MAPPING A NEW LOCUS FOR NON-SYNDROMIC STRABISMUS
WITH HIGH-THROUGHPUT GENOME ANALYSIS

by

Xin Ye

B.Sc., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2014

© Xin Ye, 2014
Abstract

Eye misalignment, called strabismus, occurs in up to 5% of individuals. While misalignment is frequently observed in rare complex syndromes, the majority of strabismus cases are non-syndromic. Over the past decade, genes and pathways associated with syndromic forms of strabismus have emerged, but the genes contributing to non-syndromic strabismus remain elusive. Non-syndromic strabismus is highly heterogeneous, and different loci have been inferred from previous genetics studies. Only a single strabismus locus, STBMS1, on chromosome 7 has been confirmed in more than one family, but the reported inheritance patterns of this locus with disease conflict and no specific variant has been proposed. Here, I analyzed a large non-consanguineous family with multiple individuals affected by strabismus across seven generations. The hypothesis is that a single variant is responsible for the non-syndromic strabismus in this particular family displaying dominant patterns of inheritance. Whole exome sequencing (WES) was performed to uncover large-blocks of variations within protein-coding regions of the genome shared by two affected distant relatives. In parallel, chromosome regions segregating with the strabismus phenotype in the family were identified using linkage analysis on 12 individuals. Linkage analysis identified one specific risk locus of high confidence. Based on the lack of protein-coding alterations in the locus, whole genome sequencing (WGS) was performed to find additional shared candidate causal variants. Combining the available information, a 10 Mb region on chromosome 14 was identified with high confidence that it was associated with strabismus, within which a set of potential regulatory sequence alterations have been highlighted for further study. This study represents the first identified locus for autosomal dominant, non-syndromic, strabismus. The project utilizes next-generation sequencing (NGS), linkage
analysis, and bioinformatic analyses to prioritize and select both coding and non-coding variants, demonstrating the effectiveness of combining NGS and classical genetic approaches. The research findings improve our understanding of strabismus genetics and defines multiple paths for future research, family-specific genetic testing for early diagnosis, and consequent preventive therapy.
Preface

A version of chapter 1 has been published. (Ye, X.C., Pegado, V., Patel, M.S., and Wasserman W.W. (2014) Strabismus genetics across a spectrum of eye misalignment disorders. *Clinical Genetics*). I conducted literature research on strabismus and wrote the draft manuscript. V. Pegado and M.S. Patel provided guidance on clinical perspectives of strabismus and edited the manuscript. W.W. Wasserman supervised the research and extensively edited the manuscript.

Dr. J. Horton (University of California, San Francisco), Dr. S. Narasimhan, and Dr. V. Pegado ascertained participants. Chromosomal microarray (CMA) testing for index was performed and analyzed by the Cytogenetic laboratory at Children’s and Women’s Health Center. DNA samples for next-generation sequencing (NGS) and microarray genotyping were prepared by M. Higginson, Z. Zong, and myself. The linkage analysis was a collaborative effort with the Hospital for Sick Children in Toronto (SickKids) through the Finding of Rare Disease Genes Canada (FORGE Canada) network and performed by N. Roslin. WES and WGS data analyses were performed by myself, with assistance from C. Shyr and other Wasserman lab members for computational support. The variant density plot script was programmed by Dr. E. Nosova. Allele-specific PCR was designed and performed by myself, while standard PCR was designed and performed by X. Han.

The collection of the samples for these studies was approved by the University of British Columbia Children’s & Women’s Research Ethics Board, approval number CW10-0317/H10-03215.
Table of Contents

Abstract ........................................................................................................................................ ii
Preface ........................................................................................................................................ iv
Table of Contents .......................................................................................................................... v
List of Tables ................................................................................................................................ viii
List of Figures ............................................................................................................................... ix
List of Abbreviations .................................................................................................................... xi
Acknowledgements ..................................................................................................................... xii

Chapter 1: Introduction .................................................................................................................. 1

1.1 Etiology and pathogenesis ....................................................................................................... 3
1.2 Risk factors ............................................................................................................................ 5
1.3 Family and twin studies ......................................................................................................... 6
1.4 Genetic mechanisms ............................................................................................................. 8
1.5 Linkage analysis .................................................................................................................... 8
1.6 Gene expression studies ....................................................................................................... 10
1.7 Duane retraction syndrome .................................................................................................. 11
  1.7.1 Chromosome 8q and type 1 DRS .................................................................................. 12
  1.7.2 CHN1 and type 2 DRS ............................................................................................... 13
  1.7.3 Type 3 DRS ................................................................................................................ 13
  1.7.4 Okihiro syndrome ......................................................................................................... 14
1.8 Animal models ..................................................................................................................... 15
  1.8.1 Famous strabismic animals ......................................................................................... 15
1.8.2 Cats ........................................................................................................... 15
1.9 Conclusion and future directions .................................................................. 17
1.10 Research objectives .................................................................................... 17
1.11 Study structure ............................................................................................ 18

Chapter 2: Project Description ............................................................................ 23

2.1 Introduction .................................................................................................... 23
2.2 Materials and methods .................................................................................. 25
  2.2.1 Samples .................................................................................................... 25
  2.2.2 Phenotyping .............................................................................................. 25
  2.2.3 DNA preparation ...................................................................................... 26
  2.2.4 Instrumentation ....................................................................................... 26
  2.2.5 Next-generation sequencing data processing and variant calling ............ 26
  2.2.6 WES variant prioritization ....................................................................... 27
  2.2.7 WES variant selection ............................................................................ 27
  2.2.8 Confirming variant presence ..................................................................... 28
  2.2.9 WES variant density plot ......................................................................... 29
  2.2.10 Linkage analysis .................................................................................... 29
  2.2.11 WGS variant prioritization .................................................................... 31
  2.2.12 WGS variant selection and confirmation ................................................ 31
2.3 Results .......................................................................................................... 32
  2.3.1 The index case .......................................................................................... 32
  2.3.2 Pedigree ................................................................................................... 32
  2.3.3 Phenotyping ............................................................................................. 33
2.3.4 Whole exome sequencing analysis ................................................................. 34
2.3.5 WES variant validation ...................................................................................... 35
2.3.6 WES variant density plot .................................................................................. 36
2.3.7 Linkage analysis ................................................................................................. 38
2.3.8 WGS analysis ..................................................................................................... 39
2.3.9 WGS variant prioritization ............................................................................... 40
2.4 Discussion ............................................................................................................... 42

Chapter 3: Future Directions ..................................................................................... 67

References .................................................................................................................. 72
List of Tables

Table 1.1  Selected syndromic strabismus associated genes and loci.......................... 19
Table 2.1  Coverage and variants from whole exome sequencing (WES). .................... 46
Table 2.2  Non-tolerant shared variants based on WESs from two distant relatives......... 46
Table 2.3  Selected syndromic strabismus associated loci. ......................................... 47
Table 2.4  Loci and corresponding hg19 coordinates associated with strabismus, either syndromic or non-syndromic. .................................................................................. 49
Table 2.5  Number of shared and total variants and log2(ratio) for each chromosome in strabismic and control groups. ......................................................................................... 50
Table 2.6  Number of shared variants between different individuals on chr14:20,000,000-35,000,000.................................................................................................................... 51
Table 2.7  Sequencing data quality for whole genome sequencing (WGS). .................... 52
Table 2.8  Coverage and variants of linked region from WGS with the BGI pipeline....... 52
Table 2.9  Coverage and variants of linked region from WGS with the in-house pipeline... 52
Table 2.10 Prioritized WGS variants from the linked region. ........................................... 53
List of Figures

Figure 1.1  A schematic representation of EOMs and nerve innervation with associated genes. ........................................................................................................................................... 20

Figure 1.2  Defects along the visual-motor axis can contribute to infantile esotropia .... 21

Figure 1.3  Overview of the strabismus genetic study. ................................................................................................................................. 22

Figure 2.1  Pedigree for the subject family with non-syndromic strabismus. .......... 54

Figure 2.2  Branch 1 of the subject family with non-syndromic strabismus. .......... 55

Figure 2.3  Simplified branch 1 showing the genotyped individuals (with study ID) and ancestors required to link them. ................................................................................................................................. 56

Figure 2.4  Branch 2 of the subject family. ................................................................................................................................. 57

Figure 2.5  Branch 3 of the subject family. ................................................................................................................................. 58

Figure 2.6  Sanger sequence results of GTF2IRD2 variant for multiple individuals. .... 59

Figure 2.7  Variant density plot of shared variants on chromosome 11. .......... 60

Figure 2.8  Variant density plot of shared variants on chromosome 14. .......... 61

Figure 2.9  Simulated maximum LOD scores under the alternative hypothesis of linkage, under a range of genetic models. ................................................................................................................................. 62

Figure 2.10 Screenshot of region chr14: 20,000,000 – 35,000,000 from UCSC genome browser. The linked region chr14: 22,779,843 – 32,908,192 is gene-poor. .......... 63

Figure 2.11 Venn diagram of WGS analysis for 001, 013, and 014 based on the BGI pipeline. ................................................................................................................................. 64

Figure 2.12 Venn diagram of WGS analysis for 001, 013, and 014 based on the in-house pipeline. ................................................................................................................................. 65
Figure 2.13  Single nucleotide variant chr14:28052726 A > C impacts binding score for FoxA1 transcription factor binding site.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS-PCR</td>
<td>allelic specific polymerase chain reaction</td>
</tr>
<tr>
<td>CADD</td>
<td>Combined Annotation Dependent Depletion</td>
</tr>
<tr>
<td>CFEOM</td>
<td>congenital fibrosis of extraocular muscles</td>
</tr>
<tr>
<td>CMA</td>
<td>chromosomal microarray</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variation</td>
</tr>
<tr>
<td>DRS</td>
<td>Duane retraction syndrome</td>
</tr>
<tr>
<td>EOM</td>
<td>extraocular muscles</td>
</tr>
<tr>
<td>EVS</td>
<td>Exome Variant Server</td>
</tr>
<tr>
<td>FANTOM5</td>
<td>functions annotation of the mammalian genome 5</td>
</tr>
<tr>
<td>FORGE Canada</td>
<td>Finding of Rare Disease Genes Canada</td>
</tr>
<tr>
<td>GSC</td>
<td>Michael Smith Genome Sciences Centre</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome wide association studies</td>
</tr>
<tr>
<td>IBD</td>
<td>identical by descent</td>
</tr>
<tr>
<td>IGV</td>
<td>Integrative Genomic Viewer</td>
</tr>
<tr>
<td>MeSHOPs</td>
<td>Medical Subject Heading Overrepresentation Profiles</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>SickKids</td>
<td>Hospital for Sick Children in Toronto</td>
</tr>
<tr>
<td>TFBS</td>
<td>transcription factor binding site</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
<tr>
<td>WES</td>
<td>whole exome sequencing</td>
</tr>
<tr>
<td>WGS</td>
<td>whole genome sequencing</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to acknowledge my supervisors, Drs. Millan Patel and Wyeth Wasserman, for giving me a chance to pursue my research interest, sharing their scientific enthusiasm, and supporting me through my graduate studies. I would also like to thank my committee, Drs. Anna Lehman and Colin Ross, for their advice and warm support. To all the current and past lab members who have helped me in one way or another, thank you for sharing your time generously. I owe a particular thanks to Dora who provided research management support, and Dr. Maja Tarailo-Gaovac who made key suggestions to improve the written thesis. To my MedGen peers, my gratitude for the suggestions and encouragement throughout my study. To my family, thank you for believing in me and supporting me. For financial support, my thanks to the University of British Columbia Medical Genetics Graduate Program and the Child & Family Research Institute.
Chapter 1: Introduction

Eye misalignment, called strabismus occurs in up to 5% of individuals\(^1\). While misalignment is frequently observed in rare complex syndromes, the majority of strabismus cases are non-syndromic. Over the past decade, genes and pathways associated with syndromic forms of strabismus have emerged, but the genes contributing to non-syndromic strabismus remain elusive. Genetic testing for strabismus risk may allow for earlier diagnosis and treatment, as well as decreased frequency of surgery. We review human and model organism literature describing non-syndromic strabismus, including family, twin, linkage, and gene expression studies. Recent advances in the genetics of Duane retraction syndrome are considered, as relatives of those impacted show elevated familial rates of non-syndromic strabismus. As whole genome sequencing efforts are advancing for the discovery of the elusive strabismus genes, this overview is intended to support the interpretation of the new findings.

Strabismus (eye misalignment) is one of the earliest recorded genetic disorders. More than 2400 years ago, Hippocrates observed ‘Children of parents having distorted eyes squint also for the most part’\(^1\). Strabismus can cause visual problems during development, including loss of binocular vision, amblyopia (‘lazy eye’), and abnormal retinal correspondence (shifting of the fixation point relative to the macula in one eye). Strabismus disrupts stereopsis, which impacts the performance of numerous practical tasks requiring the precise judgment of distance (e.g. driving) or depth (e.g. microscopy)\(^2\). In addition to reduced visual function, strabismus is associated with psychosocial problems impacting self-image, interpersonal relationships, performance in school and employment\(^3\). Children as
young as 5 years display a reduced tendency to interact with peers with noticeable strabismus
4,5. Strabismus negatively impacts employment rates and thus economic status 6. Strabismus
surgery has positive impact on quality-adjusted life years (QALY), increasing QALY by
2.61, while being highly cost-effective ($1632/QALY) 7. While non-surgical intervention
therapies (e.g. patching) in young children have not been similarly quantified, such practice
is intended to reduce the need for surgical intervention.

The prevalence of strabismus is 2–4% among Caucasians, 2.4% among
Hispanic/Latinos, 2.5% among African-Americans, and 1% in East-Asians 8–11. Among
Caucasians, esotropia (inward misalignment) is three times more common than exotropia,
while exotropia predominates in Cameroon black (63% of cases) and Asian populations
(more than 70% of cases) 12–15. Studies consistently report balanced distribution between
genders 16–19. In most cases, non-syndromic strabismus is characterized by non-restrictive,
non-paralytic ocular misalignment with the same magnitude in all directions of gaze, which
is known as concomitant (comitant) strabismus. Incomitant strabismus is paralytic in origin
and the angle of deviation varies in different directions. The occurrence of muscle paralysis
can be determined by the broad H test, which is scored positive if one eye lags behind the
other in at least one of the six positions of gaze 20.

While the causes of non-syndromic strabismus are largely unknown, twin studies and
family studies have demonstrated a substantial genetic contribution to strabismus 21.
Although the heritability of strabismus has long been recognized, most advances at the level
of specific genes have occurred during the past decade 8,12. Thus far, only a single non-
syndromic strabismus locus on chromosome 7 has been confirmed to act in more than one
family, and in those families the specific causal alterations have not been determined.
In this review, we summarize strabismus etiology and pathogenesis, genetic studies of non-syndromic strabismus and Duane retraction syndrome (DRS), as non-syndromic strabismus occurs at elevated rates in affected families\textsuperscript{22}, and describe model organism studies related to genetic forms of strabismus.

1.1 Etiology and pathogenesis

The mechanisms underlying strabismus may involve one of several systems or tissues (Figure 1.1). Past reports highlight the potential for disruptions in extraocular muscles (EOM), orbital connective tissues, cranial nerves, fusion centers, and the visual cortex\textsuperscript{23}. The position of the eye is determined by all the five components. Mechanical trauma, acquired inflammation or infiltration, and metabolic disorder can all lead to EOM myopathy and secondary strabismus. Abnormalities of either the location or stability of the connective tissue pulleys alter the direction of EOM pulling and contribute to both congenital and acquired strabismus. Congenital cranial dysinnervation disorders (CCDDs) have been associated with hypoplastic or misrouted motor nerves to EOMs, and additional cranial nerve abnormalities have been observed\textsuperscript{23}. Fusion centers include a convergent center at the rostral–dorsal midbrain and a divergence center that, based on acute onset of concomitant esotropia related to tumors, is likely situated in the hindbrain\textsuperscript{24,25}. Animal experiments show that abnormal early visual experience can lead to strabismus and cause changes in metabolic activity in the visual cortex\textsuperscript{26}.

The age of onset distribution for strabismus is bimodal, with approximately 22% diagnosed before the age of 12 months and approximately 43% detected between 2 and 3 years of age. Non-accommodative strabismus was more common in the first group, while
accommodative strabismus was more common in the second group (where accommodative refers to strabismus arising with altered visual acuity) \(^{27}\). Approximately 26% of first-degree relatives of patients with hypermetropic (far-sighted) accommodative esotropia were affected with strabismus \(^{28}\), suggesting that individuals with inherited hypermetropia may be predisposed to strabismus. However, a recent study demonstrated that heritability of strabismus was independent of refractive error. Bivariate analysis indicated a phenotypic correlation of only 0.20 between refractive error and eso-deviation, including tropia (constant eye misalignment) and phoria (latent eye misalignment); in other words genetic contributions to strabismus and hypermetropia are largely independent \(^{29}\).

As indicated above, pathogenesis of infantile esotropia may result from defects spanning the visual-motor axis (Fig. 1.2) \(^{30}\). Researchers have postulated about the relationship between strabismus and changes in the visual cortex. At the turn of the 20th century, Worth proposed that infantile esotropia was due to an inborn defect of fusion, as surgery on EOM could not reverse strabismus \(^{31}\). Tychsen suggested that this fusion faculty was situated within the striate cortex, and specifically proposing that congenital defects would therefore be present in disparity-sensitive, binocular neurons \(^{30}\). Using staining techniques, a paucity of such binocular connections was observed in both natural and induced strabismic monkeys while monocular connections remained. Electrophysiological measurement showed that loss of binocular responsiveness and disparity sensitivity was consistent with the reduced number of binocular connections \(^{32}\).

Hypotheses for strabismus mechanisms have been proposed which focus on the subcortical visual pathway, brainstem vergence motoneurons, the brainstem vestibule-ocular pathway, and cranial nerves \(^{32}\). On the other end of the visual-motor axis, Chavasse proposed
a ‘motor’ hypothesis, suggesting that abnormal optical input, such as weakness of the EOM, may impede development of binocular fusion thus leading to strabismus. He argued that surgery in the very young age to restore eye alignment could rescue binocular vision\textsuperscript{32,33}. Clinical data showed that shorter durations of misalignment correlated with better stereopsis, implying that muscle abnormalities lead to poor stereopsis, not vice versa\textsuperscript{33}. Examination of strabismic EOM identifies some abnormalities. A 2012 magnetic resonance imaging study of 12 concomitant esotropes and 13 controls demonstrated rectus muscle enlargement. Cross sections of medial rectus muscle were up to 39\% larger (p < 0.005), and those of lateral rectus muscle were up to 28\% larger in the esotropic cases. Moreover, medial rectus contractility was 60\% higher in exotropic individuals (p < 0.005)\textsuperscript{34}. It is inconclusive, however, whether the structural changes in EOMs are the cause of strabismus or merely reflect the adaptation to the change of motoneuron firing patterns, as observed in other skeletal muscle tissue\textsuperscript{35}. Schoeff et al. reasoned that the lack of evidence of EOM denervation or dysinnervation in non-syndromic strabismus suggested a visual cortex contribution\textsuperscript{34}. As live imaging technology advances, higher resolution examination may advance our understanding of the relative contribution of defects in muscle and nerves to the strabismus phenotype.

1.2 Risk factors

Significant strabismus risk factors include retinopathy of prematurity, low birth weight, premature birth, and smoking during pregnancy. As our focus will remain on genetic risk, the interested reader may find additional information about the other factors in the systematic review by Maconachie et al.\textsuperscript{19}.
1.3 Family and twin studies

Many early studies focused on the transmission of strabismus through families. However, findings varied in terms of heritability, inheritance mode, and the concordance of strabismic types. Surveys conducted between 1910 and 1950 indicated that hereditary factors ranged from 20% to 50% in families with esotropia. Schlossman and Priestley found that 47.5% of 158 patients with strabismus, 48.9% of 139 esotropes, and 36.8% of 19 exotropes belonged to families with two or more additional affected members. The authors suggested that the actual number might be larger since subtle alignment deviations could be missed. The highest reported familial incidence of strabismus was 65%.

A longitudinal study found that 18% of 34 babies born in families with a parent affected by convergent (i.e. esotropia) strabismus developed constant or intermittent esotropia by 6 months. As the types of assessed relatives varied between studies and there was no consideration of environment, the precise genetic risk is unclear. Nevertheless, the figures were much higher than those in general population (approximately 5%), supporting a contribution of genetics to strabismus risk. The concordance of strabismus types varied across the studies. Families with a mixture of esotropia and exotropia phenotypes were reported. One study found that 80% of strabismus cases occurring in the same family were concordant. Another study reported 54% concordance within 39 studied families.

As familial clustering of strabismus can reflect either a common genetic factor or an unrecognized environmental factor, twin studies are the key to quantify the relative genetic contribution. Twin studies of strabismus have reported higher concordance rates in monozygotic twins than dizygotic twins, suggesting a predominant genetic factor. Matsuo et al.’s twin study showed that strabismic subtypes of 67.3% of 49 pairs or sets were
concordant, and the concordance rate was higher in monozygosity (82.4%) than in multizygosity (47.6%) \(^{39}\). Wilmer and Backus performed a meta analysis, reporting monozygosity and dizygosity concordances of 54% and 14%, respectively, in studies with systemic ascertainment; and 66% and 19%, respectively, without systematic ascertainment \(^{40}\). This contradicted with Paul and Hardage’s 1994 study, but Wilmer and Backus observed that a translation error in the 1994 study led to an overestimation of dizygosity concordance \(^{40,41}\).

Podgor reported that the odds ratio for esotropia rose from 2.6 if a sibling from a preceding birth was affected to a ratio of 5.4 if a twin (or other multiple birth) was affected \(^{21}\).

Esotropia and exotropia have a strikingly different genetic risk profile. In the Podgor study, a striking odds ratio of 330 was reported for exotropia in cases of multiple birth with one affected twin, while single births had an extremely low odds ratio of 2.2, data most consistent with a strong multiple birth environmental impact on exotropia risk \(^{21}\). A study with 1462 twins suggested that genetic heritability was specific to esotropia, reporting that heritability of eso-deviation was 64% while no heritability was detected for exo-deviation \(^{29}\). Exotropia (75%) had higher observed concordance than esotropia (65.7%) in a Chinese twin study, which may reflect influence of both the multiple birth environmental influence on esotropia and potential ethnic differences in the genetic contribution to esotropia \(^{19,42}\).

A key consideration arises from twin studies. Wilmer and Backus raised the potential confounding contribution of phoria to the study of strabismus genetics. Phoria is a latent misalignment of the eyes that appears when fixation on a target is broken (which can be revealed with a cross-cover test). Wilmer and Backus observed that genetic factors were necessary for strabismus development but not for phoria development \(^{40}\). Phoria cases have...
been noted in families with strabismus, and a portion of strabismus genetics studies have included phoria as positive cases \(^{13}\).

Summarizing the above information, esotropia is most closely tied to heritable factors while exotropia has a stronger environmental component. Future studies should therefore be designed in a manner that controls for the environmental component, including multiple births.

### 1.4 Genetic mechanisms

Dominant, recessive, and sex-linked inheritance patterns have been proposed for non-syndromic strabismus in family studies \(^{19,37}\). In different families, Czellitzer reportedly suggested two recessive genes were responsible for strabismus, while Waardenburg proposed a model of a single autosomal gene \(^{37,43}\). A study using quantitative measurement of sensory and motor function rejected the theories of Mendelian inheritance of strabismus as a single trait \(^{14}\). The majority of studies have noted that simple Mendelian models cannot explain the complexity of strabismus inheritance patterns. There are multiple genetic mechanisms represented in the families described in the scientific literature. Furthermore, the high frequency of strabismus may confound family studies with some cases likely arising from environmental mechanisms. Without accurate categorization based on exquisite pathological characterization of the strabismus, and given the diversity of potential physical mechanisms, such conflicting results are not entirely unexpected.

### 1.5 Linkage analysis

Parikh et al. identified the first concomitant strabismus locus on chromosome 7p22.1 (STBMS1) in a linkage analysis of a large family. Among seven initially assessed multiplex families with non-syndromic strabismus, one family showed a significant logarithm of the
odds (LOD) score on chromosome 7. Although the pedigree suggested an autosomal dominant inheritance pattern, the haplotype data was most consistent with an autosomal recessive model or a more complex model, such as the authors’ proposed semi-dominant inheritance model. The autosomal recessive inheritance model has been subject to discussion. The other six families in the original study were not consistent with the chromosome 7 loci contributing. In the subject family, eight of fourteen siblings were affected, and seven of these eight patients had hypermetropia of varying severity. Rice et al. examined 12 additional families, of which one was consistent with an STBMS1 role. Five affected family members had primary non-syndromic comitant esotropia while 21 examined family members were unaffected. In this second STBMS1 family, the pattern of inheritance best fits a dominant mode of inheritance. In combination the reports indicate that there is at least one non-syndromic strabismus associated genetic component at the STBMS1 locus. Elucidating the causal mutations in the two families may clarify the conflict between transmission models.

The Ohtsuki group tried to identify comitant strabismus susceptibility loci through sib-pair analysis and non-parametric linkage analysis for multiple pedigrees. This initial 2003 attempt indicated multiple loci with low LOD scores. A 2008 report identified 4q28.3 and 7q31.2 loci as having significant evidence of linkage. After stratifying cases into esotropia and exotropia subgroups, they identified additional loci at 8q24.21 and 14q21.3, respectively.

A summary of reported candidate loci for comitant strabismus is presented in Table 1.1. Based on the range of findings, it appears likely that multiple genes are contributing to
familial forms of strabismus. Elucidating the specific genes remains a grand challenge for the field, but emerging genome sequencing tools may generate a new wave of insights.

1.6 Gene expression studies

Experimental approaches to elucidate molecular mechanisms related to strabismus have been pursued. Microarray analysis showed that expression of 604 genes differ significantly between 100 strabismic EOM samples and 28 normal EOM samples. Together with polymerase chain reaction (PCR) experiments, three major conclusions were drawn. Collagen and collagen-related genes were upregulated; specific myosins, such as EOM-specific myosin (MYH13) and myosin heavy chain-1 (MYH1), and related contractile genes were downregulated; genes involved in energy balance, such as mitochondrion homeostasis or regulations of energy metabolism, were dysregulated in strabismic EOMs. The conclusions should be assessed with caution, since it was not specified which forms of strabismus were represented in the samples, although the authors suggested that the sample set may have a high portion of exotropia cases 48.

In another study, expression levels of seven myogenesis-related genes in EOMs from 18 concomitant strabismus patients were compared against 12 samples from a single non-strabismic individual. Six of the genes had reduced expression levels, leading Zhu et al. to suggest that altered growth of muscles may be involved. However, it was unclear whether the patients had congenital strabismus nor the nature of the deviations involved 49. Furthermore, the two sample sets were collected in distinct ways (i.e. obtained from corrective surgery vs cadavers), which has been recognized to cause difficulty in the interpretation of gene expression studies 50,51.
1.7 Duane retraction syndrome

While the focus of this review is the genetics of non-syndromic forms of strabismus, there are familial syndromes in which strabismus rates are elevated in otherwise non-syndromic family members. About 70% of DRS cases do not exhibit other congenital abnormalities, and approximately 20% of cases have a family history of strabismus \(^{22,52}\). Overall DRS accounts for approximately 5% of strabismus cases \(^{53}\). DRS is a congenital cranial dysinnervation disorder. Based on these observations, we include DRS in this review as we perceive an opportunity to find common causal genes between non-syndromic strabismus and DRS.

Three types of DRS have been described based on clinical examination. In these studies, key attributes include abduction, movement of a body part away from the midline, and adduction, movement toward the midline. Type 1 DRS is characterized by marked limitation of abduction, type 2 DRS is characterized by marked limitation of adduction, and type 3 DRS is characterized by a combination of marked limitation of both \(^{54}\). The majority (60%) of diagnosed DRS cases are female. Up to 60% of all cases are bilateral, and up to 80% of unilateral cases are left-sided \(^{54,55}\). Wabbels et al. found predominant females cases (64%) and left eye involvement of unilateral cases (72%), whereas bilateral only accounted for 12% of cases \(^{56}\).

While most cases are sporadic, reports of familial DRS date back to 1896 \(^{57}\). Up to 10% of Duane anomalies are inherited in an autosomal dominant fashion \(^{58}\). The connection between infantile esotropia and DRS are illustrated by recent studies. In the Strabismus Inheritance Study in Tasmania (SIST), a set of 133 families with infantile esotropia was recruited, of which multiple members were affected with DRS in two families. A separate set
of 40 families with at least one case of DRS were recruited, of which 21 had a familial history of ocular motility disorders but only two had multiple members affected by DRS. Linkage analysis had previously shown linkage between 8q12-13 and Duane syndrome. The SIST study confirmed a prior association of both DRS and infantile esotropia with partial trisomy 8. Combining this information, a gene-dosage mechanism was proposed.

Separately, Khan et al. identified two susceptibility loci, 3p26.3-26.4 and 6q24.2-25.1 using multipoint linkage analysis in a consanguineous family with four affected children (one with DRS and three with non-syndromic esotropia).

1.7.1 Chromosome 8q and type 1 DRS

The focus on chromosome 8q in DRS studies has progressed to the search for a causal gene in the loci, but no clear single causal gene has been established. A de novo reciprocal balanced translocation t(6;8)(q26;q13) was identified in a patient with DRS. This patient had amblyopia and narrowing of palpebral fissures. The carboxypeptidase A6 (CPA6) gene at the previously identified DURS1 (DRS-1) locus on chromosome 8 was disrupted between the first two exons in this patient and was proposed as the causal gene. CPA6, a member of the M14 metallo carboxypeptidase family, is expressed in a limited number of tissues in mice, including the rectus muscle layer of the embryonic eye. In adult mouse, CPA6 was expressed in olfactory bulb and other parts of the brain. CPA6 knockdown using morpholino antisense oligos in zebrafish did not produce a phenotype, contradicting a dosage hypothesis. No pathogenic CPA6 mutations were identified in a set of 18 sporadic DRS patients. Two patients with microduplication of 8q12 displayed multiple congenital anomalies, including DRS. Studying a third patient with similar phenotype, including DRS, a recent study identified the minimal critical region at the loci of 1.2Mb, excluding CPA6. CHD7
duplication was suggested to be responsible for at least part of the features in resulting from the 8q12 duplication. Reported duplications and deletions in affected individuals do not overlap, suggesting either multiple contributing genes or a gene with distal regulatory regions might be responsible. Although the chromosome region 8q12-q13 has been linked to DRS1 in multiple cases, more study is required before a definite conclusion can be drawn about the causal gene.

1.7.2 CHN1 and type 2 DRS

The CHN1 gene has been more clearly demonstrated to be a causal gene for DRS2. CHN1 is located on chromosome 2 and encodes two Rac-specific guanosine triphosphatase (GTPase)-activating alpha-2-chimerin isoforms. Miyake et al. identified seven heterozygous missense mutations in seven unrelated DRS2 families co-segregating with the affected haplotypes. These mutations were neither recorded in the single nucleotide polymorphism database nor observed on 788 control chromosomes. CHN1 mutations were present in 7 of 20 (35%) examined DRS families, while no CHN1 mutations were observed in 140 sporadic DRS patients. Predicted gain-of-function mutations in CHN1 were found in two families with type 2 DRS. Overexpression of wild-type alpha-2-chimerin in the chick embryonic oculomotor nucleus led to stalling of oculomotor nerve growth and the premature axon termination adjacent to the dorsal rectus muscle, supporting a functional role for CHN1 in DRS.

1.7.3 Type 3 DRS

While loci have been established that account for a portion of type 1 and type 2 DRS, the genetic components of type 3 DRS are more elusive. It is possible that the type 3 DRS is more heterogeneous than the other two classes. In a thin-sectioned magnetic resonance
imaging (MRI) study, the abducens nerve was reliably observed in 60 eyes of 30 individuals from a control group. The abducens nerve on the affected eye was absent in 18 of 18 eyes from 16 patients with type 1 DRS, and in 2 of 2 eyes from type 2 DRS patients. The nerve was absent in only 3 of 5 eyes from five patients with type 3 DRS. The clinical heterogeneity in type 3 DRS may reflect genetic heterogeneity.

1.7.4 Okihiro syndrome

In addition to the ocular anomalies of the basic form of DRS, Okihiro syndrome (also called Duane-radial ray syndrome) is associated with additional abnormalities affecting the upper limbs and, less commonly with renal anomalies and sensorineural hearing loss. Autopsy and MRI studies of Okihiro syndrome patients have revealed hypoplasia or absence of the sixth nerve nucleus (i.e. abducens nerve) on the affected side, the ipsilateral lateral rectus being innervated by branches of the oculomotor nerve.

Mutations in the SALL4 zinc finger transcription factor gene were the first causal genetic alterations discovered for Okihiro syndrome patients. The discovery arose when Kohlhase et al. proposed that Okihiro syndrome might be due to mutations in a SALL gene family member based on phenotype overlap between Okihiro syndrome and Townes-Brocks syndrome, which is caused by mutations in the SALL1 gene. They successfully identified mutations in SALL4 gene from five of eight Okihiro families. Al-Baradie et al. identified a nonsense mutation in SALL4 gene in affected individuals originally reported by Okihiro et al. in 1977, as well as 2 additional families. The broader DRS phenotype is present in approximately 70% of SALL4 mutations carriers. A mouse model shows that Sall4 is regulated by Tbx5 transcription factor; both genes contribute to patterning and morphogenesis of the anterior forelimb and heart.
endophenotypes between Okihiro syndrome and Holt–Oram syndrome, which is associated with mutations in the TBX5 gene. Whole mount in situ hybridization analysis of Sall4 expression during mouse embryogenesis shows prominent expression in midbrain and branchial arches and suggests that a dosage reduction of Sall4 might disrupt abducens nerve development.

1.8 Animal models

Although the genetic origins of strabismus remain to be fully deciphered, several animal models of the phenotype have been studied and may serve as resources in the search for causal genes. Most of the model animals described below are albinos, with pigmentation loss ranging from partial to complete. Visual abnormalities, including strabismus, have been linked with albinism in diverse mammals such as albino primates, white tigers, and albino cats (including Siamese cats).

1.8.1 Famous strabismic animals

Animals with cross-eyes have become popular images on the Internet. Joco, a cross-eyed lion at the Erfurt Zoo (Germany) is most likely to suffer from congenital strabismus. The cross-eyed opossum Heidi at the Leipzig Zoo (Germany) became a celebrity, but the condition was likely environmentally triggered. The causes of strabismus in animals vary, with only a portion deriving from genetic influence. Finding suitable animal models for the study of non-syndromic strabismus could accelerate research efforts.

1.8.2 Cats

In Siamese cats, a temperature-sensitive mutated TYR gene encoding tyrosinase is expressed normally in cooler extremities, giving a darker color, while expression is reduced in warmer parts of the body, leading to poor pigmentation. Anatomical studies show that
axons of temporal retinal ganglion cells go to the opposite side of the brain instead of staying on the same side as observed in non-albino cats. The misrouting defects are also observed in albino mice and rabbits with TYR defects. Insertion of functional TYR genes into such albino mice and rabbits corrects for axon misrouting. Humans with ocular albinism also show abnormal decussation (crossing) of optic neurons, causing reduced or absent binocularity. This characteristic is associated with elevated prevalence of strabismus. Nevertheless, there is not yet convincing evidence that TYR mutations contribute to strabismus in humans. While the link between strabismus and axon misrouting is unknown, genes directly involved in optic chiasm development might be considered as candidates.

The unusual axon wiring pattern observed in TYR defective albino animals raises concern that these animals may not be suitable models for human strabismus. Artificially induced strabismus models, such as those established by tenotomy (tendon lengthening) and by exposure to early abnormal visual experience, may be similarly ill-suited to study genetic influences on strabismus. To evaluate the relevance of artificially induced strabismic cats, the ocular dominance distributions for cats with induced strabismus and natural strabismus were compared and found to be similar. Approximately 35% of cells were monocular in either strabismus group, but a statistically significant difference was noted with normal cats, which have 81% binocular cells. Work with the animal models continues, exemplified by a study which showed that early induced unilateral convergent strabismus in cats led to abnormal corpus callosum connection. Such experiments highlight how abnormal early visual experience impacts visual cortex development, but do not provide a clear path for using induced animal models to track down key genes. Thus the study of non-syndromic strabismus could benefit from efforts to identify additional eye misalignment animal models.
1.9 Conclusion and future directions

The causal genes predisposing to non-syndromic forms of strabismus remain to be discovered. The combination of next-generation sequencing with both large-scale populations and targeted families may soon reveal critical genes and consequently confirm or expose critical molecular mechanisms. Genome-wide association studies have been reported to be underway, while exome sequencing family-specific studies of non-syndromic strabismus are likely to emerge soon \(^{91,92}\). Aided by the background presented in this overview, the discovery of critical genes causing non-syndromic strabismus will allow earlier identification of individuals who are at high-risk and thus most likely to benefit from effective early intervention treatments.

(This is the end of the publication.)

1.10 Research objectives

In order to improve our understanding of non-syndromic strabismus, I combined classical genetics and NGS approaches to identify variants that co-segregate with non-syndromic strabismus in a large North American family. Given the observed Mendelian autosomal dominant mode of inheritance in the family, I hypothesize that this familial strabismus is monogenic. Three major objectives of this study are:

1) To analyze variants affecting protein-coding regions that co-segregate with strabismus in a pair of third cousins using a WES approach.

2) To identify genomic region linked to strabismus through linkage analysis.

3) To investigate variants affecting non-protein coding regions within the linked region using WGS approach.
1.11 Study structure

Family studies with well-defined phenotypes are crucial to advance the understanding of the genetics of non-syndromic strabismus. We identified a multi-generation strabismus family and took two parallel tracks: WES analysis and linkage analysis, which was a collaborative effort with the Hospital for Sick Children in Toronto (SickKids) through the Finding of Rare Disease Genes Canada (FORGE Canada) network. Both approaches led to the identification of the same region, but no promising protein-coding variant could be found in this region, which contains few exons (Figure 1.3). Next, we used WGS to scan the genome in a uniform manner and to generate additional candidate variants for the linked region. Lead candidates were selected through a combination of prioritization strategies and will be subjected to genetic validation in additional families and functional validation.
<table>
<thead>
<tr>
<th>Loci</th>
<th>Inheritance pattern</th>
<th>Ethnicity</th>
<th>Phenotype</th>
<th>PMID</th>
</tr>
</thead>
<tbody>
<tr>
<td>7p22.1(STBMS1)</td>
<td>Recessive</td>
<td>European</td>
<td>Esotropia in infancy or childhood, 7 of 8 affected individuals had various degree of hypermetropia</td>
<td>Parikh et al., 2003&lt;sup&gt;44&lt;/sup&gt;</td>
</tr>
<tr>
<td>7p22.1(STBMS1)</td>
<td>Dominant</td>
<td>Northern Irish</td>
<td>Primary nonsyndromic comitant esotropia</td>
<td>Rice et al., 2009&lt;sup&gt;45&lt;/sup&gt;</td>
</tr>
<tr>
<td>16p13.12-p12.3</td>
<td>Recessive</td>
<td>Saudi Arabian</td>
<td>Infantile esotropia and esotropic Duane retraction syndrome</td>
<td>Khan et al., 2011&lt;sup&gt;93&lt;/sup&gt;</td>
</tr>
<tr>
<td>4q28.3</td>
<td>Dominant</td>
<td>Japanese</td>
<td>Comitant strabismus</td>
<td>Shaaban et al., 2009&lt;sup&gt;47&lt;/sup&gt;</td>
</tr>
<tr>
<td>7q31.2</td>
<td>Recessive (Imprinting)</td>
<td>Japanese</td>
<td>Comitant strabismus</td>
<td>Shaaban et al., 2009&lt;sup&gt;47&lt;/sup&gt; (Shaaban et al., 2009&lt;sup&gt;94&lt;/sup&gt;)</td>
</tr>
<tr>
<td>6q26</td>
<td>Imprinting</td>
<td>Japanese</td>
<td>Comitant strabismus</td>
<td>Shaaban et al., 2009&lt;sup&gt;94&lt;/sup&gt;</td>
</tr>
<tr>
<td>12q24.32</td>
<td>Imprinting</td>
<td>Japanese</td>
<td>Comitant strabismus</td>
<td>Shaaban et al., 2009&lt;sup&gt;94&lt;/sup&gt;</td>
</tr>
<tr>
<td>19q13.11</td>
<td>Imprinting</td>
<td>Japanese</td>
<td>Comitant strabismus</td>
<td>Shaaban et al., 2009&lt;sup&gt;94&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Figure 1.1  A schematic representation of EOMs and nerve innervation with associated genes.
Defects along the visual-motor axis can contribute to infantile esotropia.
Figure 1.3  Overview of the strabismus genetic study.
Chapter 2: Project Description

2.1 Introduction

Non-syndromic strabismus is heterogeneous. The genetic mechanisms underlying the disorder remain largely unknown. Multiple loci scattered across many chromosomes, albeit often with weak evidence, have been associated with strabismus. The only locus confirmed in two independent families lies on chromosome 7 (STBMS1), but the families show conflicting inheritance patterns\(^{44,45}\). It is likely that most families with inherited forms of non-syndromic strabismus have family-specific variations. However, the discovery of each gene contributing to strabismus can be a great step forward in our understanding of the mechanisms underlying non-syndromic strabismus. The role of the gene/protein can be included to the known pathways, and the gene can be included in candidate gene analyses for other familial strabismus studies.

We identified a family with multiple generations affected by strabismus, with the earliest reported case dating to 1849 (and documented by a photograph). The participation of this family provides a unique opportunity to investigate the genetic component of strabismus. The research that follows seeks to identify the genetic cause of the strabismus phenotype in this particular family. Previous studies suggest locus heterogeneity in non-syndromic strabismus and simple Mendelian models cannot explain the complexity of the inheritance patterns. Although a portion of non-syndromic strabismus cases might reflect a common disease-common variant model, it does not exclude the possibility of a Mendelian model in a certain family. A highly penetrant autosomal dominant inheritance pattern is rarely reported for non-syndromic strabismus. Based on the small number of families reporting such a
phenotype, we reason that the causal mutation is rare (occurring at a frequency of ≤ 1% in controls) in the subject family.

There are multiple technological approaches to the identification of causal genetic variants for family-based studies. Linkage analysis is a traditional approach for family-based gene discovery, but it only detects broad regions that are likely to harbor variants with large effect size, which makes a great contribution to disease risk. Subsequent sequencing is required to identify the actual sequence variant(s)\textsuperscript{95}. The advancement of the Human Genome Project and, more recently, NGS technology provides a powerful new means to explore the human genome. The development and maturation of bioinformatics tools makes annotation, organization, and interpretation of millions of variants an approachable problem for family-based studies. Successful WES studies on different diseases demonstrate its efficiency,\textsuperscript{96} and more than 180 distinct disease-causing genes have been identified thus far using the technology\textsuperscript{97}. Although WES has successfully identified many disease-causing variants, the overall molecular diagnostic rate in large scale study is approximately 25%\textsuperscript{98}. Protein-coding regions only comprise approximately 2% of the human genome. The majority of the genome is missed using WES. At present WGS is less commonly used due to its inherent complexity and high cost, but it has been gaining an increasing popularity as the cost approaches WES and the interpretability increases. The WGS approach reveals more polymorphisms than WES, but we have neither a reliable means to filter these variants such as for coding mutations nor a cost-effective way to prove causation \textsuperscript{97}. As the technology and knowledge develop, new approaches will be established to improve the interpretability of WGS.

Within our study, we used all three approaches to identify the causal region for the family studied. A specific 10 Mb region on chromosome 14 was identified that segregates
with the phenotype, within which more than 300 rare variants are present in our cases, of which 3 are highlighted based on a qualitative assessment of the available genome annotation data.

2.2 Materials and methods

2.2.1 Samples

Based on the constructed pedigree and available contact information, a total of 18 family members were contacted initially, of which 13 gave written consent to participate in the research project. Eight participants reported early onset strabismus, and the other five reported no strabismus. Saliva and/or blood samples were collected from participants. Except for 012, who is married-in, all of the other 12 individuals were descendants of a common ancestor and were subjected to genotyping and linkage analyses. We subsequently recruited two additional strabismic members with full consent, 014 and 015, from two different branches of the family.

2.2.2 Phenotyping

The phenotyping study examined 13 individuals: 8 participants from the genetic study and 5 only involved in the phenotyping study. 014 was examined by Dr. V. Pegado, 009, 011, and 013 were examined by Dr. S. Narasimhan. Dr. J. Horton ascertained/re-ascertained all except for 014. All participants were questioned about the age of onset (if applicable), ocular history, and medical history. Examination included visual acuity, pupils, eye movements, ocular alignment, stereopsis, slit lamp examination, fundus examination, and intraocular pressure.
2.2.3 DNA preparation

At least 4mL blood sample or 6mL saliva samples were collected for one round of NGS, and at least 2mL saliva sample was collected from participants for genotyping. Blood samples were collected in a clinical setting while saliva samples were collected using Oragene-DNA (OG-500) saliva kit. 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 # PD-PR-015. All samples were confirmed to meet the quality requirements set by service providers. 7-10 µg DNA per sample at a concentration no less than 70 ng/µl was sent for NGS. 500 ng DNA per sample at a concentration of at least 50 ng/µl was sent for genotyping.

2.2.4 Instrumentation

Chromosomal microarray analysis (CMA) was performed on Affymetrix- Cytoscan HD for index, 011. WES was performed via the Agilent SureSelect Human All Exon 38Mb kit and Illumina HiSEQ 2000 platform (Perkin Elmer). Linkage analysis was performed on HumanOmni2.5, using the Infinium LCG assay. WGS was performed on an Illumina HiSEQ 2000 platform (BGI America).

2.2.5 Next-generation sequencing data processing and variant calling

The genomic aligners, Bowtie (version 0.12.9 for WES and version 1.0.0 for WGS) and BWA (version 0.6.1 for WES and version 0.7.5a for WGS), were used to map the reads to the hg19 reference genome\(^99,100\). All stated coordinates are based on hg19 in this thesis. The Genome Analysis Toolkit (GATK) (version 1.0 for WES and version 2.8 for WGS) performed local re-alignment, which allowed for correcting some misalignment at the
extremity of reads. The same informatics pipeline was applied to both WES data and WGS data.

SAMtools (versions 0.1.18 and 0.1.19 respectively) was applied to call variants from aligned WES and WGS reads. For comparison of two distinct informatics pipelines, WGS variant calls generated by the supplier (BGI) were also analyzed.

### 2.2.6 WES variant prioritization

The SIFT software program was used to assign annotations to variants, focusing on non-synonymous coding variants. Custom computer programs (scripts) were used to extract results that correspond to an autosomal dominant model. Although strabismus incidence has been reported to be up to 5% in a population, the causes are a mix of environmental and genetic influences. Any single variant could only explain a portion of cases. Since no other family has been reported with such a clear autosomal dominant pattern of inheritance, the variant in the subject family may be private, which means the variant is restricted to the family. We thus focused on variants with a frequency not higher than 1% in dbSNP build 135, giving rise to the list of rare variants. We further prioritized variants that are predicted by SIFT to be damaging or cannot be predicted, resulting in the final list of candidate variants for consideration.

### 2.2.7 WES variant selection

The list of rare variants was compared to pre-compiled lists of candidate genes from the research literature. These candidate genes were identified through three approaches. First, Medical Subject Heading Overrepresentation Profiles (MeSHOPs) was used to identify genes arising in a significant number of articles annotated as strabismus-related in the Medline database. Second, literature review on both non-syndromic strabismus and
syndromic strabismus (such as CFEOM and DRS) was performed to construct a locus list. Loci were converted to chromosome coordinates based on hg19 (http://www.tallphil.co.uk/bioinformatics/cytobands/). Third, genes expressed in selected mouse brain fine structures from Allen Brain Atlas (http://mouse.brain-map.org/) were extracted, including ‘superior colliculus, sensory related’, ‘tegmental reticular nucleus’, ‘suprachiasmatic nucleus’, ‘subthalamic nucleus’, ‘substantia nigra, compact part’, ‘subiculum’, ‘pedunculopontine nucleus’, ‘lateral septal nucleus, caudal (dau dorsal part)’, ‘entorhinal area’, and ‘Edinger-Westphal nucleus’

2.2.8 Confirming variant presence

The Integrative Genomic Viewer (IGV) (version 2.0.34) was used to visualize read alignment and assess variant quality. For the leading candidate, allelic specific polymerase chain reaction (AS-PCR) was used to amplify the target variant. The online interfaces with default settings were used for the tools mentioned below. Pseudogenes were first identified by literature search and BLAT with assembly hg19, and flanking regions with distinctions between target and pseudogenes were selected by inspecting a MUSCLE alignment (https://www.ebi.ac.uk/Tools/msa/muscle/). The following sequences were used as inputs: (GTF2IRD2) chr7: 74210484-74250981, (GTF2IRD2B) chr7:74508347-74548846, and (GTF2IRD2P1) chr7: 72656902-72685658. WASP, a web-based allele-specific PCR assay designing tool, was used to design initial pairs of primers, focusing on creating mismatch at the 3’ end of one primer, and Primer3 (http://primer3.sourceforge.net/) was used to optimize complementary primers. Finally, Beacon Designer (http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1) was used to provide information of optimal reaction conditions. AS-PCR was used to amplify the target with the following
primers: 5’-GGGACACTCGGTGTCTGACA and 5’-AACTCCGTCTCAACAACACAAC. The annealing temperature was optimized at 57°C. One African American, two Caucasian, and one Chinese DNA samples were used as controls. Final products were Sanger sequenced to confirm the presence of the predicted variant.

2.2.9 WES variant density plot

The identification of identity by descent (IBD) regions often relies on the use of microarrays, which probe for sequence in a contiguous region instead of the punctate exons studied by WES. In order to use WES data to identify candidate IBD regions in a pair of individuals, custom R-2.15 scripts were used to count the number of common variants for each chromosome. The results were depicted in a variant density plot, with X-axis as the chromosome position. The upper half of a plot represents one individual while the lower half represents another individual. Each vertical line represents one variant. To overcome the punctate spatial distribution of WES, the combined set of variants from both individuals for the same chromosome was divided into 50 non-overlapping intervals, with each interval containing the same number of observed variants for a given chromosome.

We generated data for all 22 autosomes for the comparison between the two subjects and that of two anonymous controls (unaffected). The value of shared variants over total variants for each chromosome in each individual was generated, and logarithm (base 2) for the ratio of corresponding value was calculated. Variant clusters were defined as candidates for IBD regions.

2.2.10 Linkage analysis

Linkage analysis was performed by our collaborator (Nicole Roslin, SickKids, Toronto ON). Multiple analyses were performed for quality control purpose, including but
not limited to call rate and relationship checking. 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 was used to simulate pedigrees under dominant and recessive models with a range of disease allele frequencies and penetrance\textsuperscript{111}, and Merlin 1.1.2 was used to analyze these pedigrees under the same model\textsuperscript{112}. For a particular model, the maximum LOD score from the analysis of 1000 simulated pedigrees was declared the maximum LOD score.

Simulations were performed using $q = 0.005, 0.01, \text{ and } 0.02$ for the dominant models, and $0.1, 0.16, \text{ and } 0.2$ for the recessive models (corresponding to estimated prevalence of 1%, 2.5% and 4% respectively). For autosomal dominant and recessive models, simulations were performed with penetrance of 50, 60, 70, 80, 90, 95 and 99%. An estimated phenocopy rate of $0.2\%$ was used for all models.

Multiple filters were applied to select a set of markers suitable for linkage analysis. First, SNPs on Y chromosome were excluded. Second, markers with alleles ambiguous for strand information (A/T and G/C variants) were removed to facilitate matching with HapMap data. Third, monomorphic markers were removed since they were uninformative for linkage under a dominant model. Fourth, only markers present in both this set and corresponding HapMap sets were retained. SNPs with minor allele frequency $> 0.45$ were retained for most linkage information. Lastly, only markers with pairwise $r^2 < 0.1$ were kept to prevent inflation of multipoint LOD scores due to linkage disequilibrium between the SNPs\textsuperscript{113}.

Merlin 1.1.2 was used to perform multipoint linkage analysis under a dominant model with a disease allele frequency of 0.005 and penetrance of 0.2, 99 and 99% for 0, 1, 2 copies.
of the disease-causing alleles, respectively. The parameters correspond to a disease prevalence of approximately 1%.

2.2.11 WGS variant prioritization

Variants that fell within the candidate region identified by linkage analysis were selected. To ensure that the three samples used for WGS were consistent with the same IBD region, the numbers of variants in the linked region were examined for enrichment. Allele frequency was assessed using dbSNP build 137 and Exome Variant Server (EVS), and variants with a frequency higher than 1% were excluded. Heterozygous variants shared across the three samples were selected, and SnpEff (with hg19 database) was applied to annotate variants. The tools and parameters in the BGI computational analysis pipeline are largely unknown, but the BGI pipeline is expected to be different than the in-house pipeline. Therefore, lists of variants from these two pipelines were analyzed independently.

2.2.12 WGS variant selection and confirmation

Candidate variants were compared against an in-house database containing 110 exomes and against the Michael Smith Genome Sciences Centre (GSC) database, containing 1580 non-cancer genomes and 785 cancer genomes. Multiple annotation databases and corresponding bioinformatic tools were used to annotate variants, including functions annotation of the mammalian genome 5 (FANTOM5) database, JASPAR, Segway, RegulomeDB and Combined Annotation Dependent Depletion (CADD)\textsuperscript{114–117,115}. FANTOM5 database allows us to obtain expression profiles and functional annotation of cell-type-specific transcriptome, and we examined whether candidate variants overlapped with transcription start site (TSS) and enhancer\textsuperscript{118,119}. JASPAR database stores matrix-based nucleotides profiles for transcription factors binding preference\textsuperscript{115}. Segway provides
chromatin state annotations at a nucleotide resolution for both conserved and non-conserved sequences\textsuperscript{114}. RegulomeDB utilizes experimentally measured evidence and computational processing evidence to provide functional assignment and was used to prioritize variants\textsuperscript{120}. CADD integrates many diverse annotations into a single C-score for SNVs or indels and was used to annotate the impact of variants\textsuperscript{117}. Manual inspection via IGV (version 2.3.2) was used to evaluate the read quality and to judge the credibility of variants\textsuperscript{105}. Variants that were likely due to a sequencing error or alignment error were assigned to a lower priority category. Based on the qualitative assessment, top prioritized variants were confirmed to be present by standard PCR in WGS subjects.

2.3 Results

2.3.1 The index case

An ophthalmologist referred the index patient to BC Children’s Hospital, where a preliminary determination of familial non-syndromic strabismus was proposed based on the family history and physical examination of the index and parents. Since the index exhibited additional minor characteristics beyond strabismus, a blood sample was obtained to prepare DNA for CMA analysis of potential copy number variation (CNV) and other structural alterations. The CMA results were unremarkable; the analysis did not reveal CNVs > 200 kb in size or any abnormally large runs of homozygosity (ROH). The 10 Mb locus on chromosome 14 mapped in this family was also examined on the array and no CNVs were detected.

2.3.2 Pedigree

From the foundation of an extensive family genealogy project, we constructed a seven generation pedigree that contains 157 individuals, including deceased individuals.
Three major branches from the same common ancestor were traced (Figure 2.1). Most participants come from branch 1 (Figure 2.2 and Figure 2.3). In branch 1a and 1b, strabismus was reported across 4 consecutive generations according to family anecdotes and/or medical records. An even distribution of strabismus cases was observed between females (8 individuals) and males (7 individuals). An autosomal dominant model with high penetrance best matches the inheritance pattern. The index case was assigned the identifier 011 and is indicated with the identifier in all the figures. Late in the research project, subject 014 from branch 2 and subject 015 from branch 3 self-identified as having a strabismus phenotype were enrolled in the study (Figure 2.4 and Figure 2.5). Ophthalmic information is not yet fully available from these two emerging branches.

2.3.3 Phenotyping

Thirteen individuals participated in clinical phenotype assessment sessions conducted by one of three vision experts, with twelve ascertained or re-ascertained on the same day by the same specialist for consistency. Five individuals who participated in the phenotyping study but not the genetics analysis were confirmed to be unaffected by strabismus. The characteristics of strabismus were not uniform across the affected relatives in the family. For several individuals, the original direction and amount of deviation can be difficult to ascertain retrospectively after multiple surgeries and/or development of other ocular conditions. The affected individuals could be grouped into two broad categories: esotropia and congenital superior oblique muscle palsy (fourth nerve palsy). Congenital superior oblique muscle palsy was reported in the re-ascertainment for 009 and 013 (neither of which had undergone surgery), while each was more generally classified as hypertropia in an initial
examination. Both esotropia and hypertropia were noted in 014 (this individual had undergone multiple surgeries).

2.3.4 Whole exome sequencing analysis

Whole exome sequencing was performed on DNA samples from subjects 001 and 011 (i.e. the index), with over 50 million reads (>5x10^9 bp of raw sequence data) obtained for each. Quality control measurements and the number of detected variations from the reference genome meeting the filtering criteria (see Methods) are listed in Table 2.1. The two distantly related individuals share 119 heterozygous non-synonymous variants as shown in the seventh column. As any causal variant for the dominant phenotype is expected to be rare in the general population, a maximum frequency threshold of 1% was applied, reducing the set of candidate variants to 60.

By design, WES focuses on coding regions, as coding sequence alterations have until now been the most frequently reported causal mutation in genetics studies (albeit also the most deeply studied). We therefore prioritized the review of coding variants in our initial analysis. The eight variants (Table 2.2) selected by using SIFT were prioritized for further consideration (damaging variants and variants with unknown effect), and the other candidates were retained as lower priority, but still potentially important candidates.

We sought to determine if any of the prioritized eight candidate variants were located within genes previously connected to strabismus or strabismus-related tissues. To find genes associated with strabismus, we used a candidate approach. Using as a query the MeSH terms ‘strabismus’, ‘esotropia’, ‘exotropia’, and ‘oculomotor muscles’, we ranked 30,251 gene and loci MeSH Over-representation Profiles (MeSHOPs) for similarity. Within this list, MeSHOPs for 23 genes and three loci had a statistical enrichment for one of the indicated
MeSH terms, however none overlap with the set of eight candidate genes arising from the exome sequencing.

Systemic literature review on syndromic forms of strabismus led to a list of associated loci (Table 2.3), which was combined with the set of non-syndromic loci (Table 1.1) to generate a final set of literature-based candidate regions (Table 2.4). Eighteen of 22 chromosomes have at least one locus linked to strabismus. The converted coordinate list does not overlap with the top eight variants on the final list.

We reasoned that genes expressed in anatomical structures associated with eye misalignment might have a higher chance to be causal. Therefore, we obtained expressed genes lists in a set of fine brain structures related to eye alignment in mouse from Allen Brain Atlas and compared to the final list. The eight genes on the final list did not express in any of the above structures.

2.3.5 WES variant validation

Literature mining was performed individually for each of the eight candidate variants on the exome list. The \textit{GTF2IRD2} variant chr7:74212518 G\textasciitilde{}A was selected for Sanger sequencing confirmation. \textit{GTF2IRD2} is adjacent to the William syndrome (WS) locus on chr7q, and the incidence of strabismus in WS is up to 78\%\textsuperscript{121}. This observation suggested a genetic link between strabismus and WS. WS is a rare neurodevelopmental disorder caused by hemizygous deletion at 7q11.23. At the telomeric end of the deletion lies three genes of the same transcription factor family (TFII-I family): \textit{GTF2I}, \textit{GTF2IRD1}, and \textit{GTF2IRD2}\textsuperscript{122}. Analysis of atypical deletion suggests that absence of \textit{GTF2IRD1} and \textit{GTF2I} may be responsible for certain WS features such as craniofacial dysmorphology, hypersociability, and visuospatial deficits\textsuperscript{123,124}. Disruption of \textit{GTF2IRD2} expression might contribute to WS
features. The frequency of chr7:74212518 G>A is 0.4% in Europeans, and both SIFT and PolyPhen predicted the P445S change to be damaging. Study of a skeletal-muscle-specific transgenic expression mouse model demonstrated that GTF2IRD2 has a role in determining muscle fiber type\textsuperscript{124}.

At least two pseudogenes in the human genome share high sequence identity with GTF2IRD2, and standard PCR could not amplify the target region specifically. An AS-PCR-based approach was performed with samples from 001 and 011, and subsequent Sanger sequencing confirmed the presence of the variant in both individuals. G/G genotype was expected for unaffected individuals, while G/A or A/A genotype was expected for affected individuals. However, the GTF2IRD2 variant does not segregate with strabismus in all family members. Subject 008 does not have strabismus, but has a G/A genotype; subject 009 has strabismus and is an obligate carrier, but a G/G genotype was observed (Figure 2.6). To confirm the findings, new samples were collected and the experiments repeated, confirming the initial findings.

We also found that 002 is homozygous for allele A. There is no consanguinity in this family, and chr7:74212518 G > A might be more common than reported in the population. Although GTF2IRD2 (chr7:74212518 G>A) was a promising candidate based on variant characteristics and literature, this gene was ruled out as the causal mutation for strabismus in this family by the follow-up individual genotyping. Further literature mining for the other seven variants did not lead to another lead candidate.

2.3.6 WES variant density plot

To extract additional information from WES data, we generated variant density plots to identify potential IBD regions. Chromosome 11 (0.32) and 14 (0.49) show a higher log2
values compared to the rest (ranges from -0.61 to 0.29), suggesting that the two strabismic individuals share more variants in these two chromosomes than the control group and that potential IBD regions might reside on these chromosomes. Variant density plots for each chromosome in both groups were examined closely (Figure 2.7 and Figure 2.8). The left panel was based on data from two strabismic individuals, and the right panel was based on data from controls. For each plot, the upper half represents one individual while the lower half represents the other individual within the group. Three rows were created for each individual (row 1 to 6 from top to bottom): the innermost rows (3 and 4) represent all variants from the hg19 reference genome in each individual; the middle rows (2 and 5) display heterozygous variants in each individual; and the outermost rows (1 and 6) are identical and indicate positions of shared heterozygous positions between the two individuals. Two shared heterozygous variant clusters were observed: chr14:24,600,000 – 32,300,000 and chr11:48,200,000 - 56,000,000. Reviewing our variant call files revealed two shared rare variants (≤1% in the EVS and dbSNP databases) within these regions: two on chromosome 14 and zero on chromosome 11. One variant was protein coding change in G2E3 but predicted to have a neutral effect by SIFT, and the other was in an intron of STRN3. Literature review did not identify specific relevance of the genes. We expect a single corresponding region for the whole family due to the observed autosomal dominant mode of inheritance. Although the two individuals can share more than one region by chance, additional information from other relatives is expected to refine the candidate loci – thus linkage analysis was pursued.
2.3.7 Linkage analysis

Samples from 12 individuals (8 affected; 4 unaffected) were genotyped using a high-density genotyping panel. A set of 17,779 SNPs was obtained after the SNP filtering step for linkage analysis. Simulations under the alternative hypothesis (linkage) generated a maximum simulated LOD score of 3.56, under an autosomal dominant model with minor allele frequency $q = 0.005$ and 99% penetrance. The LOD score curves did not change significantly with disease allele frequency, and the dominant models had consistently higher LOD scores than recessive models (Figure 2.9).

Based on the genotyping data from the subject family, the largest observed LOD score is 3.55, on chromosome 14, which is a striking score relative to the possible maximum (3.56) obtained from simulations. This is the only region with a LOD score higher than 3, and thus the only region for which rejection was made of the null hypothesis of independent assortment. The linked region on chromosome 14 spanned approximately 10Mb on the physical map and was bounded by the markers rs7146411 and rs1951187, corresponding to chr14: 22,779,843 – 32,908,192. This region identified by linkage analysis is a novel locus for non-syndromic strabismus.

The pedigree suggested an autosomal dominant pattern of inheritance for this large multi-generation family, and the linkage analysis result supports this hypothesis. The LOD score under an autosomal dominant model of the linked region approaches the maximum LOD score as demonstrated by simulations. A single region has achieved a significant LOD score across the whole genome and is present in all affected individuals. Moreover, the linked region overlaps the chr14:24,600,000 – 32,300,000 region identified via WES variant density plots, further supporting its association with the familial strabismus. Displaying the
candidate region in the UCSC genome browser reveals a sparsely annotated zone consistent with a gene-poor region (Figure 2.10).

**2.3.8 WGS analysis**

Whole genome sequencing was performed on three individuals with a specific focus on the candidate loci. The three individuals, 001, 013, and 014, were selected to represent three distinct branches of the tree, and willingness to provide further samples.

Since 014 joined the study subsequent to the linkage analysis, we compared the number of shared variants between 014 and strabismic relatives to 014 and unrelated controls (Table 2.6). The upper two rows show that two unrelated individuals (C5 means control family 5 for example) share less than 14,500 variants. Since control families 3, 4, and 5 were collected to investigate a rare disease with unknown ancestral relationship, we did not compare between families. Individuals within the same control family are in each case a parent-offspring pair, thus sharing at least one allele across the entire genome. We extracted variants from the chr14:20,000,000 – 35,000,000 region for each individual. The parent-offspring control pairs exhibit 17,000 to 20,400 variants in total within the loci, and numbers of shared variants between all pairs of strabismic individuals are within this range while the numbers of shared variants between unrelated individuals are within 13,300 to 14,100 (Table 2.6). This observation suggests that 014 shares the linked region.

More than 1 billion clean reads were processed for each of 001, 013, and 014 using two pipelines, and the number of called variants are presented in Table 2.8 and Table 2.9. We focused on the rare variants within the candidate region shared across the three distant relatives, considering each bioinformatics processing pipeline separately. From the BGI pipeline, the three individuals share 671 rare heterozygous variants (Figure 2.11 and Table
For variants based on the in-house pipeline, approximately 14% of variants were left after the 1% filter. About 1,200 variants were homozygous, and the rest were heterozygous. The three individuals share 378 rare heterozygous variants (Figure 2.12 and Table 2.9). Combing the BGI list and the in-house list and de-prioritizing variants with low read quality with IGV, a list of 675 variants was subjected to further analysis.

The combined list of candidate variants was compared against a database of NGS variants maintained at the Michael Smith Genome Sciences Centre (GSC). No ophthalmological annotations are available from GSC, but variations arising in this small collection and within our family would suggest that the variation is not rare or the sequencing technologies and bioinformatics processing are producing systematic false positives. De-prioritizing variants that were present more than 25 times in a combination of our in-house exome database and the GSC left 332 candidate variants in the rare variant list.

### 2.3.9 WGS variant prioritization

Since the linked region is gene-poor, we opted to perform variant prioritization using SnpEff rather than the more commonly used SIFT and PolyPhen2 software. SnpEff generates annotation for different isoforms of a gene and informs the location of non-coding variants in terms of nearby gene. Five of 332 variants map to exons and only one is predicted to be a non-synonymous-coding change.

We used different methods to annotate variants with regulatory information. The FANTOM5 project generated detailed annotation of promoter and enhancer regions using a high-throughput version of the cap analysis gene expression (CAGE) technique. No candidate variant overlaps with reported TSS, and no high quality variant overlaps with FANTOM5 enhancers⁰¹⁹. Comparison against the VISTA enhancer database also did not
reveal matches (http://enhancer.lbl.gov/). One variant was found to overlap with a reported JASPAR transcription factor binding sites (TFBS). chr14:28052726 A > C falls on a FoxA1 TFBS, and it is predicted to have a high impact on the TFBS binding capacity (ranked the top 89% of all possible single nucleotide substitutions) (Figure 2.13). It was reported twice in GSC normal samples and once in a GSC cancer sample.

Next, we used less direct regulatory information from ENCODE to identify potential variants. The ENCODE project provides comprehensive information for chromatin properties in multiple human cell lines, and many different tools have been developed to utilize the information. Segway labeled chr14:29247628 TAAACAAACAA > TAAACAA as a candidate repressor region, due to the presence of H3 trimethyl-lysine 27 (H3K27me3) mark in H1 human embryonic stem cell lines. H3K27me3 is associated with inactive gene, and it signals bivalent promoter with K4me3 in embryonic stem cells. Inspection through IGV revealed an AAAC deletion within C14orf23. A small region surrounding this variant is highly conserved in the genome, suggesting that it is under evolutionary selection and that a change may have a functional impact. The chr14:29247628 TAAACAAACAA > TAAACAA deletion was reported in one normal sample from GSC. Another variant suggested by Segway was chr14:32547404 A > T, which was labeled as flanking a TSS. A strong H3K27Ac peak can be observed on 7 cell lines from ENCODE data through UCSC genome browser, which suggests an active regulatory element. This variant falls in an intron of ARHGAP5, which belongs to the same protein family as CHN1 that is associated with Duane retraction syndrome type 2. Moreover, an alternative TSS of ARHGAP5 is reported in the FANTOM5 data in close proximity to the variant. Only one normal sample from GSC carries chr14:32547404 A > T.
Additional annotation tools were used to integrate information to evaluate the three selected variants. Briefly, chr14:28052726 A > C has a RegulomeDB score of 5 and a C-score of 6.274. chr14:29247628 TAAACAAACAA > TAAACAA has a RegulomeDB score of 5 and a C-score of 14.65. chr14:32547404 A > T has a RegulomeDB score of 2b and a C-score of 14.12. The scoring scheme can be found in respective references\textsuperscript{117,120}. These three variants were confirmed through Sanger sequencing to be present in the three subjects and are summarized in Table 2.10.

2.4 Discussion

This study identified a family with non-syndromic strabismus, and the pedigree shows a Mendelian autosomal dominant inheritance pattern. Two distant relatives, 001 and 011, were subjected to WES, and two potential IBD regions were identified through the process. Linkage analysis was performed in collaboration with the FORGE network, and analysis with 12 descendants from a common ancestor revealed a single significantly linked region on chromosome 14 with a maximum LOD score of 3.55. This linked region overlaps with one of the two shared IBD regions from WES, supporting its segregation with the familial strabismus. WGS provided a comprehensive inventory of variants in the linked region for three affected individuals. A list of 332 candidate variants was obtained after multiple filters, including 327 variants in non-coding regions and 5 in coding regions. Through multi-level assessment, 3 candidate variants were selected for Sanger sequencing and the genotype was confirmed in 001, 013, and 014. However, causality of specific variants remains to be proven by additional validation.
As a whole, strabismus is a complex disease, and previous studies have showed locus heterogeneity, but a subset of cases might better fit a Mendelian disease model. This observation supports the idea that rare variants can be the driving forces of common diseases and more similar to Mendelian diseases than is postulated by the common disease-common variant model\textsuperscript{125}. Our linked region is the best-identified locus for non-syndromic strabismus, and both pedigree and molecular evidence supports an autosomal dominant inheritance model. The striking pedigree and enthusiastic participants greatly contribute to this study. A genealogy project initiated by a family member collected ample information regarding the family history, including some descriptions of eye problems, fuels the genetic project. The emerging technology, NGS and microarray, allows us to perform genetic research in an efficient and cost-effective way. Applications of NGS have already successfully identified many genes for Mendelian diseases, for undiagnosed childhood genetic diseases, and for predisposition to common complex diseases\textsuperscript{126}.

Given the small sample sizes for both cases and controls in many NGS studies, it is common to see a focus on coding variants. For example, WES is based on the assumption that mutations on coding regions are more likely to influence human phenotype and contribute to high penetrant diseases on a nucleotide for nucleotide basis\textsuperscript{126}. Due to the paucity of protein coding alterations within the candidate loci, we performed WGS to seek potential regulatory sequence disrupting variants.

The nature of regulatory mutations can be large-scale structural alterations such as a copy number change, or small alterations impacting either non-coding RNA change or a cis-regulatory disruption such as a TFBS. A structural variation can be either unbalanced (copy number variants) or balanced, such as inversions or reciprocal translations, and the change
can be defined to span greater than one kilobase\textsuperscript{127}. CMA did not detect an alteration in a cytogenetic level of resolution, but smaller scale alterations below the ~50 kilobase threshold might be missed by the CMA procedure. Non-coding RNA can also play a role in human diseases. Based on annotation derived from GENCODE v7 and ENCODE, individuals contain approximately 200 non-coding RNA variants, of which 4-17 are < 0.5% and appear to be selected against\textsuperscript{128}. Nevertheless, there is no evidence for the involvement of non-coding RNA in this case.

Different defects along the visual-motor axis might contribute to strabismus, and all three variants seem to contribute to strabismus with a neuronal origin. Chr14:28052726 A > C falls on a FoxA1 TFBS, which lies 1.2 Mb away from FOXG1 and 0.9 Mb away from NOVA1. FoxA1 is a pioneer factor that differentially binds to distant enhancer, leading to cell type-specific changes in chromatin structure, and collaborates with other specific transcription factors\textsuperscript{129}. The TFBS variant might contribute to strabismus by altering FoxA1 recruitment and the downstream expression of a gene in the linked region.

The linked region is associated with Rett-like phenotypes, and FOXG1 (forkhead box G1) has been implicated in multiple clinical features. A feature in FOXG1-related encephalopathy is the high prevalence of strabismus, which is not observed in other Rett Syndrome (RTT) patients, and FOXG1 syndrome is an autosomal dominant disorder\textsuperscript{130}. FOXG1 contains only one exon (exon 1) and encodes a winged-helix transcriptional repressor, which is critical in forebrain development. Four alternative transcripts for additional exons (exon 2 to 5) are identified in fetal brain\textsuperscript{130}. Point mutations of FOXG1, deletion and duplication of region including FOXG1 have all shown similar phenotypes, indicating a dosage sensitivity effect\textsuperscript{131}. Deletions excluding FOXG1 also lead to Rett-like
phenotypes, suggesting a cis-acting regulatory element hypothesis. A minimal region of a silencer (chr14: 29,875,671 – 30,303,082 hg19) has been identified to locate 0.6 Mb from FOXG1 coding sequence, supporting the hypothesis\(^{132}\). These observations suggest that FOXG1 is regulated by multiple distal genetic elements, and FoxA1 TFBS might be one of these elements and underlies the familial strabismus.

One variant (chr14:29247628 TAAACAAAACAA > TAAACAA) falls within C14orf23 intron, which is a gene with unknown function. This position is highly conserved through evolution, forming a conserved island and likely undergoing a positive natural selection. C14orf23 is 2.5 Kb away from FOXG1 and its deletion is implicated in facial dysmorphism in a case of Rett Syndrome (RTT)\(^{133}\). This variant could also have an impact on FOXG1 expression due to its close proximity.

chr14:32547404 falls within an intron of ARHGAP5. A clear H3K27Ac peak implicates an active regulatory element may be situated in the region. ARHGAP5 and CHN1 (ARHGAP2) belong to the same protein family, and CHN1 is associated with Duane retraction syndrome type 2. ARHGAP5 encodes Rho GTPase activating protein 5 that negatively regulates Rho GTPases and mediate cytoskeleton changes. Similar to CHN1, ARHGAP5 might also responsible to a neural development pathway and contribute to strabismus.

The current study identifies a linked region of non-syndromic strabismus with high confidence and proposes three candidate variants, but the exact causal mutation remains elusive. Future work will be discussed in Chapter 3.
Table 2.1 Coverage and variants from whole exome sequencing (WES).

<table>
<thead>
<tr>
<th>Bowtie + BWA</th>
<th>Reads processed</th>
<th>Percentage of reads aligned (with good quality)</th>
<th>Mean of coverage (exons)</th>
<th>Total variants</th>
<th>Coding variants (not synonymous)</th>
<th>Shared heterozygous variants</th>
<th>&lt;= 1% shared heterozygous variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORGE_336_001</td>
<td>50,655,069</td>
<td>84.4%</td>
<td>27.42</td>
<td>96889</td>
<td>770</td>
<td>119</td>
<td>60</td>
</tr>
<tr>
<td>FORGE_336_011</td>
<td>50,576,080</td>
<td>88.6%</td>
<td>28.47</td>
<td>99068</td>
<td>758</td>
<td>119</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 2.2 Non-tolerant shared variants based on WESs from two distant relatives.

* Warning! Low confidence.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Coordinate</th>
<th>Ref</th>
<th>Alt</th>
<th>Amino acid</th>
<th>Gene symbol</th>
<th>Predicted impact</th>
<th>Allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>58385748</td>
<td>G</td>
<td>A</td>
<td>A337V</td>
<td>ZNF814</td>
<td>DAMAGING</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>74212518</td>
<td>G</td>
<td>A</td>
<td>P445S</td>
<td>GTF2IRD2</td>
<td>DAMAGING</td>
<td>0.000449</td>
</tr>
<tr>
<td>9</td>
<td>43091410</td>
<td>T</td>
<td>C</td>
<td>E805G</td>
<td>ANKR20A3</td>
<td>DAMAGING</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>69423844</td>
<td>C</td>
<td>G</td>
<td>Q714E</td>
<td>ANKR20A4</td>
<td>DAMAGING</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>32557449</td>
<td>G</td>
<td>A</td>
<td>S24F</td>
<td>HLA-DRB1</td>
<td>DAMAGING *</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>144148856</td>
<td>A</td>
<td>T</td>
<td>K94M</td>
<td>RP3-377D14.1(NBPF9)</td>
<td>DAMAGING *</td>
<td>NA</td>
</tr>
<tr>
<td>17</td>
<td>45567593</td>
<td>ATCTCTC</td>
<td>ATCTC</td>
<td><em>105</em></td>
<td>MRPL45P2</td>
<td>N/A</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>8887542</td>
<td>TAACAACS</td>
<td>TAACA</td>
<td><em>350</em></td>
<td>ERI1</td>
<td>N/A</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 2.3  Selected syndromic strabismus associated loci.

CFEOM: congenital fibrosis of extraocular muscles  
DRS: Duane retraction syndrome  
PEOA: Progressive external ophthalmoplegia, autosomal dominant

<table>
<thead>
<tr>
<th>Loci</th>
<th>Gene</th>
<th>Inheritance pattern</th>
<th>Pheno</th>
<th>PMID</th>
</tr>
</thead>
<tbody>
<tr>
<td>16q24.3</td>
<td>TUBB3</td>
<td>Dominant</td>
<td>CFEOM1, CFEOM3A</td>
<td>Tischfield et al., 2010[134]</td>
</tr>
<tr>
<td>12q12</td>
<td>KIF21A</td>
<td>Dominant</td>
<td>CFEOM1, CREOM3B</td>
<td>Yamada et al., 2003[135]</td>
</tr>
<tr>
<td>11q13</td>
<td>ARIX (PHOX2A)</td>
<td>Recessive</td>
<td>CFEOM2</td>
<td>Nakano et al., 2001[136]</td>
</tr>
<tr>
<td>13q12.11</td>
<td></td>
<td>Dominant</td>
<td>CFEOM3C</td>
<td>Aubourg et al., 2005[137]</td>
</tr>
<tr>
<td>8q12-13</td>
<td></td>
<td>Dominant</td>
<td>DRS1</td>
<td>Pizzuti et al., 2002[61] &amp; Amouroux et al., 2012[67]</td>
</tr>
<tr>
<td>2q31.1</td>
<td>CHN1</td>
<td>Dominant</td>
<td>DRS2</td>
<td>Miyake et al., 2008[69]</td>
</tr>
<tr>
<td>3q23</td>
<td>FOXL2</td>
<td>De novo</td>
<td>DRS, blepharophimosis-ptosis-epicanthus inversus syndrome (BPES)</td>
<td>Vincent et al., 2005[138]</td>
</tr>
<tr>
<td>16q12.1</td>
<td>SALL1</td>
<td>N/A</td>
<td>DRS, Townes-Brocks syndrome (TBS)</td>
<td>Barry et al., 2008[139]</td>
</tr>
<tr>
<td>20q13.2</td>
<td>SALL4</td>
<td>Dominant</td>
<td>Okihiro syndrome (Duane radial ray syndrome)</td>
<td>Kohlhase et al., 2002[76] &amp; Al-Baradie et al. 2002[77]</td>
</tr>
<tr>
<td>22q11.2-22qter</td>
<td></td>
<td>De novo</td>
<td>DRS, cat-eye syndrome</td>
<td>Gómez-Lado et al., 2006[140]</td>
</tr>
<tr>
<td>5p13.3-13.2</td>
<td></td>
<td>De novo</td>
<td>DRS, type I Chiari malformation</td>
<td>Bayrakli et al., 2010[141]</td>
</tr>
<tr>
<td>4q27-31</td>
<td></td>
<td>De novo</td>
<td>DRS1, mild learning difficulties</td>
<td>Chew et al., 1995[77]</td>
</tr>
<tr>
<td>1q42.13-43</td>
<td></td>
<td>De novo</td>
<td>DRS, febrile convulsions, dysmorphic</td>
<td>Kato et al., 2007[142]</td>
</tr>
<tr>
<td>2q13, 10q24.2-26.3, 20q13.12, 22q11.11-q11.22</td>
<td></td>
<td>De novo</td>
<td>DRS, other systemic abnormalities</td>
<td>Smith et al., 2010[143]</td>
</tr>
<tr>
<td>13q12.2-13</td>
<td></td>
<td>Dominant</td>
<td>Moebius syndrome</td>
<td>Slee et al., 1991[144]</td>
</tr>
<tr>
<td>15q25</td>
<td>POLG</td>
<td>Dominant</td>
<td>PEOA1</td>
<td>Van Goethem et al., 2001[145] &amp; Lamantea et al., 2002[146]</td>
</tr>
<tr>
<td>Loci</td>
<td>Gene</td>
<td>Inheritance pattern</td>
<td>Phenotype</td>
<td>PMID</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>---------------------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>4q35</td>
<td>ANT1</td>
<td>Dominant</td>
<td>PEOA2</td>
<td>Kaukonen et al., 2000&lt;sup&gt;147&lt;/sup&gt;</td>
</tr>
<tr>
<td>10q24</td>
<td>C10orf2</td>
<td>Dominant</td>
<td>PEOA3</td>
<td>Suomalainen et al., 1995&lt;sup&gt;148&lt;/sup&gt;, Li et al., 1999&lt;sup&gt;149&lt;/sup&gt; &amp; Spelbrink et al., 2001&lt;sup&gt;150&lt;/sup&gt;</td>
</tr>
<tr>
<td>17q23</td>
<td>POLG2</td>
<td>Dominant</td>
<td>PEOA4</td>
<td>Longley et al., 2006&lt;sup&gt;151&lt;/sup&gt; &amp; Young et al., 2011&lt;sup&gt;152&lt;/sup&gt;</td>
</tr>
<tr>
<td>8q23</td>
<td>RRM2B</td>
<td>Dominant</td>
<td>PEOA5</td>
<td>Tyynismaa et al., 2009&lt;sup&gt;153&lt;/sup&gt; &amp; Fratter et al., 2011&lt;sup&gt;154&lt;/sup&gt;</td>
</tr>
<tr>
<td>10q21</td>
<td>DNA2</td>
<td>Dominant</td>
<td>PEOA6</td>
<td>Ronchi et al., 2013&lt;sup&gt;155&lt;/sup&gt;</td>
</tr>
<tr>
<td>11q24.2</td>
<td>ROBO3</td>
<td>Recessive</td>
<td>Gaze palsy, horizontal, with progressive scoliosis</td>
<td>Jen et al., 2004&lt;sup&gt;156&lt;/sup&gt;</td>
</tr>
<tr>
<td>7p15.2</td>
<td>HOXA1</td>
<td>Recessive</td>
<td>Athabaskan brainstem dysgenesis syndrome, Bosley-Salih-Alorainy syndrome</td>
<td>Tischfield et al., 2005&lt;sup&gt;157&lt;/sup&gt; &amp; Bosley et al., 2008&lt;sup&gt;158&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Table 2.4  Loci and corresponding hg19 coordinates associated with strabismus, either syndromic or non-syndromic.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>p arm</th>
<th>q arm</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>q42.13-q43</td>
<td>227000000-243700000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>q13, q31-q32.1</td>
<td>110200000-114400000, 169700000-189400000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>q21-q22,q23</td>
<td>121900000-142800000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>q27-q31,q35</td>
<td>120800000-155600000, 183200000-191154276</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>p13.3-13.2</td>
<td>33800000-33800000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>q26</td>
<td>161000000-164500000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>p15, p22.1</td>
<td>q31.2</td>
<td>209000000-28800000, 4500000-7300000, 114600000-117400000</td>
</tr>
<tr>
<td>8</td>
<td>q12-q13, q23, q24.21</td>
<td>55500000-73900000, 106200000-117700000, 127300000-131500000</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>q21.3-q22.1, q24-q26.3</td>
<td>645000000-74900000, 97000000-135534747</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>q13, q23, q24.2</td>
<td>63400000-77100000, 110400000-121200000, 123900000-127800000</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>q12, q24.32</td>
<td>38200000-46400000, 125900000-129300000</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>q12.2-q13</td>
<td>27800000-40100000</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>q25</td>
<td>78300000-89100000</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>p13.2-p12.3</td>
<td>q12.1, q24.3</td>
<td>79000000-21200000, 47000000-52600000, 88700000-90354753</td>
</tr>
<tr>
<td>17</td>
<td>q23</td>
<td>57600000-62600000</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>q13.11</td>
<td>32400000-35500000</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>q13.12-q13.2</td>
<td>42100000-55000000</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>q11.1-11.22</td>
<td>14700000-23500000</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.5  Number of shared and total variants and log2(ratio) for each chromosome in strabismic and control groups.

Strab S – number of shared variants between two strabismic individuals; Strab T – number of total variants from two strabismic individuals; Strab S/T – number of shared variants over number of total variants for two strabismic individuals;
Control S – number of shared variants between two control individuals; Control T – number of total variants from two control individuals; Control S/T – number of shared variants over number of total variants for two control individuals;

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Strab S</th>
<th>Strab T</th>
<th>Strab S/T</th>
<th>Control S</th>
<th>Control T</th>
<th>Control S/T</th>
<th>Log2[(Strab S/T)/(Control S/T)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr22</td>
<td>135</td>
<td>3057</td>
<td>0.0442</td>
<td>230</td>
<td>3407</td>
<td>0.0675</td>
<td>-0.612</td>
</tr>
<tr>
<td>chr1</td>
<td>834</td>
<td>15045</td>
<td>0.0554</td>
<td>1249</td>
<td>16224</td>
<td>0.0770</td>
<td>-0.474</td>
</tr>
<tr>
<td>chr20</td>
<td>164</td>
<td>3845</td>
<td>0.0427</td>
<td>221</td>
<td>4122</td>
<td>0.0536</td>
<td>-0.330</td>
</tr>
<tr>
<td>chr21</td>
<td>191</td>
<td>3494</td>
<td>0.0547</td>
<td>257</td>
<td>3822</td>
<td>0.0672</td>
<td>-0.299</td>
</tr>
<tr>
<td>chr6</td>
<td>526</td>
<td>9269</td>
<td>0.0567</td>
<td>652</td>
<td>9627</td>
<td>0.0677</td>
<td>-0.255</td>
</tr>
<tr>
<td>chr4</td>
<td>334</td>
<td>7707</td>
<td>0.0433</td>
<td>428</td>
<td>8537</td>
<td>0.0501</td>
<td>-0.210</td>
</tr>
<tr>
<td>chr8</td>
<td>283</td>
<td>5674</td>
<td>0.0499</td>
<td>348</td>
<td>6129</td>
<td>0.0568</td>
<td>-0.187</td>
</tr>
<tr>
<td>chr12</td>
<td>455</td>
<td>7340</td>
<td>0.0620</td>
<td>552</td>
<td>7946</td>
<td>0.0695</td>
<td>-0.164</td>
</tr>
<tr>
<td>chr16</td>
<td>379</td>
<td>6356</td>
<td>0.0596</td>
<td>443</td>
<td>6670</td>
<td>0.0664</td>
<td>-0.156</td>
</tr>
<tr>
<td>chr15</td>
<td>305</td>
<td>5032</td>
<td>0.0606</td>
<td>368</td>
<td>5512</td>
<td>0.0668</td>
<td>-0.139</td>
</tr>
<tr>
<td>chr3</td>
<td>488</td>
<td>8572</td>
<td>0.0569</td>
<td>570</td>
<td>9322</td>
<td>0.0611</td>
<td>-0.103</td>
</tr>
<tr>
<td>chr5</td>
<td>503</td>
<td>7122</td>
<td>0.0706</td>
<td>553</td>
<td>7414</td>
<td>0.0746</td>
<td>-0.079</td>
</tr>
<tr>
<td>chr19</td>
<td>540</td>
<td>7887</td>
<td>0.0685</td>
<td>566</td>
<td>8126</td>
<td>0.0697</td>
<td>-0.025</td>
</tr>
<tr>
<td>chr17</td>
<td>491</td>
<td>7370</td>
<td>0.0666</td>
<td>518</td>
<td>7783</td>
<td>0.0666</td>
<td>0.001</td>
</tr>
<tr>
<td>chr2</td>
<td>618</td>
<td>10653</td>
<td>0.0580</td>
<td>674</td>
<td>11856</td>
<td>0.0568</td>
<td>0.029</td>
</tr>
<tr>
<td>chr7</td>
<td>478</td>
<td>7212</td>
<td>0.0663</td>
<td>504</td>
<td>7934</td>
<td>0.0635</td>
<td>0.061</td>
</tr>
<tr>
<td>chr13</td>
<td>167</td>
<td>3966</td>
<td>0.0421</td>
<td>171</td>
<td>4272</td>
<td>0.0400</td>
<td>0.073</td>
</tr>
<tr>
<td>chr9</td>
<td>370</td>
<td>6423</td>
<td>0.0576</td>
<td>346</td>
<td>6981</td>
<td>0.0496</td>
<td>0.217</td>
</tr>
<tr>
<td>chr10</td>
<td>488</td>
<td>7644</td>
<td>0.0638</td>
<td>444</td>
<td>8148</td>
<td>0.0545</td>
<td>0.228</td>
</tr>
<tr>
<td>chr18</td>
<td>188</td>
<td>3135</td>
<td>0.0600</td>
<td>172</td>
<td>3375</td>
<td>0.0510</td>
<td>0.235</td>
</tr>
<tr>
<td>chr11</td>
<td>510</td>
<td>8043</td>
<td>0.0634</td>
<td>445</td>
<td>8784</td>
<td>0.0507</td>
<td>0.324</td>
</tr>
<tr>
<td>chr14</td>
<td>316</td>
<td>4525</td>
<td>0.0698</td>
<td>234</td>
<td>4709</td>
<td>0.0497</td>
<td>0.491</td>
</tr>
</tbody>
</table>
Table 2.6  Number of shared variants between different individuals on chr14:20,000,000-35,000,000.

C3-1: first member of control family 3; C3-2: second member of control family 3;
C4-1: first member of control family 4; C4-2: second member of control family 4;
C5-1: first member of control family 5; C5-2: second member of control family 5.

<table>
<thead>
<tr>
<th>Comparison pair (unrelated)</th>
<th>013 &amp; C4-2</th>
<th>013 &amp; C4-1</th>
<th>013 &amp; C5-1</th>
<th>013 &amp; C5-2</th>
<th>013 &amp; C3-1</th>
<th>013 &amp; C3-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of shared variants</td>
<td>13316</td>
<td>13501</td>
<td>13668</td>
<td>13809</td>
<td>13899</td>
<td>14057</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison pair (IBD)</th>
<th>C5-1 &amp; C5-2</th>
<th>001 &amp; 014</th>
<th>001 &amp; 013</th>
<th>013 &amp; 014</th>
<th>C4-1 &amp; C4-2</th>
<th>C3-1 &amp; C3-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of shared variants</td>
<td>17107</td>
<td>18459</td>
<td>18957</td>
<td>19499</td>
<td>20162</td>
<td>20396</td>
</tr>
</tbody>
</table>
Table 2.7  Sequencing data quality for whole genome sequencing (WGS).

<table>
<thead>
<tr>
<th>BGI pipeline</th>
<th>Clean Reads</th>
<th>Percentage of reads aligned</th>
<th>Average of coverage (exons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>1,249,733,158</td>
<td>96.2%</td>
<td>38.86</td>
</tr>
<tr>
<td>013</td>
<td>1,058,300,000</td>
<td>95.46%</td>
<td>37.27</td>
</tr>
<tr>
<td>014</td>
<td>1,069,350,000</td>
<td>96.13%</td>
<td>38.46</td>
</tr>
</tbody>
</table>

Table 2.8  Coverage and variants of linked region from WGS with the BGI pipeline.

<table>
<thead>
<tr>
<th>Bowtie + BWA</th>
<th>Number clean reads in the region</th>
<th>Mean of coverage (chr14:200,000,000-350,000,000)</th>
<th>Total variants</th>
<th>Variants with a frequency of 1% (dbSNP)</th>
<th>Rare heterozygous variants</th>
<th>Rare shared heterozygous variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>6,456,829</td>
<td>41.14</td>
<td>26,046</td>
<td>4,727</td>
<td>3,142</td>
<td>671</td>
</tr>
<tr>
<td>013</td>
<td>5,874,158</td>
<td>38.75</td>
<td>26,689</td>
<td>4,948</td>
<td>3,352</td>
<td>671</td>
</tr>
<tr>
<td>014</td>
<td>5,794,351</td>
<td>38.25</td>
<td>25,836</td>
<td>4,748</td>
<td>3,077</td>
<td>671</td>
</tr>
</tbody>
</table>

Table 2.9  Coverage and variants of linked region from WGS with the in-house pipeline.

<table>
<thead>
<tr>
<th>Bowtie + BWA</th>
<th>Number of clean reads in the region</th>
<th>Mean of coverage (chr14:200,000,000-350,000,000)</th>
<th>Total variants</th>
<th>Variants with a frequency of 1% (dbSNP)</th>
<th>Rare heterozygous variants</th>
<th>Rare shared heterozygous variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>5,814,112</td>
<td>38.76</td>
<td>21,516</td>
<td>3,057</td>
<td>1,849</td>
<td>378</td>
</tr>
<tr>
<td>013</td>
<td>5,480,743</td>
<td>36.53</td>
<td>21,886</td>
<td>3,057</td>
<td>1,880</td>
<td>378</td>
</tr>
<tr>
<td>014</td>
<td>5,431,398</td>
<td>36.21</td>
<td>21,551</td>
<td>2,950</td>
<td>1,756</td>
<td>378</td>
</tr>
</tbody>
</table>
Table 2.10  Prioritized WGS variants from the linked region.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Number of times seen in 2365 GSC samples</th>
<th>Feature</th>
<th>RegulomeDB score</th>
<th>CADD score</th>
<th>Gene of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr14: 28052726 A &gt; C</td>
<td>3</td>
<td>potential FoxA1 TFBS</td>
<td>5</td>
<td>6.274</td>
<td>1.2 Mb away from FOXG1</td>
</tr>
<tr>
<td>chr14:29247628 TAAACAAAACAA &gt; TAAACAAA</td>
<td>1</td>
<td>H3K27me3 mark</td>
<td>5</td>
<td>14.65</td>
<td>2.5 Kb away from FOXG1</td>
</tr>
<tr>
<td>chr14:32547404 A &gt; T</td>
<td>1</td>
<td>H3K27Ac peak</td>
<td>2b</td>
<td>14.12</td>
<td>ARHGAP5</td>
</tr>
</tbody>
</table>
Figure 2.1  Pedigree for the subject family with non-syndromic strabismus.
Figure 2.2  Branch 1 of the subject family with non-syndromic strabismus.
Figure 2.3  Simplified branch 1 showing the genotyped individuals (with study ID) and ancestors required to link them.
Figure 2.4  Branch 2 of the subject family.
Figure 2.5  Branch 3 of the subject family.
Figure 2.6  Sanger sequence results of *GTF2IRD2* variant for multiple individuals.
Figure 2.7  Variant density plot of shared variants on chromosome 11.

The left panel is based on data from both strabismic individuals, and the right panel is based on data from two unrelated controls. The region (chr11: 48,200,000-56,000,000) that is enriched with shared variants is boxed.
Figure 2.8  Variant density plot of shared variants on chromosome 14.

The left panel is based on data from both strabismic individuals, and the right panel is based on data from two unrelated controls. The region (chr14: 24,600,000-32,300,000) that is enriched with shared variants is boxed.
Figure 2.9  Simulated maximum LOD scores under the alternative hypothesis of linkage, under a range of genetic models.
Figure 2.10 Screenshot of region chr14: 20,000,000 – 35,000,000 from UCSC genome browser. The linked region chr14: 22,779,843 – 32,908,192 is gene-poor.
Figure 2.11  Venn diagram of WGS analysis for 001, 013, and 014 based on the BGI pipeline.
Figure 2.12  Venn diagram of WGS analysis for 001, 013, and 014 based on the in-house pipeline.
Figure 2.13  Single nucleotide variant chr14:28052726 A > C impacts binding score for FoxA1 transcription factor binding site.
Chapter 3: Future Directions

The recruitment of additional family members and, ideally additional families with similar genetic transmission patterns, is critical to validate a variant as the causal mutation. Two sets of distant relatives are currently being recruited for participation in order to refine the linked region. The recruiting process can be challenging, but we have seen increased interest in the family as the project progresses. As the number of participants grows and more phenotypic information is collected, it will be interesting to calculate penetrance based on a Bayesian method. In addition, we have recently identified a three-generation family with genetic strabismus and similar clinical description as in our subject family; enrolment of the family is underway. By publishing our initial findings, we expect to draw attention from clinicians and find additional families with the same transmission pattern. The same linked region might be identified in a proportion of these families, especially families with matching phenotypes. Such findings could support our finding and facilitate a greater understanding of the genetic mechanism underlying non-syndromic strabismus.

Another approach is to analyze locus transmission in parent-child pairs (affected and controls) from many distinct families. Statistical evidence can be collected to show that the identified linked region is involved in a subset of cases with a genetic component. Data generated by existing genome-wide association studies (GWAS) for strabismus can be reprocessed for such statistical analysis. Moreover, samples from GWAS collection can be subjected to sequencing or targeted position analysis to obtain the ratio of variants across strabismus and control populations. One strabismus GWAS has been reportedly underway led by a group at Harvard University, although no publications have appeared with results.
Contradictory to its name, whole genome sequencing does not cover the complete genome (approximately 98% coverage was generated for the three WGS datasets in this study). Missing regions typically include difficult to sequence regions with high GC content or regions that can be mapped to more than one location and thus discarded in the alignment process. Among the sequenced regions, some are made up of simple repeats. One study has shown that variants in poorly covered regions (due to the repetitive nature of the sequences) can be causal\textsuperscript{161}. As sequencing capacities and alignment tools improve, we may discover changes in the linked region that we are not able to detect today.

The sensitivity of variant calling is low when the coverage is shallow. High coverage is required to determine individual genotypes with high confidence\textsuperscript{126}. Many called variants are not reliable when the depth of coverage is below 50-fold. Currently, the clinical standard is 100-fold, and this has been shown to improve calling accuracy. Ideally, we would like to obtain uniform 100-fold depth of coverage for the linked region. Alternatively, we could combine WGS raw data from 001, 013, and 014 to increase the variant detection sensitivity, but an increased false positive rate might be observed.

While we have initiated this project without substantial budget, additional NGS across family members may be pursued as the cost of NGS decreases. As \textit{de novo} genome assembly approaches mature, we expect to re-process the WGS data in order to search for small-scale structural changes. Alternatively, software may emerge to allow high-quality detection of structural changes based on alignment methods.

Detailed phenotyping is essential to study a heterogeneous condition like strabismus since multiple forms and different underlying genetic elements may be involved. We need to ascertain if participants have other forms of strabismus and refine phenotyping. As we
expand our collaboration network, we might develop a non-invasive characterization strategy with the help from ophthalmologists, such as MRI imaging of the muscle and nerve. A better understanding of the cause of the eye misalignment, muscular vs. neuronal, will allow us to prioritize genes of interest. We will then focus on a smaller profile and examine the impact of specific variants.

As I highlighted in the introduction, there is no established animal model to study non-syndromic strabismus. One challenge is the difficulty to characterize strabismus in animals. Researchers use similar criteria to measure eye misalignment in predators, such as cats and monkey. However, the measurement is more difficult in prey animals, for which eyes are not on the same plane and are more widely spread to maximize the horizontal vision to avoid predators. It is unclear whether models such as mice and zebrafish can exhibit strabismus-related phenotypes. If not, what alternative phenotypes should we look for in these model organisms when a strabismus-causing mutation is introduced? This is a crucial question to ask before we delve into validation. Without a clear predictable phenotype, we cannot test our hypothesis that the selected mutation is sufficient for the development of strabismus.

A more approachable method is the development of functional analysis for each variant in appropriate human cell lines. The CRISPR-Cas9 system is able to introduce changes in any region in the genome, including regulatory regions. This system has drawn a lot of attention due to its high efficacy and ease of operation. The lentiCRISPR system allows simultaneous transduction of Cas9 and a single-guide RNA (sgRNA), and the targeting specificity is high. Modified Cas9 can have very specific function and provide us an even more effective way to incorporate the variant to a cell line, if it is known to express the gene
of concern. We can then examine whether a specific variant leads to a change at the expression level. Although the link between the expression profile and strabismus phenotype remains unclear, it is one important step for the project before we reach out to establish animal model.

As we briefly discussed in the introduction, non-syndromic strabismus may have a strong environmental component. External forces, such as trauma, can also lead to strabismus development. Environmental factors, such as maternal smoking and decreasing socioeconomic status, can also contribute to strabismus development\textsuperscript{163–165}. In the subject family, the pedigree strongly suggests a high penetrance autosomal dominant condition, but certain requisite carriers do not report an eye phenotype. Giving the various expressivities across family members, common genetic and/or environmental modifiers might be involved. Future investigation may explore the potential correlates with the phenotype in the family, or across a broader group if additional families are identified.

In conclusion, this project builds upon a unique chance to understand the genetics of non-syndromic strabismus. While additional work will be required to bridge candidate variants to the phenotype, the linkage mapping to the specific segment of chromosome 14 will allow the global community to pursue several exciting lines of research. Based only on the results obtained in this thesis, it should be possible to assess the contribution of the region to genetic transmission of strabismus. Genotyping the region across the entire family should provide a clear measure of penetrance. And further characterization of the family may reveal specific structural properties in tissues related to strabismus.

The validation of a variant is not the end of this exciting project, which sets out to solve the 2400-year-old mystery; instead it is the beginning of a long journey to understand
vision and its related molecular network in the modern time. Ultimately, we will develop a precise understanding of the root cause of strabismus, at least in this family, and begin to contemplate improved strabismus treatment.
References


doi:10.1002/ccr3.11


