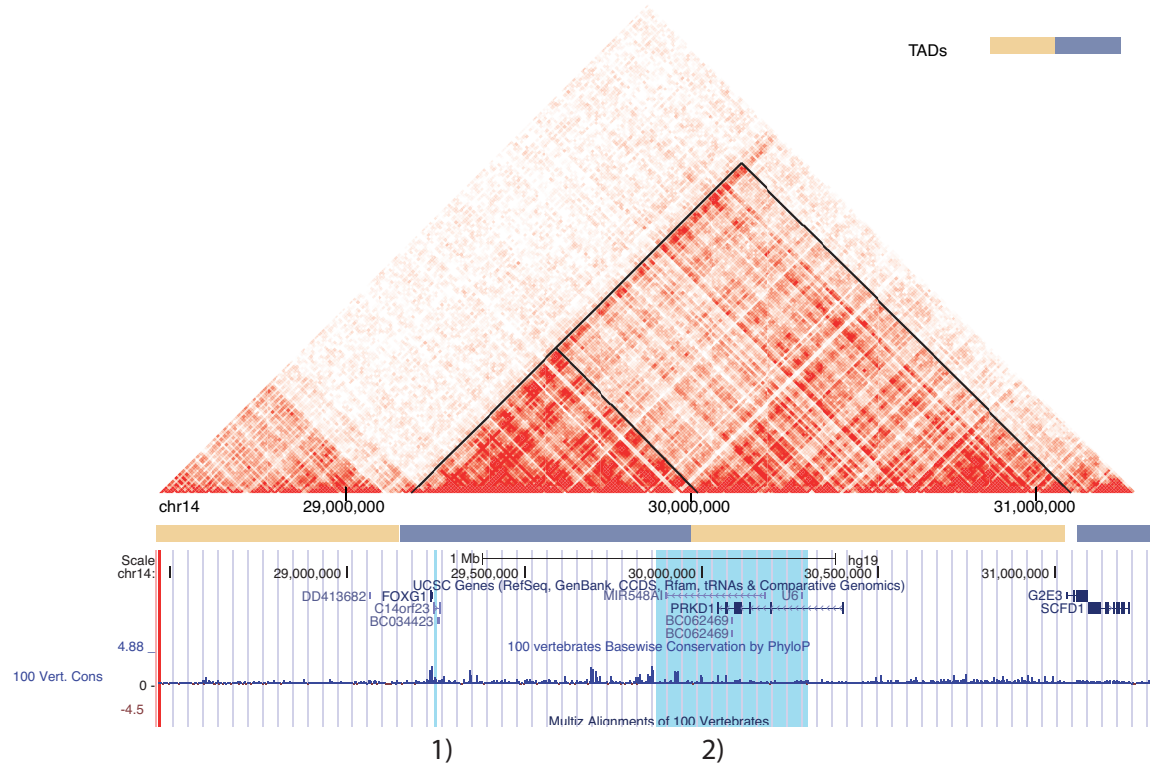


Figure 4. Topologically associating domains within the 3 Mb core region.

This heatmap illustrates the chromatin interaction based on Hi-C data (41). The deeper the red color, the stronger the intra-chromosomal interaction between corresponding segments of the DNA. FOXG1-TADs are indicated by the black triangle shapes. Three blue highlights from left to right correspond to the putative regulatory region within the FOXG1-TAD:

- 1) chr14:29247628 TAAAC > T
- 2) The SRO (smallest region of deletion overlap) regulation region affecting FOXG1 expression (21)

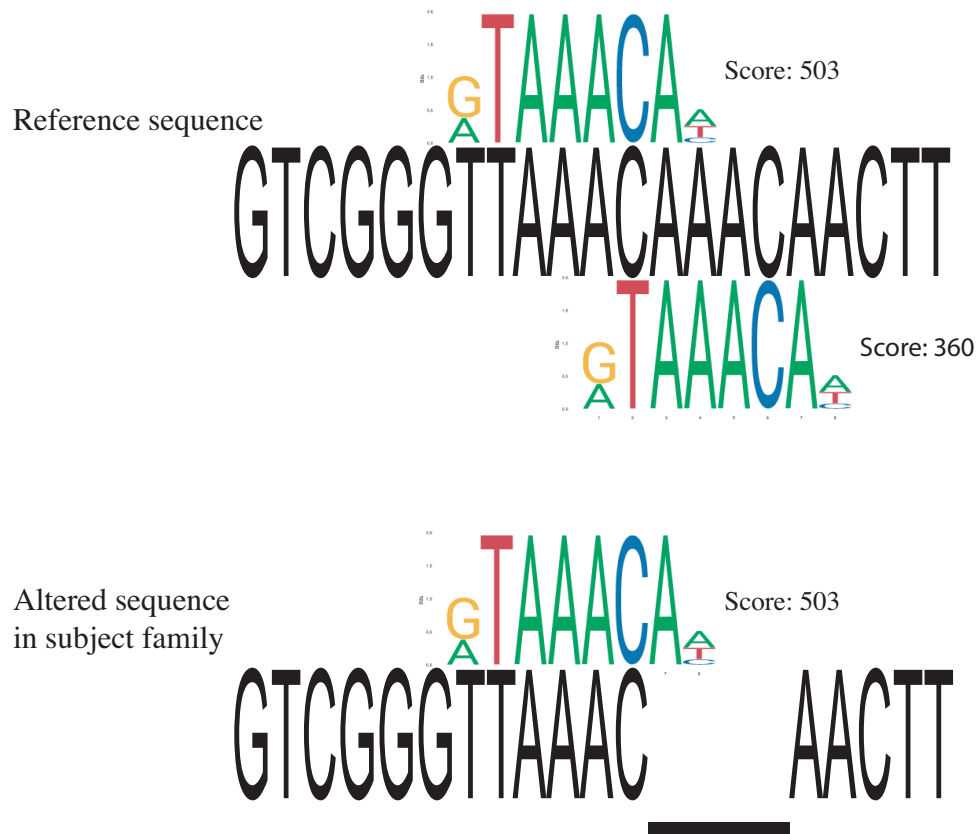
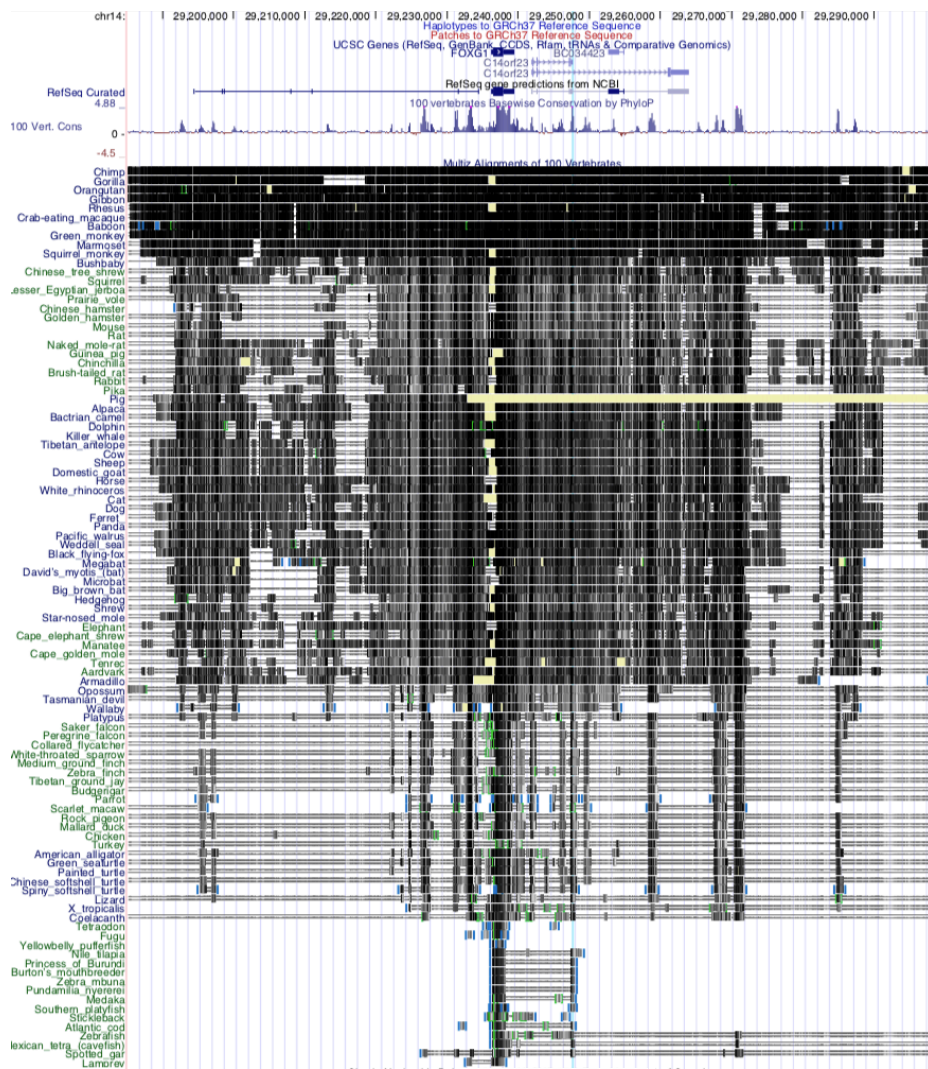


Figure 5. FOXG1 transcription factor binding site matching to reference and alternative sequence.

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>).



1) 2)

Figure 6. Ultra-conservation regions

- 1) FOXG1
- 2) The 4bp deleted region

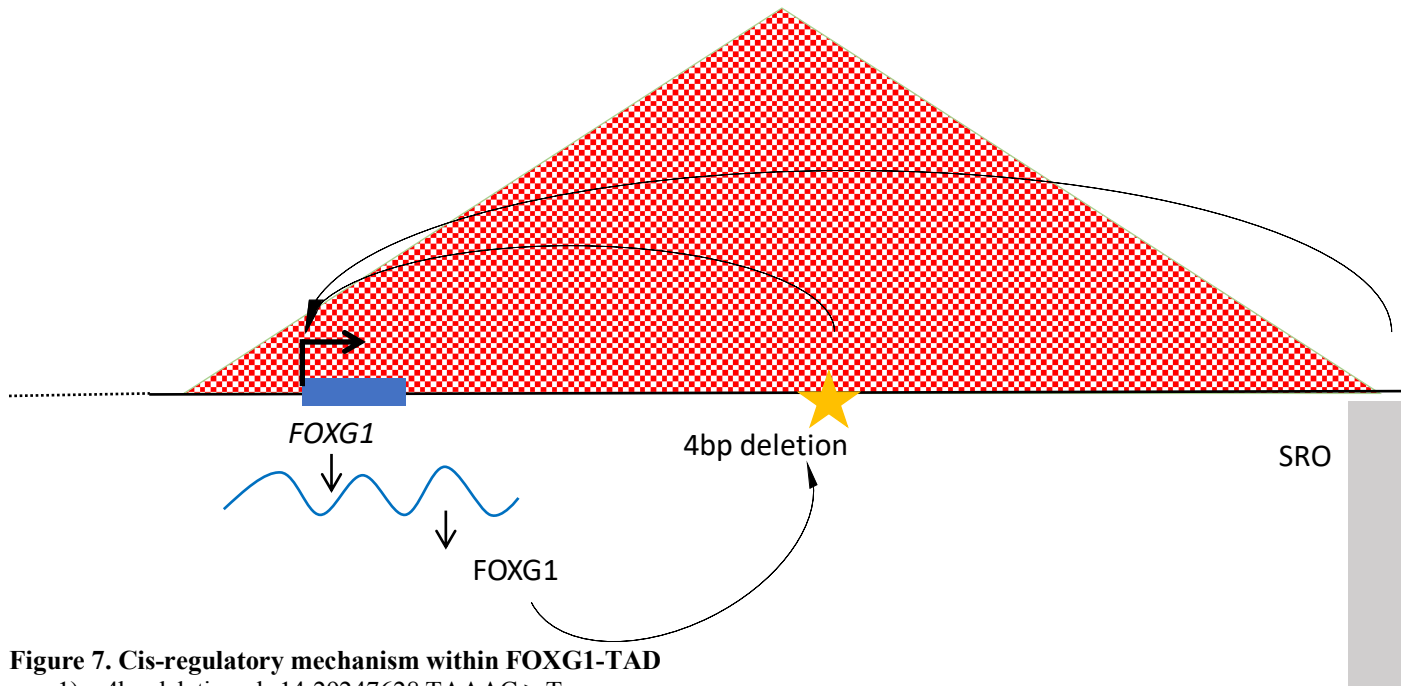


Figure 7. Cis-regulatory mechanism within FOXG1-TAD

- 1) 4bp deletion chr14:29247628 TAAAC > T
- 2) The SRO (smallest region of deletion overlap) regulation region affecting FOXG1 expression (21)

Supporting Information Legends

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.

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).

Table 1. Specific characteristics of strabismus across the affected descendants in the subject family.

Identifier	Type of reported strabismus	Age of onset	Severity	Eye movement full?	Concomitant?	Stereopsis	Strabismus surgery?	Intraocular pressure	Optic disc and macula	Other ocular conditions
001	Accommodative left esotropia with an A pattern	9 months	Severe	Y	Incomitant	Absent	bilateral strabismus surgery at age two and at age three; botulinum toxin injection at the age of 4	14 mm Hg OU	Normal	Mild amblyopia, left eye
002	Hypotropia and left exotropia. Consecutive exodeviation after initial esotropia surgery.	unclear	Severe	Y	NA	Absent	strabismus surgery at ages 3,8, and 16	16 mm Hg OU	Normal	No diplopia, suppression
004	Right esotropia	age of 3	Severe	Y	Concomitant	NA	strabismus surgery at ages 7 and 10	16 mm Hg OU	Normal	right cataract, right dense amblyopia
006	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
007	Hypertropia	unclear	cannot be ascertained	Y	cannot be ascertained	cannot be ascertained	NA	18 mm Hg OU	Optic disc normal; macular degeneration on the left eye and macular drusen on the right eye	cataract, diplopia, latent nystagmus, left eye suppression
009	Congenital right hypertropia; excyclotorsion	after 2, exact onset time unclear, aware of ocular misalignment at the age of 9	Mild	Y	Incomitant	Intact	N	16 mm Hg OU	Normal	Presbyopia
011	Childhood left esotropia	before the age of 2	Moderate	Y	Concomitant	Gross fusion & stereopsis	N	NA	Normal	Myopia
013	Congenital right hypertropia	unclear, aware of ocular misalignment at the age of 6	Mild	Y	Incomitant	Have the ability to use the 2 eyes together with stereopsis potential when the 2 images are artificially aligned	N	NA	Myopic optic discs	Myopia
014	Esotropia, right hypertropia; excyclotorsion in both eyes	at birth	Severe	Y	Incomitant	Absent	2 strabismus surgeries during 50s	NA	Normal	cataract, diplopia, latent nystagmus, left eye suppression

References

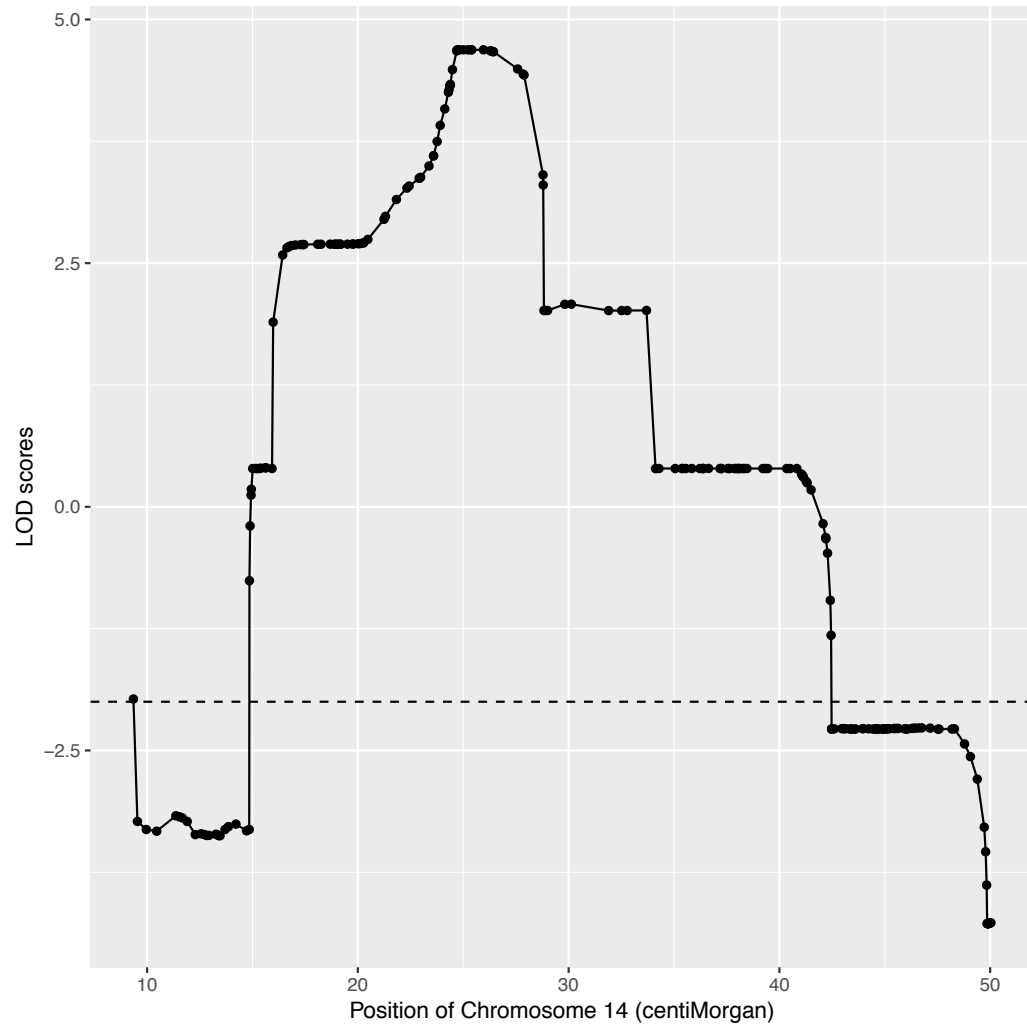
1. Demer JL. Neuroanatomical Strabismus. In: Pediatric Ophthalmology, Neuro-Ophthalmology, Genetics [Internet]. Springer Berlin Heidelberg; 2010. p. pp 59-75.
2. Lorenz B. Genetics of isolated and syndromic strabismus: facts and perspectives. *Strabismus*. 2002 Jun;10(2):147–56.
3. Ye XC, Pegado V, Patel MS, Wasserman WW. Strabismus genetics across a spectrum of eye misalignment disorders. *Clin Genet*. 2014 Aug;86(2):103–11.
4. Shaaban S, MacKinnon S, Andrews C, Staffieri SE, Maconachie GDE, Chan W-M, et al. Genome-Wide Association Study Identifies a Susceptibility Locus for Comitant Esotropia and Suggests a Parent-of-Origin Effect. *Invest Ophthalmol Vis Sci*. 2018 Aug 1;59(10):4054–64.
5. Mitter D, Pringsheim M, Kaulisch M, Plümacher KS, Schröder S, Warthemann R, et al. FOXP1 syndrome: genotype–phenotype association in 83 patients with FOXP1 variants. *Genetics in Medicine*. 2018 Jan;20(1):98–108.
6. Brooks BP, Zein WM, Thompson AH, Mokhtarzadeh M, Doherty DA, Parisi M, et al. Joubert Syndrome: Ophthalmological Findings in Correlation with Genotype and Hepatorenal Disease in 99 Patients Prospectively Evaluated at a Single Center. *Ophthalmology*. 2018 Dec 1;125(12):1937–52.
7. Chilton JK, Guthrie S. Axons get ahead: Insights into axon guidance and congenital cranial dysinnervation disorders. *Developmental Neurobiology*. 2017 Jul 1;77(7):861–75.
8. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*. 1995 Nov;11(3):241.
9. Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, et al. Exome Sequencing and the Management of Neurometabolic Disorders. *New England Journal of Medicine*. 2016 Jun 9;374(23):2246–55.
10. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016 Aug;536(7616):285–91.
11. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015 Oct;526(7571):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 Jul 1;40(W1):W452–7.
13. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010 Apr;7(4):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 Jan 8;47(D1):D886–94.
15. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012 Sep;489(7414):57–74.
16. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001 Jan 1;29(1):308–11.
17. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res.* 2018 Jan 4;46(D1):D754–61.
18. Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: key players in health and disease. *Trends in Genetics.* 2011 Jun 1;27(6):224–32.
19. Ferreira R da C, Oelrich F, Bateman B. Genetic aspects of strabismus. *Arquivos Brasileiros de Oftalmologia.* 2002 Mar;65(2):171–5.
20. 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.
21. Mehrjouy MM, Fonseca ACS, Ehmke N, Paskulin G, Novelli A, Benedicenti F, et al. Regulatory variants of *FOXP1* in the context of its topological domain organisation. *European Journal of Human Genetics.* 2018 Feb;26(2):186–96.
22. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 2018 Jan 4;46(D1):D260–6.
23. Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, et al. Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved *PAX6* Enhancer Causes Aniridia. *Am J Hum Genet.* 2013 Dec 5;93(6):1126–34.
24. Hébert JM, Fishell G. The genetics of early telencephalon patterning: some assembly required. *Nature Reviews Neuroscience.* 2008 Sep;9(9):678–85.
25. Patriarchi T, Amabile S, Frullanti E, Landucci E, Rizzo CL, Ariani F, et al. Imbalance of excitatory/inhibitory synaptic protein expression in iPSC-derived neurons from *FOXP1*^{+/-} patients and in *foxp1*^{+/-} mice. *European Journal of Human Genetics.* 2016 Jun;24(6):871–80.
26. Cargnin F, Kwon J-S, Katzman S, Chen B, Lee JW, Lee S-K. *FOXP1* Orchestrates Neocortical Organization and Cortico-Cortical Connections. *Neuron.* 2018 Dec 5;100(5):1083-1096.e5.

27. 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(2):126–34.
28. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010 Sep 2;467(7311):52–8.
29. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002 Jan;30(1):97–101.
30. 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(3):R25.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009 Jul 15;25(14):1754–60.
32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytisky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010 Sep;20(9):1297–303.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15;25(16):2078–9.
34. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 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 (Austin).* 2012 Jun;6(2):80–92.
35. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol.* 2016 01;17(1):118.
36. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol.* 2013 Jul;9(7):e1003153.
37. van Kuilenburg ABP, Tarailo-Graovac M, Richmond PA, Drögemöller BI, Pouladi MA, Leen R, et al. Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in GLS. *New England Journal of Medicine.* 2019 Apr 11;380(15):1433–41.
38. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015 Jan 5;16:22.
39. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Meth.* 2012 May;9(5):473–6.

40. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012 Sep;22(9):1790–7.
41. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biology.* 2018 Oct 4;19(1):151.

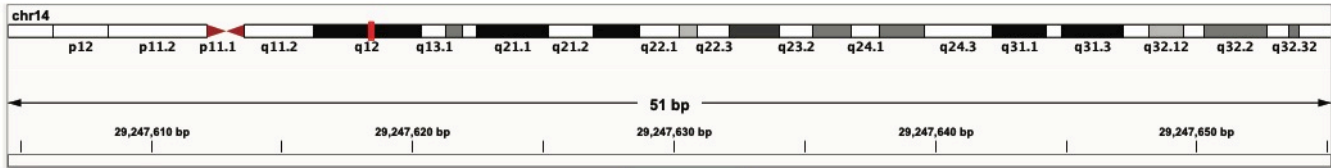
Supporting Information



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.

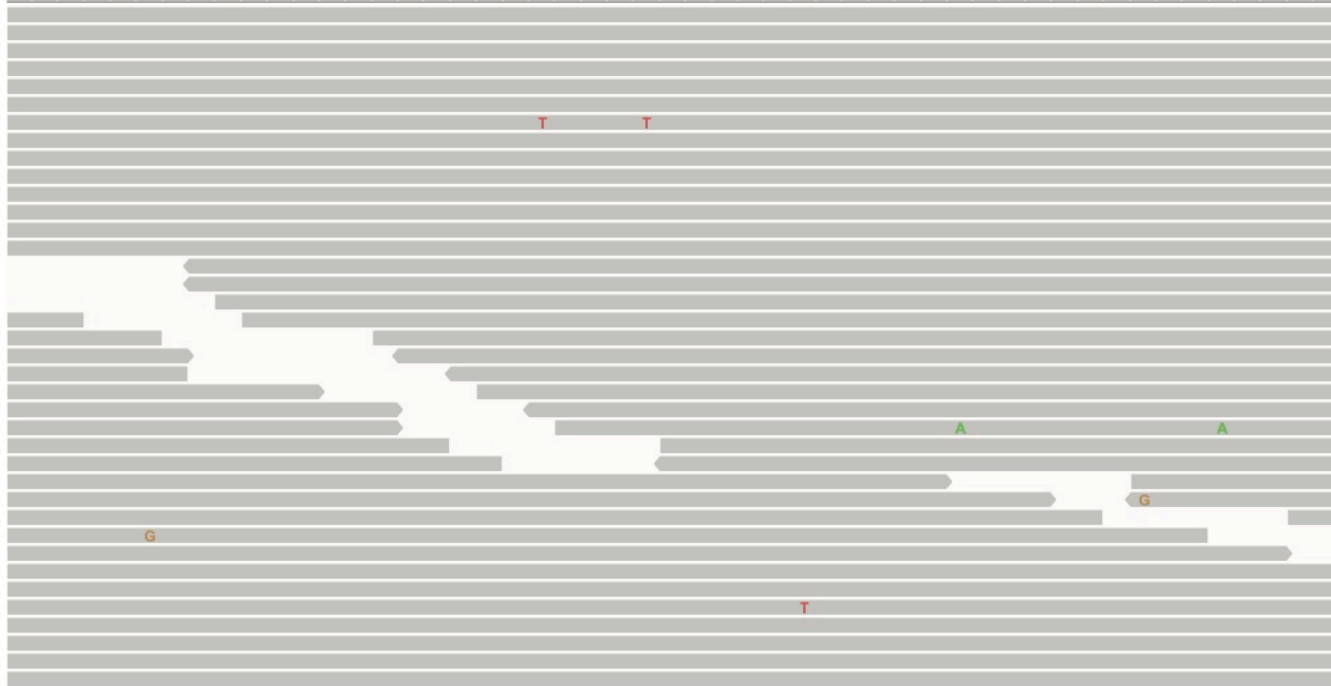
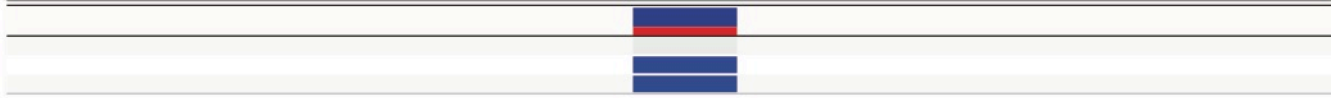


LINC01551



Variant Calls
(GATK HC)

Mother
Father
Proband



Mother (012)



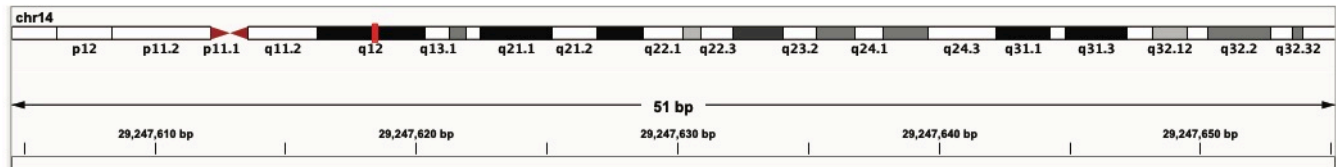
Father (013)



Proband (011)



LINC01551



Variant Calls
(GATK HC)

Mother
Father
Proband

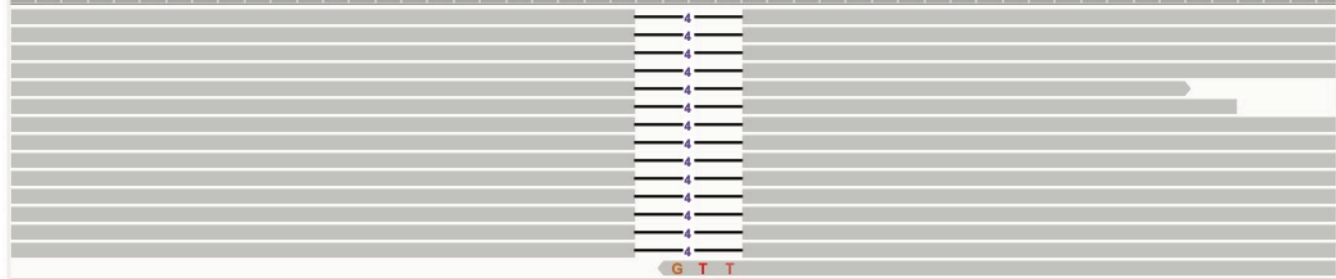
Mother (012)



Father (013)



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 T G T A C C C T G A C C A A A

LINC01551



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/>)