UNIVERSITY OF OXFORD

NEXT GENERATION SEQUENCING TO IDENTIFY NEW GENETIC CAUSES OF FAMILIAL CRANIOSYNOSTOSIS

AKIKO SONEDA HASHIMOTO

GREEN TEMPLETON COLLEGE

THEIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Trinity Term 2017

Supervisors: Prof. Andrew O. M. Wilkie, Dr Stephen .R.F. Twigg
Weatherall Institute of Molecular Medicine
John Radcliffe Hospital
Oxford, United Kingdom

Statement of Originality

All the work described in this thesis is my work with the following exceptions:

Exome sequencing of III-1 (chapter 3), IV-3 (chapter 4) and III-4 (chapter 6) was undertaken by Dr Steve Twigg (Clinical Genetics Group, Weatherall Institute of Molecular Medicine (WIMM)). The SNP Chip segregation analysis of the family in chapter 3 was performed and analysed by Dr Steve Twigg.

Whole genome sequencing of III-4 (chapter 3) and IV-5 (chapter 4) was performed by Complete Genomics. Whole genome sequencing of IV-3, V-1 (chapter 4) and III-1 (chapter 6) was undertaken by Oxford Biomedical Research Centre at the John Radcliffe Hospital (Oxford University Hospitals NHS Trust, Oxford, UK).

The bioinformatics pipeline of exome and whole genome sequence data was developed by Dr Simon McGowan at Computational Biology Research Group, WIMM. The bioinformatics pipeline of resequencing of candidate genes (chapter 3 and chapter 6) was undertaken by Dr Nils Koelling (Clinical Genetics Group, WIMM).

High-resolution SNP arrays of III-4 and IV-3 (chapter 4) and III-1 (chapter 6) were run by High Throughput Genomics at Wellcome Trust Centre for Human Genetics (Oxford University Hospitals NHS Trust, Oxford, UK). The analysis was done under the supervision of Dr Samantha Knight at Wellcome Trust Centre for Human Genetics.

Array CGH of individuals in chapter 5 (I-1, I-2, II-2 and III-2) was performed by Oxford Medical Genetics Laboratories Churchill Hospital (Oxford University Hospitals NHS Trust, Oxford, UK), who identified the *de novo* duplication of chr4: 81160133-81884710.

The strategy for CRISPR/Cas9 targeting described in chapter 3 was planned together with Dr Philip Hublitz of the Genome Engineering Service, WIMM. Karyotyping and pluripotency testing of targeted mouse ES cells and microinjection into mouse blastocysts were undertaken by Mouse Transgenic and Knockout Facility, WIMM (Dr Jackie Sharp).

Capture-C described in chapter 5 was performed by Yan Zhou (Clinical Genetics Group, WIMM).

The LC-MS/MS or HPLC/UV assays for measuring retinoic acid described in chapter 6 were undertaken by Dr Maureen A. Kane at the Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland (Baltimore, Maryland, USA).

Abstract

Next generation sequencing to identify new genetic causes of familial craniosynostosis

A thesis submitted for the degree of DPhil by Akiko Hashimoto
Green Templeton College, Oxford, Trinity Term 2017

Craniosynostosis, the premature fusion of one or more cranial sutures, is one of the most common craniofacial abnormalities. There is considerable genetic heterogeneity and the underlying pathophysiological mechanisms are diverse, occurring at multiple stages of cranial suture biogenesis. A genetic diagnosis, currently made in ~25% of cases, is an important aspect of the multidisciplinary approach to clinical management and genetic counselling. The aim of this thesis is to identify novel genetic causes of craniosynostosis in unsolved familial cases.

Four families were selected and investigated using a combination of next generation sequencing and copy number analysis. Candidate changes were followed up through segregation and functional analysis to determine if they were causal. The most convincing finding was a homozygous deletion identified in the 5' UTR (untranslated region) of dehydrogenase/reductase 3 (*DHRS3*) in a consanguineous family with syndromic coronal synostosis. This deletion was shown to lead to loss of DHRS3 activity and an increase of all-trans-retinoic acid (ATRA), a known teratogen that had previously been deduced to contribute to craniosynostosis. These findings are consistent with a novel embryopathy caused by excess ATRA, leading to skeletal and cardiac defects.

In a second large family with syndromic metopic synostosis, a nonsynonymous variant in *HIST1H2AD*, encoding one of the canonical histone H2A components of the nucleosome, was identified. Although a link between histone mutation and craniosynostosis is unknown, we speculated that this variant could perturb gene expression during development of the metopic suture. To gain support for causality, two approaches were undertaken; 1. Search for additional patients with a variant in H2A genes including resequencing in a large cohort of patients with craniosynostosis. 2. Creation of Hist1h2ad mutant Embryonic Stem cells using CRISPR/CAS9; Generation of the mouse model is ongoing. Chromosomal rearrangements can be pathogenic through altering the topological associated domain (TAD) structure. A *de novo* tandem duplication on chromosome 4 including *FGF5* and *C4orf22* was identified in a third family (affected mother and daughter) with multiple suture synostosis. The duplication overlaps TAD boundaries, leading to the hypothesis that this may induce mis-expression of genes in or near the duplication. Gene expression analysis was undertaken. To investigate this further, work using Capture-C, a technique used to discover local chromatin structure and interactions, is ongoing.

In summary, I have identified a new craniosynostosis disease gene, *DHRS3*, suggesting an additional aetiology in osteogenesis of the cranial suture caused by disruption of retinoic metabolism. In addition, two loci, *HIST1H2AD* and *FGF5/C4orf22* duplication, are plausible candidates for craniosynostosis with possible roles in disturbing neural crest development and RAS/MAPK signalling, respectively. Finally, this work illustrates a combined genomic technology approach to resolve difficult cases that may be applicable apply to other diseases in which a genetic cause is suspected.

Acknowledgements

This work was carried out at the Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford.

My first and foremost thanks must go to my supervisors, Prof. Andrew Wilkie and Dr Steve Twigg, who have given me an incredible amount of advice, guidance and encouragement. I have been fortunate to learn clinical genetics from such great clinical geneticist/scientists.

I am very grateful to all the members of 4 families in my project. I also would like to thank clinicians, especially to Dr Deirdre Cilliers, Dr Louise Wilson, Dr Jenny Morton and Dr Siddharth Banka for recruiting and contacting the families.

My gratitude also goes to Simon McGowan and Nils Koelling for bioinformatics support. I would like to thank Julie Phipps for recruiting all the patients and families, Sue Butler for assistance with cell culture, and John Frankland and Tim Rostron for dideoxy sequencing services, Kevin Clark for cell sorting, Jackie Sloane-Stanley for microinjection, Maureen Kane for retinoid biochemical assay, Samantha Knight for her supervision on CNV analysis and Philip Hublitz for his great advice on CRISPR/Cas9 gene-editing.

A big thank you also goes to everyone working in the Wilkie group for their help and advice during my DPhil, including Kerry Miller, Anne Goriely, Yan Zhou, Geoff Maher, Eduardo Calpena, Vikram Sharma, Aimee Fenwick, Deborah Lloyd, Hana Mlcochova, Hannah Ralph, Marie Bernkopf, Indira Taylor and Karen Crawford. It has been a pleasure to work in such a friendly and stimulating environment.

I am also grateful to my thesis committee supervisors, Prof. Julian C Knight and Dr Zameel Cader for their scientific advice. I would like to thank my college advisor, Prof. Anna Gloyn for her help both in the scientific and others.

I was funded by Honjo International Scholarship Foundation, GlaxoSmithKline International Scholarship and Goodger and Schorstein award. I appreciated their support.

I would like to thank all my friends and family for their encouragement and support, especially my parents, Koichi and Mieko Soneda.
Finally, at from the bottom of my heart, my huge thank you goes to my husband, Kenji Hashimoto, and lovely three children, Umika, Kai and Keito. They are my greatest source of happiness.

Contents

Figures

Tables

Abbreviations

ABS	Antley-Bixler syndrome
array CGH	array-based comparative genomic hybridisation
ASD	atrial septal defects
ATRA	all-trans retinoic acid
BMP	bone morphogenetic protein
CNV	copy number variation
CSF	cerebrospinal fluid
CT	computed tomography
CYP	cytochrome P450
DDD	Deciphering Developmental Disorders
DGV	Database of Genomic Variants
DS	deleterious score
E	embryonic day
EBV	Epstein–Barr virus
ES	embryonic stem
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
FBS	foetal bovine serum
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
gnomAD	Genome Aggregation Database
HRO	heterozygote rule out
LOD	logarithm of the odds
LoF	loss of function
MIP	molecular inversion probes
MLPA	multiplex ligation-dependent probe amplification
MMS	methyl methanesulfonate
MRI	magnetic resonance imaging
NCCs	neural crest cell
NGS	next generation sequencing
OFC	occipito-frontal circumference
PAM	protospacer adjacent motif
PDA	patent ductus arteriosus
PFO	patent foramen ovale
POR	cytochrome P450 oxidoreductase
PORD	cytochrome P450 oxidoreductase deficiency
PRS	Pierre Robin sequence
RA	retinoic acid
ROH	runs of homozygosity
RT	reverse transcriptase

SSPNN	Splice Site Prediction by Neural Network
sgRNA	single-guide RNAs
SHH	sonic hedgehog
SIBD	shared identity-by-descent
SNP	single nucleotide polymorphisms
SV	structural variations
TAD	topologically associated domains
TGF-β	transforming growth factor-beta
qPCR	quantitative PCR
UTR	untranslated region
VCF	variant calling file
WES	whole exome sequencing
WGS	whole genome sequencing
WNT	wingless-related family members

Chapter 1 INTRODUCTION

1.1 OVERVIEW

Craniosynostosis, the premature fusion of the cranial sutures, is one of the most common craniofacial abnormalities with a prevalence of 1 in 2,100 – 2,300 births (Boulet et al., 2008; Lajeunie et al., 1995). Malformation of the skull may prevent brain growth and cause a variety of medical problems including both functional impairment and an abnormal craniofacial appearance. The causes are diverse and can be monogenic, polygenic, chromosomal and environmental. In terms of monogenic causes, around 60 specific genes shown to be mutated recurrently in craniosynostosis have been identified. There is considerable genetic heterogeneity with a diversity of underlying pathophysiological mechanisms.

The Clinical Genetics group in the WIMM have been recruiting craniosynostosis patients and their family members, collaborating with 4 major craniofacial units in the UK, into a research study to identify new genetic causes and molecular base of craniosynostosis. Since the number of disease genes has been identified, very few familial cases remain undiagnosed. In the course of my study, four families with a strong suspicion of an underlying genetic cause, but for which previous genetic testing failed to identify the genetic causes, were selected for the investigation. The aim of my work has been to identify genetic causes in those unsolved familial craniosynostosis cases by using next generation sequencing (NGS) combined with multiple other genetic approaches.

In this chapter, the anatomy and development of skull bones, clinical features and molecular basis of craniosynostosis and various approaches to find causative genes in rare Mendelian disorders are reviewed.

1.2 DEVELOPMENT OF SKULL AND SUTURE

1.2.1 ANATOMY AND CRANIOSYNOSTOSIS

The human skull is divided into the neurocranium (the parts protecting the brain and the 5 organs of special sense) and viscerocranium (the parts related to the digestive and respiratory systems, the facial skeleton). The neurocranium consists of two frontal bones, two parietal bones, two temporal bones, the occipital, sphenoid and ethmoid bones, forming the calvarial vault and skull base. These bones are joined together with fibrous tissue at the seams, known as sutures (Beederman et al., 2014; Norton, 2016; Shier D, 2012). Normal calvarial vault development is dependent on the coordinated expansion of the brain and overlying skeletal elements (Levi et al., 2012). The sutures are the primary area of bone formation during skull growth and as the brain expands hugely during foetal development and early childhood, continuous patency is essential. Craniosynostosis describes the premature fusion of one or more of the cranial sutures: secondary distortion of skull shape occurs because of a combination of lack of growth at the fused suture, and compensatory overgrowth at the non-fused sutures as shown in Figure 1-1 (Johnson and Wilkie, 2011).

The cranial vault is formed through intramembranous ossification, in which condensed mesenchyme cells directly differentiate into osteoblasts and form bone tissue without any cartilaginous intermediate, while most other bones including some in the skull bone are formed through endochondral ossification (Nie et al., 2006b).

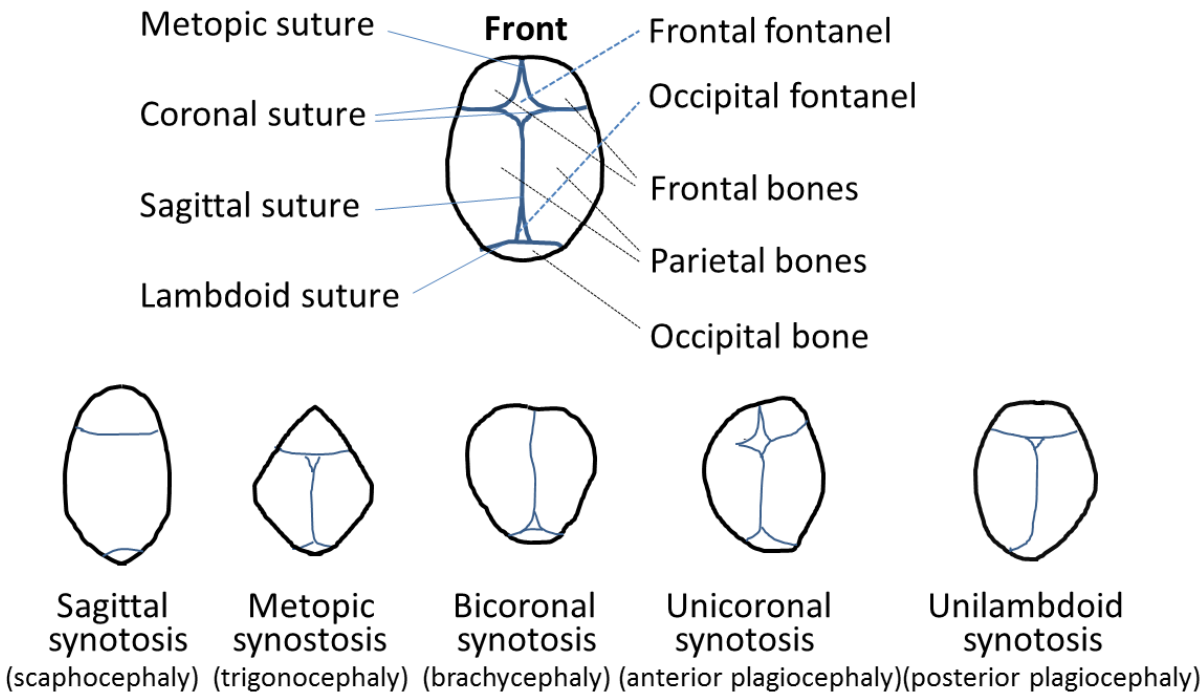


FIGURE 1-1 HUMAN CALVARIAL BONES

The top panel shows the normal human skull looking down from above. Bones forming the calvaria, sutures and fontanelles are represented. The bottom images depict the skull shapes that result from synostosis of different sutures.

1.3 CRANIAL SUTURE DEVELOPMENT

1.3.1 EMBRYOGENESIS

The structure of the skull is similar between human and other mammalian species such as rabbits, mice and rats, therefore these species have been used as research tools to examine suture biology and pathology (Levi et al., 2012). Here, suture biology is explained in terms of mouse embryogenesis.

The vertebrate embryo becomes the trilaminar embryonic disc consisting of ectoderm, mesoderm and endoderm at the end of gastrulation. Subsequently, neural folds develop from the ectoderm, the outer layer of the trilaminar disc. Part of the anterior neural fold undergoes an epithelial to mesenchyme transition forming the neural crest, derived from the closed neural tube. This process is followed by neural crest migration and in the mouse this begins at embryonic day (E) 8 and takes 2 days. In humans, it occurs between E19 and E38 (Bronner and LeDouarin, 2012; Cordero et al., 2011; Erlebacher et al., 1995; Opperman, 2000).

Stem cell specification and migration

Simultaneous with neural crest migration, cells originating from cephalic paraxial mesoderm migrate rapidly from their original paraxial location to near a population of neural crest cells (NCC) positioned just above the developing eye (this region is called the 'supraorbital regulatory centre' by Deckelbaum et al. (2010), referring to a potential role as an organizing centre) at E8.5-9.5 in mouse. This occurs in response to sonic hedgehog (SHH) signalling in the notochord (Deckelbaum et al., 2012) (Figure 1-2 A). Thus, following cell migration, developing head has mesenchyme originating from two sources; neural crest and cells from mesoderm.

Lineage Commitment and boundary formation

The NCC and mesodermal lineages at the supraorbital regulatory centre differentiate to osteoprogenitors and osteoblasts forming frontal and parietal bones, respectively (Figure 1-2 -B). These are separated by mesenchyme of mesodermal origin that constitutes the future coronal suture (Deckelbaum et al., 2012; Jiang et al., 2002a). The coronal sutures are formed by migration and expansion of supraorbital mesenchymal between E11.0 and E13.5 with osteogenic differentiation initiated from E12.0 onwards (Deckelbaum et al., 2012) (Figure 1-2-C).

Various transcription factors are believed to be involved in this lineage commitment stage; Deckelbaum et al. (2012) reported the expression of *En1* (engrailed 1), *Msx2* and *Twist1* in the supraorbital regulatory region between E11.5-13.5. Twigg et al. (2015) reported the expression of *Zic1* at E11.5-12.5 in a region overlapping the supraorbital regulatory centre. Signalling pathways involving bone morphogenetic proteins (BMPs), wingless-related family members (WNTs), and fibroblast growth factors (FGFs) are considered to be involved at this stage by interacting with these transcription factors (Mishina and Snider, 2014).

Osteogenic Proliferation and Differentiation

During intramembranous ossification, a population of mesenchymal cells in the fibrous suture remain undifferentiated and function as a growth centre, providing a source of stem cells that differentiate into osteoprogenitor cells, which subsequently proliferate and differentiate into osteoblasts (Lana-Elola et al., 2007). Osteoblast differentiation occurs at the osteogenic front (the bone margins), the main centre of cranial vault growth. Osteoblasts express collagen I and synthesize bone matrix. At this stage, key genes such as *TWIST1* are required to maintain the undifferentiated fate of mid-suture stem cells (Lana-Elola et al., 2007; Takarada et al., 2016). FGF and fibroblast growth factor receptor (FGFR) signalling maintains the balance between the proliferation of osteoprogenitor cells and osteoblastic differentiation (Morriss-Kay and Wilkie, 2005).

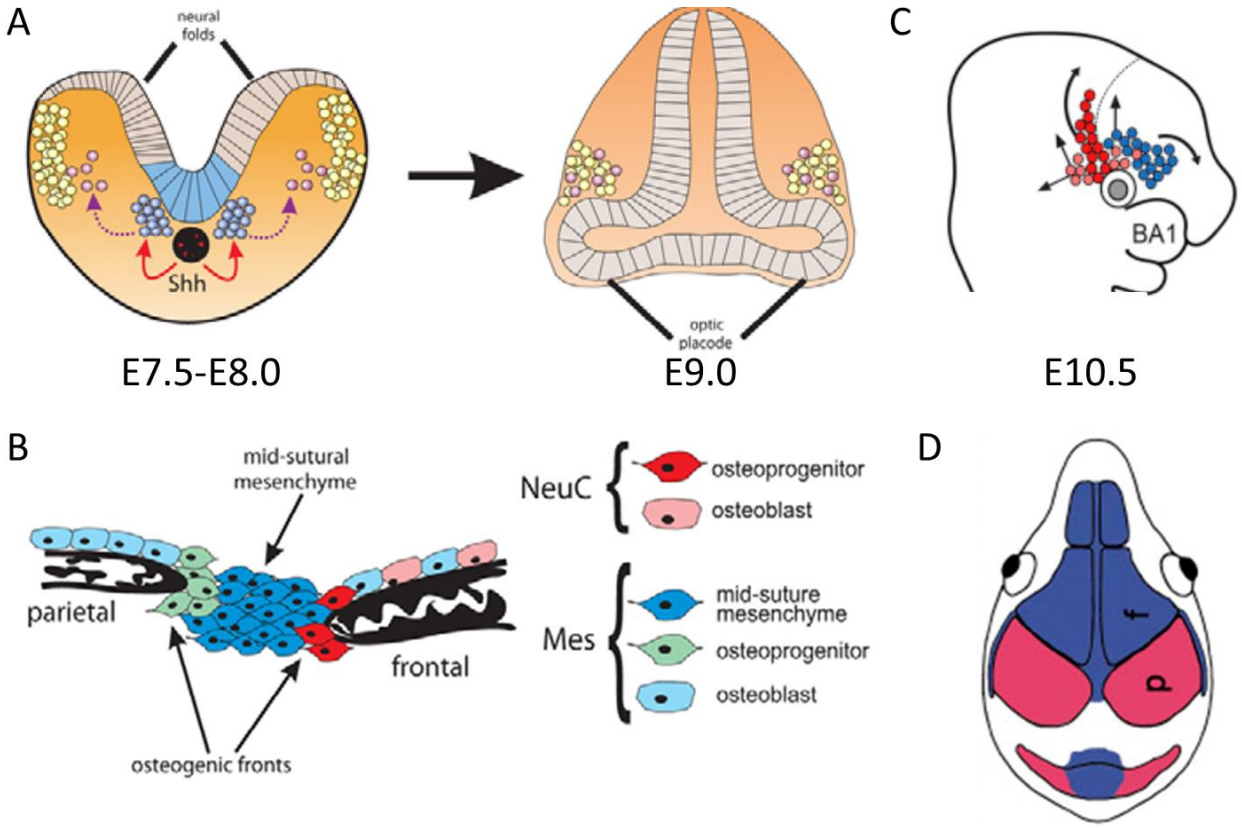


FIGURE 1-2 CRANIAL SUTURE DEVELOPMENT

A Cells originating from cephalic paraxial mesoderm migrate rapidly from their original paraxial position to near neural crest cells (yellow circles) in response to *Shh* signalling. Blue circles indicate *Shh*-responsive *Gli1*+Mesodermal cells. Red circles indicate *Gli1* lineage-derived mesenchyme.
B Schematic cross-section through developing parietal and frontal bones with sutural mesenchyme in between.
C Supraorbital regulatory centre at E10.5, showing cells of mesodermal and neural crest origin (pink/red and blue, respectively). Cells migrate to the future coronal suture (red), parietal bone (mesodermal cells, pink), and frontal bone (neural crest, blue, with a small contribution from mesoderm).
D Dual origin of skull bones, from neural crest (blue) and cephalic mesoderm (red). f, frontal bone; p, parietal bone.
A and B are from (Deckelbaum et al., 2012). C and D are from (Twigg and Wilkie, 2015), originally from (Deckelbaum et al., 2012) and (Jiang et al., 2002b).

1.3.2 METOPIC SUTURE

In the developmental anatomy of the mammalian skull, the major sutures have different tissue origins; the metopic suture is the only suture that entirely forms within neural crest, while the entire length of coronal sutures is formed by neural crest-mesoderm interfaces. The sagittal and lambdoid sutures have dual neural crest/mesodermal origins partially (illustrated in Figure 1-2–D). Therefore, pathophysiology of metopic synostosis is unique from other suture synostosis. As illustrated in Figure 1-3 (bottom, yellow boxes), metopic synostosis mainly arises through abnormalities of neural crest development.

1.3.3 HUMAN SKULL DEVELOPMENT

In human, the early formation of the cranium begins at 23-26 days of gestational age, at which time the position of the future temporal bone is marked by the optic placode. Shortly after, the optic placode is transformed into a vesicle and the boundary of future sphenoid and occipital bones is identifiable as the notochord separates from the rostral neural tube (Cohen and MacLean, 2000). Around 13 weeks of gestational age, ossification centres first appear and those mineralizing bones expand sufficiently so that their borders approach each other forming non-ossified zones, the sutures around 18 weeks. By the term, the skull bones are formed and separated by the sutures and fontanelles (Kirmi et al., 2009a; Sadler and Langman, 2006).

1.4 CLINICAL ASPECTS OF CRANIOSYNOSTOSIS

1.4.1 CLASSIFICATION

Based on clinical examination craniosynostosis can be classified in two complementary ways (Kimonis et al., 2007):

1) The number of sutures involved: simple (involving one suture) or multiple (involving two or more sutures)

2) If there is an associated phenotype: nonsyndromic (occurring without other anomalies) or syndromic (accompanied by other dysmorphic features or developmental defects).

1.4.2 THE PREVALENCE OF DIFFERENT FORMS OF CRANIOSYNOSTOSIS

A population-based study from the Metropolitan Atlanta Congenital Defects Program reported that 84% of infants with craniosynostosis presented with isolated craniosynostosis, 7% with multiple defects, and 9% with syndromes (Boulet et al., 2008). The most commonly affected suture is the sagittal suture, accounting for approximately 40%–60% of all cases and is frequently seen in a nonsyndromic form with a frequency of 1.5-2 in 10,000 children (Boulet et al., 2008; Hunter and Rudd, 1976; Lajeunie et al., 1996). Coronal synostosis occurs less often with an approximate frequency of 0.7-1 in 10,000 children. Bicoronal synostosis is the most common presentation found in patients with multiple suture synostosis and can be either syndromic or nonsyndromic. (Boulet et al., 2008; Buchanan et al., 2017; Cohen and MacLean, 2000). Historically, metopic synostosis was the third most frequent single suture synostosis with the frequency 0.67 in 10,000 children (Lajeunie et al., 1998), although, a recent marked increase in the prevalence of metopic synostosis has been reported (van der Meulen et al., 2009) Thus, the metopic synostosis is the second most common after sagittal synostosis in the more recent studies.

Genetic diagnoses accounted for 24% of all craniosynostosis cases; 77% had a single-gene mutation, and 14% had a chromosomal abnormality (Sharma et al., 2013). Cases with genetic diagnoses were found to be associated with increased rates of many complications in a cohort-based analysis carried out in Oxford. Children with nonsyndromic craniosynostosis were more likely to have a causative mutation if the synostosis was unicoronal or bicoronal than if it was sagittal or metopic (Wilkie et al., 2010).

1.4.3 UNDERLYING CAUSES

Craniosynostosis can be caused by both environmental and genetic factors.

In terms of environmental factors, several are reported as being associated with increased risk include; mechanical forces such as intrauterine constraint caused by oligohydramnios, nulliparity and multiple gestation (Graham et al., 1980; Graham et al., 1979; Higginbottom et al., 1980;

Sanchez-Lara et al., 2010), maternal thyroid disease (Rasmussen et al., 2007), teratogenic exposures to sodium valproate (Lajeunie et al., 2001) and maternal smoking (Carmichael et al., 2008).

Mutations in more than 60 genes have been identified in craniosynostosis to date. Of these, the major causative genes are described in detail in section 1.5. In addition, the contribution of polygenes is suspected, especially in midline synostosis. This is directly supported in the case of nonsyndromic sagittal synostosis, as susceptibility loci have been identified near *BMP2* and *BBS9* in a genome-wide association study (Justice et al., 2012). Both genes are known to involve in skeletal development.

1.4.4 PRESENTATION

Craniosynostosis usually becomes apparent because of progressive head shape deformity between the last third of pregnancy and the end of the first year of life (Twigg and Wilkie, 2015). The abnormal skull growth can cause various problems such as raised intracranial pressure, impaired cerebral blood flow, airway obstruction, impaired vision and hearing, motor and mental developmental delay and cosmetic issues (Wilkie, 1997). Occasionally, these may require acute intervention, so early diagnosis is crucial for the management and prevention of complications. Spiral computed tomography (CT) scan and 3D reconstructions using both bone and soft tissue windows should clearly reveal the patency, or closure of each individual suture, so that this information is available for diagnosis, surgical planning and follow-up. CT scan of the brain is also useful to detect associated anatomical abnormalities and complications, for example, agenesis of corpus callosum, ventriculomegaly and hydrocephalus (Johnson and Wilkie, 2011; Kirmi et al., 2009b).

1.4.5 NORMAL TIMING OF SUTURE CLOSURE

After birth, 70% of the adult brain weight is achieved by 18 months and 90% by age 5–8 years. Correspondingly, cranium volume increases by 17% compared to birth at 3 months old, and by 25% at 6 months old. The head size increases by about 2.5 cm during the 2nd year, with gradual expansion

until 20 years old (Huelke, 1998). The normal timing of suture and fontanelle closure is shown in Table 1-1.

TABLE 1-1 THE CLOSURE TIMING OF SUTURES AND FONTANELLES

Suture	Timing of closure	Fontanelle	Timing of closure
Metopic	3-9 months[1]	Frontal fontanelle	15-18 months
Coronal	The closure begins at 22-24 years[2], but completion continues into the third decade of life and beyond	Occipital fontanelle	3-6 months
Sagittal		Anterolateral fontanelle	3 months to 2 years
Lambdoid		Posterolateral fontanelle	3 months to 2 years

[1] Finding from 3D CT scan. (Vu et al., 2001), [2] From (Cohen and MacLean, 2000).
Adapted from (Aviv et al., 2002)

1.5 MAJOR GENETIC CAUSES

In a cohort-based study conducted in the Oxford craniofacial unit, twenty-four percent of children born between 1998 and 2006 (n=402) had a proven genetic cause for their craniosynostosis and the major contributions to the proven genetic cases (*n* = 98) were made by chromosome abnormalities (13%) and mutations in *FGFR2* (28%), *FGFR3* (19%), *TWIST1* (16%), *ERF* (5%),*FNB1* (4%) and *TCF12* (4%) (Sharma et al., 2013). Major pathogenic genes for craniosynostosis and associated clinical findings including those genes are summarised in Table 1-2.

TABLE 1-2 PATHOGENIC GENES AND DISORDERS OF CRANIOSYNOSTOSIS

Gene	Disorder	Suture affected	Inheritance pattern	Major clinical findings	Molecular mechanism	Typical mutations
EFNB1	Craniofrontonasal syndrome	Coronal	XLD	Frontonasal dysplasia with hypertelorism, cleft of the nasal tip, other finger and joint anomalies, abnormal clavicles.	Loss of function Heterozygous females more severely affected	Truncation and missense
ERF	ERF-related craniosynostosis	Multisuture	AD	Exorbitsm, midface hypoplasia, Chiari type I malformation	Loss of function	Truncation and missense
FGFR1	Pfeiffer syndrome	Coronal	AD	Marked cutaneous syndactyly, broad medially deviated thumbs and halluces	Gain-of-function Ligand dependent	p.P252R
FGFR2	Apert syndrome	Coronal, multisuture	AD (n)	Midface hypoplasia, syndactyly of the hands and feet with a tendency to the fusion of bony structures	Gain-of-function Ligand dependent	p.S252W, p.P253R
"	Beare-Stevenson syndrome	Multisuture, coronal	AD (n)	Choanal atresia, furrowed skin on the scalp and neck, acanthosis nigricans of the flexural regions in surviving children	Gain-of-function, Ligand independent	p.S372C, p.Y375C
"	Bent bone dysplasia	Coronal	AD (n)	Osteopenia, poor mineralisation of the calvaria, dysmorphic facial features, bent long bones; perinatal lethal	Gain-of-function, Ligand independent	Missense substitutions in the TM domain.
"	Crouzon syndrome	Coronal, multisuture	AD	Prominent exorbitism, midface hypoplasia with beaked nose, hands and feet are usually clinically normal	Gain-of-function, Ligand independent	Missense substitutions in IgIIIa and IgIIIc domains

"	Pfeiffer syndrome	Multisuture	AD (n)	Crouzonoid face with broad thumbs, broad great toes, brachydactyly, and variable syndactyly	Gain-of-function, ligand independent	Missense substitutions in IgIIIa and IgIIIc domains
FGFR3	Muenke syndrome	Coronal	AD	Nonspecific features – plagiocephaly, brachydactyly, cone-shaped epiphyses of phalanges	Gain-of-function, ligand dependent	p.P250R
IHH	Craniosynostosis, Philadelphia-type	Sagittal	AD	Cutaneous and distal osseous syndactyly, less commonly paraxial polydactyly of feet	Increased dosage of IHH	Duplications
IL11RA	Craniosynostosis and dental anomalies	Multisuture	AR	Maxillary hypoplasia, dental anomalies (delayed tooth eruption, supernumerary teeth) minor digit abnormalities, conductive hearing loss	Loss-of-function	Mainly missense
MEGF8	Carpenter syndrome	Metopic	AR	Hypertelorism with the broad depressed nasal bridge, epicanthus, upslanting palpebral fissures, high arched eyebrows, lateralisation defects, brachydactyly, syndactyly	Loss-of-function	Missense and truncations
MSX2	Boston craniosynostosis	Sagittal Coronal Multisuture	AD	None diagnostic; syndrome defined by specific amino acid substitutions		p.Pro148His, p.Pro148Leu
POR	Antley-Bixler syndrome with genital anomalies	Multisuture	AR	Radio-humeral synostosis, arachnodactyly, multiple joint contractures, bowed femora, abnormal steroidogenesis and genital abnormalities	Loss-of-function Some residual activity	Mainly missense, some truncations
RAB23	Carpenter syndrome	Multisuture	AR	obesity, cardiac defects, polysyndactyly, brachydactyly, genu valgum, hypogenitalism, umbilical hernia, learning disability	Loss-of-function	Nonsense and truncations
SMAD6[1]	Non-syndromic midline craniosynostosis	Sagittal Metopic	AD	Incomplete penetrance. A common risk variant near *BMP2* (rs1884302) increases the penetrance of *SMAD6* mutations.	Loss-of-function	Nonsense, truncations and missense

TCF12	TCF12-related craniosynostosis	Coronal	AD	Saethre-Chotzen-like syndrome with a dysmorphic face, ears and minor limb anomalies, incomplete penetrance	Haploinsufficiency	Mainly truncations
TWIST1	Saethre-Chotzen syndrome	Coronal	AD	Low frontal hairline, facial asymmetry, eyelid ptosis, down-slanting palpebral fissures, hypertelorism	Haploinsufficiency	Truncations, missense and deletions

AD, autosomal dominant; AR, autosomal recessive; XLD, X-linked dominant; (n) usually arises by *de novo* mutation
Modified from (Twigg and Wilkie, 2015) 1. (Timberlake et al., 2016; Timberlake et al., 2017)

1.5.1 MOLECULAR MECHANISMS

Molecular mechanisms and pathophysiology for craniosynostosis are summarised in Twigg and Wilkie, 2015 according to the cranial bone development in mouse. Figure 1-3 (Twigg and Wilkie, 2015) illustrates a pathophysiological framework for craniosynostosis.

1.5.1.1 FGF/FGFR SIGNALLING PATHWAY

The most important molecular mechanism is FGF/FGFR signalling. Gain-of-function mutations in *FGFR1, FGFR2* and *FGFR3* are a major cause of syndromic craniosynostosis, including Apert, Crouzon, Muenke and Pfeiffer syndromes. The mammalian FGF family is comprised of eighteen secreted proteins that interact with four signalling tyrosine kinase FGFRs (Ornitz and Itoh, 2015). The biology of FGFs and FGFRs is complex because of alternative splicing of the genes encoding FGFR1, -2 and -3, multiple ligands and proteoglycans activate the signalling. During embryogenesis and subsequent development, their expression occurs differentially in terms of location, timing and kinetic factors (Cohen and MacLean, 2000). In osteogenic differentiation and cranial vault growth, FGF/FGFR signalling plays a critical role. FGFs secreted by osteoblasts at the differentiating edge of the bones activate receptors involved in both osteoprogenitor cell proliferation and the conversion of these cells to differentiating osteoblasts (Morriss-Kay and Wilkie, 2005). Disruption of *Fgfr* signalling in mice alters calvarial osteoprogenitor proliferation and leads to abnormal bone growth (Eswarakumar et al., 2004).

1.5.1.2 BMP SIGNALLING PATHWAY

Although, bone morphogenetic proteins (BMP) signalling is not illustrated in Figure 1-3, this is another important signalling pathway in osteogenesis. BMPs are members of transforming growth factor-beta (TGF-β) family. TGF-β/BMPs have roles in bone formation during mammalian development and exhibit regulatory functions, through both canonical Smad-dependent pathways (TGF-β/BMP ligands, receptors and Smads) and non-canonical Smad-independent signalling. (Chen et al., 2012)

During early head development, BMP signalling participates in the formation of neural crest in cooperation with other critical signalling pathways, notably including the WNT and FGF pathways

(Stuhlmiller and Garcia-Castro, 2012). Subsequently, in craniofacial skeletogenesis, BMP signalling is an early inductive signal required for committed cell migration, condensation, proliferation and differentiation (Nie et al., 2006a).For example; Kim et al. (1998) demonstrated that *Bmp2* and *Bmp4* are present in the osteogenic fronts of cranial sutures and induce expression of *Msx1* and *Msx2* (Kim et al., 1998); Following TGF-β/BMP induction, both the Smad and MAPK pathways converge at the *Runx2* gene to control mesenchymal precursor cell differentiation. The BMP signalling pathway interacts with MAPK, Wnt, Hedgehog, Notch, and FGF in osteoblast differentiation and bone formation (Chen et al., 2012b).

1.5.1.3 RETINOIC ACID SIGNALLING CRANIAL DEVELOPMENT

Retinoic acid (RA) is a steroid derivative of vitamin A that is essential for embryonic development. RA signalling is one of the known molecular mechanisms of craniosynostosis as shown in Figure 1-3. Here, RA is discussed because a mutation in one of the RA metabolism proteins was identified in a family I have studied.

Pathogenicity relating to retinoids in humans

Vitamin A is an essential nutrient which is required for several life processes such as vision, reproduction, growth, cell differentiation, immune system and embryonic development in vertebrates. Retinoid sources come from dietary intake, nutritional supplements and some therapeutic agents and either a deficiency or an excess of vitamin A and related compounds (retinoids) can cause abnormal morphological development (Collins and Mao, 1999). Vitamin A deficiency, which is an important public health problem in developing countries, can result in xerophthalmia, immunodeficiency (especially in respiratory and digestive systems) and weight loss (Sommer et al., 1984). Whereas, excessive maternal intake of vitamin A is also teratogenic. The defects affecting tissues derived from the cranial neural crest are observed in humans, as in laboratory animals (Cohlan, 1954; Rothman et al., 1995). The evidence that humans are susceptible to retinoid teratogenesis is of concern because retinoids are used as drugs for the treatment of skin diseases and various forms of cancers. Lammer et al (1985) reported that vitamin A derivative, 13-

cis-RA (isotretinoin), which is used in the treatment of severe cystic acne, increased the risk of spontaneous abortions and congenital malformations including craniofacial, cardiac, thymic and central nervous structure. A deleterious effect on cephalic neural-crest cell activity in early embryonic periods was described as the underlying mechanism. (Lammer et al., 1985)

Retinoic acid signalling in cranial development

Studies suggested that RA induces the differentiation of osteoblasts to osteocytes during calvarial morphogenesis. This results in a bone matrix with reduced unmineralised osteoid and accelerated mineralisation, leading to craniosynostosis. RA has been suggested as of potential relevance in the maintenance of bone-mineral homeostasis (Jeradi and Hammerschmidt, 2016) (Laue et al., 2011).

Two syndromes are known to manifest craniosynostosis in association with defects involving the RA pathway. These are Antley-Bixler syndrome (ABS) (MIM: 201750), the skeletal phenotype of Cytochrome P450 oxidoreductase deficiency (PORD) caused by recessive mutation in *POR* (Cytochrome P450 Oxidoreductase), and CYP26B1 (Cytochrome P450 Family 26 Subfamily B Member 1) (MIM: 605207) deficiency due to recessive mutations in *CYP26B1* (Table 1-2). POR is a flavoprotein that donates electrons to all the steroidogenic cytochrome P450 enzymes including CYP26 enzymes (Fluck et al., 2004). *Por* $^{-/-}$ mice are early embryonic lethal and have significantly elevated levels of all-trans retinoic acid (ATRA; the major bioactive form of retinoic acid in vertebrates) (Otto et al., 2003). CYP26B1 is one of the cytochrome P450 enzymes involved in the degradation of ATRA.

The skeletal phenotype in ABS is characterised by craniosynostosis, midface hypoplasia, choanal atresia, radiohumeral or radioulnar synostosis, joint contractures, arachnodactyly, and bowing of the femora (Pandey and Fluck, 2013). An overlapping phenotype is reported in CYP26B1 deficiency patients, characterised by craniofacial abnormalities (including craniosynostosis) and radiohumeral fusion (Laue et al., 2011). Morton et al. (2016) reported a CYP26B1 deficiency patient with multisuture synostosis, radiohumeral synostosis and apparent intellectual disability, describing the phenotype as having significant similarities to ABS and Pfeiffer syndrome (Morton et al., 2016)..

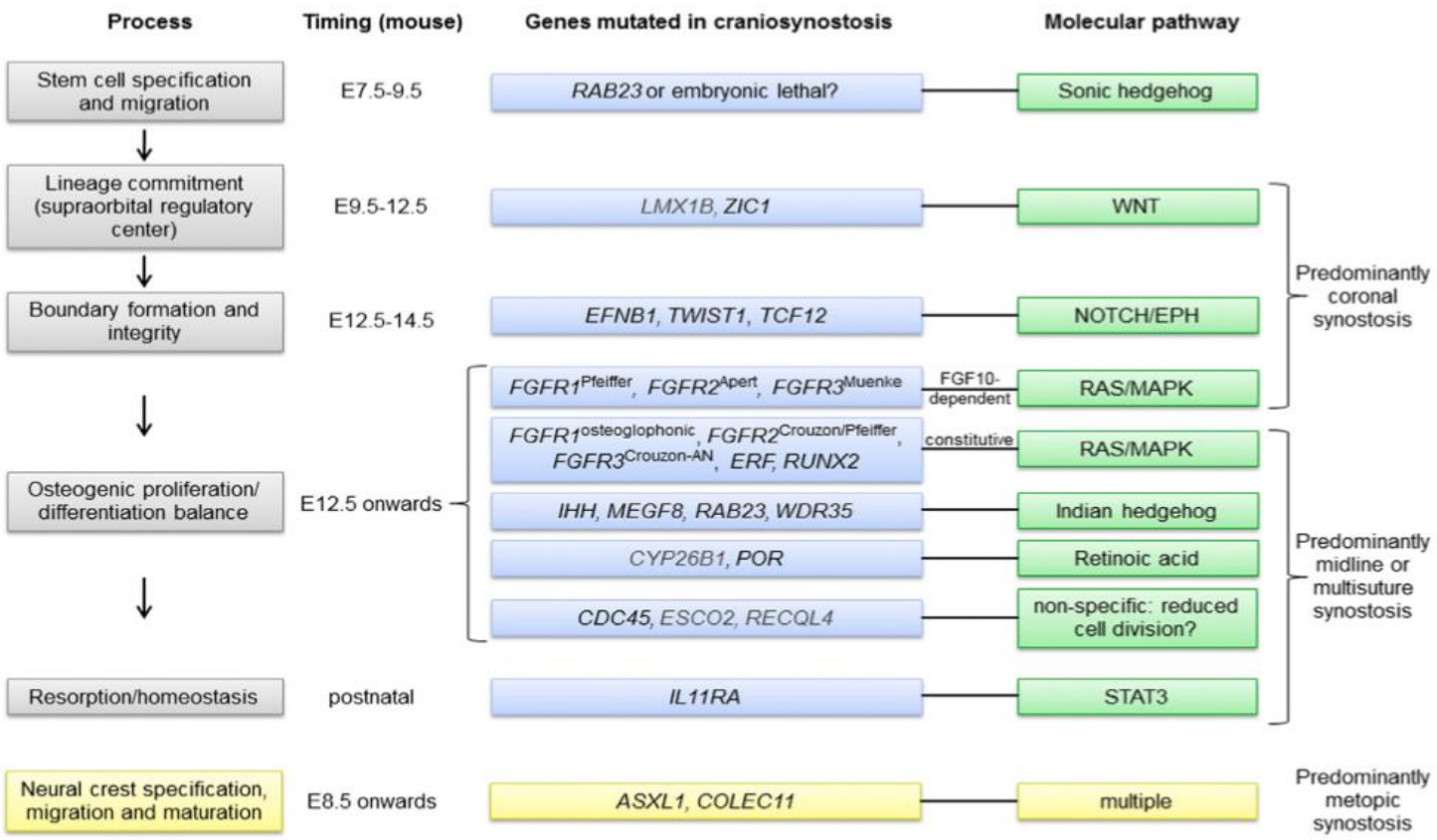


FIGURE 1-3 A GENETIC PATHOPHYSIOLOGICAL FRAMEWORK FOR CRANIOSYNOSTOSIS

Processes in suture formation are displayed on the left (grey boxes). Other panels show relative timing of events in mouse, genes involved (blue boxes; black for core genes and grey for additional genes), and signalling pathways proposed to be affected (green boxes). Patterning of the supraorbital regulatory centre and boundary formation are events particular to coronal suture development, therefore, mutations disrupting these processes lead predominantly to coronal craniosynostosis. Later developmental apply to all sutures. The origin of the metopic suture within the neural crest suggests that maturation of this tissue might be a common factor predisposing to metopic synostosis (yellow boxes, bottom). From (Twigg and Wilkie, 2015)

1.6 NEXT GENERATION SEQUENCING (NGS) IN RARE DISEASE GENETICS

Since the first whole exome sequencing (WES) experiment successfully revealed that mutations in *DHODH* cause Miller syndrome (a rare skeletal disorder involving severe micrognathia, cleft lip/palate, hypoplasia or aplasia of the posterior elements of the limbs, coloboma of the eyelids, and supernumerary nipples) (Ng et al., 2010), WES has had a major impact on the identification of causative genes for Mendelian disorders. In the 2 years after the Miller syndrome publication, mutations in more than 180 novel genes have been discovered in the same manner (Boycott et al., 2013). Subsequently, Gilissen et al. (2014) demonstrated the power of whole genome sequencing (WGS) in molecular diagnosis by applying WGS to patients with intellectual disability who had not received a molecular diagnosis after extensive genetic prescreening, including microarray-based CNV (copy number variation) studies and WES. This approach resulted in a diagnostic yield of 42% (Gilissen et al., 2014). NGS is now used widely in both clinical and research settings. Especially for rare Mendelian disorders, the pace of discovery of mutations in new genes underlying disease has increased significantly after 2010, and the proportion of discoveries made by conventional approaches were replaced by WES and WGS. From 2013 onwards, WES and WGS have discovered nearly three times as many disease genes as conventional approaches (Boycott et al., 2017).

The Clinical Genetics group at the WIMM has applied WES or WGS to investigate genetic causes in a cohort of craniosynostosis cases, selected by clinical or molecular geneticists as being high-priority cases, but in whom prior clinical genetic testing had been negative. In the past few years, seven new genes responsible for previously undiagnosed craniosynostosis cases were discovered by the group (Table 1-3). This approach also identified causal variants in known craniosynostosis-related genes where routine clinical testing failed to detect the mutation (Miller et al., 2017)

1.6.1 STRATEGIES FOR DISEASE GENE IDENTIFICATION

A key challenge in finding novel disease genes from next generation sequencing is how to distinguish pathogenic variants from the high background of non-pathogenic polymorphisms or sequencing errors. Depending on the exome enrichment set used, the sequencing platform, the algorithms used for mapping, and variant calling, typically between 20,000 and 50,000 variants are identified per sequenced exome (Gilissen et al., 2012).

Initially, all variants go through a prioritisation pipeline: 1) Filter variants with variant frequency. ExAC[1] and gnomAD[2] are applied as the polymorphism reference databases. The applied frequency filter depends on the mode of inheritance assumed, for example, under autosomal dominant inheritance model, we do not expect to find the variant within the database. 2) prioritise loss of function variants (stop mutations and frameshifts), nonsynonymous and splice-site variants. *In-silico* tools which predict possible impact of the variant on the structure and function of human proteins add further information for prioritisation. We apply a deleterious score (DS; minimum score 0, maximum 6), which is calculated by the number of prediction methods that predict a variant to be deleterious: SIFT - Deleterious (sift<=0.05) (Kumar et al., 2009), PolyPhen2 - Probably damaging (>=0.957) (Adzhubei et al., 2013), Likelihood ratio test (LRT) - Deleterious (Chun and Fay, 2009), MutationTaster - Disease causing (Schwarz et al., 2010), GERP++ - ≥5 (Davydov et al., 2010) and PhyloP - >1 (Siepel et al., 2006). This combined approach was shown to have increased discriminating power (Fu et al., 2013). 3) lastly, gene function and protein structure such as domains are taken into consideration.

As a basis for the filtering process, factors such as, the mode of inheritance of a trait, the pedigree or population structure and whether a phenotype arises owing to *de novo* or inherited variants, strongly influences both experimental design and analytical approach (Bamshad et al., 2011). Below,

[1]ExAC browser Exome Aggregation Consortium (http://exac.broadinstitute.org/): Database which provide variant information of 60,706 exome sequences from unrelated individuals.
[2]gnomAD browser, Genome Aggregation Database (http://gnomad.broadinstitute.org/): Database which provide 123,136 exome sequences and 15,496 whole genome sequences from unrelated individuals. The number of individuals is based on the data from August 2017.

several strategies commonly used for the gene identification of Mendelian diseases through NGS are discussed. These are well illustrated in the application to craniosynostosis, which has successfully employed seven different strategies to identify new disease gene.

1.6.1.1 COHORT ANALYSIS ACROSS MULTIPLE UNRELATED SAMPLES

This strategy searches for mutations within the same gene shared by multiple unrelated, affected samples with a similar phenotype. Studies have shown that combining three to four unrelated individuals was sufficient to identify a causal gene for rare diseases (Hoischen et al., 2010; Hoischen et al., 2011). Ng et al. (2010) identified *DHODH* as a causal gene for Miller syndrome with this strategy, the first gene identification made by WES. They sequenced four affected individuals from three independent families (Ng et al., 2010). Accurate phenotyping and the assumption of a single involved locus are key factors for this strategy (Bamshad et al., 2011; Boycott et al., 2013; Gilissen et al., 2012).

The cohort method has also been successfully applied to craniosynostosis, identifying causal variants in *TCF12* (Sharma et al., 2013).

1.6.1.2 TRIO ANALYSIS TO DETECT *DE NOVO* OR COMPOUND HETEROZYGOUS VARIANTS

Trio analysis is a powerful approach particularly applicable to gene discovery in disorders for which most cases are sporadic. By sequencing the exome of the patient as well as his or her parents, *de novo* candidates can be selected by filtering out all inherited variants. This generally produces a handful of potential pathogenic variants; comparison of these variants between unrelated families can further reduce these to a single candidate gene (O'Roak et al., 2012b; Vissers et al., 2010). High sequencing accuracy and equal coverage in all three samples investigated are crucial as high numbers of false positives can occur due to failure to call the corresponding germline variants in one or other parent (Bamshad et al., 2011; Boycott et al., 2013; Gilissen et al., 2012). If non-penetrance occurs, this can undermine the validity of the *de novo* approach. *ZIC*1 is a craniosynostosis disease gene found by this approach using WGS followed by analysis for coding *de novo* variants (Twigg et al., 2015).

A recessive mode of inheritance can also yield sporadically affected cases, this hypothesis (homozygous or compound heterozygous) also needs to be considered. The *CDC45* discovery is an example of compound heterozygous variants in the same gene identified in 2 sporadic craniosynostosis cases (Fenwick et al., 2016).

1.6.1.3 MULTIPLE AFFECTED INDIVIDUALS WITHIN A PEDIGREE

For a family with a suspected monogenic inherited disorder, the presence of multiple affected family members provides supportive evidence for a monogenic basis and will suggest the likely pattern of inheritance. Sequencing of multiple affected individuals can be used to substantially reduce the number of prioritised shared variants. In addition, unaffected family members can be sequenced to exclude private benign variation (Gilissen et al., 2012).

In large families, for very rare alleles, the probability of *identity-by-descent* given the observation of *identity-by-state* is high, even among distantly related individuals. For example, two first cousins share a rare allele that is *identical-by-descent* in approximately one-eighth of the genome. Sequencing the two most distantly related individuals with the phenotype of interest in the large pedigree can substantially restrict the genomic search space (Bamshad et al., 2011).

1.6.1.4 LINKAGE ANALYSIS

Linkage analysis was developed to detect excess co-segregation of marker alleles potentially linked to the underlying trait in family data, using panels of microsatellites or more recently using genotyping microarrays. As a consequence of new technical developments and substantial increase of cost-effectiveness, direct identification of causative lesions by sequencing the whole exome/genome of affected individuals has largely replaced linkage analysis (Teare and Santibanez Koref, 2014). However, linkage analysis can still be an important and powerful analysis method for the identification of genes, often in conjunction with WGS filtering approaches (Ott et al., 2015). It can provide the location of interest which enables prioritisation of variants for further analysis, a measure of the strength of evidence for a particular region of the genome being involved in disease predisposition inheritance in a particular pedigree and information on haplotypes (Bailey-Wilson and

Wilson, 2011). Particularly when variants are missed by WES or WGS, linkage information can refocus on the regions for further analysis. This can be especially relevant in regions where structural and non-coding variants were not fully detectable using current analysis strategies.

The *ERF* was discovered by the initial SNP chip analysis of 4 family members to narrow down the search space (regions where 2 affected children shared allele with the affected mother were identified), combined with the exome data (Twigg et al., 2013).

1.6.1.5 RECESSIVELY INHERITED DISORDER

In the case of suspected consanguinity, with the initial assumption of the disease being caused by a homozygous variant inherited from both parents, homozygosity analysis provides a powerful approach, especially when the degree of consanguinity is relatively distant. Homozygous variants can be prioritised by their presence in large homozygous regions of the patient's genome. Short runs of homozygosity (ROH) up to 5 Mb are considered to be consistent with an outbred population. Multiple larger ROH spread across different chromosomes provides tell-tale evidence of consanguinity (McQuillan et al., 2008; Sund et al., 2013).

The exome sequencing data, as well as single nucleotide polymorphisms (SNP) microarray analysis, itself contains sufficient numbers of informative SNPs to allow reliable homozygosity mapping (Becker et al., 2011; Gilissen et al., 2012; Walsh et al., 2010). With this approach, sequencing only one person can be sufficient to reduce the number of variants for follow-up for disease gene identification in individual cases, and does not necessarily require additional family members (Gilissen et al., 2012).

In craniosynostosis, *MEGF8* was identified by searching for homozygous region uniquely present in the proband in SNP chip analysis, followed by WES (Twigg et al., 2012).

Other approaches to identification of disease genes include suspected mosaicism and a candidate gene approach. These are illustrated by craniosynostosis examples in Table 1-3.

TABLE 1-3 RECENTLY IDENTIFIED CRANIOSYNOSTOSIS DISEASE GENES

Strategy	**Genes identified**	**Reference**
cohort across analysis	*TCF12 (CDC45)*	(Sharma et al., 2013)
Autosomal dominant	*ERF*	(Twigg et al., 2013)
Autosomal recessive	*MEGF8*	(Twigg et al., 2012)
Compound heterozygous recessive	*CDC45*	(Fenwick et al., 2016)
De novo	*ZIC1*	(Twigg et al., 2015)
Mosaicism	*SMO*	(Twigg et al., 2016)
Candidate gene	*IL6ST*	(Schwerd et al., 2017)

1.7 CHROMOSOMAL REARRANGEMENT AND DISEASE

Structural variants (SV) include copy number neutral variants such as inversions and balanced translocations and CNV such as deletion, duplication or further multiplication of genomic segment copy on a chromosome, which occur during DNA recombination and replication-based processes (Carvalho and Lupski, 2016). The size varies from a hundred to millions of base pairs. Variants of 5MB or greater variants can be detected by standard cytogenetic analysis such as G-banded karyotyping. Molecular cytogenetic techniques, such as fluorescence in situ hybridisation (FISH), can define the extent and location involved in these chromosomal variants at a much higher resolution. Using genotyping array technologies, such as single nucleotide polymorphism (SNP) arrays and array-based comparative genomic hybridisation (array CGH), a large number of submicroscopic genomic imbalances have been identified (Iafrate et al., 2004; Rodriguez-Revenga et al., 2007).

1.7.1 DETECTION OF STRUCTURAL VARIANTS IN DATA FROM NEXT GENERATION SEQUENCING

The currently available array designs contain probe numbers that range from hundreds of thousands to several million. These are efficient at detecting CNVs, however, the detection can be restricted by the resolution of the arrays and the probe spacing that tiles the reference genome. Recently, in parallel to the broadening application of NGS in genome research, the investigation of SVs using WES or WGS data has become attractive in anticipation of replacing microarrays as the leading platform for the investigation of genomic rearrangements. Gonzaga-Jauregui et al. (2012) reported that by combining WGS with high-resolution array CGH, a higher number of CNVs in the low size range (<5 kb), especially averaging ~2 kb, were detected. This suggests the smaller size CNVs can be beyond the capability of conventional array CGH detection, and that the frequency of CNVs in the human genome can be much higher for smaller sized (<1 kb) CNVs (Gonzaga-Jauregui et al., 2012). Redin et al. (2017) searched balanced cytogenic abnormalities (BCAs) at nucleotide resolution in 273 subjects with a spectrum of congenital anomalies. WGS revised 93% of the karyotypes and revealed the limitations of conventional cytogenetic approaches. At least 33.9% of BCAs resulted in gene disruptions that were likely causative, 5.2% were associated with pathogenic genomic imbalances,

and 7.3% disrupted topologically associated domains (TADs) encompassing known syndromic loci (Redin et al., 2017).

To detect SVs using NGS data, four different strategies have been reported; Read-pair (RP), Read-depth (or read count, RC), Split-read (SR) methods and *de novo* assembly (AS) (Details are reviewed in Tattini et al., 2015). Each method has reported advantages and disadvantages, so that none of them fully captures SVs with high sensitivity and specificity. In our analysis, we applied DELLY (Rausch et al., 2012), which uses RP and SR approaches, on our WGS data (described in chapter 4). This is thought to be suitable for detecting copy-number variable deletion and tandem duplication events as well as balanced rearrangements such as inversions or reciprocal translocations (Rausch et al., 2012; Tattini et al., 2015).

1.7.2 POTENTIAL MECHANISMS FOR REARRANGEMENTS ASSOCIATED WITH DISEASE

Pang et al. (2010) reported that indels/CNVs and inversions affect approximately 1.2% and 0.3%, respectively, of average human genomes, impacting 4,867 genes, whereas, SNPs affect approximately 0.1% of the human genome (Pang et al., 2010). In mutations underlying severe intellectual disability, Gilissen et al. (2014) demonstrated that the cumulative estimate for WGS to reach a conclusive genetic diagnosis is 62%, of which 21% is accounted for by CNVs, while 39% is accounted for by single nucleotide variants (Gilissen et al., 2014). SVs are therefore a major source of genetic variation in human genomes and disease.

Whereas the pathogenicity of many SVs can be explained through the direct effect on gene expression levels caused by altering the dosage of a particular gene, more complicated mechanisms can underlie the consequences of SVs in non-coding regions. Such variants have the potential to disrupt the integrity of the genome, causing changes in the regulatory architecture that lead to pathogenic alterations of gene expression levels and pattern (Lupianez et al., 2015; Rodriguez-Revenga et al., 2007).

Amongst many disease-associated CNVs that have been reported, several are associated with craniosynostosis. Exonic deletions (of sizes 5-85 kb), an 11.3 kb tandem duplication and a 3.64 Mb deletion that lead to haploinsufficiency of *TCF12* have been identified in coronal synostosis cases (Goos et al., 2016; Le Tanno et al., 2014). Microdeletions that map to 7p21.1 lead haploinsufficiency of *TWIST1* and cause Saethre-Chotzen syndrome (Johnson et al., 1998). By contrast, a 1.93 kb deletion removing exon IIIc and substantial portion of the flanking introns, and a 1.37 kb deletion between exons IIIb and IIIc, reported in Apert syndrome cases are believed to act by gain-of-function through a switch in tissue-specific alternative splicing (Bochukova et al., 2009b; Fenwick et al., 2011). Furthermore, transportable element insertions (*Alu*-elements) upstream or within exon 9 of *FGFR2* have also been reported to cause Apert syndrome (Bochukova et al., 2009a; Oldridge et al., 1999).

By contrast, distant duplications or deletions of noncoding elements can also be pathogenic. For example, Dathe et al. (2009) identified a microduplication of approximately 5.5 kb at 110 kb downstream of *BMP2* in autosomal-dominant brachydactyly type A2 (BDA2), a limb malformation characterised by hypoplastic middle phalanges of the second and fifth fingers. The duplicated region contains evolutionary highly conserved sequences. By comparing expression profiles of X-Gal in the transgenic model and *Bmp2* expression, limb-specific enhancer of *Bmp2* was considered to locate within the identified duplication. Conditional ablation of *Bmp2* in the mouse limb did not cause limb phenotypes and patients with large duplication, encompassing *BMP2* and the regulatory sequence, were reported to have a *BDA2* phenotype in association with other malformations. Considering these, the duplication of the regulatory element was suggested to lead the misexpression of *Bmp2* in the developing limb, and cause the BDA2 phenotype. (Dathe et al., 2009).

Another example is provided by *SOX9*, located at 17q24 close to a locus for Pierre Robin sequence (PRS; a subgroup of cleft palate with the manifestation of micrognathia and glossoptosis) originally mapped to 17q24 by linkage (Benko et al., 2009). Further analysis identified a 160 kb clustering of translocation breakpoints associated with PRS on chr 17, between *KCNJ2* and *SOX9*. Array CGH of

PRS cases identified further deletions within ~2.0 Mb centromeric and telomeric of *SOX9*, covering highly conserved non-coding sequences. Another heterozygous point mutation was then identified in one of the conserved regions, which increased binding affinity for the transcription factor *MSX1*. Strong activity of this putative enhancer in mandible region was observed in a transgenic mouse carrying the mutation (Benko et al., 2009). Notably, deletions and duplications in either the upstream or downstream region of *SOX9* can also result in other clinical phenotypes; sex reversal and brachydactyly-anonychia (Kurth et al., 2009; Pop et al., 2004). These cases and several others are summarised in Table 1-4.

TABLE 1-4 MUTATIONS OF NON-CODING REGULATORY REGIONS IN HUMAN DISEASE

Phenotype	**Gene affected**	**Abberations**	**Distance from gene**	**References**
autosomal-dominant brachydactyly type A2	*BMP2*	5.5 kb duplication in non-coding sequence	110 kb downstream	(Dathe et al., 2009)
Pierre-Robin sequences	*SOX9*	Translocation breakpoint	1.06-1.23 Mb upstream	(Benko et al., 2009)
		Microdeletions	1.58 Mb upstream 1.56 Mb downstream	
		Point mutation	1.44 Mb upstream	
Male to female sex reversal	*SOX9*	1.5 Mb deletion	0.5 Mb downstream	(Pop et al., 2004)
Female to male sex reversal	*SOX9*	Duplication	0.5 Mb upstream	(Franke et al., 2016)
Brachydactyly anonychia	*SOX9*	Microduplications	~ 2 Mb upstream	(Kurth et al., 2009)
CMT1 neuropathy	*PMP22*	186 kb duplication	33 kb downstream	(Zhang et al., 2010)
Short stature	*SHOX*	47 kb deletion	160 kb upstream	(Benito-Sanz et al., 2012)
Deafness	*POU3F4*	8 kb deletion	900 kb downstream	(de Kok et al., 1996)

1.7.3 STRUCTURAL VARIANTS AND TOPOLOGICAL ASSOCIATED DOMAINS

Mammalian genomes are highly compacted to fit within nuclei by being folded in three dimensions, at the same time individual genomes must be accessible to the transcriptional machinery to allow appropriate expression in different cell types and at different times of development. The three dimensional nuclear organisation of DNA includes topological associated domains (TAD). These can be identified by using chromosome conformation capture (section 1.7.4). TAD can range from kilobases to megabases in size and are separated from each other by boundary regions (Sexton and Cavalli, 2015). The boundaries subdivide the genome into discrete regulatory units, and restrict and control the contacts between enhancers and their target genes. Recently, SV mediated disruption of TADs involving *SOX9* and *IHH* was shown to be pathogenic (Franke et al., 2016; Lupianez et al., 2015).

Franke et al. (2016) demonstrated a correlation between different phenotypes and duplications of the region adjacent to *SOX9* in association with TADs. Duplications near to *SOX9* were previously known to lead to radically different clinical phenotypes; skeletal disorders and sex reversal (Table 1-4). By inspection of the three-dimensional structure of the genome in Hi-C interaction profiles of H1 embryonic stem cells, two large TADs were identified in the region. One of them contains the *SOX9* gene and its regulatory elements, the other contains two genes, *KCNJ2* and *KCNJ16*, encoding potassium channels, and their regulatory elements (Figure 1-4 A). The analysis of TAD structures using chromosome conformation capture (capture Hi-C and 4C-seq methods) in transgenic mice engineered to contain genomic duplications mimicking those in patients, demonstrated a correlation between the patient phenotype and the genomic interactions identified. The duplications causing sex reversal, contained entirely within the mouse *Sox9* TAD, showed increased contact between regulatory elements and Sox9 within the TAD, but the TAD structure was not affected (Figure 1-4 B). The larger duplications that include this region but extend up to the neighbouring genes *KCNJ2* and *KCNJ16* have no clinical effects on sex determination, but cause Cooks syndrome (abnormality of the limbs characterised by the absence of fingernails and shortened fingers and toes) (Figure 1-4C). Capture Hi-C was performed on limb buds of wild type and mutant mice (at E12.5) and a new

chromatin domain (neo-TAD) was found to be formed in the mutant limb buds. In the neo-TAD, due to ectopic contacts between the next flanking gene *Kcnj2* and the *Sox9* regulatory region, misexpression of *Kcnj2* was induced in the limb buds, which led to a limb malformation phenotype. The duplications that cover parts of two TADs and their boundary result in the formation of a neo-TAD, that is however, insulated from its neighbours and generates no phenotype (Franke et al., 2016).

These studies illustrate that genomic regulatory elements are controlled and stabilised by TADs and that the disruption of TADs by SVs can cause aberrant regulatory effects at the wrong time and place during development, giving rise to abnormalities.

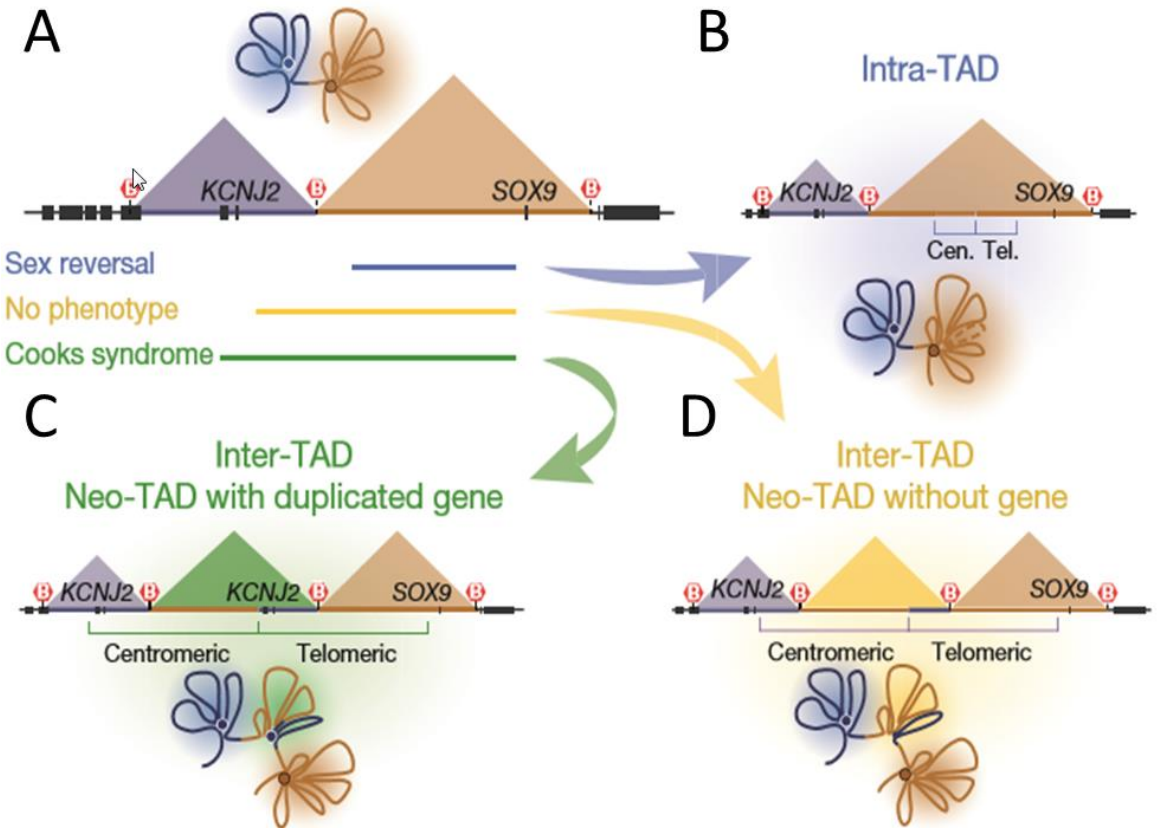


FIGURE 1-4 ILLUSTRATION OF DUPLICATION AND TADS

A Schematic image of TADs and the region of duplication in 3 different clinical phenotypes.
B Intra-TAD duplications (blue) do not change the TAD conformation but leads abnormal gene regulation and causes sex reversal.
C and D The tandem duplications cross TAD-boundaries; Inter-TAD (green and yellow). Both result in the formation of new chromatin domains (neo-TADs). In the case of C, the duplicated gene in neo-TAD, arise a novel regulatory landscape and can result in gene misexpression (Cooks syndrome).
The figure is from (Franke et al., 2016)

1.7.4 TECHNIQUES TO MAP GENOME INTERACTION

Various genomic technologies to detect potential genome interactions have been developed recently. Chromosome Conformation Capture (3C) enables the study of chromatin looping and genome architecture in three dimensions. The 3C technique quantifies the interactions between two specific fragments (one-vs-one interaction). Subsequently, 4C (Chromosome conformation capture-on-chip), Capture-C, 5C (Chromosome conformation capture carbon copy) and Hi-C were developed from the 3C. The 4C and Capture-C technology can investigate interactions between one locus and all other genomic loci. While 4C can only investigate a single region of interest at a time, Capture-C is able to analyze interactions between *cis*-regulatory elements at hundreds of selected loci at high resolution in a single assay by combining with high throughput sequencing (Hughes et al., 2014). Hi-C can detect all vs all genomic interaction, to gain a comprehensive map of genome-wide chromatin interactions with kilobase resolution, which can demonstrate overall TAD structure (reviewed in (Zhang and Lupski, 2015)).

1.8 RESEARCH AIMS

The aim of this DPhil was to define the molecular basis and pathogenesis of craniosynostosis in four families in which routine genetic investigations had proved uninformative. Focusing on families with multiple affected individuals increased the probability that the pathology indeed had a single genetic basis, with polygenetic or environmental factor (section 1.4.3) playing only minor roles. Of the four families studied, two showed autosomal dominant inheritance (Chapter 3 and 4), one showed likely autosomal recessive inheritance associated with consanguinity (Chapter 6) and one showed dominant inheritance associated with a new structural rearrangement in the grandparental germline (Chapter 5). Hence, different strategies for disease gene identification were required in the different families.

In the process, I also aimed to 1) Combine cutting-edge technologies and traditional genetic approaches to solve these difficult cases, 2) Undertake further functional analysis of significant candidate variants, and 3) Screen the craniosynostosis cohort panel to identify additional cases using NGS techniques.

Chapter 2 METHODS

2.1 ETHICAL APPROVAL

All families were recruited to the study using the ethically agreed protocol: *The genetic basis of craniofacial malformation,* chief investigator prof. Andrew Wilkie, REC reference number: 09/H0706/20 (assessed by Riverside Research Ethics Committee). The patients who have undergone WGS were recruited *to Molecular Genetic Analysis and Clinical Studies of Individuals and Families at Risk of Genetic Disease* (MGAC), REC reference number: 13/WM/0466.

2.2 REAGENTS, ENZYMES AND OLIGONUCLEOTIDES

All chemicals, unless stated otherwise, were obtained from Sigma-Aldrich. Restriction enzymes were from New England Biolabs, Promega or Roche. Oligonucleotides were ordered from either Sigma-Aldrich or Integrated DNA Technologies.

2.3 GENERAL LABORATORY METHODS

2.3.1 DNA EXTRACTION

DNA was extracted from whole blood samples collected into EDTA tubes, lymphoblastoid cell lines or fibroblast cell lines established from scalp skin obtained during surgery. DNA was extracted using the Nucleon Blood and Cell Culture (BACC) DNA extraction kit (Gen-Probe Inc.), according to the manufacturer's protocol.

2.3.2 RNA EXTRACTION

RNA from human samples was extracted from whole blood, Epstein–Barr virus (EBV) transformed lymphoblastoid cell lines or fibroblast cell lines. Blood samples were collected into PAXgene Blood RNA tubes and the PAXgene Blood RNA Kit (PreAbalytiX) was used to extract the RNA. RNA was extracted from lymphoblastoid or fibroblast cell lines using Trizol reagent (Invitrogen) and the RNeasy kit (Qiagen).

2.3.3 cDNA SYNTHESIS

cDNA synthesis was performed with RevertAid First Strand Synthesis kit (Thermo Fisher Scientific) using either oligo (dT) primers or random hexamer primers according to the manufacturer's instructions.

2.3.4 POLYMERASE CHAIN REACTION (PCR)

Unless otherwise stated, reactions were performed using FastStart Taq DNA Polymerase (Roche). The reaction was carried out using 20 ng of genomic DNA in a total volume of 20 µl, containing 2 µl of PCR reaction buffer (10x) with 20 mM $MgCl_2$, 1 µl of Nucleotide Mix (25 µM each dNTP), 1 µl each of 0.2 µM forward and reverse primers, 0.15 ul of FastStart Taq DNA Polymerase (5U/µL) and H_2O. Cycling conditions consisted of an initial step of 5 min denaturation at 94˚C, followed by 35 cycles of 94˚C for 30 s, annealing at 63˚C for 30 s and extension at 72˚C for 30 s, and a final extension step at 72˚C for 7 min.

Primers used for PCR are listed in Appendix 1. Primers were designed using Primer3 web version 4.0.0.0 (http://primer3.ut.ee/) and confirmed that the primers do not design over SNPs and those are specific by using UCSC In-Silico PRC program.

2.3.5 VISUALISATION OF PCR PRODUCTS

PCR products were visualised by electrophoresis through a 1-3 % Agarose gel which was made with 1-3% of UltraPure™ Agarose (Invitrogen) in 1x TBE buffer (8.9 mM Tris-base, 8.9 mM boric acid, 2.5 mM EDTA). Ethidium bromide (1 µg/ml) was added to the gels to allow visualisation of DNA by transillumination with UV light. Samples were loaded in loading buffer (6x: 0.25% bromophenol blue, 30% w/v sucrose) and run alongside the appropriate ladder (100 bp or 1 kb; NEB).

2.3.6 DIDEOXY SEQUENCING

Firstly, amplified PCR products were treated with *ExoI* (NEB) and Shrimp Alkaline Phosphatase (FastAP; Thermo Scientific) to remove PCR primers and dNTPs, respectively, by incubating at 37˚C for 30 mins, followed by 85˚C for 15 mins to inactivate the enzymes. Subsequently, sequencing reactions were carried out using the BigDye Terminator v3.1 cycle sequencer kit (Applied Biosystems). The DNA was precipitated with ethanol and sequenced on an ABI-3730 DNA analyser by John Frankland and Tim Rostron at the DNA Sequencing Facility in the Weatherall Institute of Molecular Medicine, John Radcliffe Hospital. Chromatograms were viewed in Sequencher 5.1.

2.3.7 DNA DIGESTION

Digest reactions were carried out in a total volume of 20 µl, containing 8 µl PCR product, 10 U of enzyme, 1 x buffer solution and 10 % of 1 x bovine serum albumin (BSA) if necessary. Incubation was performed at the recommended temperature for 2 hours to overnight.

2.4 WHOLE EXOME SEQUENCING AND WHOLE GENOME SEQUENCING

DNA library preparation for whole exome sequencing was performed using the exome capture kits specified in Table 2-1 according to the manufacturer's protocols. Massively parallel sequencing was undertaken at the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics), using the HiSeq 2000 platform (Illumina).

DNA samples for whole genome sequencing (WGS) with an OD 260/280 ratio of between 1.8 and 2 were diluted to a concentration of 50-100 ng/µl (measured by a Qubit fluorometer) and a minimum of 3-5 µg of DNA were provided for library preparation and sequencing to the Oxford Medical Genetics Laboratories (The Churchill Hospital, Oxford) or Complete Genomics. WGS performed by Complete Genomics was analysed using their proprietary software tools.

The software tools used for mapping and variant calling are shown in Table 2-1.

TABLE 2-1 CAPTURE METHODS AND ANALYSIS DETAILS FOR EACH WES AND WGS EXPERIMENT

				GRCh37 (hg 19)		GRCh38 (hg 38)	
	Case	**Gender**	**Exome Capture**	**Alignment**	**Variant Calling**	**Alignment**	**Variant Calling**
Chapter 3 Family A	II-1 (SM4796)	male	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
	III-1 (AM2587)	male	SureSelect V5	Novoalign	SAMtools v0.1.16	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
	III-2 (CM2959)	male	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
	III-4 (SM4347)	male	(Whole Genome)			Complete Genomics	Complete Genomics
Chapter 4 Family B	III-4 (TE5853)	female	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
	IV-3 (CC5360)	female	TruSeq	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
			(Whole Genome)	-	-	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
	IV-5 (MS6523)	male	(Whole Genome)	-	-	Complete Genomics	Complete Genomics
	IV-6 (JS5885)	male	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
	V-1 (FC5808)	male	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
					Platypus v0.5.2		Platypus v0.5.2
			(Whole Genome)	-	-	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
Chapter 5 Family C	I-1 (SB5750)	male	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
	II-1 (NB5827)	female	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
	II-2 (LM5471)	female	SureSelect V5	Novoalign	SAMtools v0.1.19	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
	III-2 (LM5380)	female	TruSeq	Novoalign	SAMtools v0.1.16	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
Chapter 6 Family D	III-1 (AH6740)	female	(Whole Genome)	-	-	Bowtie2	SAMtools v1.1
							Platypus v0.5.2
	III-4 (MH4737)	male	TruSeq	Novoalign	SAMtools v0.1.16	Bowtie2	SAMtools v1.1
							Platypus v0.5.2

Mapping was initially to GRCh37 genome build, but subsequently from 2015, to GRCh38.

2.5 METHODS USED IN CHAPTER3

2.5.1 EXPRESSION OF *HIST1H2AD*

RNA was extracted from fibroblast cells of III-2 and control and *HIST1H2AD* was amplified in synthesised cDNA using the following primers.

TABLE 2-2 PRIMERS TO AMPLIFY *HIST1H2AD* CDNA

Gene	Chr	Position	Change	Primer sequence 5'→3'		Size (bp)	Detected by
				Forward	Reverse		
HIST1H2AD	chr6	26,199,204	A>G	GAGCAGTACAGCCTGGATGTTG	GCTCCGCAAGGGCAACTACT	250	Seq

2.5.2 RESEQUENCING OF H2A GENES WITH MOLECULAR INVERSION PROBES (MIPS)

The sixteen canonical H2A genes were resequenced using MIPs (O'Roak et al., 2012a). Probes consisted of a universal 35 nucleotide motif (5'-CTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNN-3'), including a series of 5 N nucleotides for the identification of unique molecules, flanked by unique nucleotide sequences specific to the target region ('ligation' and 'extension' arms, in total 40 bp). Phosphorylation and capture of MIP pools were performed as described previously (Hiatt et al., 2013), with modifications. Individual MIPs were combined in an equimolar ratio and 1.6 µl of this pool (0.1 µl aliquot of each oligonucleotide) was phosphorylated with 2 U of T4 Polynucleotide Kinase (NEB) in 1x T4 DNA Ligase Buffer to a final volume of 20 µl, with incubation at 37˚C for 45 min. Inactivation of the T4 kinase was performed by incubation at 65˚C for 20 min. Capture efficiency was determined using test captures of 100 ng genomic DNA and the equimolar MIP pool. A re-balanced MIP pool was determined from sequencing results on the equimolar pool and adjusted accordingly, to improve target uniformity. DNA capture, gap-filling and ligation of circularised captured regions were performed in a single reaction. A total of 100 ng patient sample DNA was incubated at 95˚C for 10 min, then 60 °C for 22-24 hrs with 0.2 fM of the re-balanced MIP pool, 8 µM dNTPs (Bioline), 3.2 U Hemo KlenTaq® (NEB), 1 U Ampligase® DNA Ligase (Epicentre) and 1x Ampligase Buffer to a final volume of 25 µl. Uncaptured genomic DNA and uncircularised MIP probes were digested with the addition of 0.25 U Exonuclease I (NEB), 1.25 U Exonuclease III (NEB)

and 1x Ampligase buffer to each capture reaction and incubated at 37˚C for 45 min, followed by the incubation at 95˚C for 2 min to inactivate the exonuclease enzymes. Control DNA reactions were performed as described above and 5 µl of each sample was amplified in 25 µl reactions containing 1X iProof HF Master Mix (Bio-Rad Laboratories), 0.1x SYBR Green I Nucleic Acid Gel Stain (ThermoFisher), 0.5 µM Forward Primer (5'-AATGATACGGCGACCACCGAGATCTACACATACGAGATCCGTAATCGGGAAGCTGAAG-3'), 0.5 µM Reverse Primer (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGTCGATATCCGACGGTAGTGT-3') to determine the optimal number of amplification cycles. Reactions were amplified using the LightCycler® 480 Real-Time PCR System (Roche), with an initial denaturation step of 98°C for 30 s prior to 35 cycles of 98˚C for 10 s, 60 °C for 30 s and 72˚C for 30 s and a final extension of 72°C for 2 min. For each patient DNA capture reaction, samples were processed as above for 26 cycles, without SYBR Green I Nucleic Acid Gel Stain and with an Ion PGM™ sequence-specific forward primer (5'-CCTCTCTATGGGCAGTCGGTGATATCGGGAAGCTGAAG-3') and an Ion PGM™ sequence-specific, indexed reverse primer (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-[BC]-CGATATCCGACGGTAGTGT-3'). Captured reactions were pooled and purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research) and AMPure XP beads (Beckman Coulter Ltd.) to remove adapter dimers. Purified libraries were quantified using the 2200 TapeStation (Agilent Technologies) with High Sensitivity D1000 ScreenTape. Pooled and indexed PCR products were sequenced using the Ion PGM™ System (Life Technologies Ltd.). Pooled libraries were diluted to a final concentration of 10 pM and emulsion PCR and enrichment were performed with the Ion PGM™ Template OT2 200 Kit (Life Technologies Ltd.) according to the manufacturer's instructions. Data were processed with Ion Torrent platform-specific pipeline software v4.2.1, including coverage analysis and variant calling.

The library sequenced on the MiSeq was made in the same manner as above, but each patient DNA capture reaction was processed with a forward primer (5'-

AATGATACGGCGACCACCGAGATCTACACATACGAGATCCGTAATCGGGAAGCTGAAG-3'), and reverse Primer (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGTCGATATCCGACGGTAGTGT-3'). Variant calls and coverage information for samples were obtained using a custom pipeline by Dr. Nils Koelling (Clinical Genetics Group, WIMM): Reads were trimmed to remove primers and low-quality basecalls and aligned to GRCh37 using *bwa mem* version 0.7.12 (Li and Durbin, 2009). Coverage was calculated using *bedtools* version 2.25.0 (Quinlan and Hall, 2010), requiring a minimum coverage of 10 to consider a target region fully covered. Variants were called with *Platypus* version 0.8.1(Rimmer et al., 2014) and annotated using *Annovar* version 2015-06-17 (Wang et al., 2010). Further processing and annotation were performed using custom scripts written in Python 3.5.3 with *pysam*, *biopython* (Cock et al., 2009) and *pandas* (McKinney, 2010). The entire pipeline was built using *Snakemake* version 3.11.2 (Koster and Rahmann, 2012).

TABLE 2-3 THE DESIGN OF MOLECULAR INVERSION PROBES (MIPS)

Target	Ligation arm sequence	Extension arm sequence	Target length	Target Start	End
HIST1H2AA	GCAGGTGGCGGGGAATAATGCGAGTTTTT	A.CAATGACAACCTTA	144	25726342	25726531
HIST1H2AB	GCACCGCCGCGAGATACA	CACACGCCCCAAGAGTTTATTAAGC	111	26033494	26033648
HIST1H2AC	ACTCTAACACCGCCGCCAGGTAC	CCCGGCCTAGCAGTTTG	111	26124632	26124745
HIST1H2AD	GCTGCA.GTGTCGGGGGATGATGCGGGT	AAAAGCCAATATAAGA	145	26199054	26199243
HIST1H2AE	GCACCGCTGCCAGGTACACT	GACCTAGAAGCTTATTTAGC	111	26217368	26217481
HIST1H2AG	GCGGATGGCCAGTTGGAGATG	ACAGATAGTTACTTGCCCTTCGC	110	27101117	27101229
HIST1H2AH	GCGGATGGCCAGTTGGAGGTGA	ACAACCTCGCTCCTTATTTGGC	110	27115174	27115286
HIST1H2AI	CGAGTCTTCTTGTTGTCGCGGGCCGCGTT	TTCTTGGGCAGTAGCA	109	27776220	27776331
HIST1H2AJ	ATGCGAGTCTTCTTGTTGTCGC	CTCAGTTTTCTTTGGCAG	111	27782153	27782304
HIST1H2AK	GCGGTTAGGTACTCCAACACCGCCGCCAG	ATGGTCACTTTACCAA	109	27805810	27805964
HIST1H2AL	GCTGCAAGTGGCGCGGGATAAT	TTGCCTTTGGCCTTGTGG	114	27833388	27833504
HIST1H2AM	GTCTTTTTGTTGTCGCGGGCTG	AGTCTTCTTGGGGAGCAG	114	27860565	27860719
HIST2H2AA	GACCGCAGCCATGTAGACG	CTTTGCCCAGCAGCTTGTTCAGTTCC	109	149814010	149814119
HIST2H2AC	GTGACGAGGGATGATGCGCGTCTTCTTGT	TTGGCTTTGTGGCTTT	114	149858775	149858889
HIST2H2AB	GCTAGTTGCAGATGGCGAGGGATGA	CTTGTTCTTGCCAGGC	113	149859118	149859231
HIST3H2A	GATCTCGGCAGTCAAGTACTCGAG	CCCTGCGCGATGGTCA	111	228645242	228645353

2.5.3 MOUSE MODEL GENERATION

2.5.3.1 CRISPR/CAS9 DESIGN TO INTRODUCE C.268A>G; P.N90D INTO MOUSE *HIST1H2AD*

The chr13:23,574,738 [c.268 A>G (p.N90D)] *Hist1h2ad* A>G variant was introduced into the mouse genome using CRISPR/Cas9 gene editing. To avoid off-target effects particularly in other H2A genes, using 2 single-guide RNAs (sgRNAs), along with a donor template for homology-directed repair, was planned. Upstream sgRNAs were designed to the 90 bp 5′ untranslated region (UTR) of *Hist1h2ad* so as not to disrupt *Hist1h2bf* which is on the reverse strand and locates 190 bp upstream of the *Hist1h2ad* 5′ UTR. Downstream sgRNAs were designed to the intergenic region between *Hist1h2ad* and *Hist1h3d.* SgRNA were designed using CRISPRscreen (http://apps.molbiol.ox.ac.uk/CRISPR/) and E-CRISP Design (http://www.e-crisp.org/E-CRISP/), and selected so as not to create off-target effects. (Please refer to section 2.5.3 for further details.) Three upstream and 3 downstream sgRNAs were designed (Table 2-4). The construct of CRISPR/Cas9 is shown in Figure 2-1.

SgRNAs were designed to have the configuration of forward oligo: 5′-CACCG(20bp target sequence)-3′and Reverse oligo: 5′-AAAC(20 bp target sequence)C-3′. The forward and reverse strands pair with each other and have overhangs (5′-CACC and 5′-AAAC) to compatible with ligation to BbsI digested pX458 vector.

TABLE 2-4 SGRNAS DESIGNED FOR CRISPR/CAS9 GENOME EDITING

Position in the genome	Name	Primer sequence 5′ →3′		strand
		Forward	Reverse	
Upstream sgRNA	**sgRNA3**	**AAACttttcgcgcccagcgaataC**	**CACCGtattcgctgggcgcgaaaa**	**rev**
	sgRNA4	AAACtttatattattcgctgggcgc	CACCgcgcccagcgaataatatataaa	
	sgRNA7	AAACagcgaataatataaaggctC	CACCGagcctttatattattcgct	rev
Downstream sgRNA	sgRNA5	AAACatgaaatattaaccgtgatC	CACCGatcacggttaatatttcat	rev
	sgRNA6	**AAACggctatagacaattgttgaC**	**CACCGtcaacaattgtctatagcc**	
	sgRNA2	AAACggcgcggttcgttttggggC	CACCGccccaaaacgaaccgcgcc	rev

Primer sequences; Upper cases indicate the overhang sequences. Lower cases indicate the target sequences.

2.5.3.2 SGRNA CLONING INTO PX458

Equal amounts of 50 µM forward and reverse sgRNA primers were annealed at 96°C for 5 min, followed by ramping down to 25°C at 0.1 °C per second with a final step at 25°C for 1 min. Annealed

sgRNAs were diluted to 10 µM and a ligation reaction was performed with 1 µl sgRNA, 1 µg pSpCas9(BB)-2A-GFP (PX458), 10 units BbsI, 1.5 µl T4 DNA ligase and 2 µl Fermentas buffer (x10) in a total volume of 15 µl, incubating at 37˚C for 1 hr. Then, 5 µl of ligation mix was transformed into NEB 5-alpha competent *E. coli* according to the manufacturer's protocol. Plasmid DNA was purified from cultured *E. coli* cells using QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's protocol. The cloning was confirmed by PCR (standard PCR as described in 2.3.4) and dideoxy-sequencing.

The sgRNAs were transfected into B16f10 Black 6 cell lines and the targeting efficiency was verified by Dr. Philip Hublitz (the Genome Engineering Service, WIMM) by using PCR and surveyor assays. SgRNA3 and sgRNA6 were found to be the most efficient (data not shown) and were used for the following experiments.

gtatgaccgtgggcccgagactgctcttgaataaaaatcctcttgcaatttgcagcaagaccgtttctcgttgattttggggtgtcgcctctcctga

gtcagaaggtggtggagtcctcacgttgtgggtctttcattacgccaatgaatttacaggcacgtctttttttttttctacacagaacataaaagtaa

caactgttactaagtatcaacatccaatgttacaatttttcttggaggattaaacagtattcggtgtctggcaacagctccctttttgaaagggtag

5′ homology arm (1009 bp)

gtggctctgaaaagagcctttgggttaggagtgtgtgttaaaggaac**tcacttggagctggtgtacttggtgacagccttggtgccctccgacac**

3′ UTR

ggcgtgcttggccagctccccgggcagcagcaggcgcacggccgtctggatctcccgggacgtgatggtcgagcgcttgttgtaatgcgccagg

***Hist1h2bf* gene**

cgggaagcctcgctcgcgatgcgctcgaagatgtcgttcacaaacgagttcatgatgcccatggccttggaggagatgccggtgtcggggtgc

acttgcttcagcaccttgtacacgtacaccgagtagctctccttgcggctgcgcttgcgcttcttgccgtccttcttctgggccttggtcacggcctt

cttggagcccttcttcggggcgggagcggacttagcaggctcaggcatggttggtgagaatttctggaactttcacggaggcaactggtcttgtct

agttgatgaggtcttattcaaatgagggatgtgaaagggcacctttgattggtttatgacggatggcatacgaagtcttgcccaatcaacgtgcgc

G>C to inactivate PAM
sgRNA4 target sequence

Cccgaatgcttctcattggtctaggaaaaaagacgtcagccaatggcataggtcctttttcgcgcccagcgaataatataaaggctacgttcttgt

sgRNA3 target sequence
C>G to inactivate PAM

sgRNA7 target sequence
CC>GG to inactivate PAM

ttcaaaaccattagttcttagtaactgcttaatctgtaca**atgtctggacgcggcaaacagggtggcaaggctcgcgccaaggccaagacccgc**

5′ UTR

tcctcccgggccggcctgcagttccccgtgggccgcgtgcaccggctgctccgcaagggcaactactcggagcgcgtgggcgccggcgccccgg

Hist1h2ad

tgtacctggcggccgtgctggagtacctgacggccgagatcctggagctggcgggcaacgcggcccgcgacaacaagaagacgcgcatcatc

Amplicon2 (Forward primer)

ccgcgccacctgcagctggccatccgcaacgacgaggagctcaacaagctgctgggccgcgtgaccatcgcgcagggcggcgtcctgcccaa

chr13:23,574,738 c.268 A>G (p.N90D)

catccaggccgtgctgctgcccaagaagaccgagagccaccacaaggccaaggggaaataagattggtcttctgaacttgtataaaaacaaa

3′ UTR

ggctcttttcagagccatccacttcgtcatttaaaagagccatttgctgcattttccaatgaaatattaaccgtgatgtaatcactgtcatttgtctta

sgRNA5 target sequence
C>G to inactivate PAM

FIGURE 2-1 THE DESIGN OF CRISPR/CAS9 GENOME EDITING TARGETING *HIST1H2AD*

The sequence shown corresponds to the donor cassette used in the CRISPR/Cas9 targeting. 5' and 3' homology arms were underlined in purple. The positions of histone genes are shown in different colours; *Hist1h2bf*, green. *Hist1h2ad*, blue. *Hist1h3d*, red and open reading frame sequences are in bold. The residue in green in *Hist1h2ad* is chr13:23,574,738 c.268A (which is mutated to G in donor cassette). The positions of sgRNA target sequences and primers used for verifying the substitution are underlined in yellow and right blue, respectively. In the sgRNA sequence, the yellow rectangles indicate the protospacer adjacent motif (PAM) sequence and the residues in blue were mutated to inactivate the PAM site in donor cassette as shown in the figure. The triangles show the new SacII restriction site in donor cassette which was created when inactivating the PAM.

2.5.3.3 DESIGN OF DONOR TEMPLATE

As shown in Figure 2-1, the 5' and 3' homology arms were 1009 bp and 865 bp, respectively. CRISPR nucleases recognise short sequences called the protospacer adjacent motif (PAM) as a cutting site, thus, the second G in PAM sites was mutated to C to prevent sgRNA-directed cutting of the donor template. In total eight point mutations were created in the donor template; 7 bases to inactivate PAM sequences and a base to insert c.268 A>G. A SacII restriction site was generated by inactivation of the PAM sequence of sgRNA6.

The donor template was 2979 bp in length and synthesised by Thermo Fisher Scientific.

2.5.4 MOUSE ES CELL CULTURE

Mouse E14 embryonic stem (ES) cells, E14Tg2a ES cells (129/Ola), were cultured in 2% gelatine (Sigma)-coated wells or flasks in complete media (500 ml BHK-21 Glasgow medium (Gibco), supplemented with 60 ml foetal bovine serum (FBS) (Gibco) to give a final concentration of 10 %, 30,000 U each of streptomycin and penicillin, 1.2 mM glutamine (Gibco), 1 mM sodium pyruvate (Gibco), 1 mM non-essential amino acids (Gibco), 0.1 mM mercaptoethanol (Sigma), Leukemia inhibitory factor 120 µl (10^7 units/ml) (Gibco) in a 5 % CO_2 atmosphere at 37˚C. Collection of cells for DNA and RNA was carried out by adding 0.25 % of trypsin-EDTA (Gibco) to strip adhered cells followed by quenching in media and centrifugation at 1000 x g for 5 mins.

2.5.5 TRANSFECTION

Transfections were carried out in using mouse E14Tg2a ES cells (129/Ola). Cells were split into 3 wells of gelatine coated 6 well plate at $3x10^5$ cells per well on the day prior to transfection. The cells were co-transfected with 1.65 µg of each px458-sgRNA3 and px458-sgRNA6, along with 3.3 µg of the donor template plasmid using Promega FuGENE 6 transfection reagent. After 48 hrs incubation, transfected cells were collected into 15 ml. Subsequently, GFP-positive cells were sorted using Fluorescent Activated Cell Sorting (FACS) by Dr. Kevin Clark at the WIMM Flow Cytometry Facility. Sorted cells were plated in 3 wells of a 6 well plate and incubated for 3 days. After 3 days, cells were collected again into a 15 ml tube from which individual live cells were individually sorted into 96 well

plates. In total six 96 well plates were prepared and incubated for a week. Cells were split into 2 of 96 well plates; one was used for genotyping and the other was frozen down in medium containing 90% FBS and 10% Dimethyl sulfoxide (DMSO) at -80 °C for later use.

2.5.6 DNA EXTRACTION FROM 96 WELL PLATES

Wells with growing clones were labelled before DNA extraction. After aspirating medium from the plates, cells were incubated with 100 µl lysis buffer (Tris 100 mM pH 8.5, EDTA 5 mM, NaCl 200 mM) containing 100 µg/ml proteinase K at 37 ˚C overnight. The lysed solution was transferred into V bottom 96 well plates and a 2 times volume of ice cold isopropanol was added to each well to precipitate the DNA. The plate was centrifuged at 3000 rpm for 15 mins at 4 ˚C and the supernatant was removed by carefully turning the plate upside down. Fifty µl of 70 % ethanol was added to the plate to wash the DNA with centrifugation at 3000 rpm for 10 min. Ethanol was carefully tipped off and the DNA was left to dry at room temperature before addition of 30 µl of Tris-EDTA (TE) into each well. Plates were stored at -20 °C.

2.5.7 IDENTIFICATION OF TARGETED MOUSE ES CELLS

PCR amplification was performed using KAPA Taq HotStart PCR Kit (KAPABIOSYSTEMS). The reaction was carried out using 40 ng of genomic DNA in a total volume of 25 µl, containing 5 µl 5xKAPA Taq HotStart Buffer, 1.5 mM $MgCl_2$, 200 µM each dNTP, 0.4 µM primers and 0.5 units Polymerase. Cycling conditions consisted of an initial step of 3 min denaturation at 95˚C, followed by 35 cycles of 98˚C for 20 s, annealing at 63˚C for 15 s and extension at 72°C for 60 s/kb, and ended with a final extension step at 72˚C for 1 min/kb.

Mutant clones were screened by PCR using Amplicon1 primers (Table 2-5) followed by digestion with SacII. DNA from digestion-positive clones was PCR amplified and sequenced to confirm the presence of c.268A>G using Amplicon 2 primers (Table 2-5), where the forward primer is upstream of c.268A>G.

TABLE 2-5 THE PRIMERS USED TO CONFIRM THE PRESENCE OF C.268A>G IN *HIST1H2AD*

Primers	Forward	Reverse	Size (bp)
Amplicon 1 - for SacII restriction site	ACGCGTGTTTAAAATTACTCCTTTACT	CCAGCAACACTAAAAAC TGTGGA	1042
Amplicon 2 - for detecting the mutation by dideoxy-sequencing	GAGTACCTGACGGCCGAGAT		1784

Mutant clones were revived and passed to the WIMM transgenic core after genotype verification. Pluripotency and karyotype testing was undertaken before microinjection into C57BL/6 blastocysts.

2.6 METHODS USED IN CHAPTER 4

2.6.1 MENDELSCAN

Rare heterozygote rule out (RHRO) and shared identity-by-descent (SIBD) were applied to our WES data of families, described in chapter 3 and 4, using MendelScan (http://gmt.genome.wustl.edu/packages/mendelscan/index.html)(Koboldt et al., 2014).

Pedigree file in PED format that indicated the name, gender, affected the status of the samples in the variant calling file (VCF) were generated. MendelScan RHRO was run with the pedigree file, family VCFs from WES (annotated with dbSNP information) and a BED file of chromosome centromere coordinates. Two output files were generated from the command: 1. all informative variants (rare heterozygotes shared by affected or variant positions with homozygous differences between affected pairs) 2. a window of RHRO regions that are consistent with autosomal dominant inheritance given the inputs.

To run the SIBD program, family VCF files were converted to BEAGLE format by running BEAGLE FastIBD (v.3.3.12) (Browning and Browning, 2007) on a per-chromosome basis. MendelScan SIBD command was run with the pedigree file, the BEAGLE makers file and the BEAGLE output files (*.fibd). In the output file, the genome was segmented into nonoverlapping windows of 100 kb. For each window, the total number of unique affected pairs with a segment in common was computed. The results from both RHRO and SIBD were exported into Excel and plotted by chromosome.

2.7 METHODS USED IN CHAPTER 5

2.7.1 BREAKPOINT PCR AMPLIFICATION

Amplification was carried out using either 350 µM each dNTP, 300 nM primers, Expand Long Template buffer 1 and Expand Long Template Enzyme Mix 0.75 µl (Roche) for < 9 kb sized products or 500 µM each dNTPs, 300 nM primers, Expand Long Template buffer 3 and Expand Long Template Enzyme Mix 0.75 µl (Roche) for > 9 kb products, in a total volume of 50 µl. Cycling conditions consisted of an initial step of 2 min denaturation at 95˚C, followed by 10 cycles of 95˚C for 10 s, annealing at 58˚C for 30 s and extension at 68˚C for 20 s, 25 cycles of 95˚C for 10 s, annealing at 63˚C for 30 s and extension at 68˚C for 20 s and a final extension step at 68˚C for 7 min.

TABLE 2-6 PRIMERS FOR BREAKPOINT PCR AMPLIFICATION

Primer sequence 5'→3'			
Forward primers		**Reverse primers**	
F1	CTATAGGAGCTTGATGGGGATAGCATTG	R1	AAGAAAGGGTTTGATCTGCTAGTTACC
F2	CCAAAGGGGATCATGTCAGTAGAGTAT	R2	GATAGTCCTGAAGTTACAGTGAGACCAG
F3	ATACTCTTTCCCTCTCCCTCTCCCTATC	R3	AGTCTTTGTCTCAGGCTCTACTTTCTG
F4	GGTATCTCTTCATGTCTTTGCCCACTTC	R4	AGAATAGAGGGACCAGCAGCTTGTAT
F5	GAAAGTAAAGTCTTCTCCCTCACAGTCC	R5	CTAAGAACACTGGTCAATTGGGTTTC
F6	GTACATTTAGGTGTACCTTTCCCCTCAT	R6	CCTAGGCTCAGAAAGACCTCATGTTAG
F7	CAGCCAGGTGACAACTCTACTAGAATCA	R7	GTGTGTCCCAGCTAAAGTTATGGAAAG
		R8	AAGTATTATCTGCAGTGTCAGGCATCC
		R9	GGGCAAGCTTAAGGTAATTTAGGGGAC

2.7.2 COMPARISON OF ALLELIC BALANCE ON CHR4

To compare the allele balance of the genes within and near the duplication, 3 informative SNPs in *FGF5, C4ORF22* and *BMP3* were identified through the WES of individual III-2 (Table 2-7). Control samples which carry those informative SNPs were searched in our DNA stock by using the primers described in Table 2-8. Individuals JA4086 (*FGF5*, fibroblast), CL4171 (*BMP3*, fibroblast), PC4685 (*BMP3,*lymphoblastoid), JB5229 (*FGF5* and *PRKG2,* fibroblast), JM4101 (*PRKG2*,fibroblast) and CC4153 (*PRKG2*, fibroblast) were identified to have the appropriate SNPs, therefore, their samples (genomic DNA and RNA from indicated materials) were used to compare the allele balance with the affected individual III-2 in the following experiments.

DNA was extracted from blood and RNA extracted from skin fibroblast and/or lymphoblastoid cell lines. cDNA was synthesized as described in 2.3.3. By using primers in Table 2-8 and Table 2-9, PCR was undertaken as described in 2.3.4. Allelic ratios of *FGF5*, *BMP3* and *PRKG2* in both in DNA and cDNA were compared between individual III-2 and control samples by using the informative SNPs.

TABLE 2-7 THE INFORMATIVE SNPS

Gene	SNP	Position	ref	variant	Restriction enzyme
FGF5	rs3733336	chr4:81207963	A	G	*TspRI*
BMP3	rs3733549	chr4:81045996	A	G	*TaqI*
PRKG2	rs17005080	chr4:81171716	A	G	*MfeI*

TABLE 2-8 PRIMERS FOR GENOTYPING AND ALLELE DOSAGE ANALYSIS OF GENOMIC DNA

Gene	SNP	Primer sequence 5'→3'		Size (bp)
		Forward	Reverse	
FGF5	rs3733336	CCCTCAGGAGTTTCTATAGGTGTCTTC	TATCCCTCAAGCATAAAGCAAAAGG	215
BMP3	rs3733549	AGGACTGTATATCTTCAATCTG	GTATTCTGCCCCAGGAAGCTCGTT	632
PRKG2	rs17005080	AAAAGCCATTTTACATTTTATACACAGG	AAGTTTCTCAAGGCAACCCAGCTAA	546

Annealing temperature: 65˚C

TABLE 2-9 PRIMERS FOR ALLELE DOSAGE ANALYSIS OF CDNA

Gene	SNP	Primer sequence 5'→3'		Size (bp)
		Forward	Reverse	
FGF5	rs3733336	GTGTCTCAGGGGATTGTAGGAAT	GTCAGAAAAATACGTAGTCCCTG	745
BMP3	rs3733549	GAAACATCAGCCTGAGTTGTCCAGT	ACCAAGATATAAGGCTCTGGAAAAGGT	410
PRKG2	rs17005080	GAGAAGAAGCTCATTACAGATGCCC	CTCCTTTGTCATAGTATTCCACTTCC	428

Annealing temperature: 63˚C

2.7.3 DEEP SEQUENCING TO COMPARE ALLELIC BALANCE

Ion PGM™ System (Life Technologies Ltd.) was used to perform deep sequencing of informative SNPs in *FGF5, BMP3* and *PRKG2*. To run bidirectional amplicon-tagged sequencing on the system, the Ion Torrent A1 adaptor (5'- CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and P1 adaptor (5'-CCTCTCTATGGGCAGTCGGTGAT-3') sequences were added to the 5'- end of all target-specific forward and reverse primers, respectively. Ten bp Illumina barcode sequences were also added to 5'- of the A1 adaptor to enable pooling the PCR product for sequencing. Amplification was undertaken using the FastStart Taq DNA Polymerase kit (Roche). Amplification products (roughly equal amounts as judged by EtBr staining in agarose gels) were combined and then purified with

AMPure XP beads (Beckman Coulter Ltd.) to remove adaptor dimers. Libraries were quantified using the 2200 TapeStation (Agilent Technologies) with a D1000 ScreenTape. Emulsion PCR and enrichment was performed with the Ion PGM Template OT2 200 Kit (Life Technologies) according to the manufacturer's instructions. Sequencing of enriched templates was carried out on the Ion Torrent PGM for 125 cycles using the Ion PGM Sequencing 200 kit v2 and Ion 314 or 316 chips. Data were processed with Ion Torrent platform-specific pipeline software v4.2.1. The number of each allele was counted using SAMtools v0.1.19.

TABLE 2-10 PRIMERS FOR DEEP SEQUENCING

Gene	SNP	Primer sequence 5'→3'		Size (bp)
		Forward	Reverse	
genomic DNA				
FGF5	rs3733336	ACAGACTCAAGTTTCGCTTTGGA	TCCCTCAAGCATAAAGCAAAAGG	281
BMP3	rs3733549	ATCAGCCTGAGTTGTCCAGTGTC	GGAAAAGGTAACCTCCTCTTTGG	284
PRKG2	rs17005080	AGGGTGCAAATGTGACTGGACTA	ACTTTGTGTGCTTTATGATTGTTTT	291
cDNA				
FGF5	rs3733336	GCACCTCGGAAAAATACCAACTC	CTTGCTGAACTGAAGGAGTGTGC	281
BMP3	rs3733549	ATCTTTCTGCATGGACCCTCAA	TTGATACCACACTTTCTGGCTCA	282
PRKG2	rs17005080	CTAGAGGTGTTCCAAGGGGAGAA	TCCACTTCCAAGCAGTCAATGAT	280

Annealing temperature: 63˚C
The Ion Torrent A1 adaptor (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and P1 adaptor (5'-CCTCTCTATGGGCAGTCGGTGAT-3') sequences were added to the 5' end

2.7.4 REAL-TIME QUANTITATIVE PCR

To compare the expression level of genes in the region of duplication, real-time qPCR was performed in the same manner as section 2.8.3.using the primers listed in Table 2-11. The results were analysed by the comparative Ct method (also known as the delta-delta Ct method) (Schmittgen and Livak, 2008)

TABLE 2-11 PRIMES FOR QUANTITATIVE REAL-TIME PCR

	Primer sequence 5'→3'		Size (bp)
	Forward	Reverse	
BMP3	CGAAGGCAACACGGTTC	AGCGATGTCAGATTGAAGATATA	88
FGF5	ATCTACCCGGATGGCAAAG	CTCCTCGTATTCCTACAATCCC	79
PRKG2	CAGTGTGGATTTCTGGTCACT	CGTTTGAGAATCAAATTGTAGGT	110
C4ORF22	GATTTCCTGGACTCGCAGAT	CTCTCCAGTCCCTCGGTAG	103

2.8 METHODS USED IN CHAPTER 6

2.8.1 PRIMERS TO VERIFY THE DELETION IN *DHRS3*

PCR was performed as described in section 2.3.4, but using 1 µl of each of 3 primers (a forward primer and 2 reverse primers; R1-outside/downstream of the deletion, R2-inside the deletion) in a reaction. The primers were designed so that 506 bp products were amplified from F-R1 primer pairs, while 413 bp products were amplified from F-R2 primer pairs.

TABLE 2-12 PRIMER SEQUENCES FOR SCREENING THE *DHRS3* DELETION

Primer sequence 5'→3'		**Size (bp)**
Forward	**Reverse**	
CTGTAGAGGGAACATCACCAGCG	TGAGTACCAATGGCCTCCTCACT (R1)	4432
	CCCCTCCCTAAACTCCTCTGTCA (R2)	413

Annealing temperature: 63˚C

2.8.2 PRIMERS FOR MULTIPLEX PCR AMPLIFICATION OF *DHRS3*

Primers were designed for targeted amplification of 1) all 6 exons and intron/exon boundaries of *DHRS3 and* 2) the deletion identified in the proband. Nine target regions were amplified (exon 1 was divided into 3 regions, exon 2-6 were covered by 1 amplicon each, with a final amplicon for the deletion). Universal CS1 (5'-ACACTGACGACATGGTTCTACA-3') and CS2 (5'-TACGGTAGCAGAGACTTGGTCT-3') adaptor sequences were added to the 5' end of all target-specific forward and reverse primers, respectively, for the barcoding PCR amplification. To enable multiplex PCR amplification, primers were combined and optimised to get a relatively equal intensity of bands from each product on a 3 % agarose gel. Mixture of primer pairs (A) exon 1-2 and exon 6 with 1:4 ratio, (B) exon 1-3, 2 and 3 with 1:1:1 ratio and (C) exon 1-1 and exon 4 and 5 with 8:1:1 ratio were optimal (Table 2-13). Using these 3 primer mixes, 3 multiplex PCRs were performed with 20 ng of patient DNAs using the KAPA Taq HotStart PCR kit (Kapa Biosystems) according to the manufacturer's protocol.

PCR products from each reaction were pooled and 1 µl of a 1 in 100 dilutions of the total PCR pool was used for the barcoding PCR with the Access Array Barcode Library for Illumina Sequencers – 384 (SingleDirection) (Fluidigm,PN100-4876). The barcoding PCR was performed using Q5 High Fidelity

DNA polymerase (QIAGEN) according to the manufacturer's protocol. The barcoded products were pooled and purified using AxyPrep MagPCR Clean-up kits (Appleton Woods) to produce libraries for sequencing on the MiSeq. Variant calling and coverage analysis were undertaken using the same pipeline as described in section 2.5.2.

TABLE 2-13 PRIMERS FOR MULTIPLEX PCR

	Primer sequence 5′→3′		Size (bp)	Multiplex pool
	Forward	Reverse		
Deletion	CAGCCGTTTCCACACCATC	GGCCGAGCAATACAGGAATTAAA	388	
Exon1-1	TCCTAAGGTTGGGACGGTAAA	CAATGAGTCGGGAGACTTTCG	367	C
Exon1-2	CAGTCCGACGGCTGCTTTCA	TAGAACTTTCCACCTCCGGCTTCC	363	A
Exon1-3	TGAAACCGACTAGCTCAGCACTC	CTGGTGATGTTCCCTCTACAGATGA	323	B
Exon2	AATCAACCCTAATCCCTGCCTATGT	GCATCACTGTGGCTTAATGACCATT	382	B
Exon3	TTCGTGGACATCCCTGCTGG	AAACATATATCGTTGGCCCTGGG	299	B
Exon4	CAGGTGAGAAGGCTGGTCT	CCCTAGAGCATCATGTGTGTC	389	C
Exon5	CAAATGTGCCAGCACTCCTC	CCAAAGGGTGACCCAGACAT	324	C
Exon6	CTCACCCAGCAGAAGCATG	GGAGCACTTTCAGGAACCAG	303	A

Annealing temperature 63˚C

2.8.3 QUANTITATIVE REAL-TIME PCR

Quantitative real-time PCR was performed using the KAPA SYBR FAST qPCR kits (KAPA BIOSYSTEMS) according to the manufacturer's instructions on a LightCycler480 II. The primers used to amplify the cDNA of *DHRS3,* and *HRPT* and *ACTB* as an endogenous control, are shown below. The results were analysed by the comparative Ct method (also known as the delta-delta Ct method) (Schmittgen and Livak, 2008).

TABLE 2-14 THE PRIMER SEQUENCES FOR REAL-TIME PCR

	Primer sequence 5′→3′		Size (bp)
	Forward	Reverse	
DHRS3	TGCACAGCTTCCACTGTC	GAGATGTTCCAGGGCATGAG	95
HPRT	GACCAGTCAACAGGGGACAT	GTCAAGGGCATATCCTACAAC	261
ACTB	CCATGTACGTTGCTATCCAGG	CTCCTTAATGTCACGCACGAT	251

2.9 GENOME BUILDS IN EACH CHAPTER

The information of the genome builds presented in this thesis is summarised in Table 2-15. The GRCh37 was used mainly in Chapters 3 to 5 and GRCh38 was used in Chapter 6 apart from the following exceptions.

TABLE 2-15 GENOME BUILDS USED IN CHAPTERS

Chapter	**Builds**	**Exceptions**	
Chapter 3	GRCh37		
Chapter 4	GRCh37	Table 4-10 Table 4-15	hg18 GRCh38
		Table 4-18	GRCh38
Chapter 5	GRCh37		
Chapter 6	GRCh38	Table 6-1 Table 6-5	GRCh37 GRCh37

Chapter 3 A VARIANT IN A CORE HISTONE PROTEIN CAUSING CRANIOSYNOSTOSIS?

3.1 INTRODUCTION

The family consists of 3 generations with 6 affected individuals with craniosynostosis. The phenotype of affected individuals is unique and with two distinctive apparently cosegregating phenotypes; metopic synostosis and severe psoriasis. Familial metopic synostosis is extremely rare and the combination with psoriasis is apparently unique. From the family, 5 affected and 3 unaffected individuals were recruited to the research. We undertook SNP chip segregation analysis, WES and WGS to identify a causal variant. A segregating nonsynonymous variant in *HIST1H2AD* was identified from affected individuals' WES. A link between histones and craniosynostosis is unknown, however, the variant identified is a unique change in a core histone. Therefore, the pathogenicity of the variant is further investigated in this chapter.

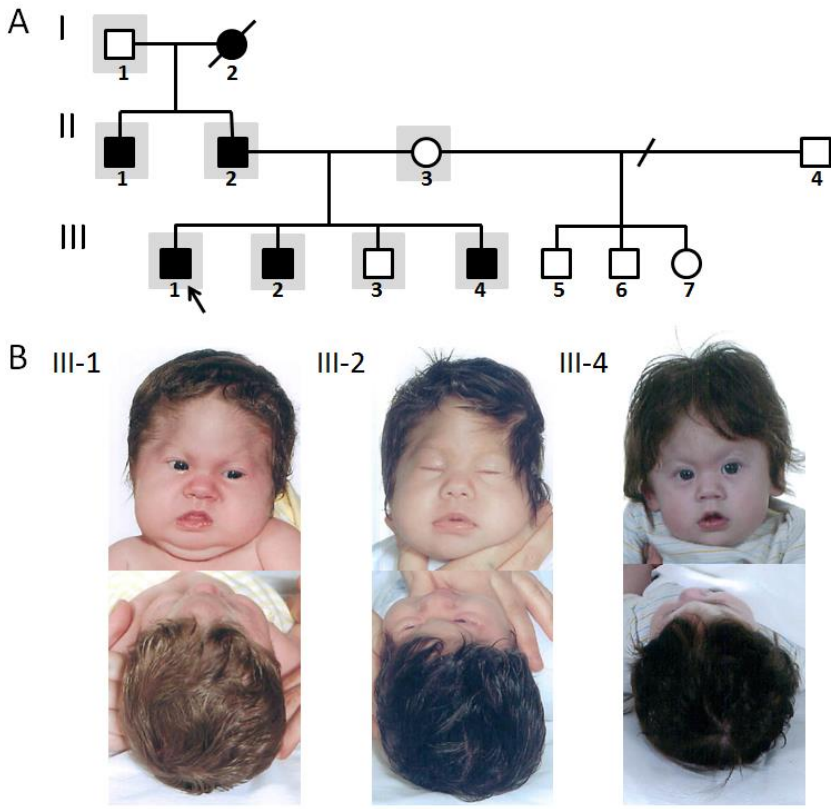


FIGURE 3-1 FAMILY TREE AND CLINICAL PHOTOS

A. *Squares*, Male family members. *Circles*, female family members. *Black symbols*, affected subjects. *White symbols*, unaffected subjects. *Grey shades*, DNA obtained. The *arrow* indicates the proband
B. Clinical photos of affected individuals (III-1, III-2 and III-4). All of them had trigonocephaly.

3.2 CLINICAL INFORMATION

3.2.1 III-1 (PROBAND)

He was born at full term by normal delivery. During the pregnancy, his mother (who had previously three children by another partner) experienced mild bleeding per vaginam and significant pelvic discomfort between 32 and 38 weeks. At 3 weeks after birth, he was referred to the craniofacial unit in Oxford for the assessment of his head shape. On examination, he had marked trigonocephaly, mild hypotelorism and low-set ears. His occipitofrontal circumference (OFC) was 38.5cm (95^{th} centile). No dysmorphic features were found in his limbs and body. A persistent cardiac murmur was noticed at 6 weeks of age and 2 large atrial septal defects (ASD) and a small patent ductus arteriosus (PDA) were detected by echocardiogram. A surgical closure for ASD and PDA was undertaken at the age of 13 months. A head CT scan at 13 months showed metopic suture synostosis, but the remaining cranial sutures were normal and no intracranial abnormality was detected. A fronto-orbital advancement and remodelling for metopic craniosynostosis was undertaken at 14 months. At 18 months, his weight was 10.9 kg (25th centile), height was 75.3 cm (10th centile) and OFC was 51.5cm (50th centile). Griffith Mental Development Scale showed age equivalent to 11 ½ months (chronological age, 18 months) and his overall developmental quotient (DQ)[3] was 71.8 (mild to moderate delay). Especially, a significant delay in receptive and expressive language skills was noted.

His head growth was good and follow-up CT scans detected no post-operative complications, however, he had a severe eating disorder and poor weight gain. At 8 years old, he was diagnosed with psoriasis and methotrexate treatment was initiated. He had joint problems on his ankles, which were attributed to psoriatic arthritis. At 9 years old, approximately 80% of the skin over his whole body was affected by psoriasis, after stopping methotrexate treatment for vaccination. By 12 years old, psoriasis was controlled with moisturising creams and stockings to protect his skin, but he had recurrent joint dislocation of his ankles.

[3] developmental quotient = ((developmental age) / (chronological age)) * 100
>= 85 normal, 71-84 mild to moderate delay, <=70 severe delay

During his adolescence, he continued to have complex psychological problems including eating and sleeping disorders as well as language delay and poor comprehension skills. Audiological testing was normal and post-operative follow up for craniosynostosis did not identify any features to suggest raised intracranial pressure.

3.2.2 III-2 (SIBLING OF III-1)

He was born at 36 weeks of gestation by normal delivery. His birth weight was 2.6 kg (10^{th} to 25^{th} centile). At birth, he was noted to have trigonocephaly, palpable symmetrical fontanelle, no obvious cleft or submucous cleft palate, bilateral inguinal hernias, bilateral undescended testes and absence of the left kidney on ultrasound scanning. His hernia was repaired at 7 weeks old. On assessment at 8 weeks, his OFC was 38 cm (10^{th} centile) and a small patent foramen ovale (PFO) and ASD were detected by echocardiogram, which are followed up and closed spontaneously by 10 years old. At 15 months, fronto-orbital advancement and remodelling was undertaken for metopic synostosis.

His developmental assessment at 14 months using Griffith Mental Development Scale showed developmental level equivalent to 10-12 months old. At 2 years and 11 months, his development was assessed by Schedule of Growing Skills, which showed general developmental problems, most strikingly in the area of speech and language. Psoriasis became significant from the age of 4 years old with joint problems, such as the fine motor problems and restriction of the joint movements in his hand, which persisted to the age of 12. Orchidopexy for undescended testes was undertaken at the age of 5.

From 2 years and 3 months old, he had difficulty in eating, probably due to gastro-oesophageal reflux. The problem persisted and he required nasogastric feeding until 7 years old. Post-operative course and skull development detected no problems.

3.2.3 III-4 (SIBLING OF III-1)

Due to the family history of craniosynostosis, his mother had repeated prenatal with sonography and his trigonocephaly was detected at 31 weeks of gestational age. He was born by normal delivery at 37 weeks with birth weight 3.1 kg (25^{th}-50^{th} centile). He had trigonocephaly, low set ears, high

arched palate, submucous cleft palate and large bilateral inguinal hernias at birth. A systolic murmur was detected at 3 months old and a small ASD, PFO and PDA were detected by echocardiogram. Those closed spontaneously by 18 months of age. At 3 months old, he underwent bilateral inguinal bilateral hernia repair. At 16 months old, a fronto-orbital advancement and remodelling was undertaken for his craniosynostosis. His developed severe psoriasis by the age of 16 months and had joint problems.

He had a moderate developmental delay, smiling at 4 months, sitting at 13 months and crawling at 16 months. His developmental assessment at the age of 16 months showed that language skills were equivalent to 9-12 months old. He had good progress on his global development after the surgical treatment for craniosynostosis, however, significant speech and language delay were observed. At 6 years old, the comprehension section of preschool language scales showed the low average range for his level of understanding (percentile rank of 18).

At 7 years old, his weight was 19.5kg (9^{th} centile) and height was 114.9cm (9^{th} centile).

3.2.4 OTHER AFFECTED FAMILY MEMBERS

There was a paternal family history of unusual head shapes. The proband's father, II-2, had a slightly unusual skull shape from birth, but has had no surgical history and clinical manifestations. On examination, II-2 had hypotelorism, flat midface, mild trigonocephaly, low set ears, high-arched palate and crowded teeth. He had a long proximal phalanx of the hands and clawed toes. From his childhood, he was affected by psoriasis and hearing and speech impairment. His brother (II-1) had similar features including a similar craniofacial appearance and psoriasis, and their deceased mother (I-2) reportedly had similar features. Given these, I-2, II-1 and II-2 were thought to be affected individuals. Summary of clinical phenotype is shown in Table 3-1.

TABLE 3-1 SUMMARY OF CLINICAL MANIFESTATION OF AFFECTED INDIVIDUALS

	II-1	II-2	III-1	III-2	III-4
Relation	Uncle	father	proband	sibling	sibling
Craniosynostosis	Metopic*	Metopic*	Metopic	Metopic	Metopic
Surgery for craniosynostosis	-	-	+	+	+
Congenital heart disease	-	-	ASD, PDA	PFO, ASD	ASD, PFO, PDA
Inguinal Hernia			-	+	+
Psoriasis	+	+	+	+	+
Developmental delay			+	+	+
Speech delay	+	+	+	+	+

*Clinically diagnosed with trigonocephaly.

3.2.5 DISEASE MODEL

The pedigree suggests an autosomal dominant inheritance pattern and the phenotypes of affected individuals in the family were very similar and distinctive; 1. Craniosynostosis was affecting only the metopic suture, 2. Affected cases were suffered from severe psoriasis, 3. Affected individuals have problems with speech and language interpretation. 4. III-1, III-2 and III-4 had congenital heart disease. 5. III-2 and III-4 had bilateral inguinal hernias. From the inheritance pattern, heterozygous variants segregating in the family were assumed as potential candidates. For the phenotype combination, initially, 2 possibilities were considered; one variant explains all phenotypes or 2 variants cause 2 different phenotypes (digenetic condition).

3.3 RESULTS

3.3.1 SNP CHIP ANALYSIS

Prior work on this family was conducted in 2010 by Dr Steve Twigg in the lab, which comprised analysis of the SNP chip segregation of 5 individuals (II-2, II-3, III-1, III-2 and III-4) combined with exome sequence of one of them (III-1). The regions that segregated, as identified by SNP chip, are shown in Table 3-2.

TABLE 3-2 INCLUDED AND EXCLUDED REGIONS FROM SNP CHIP

Event	Chr	Hg19/GRCh37				
		Included regions			Excluded regions	
		Start	End	Size (bp)	Size (bp)	Note
excluded	1				246,426,949	entire chromosome
excluded	2				242,616,062	entire chromosome
excluded	3				8,165,732	
included	3	8,232,626	188,895,382	182,170,450		
excluded	3				8,951,554	
included	4	0	13,553,129	13,123,944		
excluded	4				54,973,557	
included	4	68,453,189	89,835,017	21,918,256		
excluded	4				101,108,311	
excluded	5				3,624,321	
included	5	3,662,460	35,503,887	31,824,184		
excluded	5				137,969,091	
included	5	173,576,124	176,977,611	3,401,482		
excluded	5				3,727,928	
excluded	6				7,638,992	
included	6	7,788,602	106,051,024	98,424,116		
excluded	6				59,789,933	
included	6	166,027,660	171,115,066	4,792,824		
excluded	7				32,055,380	
included	7	32,230,177	157,637,962	125,134,021		
excluded	7				1,481,524	
included	8	0	3,505,698	3,326,288		
excluded	8				137,798,724	
included	8	141,222,648	146,293,086	4,972,060		
excluded	9				18,714,512	
included	9	18,761,099	81,671,791	62,110,512		
excluded	9				59,302,699	
excluded	10				135,154,325	entire chromosome
excluded	11				134,255,428	entire chromosome
included	12	0	975,276	783,657		
excluded	12				127,765,620	
included	12	130,045,204	130,320,725	275,521		
excluded	12				3,394,370	
included	13	19,263,735	24,201,883	4,938,148		
excluded	13				91,025,215	
excluded	14				4,094,810	
included	14	24,308,747	37,457,827	13,148,991		
excluded	14				60,713,500	
included	14	98,171,325	107,349,539	9,112,404		
excluded	15				3,301,899	

included	15	24,172,192	25,055,278	883,086		no known shared allele
excluded	15				77,608,988	
excluded	16				68,037,878	
included	16	69,525,697	90,354,752	20,602,258		
included	17	0	5,944,783	5,866,606		
excluded	17				72,755,347	
included	18	0	499,239	486,397		
excluded	18				75,626,315	
excluded	19				12,262,732	
included	19	12,619,771	56,506,417	48,717,458		
excluded	19				2,599,825	
excluded	20				11,108,197	
included	20	11,171,441	31,505,292	19,849,512		
excluded	20				31,411,399	
included	21	14,687,571	36,614,416	21,926,844		
excluded	21				11,386,966	
excluded	22				35,066,413	entire chromosome

3.3.2 WHOLE EXOME SEQUENCING (WES) OF III-1

Initially, 41,028 variants were listed from WES of III-1. The filtering process is shown in Table 3-3.The segregation data from SNP chip analysis was combined with the WES result and excluded 64 % of variants at the 4th step of filtering.

TABLE 3-3 FILTERING STRATEGY FOR WES OF III-1

Filtering strategy	No. of variants
Total number of variants	41,028
Number of exonic variants	20,072
dbSNP / 1000G filtering	1,168
SNP array segregation analysis	415
Artefact exclusion (due to highly homologous sequence)	200
Number of synonymous changes remaining	102
Number of nonsynonymous changes remaining	98
Excluding non-synonymous changes found in other 2 patients (likely artefacts)	48
Predicted damaging (PolyPhen-2)	18
Candidates remaining after experimental verification in pedigrees [1]	9

1 Genotype of 3 other individuals (I-1, II-1 and III-3) was obtained by dideoxy-sequencing.

From the last remaining 9 variants, a single segregating variant, *KIF16B* c.487C>T, p.R163W at chr20:16,492,132 was listed as the top candidate. *KIF16B* is a modulator of receptor recycling and degradation (Hoepfner et al., 2005) and the p.R163 is highly conserved, however, subsequently this variant appeared in the Exome Sequencing Project (ESP) and a later version of dbSNP as rs201732566. Considering the distinctive phenotypes of affected individuals, the variant was concluded to be unlikely to be causative.

3.3.3 HIGH RISK ALLELE FOR PSORIASIS SEGREGATES IN THE FAMILY

Psoriasis is a chronic autoimmune inflammatory skin disease, caused by a complex interplay between environmental and genetic factors (Nestle et al., 2009).

Classic genome-wide linkage analysis has identified *PSORS1* on chr 6 as a chromosomal locus with statistically the most significant linkage to psoriasis, this contains *HLA-C* segregating putative immune mechanism (Bowcock and Krueger, 2005). Subsequently, HLA-C 06:02 was identified as a major susceptibility *HLA-C* (Chr6:32,236,526-31,239,913) haplotype for early onset psoriasis (Gudjonsson et al., 2006). Since, the SNP chip segregation analysis on this family suggested the region of the *HLA-C* on chromosome 6 was segregating in all affected individuals (Figure 3-2), the *HLA-C* haplotype in the family was examined. The HLA-haplotyping using direct sequencing of *HLA-C* exon 2 and exon 3 was undertaken by Dr. Tim Robertson at John Radcliffe Hospital. As a result, HLA-C 06:02 was identified as the HLA-C haplotype which segregated in all individuals with craniosynostosis and psoriasis in this family (Figure 3-2).

This finding may well explain the segregation of psoriasis but would not explain the craniosynostosis, hence I continued the search for causal variants for craniosynostosis by undertaking exome sequencing of additional individuals in the family.

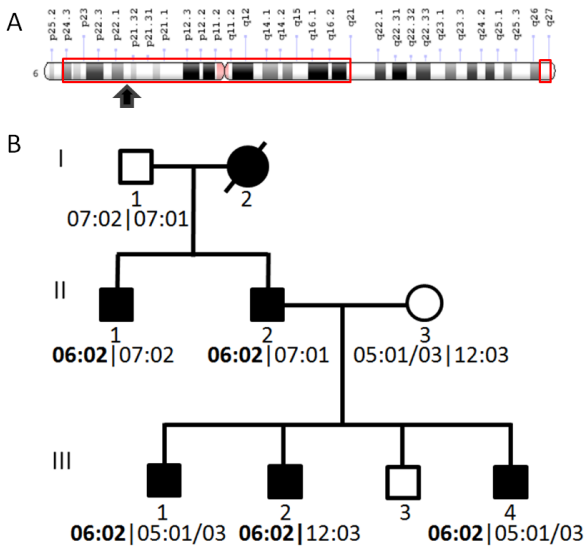


FIGURE 3-2 *HLA-C* HAPLOTYPE OF FAMILY A

A. Schematic diagram of chr6. Red rectangle, shared region detected by SNP chip analysis. Black arrow, the location of *HLA-C*.
B. Family tree with HLA haplotype. Black symbols, affected subjects. White symbols, unaffected subjects.

3.3.4 WES OF II-1 AND III-2

Considering that the quality of exome sequence of III-1 was relatively poor (63.8% x10 coverage and 37.4% x30 coverage), WES of additional affected individuals (II-1 and III-2) was undertaken as the next step to identify a causal variant for craniosynostosis.

The Agilent SureSelect Human All Exon V1 was used for III-1 and SureSelect Human All Exon V5 was used for II-1 and III-2. The coverages of WES of II-1 and III-2 were 85.9% x10 coverage and 77.8% x30 coverage and 80.6% x10 coverage and 67.9% x30 coverage, respectively.

The exome sequence data from II-1 and III-2 were put through our in-house bioinformatics pipeline and in total 6,834 variants in 3 exome sequences were listed[4]. The variants were filtered as below (Table 3-4) and the genotypes of unaffected individuals for the remaining 15 variants (at 6th step) were examined by dideoxy-sequencing. At this stage, we obtained a new DNA sample from III-3 (the unaffected brother), which increased the power of segregation analysis. As a result, 5 variants fit the segregation as shown in Table 3-4 and Table 3-5.

[4] Filtering condition; ESP 0.05, 1000G project 0.05, Max number of variants per gene 5

TABLE 3-4 DETAILS OF THE NUMBER OF VARIANTS DETECTED BY EXOME SEQUENCING OF THE FAMILY

Filtering strategy		No, of variants
total variants in II-1, III-1 and III-2		**6,834**
variants shared in II-1, III-1 and III-2 or II-1 and III-2		**1,245**
variants in the shared region determined by SNP chip		**400**
variants in coding regions		**74**
	frameshift	3
	nonframeshift	7
	nonsynonymous	32
	synonymous	28
	unknown	4
variants not in ESP, 1000G or dbSNP		**27**
variants after excluding in other exomes or mapping artefact		**15**
	frameshift	2
	nonsynonymous	7
	synonymous	6
variants not segregating in unaffected individuals (I-1 and III-3)		**5**

TABLE 3-5 CANDIDATE VARIANTS FROM WES OF II-1, III-1 AND III-2 AND ITS SEGREGATION IN THE FAMILY

Gene	Chr[1]	Position	Exonic_function	Variant		DS[2]	II-1	III-1	III-2	III-4	III-3[3]	I-1[3]	gnomAD
CHRD	**chr3**	**184,105,253**	**synonymous**	**c.2439C>G**	**p.V813V**	**0**	**GC**	**GC**	**GC**	**GC**	**CC**	**CC**	**not reported**
EVC2	chr4	5,710,119	nonsynonymous	c.122C>A	p.P41H	3	GT	GT	GT	GT	GT	GT	402/120256
CRMP1	chr4	5,868,475	synonymous	c.390A>G	p.R130R	0	AG	AG	AG	AG	AG	AG	2/244390
HIST1H2BE	**chr6**	**26,184,044**	**frameshift deletion**	**c.21delC**	**p.S7fs**	**0**	**fs**	**fs**	**fs**	**fs**	**no fs**	**no fs**	**359/277238**
HIST1H2AD	**chr6**	**26,199,204**	**nonsynonymous**	**c.268A>G**	**p.N90D**	**3**	**AG**	**AG**	**AG**	**AG**	**AA**	**AA**	**not reported**
ASCC3	**chr6**	**100,988,240**	**nonsynonymous**	**c.5574G>T**	**p.L1858F**	**2**	**AC**	**AC**	**AC**	**AC**	**CC**	**CC**	**1/245372**
ZNF212	chr7	148,950,727	nonsynonymous	c.709A>G	p.I237V	0	AG	AG	AG	AG	AG	AG	1/246252
ZNF16	chr8	146,156,530	nonsynonymous	c.1643C>T	p.T548I	2	AG	AG	AG	AG	AG	GG	3/276966
TLN1	chr9	35,699,085	nonsynonymous	c.6943G>A	p.A2315T	5	CT	CT	CT	CT	CC	CT	2/245522
PRUNE2	chr9	79,323,536	synonymous	c.3654T>C	p.S1218S	0	AG	AG	AG	AG	AA	AG	1/245700
PKD1L3	**chr16**	**72,020,171**	**synonymous**	**c.783C>T**	**p.P261P**	**0**	**TC**	**TC**	**TC**	**TC**	**CC**	**CC**	**not reported**
CCDC105	chr19	15,121,692	frameshift deletion	c.55_56del	p.19_19del	0	del	del	del	del	no del	del	133/186368
MED26	chr19	16,688,251	synonymous	c.390C>T	p.R130R	0	CT	CT	CT	CT	CC	CT	not reported
SCN1B	chr19	35,530,094	synonymous	c.522C>T	p.L174L	0	CT	CT	CT	CT	CC	CT	1/246252
EIF3K	chr19	39,111,001	synonymous	c.84C>G	p.T28T	0	CG	CG	CG	CG	CC	CG	1/246128

1 chromosome, 2 deleterious score, 3 unaffected individuals: genotypes were examined by dideoxy-sequencing, 4 not examined.
Five variants fitting the expected segregation pattern are shown in bold. Grey shades indicate the genotypes that didn't match the phenotype.

Of the 5 segregating variants, 2 variants are synonymous (*CHRD* c.2439C>G, p.V813V and *PKD1L3* c.783C>T, p.P261P), which have no splicing effects by Splice Site Prediction by Neural Network (SSPNN) (http://www.fruitfly.org/seq_tools/splice.html). The *HIST1H2BE* c.21delC is reported in gnomAD with the frequency of 0.00129. The *ASCC3* (Activating Signal Cointegrator 1 Complex Subunit 3) encodes a protein that belongs to a family of helicases that are involved in the ATP-dependent unwinding of nucleic acid duplexes. The Leu1858 locates in the SEC63 binding domain, but the residue is not conserved in *D. rerio* or invertebrates. Polyphen 2 predicted this change as probably damaging (0.965), however, the p.L1858F was reported in gnomAD (1/245372).

Given these findings, the *HIST1H2AD* c.268A>G (p.N90D, segregating perfectly in the family (Figure 3-3) without any report in ExAC or gnomAD was considered as the top candidate.

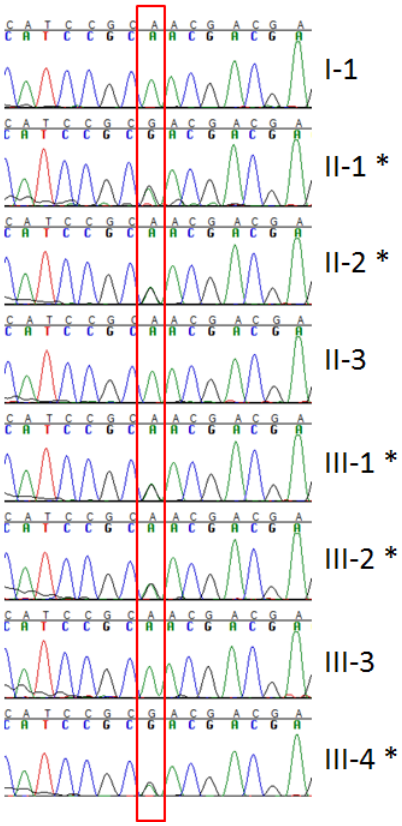


FIGURE 3-3 GENOTYPE OF *HIST1H2AD* C.268A>G AT CHR6 26,199,204 IN THE FAMILY

*indicates affected individuals. Only the affected individuals are heterozygotes of the variant.

3.3.5 SEGREGATION ANALYSIS BY USING WES OF II-1, III-1 AND III-2

To identify the regions that segregate with disease in the family, segregation analysis was carried out using 2 methods. Firstly, the list of 1245 variants from Table 3-4 was used combined with the SNP chip segregation data and blocks of segregating variants were identified. The DNA from other unaffected individuals (I-1 and III-3) was used to refine the regions by checking for non-segregation of the selected variants in the blocks. A second approach utilised the variant analysing software, MendelScan. The exome sequencing data of II-1, III-1 and III-2 were processed to find the regions of "absence of homozygous differences". Figure 3-4 shows the results of both segregation analyses. Together with the previous SNP chip data, the results suggested at least 6 regions were segregating in this family; Chr3:132,193,853-189,691,738, Chr5:32,249,097-34,802,939, Chr6:26,199,204-44,198,312 and 100,988.240-106,975,095, Chr16:69,996,751-90,002,074 and Chr20:10,624,963-31,654,699. The large segregating region on Chr6 covers both *HLA-C* and *HIST1H2AD* separated by 5 Mb in the distance, so that the *HLA-C* 06:02 and *HIST1H2AD* c.268A>G variants were co-segregating in as without any cross in the family.

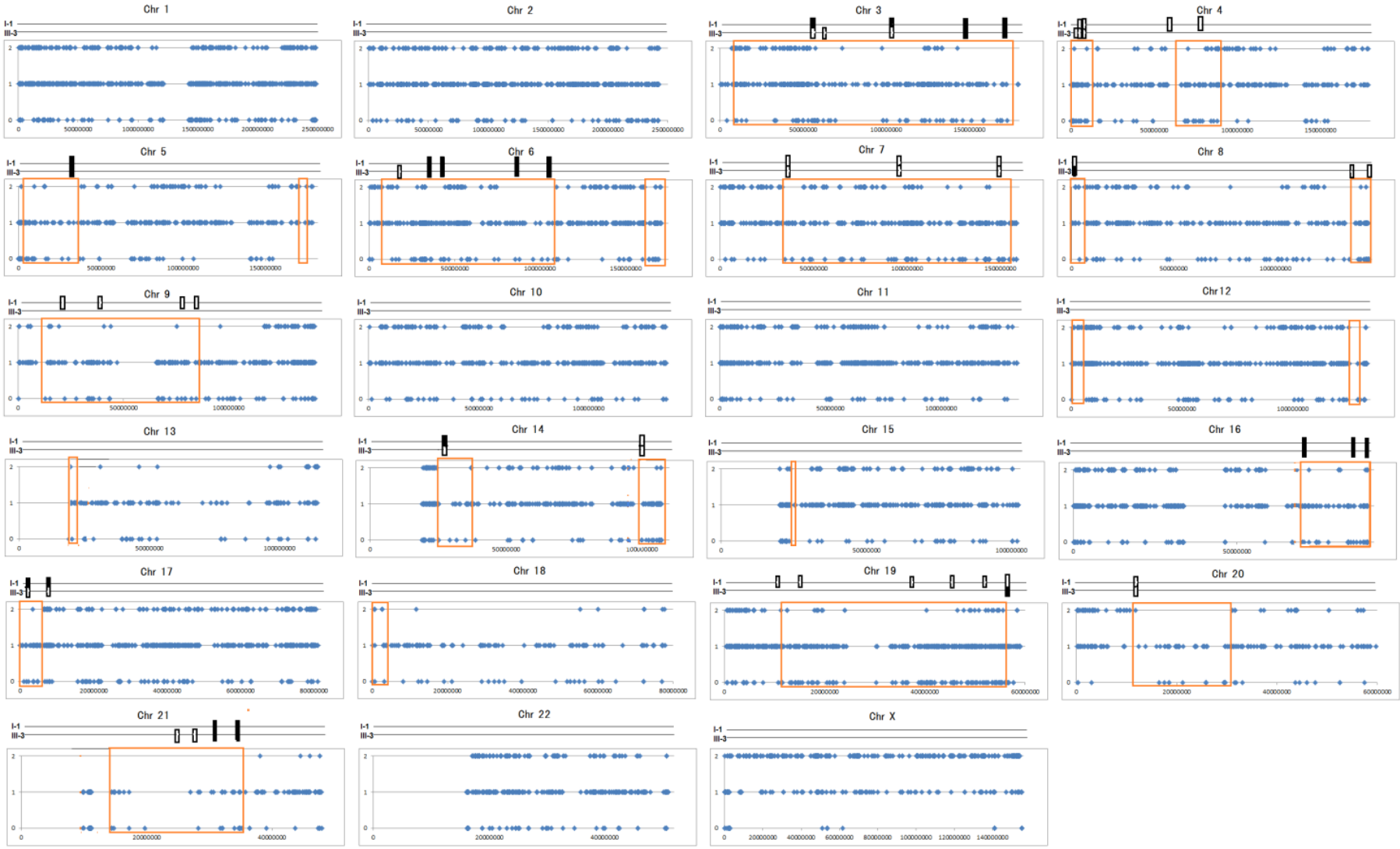


FIGURE 3-4 THE RESULTS OF SEGREGATION ANALYSIS

For each chromosome, the result of Rare Heterozygote Rule Out by MendelScan is shown. Light blue marks in graphs show variants along chromosome position. The top row of the graph is variants with homozygous differences between affected pairs (rule outs). Middle rows show variants shared between 2 individuals. Bottom rows are variants shared by 3 individuals. Orange rectangles show segregating regions detected by SNP chip analysis (II-2, II-3, III-1, III-2 and III-4). Black and white rectangles show the variants examined by dideoxy sequencing. White rectangles show the variants which segregated in the unaffected individuals I-1 or III-3, excluding the particular locus. Black rectangles show the variants which didn't segregate in I-1 or III-3, meaning that the locus was not excluded.

3.3.6 *HIST1H2AD* p.N90D AS A CANDIDATE VARIANT

In the eukaryotic nucleus, DNA is organised together with nucleosomes into a highly complex nucleoprotein structure called chromatin. The nucleosome consists of 2 each of the core histones H2A, H2B, H3 and H4 to form a histone octamer, which wraps and binds approximately 146 base pairs of DNA. In mammals each histone protein is encoded by multiple genes that are categorised into 2 major groups; replication-dependent (canonical) and replication-independent (replacement) histones (Banday et al., 2014) (Marzluff et al., 2002). Among core histones, the H2A family exhibits highest sequence divergence, resulting in the largest number of variants known (Bonisch and Hake, 2012), in which, *HIST1H2AD* encodes one of the 17 replication-dependent histone H2As, located on the largest histone gene cluster on chromosome 6p22-p21.3.

HIST1H2AD p.N90 is a highly conserved residue among all canonical H2A genes as well as most of replacement H2A variants and through evolution down to yeast (Figure 3-6 A). In gnomAD, the p.N90D variant has not been reported either in *HIST1H2AD* or in any of 16 canonical H2A genes, suggesting this is a unique variant to the family. The variants reported in gnomAD at position p.N90 in canonical H2A histones are listed in Figure 3-6 B. As illustrated in Figure 3-6 C, on the surface of nucleosome, the H2A-H2B dimer is known to form a negatively charged acidic patch (red colouring), which consists of acidic side chains of H2A (E56, E61, E64, D90, E91, E92[5]) and H2B (E110, E112) (Luger et al., 1997). The acidic patch has the shape of a narrow groove. The studies to date have shown that the acidic patch mediates the higher order structure of the chromatin fibre by interacting with the N terminal region of histone H4 as well as nucleosome-binding proteins such as LANA, IL-33, RCC1, Sir3 and HMGN2 (Kalashnikova et al., 2013; Mattiroli et al., 2014).

Matsubara et al. (2007) constructed 320 yeast mutant strains to understand the functional surface of a core histone. In the study, they showed H2A N90 is a functional surface residue with sensitivity to methyl-methanesulfonate (MMS) induced a double-strand break in DNA. (Matsubara et al., 2007) Moreover, in this study, 8 out of 320 point mutants showed lethality. Of these, four point mutants

[5] Corresponds to E57, E62, E65, D91, E92, E93 in (Matsubara et al., 2007) and our sequence

(H2A Y58, E62, D91 and H2B L109) located at the acidic patch and its nearby region, suggesting the acidic patch is functionally important in core histones. (Matsubara et al., 2007)

Interestingly, some replication-independent H2A variants have the variability in the acidic patch. For instance, the H2A.Z has an additional negatively charged residue (NDEELD vs NDEELN motif [6]) that interacts more tightly with the H4 tail in the array to promote intramolecular folding (Fan et al., 2004). The H2A.Bbd variant lacks residues involved in the acidic patch (NDRLLS instead of NDEELN) resulting in a decreased tendency of chromatin fibre folding (Zhou et al., 2007). The variability in the acidic patches confers the various degree of chromatin compaction and consequently, may cause changes in the regulation of transcription and replication (Shaytan et al., 2015).

The N90 locates just next to the acidic patch on the surface as shown in Figure **3-6** C and the change from asparagine to aspartic acid adds extra negative charge to the acidic patch (NDEELN to DDEELN). Given that the acidic patch modulates higher-order chromatin structure, this suggests that p.N90D could have a deleterious effect and be pathogenic.

The alignment of canonical H2A proteins differs mainly in the C-terminal six amino acids (Marzluff et al., 2002). So far, functional specialisation of canonical H2As has not been clearly demonstrated. However, some studies suggested canonical histone isoforms can possess distinct cellular function or show different expression levels across development (Banday et al., 2014; Singh et al., 2013).

The link between genes encoding histones and craniosynostosis is unknown to date, however, Cox et al. (2012) reported that a dominant D123N mutation in h3f3, one of the zebrafish variant histone H3.3 genes, causes cranial neural crest defects arising from insufficient H3.3 nucleosomal incorporation by forming aberrant H3 homodimers (Cox et al., 2012). Interestingly, the metopic suture uniquely forms within neural crest, while other sutures are of neural crest and/or mesoderm origin (Morriss-Kay and Wilkie, 2005).

[6] The first N is equivalent to p.N90.

The distribution or expression of canonical H2As is not clear in tissues or development. There is no expression data available in GTEx (https://www.gtexportal.org/home/). To see whether *HIST1H2AD* is expressed in human, RNA was extracted from fibroblast cells of III-2 and control and *HIST1H2AD* was amplified in synthesised cDNA. Primers were designed to be specific to *HIST1H2AD*, which was confirmed by running *in silico* PCR and the inspection of canonical H2As alignment. (It confirmed that primer sequences of *HIST1H2AD* have 2-7 bp differences to other canonical H2As' alignment.) Since *HIST1H2AD* has a single exon, RT-PCR without reverse transcriptase (RT) was performed to exclude the possibility of DNA contamination. As a result, the *HIST1H2AD* products were amplified successfully only in RT (+) samples. The following dideoxy-sequencing confirmed the variant in sample of III-2 (Figure 3-5). The result suggested that *HIST1H2AD* is expressed in human fibroblast.

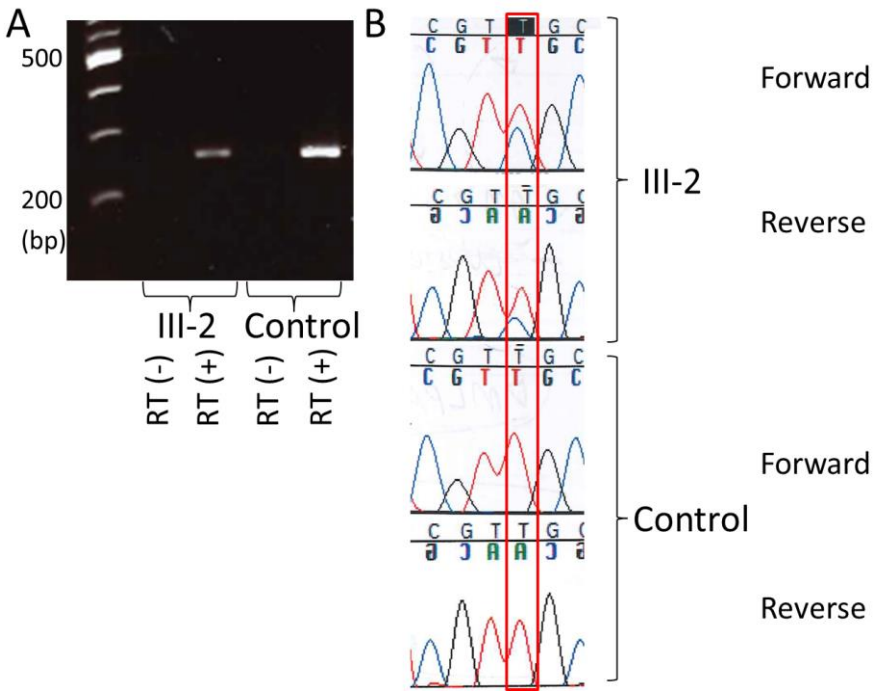
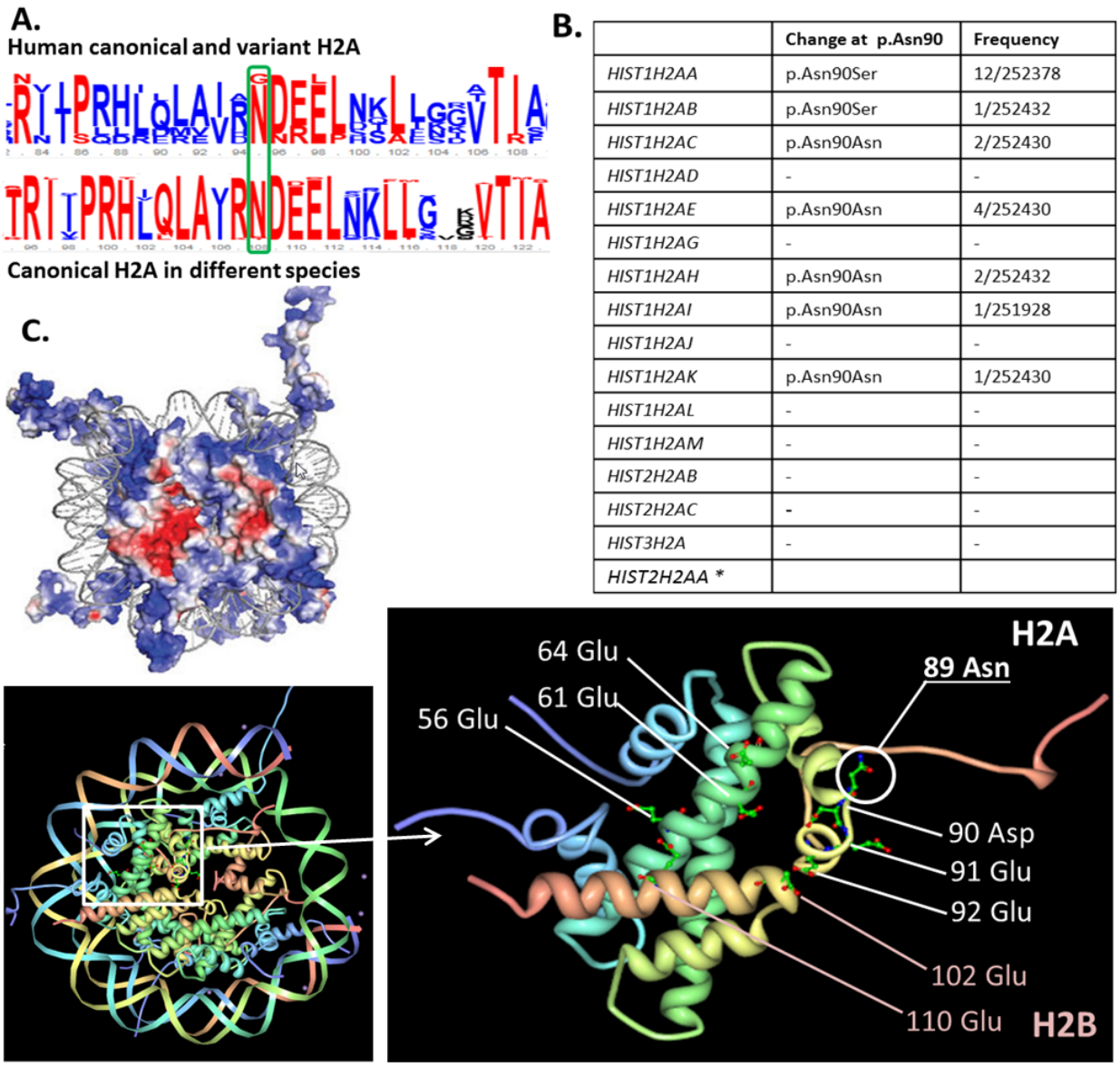


FIGURE 3-5 EXPRESSION OF *HIST1H2AD*

A. Amplified products on 3% agarose gel. Clear bands are detected in both RT(+) samples. RT(-) samples showed no amplification. RT; Reverse transcriptase.
B. Chromatogram of the products from A. Mutated allele are observed in dideoxy-sequencing of III-1.

Although *HIST1H2AD* p.N90D is the top candidate variant in this family and the gene is actually expressed in the human tissue, there is not enough evidence to determine this is a causative variant. To gather further evidence that *HIST1H2AD* p.N90D may be pathogenic, two approaches were taken; 1. Resequencing 16 genes encoding canonical H2As (the regions around the variant were focused on) in our craniosynostosis patient cohort, 2. Generation of a mouse model to assess the phenotype caused by the variant.



	Change at p.Asn90	Frequency
HIST1H2AA	p.Asn90Ser	12/252378
HIST1H2AB	p.Asn90Ser	1/252432
HIST1H2AC	p.Asn90Asn	2/252430
HIST1H2AD	-	-
HIST1H2AE	p.Asn90Asn	4/252430
HIST1H2AG	-	-
HIST1H2AH	p.Asn90Asn	2/252432
HIST1H2AI	p.Asn90Asn	1/251928
HIST1H2AJ	-	-
HIST1H2AK	p.Asn90Asn	1/252430
HIST1H2AL	-	-
HIST1H2AM	-	-
HIST2H2AB	-	-
HIST2H2AC	-	-
HIST3H2A	-	-
HIST2H2AA *		

FIGURE 3-6 ACIDIC PATCH AND *HIST1H2AD* P.N90D

A. The curated sequence of paralogues H2A from HistoneDB 2.0 https://www.ncbi.nlm.nih.gov/projects/HistoneDB2.0/index.fcgi/browse/
Top panel; Curated protein sequences of human canonical H2A and variant H2As (H2A.P, H2A.1, H2A.X, H2A.B and H2A.Z)
Bottom panel; Curated protein sequence of canonical H2A in 35 species, vertebrates and yeasts.

B. Changes at 90th asparagine in other canonical H2As in gnomAD. * no data are available for HIST2H2AA (including both HIST2H2AA3 and HIST2H2AA3) on gnomAD

C. Top left; the red region indicates the acidic patch in nucleosome from (Mattiroli et al., 2014). Bottom left and right; screenshots from Protein Data Bank. ProteinID-1AOI (Complex between nucleosome core particle (H3, H4, H2A and H2B) and 146bp long DNA fragment (*Xenopus*)). The right panel is an expansion of white rectangle on the left and only H2A and H2B dimer was shown with amino acid constituting acidic patch individually numbered. Note that H2A Asn89 corresponds to Asn90 in (Matsubara et al., 2007) and our sequence. The same transformation applies to the other residues of H2A listed.

3.3.7 SEARCH FOR ADDITIONAL CASES

3.3.7.1 SEARCH IN CRANIOVAR

The search in CranioVar detected 3 patients including the proband as shown in Table 3-6.

HIST1H2AD p.K119R was inherited from the unaffected mother and the same variant is in gnomAD. *HIST1H2AM* p.P81T was inherited from the unaffected mother and p.P81T was reported in other canonical H2A (*HIST1H2AG)* with a frequency of 0.000008238. In addition, the p.P81 was not found to be functional in yeast study (Matsubara et al., 2007). Therefore, these 2 cases were thought unlikely to be causative.

TABLE 3-6 H2A VARIANTS IN CRANIOVAR

Gene	Chr	Start	Function	Change	Change	DS [1]	frequency in gnomAD	Phenotype	Case ID
HIST1H2AD	chr6	26198976	nonsynonymous	c.268A>G	p.N90D	4	-	SMS[2]	proband
HIST1H2AD	chr6	26198888	nonsynonymous	c.356A>G	p.K119R	2	1/246228	Pfeiffer Cloverleaf skul	6463
HIST1H2AM	chr6	27892909	nonsynonymous	c.241C>A	p.P81T	3	-	SBC[3]	4103

1. Deleterious score, 2. Syndromic metopic synostosis, 3. Syndromic bicoronal synostosis

3.3.7.2 RESEQUENCING OF THE REGION OF VARIANT IN CANONICAL H2AS IN CRANIOSYNOSTOSIS COHORT

To find additional cases as with the variant in H2A coding genes, resequencing of 17 genes encoding canonical H2A was planned. This was challenging as 60% of the exonic sequences in 17 genes are identical. (In particular, two genes, *HIST2H2AA3* and *HIST2H2AA4* have completely identical sequences.) To enable the multiplex targeted sequencing of 17 highly similar H2A genes, molecular inversion probes (MIPs), which are reported as a practical and effective capture method for accurately detecting mutations (Hiatt et al., 2013), were used. MIPs were designed by using MIPgen http://shendurelab.github.io/MIPGEN/ (Boyle et al., 2014), and 16 unique probes to each target was selected for each of 17 genes (1 probe for identical *HIST2H2AA3* and *AA4*). To obtain the specificity to the target, MIPs were designed at different positions (not equivalent positions) among each, so that only 8 bp including the codon encoding 90th Asn were covered in all target region (Figure 3-7 B). Considering the extremely distinctive phenotype of this family, the causal variant was considered

likely to be associated with a specific phenotype, therefore, we thought focusing on the specific codon including Asn90 should be a reasonable approach.

.

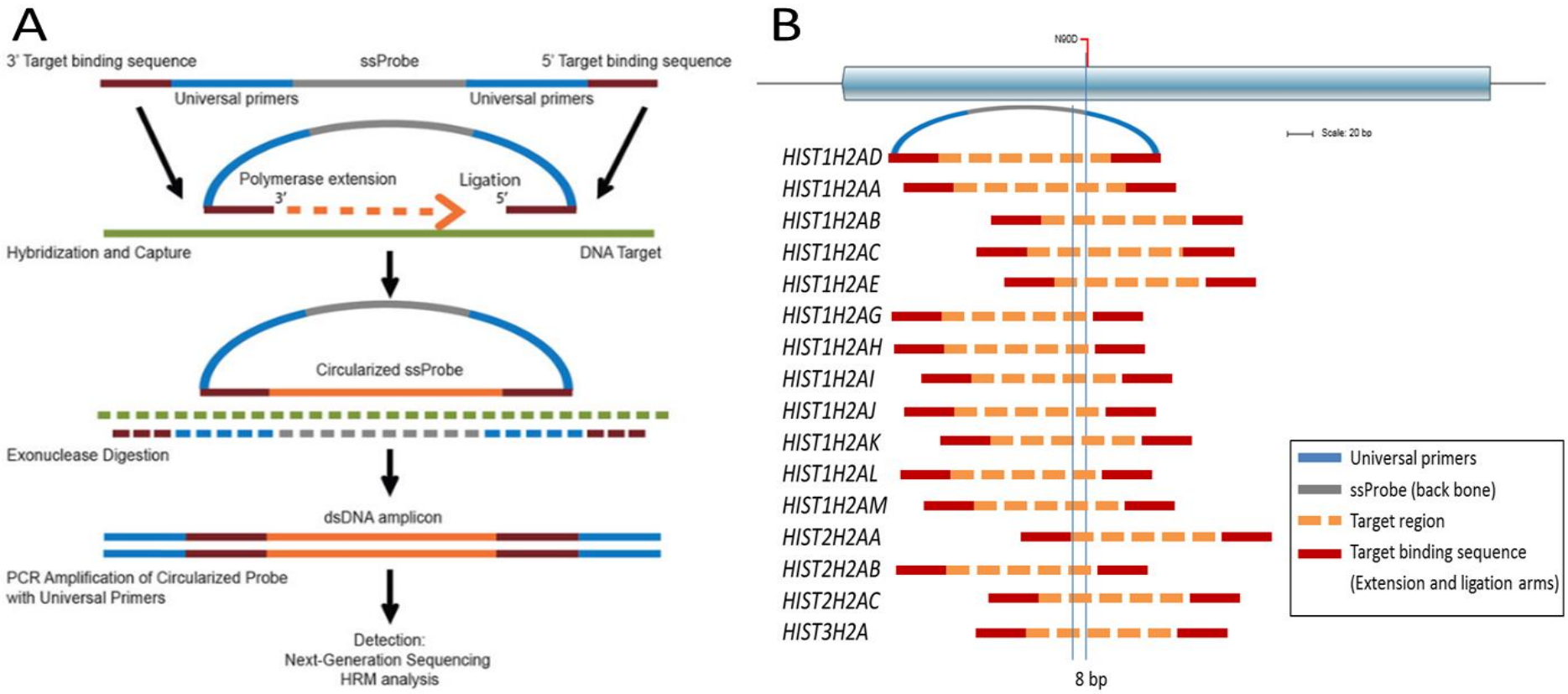


FIGURE 3-7 SCHEMATIC IMAGE OF MIPS

A First step; MIPs flow-chart from (Stefan et al., 2016). The MIPs, probes which consist of target binding sequences, universal sequences and ssProbe, hybridize to genomic DNA. Targeted regions are captured by polymerase extension and ligation. Second step; Unhybridised genomic DNA and probes are digested by exonuclease. Third and fourth step; Circularised probes are PCR amplified with universal probes and sequenced by Next Generation Sequencing.
B Schematic image of the target region of each MIP.
The blue rectangle on top shows exon of H2A gene. MIPs probes comprise universal primers, ssProbe and target binding sequences at both ends (extension and ligation arms). Targeted regions are shown in orange line (120-140 bp in length). Eight base pairs including 90th Asn codon are covered by all MIPs.

3.3.7.2.1 SUMMARY OF PHENOTYPE

In total 492 unsolved craniosynostosis cases were resequenced, of which, full coverage (more than 10 reads in all resequenced H2As, focusing on 8bp which covered in all MIPs design) was obtained in 454 cases. The phenotype of these 454 cases were summarised in Table 3-7.

TABLE 3-7 SUMMARY OF RESEQUENCING COVERAGE

	Non-syndromic	Syndromic	Combined
Metopic	131	27	158
Sagittal	45	27	72
Unilateral coronal	101	13	114
Bilateral coronal	9	11	20
Uni- or bilateral lambdoid	7	3	10
Multisuture	34	20	54
Syndromes[1]	0	20	20
Other or unknown	6		6
Combined	327	121	454

1 Clinically diagnosed as Carpenter (2 cases), CFNS (1 case), Crouzon (5 cases), Pfeiffer (3cases) and Saethre-Chotzen (9 cases) syndrome.

3.3.7.2.2 VARIANTS IDENTIFIED FROM RESEQUENCING

Considering that the pLI[7] for each canonical H2A gene is low and loss of function (LoF) variants can be replaced by other H2A copies, nonsynonymous variants were focused upon. Variants reported in gnomAD were thought to be non-pathogenic variants, considering the distinctive phenotype of the affected individuals.

Nonsynonymous variants detected from resequencing are listed in below (Table 3-8). Positive controls in the panel were successfully detected. *HIST1H2AM* p.P81T and *HIST1H2AD* p.K119R had been identified in CranioVar (section 3.3.7.1).

For others, *HIST1H2AH* p.H124Y and *HIST1H2AB* p.A71T are reported in gnomAD. *HIST1H2AC* p.T60A is not reported in gnomAD, however, p.T60A is reported in *HIST1H2AG* and *HIST1H2AJ* with the

[7] The closer pLI is to 1 (the minimum 0 to the maximum 1), the more LoF intolerant the gene appears to be. pLI >= 0.9 is considered as a LoF intolerant set of genes

frequency 2/246250 and 7/237666, respectively. As a result, no possible pathogenic variants were identified.

TABLE 3-8 VARIANTS DETECTED FROM RE-SEQUENCING

Gene	Chr	Start	Function	change	AA change	DS	frequency in gnomAD	Sample ID	Plate
HIST1H2AH	chr6	27115277	nonsynonymous	c.373C>T	p.H124Y	1	7/245988	393	MP1
HIST1H2AD	**chr6**	**26198976**	**nonsynonymous**	**c.268A>G**	**p.N90D**	**4**	**-**	**proband**	**MP2**
HIST1H2AM	chr6	27892909	nonsynonymous	c.C241C>A	p.P81T	3	-	4103/4564	MP7
HIST1H2AB	chr6	27892909	nonsynonymous	c.211G>A	p.A71T	5	2/246252	6536	MP8
HIST1H2AD	chr6	26199116	nonsynonymous	c.356A>G	p.K119R	2	1/246228	6331	Met2
HIST1H2AC	chr6	26124638	nonsynonymous	c.178A>G	p.T60A	4	-	6343	Met3

3.3.7.3 H2A VARIANTS IN DECIPHERING DEVELOPMENTAL DISORDERS

Subsequently, data on all H2A encoding genes was obtained from the Deciphering Developmental Disorders (DDD) to find any possible pathogenic variants. Cases with a nonsynonymous variant in canonical H2As were extracted from the DDD data and shown in Table 3-9. No cases of p.N90D were identified, however, some variants remained on the list even after excluding variants reported in gnomAD based on 3 criteria; 1-same nonsynonymous variant is reported in the same gene, 2-other nonsynonymous change is reported at the same position in the same gene and 3-same nonsynonymous change is reported in other H2A genes. Of these, DDDP 123440 *HIST1H2AI* p.Gly107Asp was found to be the most interesting for several reasons; (1) No amino acid changes are observed in other canonical histones in gnomAD; (2) Variant is *de novo*; (3) The amino acid is reported to be functional in the yeast study (Matsubara et al. 2007) ;(4) The case has some overlap clinical features with our case.

This case could be an additional potential candidate related to developmental disorders, so that it would be worth to be followed up, particularly when p.N90D is confirmed as pathogenic.

TABLE 3-9 THE CASES WITH CANONICAL H2A VARIANTS IN DDD DATA

proband	sex	chr	position	gene	ref/alt alleles		Gnom AD 1	Note	origin	singleton or trio	Yeast study 2	Phenotype
DDDP 110356	M	6	26,033,642	*HIST1H2AB*	A/G	p.Leu52Pro	2			singleton		
DDDP 129552	M	6	26,124,513	*HIST1H2AC*	G/C	p.Arg18Pro	3		de novo			
DDDP 120049	M	6	26,124,629	*HIST1H2AC*	G/C	p.Glu57Gln	3			singleton		
DDDP 102639	**F**	**6**	**26,124,663**	***HIST1H2AC***	**G/T**	**p.Gly68Val**		**change to Asp or Arg in other histones**	**de novo**		**not tested**	**An abnormal form of the vertebral bodies|Delayed speech and language development|Hypoplasia of the maxilla|Pectus carinatum|Seizures|Thoracic kyphosis**
DDDP 101892	F	6	26,124,725	*HIST1H2AC*	C/T	p.Arg89Cys	3		from affected father			
DDDP 101329	**M**	**6**	**26,124,737**	***HIST1H2AC***	**G/A**	**p.Glu93Lys**		**change to Asp or Arg in other histones**	**de novo**		**HU, MMS**	**Autistic behaviour |Delayed speech and language development**
DDDP 116506	F	6	26,199,156	*HIST1H2AD*	C/T	p.Gly106Ser	3		from affected father			
DDDP 111287	**M**	**6**	**26,199,304**	***HIST1H2AD***	**C/G**	**p.Leu56Phe**		**Change to Val, Pro, Arg**		**singleton**	**not tested**	**Abnormally lax or hyperextensible skin|Anterior beaking of lumbar vertebrae|Cigarette-paper scars|Deeply set eye|Delayed speech and language development|Hyperextensibility of the finger joints|Specific learning disability**
DDDP 108537	F	6	26,199,422	*HIST1H2AD*	G/C	p.Thr17Ser	1			singleton		
DDDP 110868	M	6	26,199,422	*HIST1H2AD*	G/C	p.Thr17Ser	1		from affected mother			
DDDP 117362 117361	**F**	**6**	**26,217,411**	***HIST1H2AE***	**C/T**	**p.Ala70Val**		**Change to Pro, Arg**	**de novo**		**not tested**	**Abnormal lower motor neuron morphology|Microcephaly|Sleep disturbance|Specific learning disability|Upslanted palpebral fissure**
DDDP 130396	F	6	27,100,947	*HIST1H2AG*	C/A	p.Arg33Ser	2		de novo			
DDDP 111370	F	6	27,100,962	*HIST1H2AG*	G/T	p.Gly38Cys	1		from affected father			
DDDP 117042	F	6	27,115,026	*HIST1H2AH*	A/C	p.Tyr40Ser	3			singleton		

DDDP 123440	**M**	**6**	**27,776,307**	***HIST1H2AI***	**G/A**	**p.Gly107Asp**		**no change at all**	**de novo**		**Spt,HU MMS**	**Atrial septal defect|Chordee|Cryptorchidism|Gastrostomy tube feeding in infancy|Severe global developmental delay|Ventriculomegaly**
DDDP 118998	**F**	**6**	**27,782,433**	***HIST1H2AJ***	**C/T**	**p.Gly29Asp**		**Change to Arg, Ser**		**singleton**	**no effect was seen**	**Microcephaly|Mild global developmental delay|Seizures**
DDDP 125945	F	6	27,782,449	*HIST1H2AJ*	G/A	p.Leu24Phe	3			singleton		
DDDP 111420	F	6	27,782,478	*HIST1H2AJ*	T/C	p.Lys14Arg	3		from affected father			
DDDP 116635	F	6	27,805,903	*HIST1H2AK*	C/T	p.Arg72His	3			singleton		
DDDP 110827	M	6	27,833,137	*HIST1H2AL*	C/G	p.Ser2Trp	2		de novo			
DDDP 116893	F	6	27,833,160	*HIST1H2AL*	A/G	p.Lys10Glu	3			singleton		
DDDP 117120	F	6	27,833,175	*HIST1H2AL*	GC/TC	p.Ala15Ser	1			singleton		
DDDP 112377	M	6	27,860,909	*HIST1H2AM*	G/A	p.Q7stop	LoF		from affected mother			
DDDP 118580	F	1	149,858,706	*HIST2H2AC*	C/T	p.Ala61Val	3		from affected father			
DDDP 113199	M	1	149,859,151	*HIST2H2AB*	C/G	p.Gly106Ala	2		de novo			**De novo* variant in *USP9X* was reported as well.
DDDP 129708	F	1	228,645,425	*HIST3H2A*	G/A	p.His32Tyr	3		from affected mother			

1 Variants in gnomAD; 1, same change reported in the same gene. 2, other nonsynonymous change at the same position. 3, same change is reported in other H2A genes.

2 Yeast study; data are shown in bold cases only. Spt, Suppressor of Ty phenotype. HU and MMS, sensitivities to hydroxyurea and methyl-methanesulfonate, respectively,

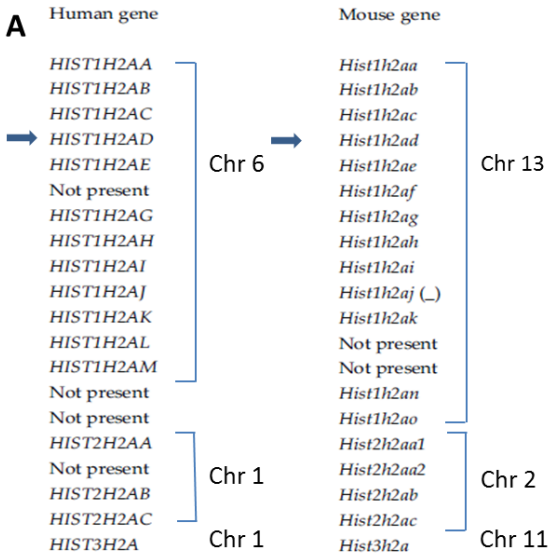
3.3.8 GENERATION OF MOUSE MODEL USING CRISPR/CAS9 TARGETING

As mentioned above, *HIST1H2AD* p.N90D is a novel variant segregating perfectly in the family and likely to have a functional effect, however, the family is not big enough to conclude *HIST1H2AD* p.N90D is the pathogenic mutation. To find the functional effect of the variant, making a histone point mutant p.N90D in *Saccharomyces cerevisiae* (yeast) or chromatin folding assay have been taken into consideration, however, those would not conclude if the heterozygous variant in just one of 17 canonical H2A genes could cause the patients' phenotypes. To answer the question more directly, making the mouse model of *HIST1H2AD* p.N90D was undertaken in parallel to the resequencing of H2A genes.

The mouse has 18 canonical H2A genes as shown in Figure 3-8 A (Marzluff et al., 2002). Firstly, the mouse orthologue of human *HIST1H2AD* was sought. The *HIST1H2AD* gene locates in the largest cluster, HIST1, which is on chromosome 6 (6p21-p22) in humans and chromosome 13 in mice. The *HIST1H2AD* in human and *Hist1h2ad* in mouse locate at the equivalent location in HIST1 and next to the same genes, *HIST1H2BF* (Human)/*Hist1h2bf* (Mouse) and *HIST1H3D* (Human)/*Hist1h3d* (Mouse) as seen in Figure 3-8 C. Subsequently, sequence similarity between *HIST1H2AD* and *Hist1h2ad* were examined. The sequences between *HIST1H2AD* and *Hist1h2ad* showed the highest similarity among the H2A genes in the same part of HIST1 (Figure 3-8 B). Considering these, *Hist1h2ad* was concluded to be the likely mouse orthologue of human HIST1H2AD. Therefore, CRISPR/Cas9 targeting was designed to target Hist1h2ad.

In terms of the matrix in Figure 3-8 B, H2A genes in mouse are more similar to each other compared to H2A genes in human. This could suggest a different evolution of the mouse and human tandem gene array, so that precise functional equivalence might still be questioned.

A



B

	HIST1H2AE	HIST1H2AC	HIST1H2AD	Hist1h2ac	Hist1h2af	Hist1h2ae	Hist1h2ad
HIST1H2AE	100	75.24	74.85	86.01	85.5	85.86	85.24
HIST1H2AC	75.24	100	73.53	87.53	87.02	87.63	87.28
HIST1H2AD	74.85	73.53	100	86.26	86.77	86.11	87.02
Hist1h2ac	86.01	87.53	86.26	100	97.18	97.96	98.46
Hist1h2af	85.5	87.02	86.77	97.18	100	97.96	98.72
Hist1h2ae	85.86	87.63	86.11	97.96	97.96	100	99.24
Hist1h2ad	85.24	87.28	87.02	98.46	98.72	99.24	100

C

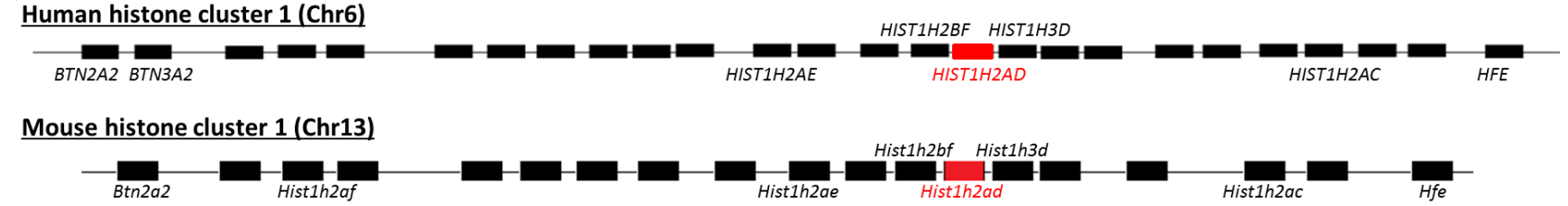


FIGURE 3-8 THE COMPARISON BETWEEN HUMAN *HIST1H2AD* AND MOUSE *HIST1H2AD*

A. Mammalian replication-dependent histone genes modified from (Marzluff et al., 2002).
B. Percent Identity Matrix among mouse and human H2A genes in the same part of the cluster 1- created by Clustal2.1. Comparison of exonic sequence (only open reading frames).
C. Schematic image of the positions of *HIST1H2AD* (human) and *Hist1h2ad* (mouse) in histone cluster 1.

3.3.8.1 *HIST1H2AD* TARGETING

The coding sequences of the canonical H2A genes are very highly conserved, especially in mouse; 129 amino acids encoded by *Hist1h2ab, Hist1h2ac, Hist1h2ad, Hist1h2ae, Hist1h2ag, Hist1h2ai, Hist1h2ak, Hist1h2an* and *Hist1h2ao* are identical (Marzluff et al., 2002). Therefore, taking a direct targeting approach with a sgRNA designed against the coding region was found to be extremely difficult; sgRNA searches showed that any sgRNA close to the variant detected 20 or more target hits. With the support of Dr. Philip Hublitz, an alternative approach utilizing 2 sgRNAs targeting surrounding unique sequences (one in the upstream 5'UTR and the other in the downstream intergenic region) and a larger donor template (2979 bp -Figure 2-1) including about 1 kb homology arms and chr13:23,574,738 c.268A>G p.N90D were used for homology direct repair (HDR) to target *Hist1h2ad* specifically and avoid off-target effects.

In the gRNA design process, care was taken to ensure that (1) the promoter of *Hist1h2ad* remained intact to avoid disrupting the gene expression, and (2) inactivating the expression of upstream and downstream genes did not occur. The sgRNAs were designed using CRISPRscreen (http://apps.molbiol.ox.ac.uk/CRISPR/) and E-CRISP Design (http://www.e-crisp.org/E-CRISP/) and no off-targets were predicted.

The donor template was designed to contain the target variant chr13:23,574,738 c.268A>G, inactivated PAM sites (so that donor template was not cut by sgRNAs) and included a novel SacII restriction site useful for detecting the template (Figure 2-1) As shown in Figure 3-9, primers were designed to screen for correct targeting of *Hist1h2ad*.

Targeting efficiency is shown in Table 3-10. With this strategy, the variant was successfully introduced into *Hist1h2ad* of mouse ES cell E14 (Figure 3-9 gel image and chromatogram of the bottom panel). The genotype of mutant clones was verified by dideoxy sequencing. Karyotyping and pluripotency tests were carried out and blastocyst injection of targeted ES cells was undertaken by the Genome Engineering Facility in the WIMM.

TABLE 3-10 TARGETING EFFICIENCY IN EACH 96 WELL PLATE

	Alive cells / 96[1]	PCR amplified [2] / Alive cells (%)	Mutant clones / PCR amplified (%)	
Plate 1	40	28 (70 %)	0 (0 %)	
Plate 2	47	26 (55 %)	0 (0 %)	
Plate 3	43	39 (90 %)	5 (13 %)	4 homozygous, 1 heterozygous
Plate 4	53	33 (62 %)	5 (15 %)	5 homozygous
Plate 5	52	32 (62 %)	1 (3 %)	1 homozygous
Total	**235 / 480 (49 %)**	**158 / 235 (67 %)**	**11 / 158 (7%)**	

1 The number of clones alive after single cell sorting

2 The number of clones from which DNA was extracted and PCR was successfully used to amplify products assay for the insertion of the mutation.

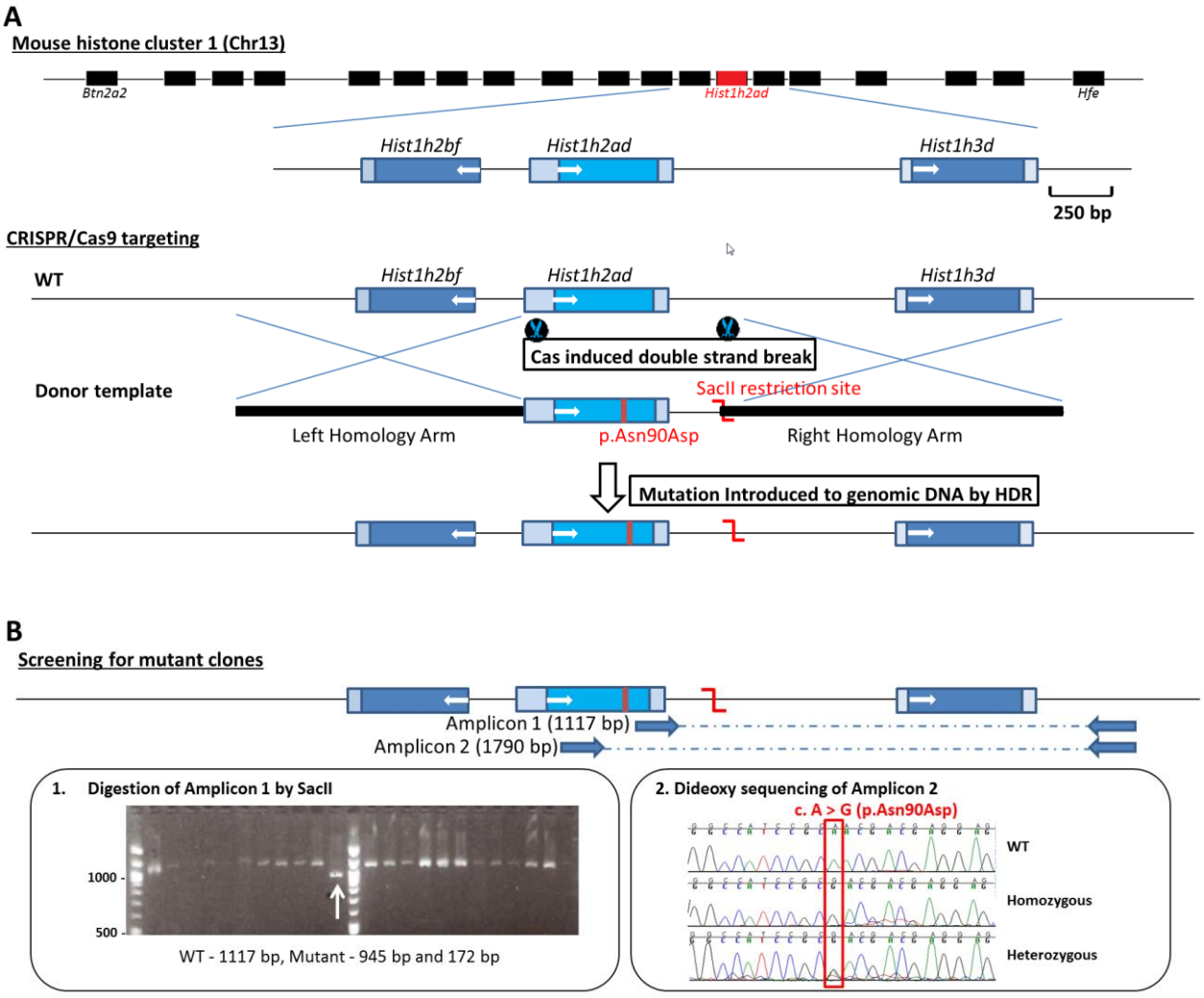


FIGURE 3-9 SCHEMATIC IMAGE OF CRISPR/CAS9 DESIGN TO INTRODUCE P.N90D C.268A>G INTO *HIST1H2AD*

A Top panel; schematic diagram of the target region for CRISPR/Cas9 gene editing. *Black rectangles*, genes; *red rectangle*, the target gene, *Hist1h2ad*. *Grey rectangles* indicate UTRs. *White arrows* indicate the direction of gene transcription.
Middle panel; diagram of Cas9 induced double-strand break and homology direct repair (HDR). The sgRNA target positions are indicated by scissors and the vertical red line shows the position of the mutation. A SacII restriction site, which inactivates the PAM site of the downstream sgRNA sequence was included in the donor template.
B Screening for targeted ES cell clones.
Blue arrows indicate the location of forward and reverse primers used for screening of clones. To detect the donor template inserted into the targeted region, the reverse primer was designed on the downstream sequence of the end of the donor template. The forward primer of Amplicon 1 locates upstream of the SacII restriction site. The forward primer of Amplicon 2 locates upstream of chr13:23,574,738 c.268A>G. The screening of mutant clones was done in 2 steps; the first screening was done with Amplicon 1 and digested clones were amplified again using Amplicon 2 followed by sequencing of the PCR products.
The white arrow in the gel image is an example of a clone with a homozygous SacII restriction site. The chromatogram shows the dideoxy sequencing of wild type, homozygous and heterozygous clones.

3.3.8.2 CHIMERIC MICE

From the microinjection, in total 3 chimeric mice (2 females and 1 male) were generated. These were bred to C57Bl/6 to produce the first generation of heterozygous mutants, however, only wild type first generation mice were born to date (September 2017).

3.4 DISCUSSION

The phenotype of this family is clearly unusual to have segregation of metopic synostosis such high penetrance. In addition to the metopic synostosis, syndromic features including congenital heart disease, inguinal hernia and psoriasis were segregating in the family. It was initially unclear whether all features had the same underlying cause. However, at an early stage of the investigation, the evidence that a high-risk HLA-C allele for psoriasis is segregating in the family was obtained. This association is well described and affected individuals do not have additional syndromic features, concluded that another variant must account for the metopic synostosis and other feathers.

As described, the *HIST1H2AD* c.268A>G p.N90D is considered to be the promising candidate, however, investigations to support the variants' pathogenicity are ongoing.

Resequencing of the specified position across all canonical H2A genes could not identify additional cases with possible pathogenic variant. Highlighting on metopic synostosis, the patients screened with full coverage, 27 and 131 cases were syndromic and nonsyndromic metopic synostosis, respectively. However, considering that phenotype of this family is extremely distinctive (no other families in our database exhibit syndromic metopic synostosis with apparent autosomal dominant inheritance), the variant might be very specific to the family.

Similarly, over 20 cases with a candidate pathogenic variant in canonical H2A were reported in DDD, however, most of them seem to be incidental or subclinical. After filtering out previously reported variants or previously described amino acid change in other H2As in gnomAD, only handful variants were remained (Table 3-9). Of those, DDDP123440 (Chr:27,776,307 G>A p.Gly107Asp) has some overlapping with our case in terms of phenotype and arouse our interest. However, same as our

case, generating a mouse model would be required to investigate the variants' effect causing the phenotype. Thus, we decided to assess the phenotype of the mutant mouse carrying *HIST1H2AD* c.268A>G and pathogenicity of the variant first, before investigating the case, DDDP123440, further.

Several projects are still ongoing to determine the pathogenicity of *HIST1H2AD* c.268A>G p.N90D including the generation of the mouse model and RNA-seq, by using the mutant (*Hist1h2ad* c.268A>G p.N90D) and wild type mouse ES cells which were generated by CRISPR/Cas9 targeting.

A nucleosome is a basic unit of DNA package to compact DNA and the higher order structure made by the folded nucleosome is important in regulating correct gene expression. As described in section 3.3.6, *HIST1H2AD* p.N90 locates just next to the acidic patch, a key element of the nucleosome surface to compact DNA, so that the variant may affect gene expression and regulation of other genes. Currently, RNA-seq on mutant and wild type mouse ES cells are ongoing to compare the overall gene expression and see what kind of genes are differentially affected.

On the other hand, the disruption caused by the variant could be tissue-specific. In terms of the mouse model, when the heterozygous mutant mice are born, in addition to the observation of phenotype and histological analysis of the suture, suture materials from wild type and mutant mice would open up the possibility of further analysis such as comparison of gene expression and regulatory interaction. Breeding to homozygous mice would provide additional opportunity to explore effects of gene dosage. However, even if no positive findings are observed from these experiments, there might be evolutionary differences between human and mouse core H2As (As described in Figure 3-8, sequence similarity is not equal between human and mouse). In this case, RNA seq using human suture samples (either derived from iPS cells or transgenic) would be another option, but probably only if the pathogenicity of the variants were strongly suggested by finding additional cases with the similar mutation.

For the CRISPR/Cas9 targeting, although it was designed to target *Hist1h2ad* specifically, there is still a possibility that of homology-directed repair would have occurred at an inappropriate region

(random insertion). A screen for random insertion has not been undertaken so far. This would be undertaken by Southern blotting.

Recently, germline mutations affecting the histone H4 core are reported to cause a developmental syndrome by altering DNA damage response and cell cycle control. *De novo* heterozygous missense mutations (p.K91Q and p.K91R) in *HIST1H4C* were identified in subjects presenting with similar phenotypes, comprising short stature, microcephaly, intellectual disability, characteristic facial features and foot anomalies. By generating and investigating K91Q- and K91R- expressing zebrafish embryos, the authors suggested that an increase of DNA double-strand breaks in mutant embryos lead to abnormal cell cycle progression and apotosis as the basis for the syndromic phenotypes of their patients (Tessadori et al., 2017). Some overlaps are seen in our patient and these cases; (1)the syndromic phenotype (especially craniofacial abnormality and intellectual disability) (2) H2A N90 showed sensitivity to methyl-methanesulfonate (MMS) induced double-strand break in DNA in the yeast study (Matsubara et al., 2007) and (3) The structure and protein alignment are not identical between H2A and H4, however, H4 is p.K91 and H2A p.N90 are still in the similar location at the structure of nucleosome. (The H4 p.K91is the last residue of α3 helix The H2A p.N90 locates just after α3 helix, although, the H4p.K91 is not on the surface of nucleosomes, while H2A p.N90 locates on the surface).

Another link between histone and congenital abnormalities is explained by Zaidi et al. (2013). The incidence of *de novo* mutations in 362 severe congenital heart disease cases was analysed by WES and a marked excess of *de novo* mutations in genes involved in the H3K4 methylation and ubiquitylation of H2B K120 (in H3K27 methylation pathway) were identified (Zaidi et al., 2013). In histones, histone N tails are the main loci that those methylation events occur, however, the acidic patch is suggested to regulate the H2B K123 ubiquitylation in yeasts, suggesting the variant in the acidic patch may also disrupt methylation and affect normal heart development (Cucinotta et al., 2015). Given these, the investigation of pathogenicity of the variant in core histone is ongoing.

Chapter 4 INVESTIGATION OF A CAUSAL VARIANT IN A LARGE PEDIGREE WITH CRANIOSYNOSTOSIS BY COMBINING MULTIPLE GENETIC APPROACHES

4.1 INTRODUCTION

This family has multiple individuals affected by coronal or multiple suture synostosis, segregating through at least 4 and probably 5 generations. The family is the largest family in our patient panel and 11 individuals (6 affected, 3 unaffected first degree relatives, 2 spouses) were recruited into the study. Getting information from 7 informative meioses suggested a great advantage to identify a possible candidate. However, to obtain a unique linkage, 10 meioses would be required. The pedigree displays an autosomal dominant inheritance pattern and the craniosynostosis present in affected individuals was severe as most of them needed surgical intervention; hence I have looked for heterozygous variants segregating only in affected individuals and assuming complete penetrance within the pedigree. To find the causal variant, WES, WGS, SNP chip and array CGH were undertaken in combination with segregation analysis.

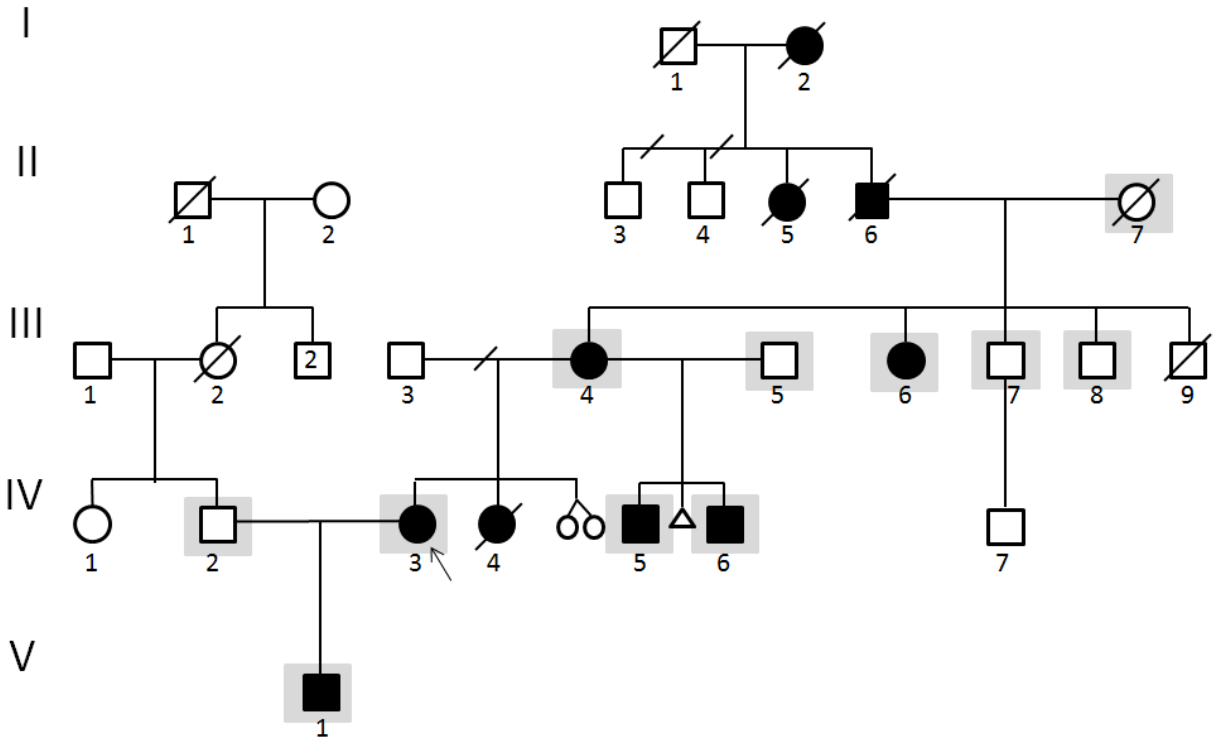


FIGURE 4-1 FAMILY TREE

Squares, male family members: circles, female family members. *Black symbols*, affected subjects; *white symbols*, unaffected subjects. *Grey shades*, DNA obtained. The *arrow* indicates the proband.

4.2 CLINICAL INFORMATION

Initially, the family was referred to the clinical genetics department at Birmingham Maternity Hospital in the mid-1990s and clinical information and DNA samples for II-7, III-4, III-5, III-6, III-7, III-8, III-9 and IV-3, IV-5, IV-6 were collected. The investigation into the genetic cause was conducted by Dr. Willie Reardon at the Institute of Child Health, London between 1993 and 2000. The detailed results were as stated below. In conclusion, a possible candidate was not identified.

- Normal karyotype. In particular, no abnormality of chromosome 7 (1993)
- No common Pfeiffer mutations in *FGFR1* Exon 5. No mutations in several of the exons encoding the extracellular domain of *FGFR2*, but not including transmembrane and intracellular regions of *FGFR* (1995)
- Several markers in *FGFR2, FGFR1* and *MSX1* in samples from the family were analysed and these genes were excluded from possible loci. Three individual markers from the Saethre-Chotzen region on chromosome 7p were tested against the family, but all were uninformative (1996).
- *FGFR1, FGFR2, FGFR3, FGFR4, MSX2, TWIST1, GLI3* and many other candidate loci known to cause definite craniosynostosis were all negative (2000)
- No mutation of *FGFR3* p.Pro250Arg. Linkage analysis also suggested that this gene was not involved (2000).

In 2012, this family was re-recruited to the craniofacial research study in Oxford, using the samples from the 1990s and with the newly obtained research consent. At this time, II-7 was deceased and V-1 was born to the family. V-1 was affected by craniosynostosis and needed ongoing medical care as detailed below. As most of the family members are not currently in contact with any genetics department, the clinical manifestations of the family are mainly based on the clinical notes in the 1990s. However, individuals III-7 and III-8 were visited at home by me, accompanied by Prof. Andrew Wilkie, to obtain further information about the family history and confirm that they were clinically

unaffected. Consent for the use of the sample stored from the deceased individual II-7 was given after following a discussion within the family.

Clinical information on all individuals in the family is summarised in Table 4-1.

4.2.1 I-2, II-5, II-6 AND II-7

II-6 was seen by Dr. Jack Insley at Birmingham Children's Hospital and radiographs were taken on him in the 1970s (radiographs are not available now). II-6 was concluded to have mild craniosynostosis on the basis of the skull radiographs; this was thought to be inherited from II-4's mother (I-2). II-7, who thought to be unaffected, was consented to the research in 1995 and DNA extracted from blood was preserved.

As accurate phenotypic information on II-6 and II-7 were essential to conduct segregation analysis, we re-assessed photographs of II-6 and II-7 offered from the family (Figure 4-3). Although it was difficult to judge craniosynostosis from the pictures, while II-7 showed no abnormal facial features, II-6 had an asymmetrical face, deep set eyes, irregular dental alignment and a low hairline. He was described to have a similar facial appearance to IV-5 by other family members (III-7 and III-8). Based on the observations, we assigned II-6 as affected and II-7 as unaffected subjects, so that DNA of II-7 was used for excluding candidate variants.

For other members of generation I and II, II-3 refused to be enrolled in the research in 1995 and the family had no contact with II-4. II-5 had died of liver cancer at the age of 44. He was recorded as affected, but no further information was obtained from the family and no DNA sample was preserved from him.

4.2.2 III-4 (GRANDMOTHER OF V-1)

At the age of 42, she was seen by Dr. Carol McKeown at Birmingham Maternity Hospital in 1993. Her height was 159.5 cm (30th centile), OFC 47 cm (below 3rd centile). She had bicoronal synostosis, normal forehead, high arched palate and normal ears. Her hands are normal apart from longitudinal

nail ridges. In the feet, she had bilateral partial 2nd and 3rd toe syndactyly and normal halluces. She had normal intelligence and no history of operation.

She had 4 children with craniosynostosis and 2 other pregnancies with one ending in miscarriage of twins after IV-4 and another as an ectopic pregnancy between IV-5 and IV-6.

4.2.3 III-6

She was born with a cloverleaf skull. The birth weight was 3.6 kg (50th centile). She had cerebrospinal fluid (CSF) rhinorrhoea from her left nostril from birth. Her first operation, a bilateral coronal strip craniectomy, was performed at the age of 4 months. After the first procedure the previously noted CSF rhinorrhoea from the left nostril became more profuse, so that 2 weeks later, an anterior fossa exploration was undertaken through a left frontal craniotomy, which found a concavity just behind the left olfactory bulb with an adherent brain, repaired using muscle fascia.

A further craniectomy was undertaken at 10 months old. At this stage, she had two episodes of meningitis and required long-term antibiotic treatment. She developed epilepsy, which was controlled with medication.

At the age of 28 years, she presented with a 2-year history of intermittent left sided rhinorrhoea, deterioration in memory and gait, and headaches. She was noted to have a large skull with cloverleaf shape with no focal neurological deficit. A magnetic resonance imaging (MRI) revealed extensive and severe hydrocephalus with a hindbrain herniation. After a ventriculoperitoneal shunt was inserted, her symptoms improved.

At the age of 32 years, her height was 168 cm (85th centile), OFC 53.5 cm (25th centile). She had a midline scar across the top of the head from her previous craniofacial procedures. She had a narrow sloping forehead, prominent narrow nose, bilateral ptosis and proptosis (Figure 4-3). The ears were normal and she had a normal palate. Her hands had no radial abnormality, but had webbing and ulnar deviation. She had a bilateral partial 2nd and 3rd syndactyly on her feet. Although she had a speech deficit, her intelligence was normal.

4.2.4 III-7, III-8 AND III-9 (UNAFFECTED INDIVIDUALS)

III-7, III-8 and III-9 were described as unaffected subjects in the clinical note made in 1993. To exclude the possibility of the mildly affected clinical case, we undertook a home visit to reassess III-7 and III-8. Their pictures are in Figure 4-3 and Figure 4-4 and the detailed measurements are in Table 4-1. Neither has had any a particular health problem or surgical history. Their intelligence is normal and they had no clinical features to suggest they were affected by craniosynostosis. We concluded that both III-7 and III-8 are unaffected and used their DNA samples to exclude variants as candidates.

The family had no contact with III-9, so we were not able to obtain consent for using a stored DNA sample from III-9. He is now deceased.

4.2.5 IV-3 (MOTHER OF V-1)

She was born at full term with a birth weight of 3.3 kg (25^{th}-50^{th} centile). She had multiple synostosis which resulted in an oxycephalic skull shape with forehead retrusion and exorbitism and associated with bilateral ptosis. She walked at 2 years old, but overall her development was normal. She had no history of craniofacial surgery.

At the age of 21 years, she was seen by a clinical geneticist. Her height was 165.5 cm (70^{th} centile), OFC 50.0 cm (below 3^{rd} centile) with a narrow and sloping forehead, exorbitism and mild ptosis. She had a narrow high palate, prominent narrow nose and slightly prominent ear crura. Her hands and feet were normal, without radial abnormality or webbing (Figure 4-3).

She had coeliac disease, however apart from this, she had normal general health and normal intelligence.

Her phenotype was considered to be the most Saethre-Chotzen syndrome-like in the family, but was milder in comparison with her mother (III-4) and aunt (III-5).

4.2.6 IV-4 (AUNT OF V-1)

She was born with cloverleaf skull and CSF rhinorrhoea was observed from birth. Although her speech was slightly delayed, general development was normal and she attended a normal school. She died of meningitis at the age of 8 years. A clinical photograph obtained from the family (Figure 4-1) suggested that she had a cloverleaf skull, a narrow forehead, prominent narrow nose, bilateral ptosis and proptosis. No DNA sample was available from her.

4.2.7 IV-5 (UNCLE OF V-1)

He was born with cloverleaf skull. His birth weight was 3.32 kg (50th centile), length 54 cm (90th centile) and OFC was 31 cm (below 3rd centile). The first operation for skull release was performed at 6 weeks in 1983. Two lateral craniotomy free bone plates were made in an attempt to reconstitute the coronal, parasagittal and lambdoid sutures.

At 6 months old, surgery for correction of posterior craniosynostosis was undertaken to release the constriction ring around the back and side and head. After the surgeries, his development was good and he had normal intelligence. (Figure 4-3)

4.2.8 IV-6 (UNCLE OF V-1)

Prenatal ultrasound showed a cloverleaf skull deformity. He was born at full term by normal delivery. His birth weight was 2.82 kg (10th centile), length 51cm (50-75th centile), OFC 34 cm (10-25th centile). He had prominent and divergent eyes. Skull radiographs at birth showed that all vault sutures were fused with typical cloverleaf appearance. A CT scan at 6 weeks showed that both lateral ventricles, particularly the temporal horns were mildly dilated. He had surgery for a posterior skull release at 6 weeks old. After the surgery, his developmental progress was good with relatively acceptable cosmetic appearance and normal intelligence. (Figure 4-3)

4.2.9 V-1 (PROBAND)

During the pregnancy, ultrasound scans indicated that he had brachycephaly which became more marked as the pregnancy progressed. The appearance of the brain was normal and there was no hydrocephalus. He was born by elective caesarean section at 39 weeks' gestation, weighing 4.3 kg

(90-95th centile) and there were no significant neonatal problems. He had a high palate and had difficulty in breastfeeding but managed to be bottle fed. He smiled at 6 weeks. He passed his neonatal hearing screen.

At the age of 14 weeks, his height was 59.1 cm (9-25th centile) and weight was 6.22 kg (25-50th centile). He had a complex multisuture craniosynostosis with an open anterior fontanelle, a turricephalic head shape with ridging over his metopic suture and flattening of the occiput similar to his mother. He had mild exorbitism and an intact palate which was high and narrow anteriorly. His ears were small with thickened helices. He had a normal neck, chest shape, nipples, heart sound and abdominal examination. He had a small penis mainly due to a suprapubic fat pad and testes were descended. His limbs were normal with deep creases on the soles. He had some mild eczema and a single café au lait patch on his right abdomen. His spine looked normal. A CT scan at 1 year old showed the presence of multisuture synostosis with severe copper beaten skull (Figure 4-2). The basal cisterns were small and there were features of early hydrocephalus. Ophthalmologic examination showed bilateral papilledema. Therefore, biaxial wedge osteotomies with the insertion of 4 springs for asymmetric posterior left sided vault expansion were undertaken at the age of 1 year. At the age of 2 years and 7 months, fronto-orbital remodelling and removal of springs were undertaken. By the age of 5 years and 6 months, he was assessed as autistic, with delayed speech and language, social, early learning and self-care skills. His motor development was within the normal range.

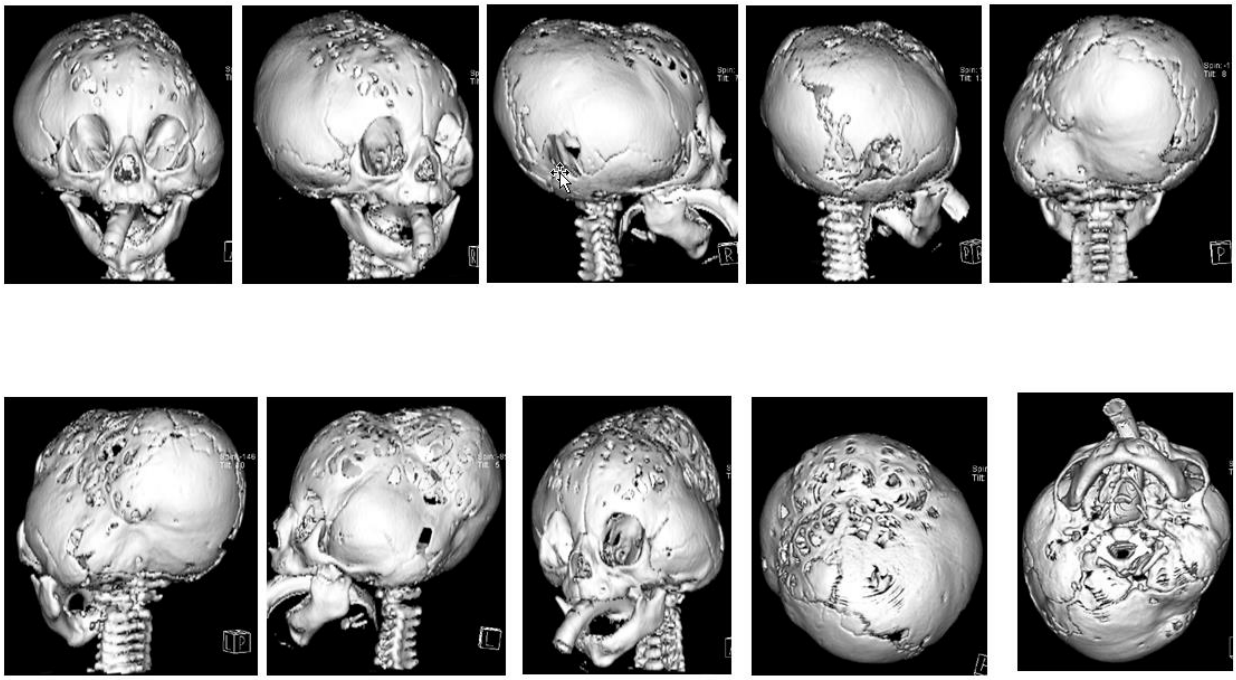


FIGURE 4-2 3D CT IMAGES OF V-1

CT was performed at the age of 1 year. Multisuture craniosynostosis (sagittal, left coronal and lambdoid synostosis) and asymmetrically shaped skull with severe copper beating findings are observed. The right coronal suture is partially fused. An intubation tube is inserted into the oral cavity.

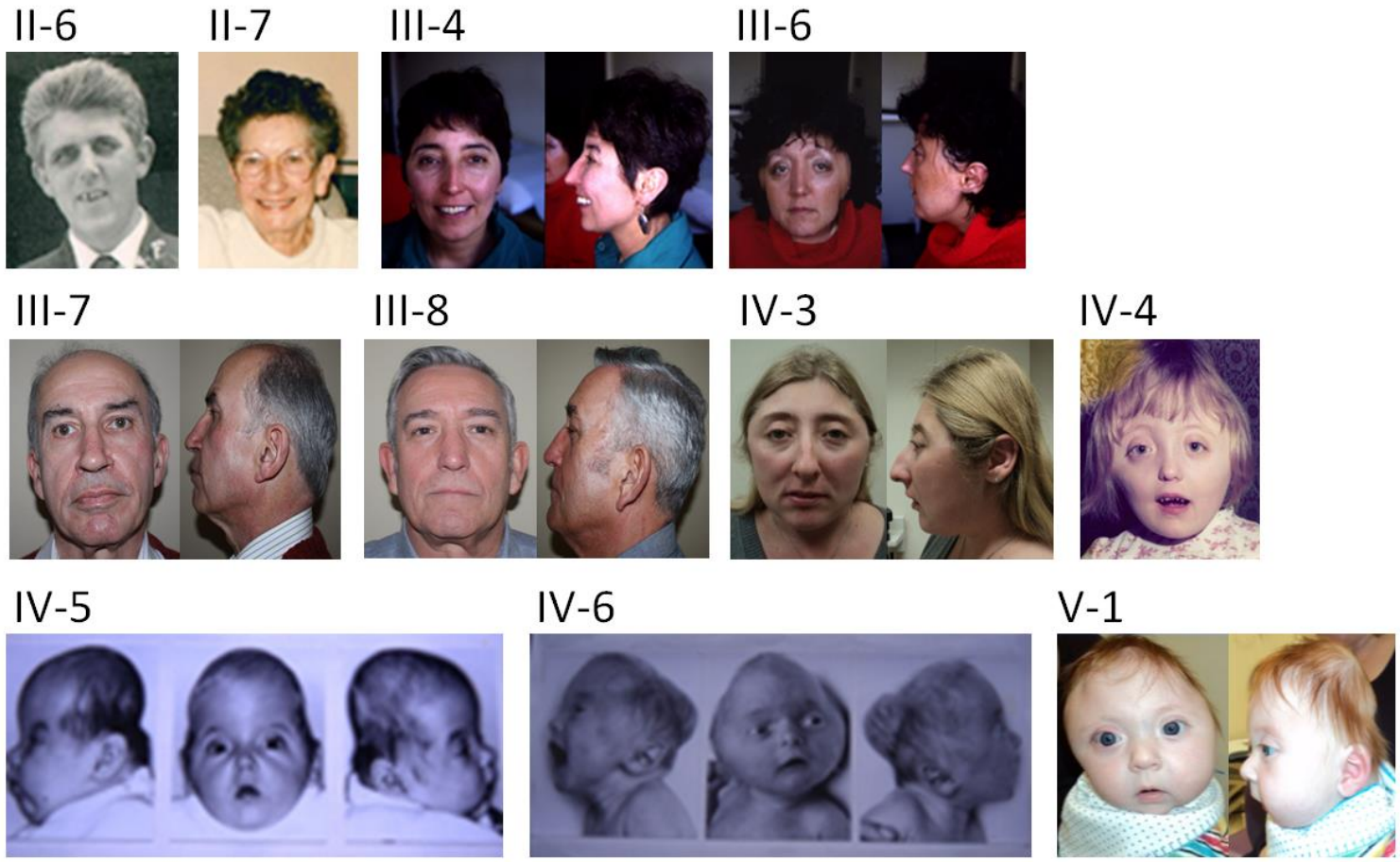


FIGURE 4-3 CLINICAL PHOTOS – 1

Facial appearance of affected and unaffected family members. II-6, III-4, III-6, IV-3, IV-4 IV-5, IV-6 and V-1 are affected individuals and II-7, III-7 and III-8 are unaffected individuals.

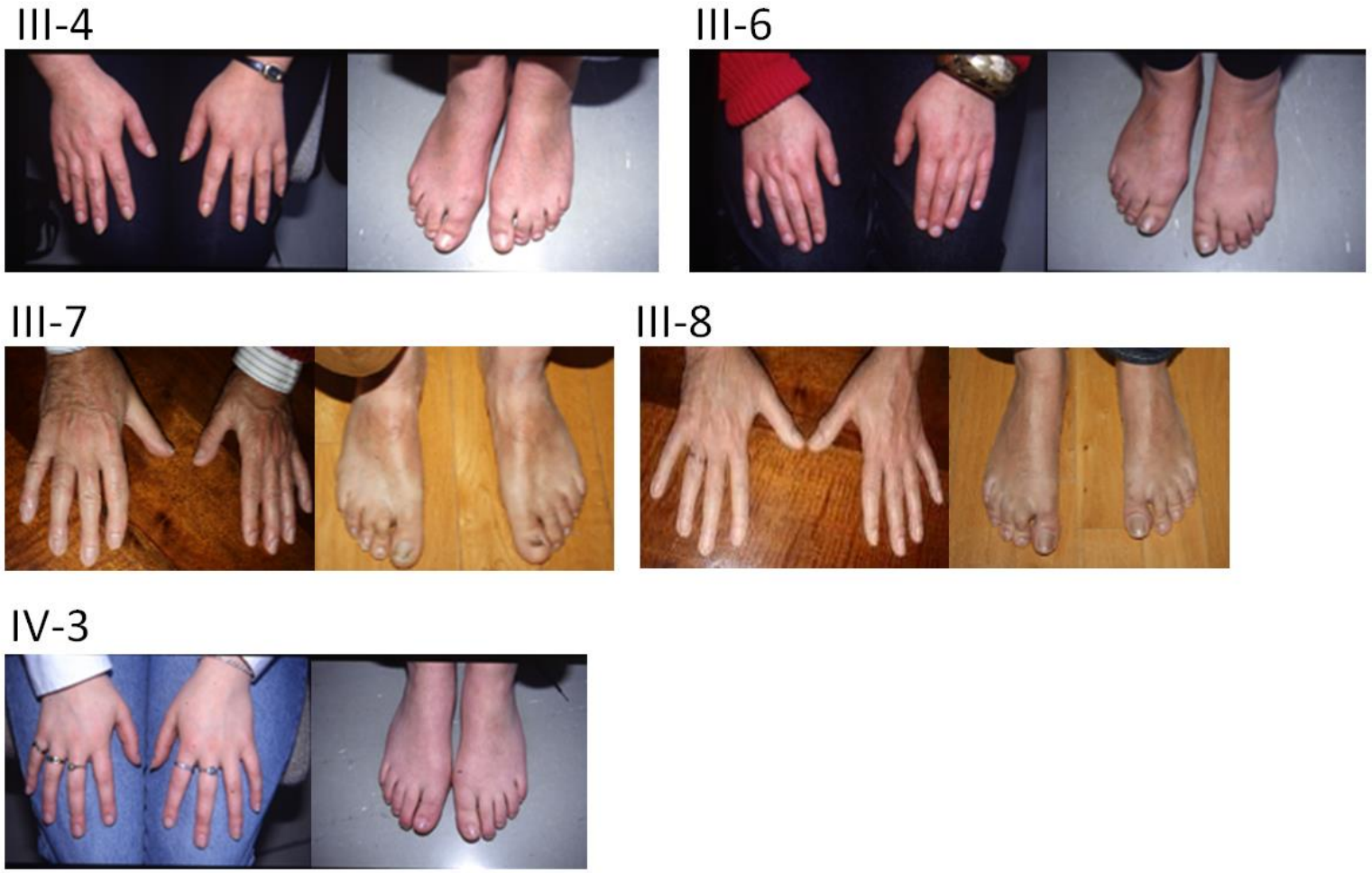


FIGURE 4-4 CLINICAL PHOTOS - 2

Hands and feet from affected and unaffected members. III-4, III-6 and IV-3 are affected individuals and III-7 and III-8 are unaffected individuals.

TABLE 4-1 SUMMARY OF CLINICAL INFORMATION

	V-1	IV-3	IV-4	IV-5	IV-6
			died when 8 years old of meningitis		
craniosynostosis	asymmetric turricephaly	Saethre-Chotzen like brachyturricephaly	cloverleaf skull and rhinorrhoea	cloverleaf skull	cloverleaf skull
WES or WGS	WES, WGS	WES, WGS		WGS	WES
DNA	○	○	×	○	○
the prenatal US					cloverleaf skull
delivery					full term
OFC at birth (cm)				31 (below 3rd centile)	34 (10-25th centile)
BW at birth (kg)		3.3 34 (25-50th centile)		3.32 (50th centile)	2.82 (10th centile)
Length at birth (cm)				54 (90th centile)	51 (50-75th centile)
Operation	yes	no	no	yes	yes
Epilepsy	no	no	yes	no	no
Meningitis	no	no	yes	no	no
Age		23			
Ht (cm)		165.5 (70th centile)			
OFC (cm)		50 (below 3rd centile)			
ICD (cm) inner canthal distance		3.3 (+1 SD)			
OCD (cm) outer canthal distance		10 (-1 SD)			
IPD (cm) interpupillary distance		6.2 (75th centile)			
Head and Face					
High arched palate	yes	yes			
Bilateral ptosis			yes		
Proptosis	yes	yes	yes		
Ears size (cm)		6 (Mean)			
shape		normal			
Hands		normal			
Hand (cm)		19 (55th centile)			
Palm (cm)		10.5 (40th centile)			
Feet		normal			
Speech		normal	talked late		
Development			walked at 11/12 months		
Intelligence		normal	normal	normal	normal

	III-4	III-6	II-6	II-7	III-7	III-8
			deceased	deceased		
craniosynostosis	bilateral coronal synostosis	cloverleaf skull and rhinorrhoea	mildly affected on the basis of skull X-ray	not affected	not affected	not affected
WES or WGS		WES				
DNA	○	○	×	○	○	○
the prenatal US						
delivery						
OFC at birth (cm)						
BW at birth (kg)		3.6 (50th centile)				
Length at birth (cm)						
Operation	no	yes	no	no	no	no
epilepsy	no	yes	na	na	no	no
meningitis	no	yes	na	na	no	no
age	44	34			61	59
Ht (cm)	159.5 (30th centile)	168 (85th centile)			172.7 (25th centile)	177.8 (50th centile)
OFC (cm)	47 (below 3 centile)	53.5 (25th centile)			58.0 (98th centile)	57.0 (90th centile)
ICD (cm) inner canthal distance	3.4 (+1SD)	3.8 (+3.4SD)			3.5 (+1 SD)	3(Mean)
OCD (cm) outer canthal distance	10 (-1SD)	10.5 (mean)			9.2 (-1.5 SD)	8.7 (-2 SD)
IPD (cm) interpupillary distance	7 (over 97th centile)	7 (over 97th centile)			6.2 (75th centile)	6.3 (75th centile)
Head and Face						
High arched palate	yes	yes			no	no
Bilateral ptosis	yes	yes			no	no
proptosis	yes	yes	deep set		no	no
Ears size (cm)	6.4 (+1.0 SD)	7.1 (+2.75 SD)			7.2 cm (+2.75 SD)	7.5 cm (+2.75 SD)
shape	normal	normal			normal	normal
Hands	longitudinal nail ridges	webbing, deviation			normal	normal
Hand (cm)	17.3 (20th centile)	18.7 (50th centile)			16.6 (15th centile)	17.3 (20th centile)
Palm (cm)	9.6 (3rd centile)	11 (70th centile)				
Feet	partial 2-3syndactyly bilateral	partial 2-3syndactyly bilateral			normal	normal
Speech		speech difficulty			normal	normal
Development						
Intelligence	normal	normal			normal	normal

US, Ultrasound sonography. OFC, occipitofrontal circumference. BW, body weight. Ht, height.

4.3 RESULTS

4.3.1 WES OF IV-3

Initial analysis of this family started with WES of individual IV-3, anticipating that this might yield an obvious pathogenic mutation. The details of exome capture kit and software used for alignment and variant calling are listed in Table 2-1.

The variant list started with 1742 variants, which were filtered out if the total number of variant reads was less than 5 or variants were previously identified in the ESP or WGS500 database. Subsequently, gene function and the position of the variants (whether it locates in the domains) were evaluated to prioritise the variants.

Twenty-eight apparently heterozygous variants remained in the list after filtering including 21 nonsynonymous, 2 stop gain, 2 frameshift and 2 inframe variants and single splicing and unknown function variants.

Firstly, a nonsynonymous variant in *BMPR1B* (the bone morphogenetic protein receptor type IB) c.995A>T, p.D332V appeared to be a promising candidate. The deleterious score of the variant was 6 and it was not reported in ExAC or gnomAD. The p.D332 is a highly conserved amino acid in vertebrate species. A homozygous mutation in *BMPR1B* is the cause of acromesomelic chondrodysplasia with genital anomalies (AMDGA) [MIM:609441]. Acromesomelic chondrodysplasias are rare hereditary skeletal disorders characterised by short stature, very short limbs, and hand/foot malformations (Demirhan et al., 2005; Graul-Neumann et al., 2014). In addition, heterozygous missense mutations in *BMPR1B* cause brachydactyly type A2 (BDA2) [MIM:112600]. The study demonstrated that these mutations function as dominant negative in vitro and in vivo (Lehmann et al., 2003). However, segregation analysis of *BMPR1B* in V-1, affected son of IV-3, revealed the variant was not inherited in V-1, excluding the variant as a candidate. In addition, the variant was inherited from the unaffected grandmother, II-2 (Figure 4-5).

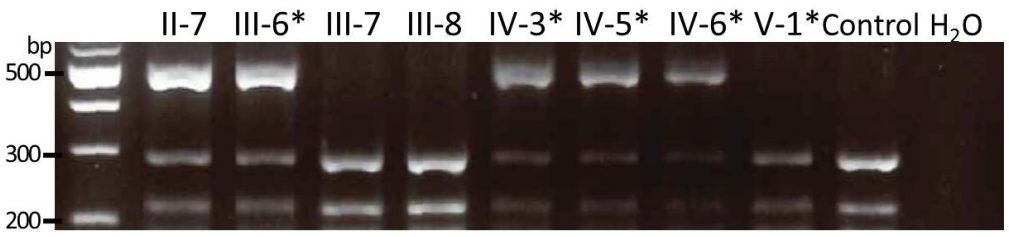


FIGURE 4-5 SEGREGATION OF THE *BMPR1B* C.995A>T, P.D332V IN THE FAMILY.

Digest of PCR product with BstYI. Wild type allele; 289 and 216 bp. Mutant allele; 505 bp.
* shows affected individuals.

Subsequently, the segregation of the other 27 variants was examined by dideoxy sequencing of III-6 (unaffected), III-7 (unaffected), IV-6 (affected) and V-1 (affected). The variants were excluded except a single nonsynonymous variant, *RNF17* c.173T>C p.L58P, on the list.

However, the *RNF17* p.L58 was not a highly conserved residue and p.L58P is reported with a frequency 0.0002234 in gnomAD, leading to the decision to carry out WES of further samples from affected individuals. (The variant was confirmed to be not inherited by IV-5 (affected) and inherited by III-7 (unaffected) by the later analysis.

4.3.2 WES OF III-6, IV-6 AND V-1

To proceed to the further analysis, WES of 3 other affected individuals (III-6, IV-6 and V-1) was undertaken. The previous WES of IV-3 was run with the TruSeq Exome Enrichment Kit, however, the Agilent SureSelect Human All Exon V5 was used for WES of III-6, IV-5 and V-1, so that covered regions were not exactly identical to those in the original sample IV-3.

From the bioinformatics pipeline, 6872 variants were listed from the 4 exome sequences.[8] The variants were filtered shown in Table 4-2.

[8] Filtering condition; ESP 0.01, 1000G genomes project 0.01, Max number of variants per gene 5

TABLE 4-2 DETAILS OF THE NUMBER OF VARIANTS DETECTED BY WES THE FAMILY IN CHAPTER 4

total variants in III-6, IV-3, IV-6 and V-1		6872
Variants shared in 4 of them or III-6, IV-6 and V-1		632
variants after excluding common variants and mapping artefact		109
Variants in coding regions		30
	stop gains	2
	nonsynonymous	20
	Synonymous	8
variants not reported in 1000G, ESP		16
	stop gains	2
	nonsynonymous	10
	Synonymous	4

For the remaining 16 variants, the genotypes of additional samples from the affected individual (IV-5) and unaffected individuals (II-7, III-7 and III-8) was examined by dideoxy sequencing to find the variants that segregated only in affected individuals (Table 4-3).

TABLE 4-3 SHARED VARIANTS FROM WES OF IV-3, III-6, IV-6 AND V-1 AND THEIR GENOTYPE IN OTHER FAMILY MEMBERS (GRCH37)

Gene	Chr	Position	DS	change		Exonic_function	IV-3*	III-6*	IV-6*	V-1*	IV-5*	III-7	III-8	II-7	frequency in gnomAD	Note
ALG6	chr1	63876987	3	c.665G>A	p.G222D	nonsynonymous	GA	GA	GA	GA	GA	GG	GA	GG	107/276522	
FAM82A1	chr2	38156623	0	c.203T>C	p.F68S	nonsynonymous	TC	TC	TC	TC	TC	TC	na	na	4/149320	
NRXN1	**chr2**	**50724758**	**0**	**c.259C>T**	**p.N864N**	**synonymous**	**CT**	**CT**	**CT**	**CT**	**CT**	**CC**	**CC**	**CC**	**not reported**	
ECEL1	**chr2**	**233351209**	**2**	**c.155T>C**	**p.L52P**	**nonsynonymous**	**TC**	**TC**	**TC**	**TC**	**TC**	**TT**	**TT**	**TT**	**202/51602**	
CPVL	chr7	29152356	3	c.252C>A	p.Y84X	stopgain	CA	CA	CA	CA	CA	CA	na	na	6/246194	
GDF6	chr8	97157413	1	c.746C>A	p.A249E	nonsynonymous	CA	CA	CA	CA	CA	CC	na	CA	245/126108	
KANK1	chr9	744589	0	c.3522G>A	p.P1174P	synonymous	GA	GA	GA	GA	GG	GA	na	na	28/277216	
KCNV2	chr9	2718341	3	c.602T>A	p.I201N	nonsynonymous	na	TA	TA	TA	TT	TA	TA	TA	4/220448	
CDC37L1	chr9	4688573	4	c.475A>G	p.M159V	nonsynonymous	AG	AG	AG	AG	na	AG	na	na	3/193426	
GRK5	chr10	121203141	0	c.1143C>T	p.G381G	synonymous	CT	CT	CT	CT	CT	CC	CT	CC	3/245446	
DMBT1	chr10	124390666	0	c.3944C>T	p.S1315F	nonsynonymous	CT	CT	CT	CT	CT	CC	CT	CC	6/246230	
GJA3	chr13	20717217	5	c.211C>G	p.P71A	nonsynonymous	CG	CG	CG	CG	CC	CG	na	na	not reported	
IFT88	chr13	21218984	0	c.1836A>G	p.S612S	synonymous	AG	AG	AG	AG	na	AG	na	na	264/235864	
RNF17	chr13	25341452	2	c.173T>C	p.L58P	nonsynonymous	TC	TC	TC	TC	TT	TC	na	na	55/246164	
PLCB2	chr15	40590557	3	c.1022C>A	p.S341X	stopgain	CA	CA	CA	CA	na	CA	na	na	not reported	pLI=0.01
C17orf85	chr17	3721680	2	c.1187G>A	p.S396N	nonsynonymous	GA	GA	GA	GA	GG	na	na	GA	4/246232	

DS; deleterious score. na; genotype was not examined. Two variants in bold are segregating completely according to phenotypic status in the tested samples. Shades indicate discordances in the segregation with the phenotype. * indicates affected subjects.

Two variants segregate according to phenotypic status perfectly; *ECEL1* (c.155T>C, p.L52P) and *NRXN1* (c.2592C>T, p.N864N).

ECEL1 is the causal gene for distal arthrogryposis type 5D, which is inherited as an autosomal recessive trait (Dieterich et al., 2013; McMillin et al., 2013), but *ECEL1* p.L52P presented in one of our solved proband's exome data and was reported in gnomAD with a frequency of 0.00391, suggesting the change is unlikely to be pathogenic in the heterozygous state.

The synonymous variant in *NRXN1* (Neurexin 1) has not been reported in gnomAD. Neurexins are cell-surface molecules that bind neuroligins to form heterophilic, Ca^{2+}-dependent complexes at central synapses (Reissner et al., 2008). CNVs, mainly deletions, in the *NRXN1* region cause learning disability, autism spectrum disorders or seizures (Kim et al., 2008; Schaaf et al., 2012; Zahir et al., 2008). Splice effect caused by the synonymous variant was examined in multiple *in silico* prediction tools by using Alamut® Visual (http://www.interactive-biosoftware.com/doc/alamut-visual/2.6/splicing.html), which is a prediction module integrating seven prediction methods;(1. Splice signal detection: SpliceSiteFinder (Zhang, 1998), MaxEntScan (Yeo and Burge, 2004), SSPNN, GeneSplicer (Pertea et al., 2001) and Human Splicing Finder (HSF) splice (Desmet et al., 2009), 2. Exonic splicing enhancers (ESE) binding site detection: ESEFinder (Cartegni et al., 2003) and RESCUE-ESE(Fairbrother et al., 2004)). Of these 7 tools, only RESCUE-ESE (http://genes.mit.edu/cgi-bin/rescue-ese_new.pl) predicted a slight change between wild type and the variant as the variant sequence would delete 3 interaction sites with ESEs, but create a new single interaction.

Considering that the *NRXN1* c.2592C>T, p.N864N is the unique variant to the family that segregates perfectly with the phenotype and is predicted that it might disrupt some potential exon splice enhancer sequences, I attempted to confirm splicing effects of the variant in *vivo* by examining the expression of cDNA products from the transcript obtained from an affected individual's (IV-3) cell. RNA extracted from PAX gene whole blood and lymphoblastoid cell lines were used to synthesise cDNA, however, transcripts of *NRXN1* were not detected in either control or patient samples. The

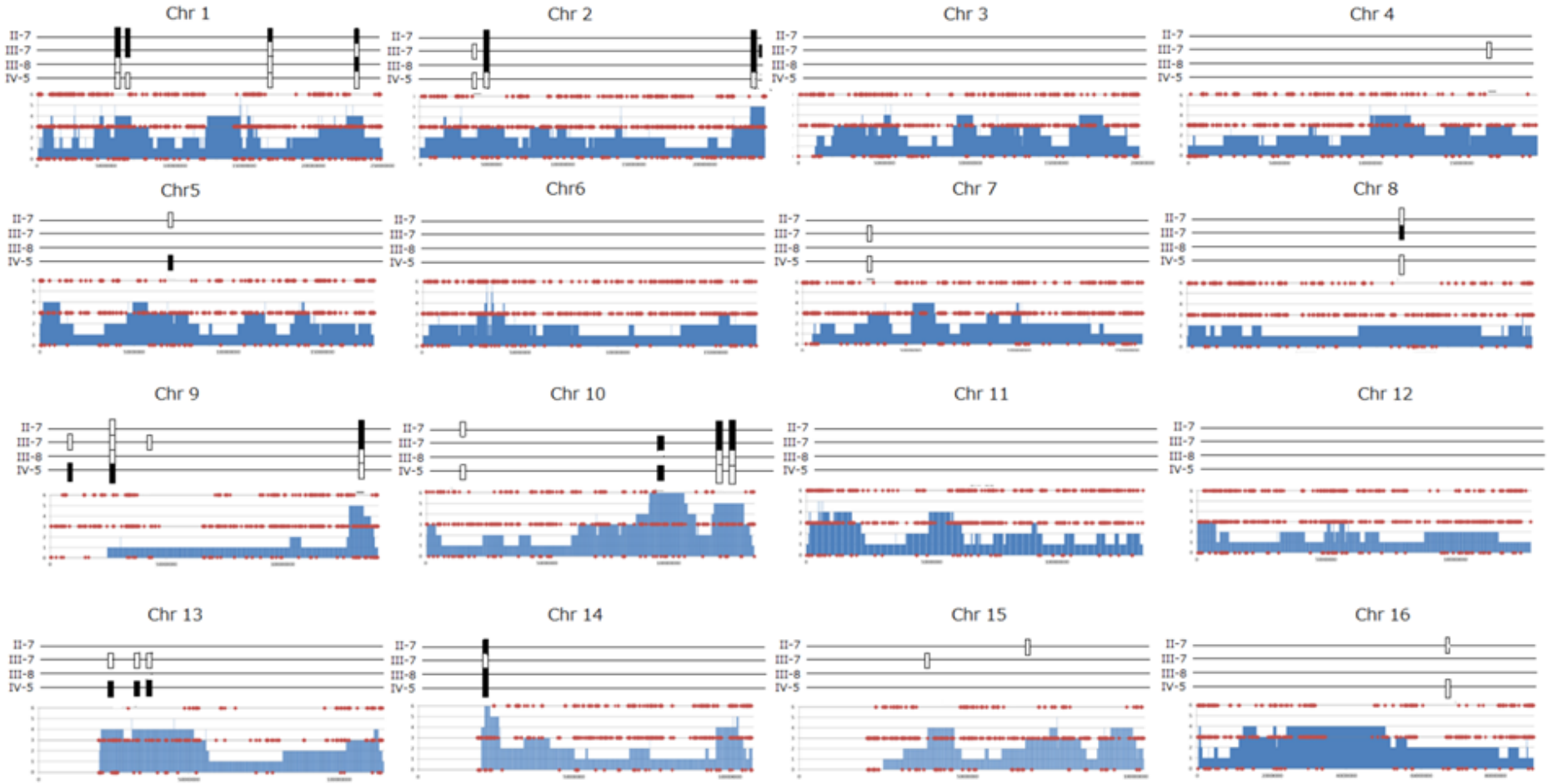
major expression of *NRXN1* is in the brain and Yasuda et al. (2011) also reported that expression levels of *NRXN1* mRNA could not be quantitated in lymphoblastoid cells (Yasuda et al., 2011). Gene expression data in GTEx also show that main tissues with high *NRXN1*expression are brain and nerve, while expressions in lymphoblastoid and fibroblasts are low.

4.3.3 SEGREGATION ANALYSIS BY MENDELSCAN

As 4 of the affected members from the family had been sequenced, segregation analysis could help to narrow down the disease region. To undertake this, I used MendelScan v.1.2.1 (Koboldt et al., 2014) (http://gmt.genome.wustl.edu/packages/mendelscan/), which is a software package for analysing the sequencing data in family studies of inherited disease. MendelScan has mainly 3 features; variant calling and prioritisation, rare RHRO and SIBD. Therefore, I decided to try RHRO and SIBD to obtain more detailed information about the segregating regions in the family.

Exome sequence data from III-6, IV-3, IV-6 and V-1 were used as input. The output from the MendelScan analysis was combined with dideoxy sequencing of other family members, IV-6 (affected) and III-7, III-8 and II-7 (unaffected) and results are shown in Figure 4-6.

From this analysis, two regions of chromosome 2 were found to segregate with the phenotype in the family. This is consistent with the results shown in Table 4-3 in which *NRXN1* and *ECEL1*, both located in different regions on chr2.


Chr 1
Chr 2
Chr 3
Chr 4
Chr5
Chr6
Chr 7
Chr 8
Chr 9
Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
Chr 16
II-7
III-7
III-8
IV-5

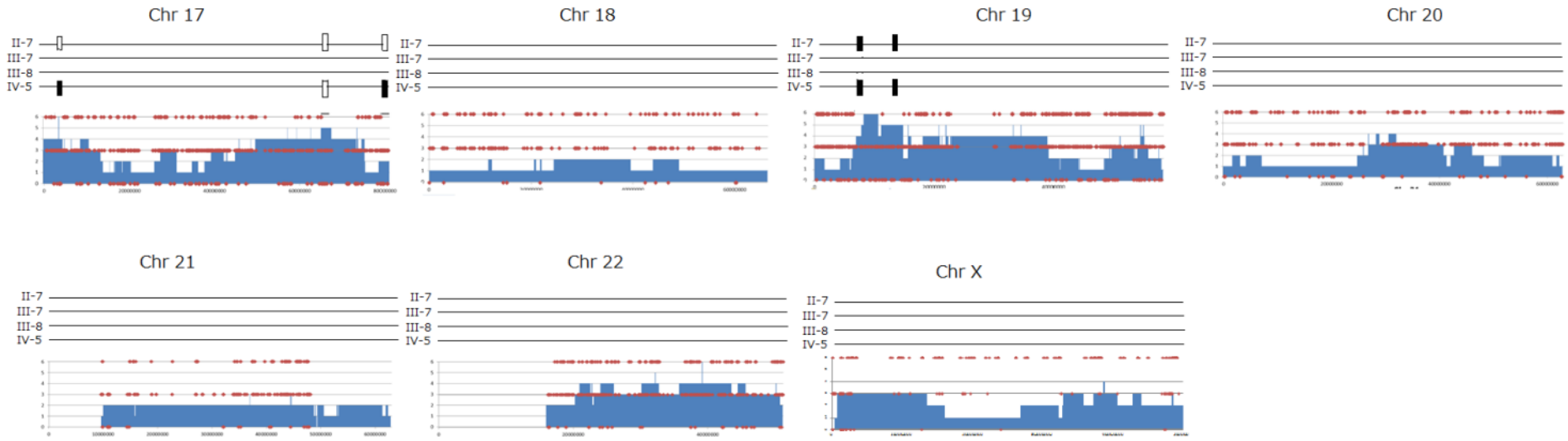


FIGURE 4-6 SEGREGATION ANALYSIS BY INDIVIDUAL GENOTYPING AND MENDELSCAN

For each chromosome, the lower half of the diagram shows the result of Rare Heterozygote Rule Out (red markers) and Shared Identity-by-Descent (blue bar graph) by MendelScan which was run with exome sequencing data of IV-3, V-1, IV-6 and III-6 is shown at the corresponding chromosome position. The red marks on the top row are variants with homozygous differences between affected pairs (rule outs). Middle rows are variants shared between 2 or 3 individuals. Bottom rows are variants shared by all.
Black and white rectangles on the upper half of the diagram show the variants examined by dideoxy sequencing of II-7, III-7, III-8 (unaffected individuals) and IV-5 (affected individual). White rectangles show the variants which segregated to each individual and black rectangles show the variants which didn't segregate. As II-7, III-7 and III-8 are unaffected individuals and IV-5 is affected, the combination of the white rectangle (the variants which segregated) in IV-5 and black rectangles (the variants which didn't segregate) in II-7, III-7 and III-8 show the loci which were not excluded. The loci examined are listed in Appendix 2.

4.3.4 SEGREGATION ANALYSIS BY RARE VARIANTS FROM WES OF IV-3, III-6, IV-6 AND V-1 COMBINED WITH GENOTYPING OF ADDITIONAL MEMBERS

To obtain precise location of segregating regions in the family, all the variants with frequency < 0.1 were listed up by chromosomal order from WES of IV-3, V-1, IV-6 and III-6. For the regions which have rare variants shared by III-6, IV-6 and V-1, the genotypes of other family members (II-7, III-7, III-8 and IV-5) were examined by dideoxy sequencing to identify the segregating regions. Following this, the genotype of rare variants upstream and downstream of the segregating regions were examined to find the position of recombination to further narrow down the segregating regions (The variant list is in Appendix 3).

The Table 4-4 shows the detailed position of the detected regions which confirmed segregation in the family or regions with no informative SNPs to exclude the segregation. The first two regions listed in Table 4-4, chr2:50,724,758-55,481,137 and chr2:224,824,820-237,374,374 were confirmed as segregating in the family from the analysis. These regions corresponded to similar regions identified by MendelScan.

Table 4-4 The end positions confirmed as segregating or not segregating

GRCh37				
chromosome	**5' position (Start)**		**3' position (end)**	
	excluded	**confirmed segregation**	**confirmed segregation**	**excluded**
chr2	**46,842,119** Inherited in III-8 (unaffected)	**50,724,758**	**55,481,137**	**61,007,782** HRO between IV-6 and V-1
chr2	**220,413,658** HRO between III-6, V-1 and IV-6	**224,824,820**	**237,374,374**	**240,900,607** HRO between IV-6 and V-1
chr1	**63,876,987** segregates in III-7 (unaffected)			**66,829,262** HRO between IV-6 and V-1
chr1	**237,958,564** segregates in III-8 (unaffected)			**244,773,803** HRO between III-6 and IV-6
chr2	**207,003,310** segregates in II-7 (unaffected)			**208,989,037** HRO between III-6 and IV-6
chr10	**28,490,983** segregates in II-7, III-8 (unaffected)			**31,134,557** HRO between V-1,IV-3 and V-1
chr10	**94,409,800** HRO between IV-6 and V-1			**98,281,892** segregates in III-7 (unaffected)
chr10	**124,390,666** segregates in III-7 (unaffected)			**131,906,096** HRO between IV-6 and V-1
chr15	**34,060,740** HRO between III-6, V-1 and IV-6			**40,331,182** Inherited in III-7 and III-8 (unaffected) not inherited in IV-5
chr15	**66,170,128** Inherited in II-7 (unaffected) not inherited in III-4 (affected)			**71,184,278** HRO between IV-6 and V-1

HRO, Homozygous Rule Out.

Not to miss any regions which could possibly segregate with the phenotype in the family, segregating regions were listed as shown in Table 4-5. The first 2 regions are confirmed to be perfectly segregating. For the other eight regions, the pattern of recombination between the two excluded loci could not be resolved, so it remains possible that some of these contain regions that segregate with the phenotype, although not strongly supported as for the two regions above (Table 4-5).

In further investigations, the focus was on the 10 regions listed in Table 4-5, particularly the definite segregating regions on chr2.

TABLE 4-5 SEGREGATING REGIONS

GRCh37			GRCh38		
chromosome	**start**	**end**	**chromosome**	**start**	**end**
chr2	**46,842,120**	**61,007,781**	**chr2**	**46,614,981**	**60,780,646**
chr2	**220,413,659**	**240,900,606**	**chr2**	**219,548,937**	**239,961,189**
chr1	63,876,988	66,829,261	chr1	63,411,315	66,363,578
chr1	237,958,565	244,773,802	chr1	237,795,265	244,610,500
chr2	207,003,311	208,989,036	chr2	206,138,587	208,124,312
chr10	28,490,984	31,134,556	chr10	28,202,055	30,845,627
chr10	94,409,801	98,281,891	chr10	92,650,044	96,522,134
chr10	124,390,667	131,906,095	chr10	122,631,151	130,107,831
chr15	34,060,741	40,331,181	chr15	33,768,540	40,038,980
chr15	66,170,129	71,184,277	chr15	65,877,791	70,891,938

Two regions on chromosome 2 on top of the list (in bold) are confirmed to be segregating while other regions are considered to be possibly segregating.

4.3.5 WGS OF IV-5

As the analysis had not identified a clearly pathogenic mutation in any segregating region, WGS was performed on one sample from an affected individual, IV-5, not previously analysed by exome sequencing. I hypothesised that the causative variant may have been poorly covered by the exome sequence or might be in non-coding sequence.

WGS of IV-5 was carried out by Complete Genomics. ANNOVAR was used for annotation of the variant list which was combined with the results of WES of III-6, IV-3, IV-6 and V-1 by the in-house bioinformatician.

In total 247 identified variants are in the segregating regions, including exonic, 3' and 5' UTR, intronic (mainly within 100 bp from the exon-intron boundary), intergenic and up-downstream variants, in which 18 variants were in coding regions or within 10 bp of exon-intron boundaries (Table 4-6). All 18 variants were either not segregating in all of III-6, IV-3, IV-6 or V-1 (all affected individuals, the genotypes were confirmed from WES data) or reported in either WGS500, dbSNP or gnomAD, except for the synonymous variant in *NRXN1* c.2592C>T, p.N864N at Chr2:50,724,758 which was mentioned previously (section 4.3.1).

Subsequently, all shared variants in affected individuals (III-6, IV-3, IV-6, V-1 and IV-5) not previously reported in WGS500, ESP, 1000G or dbSNP in segregating regions were examined. Three variants including *NRXN1* c.2592C>T, p.N864N were identified (Table 4-7).

The C>G at chr2:54,586,036 in an intron of *SPTBN1* is reported in gnomAD with the frequency of 2/188332, and predicted no splice effect by SSPNN. The A>G at chr15:37,100,799 in *CSNK1A1P* which locates within the intronic region of a ncRNA gene, has no prediction of allelic splicing by SSPNN and di-deoxy sequencing demonstrated that the variant was inherited by the unaffected individual III-8.

TABLE 4-6 EXONIC VARIANT LIST FROM WHOLE GENOME SEQUENCING OF IV-5 (GRCH37)

Gene	Chr	Start	Ref	Alt	Function.	Codon Change	Amino acid change	DS[1]	HM[2]	IV-3	V-1	III-6	IV-6	dbSNP	gnomAD[4]
FMN2	chr 1	240,371,106	C	T	synonymous	c.2994C>T	p.P998P	0	Y	C/T	na	na	C/T	rs11586155	26055/98828
FMN2	chr 1	240,371,145	TCCCCCT	TCCCCCA	nonframeshift substitution	c.3033_3039 TCCCCCA		0		TCCCCCT / TCCCCCT	na	na	TCCCCCT / TCCCCCT	rs71646892	1270/26472
NRXN1	**chr 2**	**50,724,758**	**G**	**A**	**synonymous**	**c.2592C>T**	**p.N864N**	**0**		**G/A**	**G/A**	**G/A**	**G/A**		**Not reported**
STK11IP	chr 2	220,473,923	C	T	synonymous	c.1947C>T	p.H649H	0		C/T	C/T	C/C	C/C	rs185836682	57/255398
RHBDD1	chr 2	227,778,999	C	T	nonsynonymous	c.788C>T	p.T263M	0		C/C	C/C	C/C	C/T	rs34134244	1046/276590
MFF	chr 2	228,194,480	AG	TT	nonframeshift substitution	c.19_20TT		0		TT/TT	TT/TT	AG/TT	AG/TT	rs3211097	86263/272028,68201/271918
AGFG1	chr 2	228,356,365	T	-	Intronic +9			0		-/-	-/-	-/-	-/-	rs11307362	50599/97610
SP100	chr 2	231,328,813	T	C	synonymous	c.1014T>C	p.N338N	0		C/T	C/T	C/T	C/T	rs140194556	570/276928
SP100	chr 2	231,363,253	A	-	Intronic *3			0	Y	-/-	A/-	-/-	-/-	rs34979917	45022/74040
ARMC9	chr 2	232,087,474	AT	GA	nonframeshift substitution	c.538_539 GA		0	Y	GA/GA	GA/GA	GA/GA	GA/GA	rs162450, rs162451	273915/273938,273886/273918
GIGYF2	chr 2	233,568,121	T	TT	Intronic -10			0		T/T	na	na	na		Not reported
UGT1A7	chr 2	234,590,974	CG	AA	nonframeshift substitution	c.391_392 AA		0		CG/GA	CG/GA	CG/GA	CG/GA	rs17863778, rs17868324	158853/267870,758859/268032
UGT1A4	chr 2	234,627,937	TGG	CGG	nonframeshift substitution	c.471_473 CGG		0		CGG/CGG	CGG/CGG	CGG/CGG	CGG/CGG	rs2011404	240062/277082
RAB17	chr 2	238,485,990	G	C	nonsynonymous	c.345C>G	p.D115E	1		G/G	G/G	G/G	G/G	rs140729649	752/27692
SVIL	Chr 10	29,784,072	G	C	nonsynonymous	c.2425C>G	p.P809A	3	Y	G/G	G/C	G/G	G/G	rs2368406	194296/253508

ACADSB	Chr 10	124,812,686	T	-	Intronic +10				Y	-/-	-/-	-/-	-/-	rs11307362	50599/97610
CHST15	Chr 10	125,780,757	GGG	GC	frameshift substitution	c.1360_1362 GC		6		GGG/ GC	GGG/ GC	GGG/ GC	GGG/ GC	rs5788645 rs4072194	[3]
FAM53B	Chr 10	126,311,908	G	A	nonsynonymous	c.1172C>T	p.A391V	0		G/G	G/G	G/G	G/A	rs147886380	852/270172

IV-3, V-1, III-6 and IV-6 are all affected individuals. na; genotypes were not available

1 DS; deleterious score, 2 HM: Homozygous Mutations - Y indicates variants were detected as homozygous. 3 No frequency in gnomAD, but G=0.1929/966 in 1000G. 4 Some of the variants were presented with high frequency in gnomAD. Most of these were deletion or insertion and the annotated position was not exactly the same, therefore failed to filter out in the previous steps.

TABLE 4-7 SHARED VARIANTS IN III-6, IV-3, IV-6, V-1 AND IV-5 FROM WGS (GRCH 37)

Gene	**Chr**	**Start**	**Ref**	**Alt**	**Function**	**Details**	II-7	**III-7**	**III-8**	**gnomAD**
NRXN1	chr2	50,724,758	G	A	exonic	synonymous, c.2712C>T, p.N904N	GG	GG	GG	not reported
SPTBN1	chr2	54,856,036	C	G	intronic	distance to exon -33	GG	GG	GG	2/188332
CSNK1A1P1	chr15	37,100,799	A	C	ncRNA_ intronic		AA	AA	AC	not reported

II-7, III-7 and III-8 are unaffected individuals.

4.3.6 INTRONIC VARIANTS ANALYSIS IN WGS DATA

In the previous search, the variants in introns were not fully covered, so that intronic variants in WGS of IV-5 were investigated further. A list of intronic and intergenic variants with splice prediction scores of wild type and the variant by SSPNN were generated for the 2 segregating regions on chromosome 2. The 2856 variants on the list were filtered as shown in Table 4-8, leaving 7 variants remaining. All 7 variants locate more than 1 kb away from exons, so that WES of III-6, IV-3, IV-6 and V-1 didn't capture the regions. For these 7 variants, the segregation in the family was examined using DNAs from IV-5 (to validate the WGS result), III-4 (the affected mother of IV-5), III-5 (the unaffected father of IV-5) and V-1 (one of the other affected) by dideoxy sequencing (Table 4-9). As a result, 2 variants were considered to segregate with the phenotype; chr2: 224,858,544 C>T and chr2: 229,895,749 T>C, while other variants were inherited from III-5, the unaffected father of IV-5, or a false positive.

The C>T at Chr2:224,858,544 locates 1.8 kb upstream from exon 4 of *SERPINE2* creates new splice site with the score of 0.56. The patterns of *SERPINE2* transcripts were compared between RNAs from one of the affected subjects (IV-3) and control, however, no changes were observed.

The T>C at Chr2:229,895,749 locates about 4.7 kb away from exon 1 of *PID1* and is predicted to create new splice site with a score of 0.37 on the positive strand, however, considering *PID1* is on the reverse strand, this variant was not further taken into consideration.

TABLE 4-8 THE NUMBER OF INTRONIC AND INTERGENIC VARIANTS WITHIN SEGREGATING REGIONS ON CHROMOSOME 2 IN IV-5

Variants in the list	2856
Variants with splice prediction score change from WT to Mut	303
Exclude intergenic variants (only intronic variants were left) Exclude variants seen repeatedly in multiple sequenced affected and unaffected individuals (more likely to be an artefact) Exclude variants not segregating in WES of III-6, IV-3, IV-6 and V-1	40
Variants which create new splice site – followed by di-deoxy sequencing of IV-4,III-4,III-5 and V-1 to validate the variant and determine the segregation	7

TABLE 4-9 SEVEN INTRONIC VARIANTS WHICH CHANGED SPLICE SITE PREDICTION SCORES

Chr	Start	Ref	Variants	Nearest exon	Distance to exon (bp)	Genotype by de-deoxy sequencing				ref score	var score
						IV-5	III-4	III-5	V-1		
chr 2	49,220,601	A	G	*FSHR*	2,819	AG	AA	AG	AA	0	0.9
chr 2	50,228,720	G	C	*NRXN1*	27,370	GC	GG	GC	GG	0	0.57
chr 2	50,541,129	A	G	*NRXN1*	32,720	AG	AA	na	AA	0	0.59
chr 2	**224,858,544**	**C**	**T**	***SERPINE2***	**1,805**	**TC**	**TC**	**CC**	**TC**	**0**	0.56
chr 2	**229,895,749**	**T**	**C**	***PID1***	**4,757**	**TC**	**TC**	**na**	**TC**	**0**	0.37
chr 2	238,625,033	C	CCTTTTT GTACTCA TGTA	*LRRFIP1*	2,104	no ins[1]	na	no ins	no ins	0	0.42
chr 2	240,059,479	G	A	*HDAC4*	1,890	AG	GG	AG	GG	0	0.62

[1] no insertion was detected by di-deoxy sequencing, suggesting the WGS result was a false positive.

4.3.7 SNP CHIP ANALYSIS TO DETECT COPY NUMBER VARIANTS

After the investigation into single nucleotide change could not identify a candidate variant from the family, as a next step, we moved into the investigation copy number variants (CNVs).

For this family, array CGH had been performed on IV-3 and V-1 (the result is shown in Table 4-10 and Table 4-11) by North East Thames Regional Cytogenetics Laboratory, however, none of the CNVs identified were shared or in the segregating regions.

TABLE 4-10 ARRAY CGH RESULT ON IV-3 (ARRAY TYPE: NIMBLEGEN 135K WG, HG18)

Type	**Chr**	**Location**	**Length (bp)**	***Genes***	**Comments**
Loss	6	220,109-308,939	88,829	*DUSP22*	Known CNV
Gain	7	143,528,320-143,705,123	176,803	*OR2A42,ARHGEF5,OR 2A7,OR2A1*	Known CNV
Loss	17	18,303,140-18,320-039	16,899	-	Known CNV
Gain	17	41,541,431-41,720,878	179,447	*KIAA1267*	Known CNV
Loss	17	42,011,291-42,121,240	109,958	*NSF,ARL17,ARL17P1*	Known CNV
Loss	X	3,759,959-3,863,409	103,450	-	Known CNV
Gain	19	60,141,349-60,186,618	45,269	*NLRP2,NLRP7*	Unknown significance

TABLE 4-11 ARRAY CGH RESULT ON V-1 ARRAY TYPE: AFFYMETRIX 750K, ESTIMATED AVERAGE RESOLUTION 200KB HG19)

Type	Chr	Location	Length (bp)	*Genes*	Comments
Gain	7	57,149,589-57,597,640	448,051	*ZNF479, GUSBP10, ZNF716, MIR3147*	Known CNV
Gain	15	31,255,090-31,324,010	68,920	*MTMR10, TRPM1*	Known CNV
Loss	22	25,656,219-25,911,100	254,880	*IGLL3P, CRYBB2P1, LRP5L*	Known CNV

To detect smaller CNVs, a high-resolution SNP array using Illumina Infinium Omini2.5 Exome v1.2 (total number of markers; 2,612,357) was performed on III-6 and IV-3. III-6 and IV-3 were selected based on the quantity and quality of the remaining DNA samples.

The data were analysed with NEXUS Copy Number ™ under the supervision of Dr. Samantha Knight at Wellcome Trust Centre for Human Genetics, University of Oxford. The data from IV-3 were analysed first and III-6 was used to examine the segregation. The number of detected variants with the set-up filtering[9] was as below.

TABLE 4-12 THE NUMBER OF VARIANTS DETECTED BY NEXUS COPY NUMBER

	IV-3	III-6
CN Gain	11	19
CN Loss	17	12
Homozygous loss	2	3

For the segregating regions, no variants were called from the automatic filtering. By inspecting the 10 potentially segregating regions manually, there were 2 possible CNVs; chr10:126,674,471-126,674,695, homozygous copy loss and chr2:231,376,175-231,376,432, copy number (CN) Loss,

[9] Filtering conditions for the NEXUS copy number analysis;
Significance Threshold = 1.0E-6, Max Contiguous Probe Spacing (kb) = 500.0, Min number of probes per segment = 10, High Gain = 0.5, Gain = 0.12, Loss = -0.12, Big Loss = -0.7, Male Sex Chromosomes Big Loss = -1.1, 3:1 Sex chromosome gain = 1.2, 4:1 Sex chromosome gain = 1.7, Homozygous Frequency Threshold = 0.98, Homozygous Value Threshold = 0.85, Heterozygous Imbalance Threshold = 0.46, Minimum LOH Length (kb) = 500, Minimum SNP Probe Density (Probes/Mb) = 0.0

however, both of variants were detected with only 1 probe and the same probes in III-6 were observed in the normal range, suggesting those were unlikely to be real CNVs.

Subsequently, all CNVs in IV-3 were examined. Most of the variants were reported variants in Database of Genomic Variants (DGV). Five variants on Table 4-13 were novel variants without previous reports in DGV. For 4 variants out of 5, same probes or the regions in III-6 were within the normal range, meaning not segregating in III-6. The CN Gain at chr1:63,595,285-63,607,207 was also detected in III-6, but this region was considered as outside of segregating regions as 2 rare variants on chr1:63,269,410 ACT>A and chr1:63,876,987 A>G had been inherited in the unaffected subject, III-8(Refer to Appendix3).

TABLE 4-13 THE NOVEL SVS DETECTED IN IV-3

Chromosome Region (GRCh37)	**Event**	**Length (bp)**	**No. of Probes**	**Notes**	**Genes in the region**	**In data of III-6**
chr1:63,595,285-63,607,207	CN Gain	11,923	22	No gains in DGV in the region		Yes
chr3:148,870,377-148,880,334	CN Gain	9,958	15	No gains in DGV in the region	*HPS3, CP*	No
chr4:32,851,796-32,871,207	CN Gain	19,412	13	1 overlapping gain in DGV		No
chrX:23,803,232-23,820,859	CN Gain	17,628	14	No CNVs in DGC in the region	*SAT1*	No
chrX:91,704,903-91,724,026	CN Gain	19,124	12	No CNVs in DGV in the region	*PCDH11X*	No

As a final confirmation, all the variants in III-6 and IV-3 were compared to find shared rare CNV including non-segregating regions (Table 4-14). Seven shared CNVs were identified, however, all of them were reported CNVs in DGV and not located in the segregating regions either, apart from previously detected chr1:63,595,285-63,607,207.

TABLE 4-14 SHARED CNVS IN III-6 AND IV-3

Position (GRCh37)	Event	Size	No. of probes	Report in DGV	Genes
chr1:63,595,285-63,607,207	CN Gain	11,923	22	No gains in DGV in the region	
chr5:45,910,864-46,403,195	CN Gain	492332	260	known CNVs	
chr7:25,376,122-25,382,638	CN Loss	6517	10	known CNVs	
chr8:5,595,499-5,606,012	CN Loss	10514	25	known CNVs	
chr17:44,165,593-44,465,113	CN Gain	299521	80	known CNVs	*KANSL1, KANSL1-AS1, LRRC37A, ARL17A, ARL17B, NSFP1*
chr19:55,300,306-55,344,791	CN Gain	44486	20	known CNVs	*KIR2DL4, LOC100287534, KIR2DS4, KIR3DL1*
chr19:55,441,378-55,504,955	CN Gain	63578	120	known CNVs	*NLRP7, RNU6-64P, NLRP2*

4.3.8 WGS OF IV-3 AND V-1

At this stage, we had undertaken WES on IV-3, III-6, IV-6 and V-1, WGS on IV-5 and high-resolution SNP array on III-6 and IV-3, however, no strong candidate mutation had been identified. One of the problems at this stage was that the WGS of IV-5 by Complete Genomics provided us only reads that different from reference instead of all raw data, therefore, there was a difficulty in merging the results with other WES data to run a comprehensive analysis. To overcome this problem, WGS on IV-3 and V-1 were undertaken by the Oxford Biomedical Research Centre Genetics Laboratory. The samples from IV-3 and V-1 were selected due to the quality and quantity of remaining DNA, although a pair of samples with more distant relationship e.g. V-1 and III-6 would narrow down the segregating regions more effectively.

Data were uploaded to the new CranioVar-GRCh38 system and variants were called by Platypus v0.5.2 and SAMtools v.1.1. (Table 2-1)

Firstly, the variants shared in WGS of V-1 and IV-5 were evaluated as these were the most distantly related individuals who had undergone WGS. A total 2,198 variants were extracted from both data[10], of which, 15 variants were shared in both V-1 and IV-5. Variants which reported in 1000G, ESP or ExAC were excluded, leaving 3 variants as Table 4-15.

TABLE 4-15 SHARED EXONIC VARIANTS BETWEEN WGSOF V-1 AND IV-5

Gene	Chr	Position (GRCh38)	Ref	Var	Function	DS
NRXN1	chr2	50497620	G	A	Synonymous	-
JUND	chr19	18281347	C	T	Synonymous	-
CREM	chr10	35171295	AT	CC	nonframeshift substitution	-

The synonymous variant in *JUND* at chr19:18,281,347 didn't segregate in IV-6 (0 variant reads, while 8 wild type reads in WES of IV-6). The nonframeshift substitution in *CREM* is the combination of 2 SNPs; rs536062567 and rs555551825 both with allele frequency 0.006, suggesting the variant is unlikely to be causative.

Secondly, the variants shared in IV-3 and V-1 were examined. In total 2544 variants were listed[11], of which 44 variants were located in segregating regions. Most of the 44 variants were mapping artefacts; 18 variants were commonly observed in other exome sequencing, 22 variants were detected only in one of IV-3 or V-1 with the variant/wild type proportion less than 20%, suggesting calling error. Of the remaining 4 variants, the variant in *NRXN1* was the only undescribed variant.

[10,4] Filtering Condition; Max.ESP freq 0.001, Min.var read depth 4, Min.proportion var reads(%) 0, Max.var observation across database 8, Max.1000Genomes freq 0.0001, Max.proportion var read (%) 100, Max.ExAC allele freq 0.001, Min. ExAC gene constraint (pLI) 0.5, Exonic variants only (stoploss, stopgain, nonsynonymous SNV, frameshift insertion, frameshift deletion, frameshift substitution, splicing, splice consensus, synonymous SNV, nonframeshift insertion, nonframeshift deletion, nonframeshift substitution, unknow.n

4.3.9 STRUCTURAL VARIANTS IN WGS

From WGS of IV-3 and V-1, SVs were investigated by 2 ways; NEXUS Copy Number™ to detect CNVs and DELLY (Rausch et al., 2012) to detect CNVs and copy neutral SVssuch as inversions and translocations. The analysis using NEXUS Copy Number was done under the supervision of Dr. Samantha Knight and analysis with DELLY was undertaken by Dr. Nils Koelling (Clinical Genetics Group, Weatherall Institute of Molecular Medicine).

During the process of uploading the data to NEXUS Copy number™, we noticed that WGS of IV-3 was PCR amplified for WGS, so that WGS of V-1 was used to examine CNVs and IV-3 to cross-reference the region.

4.3.9.1 NEXUS COPY NUMBER TO IDENTIFY COPY NUMBER VARIANTS

In total 288 deletions and duplications were picked up from the program[12]. Three gains were located in the segregating regions, however, none of them was detected in the high-resolution SNP array data of IV-3 and III-3 or observed in WGS data on GBrowse, suggesting that all of these are likely to be false positives.

Subsequently, all the variants detected in V-1 were examined. Of the 288 regions called, most were excluded because CNVs are in DGV, in segmental duplications or in telomere or centromere regions, however, 1 CN gain and 1 CN loss on chromosome 3 looked real (Table 4-16), but these regions were out of segregating regions and no variants were observed in the high-resolution SNP chip of III-6 as well.

TABLE 4-16 CNVS DETECTED FROM WGS OF V-1

Chromosome Region (GRCh37)	Event	Length (bp)	No. of Probes	Genes
chr3:134,808,192-134,821,487	CN Loss	13,296	8	*EPHB1 intron*
chr3:118,903,293-118,921,695	CN Gain	18,403	12	*UPK1B*

[12] Filtering condition; Significance Threshold = 1.0E-6, Max Contiguous Probe Spacing (kb) = 5.0, Min number of probes per segment = 2, High Gain = 1.5, Gain = 0.13, Loss = -0.12. Big Loss = -1.5, Male Sex Chromosomes Big Loss = -2.0, 3:1 Sex chromosome gain = 1.5, 4:1 Sex chromosome gain = 2.0, Homozygous Frequency Threshold = 0.88, Homozygous Value Threshold = 0.8, Heterozygous Imbalance Threshold = 0.4, Minimum LOH Length (kb) = 1000, Minimum SNP Probe Density (Probes/Mb) = 0.0, Organism = Human, Build = NCBI Build 37

Because no possible causative variants were detected in the first analysis, a higher resolution analysis[13] was undertaken subsequently, which detected 12,413 CNVs. Seven CNVs located in the 10 possible segregating regions (Table 4-17). Corresponding regions were cross-referenced in the high-resolution SNP array of IV-3 and III-6, however, none of the 7 CNVs were observed. The variants detected from the high-resolution analysis were considered to be either false positives or ones that exist only in V-1 (inherited from the unaffected father).

TABLE 4-17 CNVS DETECTED FROM WGS OF V-1 WITH HIGHER RESOLUTION FILTERING

Chromosome Region (GRCh37)	Size (bp)	Event	Probes	Gene Symbols	Genes
chr1:239,787,251-239,791,791	4,541	CN Loss	10	0	
chr1:243,318,434-243,320,499	2,066	CN Loss	5	1	*CEP170*
chr1:243,856,865-243,863,499	6,635	CN Loss	14	1	*AKT3*
chr2:207,063,169-207,068,255	5,087	CN Gain	8	2	*GPR1, GPR1-AS*
chr10:94,820,878-94,834,214	13,337	CN Gain	23	2	*CYP26C1, CYP26A1*
chr15:66,852,106-66,863,790	11,685	CN Gain	19	1	*LCTL*
chr10:98,266,167-98,275,088	8,922	CN Gain	15	1	*TLL2*

4.3.9.2 STRUCTURAL VARIANTS ANALYSIS ON WGS OF V-1 BY DELLY

In parallel to the analysis by NEXUS Copy Number™, SVs were analysed in the data of WGS of V-1. DELLY was designed to detect copy-number variable deletion and duplication events (CNVs) as well as balanced rearrangements such as inversions or translocations (Rausch et al., 2012). In this analysis, translocations and inversions were evaluated in the first instance, considering those would have been missed in the previous search by NEXUS Copy Number™.

The primary list from DELLY contained 69,503 variants, of which 825 variants were in the segregating regions. Further filtering was done by selecting heterozygous variants and quality filters, leaving 255

[13] Filtering condition; Significance Threshold = 1.0E-6, Max Contiguous Probe Spacing (Kbp) = 5.0, Min number of probes per segment = 2, High Gain = 1.5, Gain = 0.13, Loss = -0.12, Big Loss = -1.5, Male Sex Chromosomes Big Loss = -2.0, 3:1 Sex chromosome gain = 1.5, 4:1 Sex chromosome gain = 2.0, Homozygous Frequency Threshold = 0.88, Homozygous Value Threshold = 0.8, Heterozygous Imbalance Threshold = 0.4, Minimum LOH Length (KB) = 1000, Minimum SNP Probe Density (Probes/MB) = 0.0, Organism = Human

variants. Fifteen out of 255 variants were inversions and translocations (Table 4-18). Most inversions were described in DGV, apart from 2 inversions; chr2: 51,438,919-51,439,021 and chr15: 34,455,121-65,404,331. For the former, some larger losses containing the region are reported in DGV. The inversion is 400 kb away from *NRXN1* and no regulatory signals were seen in that region. Although this inversion can be visualised in GBrowse in WGS of both V-1 and IV-3 (V-1's affected mother) and the region is within the confirmed segregating regions, it's hard to judge the pathogenicity. For the latter, a part of the inversion is included in the possible segregation region, however, the majority of the inversion is out of the regions.

DELLY detected 4 possible translocations on chromosome 2 and 15, however, chromosome analysis was undertaken on III-4 and a negative result in 1993 at Birmingham Women's Hospital (based on the description in the clinical note, no actual result was recorded).

TABLE 4-18 INVERSIONS AND TRANSLOCATIONS IN WGS OF V-1 BY DELLLY

Chr	Start	End	Size (bp)	Event[1]	Function	Gene	DGV	Translocations
chr2	51,438,919	51,439,021	102	<INV>	intergenic	*NRXN1,MIR4431*	Not reported (larger losses which contains this region in DGV)	
chr1	237,402,804	237,402,905	101	<INV>	intronic	*RYR2*	nsv1140268	
chr1	237,402,807	237,402,912	105	<INV>	intronic	*RYR2*	nsv1140268	
chr2	207,645,153	207,645,454	301	<INV>	intergenic	*METTL21A,CCNYL1*	nsv1110934	
chr2	205,938,356	205,938,356	0	<TRA>	.	.		CHR2=chr1;END=178040326
chr10	29,495,313	30,689,027	1,193,714	<INV>	exonic	*KIAA1462,LYZL2,MAP3K8,MTPAP,SVIL*	nsv1129323	
chr10	29,495,050	30,689,300	1,194,250	<INV>	exonic	*KIAA1462,LYZL2,MAP3K8,MTPAP,SVIL*	nsv1129323	
chr10	125,501,849	125,502,207	358	<INV>	intergenic	*CTBP2,TEX36-AS1*	esv3426545	
chr10	125,502,109	125,508,653	6,544	<INV>	intergenic	*CTBP2,TEX36-AS1*	esv7651	
chr10	125,501,849	125,512,533	10,684	<INV>	intergenic	*CTBP2,TEX36-AS1*	nsv1148108	
chr15	39,702,423	39,702,423	0	<TRA>	.	.		CHR2=chr12;END=565596313
chr15	34,346,632	34,346,632	0	<TRA>	.	.		CHR2=chr5;END=123583553
chr15	34,455,121	99,859,452	65,404,331	<INV>	exonic	*AAGAB,ABHD17C,ABHD2,ACAN,ACSBG1,ACTC1,ADAL,ADAM10,ADAMTS7,ADAMTSL3,ADPGK,AEN,AGBL1,AKAP13, ALDH1A2, (more)*	Not reported	
chr15	39,702,436	39,702,436	0	<TRA>	.	.		CHR2=chr12;END=565595921
chr15	68,395,723	68,395,723	0	<TRA>	.	.		CHR2=chr5;END=8442858

INV, inversions. TRA, translocations. Start and End positions are described using GRCh38

4.4 DISCUSSION

Although multiple approaches were taken to identify the causal variant in the family, no strong candidate variants have been identified. The *NRXN1* (c.2592C>T, p.N864N) was identified as previously undescribed variant and completely segregates in the family, however, most of splice prediction tools predicted no splice effect apart from RESCUE-ESE. Since the expression of *NRXN1* is limited to brain, I could not demonstrate whether this variant affects the splicing. The LOD score for *NRXN1* (c.2592C>T, p.N864N) was calculated as 2.408, which is the highest LOD achievable with this pedigree. Considering that a LOD of > 3 is necessary for significant linkage, this illustrates the limitations, even for this unusually large unsolved craniosynostosis pedigree, of pinpointing a unique location of the disease gene.

Another novel 102 bp inversion on chr2:51,438,919-51,439,021[14] (intergenic region between *NRXN1* and *MIR4431*) was found to be segregating in the family (Table 4-18), however, it was not strongly considered likely to be the pathogenic variant considering the location.

The family displays a distinctive phenotype through 4-5 generation. If the pathogenic variant locates in a novel locus, there is a question why there are no other families sharing the pathogenic gene with this family. The pathogenic variant might be specific amino acid substitution in a gene or specific rearrangement in the family, considering the uniqueness of the case. An example of how disease with highly specific pathophysiology can be rare is provided by *MSX2* related craniosynostosis, which is caused by mutations specifically at the Pro148 residue in the homeodomain. This was the first molecular lesion to be described in craniosynostosis. (Jabs et al., 1993), but it took another 20 years before the next two families were identified (Florisson et al., 2013; Janssen et al., 2013).

If the pathogenic variant would be in one of the known loci for craniosynostosis, *IHH* (Indian hedgehog, encoding a secreted signalling molecule of the hedgehog family) and upstream *IHH*

[14] in GRCh38

regulatory region should be taken into account. *IHH* locates on chr2:219,919,142-219,925,238, close to one of the segregating regions in the family (chr2:220,413,659-240,900,606). CNVs involving the IHH locus are known to be associated with craniosynostosis (Barroso et al., 2015; Klopocki et al., 2011). However, no overlap between the segregating region and *IHH* and its' regulatory regions was confirmed and no CNVs were identified around the loci.

Another major craniosynostosis disease gene, *FGFR2* at chr10:123,237,844-123,357,972 also locates near the segregating region on chr10: 124,390,667-131,906,905, however, no shared variants were identified in *FGFR2*.

Other possible reasons why the pathogenic variants were not revealed, these include (1) The variant was not called due to insufficient coverage in WES or WGS, (2) Rearrangements are still challenging issue to call, (3) Analysis of segregating regions has not identified all the segregating regions. This could occur if some gaps exist between rearrangements, (4) Phenotype assumption was not correct. For example, non-penetrance in III-7 or III-8 leads the exclusion of pathogenic variant from the candidate list. If one of the affected individuals is a phenocopy, the pathogenic variant can't be identified.

Identification of SVs, in particular balanced chromosomal abnormalities, has been shown to be difficult. Redin et al. (2017) applied WGS onto 273 subjects with congenital anomalies and which revised 93 % of karyotypes previously reported and limitations in conventional cytogenic approaches was noted. Particularly considering cytogenetics testing of the family was undertaken in 1993, there might have been something missed. SV analysis in WGS has not detected possible SVs in our analysis. Currently, as a next step, *de novo* assembly is being attempted to identify previously occult SVs.

Chapter 5 A MOTHER AND DAUGHTER WITH MULTIPLE SYNOSTOSIS WITH *DE NOVO* DUPLICATION ON CHROMOSOME 4

5.1 INTRODUCTION

In the previous chapters, large families with autosomal dominant inheritance were described. The family described in this chapter comprises 2 affected individuals, a mother and daughter (II-2 and III-2), with very similar phenotype; both had clover leaf skulls, the most severe craniosynostosis phenotype, as shown in Figure 5-1. Although, autosomal dominant segregation was suspected to have occurred between II-2 and III-2, the maternal grandparents (who were both examined by a consultant clinical geneticist) manifest no pathological appearances. Therefore, I applied a *de novo* disease model to this family, considering that a *de novo* variant that arose in the affected mother (II-2), and was passed on to the proband (III-2) would be a candidate. A *de novo* duplication with this inheritance pattern was detected by array CGH. In this chapter, I have focused on this duplication and investigated whether it may be pathogenic.

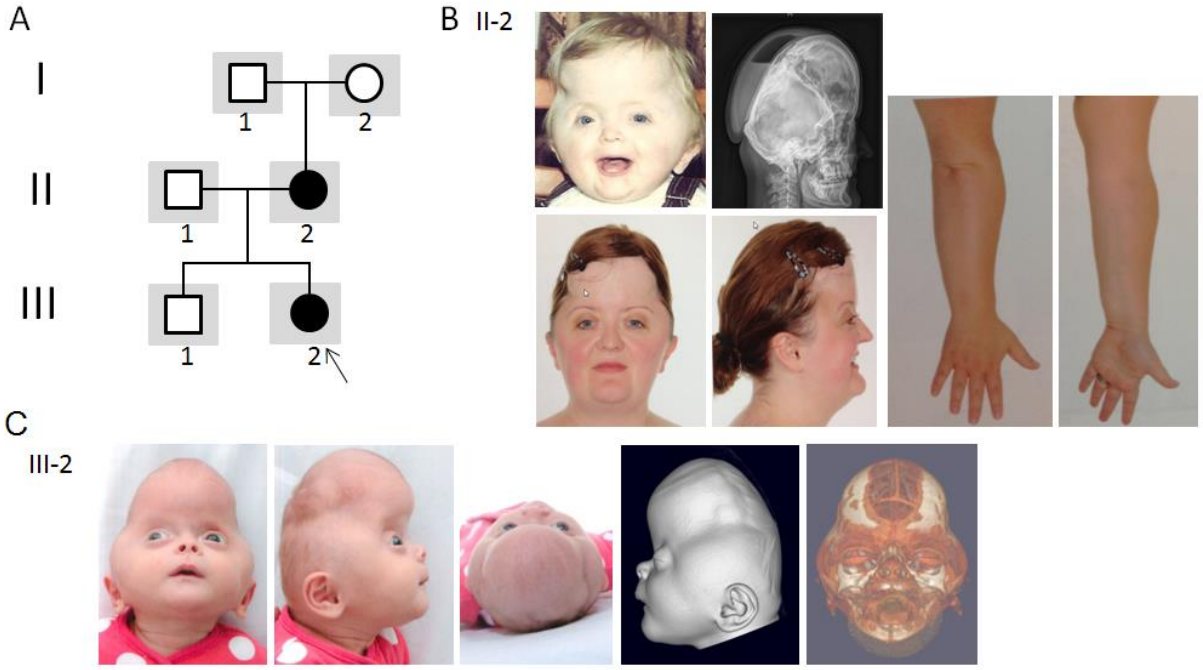


Figure 5-1 Pedigree and clinical photos
A, Family tree. The arrow indicates the proband. Squares, male family members; circles, female family members; black symbols, affected subjects; white symbols, unaffected subjects, grey shading, DNA obtained.
B Clinical pictures of II-2. upper panel; clinical picture and skull X-ray at 1 year old. Severe skull deformity with cloverleaf skull shape is observed. Lower panel; clinical pictures at the age of 33 years. The skull deformity persisted. She had short forearms and fingers.
C Clinical pictures of III-2. Clinical pictures and 3D-CT scans at the age of 1-month-old. Cloverleaf skull, hypertelorism, low set ears and a depressed nasal bridge are observed. 3D CT scan showed multiple synostosis with a severe skull deformity.

5.2 CLINICAL INFORMATION

5.2.1 III-2 (PROBAND)

The III-2 was detected a cloverleaf skull by transabdominal sonography at the 23rd weeks of gestation and the mother was referred for a prenatal assessment by the craniofacial team. Additional concerns were raised about airway obstruction and proptosis based on antenatal MRI. Her mother (II-2) was also affected multiple synostosis and underwent surgeries after her birth as below. III-2 was born by an elective caesarean section at 37 weeks. She had a cloverleaf skull with a widely open anterior fontanelle, hypertelorism, low set ears and a depressed nasal bridge, but no airway problems and no proptosis. Her anterior fontanelle was bulging but not tense. OFC was 29cm (below 2nd centile). CT scan showed a cloverleaf deformity of the head with synostosis of the lambdoid and coronal sutures, a severe constriction of the brain, temporal horn dilatation and aberrant venous anatomy in the posterior fossa (Figure 5-1-C). The degree of ventricular enlargement was suggestive of a degree of long-standing CSF flow obstruction. At 1 month of age, she had a posterior craniectomy, followed by insertion of endoscopic third ventriculostomy at 7 months and a posterior craniectomy and a partial remodelling at 15 months old. At 3 years and 3 months, she underwent a fronto-orbital distraction advancement. An assessment of her communication development at 3 years and 3 months was reported as mildly delayed, while her cognitive abilities were developing well equivalent to her age. She had a mild delay in motor and language development at the age of 4 years.

5.2.2 II-2 (MOTHER OF III-2)

The proband's mother (II-2) was born at 38 weeks of gestational age after a normal pregnancy. She was born with a significant cloverleaf skull abnormality. She had small eyes with proptosis and small ears. Her length was 48 cm (25th centile) and OFC was 29 cm (below 2nd centile). Neurological and general clinical examinations were normal. MRI showed cloverleaf deformity and minor symmetrical dilation of the lateral ventricles, suggesting mild hydrocephalus. At 3 months of age, she had the first stage of craniectomy followed by the 2nd stage of craniectomy at 4 months and the 3rd stage at 5

months of age. Her postoperative course was good and her development and intelligence were normal.

She had an assessment for short stature by a paediatric endocrinologist at 12 years old. At that time, her height was 135.9 cm (2nd centile) and bone age corresponded to her chronological age, suggesting that her shortness was constitutional. Her final height was 144.8 cm (below 2nd centile) and had short arms (Figure 5-1).

The genetic diagnostic tests for *FGFR1* exon 7, *FGFR2* exons IIIa and IIIc, *FGFR3* exons 7 and 10, *TWIST1* exon1, *FGFR2* exons 3, 5, 11, 14, 15, 16 and 17 were requested, but no pathogenic variants were detected. MLPA (Multiplex Ligation-dependent Probe Amplification) analysis of *TWIST1, RUNX2, ALX1, ALX3, ALX4, MSX2,* and *EFBN1* failed to detect any deletions or duplications.

5.2.3 OTHER FAMILY MEMBERS

Since the phenotype of the proband's maternal grandparents (I-1 and I-2) was critical in deciding the disease model, their clinical features were carefully examined by one of the clinical geneticists. Like the mother of the proband (II-2), the grandmother (I-2) had short stature (149.8 cm, below 2nd centile), however there were no other features that suggested she was affected and proportionally normal. The proband's great-grandmother was also short stature, 147.3 cm (below 2nd centile). Although, short stature is a common feature in mother's side of the family, the proportion of I-2 and II-2 were different. Given these, the conclusion was that the grandparents were unaffected and short stature in I-2 and her mother was constitutional rather than related to the craniosynostosis disease phenotype.

5.3 RESULTS

5.3.1 ARRAY CGH

Following the negative results of clinical genetic testing, array CGH was carried out on 4 individuals (I-1, I-2, II-2 and III-2) by Oxford Medical Genetics Laboratories. The array CGH analysis (using an Agilent ISCA 180 kb oligo array) showed that II-2 and III-2 have 2 imbalances; a 670 kb duplication of 4q21 from base pair 81,160,133 to 81,884,710; and a 220 kb deletion of 22q12.3 from base pair

34,211,967 to 34,436,983 (GRCh37). The array CGH of I-1 and I-2 identified that the deletion was inherited from I-2 and the duplication was *de novo* arising in II-2. The array CGH data showed that the *de novo* duplication covered the fibroblast growth factor gene *FGF5* and part of *C4ORF22* (Figure 5-2).

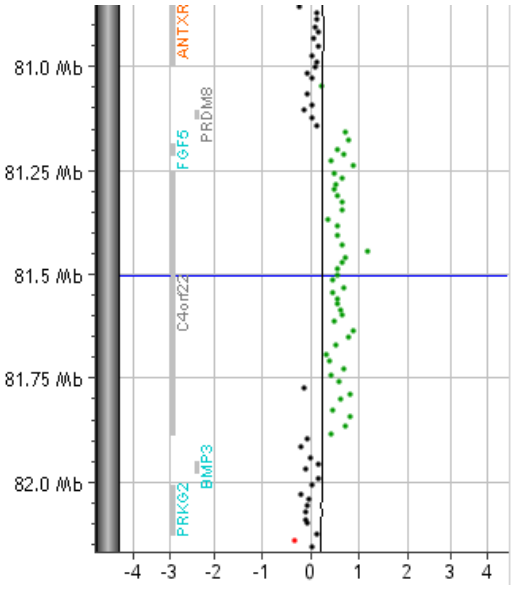


FIGURE 5-2 GRAPHICAL VIEW OF THE DUPLICATION DETECTED BY ARRAY CGH

Gene view of chr4:80,839,991-82,167,175
(provided by Dr. Carolyn Campbell, Oxford Medical Genetics Laboratories)

No gains or losses in the exact same region are reported in DGV. In DECIPHER (https://decipher.sanger.ac.uk/), a number of deletions are reported in the region, while only three duplications which overlap with the *de novo* duplication have been reported (Table 5-1).

FGF5 encodes one of the Fibroblast growth factors (FGFs), however, its role is mainly reported as a regulator of the hair growth cycle (Hebert et al., 1994; Housley and Venta, 2006). Little is known about the function of *C4ORF22*. *BMP3*, encoding bone morphogenetic protein 3, locates 67 kb downstream of the duplication. Schoenebeck et al. (2012) identified a missense substitution p.Phe452Leu in *BMP3*, that is nearly fixed among extreme brachycephalic dog breeds, by WGS analysis from eleven dog breeds of widely varying skull shapes (Schoenebeck et al., 2012). *BMP3* functional assays using a zebrafish model suggested that *BMP3* is indispensable for normal

craniofacial development (Schoenebeck et al., 2012). BMPs are members of the transforming growth factor-beta (TGF-beta) superfamily. Several papers reported that *Bmp3* works as a negative determinant of bone density or an inhibitor of cartilage repair (Daluiski et al., 2001; Zhang et al., 2015). In mouse, the minimal promoter required for expression resides within a 0.8 kb region upstream of *Bmp3* that is conserved among mammalian genomes (Lowery et al., 2013), but the *de novo* duplication does not include *BMP3* and its promoter. On the other hand, distant duplications or deletions of noncoding 5' and 3' elements are known to be associated with some phenotypes, for example, in *BMP2*, it is reported that duplications downstream of *BMP2* are associated with brachydactyly type A2 (Dathe et al., 2009) and Duplication of noncoding elements of *SOX9* are also reported to be associated with brachydactyly-anonychia (Kurth et al., 2009). As described in section 1.7.2 the distance to the promoter from CNVs can be up to 1.5 Mb (Benko et al., 2009; Kurth et al., 2009; Liu et al., 2014). Therefore, the duplication might influence the expression of downstream *BMP3*.

In terms of the deletion which was detected in the family, the 220 kb deletion at 22q12.3, partially encompasses the gene *LARGE* (approx. 660 kb in size). No losses or gains in the same region are described in DGV. Homozygous and compound heterozygous mutations and deletions of *LARGE* have been reported in patients with muscular dystrophy-dystroglycanopathy (OMIM603590), however, craniofacial malformations are absent. In DECIPHER, 4 deletion cases overlap with the deletion identified in this family: 1) a *de novo* 5.5 Mb deletion at chr22:33,495,795-38,995,208 with abnormality of lung morphology, atrial septal defect, hearing impairment, but no craniofacial abnormality; 2) a 12 Mb deletion case with unknown inheritance and clinical information; 3) and 4) two deletions inherited from normal parents, 68.3 kb and 351.3 kb in size. Considering the deletion was heterozygous and inherited from the unaffected maternal grandparental (I-2) origin, the deletion was considered not to be pathogenic, at least as far as the craniosynostosis is concerned.

TABLE 5-1 DUPLICATIONS OVERLAPPING WITH OUR CASE IN DECIPHER

Patient	position	length (bp)	Inheritance	Pathogenicity	Genes involved	Phenotype
250353	chr4:74432894-81454157	7,021,264	unknown	Unspecified	80	Iris coloboma, Cryptorchidism, Abnormality of the face, microcephaly, Delayed eruption of teeth, Intellectual disability
256563	chr4:74201207-110610111	36,408,905	Imbalance arising from a balanced parental rearrangement	Unspecified	300	Growth retardation, Coloboma, Absent speech, Intellectual disability
248919	chr4:78832337-84479993	5,647,657	De novo	Unspecified	56	Enuresis, Obesity, Finger joint hypermobility, Genu valgum, Intellectual disability

5.3.2 EXOME SEQUENCING

The array CGH identified a *de novo* duplication that is a possible causative for the craniosynostosis found in the family, however, to confirm that this is the only *de novo* mutation to focus on, exome sequencing was performed.

Initially exome sequencing was undertaken on III-2. Thirty-five variants were selected as candidates based on gene function and deleterious scores and by excluding common variants present in the ESP. These variants were validated by dideoxy-sequencing of samples from I-1, I-2 and II-2 to test for maternally transmitted *de novo* mutations. Nine of 35 variants were transmitted from II-2 to III-2, however, all were inherited from I-1 or I-2. In a more comprehensive analysis, exome sequencing of the mother (II-2) and maternal grandparents (I-1, I-2) was carried out. Although qualities of the exome sequence of I-1, I-2 and II-2 were relatively poor as shown by the low percentage of exons at 10x coverage compared to of III-2 in Table 5-2, no *de novo* mutations were found from the trio analysis.

TABLE 5-2 SUMMARY OF EXOME SEQUENCING

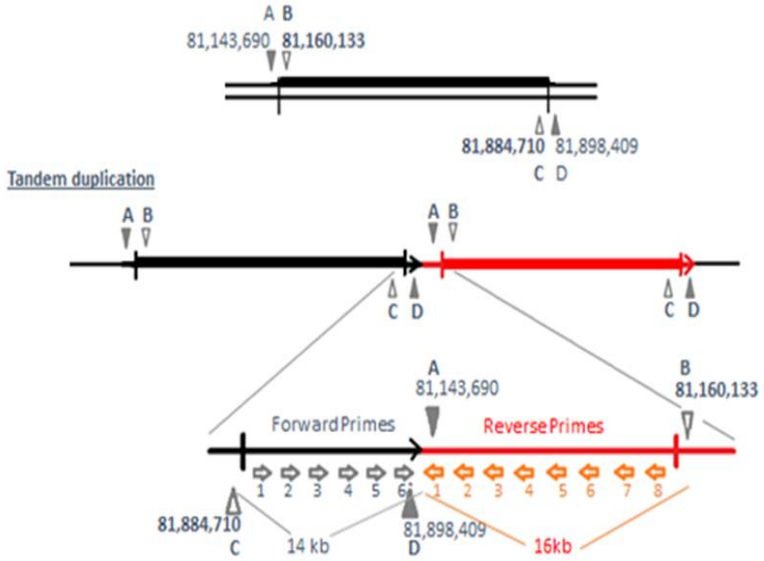
	Capture	Exons at 10x coverage	Exons at 30x coverage	Variant Calling
I-1	SureSelect V5	74.8 %	58.4 %	SAM tools v0.1.19
I-2	SureSelect V5	68.5 %	49.5 %	SAM tools v0.1.19
II-2	SureSelect V5	69.9 %	51.0 %	SAM tools v0.1.19
III-2	True Seq	85.2 %	56.8 %	SAM tools v0.1.19

5.3.3 MOLECULAR NATURE OF THE DUPLICATION

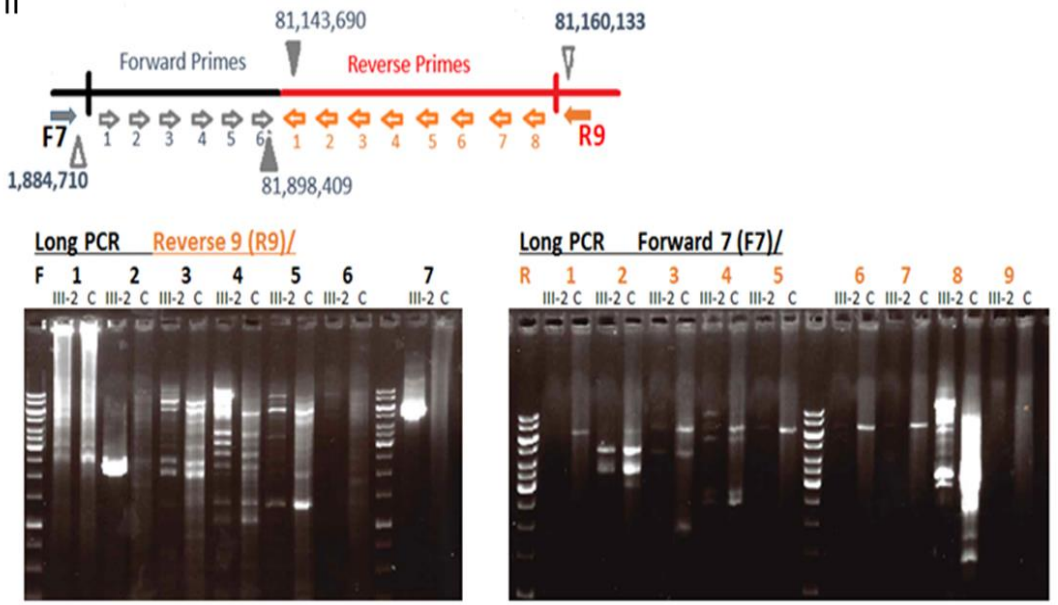
To determine the exact molecular nature of the *de novo* duplication, a long PCR analysis using nested primers was undertaken, assuming tandem duplication. The duplication coordinates were estimated as chr4:81,160,133-81,884,710 by array CGH, with the most distal point of the probe immediately proximal to the region of duplication at chr4: 81,143,690 and the most proximal point of the probe immediately distal to the region of duplication at chr4: 81,898,409. In total 14 primers were designed in the region of chr4:81,143,690-81,160,133 and chr4:81,884,710-81,898,409 as shown in Figure 5-3.

The data indicated that the duplication was a tandem duplication most likely caused by non-homologous end joining and its breakpoint was chr4:81,159,506-81,889,443 (with 3 nucleotide overlaps). The array CGH had shown the duplication covers entire *FGF5* and a part of *C4ORF22*, but this analysis revealed the duplication covers the entirety of *FGF5* and *C4ORF22.*

I


A B
81,143,690 81,160,133
81,884,710 81,898,409
C D
Tandem duplication
A B
A B
C D
C D
A
81,143,690
B
81,160,133
Forward Primes
Reverse Primes
1 2 3 4 5 6i 1 2 3 4 5 6 7 8
81,884,710
C
14 kb
81,898,409
D
16kb

II


81,143,690
81,160,133
Forward Primes
Reverse Primes
F7
1,884,710
1 2 3 4 5 6i 1 2 3 4 5 6 7 8
R9
81,898,409
Long PCR Reverse 9 (R9)/
F 1 2 3 4 5 6 7
III-2 C
Long PCR Forward 7 (F7)/
R 1 2 3 4 5 6 7 8 9
III-2 C

III

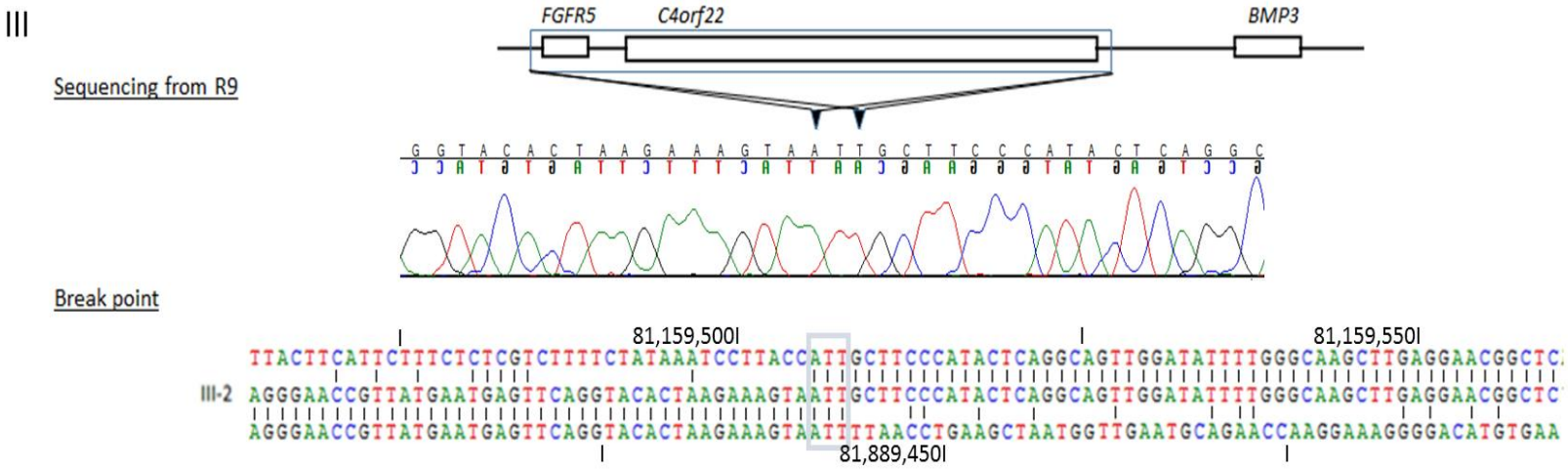

FGFR5
C4orf22
BMP3
Sequencing from R9
Break point
81,159,500l
81,159,550l
III-2
81,889,450l

FIGURE 5-3 LONG PCR AND WALKED IN PRIMERS

I. Upper panel; a schematic representation of the duplication. B and C indicate the duplication coordinates from array CGH. A shows the distal point of the probe immediately proximal to the region of duplication, while D shows the proximal point of the probe immediately distal to the region of duplication.
Middle panel; the schematic image of the tandem duplication. The thickened black and red regions indicate the assumed duplicated region.
Bottom panel; expansion of the region between C and B. Eight forward primers were designed in the region of chr4:81,143,690-81,160,133 and 6 reverse primers in the region of chr4:81,884,710-81,898,409 at approximately 2 kb intervals.
II. Upper panel; positions of primers.
Lower panel; long PCR products. PCR amplification using primer combinations R9/F2 and R9/F7 generated product sizes of 8.7 kb (Lane 2 III-2) and 30 kb (Lane 7 III-2), respectively.
III. DNA sequence characterisation of the duplication, sequenced the 8.7 kb product in II. The upper panel shows the positions of genes in the duplication (rectangular box).
Below, the DNA sequence chromatogram spanning the breakpoint. The grey rectangle indicates the range of positions of the breakpoint within the three nucleotide identity shared by both normal sequences.

5.3.4 SEGREGATION OF THE DUPLICATION

From exome sequencing of II-2 and III-2, three informative SNPs in *FGF5, BMP3* and *PRKG2* which were heterozygous in one or both of the affected individuals were identified (Table 5-3). Genotyping of these 3 SNPs in the family (II-1, II-2, III-1, III-2, IV-1 and III-2) was undertaken to investigate the segregation of these SNPs, especially to see whether this duplication segregates in the unaffected brother (III-1). As shown in Figure 5-4, the duplication only segregates in II-2 and III-2. The duplicated allele was of grandpaternal origin (I-1).

TABLE 5-3 GENES, INFORMATIVE SNPS AND GENOTYPES OF EACH INDIVIDUAL FROM EXOME SEQUENCING

gene	*FGF5*	*BMP3*	*PRKG2*
SNP	rs3733336	rs3733549	rs17005080
Sample		genotype	
III-2	GGA	GA	GA
II-2	GGA	GG	GA
I-1	GA	GG	AA
I-2	GA	GG	GA

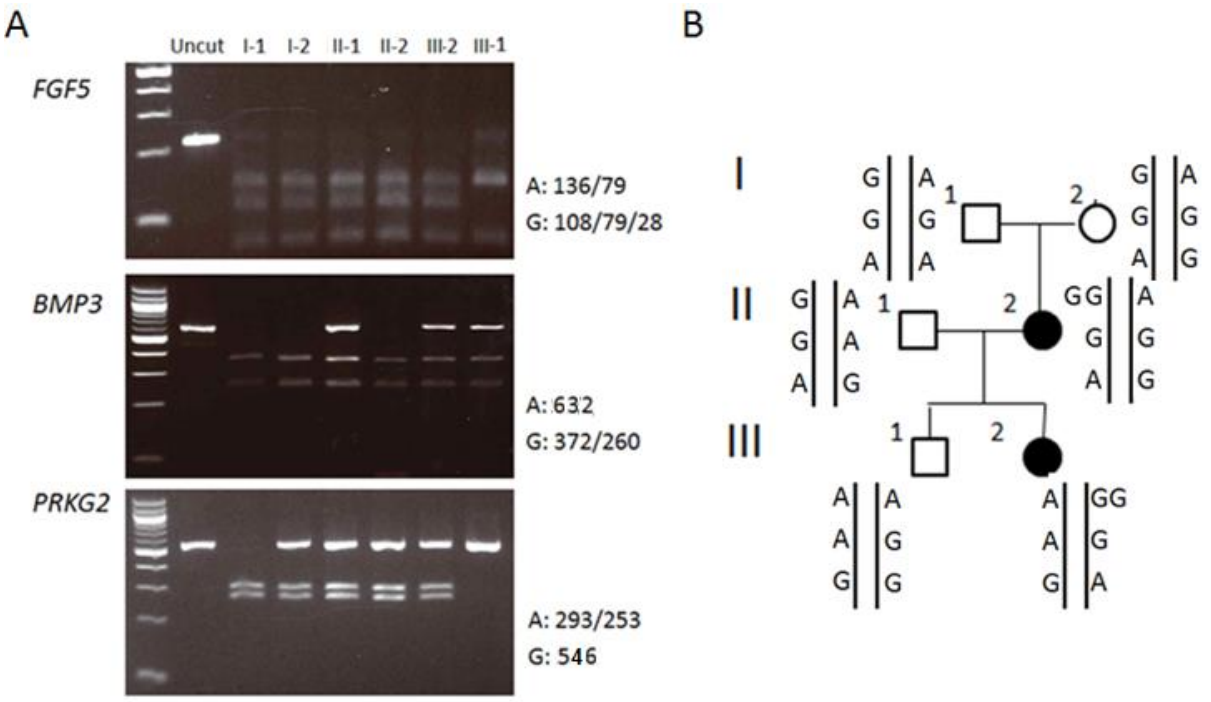


FIGURE 5-4 GENOTYPING OF INFORMATIVE SNPS IN *FGF5*, *BMP3* AND *PRKG2*

A. From the top; *TspRI* digest of *FGF5*. The enzyme cuts the A allele into fragments of 136 and 79 bp, G allele into fragments of 108, 79 and 28 bp. In II-2 and II-3, the 108 bp bands are observed as same intensity as 136 bp bands. This suggests dosage increase of G alleles in II-2 and III-2.
TaqI digest of *BMP3*. The enzyme cuts only G allele into 372 and 260 bp. The intensity of bands seen at 108bp in II-2 and III-2 are relatively stronger than bands at 136 bp, suggesting the G allele is duplicated.
MfeI digest of *PRKG2*. The enzyme cuts only A allele into 293 and 253 bp.
B. Family tree with the haplotypes of rs3733336 on *FGF5*, rs3733549 on *BMP3* and rs17005080 on *PRKG2*, based on the results in panel A.

5.3.5 QUALITATIVE ANALYSIS OF ALLELIC RATIOS OF *FGF5, BMP3* AND *PRKG2*

The above results have shown that the exact molecular nature of the duplication as a tandem duplication and its segregation is consistent with the pattern of affected individuals in the family, so the next question was whether I would gather functional evidence that the duplication is altering the function of cranial sutures. The most likely effect will be on expression, either directly for genes within the duplicated segment or indirectly by altering the regulation of genes outside the duplication (Franke et al., 2016; Lupianez et al., 2015). To quantify the expression level of the genes around the duplicated region, I decided to look at the allelic ratio in which one allele serves as a

control for the other. This was possible for 3 of 4 genes in the region of the duplication by using informative SNPs in *FGF5*, *BMP3* and *PRKG2* (Table 5-3). No informative SNPs were identified in *C4ORF22* from the exome sequencing. DNA and RNA extracted from skin fibroblasts and EBV transformed lymphoblastoid cell lines from III-2 were used for the experiment.

Allelic ratios of *FGF5, BMP3* and *PRKG2* in both genomic DNA (gDNA) and cDNA were compared between III-2 and control (Figure 5-5). Qualitatively, the duplicated G allele of *FGF5* from III-2 was seen to be stronger than the A allele in both genomic gDNA and cDNA compared to control samples, while *PRKG2* alleles in DNA and cDNA were seen with almost the same ratio between III-2 and control. The G allele and the A allele of *BMP3* were qualitatively observed as a 1:1 ratio in gDNA from both the control and III-2. In cDNA, the A allele of *BMP3* cDNA was seen to be stronger both in III-2 and control in fibroblast, which was same as in control lymphoblastoid cell line, while the G allele is seen to be stronger in lymphoblastoid cDNA of III-2. In summary, allelic expression in *BMP3* between lymphoblastoid and fibroblast cell lines appeared inconsistent, whereas relative expression of alleles for *FGF5* and *PRKG2* appeared to be consistent with duplication status.

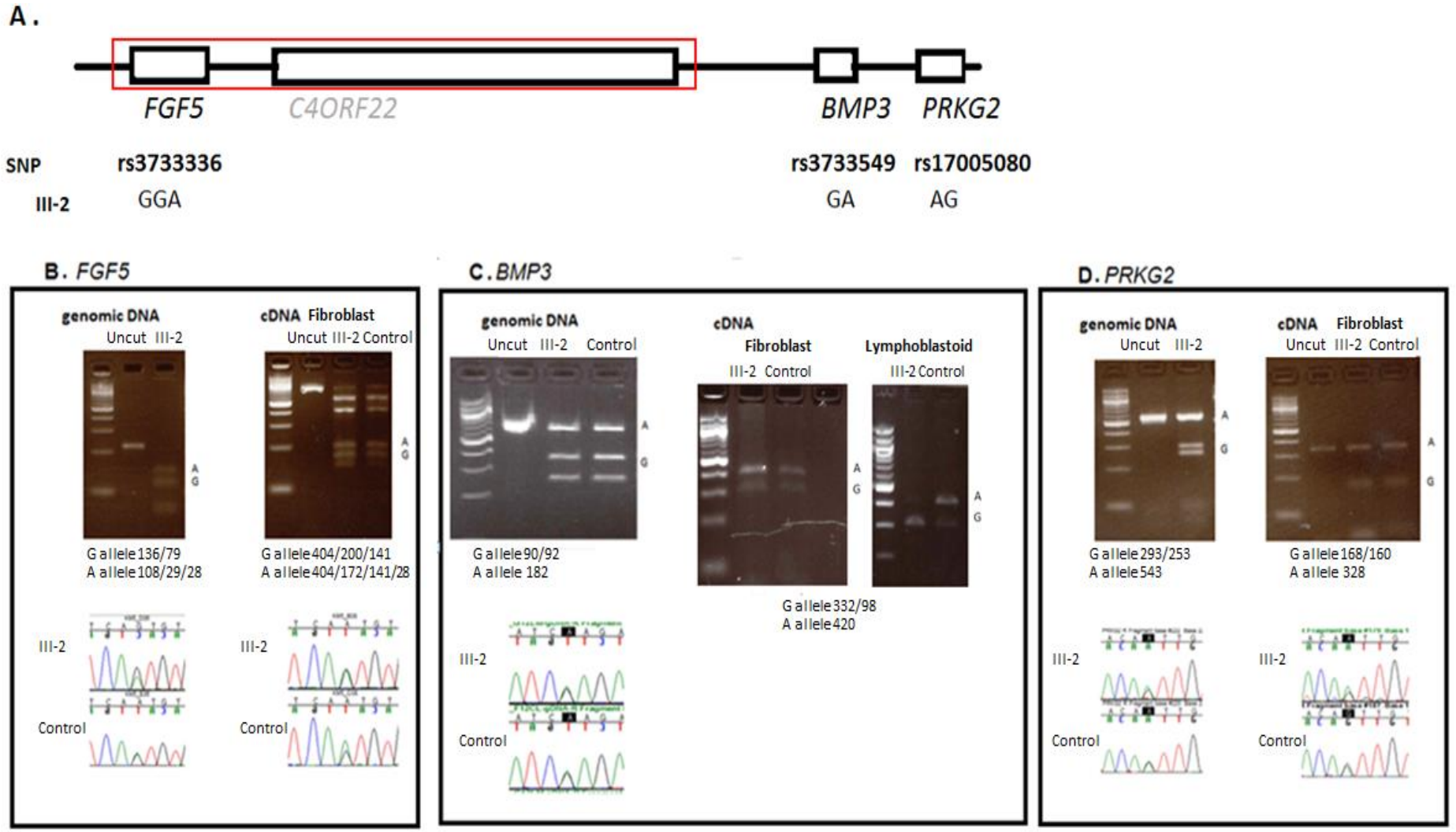


FIGURE 5-5 QUALITATIVE COMPARISON OF ALLELE EXPRESSION BETWEEN CONTROLS AND III-2 IN BOTH DNA AND CDNA USING INFORMATIVE SNPS IN *FGF5, BMP3* AND *PRKG2*

Informative SNPs shown in Table 5-3 were used to compare the allelic expression. Restriction enzymes for each gene were; TspRI for *FGF5, TaqI* for *BMP3* and *MfeI* for *PRKG2*. A. shows the duplication and the position of *FGF5, BMP3* and *PRKG2*. B, *FGF5* genomic DNA: The duplicated G allele appears stronger than the A allele in both the digest and the chromatogram (III-2). *FGF5* cDNA: The duplicated G allele was stronger than the A allele in the digest of III-2, while the A allele is stronger in the digest of control. The chromatograms show the higher A peak in control, while same peak heights were seen in III-2. C, *BMP3* genomic DNA: The A allele and G alleles appear in the same ratio both in the digest and the chromatogram. *BMP3* cDNA: Fibroblast -The A allele (uncut band) appears stronger than G allele in both III-2 and Control. EBV-The A allele was not seen in III-2, while a clear uncut band was seen in control cDNA. D, *PRKG2* genomic DNA: The ratio of A allele and G allele was estimated as the almost same level in the digest. The same peak height was seen in the chromatograms. *PRKG2* cDNA: The bands seen in III-2 and control digest were an almost same ratio. The peak heights of the G allele and the A allele between III-2 and control were seen at the same level.

5.3.6 QUANTITATIVE ANALYSIS OF ALLELIC RATIOS OF *FGF5*, *BMP3* AND *PRKG2*

Following the qualitative analysis in section 5.3.5, I measured the allelic ratio quantitatively by deep sequencing using the Ion Torrent PGM system. Allelic ratios of *FGF5, BMP3* and *PRKG2* both in DNA and cDNA were compared between III-2 and controls. The informative SNPs in Table 5-3 were used.

The raw data are presented in Table 5-4 and results are summarised in Figure 5-6. The allelic ratio of genomic *BMP3* and genomic *PRKG2* were almost equal both in proband (III-2) and controls, while the G allele of genomic *FGF5* in III-2 was counted about 65% of the total alleles. This was considered to reflects the duplication in III-2. The allelic ratios in genomic *FGF5* in controls were approximately 50:50.

In cDNA, the allelic ratios in *BMP3* showed the opposite results between fibroblast and lymphoblastoid cells with the majority of cDNA from fibroblast was the A allele while the G allele was the majority allele in lymphoblastoid cDNA. In control cDNA, both fibroblast and lymphoblastoid cells showed more of the A allele compared to the G allele. In *FGF5*, the A allele counts were higher than the G allele both in III-2 and controls, approximately 60%. The allelic ratios in *PRKG2* were inconsistent in control cells. As a result, the allelic ratios of cDNA were difficult to interpret as there was no consistency between the cell types in *BMP3,* in addition to the variability in *PRKG2* and no reflection of the duplication in *FGF5* expression.

TABLE 5-4 THE RESULTS OF DEEP SEQUENCING

***BMP3* genomic** rs3733336

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2	1	5111	2442	47.8	2669	52.2
	2	15925	7439	46.7	8486	53.2
	3					
Control 1	1	9909	4956	50.0	4953	49.9
	2	13200	6327	47.9	6873	52.1
	3					
Control 2	1	10091	4884	48.4	5207	51.6
	2	28785	13382	46.5	15403	53.5
	3					

***BMP3* cDNA**

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2 Fibroblast	1					
	2	8410	107	1.3	8303	98.7
	3	15602	1278	8.2	14324	91.8
Control 1 Fibroblast	1	1899	386	20.3	1513	79.7
	2	9117	2386	26.2	6731	73.8
	3	8767	3630	41.4	5137	58.6
III-2 Lymphoblastoid	1	8453	6255	74.0	2198	26.0
	2	13705	11927	87.0	1778	13.0
	3	3464	2179	62.9	1285	37.1
Control 2 Lymphoblastoid	1	3814	1902	49.9	1912	50.1
	2	2199	556	25.3	1643	74.7
	3	9457	2022	21.4	7435	78.6

***FGF5* genomic** rs3733549

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2	1	6617	4281	64.7	2336	35.3
	2	13684	8490	62.0	5194	38.0
	3					
Control 1	1	9392	4535	48.3	4857	51.7
	2	10439	4910	47.0	5529	53.0
	3					
Control 2	1					
	2	8700	3847	44.2	4853	55.8
	3					

***FGF5* cDNA**

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2 Fibroblast	1	9041	4021	44.5	5020	55.5
	2	9068	4054	44.7	5014	55.3
	3	17690	8095	45.8	9595	54.2
Control 1 Fibroblast	1	6234	2609	41.9	3625	58.1
	2	27694	10709	38.7	16985	61.3
	3	15526	6262	40.3	9264	59.7
Control 2 Fibroblast	1					
	2	6750	2783	41.2	3967	58.8
	3	11355	4515	39.8	6840	60.2

***PRKG2* genomic** rs17005080

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2	1	8675	3566	41.1	5109	58.9
	2	8557	3897	45.5	4660	54.4
	3					
Control 1	1	2050	1123	54.8	927	45.2
	2	8639	3906	45.2	4733	54.8
	3					
Control 2	1					
	2	6829	3110	45.5	3719	54.5
	3					
Control 3	1					
	2	6442	3109	48.3	3333	51.7
	3					

***PRKG2* cDNA**

Case	Run	total reads	G		A	
			reads	(%)	reads	(%)
III-2 Fibroblast	1	9042	5040	55.7	4002	44.3
	2	2882	1738	60.3	1144	39.7
	3	15914	2196	13.8	13718	86.2
Control 1 Fibroblast	1	6234	2609	41.9	3625	58.1
	2	931	183	19.7	748	80.3
	3					
Control 2 Fibroblast	1					
	2	2422	1401	57.8	1021	42.2
	3	4902	2416	49.3	2486	50.7
Control 3 Fibroblast	1					
	2	671	671	100.0	0	0.0
	3					

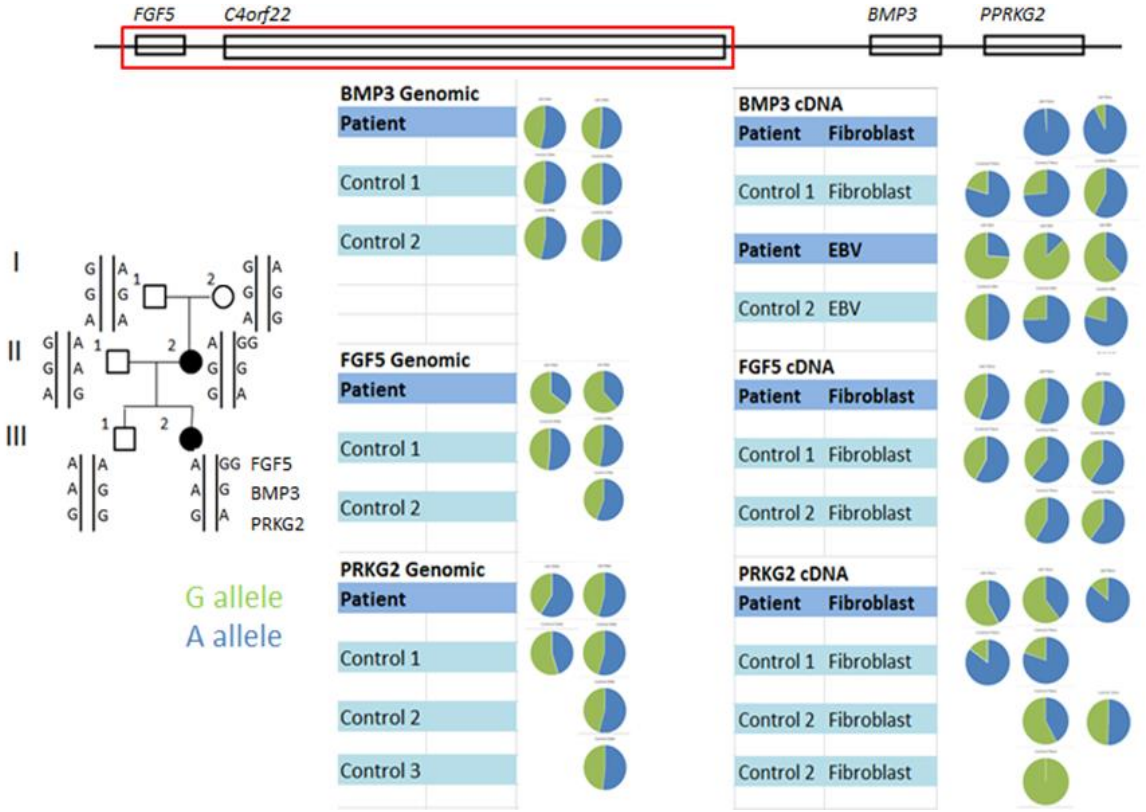


FIGURE 5-6 COMPARISON OF ALLELE EXPRESSION BETWEEN CONTROL AND III-2 IN BOTH DNA AND CDNA USING INFORMATIVE SNPS IN *FGF5, BMP3* AND *PRKG2.*

Top panel; schematic diagram of the duplication.
Bottom left; genotype of three SNPs on *FGF5, BMP3, PRKG2*
Bottom right; the pie charts illustrate the allelic ratios of the G allele (green) and the A allele (blue). Patient, individual III-2. EBV, Lymphoblastoid cell line. The experiments were undertaken 1 to 3 times as the number of the pie charts.

5.3.7 qPCR TO EVALUATE GENE EXPRESSION

Following the ambiguous results from the comparison of allele expression in sections 5.3.5 and 5.3.6, cDNA expression of the genes (*FGF5, C4ORF22, PRKG2* and *BMP3*) in III-3 and control was measured by quantitative PCR (qPCR). Overall expression levels of 4 genes were detected with very low levels compared to gene expression of *ACTB* (Figure 5-7 A and B, data was shown in relative expression of *ACTB* x10000). The expression levels of both *BMP3* and *FGF5* were low compared to the expression levels of *C4ORF22* and *PRKG2* in both from lymphoblastoid and fibroblast cells (Expression of *FGF5* was undetectable in lymphoblastoid cells). There were no significant differences in the *BMP3* expression levels between III-2 and the control. The expression level of *FGF5* from III-2 was higher than the level of control, however the levels of *C4ORF22* were detected in almost same level (Both genes should locate in the duplication). The expression levels of *BMP3* and *FGF5* in lymphoblastoid and fibroblast cells are referred in GTEx and both were recorded relatively low. Considering these, the inconsistent allele expression results (section 5.3.6) were considered to be possibly due to the low gene expression levels in cells inducing the overestimation of the allelic balance.

5.3.8 CAPTURE-C TO EXPLORE CHANGES IN POTENTIAL REGULATORY ELEMENTS RELATED TO THE DUPLICATION

To address the question whether the duplication might disrupt *BMP3* regulation, Capture-C, which is an advanced Chromosome Conformation Capture (3C) technique to detect distal regulatory elements controlling genome expression, was undertaken (Hughes et al., 2014). As a preliminary experiment, to determine which cell type is suitable to undertake Capture-C, expression levels of *Bmp3, Fgf5* and *Prkg2* in different mouse cell types were examined by qPCR. The expression levels in mouse day 14 ES cell (E14), mouse mesenchymal stem cells (MSC) and mouse bone cells which are derived from mouse coronal suture were measured. Of these, the expression level of *Bmp3* and *Fgf5* in mouse bone cells was relatively higher than other 2 cell types, so that mouse bone cells were used for Capture-C experiments (Figure 5-7 C).

The result of Capture-C, which was undertaken by Dr. Yan Zhou at the WIMM, is shown in Figure 5-7-D. No clear peaks suggesting interaction with *Bmp3* were detected in the duplicated regions.

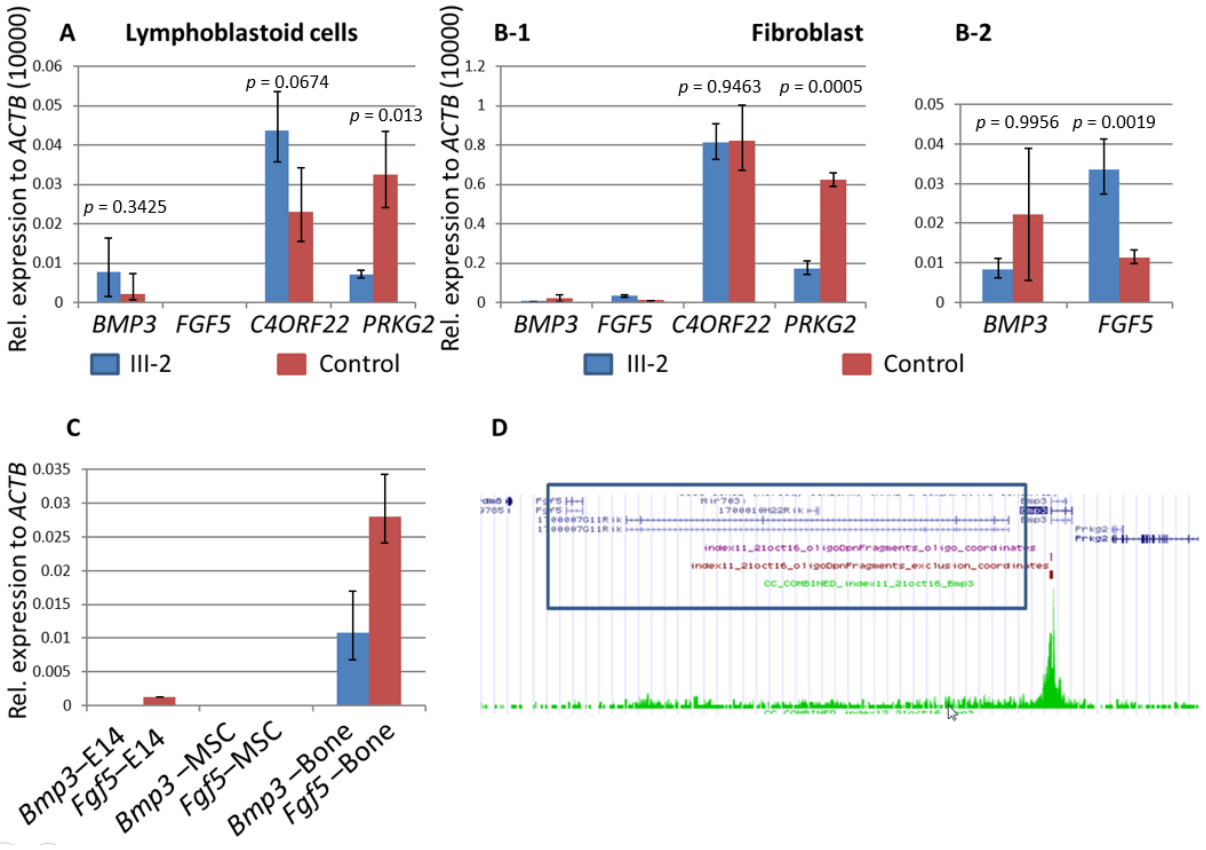


FIGURE 5-7 GENE EXPRESSION LEVELS OF GENES WITHIN OR NEAR DUPLICATION AND CAPTURE C

A Gene expression of *BMP3, FGF5, C4ORF22* and *PRKG2* in III-2 and control lymphoblastoid cells. Blue bar - III-2, Red bar – Control. Data are in triplicate and shown as relative expression to *ACTB* x10000. Expression of *FGF5* was undetectable. Error bars show standard deviations. *p* values between III-2 and Control are calculated by two-tailed *t*-test. No significant differences are observed in *BMP3* expression between III-2 and the control.

B 1 Gene expression of *BMP3, FGF5, C4ORF22* and *PRKG2* in III-2 and control fibroblast cell line. Blue bar - III-2, Red bar – Control. Data are from triplicate and shown as relative expression to *ACTB* x10000. B-2 is an expansion of *BMP3* and *FGF5* expression in B-1. Error bars show standard deviations. *p* values between III-2 and Control are calculated by two-tailed *t*-test.
Expression levels of *BMP3* and *FGF5* are low in both III-2 and the control. Expression of *FGF5* in the control is significantly higher than III-2, while expression of *C4ORF22* was detected in the same level between III-2 and the control.

C Gene expression levels of *Bmp3* (red bars) and *Fgf5* (blue bars) in mouse suture derived bone cells (Bone), mesenchymal stem cells (MSC) and E14 cells (E14), relative to *ACTB* expression. Data from triplicates.

D Capture-C from mouse suture derived bone cell (Bone). The blue rectangle indicates the equivalent position of *de novo* duplication in the family. The green signals below present CaptureC from mouse-derived bone cell from using *Bmp3* as viewpoint. No strong interaction profiles from the *Bmp3* viewpoints are seen.

5.4 DISCUSSION

A ~720 kb *de novo* duplication on chromosome 4, which is not described previously in DGV, was identified in the family. The duplication segregates with the phenotype and its molecular basis was confirmed as a tandem. Initially, the duplication causing the change in gene dosage of the downstream gene, *BMP3*, was regarded as the possible pathogenicity of this case, however, the expression analysis using the patient and control cDNA could not demonstrate significant changes in the expression level as shown in 5.3.5 and 5.3.6.

Since the Capture-C experiments in 5.3.8 did not detect clear interactions between *Bmp3* and the duplicated region, it was difficult to explain how the duplication can be pathogenic, however, a new molecular basis explaining the pathogenicity of duplication was reported by Franke et al. (2016) as described in section 1.7.3. As shown in Figure 5 8 A, the duplication of our case on chr4 locates over 2 clear TADs (Hi-C data was obtained from http://genome.ucsc.edu/cgi-bin/ (Dixon et al., 2012)). By assuming that this duplication creates a new TAD in the same manner as the report by Franke et al. (2016), the new TAD would encompass *FGF5* and *C4ORF22*. By deleting the boundary between 2 TADs, misexpression of *FGF5* (or *C4ORF22*) is possibly induced. Although, the previous Capture-C experiment using mouse suture derived bone cells did not identify any regulatory signals in the duplication, regulatory elements of *C4ORF22* or tissue specific (e.g. sutures in this case) regulatory elements could locate in the region. Based on this hypothesis and expecting the differences can be observed between III-2 and a control, Capture-C and Hi-C by using fibroblast and lymphoblastoid cell lines are currently undertaken. In terms of Capture-C, multiple probes were designed on the promoter regions of *FGF5, BMP3, C4ORF22* and *PRKG2*, and CTCF sequences around the duplication to capture structural differences.

Significant work still needs to be done to support the assumption of TAD disruption caused by the duplication. Franke et al. (2016) demonstrated the discrepancy in 4C seq profiles initially using human fibroblasts from 3 cases (a control and 2 cases with duplication presenting sex reversal or no phenotype). Thus, we expect to see different interaction profiles between III-2 and the control,

however, tissue specificity could be one of the issues, considering neither *FGF5* or *BMP3* showed high expression levels in fibroblast and lymphoblastoid cell lines (section 5.3.7). Subsequently, Franke at el 2016 demonstrated misexpression of *Kcnj2* and new TAD created by duplication by analysing the mutant mouse limb buds using Capture-C. Generating transgenic mouse carrying the duplication would be considered to overcome the tissue specificity in our case as a next step.

Another approach is currently taken by undertaking MLPA. The MLPA probes were designed to target exons of *FGF5* and *BMP3* to detect abnormal copy numbers and to screen in our craniosynostosis cohort. Finding additional patients with CNVs would support pathogenesis of this duplication and narrow down the region of interest which leads to abnormal phenotypes.

FGF5 is one of the FGF ligands, but its major role has been recognised in hair growth. Hajihosseini et al. (2002) undertook the RT-PCR expression analysis of all FGFs in mouse cranial sutures, but no *FGF5* expression was observed in mouse coronal sutures (Hajihosseini and Heath, 2002). Although, a link between *FGF5* and craniosynostosis or bone formation is not clear, some links are suggested between *FGF5* and craniofacial abnormality; a recent study, which analysed microRNA binding site SNPs in *FGFs* and *FGFRs,* reported that *FGF5*/rs3733336 is one of the susceptibility loci for non-syndromic orofacial cleft (Li et al., 2016).

The links between increased dosage of other FGFs and craniosynostosis have been reported. Grillo et al. (1998) reported that increased dosage of the *FGF3* and *FGF4* is a risk factor for craniosynostosis (Carlton et al., 1998; Grillo et al., 2014). Heterozygous *Bey* mice, which *Bey* mutation is inserted in the intragenic region between *Fgf3* and *Fgf4*, exhibit dysmorphic facial appearance and craniosynostosis. The expression analysis of cranial sutures showed up-regulation of *Fgf3* and F*gf4* (Carlton et al., 1998).

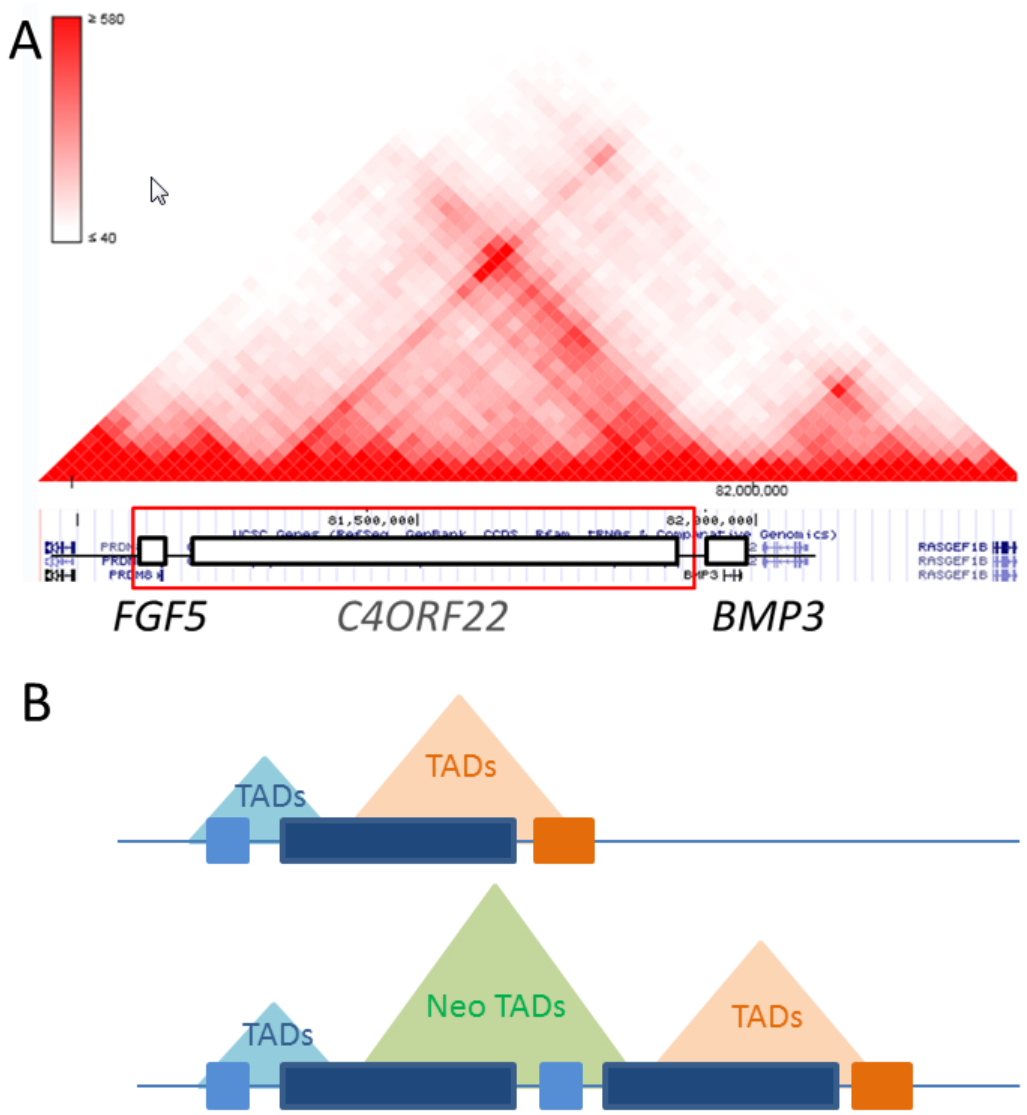


FIGURE 5-8 TAD AND THE DUPLICATION ON CHR4

A. Hi-C data of the region of the duplication in hESC show 2 TADs overlying the duplication. The duplicated region is shown in the red rectangle. The positions of genes are schematically illustrated.

B. Top panel; TADs and 3 genes (FGF5-blue, C4ORF22-Dark blue and BMP3 orange) from the panel are depicted. Bottom panel; the schematic image of tandem duplication and neo TAD created from the duplication. Neo TADs is assumed to contain genes.

Chapter 6 A SYNDROMIC CRANIOSYNOSTOSIS CASE CAUSED BY A NOVEL DEFECT OF DEHYDROGENASE/REDUCTASE 3 (*DHRS3*) IN RETINOIC ACID METABOLISM

6.1 INTRODUCTION

The family is consanguineous with first cousin parents and 2 children with syndromic craniosynostosis. The inheritance pattern was assumed as autosomal recessive and homozygous variants were searched as a candidate for the causative mutation. A combined approach using WES, WGS and SNP chip analysis identified a novel candidate gene encoding a component of retinoic acid metabolism. In this chapter, the process to identify the candidate variant and subsequent follow-up works, including resequencing of the candidate gene in the craniosynostosis cohort and functional assays of the gene are described.

6.2 CLINICAL INFORMATION

The family consists of two affected children with syndromic coronal synostosis who were born to first cousin parents of Pakistani origin. The family tree is shown in Figure 6-1. A history of 7 miscarriages in II-2 was thought due to antiphospholipid antibody syndrome (APS). There was no additional family history of craniosynostosis and the parents (II-1 and II-2) did not have craniosynostosis or other abnormal features, thus they were thought to be unaffected. III-2 was examined by a clinical geneticist and confirmed as unaffected. III-3 was born at 29 weeks of gestation and died at 1 month of age, however, parents don't know the cause of his death and details were not confirmed.

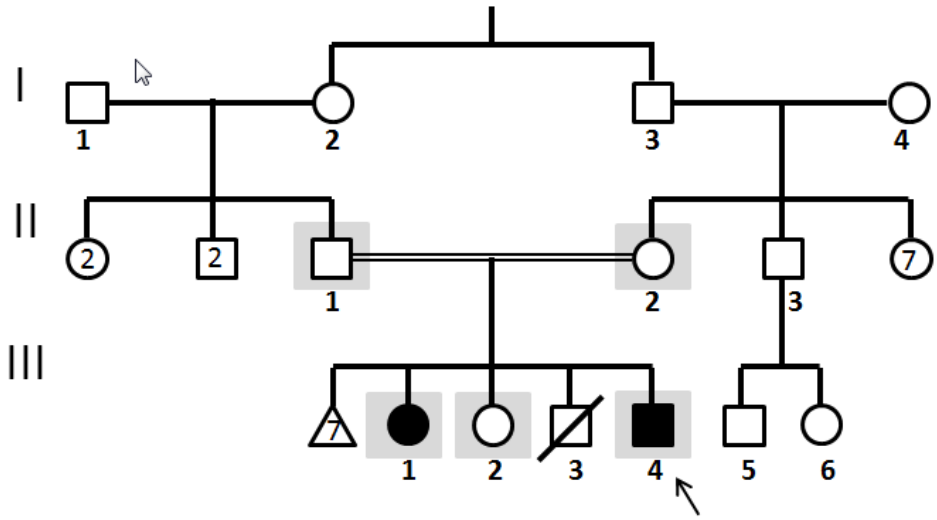


FIGURE 6-1 FAMILY TREE

Squares, male family members: circles, female family members. *Black symbols*, affected subjects; *white symbols*, unaffected subjects. *Grey shades*, DNA obtained. The *arrow* indicates the proband.

6.2.1 III-1 (SIBLING OF III-1)

During the pregnancy of III-1, her mother (II-2) was on anticoagulant therapy with heparin in the management of APS. III-1 was born at 35 weeks of gestation by a caesarean section due to poor indices on Doppler monitoring. The birth weight was 2.04 kg (9^{th}-25^{th} centile). Brachycephaly and atrial septal defect were diagnosed on day one of life. A CT scan at 1 year old showed bicoronal synostosis, possible left lambdoid synostosis and fused anterior skull base, but no brain abnormality. Clinically, she had a slightly hooked nose, a high palate and watering eyes. Her hands and feet were small with normal digits and a normal motion range of joints. She was diagnosed as having Crouzon syndrome and underwent a fronto-orbital advancement for biconal synostosis at the age of 2 years. By the age of 8 years, she had developed a class III malocclusion with midfacial hypoplasia. The atrial septal defect was closed using a catheter procedure (with Amplatzer device) by the age of 10 years. At 10 years of age, she was diagnosed with scoliosis, which required a brace treatment and posterior instrumented correction at 12 years old. She was diagnosed with psoriasis at the age of 12.

She required hearing aids temporarily, but hearing loss was thought to be caused by glue ear.

She had a mild delay in motor development during infancy, sitting independently at 1 year old, crawling at 1 year old and independently walking after 2 years old. She had a mild to moderate speech delay which required speech therapy, however, she attended normal school and adapted to school well at the age of 15 years. For her growth, she had difficulty to gain her weight. Her weight and height at 12 years old were 23.1 kg (below 0.4th centile) and 134.4 cm (2nd - 9th centile), respectively.

Conventional genetic tests were performed for *FGFR3* Pro250Arg, *FGFR2* exon 8 and 10, MLPA and sequencing for *TWIST1*, however, no causal variants were identified.

6.2.2 III-4 (PROBAND)

During the pregnancy, his mother (II-2) was on anticoagulant therapy with heparin in the management of APS. III-4 was born at 36 weeks of gestation by a caesarean section. The birth weight was 2.31 kg (9th to 25th centile) and OFC was 35.6 cm (2nd centile). At birth, he had asymmetry of the skull, prominent eyes, pointed long great toes and normal hands. A CT scan at the age of 1 month showed dilated ventricles without apparent obstruction, enlarged CSF spaces, vermian hypoplasia and right coronal synostosis. He underwent a fronto-orbital advancement for bicoronal synostosis at the age of 1 year. A small secundum ASD and a small PFO were detected at the age of 2 years. At the age of 3 years, he underwent bilateral cryptorchidism repair. The possibility of scoliosis was raised at 6 years old.

His development was mild to moderately delayed, sitting independently at the age of 1 year, crawling at the age of 2 years and independently walking at the age of 2 and a half years. He had mild to moderate speech delay which required speech therapy by the age of 5 years. At primary school, he required extra support for literacy.

He had difficulties in eating and required pureed nutrients. His poor weight gain has been a persisting concern since the age of 3 years old. At the age of 6 and a half years, his height was 110.8

cm (10th centile), weight was 14.4 kg (<1st centile), OFC was 49 cm (<1st centile) and body mass index (BMI) was 11.7.

The microarray was undertaken and a heterozygous 200 kb deletion at 22q12.3 involving *LARGE1* was detected. This deletion was identified in the unaffected father (II-1), so the deletion was considered likely to be non-pathogenic.

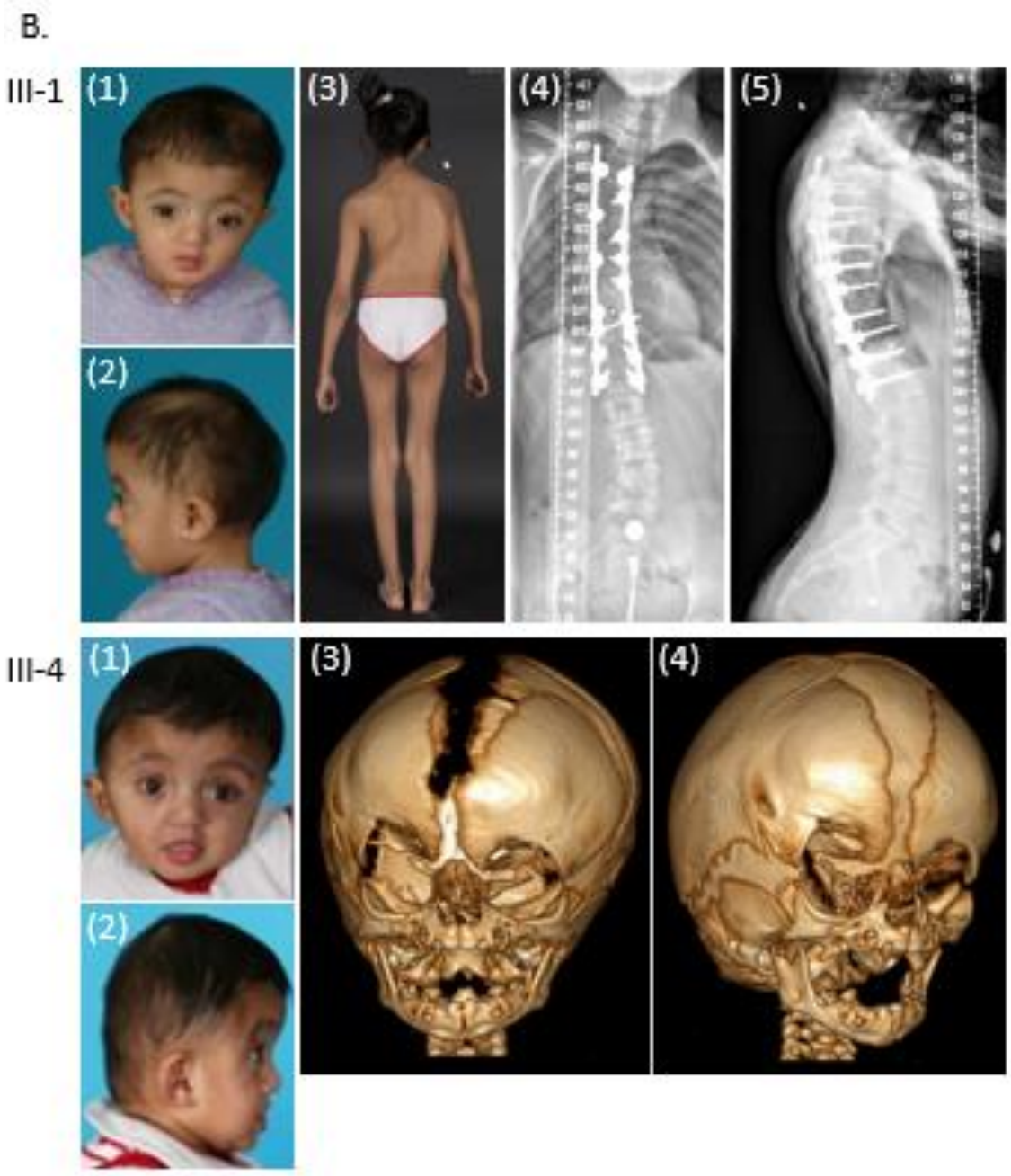


FIGURE 6-2 CLINICAL PHOTOGRAPHS OF III-1 AND III-4

III-1 - (1) and (2), clinical photographs at the age of 6 months. (3) At the age of 10 years, she was diagnosed with scoliosis and required posterior instrumented correction. 4) and 5), post-operative spine radiograph at the age of 12 years.
III-4 - (1) and (2), clinical photographs at the age of 6 months. (3) and (4), 3D-CT scan at the age of 1 months. Right coronal synostosis was observed.

6.3 RESULTS

6.3.1 EXOME SEQUENCING ON III-4

Firstly, WES on III-4 was undertaken using TruSeq and mapped to GRCh37 by Novoalign. SAMtools v0.1.16 was used for the variant calling (Table 2-1). After filtering the variants by 1000G frequency < 0.2, 6040 variants were listed, of which 609 variants were homozygous (range of variant allele frequency > 90%). Initially, a missense variant in *MMP14* (c.G269A, p.R90H) came up as a strong candidate. *MMP14* encodes a membrane-tethered metalloproteinase MT1-MMP, capable of remodelling the extracellular matrix and modulating receptors on the cell surface. The MT1-MMP acts as a critical negative modulator of ADAM9 activity to maintain FGFR2 signalling in calvarial osteogenesis and loss of MT1-MMP results in craniofacial abnormalities (Chan et al., 2012). However the affected sister III-1 did not inherit the variant (homozygous wild type) on follow-up dideoxy-sequencing. A further 25 nonsynonymous variants were prioritised after filtering using SNP frequency in ESP and 1000G (<0.005), deleterious scores (>1) and the number of total reads (>5). The genotypes of the unaffected parents (II-1 and II-2) and the affected sister (III-1) were determined by dideoxy-sequencing and this left 4 variants that were homozygous in III-1 and heterozygous in II-1 and II-2 as listed in Table 6-1. Initially, these variants were all considered as candidates, however, a sample from the unaffected sister (III-2) was subsequently obtained for testing and surprisingly all four variants were also present in homozygous state, excluding all of them. Hence, no credible candidates remained in the data on III-4 so that we decided to perform WGS on the affected sister (III-1).

TABLE 6-1 FIRST CANDIDATES FROM WES ON III-4

Gene	Variant		Position (GRCh37)	Frequency In ExAC [1]	DS[2]	Genotype by dideoxy-sequencing				
						III-4*	II-1	II-2	III-1*	III-2
MMP14	c.269G>A	p.R90H	chr14:23,311,133	0.00003296	6	AA	AT	AT	GG	-
NEO1	c.3246T>G	p.H1082Q	chr15:73,575,321	0.0006263	4	GG	GT	GT	GG	GG
PLOD1	c.535G>A	p.D179N	chr1:12,012,748	0.0005366	5	AA	AG	AG	AA	AA
SEC23IP	c.694C>T	p.P232S	chr10:121,658,469	0.0004230	3	TT	AT	AT	TT	TT
PKM	c.62T>C	p.I21T	chr15:72,511,424	0.0006515	1	CC	CT	CT	CC	CC

*indicates affected individuals.
[1] DS; deleterious score, [2] Frequency in ExAC – all of the variants had no report of homozygous cases.

6.3.2 WHOLE GENOME SEQUENCING ON III-1

Based on the hypothesis that the WES had not detected the causative variant in the family, WGS was performed on III-1. At this stage, based on WES of III-4 and WGS of III-1, shared homozygous regions could be identified. From WGS of III-1, homozygous variants were prioritised based on the deleterious scores. For each variant, the genotype of III-4 was referenced on WES results and also examined whether the variant located in the shared homozygous regions between III-1 and III-4. If the variant passed all of these conditions, the genotype of parents and unaffected sister were examined by dideoxy-sequencing. The frequencies in ESP or ExAC were referenced to exclude the variants with MAF of 0.005 or higher. Finally, two variants (1 frameshift and 1 intronic splice site change) were left on the list (Table 6-2). The variant in *ZSCAN1* was already described variant with homozygous cases. The intronic splice variant in *FAM129* was not described homozygous variants, but 10 homozygous LoF cases are reported in ExAC. Therefore, both variants were considered unlikely to be pathogenic.

TABLE 6-2. CANDIDATE VARIANTS FROM WGS ON III-1

Gene	Ref [1]	Var [2]	change	Position (GRCh38)	Frequency In gnomAD	Homo. [3]	DS [4]	Genotype by dideoxy-sequencing				
								III-1 *	III-4 *	II-1	II-2	III-2
ZSCAN1	C	-	Frameshift deletion	Chr19: 58,038,298	1069/148752	14	6	-/-	-/-	C/-	C/-	C/-
FAM129C	-	T	Intronic splice site	chr19: 17,552,096	Not described (T=0.3179/1592 (1000G)	10 LoF cases	6	T/T	T/T	-/T	-/T	-/T

1, Ref; reference genotype, 2, Var; variant, 3, Homo,; The number of the homozygous variants.
4, DS; deleterious score, * indicates affected individuals.

6.3.3 SNP CHIP ARRAY TO IDENTIFY HOMOZYGOUS DELETIONS

Because no candidate variant was identified from the WES and the WGS, a high-resolution SNP array using Illumina Infinium Omini2.5 Exome v1.2 (Total number of markers; 2,612,357) was carried out on III-1 to search for copy number variants. Nexus Copy Number was used for the analysis under the supervision of Dr. Samantha Knight.

As homozygous variants were assumed more likely to be causative, homozygous deletions were examined in the first instance. Twenty-seven homozygous deletions were identified (Table 6-3). These 27 deletions were evaluated by using the following three criteria;

1. The deletions locate in shared homozygous regions in both III-1 and III-4.
2. If the homozygous deletion was detected in WGS of III-1 (there should be no reads aligned on the regions if it is a true homozygous deletion.).
3. Homozygous deletions previously reported in DGV should be excluded.

As seen in Table 6-3, 5 homozygous deletions located in homozygous regions in both III-1 and III-4, of which 2 deletions at chr16:82,974,702-82,976,354 and chr15:63,338,895-63,339,280 were not observed in WGS of III-1 on GBrowse, suggesting that these were errors (both had been detected by only 1 or 2 probes on SNP array.) The remaining 3 deletions were examined in DGV to examine overlapped deletions previously reported and 2 deletions at chr1:12,618,127-12,623,043 and

chr16:69,007,782-69,011,396 were identified as undescribed variants. Lastly, the distance from the homozygous deletions to the nearest genes, the conservation of the region in other species and ENCODE data were examined to see whether the deletions could interrupt enhancer-promoter regions (Table 6-3 in the additional information column). As a result, a homozygous deletion at chr1:12,618,127-12,623,043 which contains the 5' untranslated region (UTR) region of *DHRS3* (Dehydrogenase/Reductase 3) came up as the top candidate (Figure 6-3). In DGV, some deletions which cover *DHRS3* are reported, but all of them are present at low frequency (Table 6-4). In ExAC, only 1 LoF variant (chr1:1,632,755 c.824+1 G>A splice donor) is reported with the allele frequency of 0.00002397 (1/41720) and no homozygotes. In ENCODE, the region of homozygous deletion shows high peaks of H3K4Me1, H3K4Me3 and H3K27Ac marks, which suggests that although the coding region of *DHRS3* remains intact, the deletion could seriously abrogate expression of *DHRS3* by removing the promoter region.

Considering the information above, *DHRS3* was thought to be a leading candidate and segregation in the family was examined (Figure 6-4).

TABLE 6-3 THE LIST OF HOMOZYGOUS DELETIONS DETECTED FROM SNP ARRAY AND FILTERING DETAILS

Position (GCRh38)	Length (bp)	No. of probes	Genes in the region	Filtering				
				Homozygous regions of		Verified in WGS of III-1	CNVs in DGV	Additional information (the distance to genes, conservation and the regulatory signals in ENCODE)
				III-1	III-4			
chr1:8295859-8306027	10169	10		yes	yes	yes	yes	intergenic, 16kb away from *SLC45A1*, conserved down to zebrafish
chr1:12618127-12623043	**4917**	**10**		**yes**	**yes**	**Yes**	**no exact match**	**5'UTR of *DHRS3*, partially conserved down to *x.tropicalis* High H3K4Me1, H3K4Me3 and H327Ac in ENCODE**
chr1:14580568-14582946	2379	1		no	no			
chr2:17574696-17575348	653	1	*VSNL1*	no	no			
chr2:55288607-55289358	752	2	*CCDC88A*	yes	no			
chr2:214780669-214780864	196	5	*BARD1*	no	no			
chr3:65132184-65132492	309	1		no	no			
chr3:194358639-194358756	118	2	*LRRC15*	yes	no			
chr3:198113116-198113277	162	2		yes	no			
chr4:92141101-92141600	500	1		yes	no			
chr5:145130840-145133264	2425	1		yes	no			
chr6:15438419-15441336	2918	2	*JARID2*	yes	no			
chr6:162379546-162379698	153	2	*PARK2*	yes	no			
chr7:156594462-156603008	8547	10	*LINC01006*	yes	no			
chr8:19093538-19093695	158	1		yes	no			
chr8:23122999-23132105	9107	10		yes	no			

chr10:77565652-77566538	887	1	*KCNMA1*	yes	no			
chr10:79790853-79851251	60399	1	*NUTM2B-AS1 LOC642361*	yes	no			
chr11:81788506-81810088	21583	12		no	yes			
chr14:80664561-80670434	5874	8	*CEP128*	yes	no			
chr15:63338895-63339280	386	1	*CA12*	yes	yes	no		
chr16:69007782-69011396	3615	2	*TANGO6*	yes	yes	yes	no exact match	intronic, approx. 11.5kb away from TANGO6 exon 16, not conserved in mouse and other species
chr16:82974702-82976354	1653	2	*LOC101928417 CDH13*	yes	yes	no		
chr18:50604720-50605268	549	1	*MAPK4*	yes	no			
chr18:62498088-62500268	2181	2		yes	no			
chr19:20417027-20534411	117385	24	*ZNF826P*	yes	no			
chr22:25054828-25056756	1929	1	*KIAA1671*	yes	no			

Shades indicate reasons for excluding the variant from candidates.

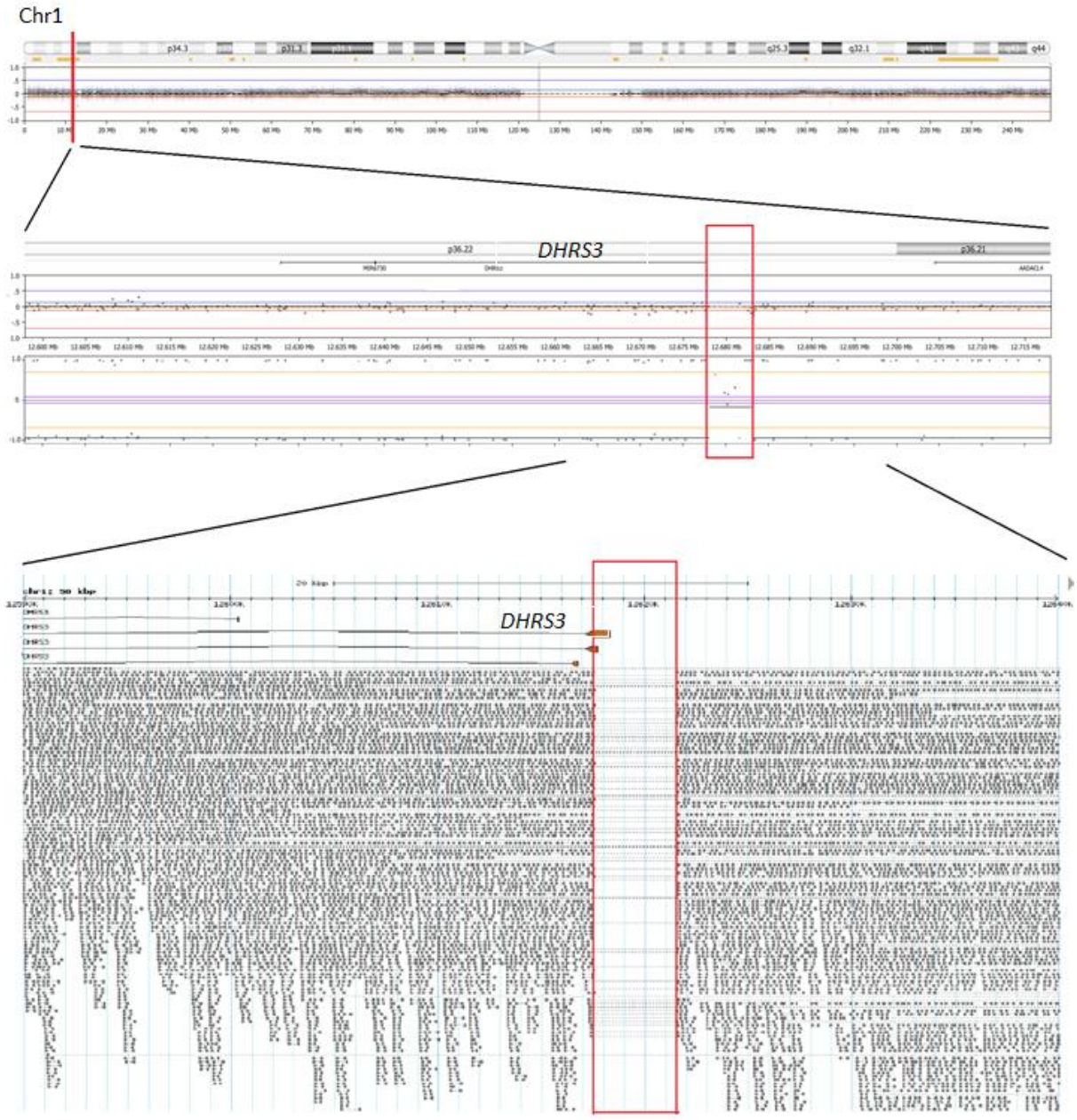


FIGURE 6-3 THE REGION OF THE HOMOZYGOUS DELETION ON CHROMOSOME 1 IN III-1

Top panel; SNP Chip analysis of III-1. View of chromosome 1. Yellow lines under chromosome indicate loss of heterozygosity regions. Red line shows the location of the deletion.
Middle panel; SNP Chip of the deleted region, expansion from the top panel. *DHRS3* gene locates in the middle. Red rectangle shows the region of the deletion. Ten points are missing in the panel of Log R Ratio (top). B allele frequency (bottom) shows random plots in the region of the deletion.
Bottom panel; WGS analysis of III-1. *DHRS3* exon 1 is shown in red blocks in the middle. Black plots indicate the reads outcome from WGS. Rectangle shows the region of homozygous deletion in the patient. Note multiple read pairs spanning the deleted region (joined by dotted lines).

TABLE 6-4 DELETIONS OVERLAPPING *DHRS3* REGION IN DGV

ID	Chromosome	Start	End	Size (bp)	Detected by	Frequency
nsv 521737	chr1	12539396	12624831	85435	SNP array	1/2026
nsv466172	chr1	12565072	12610391	45319	SNP array	1/1557
nsv545442	chr1	12565072	12610391	45319	SNP array	1/17421
nsv10011540	chr1	12601623	12832443	230820	SNP array	1/29084

No homozygous deletions were described.

6.3.4 SEGREGATION OF DHRS3 5'-UTR DELETION IN THE FAMILY

Breakpoints of deletion were examined from WGS. The segregation of the deletion in the 5'UTR region of *DHRS3* was examined within the family by using 3 primers; a forward primer was designed upstream of the deletion (F1) and the first reverse primer was downstream of the deletion (R1) while the second reverse primer was designed inside the deletion (R2). The result (Figure 6-4 A) showed only affected siblings (III-1 and III-4) carried the deletion in the homozygous state, while the parents (II-1 and II-2) and the unaffected sibling (III-1) were heterozygous for the deletion. The breakpoint was identified with 2 bp ambiguity because of a shared C-C dinucleotide at the breakpoint as chr1:12,617,576_12,621,501del (Figure 6-4 B) with non-homologous end joining (or microhomology-mediated break-induced replication MMBIR). The deletion was 3,926 bp in size.

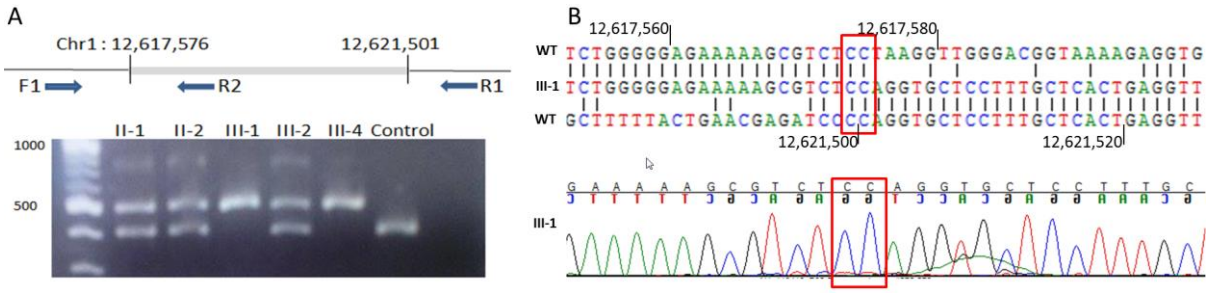


FIGURE 6-4 SEGREGATION ANALYSIS OF THE DELETION IN *DHRS3*

A. Upper panel; Schematic location of the primers. The thick grey line between Chr1:12,617,576 and 12,621,501 indicates the deleted region. Blue arrows (F1, R1 and R2) indicate primer positions and directions. The product size between F1 and R1 with deletion is 506 bp and the product size between F1 and R2 is 413 bp.
Lower panel; The PCR result using F1, R1 and R2 primers. Only the 506 bp band was seen in III-1 and III-4, while both the 506 bp and 413 bp bands were observed in II-1, II-2 and III-3, suggesting III-1 and III-4 are homozygous for the deletion allele and II-1, II-2 and III-3 are heterozygous for the deletion. Control showed only a normal 413 bp fragment.
B.Dideoxy-sequencing of the PCR product from III-1.
Upper panel; alignment of III-1 was aligned with reference sequences. The red rectangle shows a break point with two base pair ambiguity.
Lower panel; chromatogram. The red rectangle indicates the breakpoint.
(GRCh38)

6.3.5 *DHRS3* AS A CANDIDATE GENE

DHRS3 (dehydrogenase/reductase 3) encodes a member of the short-chain dehydrogenase/reductase and related enzymes (SDR) family, which constitutes one of the largest enzyme superfamilies. This family is involved in the metabolism of a large variety of compounds, however, to date, only all-trans retinaldehyde has been identified as a DHRS3 substrate (Lundova et al., 2015; Persson et al., 2009). In retinoid signalling, all-trans retinoic acid (atRA) is a major active cellular retinoid metabolite. AtRA is synthesised from cellular stores of all-trans retinol in a two-step reaction; the first step is the oxidation of retinol (vitamin A) to retinaldehyde and the second step is the oxidation of retinaldehyde to atRA. The first step is reversible and determines the overall rate of atRA synthesis, while the second step is irreversible. DHRS3 has a role to catalyze the reverse reaction from retinaldehyde to retinol in the pathway (Adams et al., 2014; Blomhoff and Blomhoff, 2006; Kedishvili, 2013). Retinoic acid signalling plays an important role during vertebrate development from cell proliferation and differentiation to embryonic development and organogenesis (Cunningham and Duester, 2015; Mammadova et al., 2016; Rhinn and Dolle, 2012).

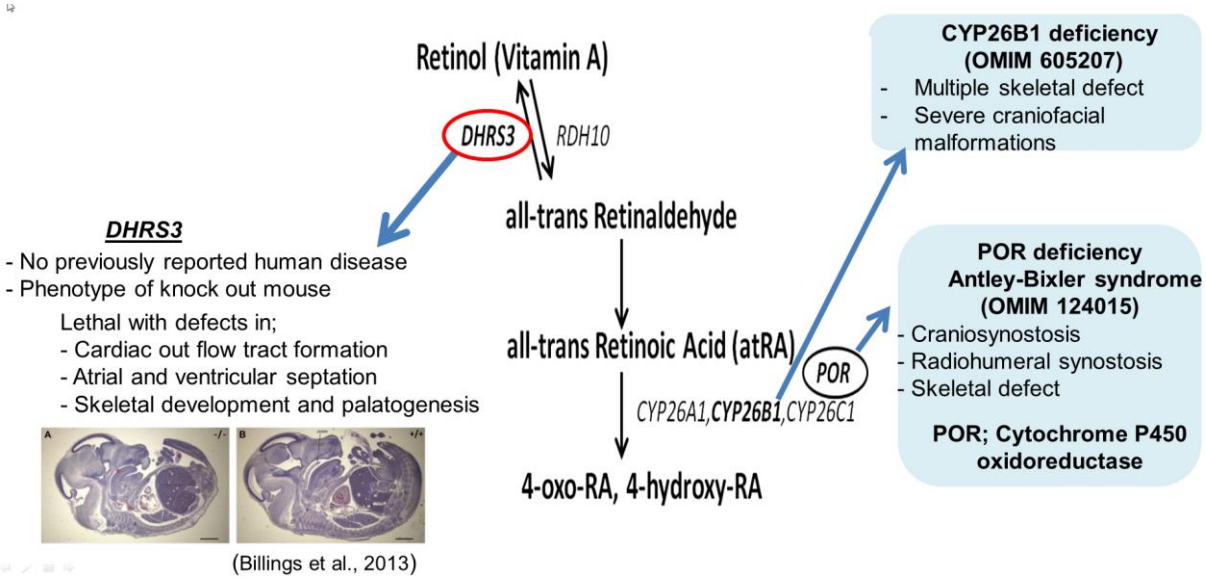


FIGURE 6-5 RETINOIC ACID METABOLISM AND RELATING PHENOTYPES IN HUMAN AND MOUSE

Retinoic metabolism is illustrated with involved genes. *DHRS3* catalyze the reverse reaction from retinaldehyde to retinol. Left bottom pictures show the phenotype of *DHRS3* $^{-/-}$ mouse. On the right-hand side, 2 human diseases mutations, PORD and CYP26 B1, are shown. By disrupting ATRA degrading, the level of ATRA increases in those 2 diseases.

No human disease caused by DHRS3 deficiency was reported so far, however, there are good biological links between *DHRS3* and craniosynostosis and other phenotypes present in the two affected siblings in the family. In the Dhrs3 deficient mouse, *Dhrs3*$^{-/-}$ embryos showed an increased level of atRA associated with alterations in the expression of several developmentally important atRA target genes. The Dhrs3 deficient mouse was lethal late in gestation with defects in cardiac outflow tract formation, atrial and ventricular septation, skeletal development and palatogenesis (Billings et al., 2013). In *Xenopus*, knockdown of *Dhrs3* resulted in a phenotype of shortened anteroposterior axis, reduced head structure and perturbed somitogenesis, which were also found in embryos treated with an excess of atRA (Kam et al., 2013).

Two human disorders, Cytochrome P450 oxidoreductase (POR) deficiency and CYP26B1 deficiency, both of which involve elevated retinoid acid by disturbing the atRA degradation pathway (Fukami et al., 2010; Reijntjes et al., 2007; Yashiro et al., 2004), have overlapping phenotypes with our case.

POR deficiency (PORD) is a disorder of steroidogenesis with a phenotypic spectrum ranging from cortisol deficiency to skeletal phenotype. The cases characterised by craniofacial abnormalities (craniosynostosis and midface hypoplasia), radiohumeral synostosis, and multiple joint contractures are referred to ABS. Fluck et al. (2004) reported the first PORD cases presenting congenital adrenal hyperplasia with apparent combined CYP17 and CYP21 deficiency (Fluck et al., 2004). POR is a membrane-bound protein and catalyses electron transfer from NADPH to all microsomal cytochromes P450, such as CYP1-3 in Hepatic phase 1 drug metabolism, CYP17,18 and 21 in steroid synthesis and CYP26 and 51 for retinoic acid metabolism and sterol synthesis (Tomalik-Scharte et al., 2010).

The CYP26 family of enzymes (CYP26A1, B1, and C1) control catabolism of atRA (Pennimpede et al., 2010). Cyp26a1 knock out mice are lethal during mid-late gestation and show a number of major morphogenetic defects, especially in caudal regions (spina bifida, and/or sirenomelia, and in severe cases caudal regression) (Abu-Abed et al., 2001). Cyp26b1 knock out mice are neonatal lethal and foetuses exhibit severe limb malformation and craniofacial abnormalities (MacLean et al., 2007; Yashiro et al., 2004). In humans, 3 cases with *CYP26B1* mutations have been reported. Laue et al. (2011) identified 2 homozygous nonsynonymous mutations in CYP26B1 in patients with skeletal (oligodactyly, femoral bowing, narrow thorax, small pelvic bones, and radiohumeral synostosis) and craniofacial anomalies (cranium bifidum, coronal, and lambdoid synostosis.) (Laue et al., 2011) Morton et al. (2016) reported a case with homozygous nonsynonymous mutation in *CYP26B1*, which was associated with multisuture synostosis, radiohumeral synostosis, normal bone mineral density, and apparent intellectual disability. The authors highlighted the phenotypic similarities of this case to Antley-Bixler and Pfeiffer syndromes (Morton et al., 2016).

Given these considerations, *DHRS3* was thought to be a plausible candidate gene. The entire promotor and part of 5'UTR is deleted, but the entire open reading frame remains intact. This arises the possibility of the mutation being hypomorphic, with a small amount of structurally normal protein being synthesised. To check the further evidence for pathogenicity, investigation to find additional patients with pathogenic variants in *DHRS3*, measure the gene expression among the

family and biochemical assays to measure retinoic acid distribution in patients and controls were undertaken.

6.3.6 RESEQUENCING OF *DHRS3*

6.3.6.1 SCREENING FOR THE *DHRS3* DELETION IN ASIAN SAMPLES

Since affected children's parents are originally from Pakistan, to identify additional cases with the identical deletion, the Asian samples with craniosynostosis in our patient panel were selected and evidence for the deletion was sought by using the same primers as used for the segregation analysis as described in section 6.3.4. In total 35 Asian samples were examined[15], but no heterozygous or homozygous cases with the deletion were identified.

6.3.6.2 SCREENING OF *DHRS3* IN THE CRANIOSYNOSTOSIS COHORT BY MULTIPLEX PCR

To search for additional individuals with variants in *DHRS3*, resequencing of *DHRS3* in a craniosynostosis patient cohort was undertaken by using multiplex PCR. Eight amplicons were designed over *DHRS3* 5'UTR and from exon 1 to exon 6, with a margin of at least 20 bp into introns. The primers to amplify the deletion of chr1:12,617,575-12,621,500 were also added to the multiplex PCR to detect the cases with the same deletion.

In total 432 craniosynostosis cases were resequenced, but no deletion was identified apart from the products amplified by positive control (III-1). In total 389 previously undiagnosed craniosynostosis cases were screened, of which 382 cases were screened with full coverage (all exons were covered with more than 10 reads). The phenotypes of those 382 cases were shown in Table 6-5.

[15] Screened Asian samples' ID

777	2087	3311	4565	5075	5581	5694	5850	6041	6378	6501	7343
1269	2967	4102	4578	5098	5614	5779	5861	6086	6381	6820	7515
1786	3242	4275	4586	5285	5692	5782	5863	6108	6434	7210	

TABLE 6-5 THE DETAILS OF FULLY SCREENED PATIENTS WITH FULL COVERAGE

	Non-syndromic	Syndromic	Combined
Metopic	45	20	65
Sagittal	48	28	76
Unilateral coronal	107	13	120
Bilateral coronal	12	11	23
Uni- or bilateral lambdoid	8	3	11
Multisuture	36	24	60
Syndromes[1]	0	20	20
Other or unknown	7		7
Combined	283	99	382

1 Clinically diagnosed as Carpenter (2 cases), CFNS (1 case), Crouzon (5 cases), Pfeiffer (3 cases) and Saethre-Chotzen (9 cases) syndrome.

Variants shown in Table 6-6 were detected including some rare nonsynonymous variants, but no homozygous or compound heterozygous were identified. As for the bottom 6 variants in Table 6-6, both variants in the same case were confirmed to have been inherited from one of the unaffected parents, in addition significant splicing effects were not identified using an in silico prediction tool (SSPNN).

TABLE 6-6 VARIANTS DETECTED FROM RESEQUENCING OF DHRS3IN SUBSETS WITH CRANIOSYNOSTOSIS

BC, ID	Position (GRCh37)		Base		DS *	Function or position from exon	Variant	Frequency		Comment
			Ref *	Alt *				ESP	ExAC	
1-007, 1140	chr1	12639416	T	A	6	nonsynonymous	p.N122Y			Not validated by dideoxy sequencing
1-037, 2492	chr1	12638918	C	T	5	nonsynonymous	p.V176M	8E-05	0.0002	
1-094, 2667	chr1	12628376	C	T	5	nonsynonymous	p.R301Q		3.3E-05	
1-068, 2140	chr1	12677188	G	A	3	nonsynonymous	p.R56C	0.0025	0.0024	
1-096, 2480	chr1	12677188	G	A	3	nonsynonymous	p.R56C	0.0025	0.0024	
1-165, 5457	chr1	12677188	G	A	3	nonsynonymous	p.R56C	0.0025	0.0024	
1-060, 2112	chr1	12638973	G	A	0	synonymous	p.A157A	0.0002	3.5E-05	
1-100, 6566	chr1	12639068	C	T	0	intronic, -84			0.0012	
1-002, 4944	chr1	12677087	C	T	0	intronic, +71				
1-098, 6108	chr1	12677555	C	A	0	UTR5				
2-004, 6122	chr1	12677188	G	A	3	nonsynonymous	p.R56C	0.0025	0.0024	
2-025, 7359	chr1	12677188	G	A	3	nonsynonymous	p.R56C	0.0025	0.0024	
2-252, 6303	chr1	12640590	C	A	0	synonymous	p.R100R			from unaffected father
2-252, 6303	chr1	12628482	C	T	0	intronic, -29		8E-05	1.6E-05	from unaffected father
2-005, 7522	chr1	12639308	G	A	0	intronic,+12		8E-05	0.0004	from unaffected father
2-005, 7522	chr1	12639310	A	G	0	intronic,+10		0.0105	0.0032	from unaffected father
2-033, 7544	chr1	12632681	G	T	0	intronic,+74				from unaffected father
2-033, 7544	chr1	12640408	A	G	0	intronic,+142			0.0031	from unaffected father

Ref; reference, Alt; the base altered to (Variant), DS; deleterious Score

6.3.7 ADDITIONAL PATIENTS IN GENEMATCHER

DHRS3 was registered into GeneMatcher (https://genematcher.org/), which is a freely accessible website designed to enable connections between clinicians and researchers who share an interest in the same genes (Sobreira et al., 2015). Two potentially interesting cases with homozygous missense variants in *DHRS3* were identified (Table 6-7). We hope to collaborate with the clinician involved with case 2 and plan to verify if the variant is pathogenic. Unfortunately the clinician with case 1 was unable to recontact the family.

TABLE 6-7 CASES IDENTIFIED THROUGH GENEMATCHER

	Case 1	**Case 2**
variant	p.V171M	p.V110I
change	homozygous	homozygous
frequency in gnomAD	no	0.000158 (no homozygous cases)
consanguinity	unknown	yes
detected in	1 child	3 siblings
phenotype	Elongated facies Droopy eyelids Mild malar hypoplasia Mandibular prognathism Mild hypertelorism	Global developmental delay Speech delay
skeletal phenotype	Brachycephaly Long extremities	no
polyphen 2	probably damaging	probably damaging
conservation of amino acid	highly conserved	highly conserved

6.3.8 THE EXPRESSION LEVEL OF *DHRS3*

Given the expectation that the DHRS3 deletion would substantially reduce *DHRS3* expression, the level of *DHRS3* transcripts was quantified with quantitative RT-PCR by using RNAs extracted from whole blood and lymphoblastoid cell lines from the affected and unaffected members of the family and control samples. β-actin (*ACTB*) and hypoxanthine phosphoribosyl transferase (*HPRT*) were used as endogenous control genes. The data is shown in Table 6-8. As the expression levels between *DHRS3* and *HPRT* were relatively close compared to the levels between *DHRS3* and *ACTB*, the data normalised *HPRT* was selected to be shown in Figure 6-6.

The transcript levels of *DHRS3* in affected siblings are significantly reduced to 4-7% (III-1) and 2-3% (III-4) relative to controls both in cDNAs from whole blood and lymphoblastoid cell lines, while carrier individuals (II-2 and III-2) showed intermediate levels. Both data normalised to *ACTB* and showed similar results. Hence, these observations support the hypothesis that the *DHRS3* deletion is pathogenic.

TABLE 6-8 THE RESULT S OF REAL-TIME QPCR

RNAs from whole blood

1) *DHRS3* normalised to *ACTB*

	DHRS3 Average Ct	*ACTB* Average Ct	ΔCT *DHRS3 -ACTB*	ΔΔCT ΔCT - ΔCT Control	*DHRS3* Rel. to Control sample
Control 1	27.92±0.04	18.65±0.03	9.27±0.05	0±0.21	1.0 (0.97-1.03)
Control 2	27.54±0.02	18.56±0.02	8.98±0.03	-0.29±0.03	1.22 (1.20-1.24)
II-2	29.45±0.22	18.55±0.01	10.99±0.22	1.63±0.22	0.32 (0.27-0.37)
III-1	31.63±0.19	18.65±0.04	12.98±0.19	3.71±0.19	0.08 (0.07-0.09)
III-2	28.65±0.03	18.79±0.02	9.86±0.03	0.59±0.03	0.66 (0.65-0.68)
III-4	31.69±0.32	19.27±0.04	12.42±0.32	3.15±0.32	0.11 (0.09-0.14)

2) DHRS3 normalised to *HRPT*

	DHRS3 Average Ct	*HPRT* Average Ct	ΔCT *DHRS3 - HPRT*	ΔΔCT ΔCT - ΔCT Control	*DHRS3* Rel. to Control sample
Control 1	27.92±0.04	28.36±0..11	-0.44±0.11	0±0.11	1.0 (0.92-1.08)
Control 2	27.54±0.02	28.35±0.16	-0.82±0.15	-0.38±0.15	1.29 (1.16-1.44)
II-2	29.45±0.22	27.97±0.06	1.45±0.22	1.92±0.22	0.26 (0.28-0.38)
III-1	31.63±0.19	27.32±0.14	4.31±0.19	4.75±0.19	0.07 (0.07-0.09)
III-2	28.65±0.03	27.96±0.03	0.69±0.03	1.13±0.03	0.66 (0.65-0.68)
III-4	31.69±0.32	26.88±0.10	4.81±0.34	5.25±0.34	0.03 (0.09-0.14)

RNAs from lymphoblastoid cell line

3) *DHRS3* normalised to *ACTB*

	DHRS3 Average Ct	*ACTB* Average Ct	ΔCT *DHRS3 - ACTB*	ΔΔCT ΔCT - ΔCT Control	*DHRS3* Rel. to Control sample
Control 1	28.34±0.11	15.86±0.08	12.47±0.12	0±0.12	1.0 (0.91-1.10)
Control 2	28.26±0.28	16.18±0.03	12.08±0.28	-0.39±0.28	1.31 (1.08-1.60)
III-1	33.15±0.17	16.24±0.10	16.92±0.26	4.44±0.26	0.05 (0.04-0.05)
III-2	29.18±0.08	16.10±0.06	13.08±0.09	0.61±0.09	0.66 (0.61-0.70)
III-4	34.41±0.51	15.88±0.06	18.52±0.52	6.05±0.52	0.01 (0.01-0.02)

4) *DHRS3* normalised to *HPRT*

	DHRS3	*HPRT*	ΔCT	ΔΔCT	*DHRS3*
	Average Ct	Average Ct	*DHRS3 -HPRT*	ΔCT - ΔCT Control	Rel. to Control sample
Control 1	28.34±0.11	22.00±0.04	6.34±0.14	0±0.14	1.0 (0.92-1.09)
Control 2	28.26±0.28	21.62±0.02	6.43±0.28	0.30±0.28	0.81 (0.67-0.98)
III-1	33.15±0.17	22.41±0.19	10.74±0.20	4.40±0.20	0.04 (0.04-0.05)
III-2	29.18±0.08	22.22±0.03	6.96±0.10	0.62±0.10	0.65 (0.61-0.69)
III-4	34.41±0.51	22.30±0.01	12.11±0.52	5.77±0.52	0.02 (0.01-0.03)

The levels of *DHRS3* transcripts measured with quantitative RT-PCR. The transcript levels in RNAs extracted from whole blood (1) and 2)) and lymphoblastoid cell lines (3)and4)) were normalised to *HPRT*(2) and 4)) *or ACTB* (1) and 3)). The $\Delta\Delta$Ct method was used for calculating relative gene expression. Data are shown as mean±SD of triplicate wells.

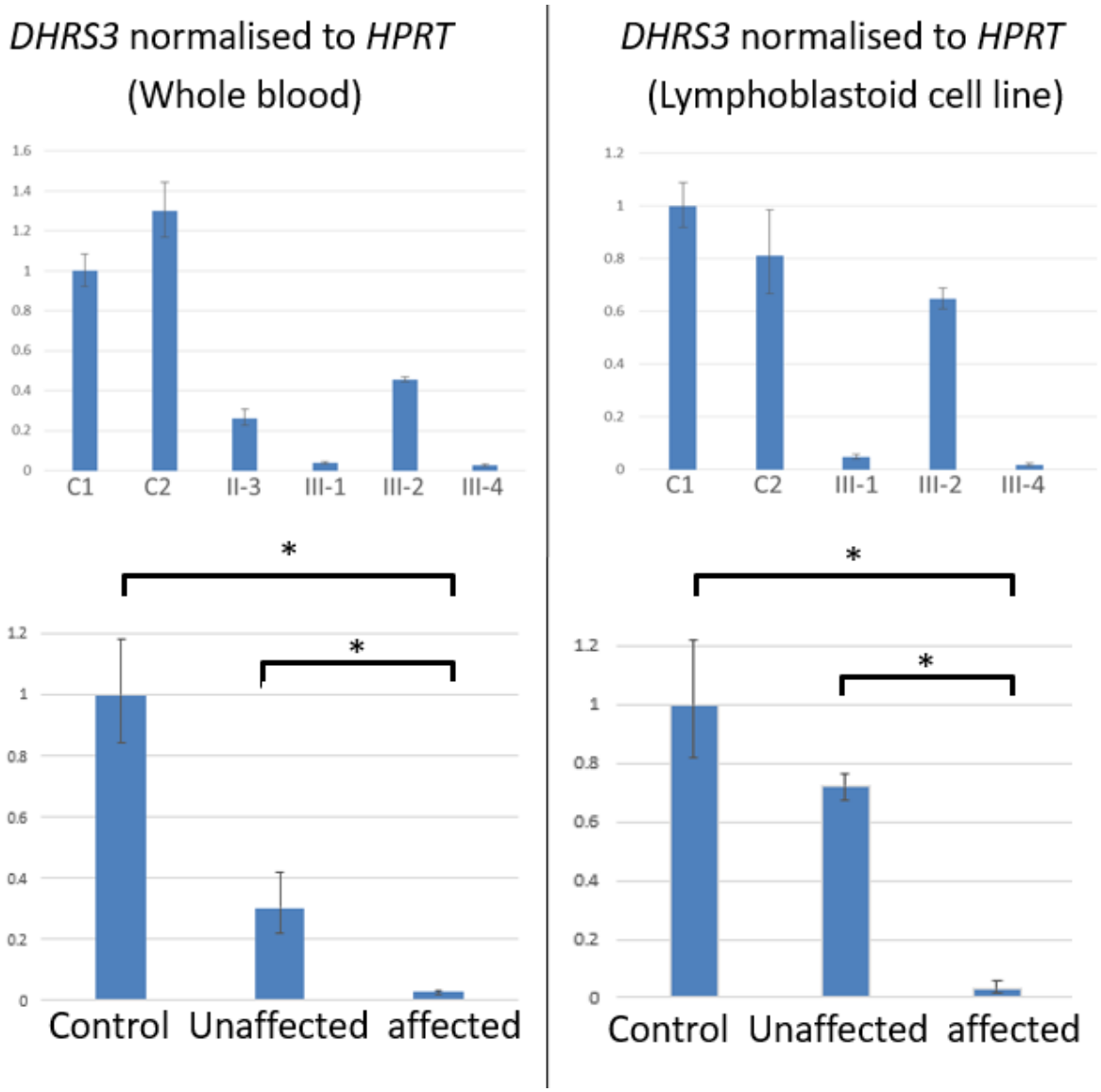


FIGURE 6-6 THE LEVELS OF DHRS3 TRANSCRIPTS IN THE FAMILY

The upper graphs indicate relative gene expression among C1 (Control 1), C2 (Control 2), II-3 (the unaffected mother, a carrier of the deletion), III-1 (the affected sister), III-2 (the unaffected sister, a carrier of the deletion) and III-4 (the affected proband) from triplicate. Error bars show standard deviations.
The lower graphs are a comparison of relative DHRS3 expression among controls (Control 1 and 2), unaffected individuals (carriers of the deletion, II-3 and III-2) and affected individuals (III-1 and III-4). Error bars show standard deviations. * $p < 0.0001$ Two-tailed *t*-test are used to calculate *p*-values. Expression levels of *DHRS3* in affected individuals are significantly low compared both controls and unaffected family members.

6.3.9 BIOCHEMICAL ASSAY TO MEASURE RETINOID INTERMEDIATES

Subsequently, retinoid intermediates were measured to assess whether there was a congruent biochemical abnormality in affected individuals. Plasma samples from the family members (II-3, II-1, II-3 and II-1) and 2 controls were obtained under conditions design to minimise exposure to ultraviolet light and the measurement of retinoid substrates was undertaken by using LC-MS/MS or HPLC/UV assays by Dr. Maureen A. Kane at School of Pharmacy, University of Maryland, USA.

Consistent with expectation, plasmas from 2 affected subjects (III-1 and III-4) demonstrated significantly increased atRA concentration (16-18 nM), while unaffected family members (II-3 and III-2, carriers of the deletion) had slightly higher levels of atRA (11-13 nM) compared to the healthy controls (9-11 nM) (Figure 6-7 and Figure 6-8-1). On the other hand, decreased retinol and retinyl esters were observed in 2 affected subjects, III-1 and III-4 (0.31-0.97 µM and 236-326 nM, respectively), while intermediate reduced levels were obtained in unaffected family members, II-3 and III-2 (1.03-1.65 µM and 253-357 nM, respectively), compared to the controls (0.05-2.76 µM and 567-697 nM, respectively).

Those findings agree with the hypothesis that 3.9 kb homozygous deletion involving *DHRS3* causes the disruption in retinoic metabolism and leads to increased levels of atRA in affected individuals. The results are consistent with observations of the *Dhrs3* deficient mouse (Billings et al., 2013).

.

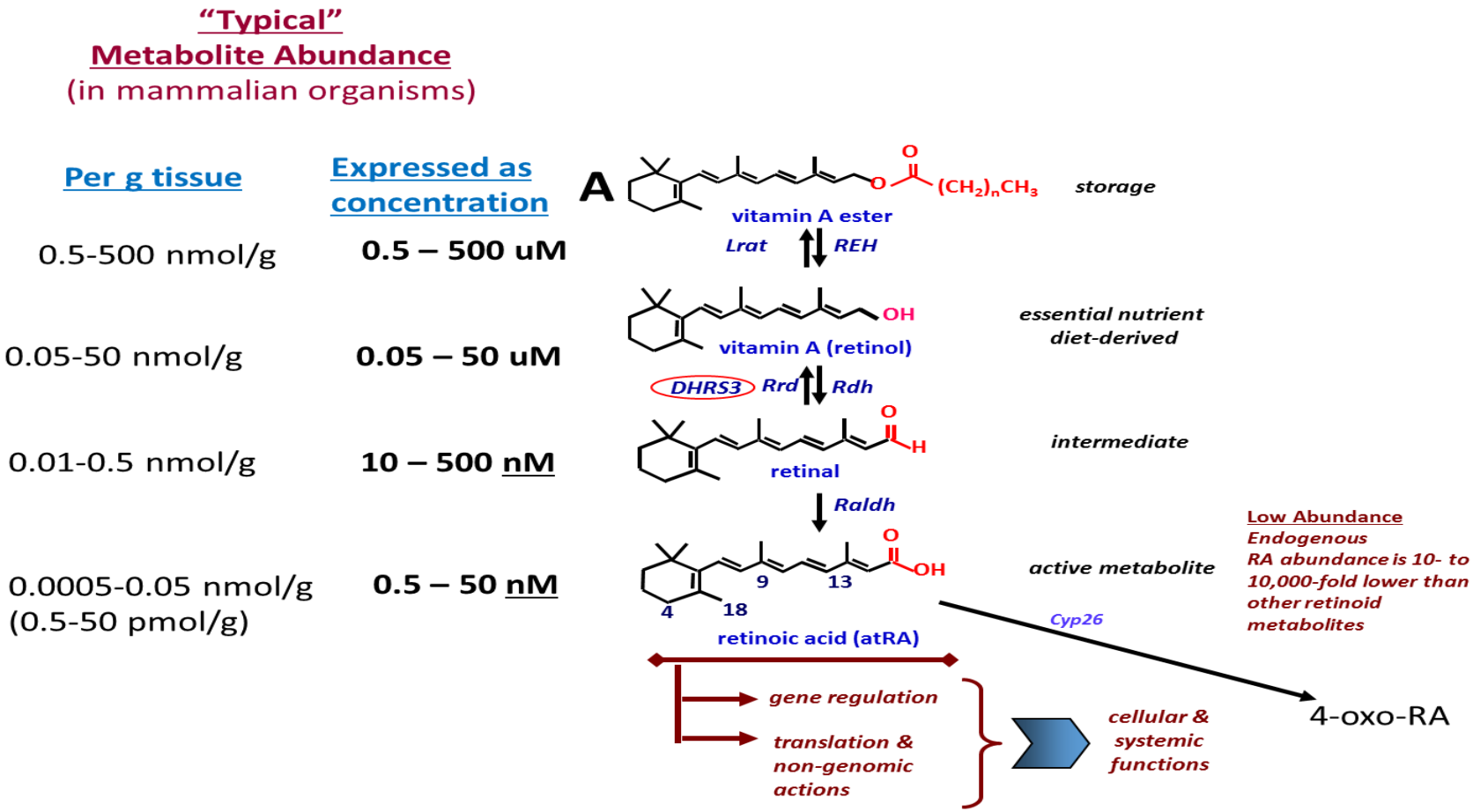


FIGURE 6-7 THE RETINOID METABOLISM AND CONCENTRATION OF EACH SUBSTRATES

Each substrate in retinoic metabolism and expected concentration in normal samples are shown. (Illustrated by Dr. Maureen A. Kane)

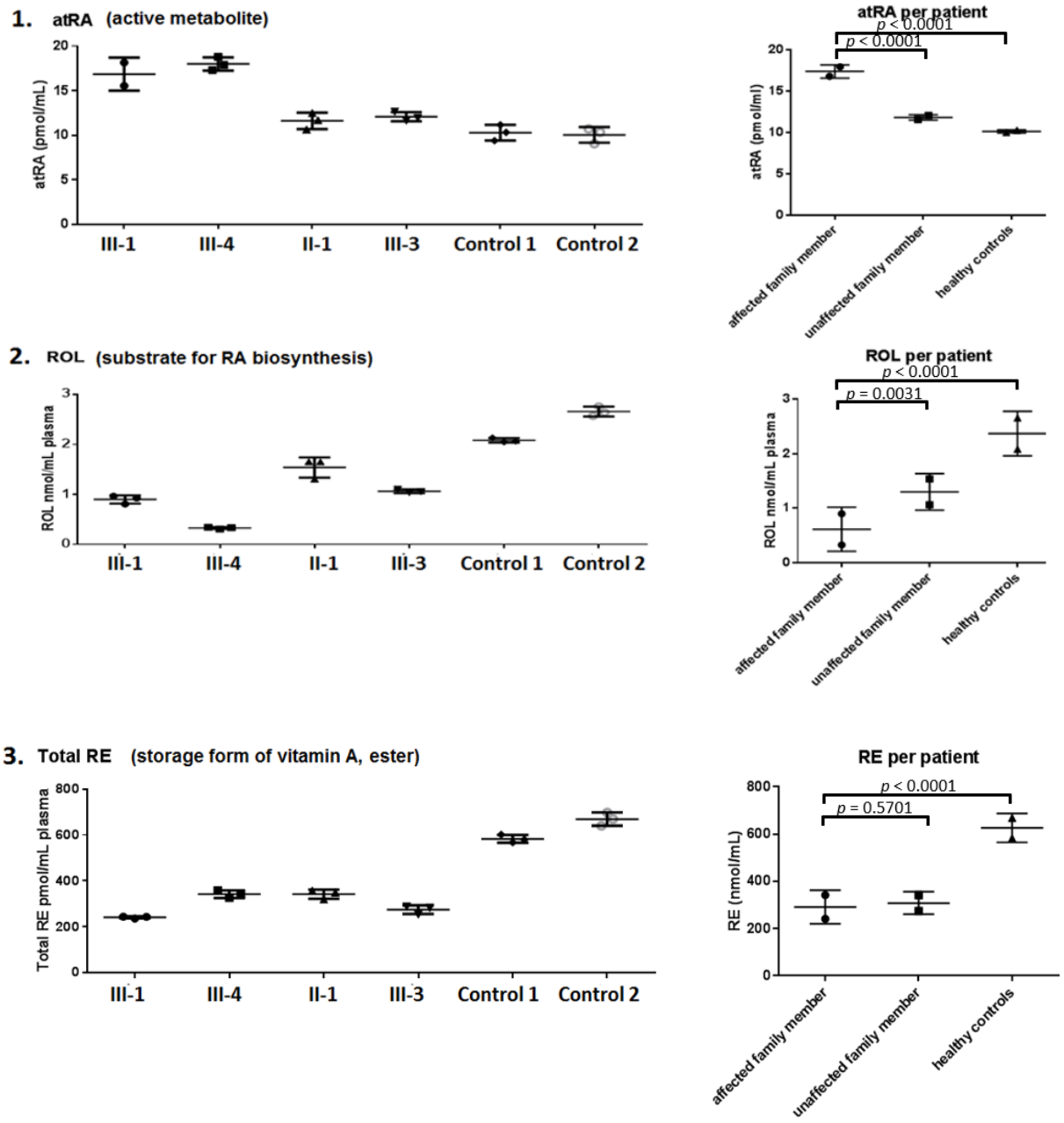


FIGURE 6-8 BIOCHEMICAL ASSAY

Each panel shows the results of LC-MS/MS or HPLC/UV assays. III-1 and III-4 are affected individuals, II-1 and III-3 are unaffected individuals in the family (heterozygous for the deletion) and Control 1 and 2 are from the normal population. Data are normalised to ml plasma where pmol/ml is equivalent to nM from triplicate wells and error bars show standard deviations.The top panels (1) show the measurement of atRA, the second panels (2) show the measurement of ROL (retinol) and the bottom panels (3) show the measurement of RE (retinyl ester).
The left panels show the results for each individual and the right panel categorises them into 3 groups; affected, unaffected family member (heterozygous) and normal controls. Two-tailed t-test is used to calculate *p* values.
(This work was undertaken by Dr. Maureen A. Kane)

6.4 DISCUSSION

The expression analysis of *DHRS3* and the biochemical assay of retinoid metabolism strongly support that the 3.9 kb homozygous deletion in the 5' UTR of *DHRS3* is a pathogenic mutation in the family. *Dhrs3* knock out mice are lethal in late gestation/perinatal (Billings et l. 2013), however, the affected subjects in the family retained 2-7 % of *DHRS3* expression because the homozygous deletion removes the promoter region of *DHRS3,* but the coding region remains intact. This indicates that the deletion is a hypomorphic allele and explains why the clinical manifestations of the two homozygous patients are not as severe as *Dhrs3*$^{-/-}$ mice.

The lethality of *Dhrs3*$^{-/-}$ mice is reported as due to cardiac malformations, including a ventricular septation defect (VSD), atrial septation defect (ASD) and a double-outlet right ventricle (DORV). The septation defect was also observed in both affected children in our case. Notably, *Dhrs3*$^{-/-}$ embryo displayed abnormality in axial skeleton development and vertebral fusion (C1 and C2), which can be corresponding to the affected sister's phenotype (severe scoliosis). For craniofacial abnormality, E17.5 *Dhrs3*$^{-/-}$ embryo displayed delayed ossification and defects in palatogenesis, instead of craniosynostosis (Bellings et al. 2013). This is an arguable point, but Cyp26b1 mutants and RA treated wild-type fish display the hyper ossification of facial and axial bones. Mineralization defects, due to increased differentiation of osteoblasts to terminal osteocytes caused by RA, are suggested as an underlying mechanism (Laue et al. 2008). The distribution and levels of RA are tightly controlled during embryonic development. For example, *Rdh10*$^{-/-}$ mouse (RDH10 synthesise all-trans retinaldehyde from retinol, reverse role to DHRS3) displays similar phenotype as *Dhrs3*$^{-/-}$ mouse including lethality in E10.5-14.5 with the craniofacial abnormality (severe nasal/facial clefting, truncated nasal process), cardiac defect and defects in organogenesis. A relation between RA and suture development are still not clearly explained and our case could give additional insight into suture development and embryopathy caused by RA.

The genes involved in retinoic metabolism were searched in CranioVar to identify cases with possible candidate variants, but no cases were identified. To date, no other congenital abnormalities are reported in retinoic metabolism apart from PORD and CYP26B1 deficiency relating to craniosynostosis, however, the genes involving retinoic metabolism should be kept in our search list.

No additional craniosynostosis cases were identified from resequencing of *DHRS3*. Two potential cases were identified through GeneMatcher. Those cases need to be followed up as those are homozygous nonsynonymous variants and it cannot be assumed that the nonsynonymous variants are deleterious. Similar expression analysis of *DHRS3* and biochemical assays are planned to assess the pathogenicity.

The deletion was not identified either by WES or initially by WGS during the investigational workup. Because the deletion was in the 5' UTR of *DHRS3*, WES did not capture the region at all. Although WGS covered the region, an initial SV analysis on WGS generated 30,000 lines of variants with nothing obvious to signal on the interesting homozygous deleted segments. Thus, it was not possible to spot on the deletion on *DHRS3*. Later, the deletion was identified successfully by the follow-up analysis by using DELLY and annotating SVs, most importantly with: the functional impact, whether the SVs was similar to a previously reported SVs in DGV and the pLI for any affected gene from ExAC. This case implies that SVs can be missed unless appropriate analysis was undertaken. Heterozygous deletions and duplications could be even more difficult to identify compared to homozygous deletions.

The 3.9 kb homozygous deletion in the 5' UTR of *DHRS3* in the family was not reported in DGV and other databases, however, this can be due to under-representation of South Asian ethnicity in this correlation the parents are of Pakistan origin. It would be interesting to know how frequent the deletion in the 5' UTR of *DHRS3* is in the selected ethnicity. In our craniosynostosis cohort, we have limited number of Asian cases. We screened the deletion in 36 Asian craniosynostosis patients, but this failed to identify additional cases.

In conclusion, these findings confirm an additional aetiology in osteogenesis in the suture caused by a disruption in retinoic metabolism following POR deficiency and CYP26B1 deficiency. The combination of genetic approaches was useful to identify the deletion.

Chapter 7 CONCLUSIONS

Various factors are known to cause craniosynostosis, including both genetic (chromosomal, monogenic, digenic, SVs) and environmental insults. In this study, 4 families with a strong suspicion of an underlying genetic cause were selected for investigation. The four families exhibited different inheritance patterns (two autosomal dominant, 1 autosomal recessive and 1 *de novo*) and multiple approaches were taken to identify causal variants including WES and WGS, combined with SNP chip, arrayCGH and segregation analysis. We initially undertook WES on one of the affected members from each family, expecting that this might yield an obvious pathogenic mutation, however, multiple methods were required to find plausible candidate variants. The genetic causes of craniosynostosis are very heterogeneous, therefore, the genetic causes in the families without a genetic diagnosis from conventional testing likely to be caused by more complicated molecular basis, such as non-coding variants or SVs, rather than single nucleotide change in coding regions.

7.1 PATHOPHYSIOLOGY OF CRANIOSYNOSTOSIS

Pathogenic or candidate variants with a different molecular basis were identified in 3 of the families; a missense substitution in a histone protein (Chapter3), a copy number variant (duplication of *FGF5* and *C4ORF22*) that may affect 3D chromatin structure (Chapter5) and a dosage effect caused by homozygous deletion in the 5' UTR of *DHRS3* (Chapter6). The underlying mechanisms would each fit into different pathways of the pathophysiology of craniosynostosis (Figure 1-3); *DHRS3* in retinoic acid metabolism*, HIST1H2AD* p.N90D into neural crest development and duplication of *FGF5* and *C4ORF22* potentially into FGF/FGFR signalling. Thus these findings confirm that perturbation of a relatively wide range of pathways can cause craniosynostosis.

7.1.1 RETINOIC METABOLISM

In chapter 6, mutation of a novel pathogenic gene (*DHRS3*) was identified in a consanguineous family with syndromic coronal synostosis. *DHRS3* expression was significantly lower than normal, and the level of retinoic acid subsequently increased in the 2 affected siblings. The phenotype partly overlaps those previously described in deficiencies of POR (Antley-Bixler syndrome) and CYP26B, both of which disrupt retinoic metabolism, likely reflecting shared pathogenesis. These findings identify a

novel embryopathy caused by excess ATRA, which leads to multiple organ defects including skeletal and cardiac anomalies. The resequencing of *DHRS3* in our craniosynostosis cohort did not identify additional patients with mutations in this gene. Considering the frequency of craniosynostosis (1 in 2,100 – 2,300 births), the fact that to cause disease both alleles should be mutated, but probably not lead to complete loss-of-function (in the mouse, knock-out is lethal), the prevalence of *DHRS3*-related craniosynostosis is likely to be very low. In such rare cases, GeneMatcher was found to be a helpful tool to find additional cases. The two homozygous missense cases identified are apparently interesting cases to be followed up. The accumulation of further cases with *DHRS3* mutations will lead to more insights into the role of retinoic acid metabolites in osteogenesis and cranial development. Furthermore, future analysis of NGS data in craniosynostosis should pay attention to interesting variants present in genes involved in retinoic acid metabolism.

7.1.2 THE NONSYNONYMOUS VARIANT IN *HIST1H2AD*

In the family with syndromic metopic synostosis and psoriasis (chapter 3), we started off from the point where we did not know whether the phenotype was caused by a single mutation or whether multiple genes were involved. *HLA-C*:0602, the high-risk allele for psoriasis, was identified as co-segregating with the phenotype, and considered likely to account for the presence of psoriasis in this family. On the other hand, *HIST1H2AD* c.268A>G, p.N90D was identified as a possible candidate for syndromic metopic synostosis. There were multiple lines of evidence supporting pathogenicity of this variant including: (1) the uniqueness of the variant (no identical change has been reported in the H2A gene family which consists of 17 copies), (2) the variant locates immediately adjacent to the acidic patch which is essential to maintain 3D chromatin structure, (3) the link between a nonsynonymous variant in another histone, *h3.3*, and a neural crest defect (Cox et al., 2012) as the metopic suture is the only suture which is exclusively derived from NCC, and (4) the recently reported cases with heterozygous missense mutation of histone H4 (Tessadori et al., 2017). As described in section 3.4, there are some overlaps between this case and our case, including the clinical phenotype and the position of the mutation, pointing to the importance of mutations in

histones in embryogenesis. . However, in the absence of a second family or case confirmation of pathogenicity requires further work and to gain this proof we aimed to generate a mouse carrying the identical mutation, *Hist1h2ad* p.N90D.

Canonical H2As are encoded by multiple highly similar genes both in human and mouse, thus targeting a single H2A gene is not straightforward. With our strategy, CRISPR/Cas9 targeting successfully introduced the variant only into the targeted gene in mouse E14 ES cells. The generation of this mouse model, as well as RNA-seq in targeted ES cells to investigate effects of the variant on gene expression, is currently ongoing. Chimeric mice have been generated, hopefully germline transmission will be achieved soon.

Due to the sequence similarity among H2A genes especially in mouse (Figure 3-8), specifically detecting *Hist1h2ad* cDNA is especially challenging. Probably for the same reason, the localization and timing of expression of each of the canonical H2A genes is still unrevealed to date. Relating to this, and as discussed in section 3.4, *HIST1H2AD* expression in mouse and human may not be completely equivalent. These could be limiting factors on the further investigation even if the human variant is pathogenic.

7.1.3 THE DE NOVO DUPLICATION ON CHR4

The *de novo* duplication on chr4 encompassing *FGF5* and *C4ORF22*, was found to co-segregate with the phenotype and is a good candidate for the cause of craniosynostosis in the studied family. Furthermore, this tandem duplication has not previously been reported in DGV. Initially, we thought that the duplication might affect the downstream *BMP3* gene expression, however gene expression analysis using fibroblast and lymphoblastoid cell lines derived from the family were inconclusive, probably due to the low levels of expression of *BMP3* and *FGF5* in these cells. Over the course of this thesis, there have been many reports on the pathogenic roles of genomic duplications and deletions, including those that disrupt TADs (Franke et al, 2016) (Will et al., 2017). To investigate whether a similar mechanism may be at play here, potentially leading to aberrant expression of FGF5,

chromosome conformation capture experiments are being undertaken to see whether the tandem duplication disrupts the TAD domain locating this region. A possible scenario leading to craniosynostosis might be disruption of the normal TAD structure resulting in ectopic expression of *FGF5* within cranial sutures during development. If the chromosome conformation experiments indicate a disrupted TAD, then analysis of mutant mouse embryos carrying the duplication could be carried out to check this hypothesis.

7.1.4 THE LARGE PEDIGREE WITH DOMINANT MULTIPLE SYNOSTOSIS

Despite considerable efforts to find the causal mutation in the multi-generation family described in chapter2, a candidate variant was not identified. Although multiple samples were analysed by WES and WGS, perhaps the different capture methods used for WES and sample availability (the most distant relation would have provided the most informative information.) limited the power of analysis. Providing that there is complete penetrance in this family, then a possible explanation for the inability to detect a mutation, is that it is non-coding (the coding genome was analyzed repeatedly by WES and WGS within the family), and potentially involves SVs potentially neutral copy, that would be hard to detect. It is still possible that a polygenic cause or gene-environmental interaction may be involved. Hopefully continuous work on this family including ongoing CNVs analysis will discover the cause, for example using 10X Genomics technology to enable phasing across very large regions of DNA to discover structural variants in previously inaccessible and complex regions of the genome.

7.2 COMBINATION OF TECHNOLOGY

In addition to adding to the craniosynostosis pathophysiological framework, the case with the deletion in *DHRS3* provided valuable lessons in interpreting NGS data and the approaches that can be taken to complicated cases. The homozygous deletion was detected by the WGS, however, the primary analysis did not annotate the deletion as a strong candidate, and it was necessary to analyse SNP array data to pinpoint it. Since WGS is widely used in a clinical and laboratory setting, CNV and SVs detection using WGS data is getting more common, and methods for analysis are improving.

Various software is developed to detect the CNVs from WGS, however, each method (described in section1.7.1) have detection limit and the combination of multiple methods or analysis are useful in the situation. SV detection using this data needs to be improved to report such candidates. Furthermore, new technologies that allow long-range phasing will undoubtedly help.

Overall, the candidate variants from the 4 families analysed in this thesis were identified by a combination of multiple approaches, rather than single WES or WGS. These cases illustrate the effectiveness of using combined genomics technologies and approaches to resolve difficult cases.

Chapter 8 APPENDIX

8.1 PRIMER SEQUENCES

8.1.1 PRIMERS USED IN CHAPTER 3

TABLE 8-1 PRIMERS TO DETECT THE VARIANTS IN TABLE 3-5

Gene	Chr	Positon	Change	Primer sequence 5'→3'		Size (bp)
				Forward	Reverse	
CHRD	chr3	184,105,253	c.2439C>G	GACAAGAAGGGGAGAGTATATAGGGGG	TATCGTTTTGATTCTTGAGCCCTCCTC	422
EVC2	chr4	5,710,119	c.122C>A	TTCTTTGCGAAATACTAAGGGTCCTCC	GGGTGTTTGTTCCAGAAAACTTTCGG	458
CRMP1	chr4	5,868,475	c.390A>G	ACAAAACAGGAAACGTGTCTGCATTTT	AAGTATTGAAGCTGGTGATGGATCTGG	528
HIST1H2BE	chr6	26,184,044	c.21delC	TGTCTACAAATTCCGAGGATCAGGAGA	AAGTTTACTTGGAGCTGGTGTACTTGG	457
*HIST1H2AD**	chr6	26,199,204	c.268A>G	AGGAATACATGGGTGGCTCTGAA	AGCGTTATTTACAATTCGACCTCCA	577
ASCC3	chr6	100,988,240	c.5574G>T	ATACCTGACATACTCTGAGAGCTTGGT	ATCTACCTCTCAGTTTCTGGTTCACTG	540
ZNF212	chr7	148,950,727	c.709A>G	TGTCATCCTCCTCCACGTTATCCTAAT	CTCTACTACCAACAGGCTCCTCTAGAG	336
ZNF16	chr8	146,156,530	c.1643C>T	TTATCTGATGCTGAATGAGGTGTGAGC	TGACTGTGGGAAAGCATTTAGTCAGAG	557
TLN1	chr9	35,699,085	c.6943G>A	TTCTGGGTATTTAAGCACTCACCTTCC	AAACCCTTATCATGGTGTGATCTCTGC	538
PRUNE2	chr9	79,323,536	c.3654T>C	GGTGTGGAAAATACCCTGGATTCTCAA	ATCAGGAAGCAAATCAGGTAGATTGGG	402
PKD1L3	chr16	72,020,171	c.781C>T	AAAGCTACCTGTGACTTGAATTTCCCA	AGAGTATTTGGGGAAATGCTGCCTTAA	366
CCDC105	chr19	15,121,692	c.55_56del	GTTCCAAAGCAAAGACTCCATCCTACT	CCTTTGATCATCTCCACGCGGAAG	431
MED26	chr19	16,688,251	c.390C>T	GATCTTGCCACTGTGCTTGTCATTCTC	CTCTGGTTGTAGGAAACACGACTTGG	528
SCN1B	chr19	35,530,094	c.522C>T	AGGCTGGGTATTAATAATCACAGTGCA	GACGACCCCTATCCTGTGTCTGAC	452
EIF3K	chr19	39,111,001	c.84C>G	TAGGCGAGAGGTCTGGGATAAGAATAA	CCAACTTCCTTGGCTTCCTAGTCTTG	442

*The primers for *HIST1H2AD* were also used for genotyping family members.

Annealing temperature: 63 °C

TABLE 8-1-2 PRIMERS TO DETECT THE VARIANTS IN VARIANTS DETECTED BY RE-SEQUENCING

Gene	Chr	Primer sequence 5'→3'		Size (bp)
		Forward	Reverse	
HIST1H2AA	chr6	AGACGGTAGGTGGCTCTGAAAAG	ATAGCAAATGCCAACCTCAGTCA	563
HIST1H2AB	chr6	AATCCAAAGCACAACAGCTCATT	ACATTTTCTTGTGGCGATTTTCC	550
HIST1H2AC	chr6	ACTACCTATAAAAGCGGCCATGT	AGCTCTTCCTTTGAAACGGTAGGG	575
HIST1H2AE	chr6	ACTAGCTTCTTATTCCTCCATCGA	CAGAAAAGGTGGGTGGCTCTGAA	554
HIST1H2AG	chr6	TTGTGGTTGCTCGTAGTGAGTTG	TGTCCATGTCTCACTTTCTTTCACA	553
HIST1H2AH	chr6	TTTTCGCGCCCAATAGTGTTTAT	GCGAAGCAAACATGAAGTGAAAC	570
HIST1H2AI	chr6	CGCGCCCAGTAGAGGCTATAAAA	ACATAGTGCTCAGTTCTTCCACA	571
HIST1H2AJ	chr6	GTGAAGCAGCTGAAAACGGACTC	CCAATCGCACAGCTTCCTTTTCG	588
HIST1H2AK	chr6	GCCATTGCAAAATGACTCGACTA	TAGCGGAGCTTTAGCGTTACGTC	580
HIST1H2AL	chr6	TTCGCGCCCAGTATTGACTATAA	AGCTCTCCCGCAGAAATAGTAGG	550
HIST1H2AM	chr6	CGTAGAAATGGTGGGTGGCTCTG	CAGCTGCTATAAAATGCGCGTCC	536
HIST2H2AB	chr1	GCGCTATTTCGTGATCAGCCCTT	CAAACGACTTCCCTTCACCCCTC	595
HIST2H2AC	chr1	GAGGTTCTGAGCGTTGTCTGTGT	TGTCAAATGAAGGGCTGATCACGA	532
HIST2H2AA3	chr1	ATTTGAAAACGTGGGTGGCTCT	GGACAATCGCCGTTCTGTCTAT	594
HIST3H2A	chr1	CGGCCTATAAAGAACTGCAGGAT	TTTTAGTCTCGCTTTTCGGTTGC	551

Annealing temperature: 60 °C

8.1.2 PRIMERS USED IN CHAPTER 4

TABLE 8-1-3 PRIMERS TO DETECT THE VARIANT IN *BMPR1B* IN SECTION 4.3.1

Gene	Chr	Positon	Change	Primer sequence 5'→3'		Size (bp)
				Forward	Reverse	
BMPR1B	chr4	96,052,582	c.995A>T	GCTTTCATTGCTGCAGATATCAAAGGG	CTAAGCAATGTCTCTGATGAGCATAGAAAC	505

Annealing temperature: 63˚C

TABLE 8-1-4 PRIMERS USED IN SECTION 4.3.5

Gene	Chr	Positon	Change	Primer sequence 5′→3′		Size (bp)
				Forward	Reverse	
ALG6	chr1	63876987	c.665G>A	TGGGGTGTTCTTTGGAATATCTTGTGAC	TCTAACTACCCAATCCACACTACCAGT	397
FAM82A1	chr2	38156623	c.203T>C	AACCAGGGATAGCAATGAAGTTACCTG	AACTTACCCTCCTTCACTTTCTGCTTC	365
NRXN1	chr2	50724758	c.259C>T	TGGTCTTGAAGGTGACAGGA	TGGCAGGTGATCATACTAGGC	231
ECEL1	chr2	233351209	c.155T>C	GAACGAGTAGAAGTCCTGGCATGG	ATAGTCCCTGGTTTCCGCAGTGAATG	662
CPVL	chr7	29152356	c.252C>A	AAATGATAGGAAGAAGACTGTGGCAGG	TTCATCCAGACTGAGATAGATGGCTCA	497
GDF6	chr8	97157413	c.746C>A	GGTGAATACCACCAGCAGGGCC	CAGACGATCTCTCGCACACTCCTCTCC	443
KANK1	chr9	744589	c.3522G>A	GAAGACCGAACGAGTAGGGATATTTGG	GCTGCTATCTGGATAATTCTGCACAGA	550
KCNV2	chr9	2718341	c.602T>A	CCAAGACGCGCCTAGGTCG	ATCTCCTCCACGGTGTTGAGC	499
CDC37L1	chr9	4688573	c.475A>G	ACCCGGCCTCATACATCATTTTAAACA	GATCTATACAGCTGTGCCAGAAGCTAC	457
GRK5	chr10	121203141	c.1143C>T	CAGAGGTCCTGAACAACCAGAGGTAC	CTCCTCGGAGAACTTGTGGGAGTAC	185
DMBT1	chr10	124390666	c.3944C>T	TGGCCATGATCTCTCAGTTAATGTGTC	TCACGATGATTACAGTTGTGGGAGAAC	331
GJA3	chr13	20717217	c.211C>G	CGAACAGCGTCTTGAAGATGATG	TGGGAAGACTCTTAGAAAATGCACAGG	459
IFT88	chr13	21218984	c.1836A>G	TAATTGAAAAGTGTGTCAAGCAAAGCT	AGAAGCTCTTTCAAAGTACTGAATAGCT	309
RNF17	chr13	25341452	c.173T>C	CATTAACACACTGTGGGCAGTTTAATCAC	CCAACTTTTTAGTTTCTAACTAGAAGTACGTTT	440
PLCB2	chr15	40590557	c.1022C>A	CATTCATCTTCTCTAGCTTCAGCCCTC	CCCTGAGAATAACACAGACCATGATGG	356
C17orf85	chr17	3721680	c.1187G>A	TTTCAACTGAGATTCCACTTCGTCAGC	ATTTCTTCAGGGCTAGTGAATGTTCCC	312

Annealing temperature: 63˚C

TABLE 8-1-5 PRIMES USED IN SECTION 4.3.5 FOR SHARED VARIANTS

Gene	Chr	Positon	Change	Primer sequence 5′ → 3′		Size (bp)
				Forward	Reverse	
SPTBN1	chr2	54,856,036	c.2712C>T	TCACTTTGTCTGCGGTTTGG	TTCTCCCGTATCCAGCCTTC	400
CSNK1A1P1	chr15	37,100,799	C>G	CTCAAAGCCTGTTTCCCCAC	ATCACACCTTTCAGTCTTCACA	404

Annealing temperature: 63˚C

TABLE 8-1-6 SEQUENCES USED IN SECTION 4.3.6

Gene	Chr	Position	Change	Primer sequence 5′ → 3′		Size (bp)	Detected by
				Forward	Reverse		
FSHR	Chr2	49,220,601	A>G	GGCTTGGCTTCTTCTTTCTCACG	TGAAAAGAAGGCCCAGAGTGCT	257	Seq
NRXN1	Chr2	50,228,720	G>C	TCACCAGGGGCTTAAATATTAAAGT	ATACGTAGAAGTTGGTATGTGCC	302	Seq
NRXN1	Chr2	50,541,129	A>G	TCACCAGGGGCTTAAATATTAAAGT	ACGTAGAAGTTGGTATGTGCC	374	Seq
SERPINE2	Chr2	224,858,544	C>T	ACACTGAAGTTTCCCACAGCGT	CAGAGGCAAGAAACAACATTCACT	252	Seq
PID1	Chr2	229,895,749	T>C	AATACTGTTAATGGCCGGGTG	ACGCCCAGCCCCTATATTTCT	362	Seq
LRRFIP1	Chr2	238,625,033	C>CCTTTTTGTACTCATGTA	TGCATTTCCCTGACATTGAAGT	GAAGGGCGGCCAGATTTAAAA	291	Seq
HDAC4	Chr2	240,059,479	G>A	CCCAGGGTGAACAGAAAAGGCA	AGAACATGTAGAAGGAGTTACAGGA	352	Seq

Annealing temperature: 60 °C

8.1.3 PRIMERS USED IN CHAPTER 6

TABLE 8-1-7 PRIMERS USED IN TABLE 6-1

Gene	Chr	Positon	Change	Primer sequence 5'→3'		Size (bp)
				Forward	Reverse	
MMP14	chr14	23,311,133	c.269G>A	CATATTCACCCTGACCTCATAACCTTGG	CAAGGGATCCTCAGGCTCTGGCTCTC	473
NEO1	chr15	73,575,321	c.3246T>G	CATCTAGAGAAGAGAGCCATGTTCACG	GGCTTCTTACTTTTTCTGGTGAGAGGT	319
PLOD1	chr1	12,012,748	c.535G>A	CACCATTGGTCTTATTCTTGGGGTCAT	GATTAGTGTCCTTATCAATGAGGCCCC	671
SEC23IP	chr10	121,658,469	c.694C>T	CTTATCTGCCTTCTCAGCCAAGTAGTC	ACCGCTTCAATTCTCCTATGTCTCTAG	550
PKM	chr15	72,511,424	c.62T>C	ATCTGTGTAGTATCCCATGTAGCACCT	TCCTTTCCATTGTTCACAATCTGGTGA	489

Annealing temperature: 63˚C

TABLE 8-1-8 PRIMERS USED IN TABLE 6-2

Gene	Chr	Positon	Change	Primer sequence 5'→3'		Size (bp)
				Forward	Reverse	
ZSCAN1	chr19	58,038,298	del C	CTCACACAGATGTGCCAGCAG	GGAGGGATCTTTAATCACCGACT	253
FAM129C	chr19	17,552,096	ins T	AGTGGAATTTCTGGGCCAAAACG	AAGGACAGGAGTTCGAGACCAGC	390

Annealing temperature: 63˚C

8.2 APPENDIX2

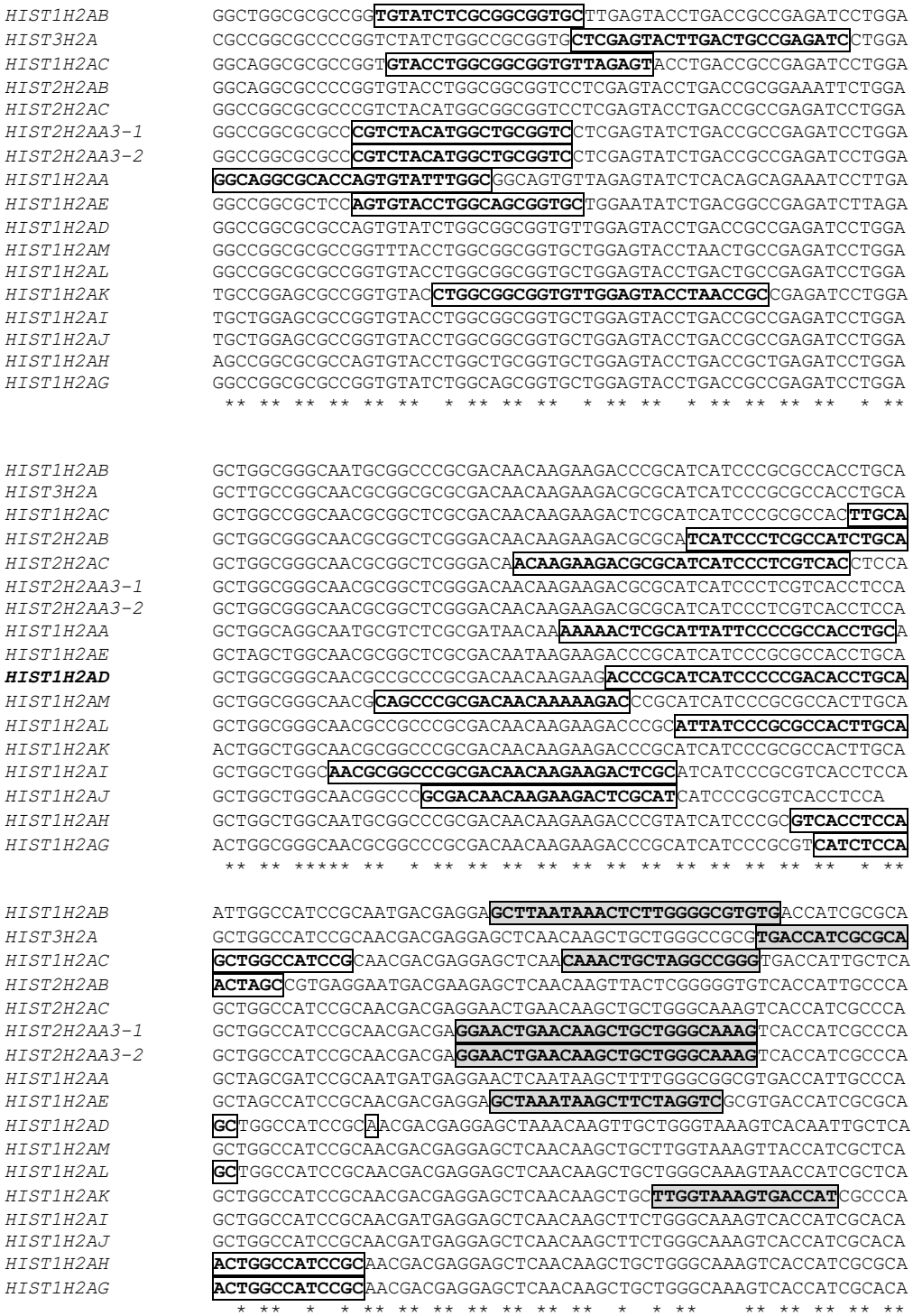
8.2.1 ALIGNMENT OF HUMAN CANONICAL H2A GENES

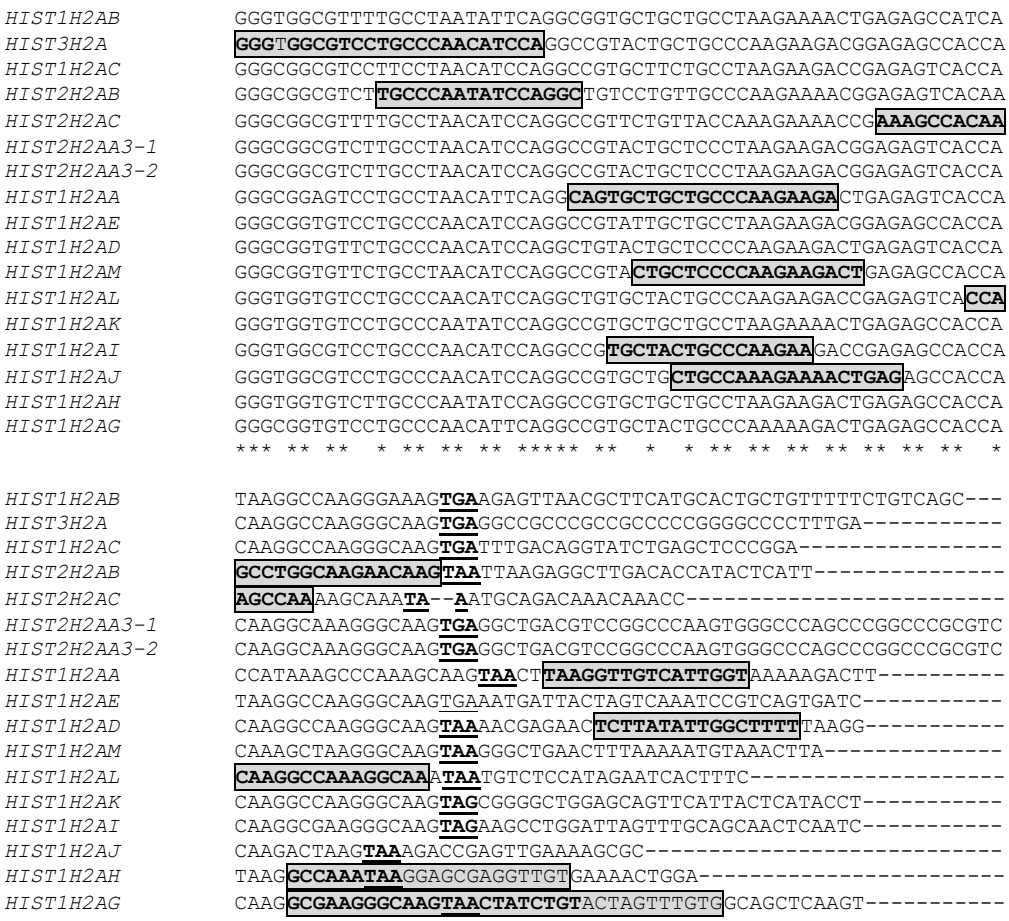
The sequence alignments of canonical H2As are shown along with the position of MIP probes.

```
HIST1H2AB         CA---TTACATTTTCTTGTGGCGATTTTCCCTTATCAGAAGTAGTTATGTCTGGTCGCGG
HIST3H2A          TAGTCTCGCTTTTCGGTTGCCGTTGTCTTTTTTCCTTGACTCGGAAATGTCCGGTCGTGG
HIST1H2AC         CTCGTGAGCTTAGGCCGCTGGTTTTGGTGATTTTTGTCTGATTGCAATGTCTGGACGTGG
HIST2H2AB         CTC-------CTTGTAAAAGAAGCTGTAACTGCAATTTTAGCCGTAATGTCAGGACGCGG
HIST2H2AC         GTTGCGAGGTTCTGAGCGTTGTCTGTGTTTAACCTTGATTTCAGTCATGTCTGGTCGTGG
HIST2H2AA3-1      CCCGATCGCCAGGCAGGAGTTTCTCTCGGTGACTACTATCGCTGTCATGTCTGGTCGTGG
HIST2H2AA3-2      CCCGATCGCCAGGCAGGAGTTTCTCTCGGTGACTACTATCGCTGTCATGTCTGGTCGTGG
HIST1H2AA         GCCCTAGAAATGTGGTTCTCGTGTTGCTGCAATTTGATGCGAGGAGATGTCTGGACGAGG
HIST1H2AE         TCTTTATTCAGTGGATTGTTAGTTCTTCTGCTGTTAGGAAGCCACTATGTCTGGACGTGG
HIST1H2AD         TTACATTTTTGTCTTCATTGCTTAACATCGTTTTTAAGAATTCAAAATGTCCGGACGCGG
HIST1H2AM         TTCCTTTC--ACTCACTTTCTGACTTAGGCCACAGGTCGTTTTACCATGTCTGGACGTGG
HIST1H2AL         ACTTCTCTCGGCTGTTCTCAGTTCCTCCATTTATCGTTTCTTCGTCATGTCGGGACGCGG
HIST1H2AK         GTTACGT-C-AGTCTCAGGAACGTTACTGATTGAATTTTTTACGTCATGTCGGGACGTGG
HIST1H2AI         TTCACTT-CAGTCTTTCTTGACCGTATAGATAATAGGCCTTTTGCCATGTCTGGGCGTGG
HIST1H2AJ         TTCATTT-CAGTCTTTCTTGACCGTAAAGGTAATAGACCTTTTGCCATGTCTGGGCGTGG
HIST1H2AH         GTT--------GG--------CTACTTTCAGTAAGTTGTGACCAGTATGTCTGGACGTGG
HIST1H2AG         TTCACTT-TGTGG--------TTGCTCGTAGTGAGTTGCGCTCGCTATGTCTGGACGTGG
                                                              ***** ** ** **

HIST1H2AB         CAAACAAGGCGGTAAAGCTCGCGCCAAGGCTAAGACTCGGTCTTCTCGTGCAGGTTTGCA
HIST3H2A          TAAGCAGGGTGGCAAGGCGCGCGCCAAGGCTAAGTCGCGCTCGTCGCGCGCGGGGCTGCA
HIST1H2AC         TAAGCAAGGAGGCAAAGCTCGCGCCAAAGCGAAATCCCGCTCTTCTCGCGCTGGTCTCCA
HIST2H2AB         AAAGCAGGGAGGCAAGGCCCGCGCTAAGGCCAAGTCGCGCTCGTCCCGCGCTGGTCTCCA
HIST2H2AC         CAAACAAGGAGGCAAGGCCCGCGCCAAGGCCAAGTCGCGCTCGTCCCGCGCTGGCCTCCA
HIST2H2AA3-1      CAAGCAAGGAGGCAAGGCCCGCGCCAAGGCCAAGTCGCGCTCGTCCCGCGCTGGCCTTCA
HIST2H2AA3-2      CAAGCAAGGAGGCAAGGCCCGCGCCAAGGCCAAGTCGCGCTCGTCCCGCGCTGGCCTTCA
HIST1H2AA         GAAGCAGGGAGGAAAAGCACGCGCCAAGTCTAAGTCTCGCTCTTCTAGAGCGGGTTTGCA
HIST1H2AE         AAAGCAAGGCGGCAAAGCTCGGGCAAAAGCTAAAACGCGTTCTTCCAGGGCCGGTCTTCA
HIST1H2AD         CAAGCAAGGCGGAAAGGCCCGAGCTAAGGCTAAGACCCGCTCTTCGCGGGCCGGACTCCA
HIST1H2AM         CAAGCAGGGCGGCAAGGCTCGCGCCAAGGCCAAAACCCGCTCCTCTAGAGCTGGGCTCCA
HIST1H2AL         CAAGCAGGGAGGCAAAGCTCGCGCCAAAGCCAAGACCCGCTCTTCTCGTGCCGGTCTCCA
HIST1H2AK         CAAGCAGGGCGGCAAAGCTCGCGCCAAGGCTAAGACCCGTTCTTCAAGGGCGGGTCTTCA
HIST1H2AI         CAAGCAGGGAGGCAAAGCTCGCGCCAAGGCCAAGACCCGCTCTTCTCGGGCCGGGCTTCA
HIST1H2AJ         TAAGCAGGGAGGCAAAGCTCGCGCCAAGGCCAAGACCCGCTCTTCTCGGGCCGGGCTTCA
HIST1H2AH         CAAGCAAGGCGGTAAAGCTCGCGCCAAGGCCAAGACCCGCTCTTCTCGGGCTGGGCTTCA
HIST1H2AG         CAAGCAGGGAGGCAAAGCCCGCGCTAAGGCCAAGACTCGCTCTTCTAGGGCCGGTCTCCA
                   ** ** ** ** ** ** ** ** **  * **  * ** ** ** **  * ** **  * **

HIST1H2AB         GTTTCCTGTGGGCCGAGTGCACCGCCTGCTCCGCAAAGGCAACTACTCCGAGCGCGTCGG
HIST3H2A          GTTCCCCGTGGGCCGCGTGCACCGGTTGCTCCGCAAGGGCAACTATTCGGAGCGCGTGGG
HIST1H2AC         GTTCCCGGTGGGCCGAGTGCACCGCCTGCTCCGTAAAGGCAACTACGCAGAGCGGGTTGG
HIST2H2AB         GTTCCCGGTGGGGCGAGTGCACCGCTTGCTGCGCAAAGGCAACTACGCGGAGCGGGTCGG
HIST2H2AC         GTTCCCGGTAGGGCGAGTGCACCGCTTGCTGCGCAAAGGCAACTACGCGGAGCGGGTGGG
HIST2H2AA3-1      GTTCCCGGTAGGGCGAGTGCATCGCTTGCTGCGCAAAGGCAACTACGCGGAGCGAGTGGG
HIST2H2AA3-2      GTTCCCGGTAGGGCGAGTGCATCGCTTGCTGCGCAAAGGCAACTACGCGGAGCGAGTGGG
HIST1H2AA         GTTTCCCGTAGGCCGGATCCATCGTCTGCTTCGTAAGGGAAACTATGCAGAGCGGATAGG
HIST1H2AE         GTTTCCAGTTGGCCGTGTGCACCGCCTCCTCCGCAAAGGCAACTACTCCGAACGAGTCGG
HIST1H2AD         GTTCCCTGTGGGCCGCGTACACCGCTTGCTCCGCAAGGGCAACTACTCCGAGCGAGTCGG
HIST1H2AM         ATTTCCTGTAGGACGAGTGCACCGCCTGCTCCGCAAGGGCAACTACGCTGAGCGGGTCGG
HIST1H2AL         GTTCCCCGTGGGCCGAGTGCACCGACTGCTCCGCAAGGGCAACTATGCTGAGCGGGTCGG
HIST1H2AK         GTTCCCAGTGGGCCGAGTGCACCGACTGCTCCGCAAGGGCAACTACGCTGAGCGGGTCGG
HIST1H2AI         GTTTCCCGTAGGCCGAGTGCATCGCCTGCTCCGCAAAGGCAACTATGCGGAGCGGGTCGG
HIST1H2AJ         GTTTCCCGTAGGCCGAGTGCATCGCCTGCTCCGCAAAGGCAACTATGCGGAGCGGGTCGG
HIST1H2AH         GTTCCCCGTGGGCCGAGTGCACCGCCTGCTCCGCAAGGGTAATTATGCCGAGCGGGTTGG
HIST1H2AG         GTTCCCCGTGGGCCGAGTGCACCGCCTGCTCCGCAAAGGCAACTATGCCGAGCGGGTCGG
                   ** ** ** ** **  * ** **  * ** ** ** ** ** ** **  * ** **  * **
```


HIST1H2AB GGCTGGCGCGCCGGTGTATCTCGCGGCGGTGCTTGAGTACCTGACCGCCGAGATCCTGGA
HIST3H2A CGCCGGCGCCCCGGTCTATCTGGCCGCGGTGCTCGAGTACTTGACTGCCGAGATCCTGGA
HIST1H2AC GGCAGGCGCGCCGGTGTACCTGGCGGCGGTGTTAGAGTACCTGACCGCCGAGATCCTGGA
HIST2H2AB GGCAGGCGCCCCGGTGTACCTGGCGGCGGTCCTCGAGTACCTGACCGCGGAAATTCTGGA
HIST2H2AC GGCCGGCGCGCCCGTCTACATGGCGGCGGTCCTCGAGTACCTGACCGCCGAGATCCTGGA
HIST2H2AA3-1 GGCCGGCGCGCCCGTCTACATGGCTGCGGTCCTCGAGTATCTGACCGCCGAGATCCTGGA
HIST2H2AA3-2 GGCCGGCGCGCCCGTCTACATGGCTGCGGTCCTCGAGTATCTGACCGCCGAGATCCTGGA
HIST1H2AA GGCAGGCGCACCAGTGTATTTGGCGGCAGTGTTAGAGTATCTCACAGCAGAAATCCTTGA
HIST1H2AE GGCCGGCGCTCCAGTGTACCTGGCAGCGGTGCTGGAATATCTGACGGCCGAGATCTTAGA
HIST1H2AD GGCCGGCGCGCCAGTGTATCTGGCGGCGGTGTTGGAGTACCTGACCGCCGAGATCCTGGA
HIST1H2AM GGCCGGCGCGCCGGTTTACCTGGCGGCGGTGCTGGAGTACCTAACTGCCGAGATCCTGGA
HIST1H2AL GGCCGGCGCGCCGGTGTACCTGGCGGCGGTGCTGGAGTACCTGACTGCCGAGATCCTGGA
HIST1H2AK TGCCGGAGCGCCGGTGTACCTGGCGGCGGTGTTGGAGTACCTAACCGCCGAGATCCTGGA
HIST1H2AI TGCTGGAGCGCCGGTGTACCTGGCGGCGGTGCTGGAGTACCTGACCGCCGAGATCCTGGA
HIST1H2AJ TGCTGGAGCGCCGGTGTACCTGGCGGCGGTGCTGGAGTACCTGACCGCCGAGATCCTGGA
HIST1H2AH AGCCGGCGCGCCAGTGTACCTGGCTGCGGTGCTGGAGTACCTGACCGCTGAGATCCTGGA
HIST1H2AG GGCCGGCGCGCCGGTGTATCTGGCAGCGGTGCTGGAGTACCTGACCGCCGAGATCCTGGA
GCTGGCGGGCAATGCGGCCCGCGACAACAAGAAGACCCGCATCATCCCGCGCCACCTGCA
GCTTGCCGGCAACGCGGCGCGCGACAACAAGAAGACGCGCATCATCCCGCGCCACCTGCA
GCTGGCCGGCAACGCGGCTCGCGACAACAAGAAGACTCGCATCATCCCGCGCCACTTGCA
GCTGGCGGGCAACGCGGCTCGGGACAACAAGAAGACGCGCATCATCCCTCGCCATCTGCA
GCTGGCGGGCAACGCGGCTCGGGACAACAAGAAGACGCGCATCATCCCTCGTCACCTCCA
GCTGGCGGGCAACGCGGCTCGGGACAACAAGAAGACGCGCATCATCCCTCGTCACCTCCA
GCTGGCGGGCAACGCGGCTCGGGACAACAAGAAGACGCGCATCATCCCTCGTCACCTCCA
GCTGGCAGGCAATGCGTCTCGCGATAACAAAAAAACTCGCATTATTCCCCGCCACCTGCA
GCTAGCTGGCAACGCGGCTCGCGACAATAAGAAGACCCGCATCATCCCGCGCCACCTGCA
GCTGGCGGGCAACGCCGCCCGCGACAACAAGAAGACCCGCATCATCCCCCGACACCTGCA
GCTGGCGGGCAACGCAGCCCGCGACAACAAAAAGACCCCGCCACTTGCA
GCTGGCGGGCAACGCCGCCCGCGACAACAAGAAGACCCGCATTATCCCGCGCCACTTGCA
ACTGGCTGGCAACGCGGCCCGCGACAACAAGAAGACCCGCATCATCCCGCGCCACTTGCA
GCTGGCTGGCAACGCGGCCCGCGACAACAAGAAGACTCGCATCATCCCGCGTCACCTCCA
GCTGGCTGGCAACGGCCCGCGACAACAAGAAGACTCGCATCATCCCGCGTCACCTCCA
GCTGGCTGGCAATGCGGCCCGCGACAACAAGAAGACCCGTATCATCCCGCGTCACCTCCA
ACTGGCGGGCAACGCGGCCCGCGACAACAAGAAGACCCGCATCATCCCGCGTCATCTCCA
ATTGGCCATCCGCAATGACGAGGAGCTTAATAAACTCTTGGGGCGTGTGACCATCGCGCA
GCTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCCGCGTGACCATCGCGCA
GCTGGCCATCCGCAACGACGAGGAGCTCAACAAACTGCTAGGCCGGGTGACCATTGCTCA
ACTAGCCGTGAGGAATGACGAAGAGCTCAACAAGTTACTCGGGGGTGTCACCATTGCCCA
GCTGGCCATCCGCAACGACGAGGAACTGAACAAGCTGCTGGGCAAAGTCACCATCGCCCA
GCTGGCCATCCGCAACGACGAGGAACTGAACAAGCTGCTGGGCAAAGTCACCATCGCCCA
GCTGGCCATCCGCAACGACGAGGAGCTAAACAAGTTGCTGGGTAAAGTCACAATTGCTCA
GCTAGCGATCCGCAATGATGAGGAACTCAATAAGCTTTTGGGCGGCGTGACCATTGCCCA
GCTAGCCATCCGCAACGACGAGGAGCTAAATAAGCTTCTAGGTCGCGTGACCATCGCGCA
GCTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTTGGTAAAGTTACCATCGCTCA
GCTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCAAAGTAACCATCGCTCA
GCTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTTGGTAAAGTGACCATCGCCCA
GCTGGCCATCCGCAACGATGAGGAGCTCAACAAGCTTCTGGGCAAAGTCACCATCGCACA
GCTGGCCATCCGCAACGATGAGGAGCTCAACAAGCTTCTGGGCAAAGTCACCATCGCACA
ACTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCAAAGTCACCATCGCGCA
ACTGGCCATCCGCAACGACGAGGAGCTCAACAAGCTGCTGGGCAAAGTCACCATCGCACA



White rectangles indicate ligation arm, shaded rectangle indicate extension arms.
Underlined sequences indicate start codon (in the beginning) or stop codon (at the end).
Red mark show the position of our mutation occurred (*HIST1H2AD* c.268A>Gp.N90D).

8.3 APPENDIX3

TABLE 8-1-9 THE LIST OF VARIANTS USED TO IDENTIFY THE SEGREGATING REGIONS IN CHAPTER 4

Gene	DS	Chr	Start	IV-3*	V-1*	III-6*	IV-6*	IV-5*	III-7	II-7	III-8	location	Function
GPR157	-	chr1	9,188,226		16; 0; 27; 0	17; 0; 12; 0	18; 0; 9; 1	GG	GG	CG	CG	intronic	
PTCHD2	-	chr1	11,579,608	1; 6; 2; 7	0; 0; 21; 45		6; 9; 7; 10					intronic	
GLIS1	-	chr1	53,972,285	7; 2; 8; 4	25; 9; 18; 6	11; 4; 9; 7		GA	GA	GG	GA	UTR3	
ACOT11	-	chr1	55,075,853	6; 16; 6; 4				TC	TC	CC	TC	UTR3	
ATG4C	-	chr1	63,269,410	5; 6; 10; 1	25; 7; 20; 4	10; 3; 4; 2	12; 2; 12; 3	del	CT	CT	del	UTR5	
ALG6	3	chr1	63,876,987	0; 4; 0; 9	3; 27; 2; 38	0; 23; 0; 16	3; 27; 3; 21	AG	GG	GG	AG	exonic	nonsynonymous
DNAJC6	-	chr1	65,830,499		29; 34; 15; 39	12; 23; 6; 13	14; 15; 10; 18	AG	GG	GG		intronic	
PDE4B	-	chr1	66,839,954	12; 11; 5; 15				GA	GA	GG	GG	UTR3	
C1orf141	-	chr1	67,558,379	8; 12; 4; 9				del	del	no	no	UTR3	
GNG12	-	chr1	68,243,107	14; 3; 5; 3				AA	AA	AA	AA	intronic	
CTH	-	chr1	70,896,152		1; 13; 2; 13		0; 4; 0; 7		GA	AA		intronic	
HFM1	-	chr1	91,733,453	1; 10; 1; 4	0; 6; 0; 9	0; 0; 0; 9	0; 7; 0; 7			TG	TT	intronic	
SORT1	-	chr1	109,856,843		6; 18; 8; 34	8; 11; 7; 9	13; 8; 5; 20		TG	TG	TT	intronic	
UHMK1	-	chr1	162,473,500			10; 3; 7; 0	3; 0; 12; 1	TT	CT	TT	CT	intronic	
NME7	-	chr1	169,204,230	7; 0; 4; 0	12; 0; 5; 0		5; 0; 7; 0	TC	TC	CC	TC	intronic	
DNM3	-	chr1	171,958,073	9; 2; 11; 0	9; 0; 11; 0	7; 0; 4; 0	4; 1; 6; 0		AG	GG	AG	intronic	
SWT1	-	chr1	185,240,506	0; 0; 6; 5	7; 25; 5; 17	1; 5; 3; 3	1; 11; 0; 5		AG	GG	AG?	exonic	synonymous
KCNT2	-	chr1	196,288,561	8; 1; 10; 1	32; 7; 40; 9	16; 4; 19; 4			GA	GG	GG	intronic	
TMEM63A	1	chr1	226,055,624			6; 0; 4; 4	8; 4; 8; 3	CC	TC	TC	TC	exonic	nonsynonymous
TRIM67	-	chr1	231,334,778	10; 1; 14; 1	57; 49; 33; 33	23; 21; 17; 19	26; 33; 15; 17	CT	CT	CC	CC	intronic	
EGLN1	-	chr1	231,506,478	1; 11; 0; 6	8; 12; 3; 14	1; 3; 1; 3	4; 10; 1; 7	AG?	AG	AA	AA	intronic	
RYR2	-	chr1	237,494,052			7; 0; 5; 0		GA	GA	GG	GG	intronic	
RYR2	-	chr1	237,958,564	11; 5; 5; 2	7; 3; 10; 2	2; 1; 3; 1		AC	AC	AA	AA	intronic	
SMYD3	5	chr1	246,498,689	7; 11; 10; 7	41; 16; 41; 20		23; 4; 23; 7	CG	CC	CC	CG	exonic	nonsynonymous
CAPN14	6	chr2	31,422,398	13; 12; 9; 15	24; 27; 20; 23		13; 12; 7; 10		del	del	-	exonic	frameshift deletion
RMDN2	-	chr2	38,156,623		13; 16; 9; 14	8; 16; 6; 5	13; 13; 17; 9	TC	TC	TT	TT	exonic	nonsynonymous
HNRNPLL	-	chr2	38,813,090		4; 59; 6; 45	1; 30; 2; 31	1; 37; 5; 36	TC	TC	TT	TT	intronic	
PIGF	1	chr2	46,842,119	16; 2; 9; 4	9; 28; 7; 23	6; 16; 4; 10	5; 15; 3; 5	GA	GA	AA	AA	exonic	nonsynonymous

Gene	DS	Chr	Start	IV-3*	V-1*	III-6*	IV-6*	IV-5*	III-7	II-7	III-8	location	Function
NRXN1	-	chr2	50,724,758	8; 24; 13; 16	39; 50; 31; 26	19; 13; 9; 14	18; 23; 20; 18	TG	GG	GG	GG	exonic	synonymous
SPTBN1	-	chr2	54,856,036		18; 7; 14; 2	13; 2; 14; 4	15; 2; 12; 4	CT	CC	CC	CC	intronic	
MTIF2	-	chr2	55,481,137	5; 0; 5; 0	27; 3; 28; 2	14; 1; 24; 0	19; 2; 26; 1	CG	GG	GG	GG	intronic	
WDR75	-	chr2	190,339,023	3; 8; 5; 14	13; 23; 5; 28		6; 21; 8; 6	CT	CC	CT	CC	exonic	synonymous
NBEAL1	3	chr2	204,037,533	6; 6; 1; 3	15; 20; 14; 16	6; 10; 4; 10	7; 7; 4; 8	GA	AA	GA	AA	exonic	nonsynonymous
NDUFS1	3	chr2	207,003,310	5; 13; 4; 15	6; 19; 7; 12	2; 7; 1; 5	5; 7; 2; 10	CG	GG	CG	GG	exonic	nonsynonymous
MRPL44	-	chr2	224,824,820		0; 1; 0; 4	0; 2; 0; 4	0; 8; 0; 11	TC	CC	CC	CC	intronic	
SP100	-	chr2	231,328,813	6; 10; 7; 7	33; 29; 22; 20	13; 8; 11; 11	14; 10; 11; 8	TC	TT	TT	TT	exonic	synonymous
ECEL1	2	chr2	233,351,209		5; 3; 7; 4		1; 1; 3; 1		AA			exonic	nonsynonymous
EIF4E2	-	chr2	233,421,049	11; 0; 11; 0	6; 0; 10; 0	8; 1; 8; 0	12; 0; 13; 0		GG	GG	GG	intronic	
NGEF	-	chr2	233,748,006	6; 3; 4; 1			5; 0; 2; 2					intronic	
IQCA1	-	chr2	237,374,374	0; 14; 0; 14	3; 18; 3; 7	3; 6; 2; 9	0; 6; 0; 4	CT	CC	CC	CC	intronic	
MGC16025	-	chr2	240,116,262	6; 3; 8; 4								ncRNA_exonic	
DUSP28	-	chr2	241,503,045	4; 0; 5; 0								intronic	
DENND6A	-	chr3	57,644,825	18; 4; 16; 7		28; 2; 22; 4	23; 4; 29; 2	CT	CC	CT	CT	intronic	
SLC25A26	-	chr3	66,271,603		1; 14; 1; 14	0; 6; 0; 10	1; 8; 0; 6	TC		TC	TT	intronic	
GXYLT2	-	chr3	72,926,101	6; 2; 8; 4	0; 0; 15; 12		11; 8; 3; 4	-	del	del	del	intergenic	
STK32B	-	chr4	5,053,652		37; 54; 23; 31	14; 15; 8; 24	14; 18; 9; 24	AC		AC	AC	intronic	
FAM184B	-	chr4	17,666,297		8; 11; 0; 9	2; 2; 0; 7	2; 8; 0; 6	GG	GG		GG	intronic	
GUCY1B3	-	chr4	156,721,182	4; 12; 5; 18	21; 17; 14; 13		17; 19; 13; 13	GG	AG	GG	AG	exonic	synonymous
SAP30	-	chr4	174,292,301	2; 7; 1; 5	4; 0; 8; 2	4; 0; 3; 1	3; 1; 4; 1		AT			UTR5	
FAM149A	-	chr4	187,077,426		0; 12; 0; 5	1; 5; 0; 7	0; 8; 0; 13	AA	AA		AA	intronic	
CDH9	-	chr5	26,907,047		0; 13; 0; 14	0; 13; 0; 11	0; 13; 0; 18	CC	CC	TC	CC	intronic	
C1QTNF3-AMACR	-	chr5	33,988,222		11; 0; 11; 0	7; 0; 5; 0		CT	CT	CC	CT	ncRNA_exonic	
GZMA	3	chr5	54,405,889	5; 2; 12; 3	12; 18; 7; 17	1; 4; 2; 5		GG	GG	AG		exonic	nonsynonymous
PTCD2	-	chr5	71,622,616		4; 31; 4; 20	2; 10; 0; 16	1; 18; 0; 12	AA		AG		intronic	
RNU5D-1	-	chr5	80,521,352		8; 0; 5; 0		7; 0; 4; 0	GG	GG	GT	GT	ncRNA_intronic	
MIR3977	-	chr5	82,136,024		7; 16; 6; 24	1; 9; 1; 9		TG	TT	TT	TT	ncRNA_exonic	
C7orf31	-	chr7	25,175,885	20; 11; 9; 9	57; 32; 42; 19	34; 15; 22; 12		TT	CT	TT	TT	exonic	synonymous
CPVL	6	chr7	29,152,356	9; 10; 8; 8	1; 1; 8; 4	5; 3; 3; 1	2; 1; 8; 3	GT	GT			exonic	stopgain
POLD2	-	chr7	44,154,453	6; 9; 3; 5		16; 5; 14; 1	14; 2; 14; 0		AG	AG	GG	exonic	synonymous

Gene	DS	Chr	Start	IV-3*	V-1*	III-6*	IV-6*	IV-5*	III-7	II-7	III-8	location	Function
CYP3A4	-	chr7	99,367,339	0; 0; 19; 3	13; 0; 12; 0	8; 1; 9; 1	15; 0; 11; 1	CC	CA	CA	AA	intronic	
TRIM55	-	chr8	67,062,485	21; 8; 18; 5	29; 5; 33; 0	13; 1; 22; 1		TG	TT	TG	TT	intronic	
C8orf37	-	chr8	96,272,580	8; 0; 7; 0	6; 0; 13; 0	8; 1; 9; 0	10; 2; 8; 1	AG	AA	AG	AA	intronic	
GDF6	1	chr8	97,157,413		1; 2; 3; 5	2; 3; 4; 7	2; 6; 3; 1	TG	TT	TG		exonic	nonsynonymous
VPS13B	5	chr8	100,829,841			29; 10; 26; 11		GG	AA	GA	AA	exonic	nonsynonymous
ATP6V1C1	-	chr8	104,075,306	6; 4; 5; 8	4; 3; 6; 7	3; 1; 6; 1	3; 3; 3; 4	GT	TT	GT	TT	intronic	
CSMD3	-	chr8	113,277,850	1; 11; 0; 14	10; 37; 7; 27		4; 17; 1; 8	TC	TC	TC	TC	intronic	
KANK1	-	chr9	744,589	3; 10; 9; 12	38; 21; 16; 12	4; 8; 19; 8	14; 10; 9; 11	GG	GA			exonic	synonymous
KCNV2	3	chr9	2,718,341		4; 8; 6; 11	7; 10; 3; 7	2; 10; 3; 14	TT	AT	AT	AT	exonic	nonsynonymous
CDC37L1	4	chr9	4,688,573	5; 8; 5; 2	11; 14; 12; 7	12; 6; 3; 2	8; 4; 7; 3		AG			exonic	nonsynonymous
KIAA1432	-	chr9	5,772,537	14; 0; 6; 0	28; 9; 24; 6		25; 3; 28; 4		CT	CT	CT	intronic	
PCSK5	-	chr9	78,803,362			4; 0; 6; 0		CT	CT	CT	TT	intronic	
SPATA31D1	-	chr9	84,608,198	20; 14; 19; 12	32; 26; 31; 25	20; 21; 18; 12	28; 19; 23; 22	AC		AC	CC	exonic	nonsynonymous
ROR2	2	chr9	94,487,106	4; 15; 7; 12			25; 26; 19; 21	GA	GG	GA	GA	exonic	nonsynonymous
C9orf129	2	chr9	96,097,836	44; 19; 19; 19			24; 25; 12; 15	GA	GG	GA	GA	exonic	nonsynonymous
CIZ1	-	chr9	130,941,121	7; 11; 4; 2	37; 21; 26; 10	17; 8; 15; 8	11; 8; 20; 8	CT	TT	TT	CT	exonic	synonymous
GBGT1	-	chr9	136,036,978		4; 10; 0; 7	1; 4; 1; 6	2; 6; 2; 7	CT	TT	TT	CT	intronic	
SNAPC4	1	chr9	139,273,451	0; 11; 2; 3	33; 14; 27; 12	14; 5; 8; 2	15; 4; 19; 8	GA	GG	GG	GA	exonic	nonsynonymous
KIAA1984-AS1	-	chr9	139,700,717	9; 11; 4; 10		22; 15; 20; 10	11; 6; 14; 17		CC	CC	AC	ncRNA_intronic	
FRMD4A		chr10	14,372,134	31, 25, 18, 18	0.301	12, 12, 15, 7		TC		TC		UTR5	
MPP7	-	chr10	28,490,983	3; 0; 8; 0	27; 0; 13; 0	8; 0; 14; 1	14; 0; 11; 0	TC	TC	TC	TT	intronic	
HECTD2	1	chr10	93,220,242	12; 5; 9; 7	49; 49; 42; 40	20; 16; 19; 17		TA	TT	AT	TT	exonic	nonsynonymous
TM9SF3	-	chr10	98,281,892	7; 9; 4; 11	51; 16; 50; 7	23; 7; 19; 8	24; 5; 25; 5	AA	AA	AA	AG	UTR3	
PKD2L1	-	chr10	102,054,419	7; 14; 9; 20	38; 47; 20; 26	24; 16; 12; 18	15; 17; 20; 15			GG	AG	intronic	
GRK5	-	chr10	121,203,141	5; 11; 2; 6	17; 24; 14; 33	5; 8; 5; 12	6; 10; 6; 9	CT	CC	CC	CT	exonic	synonymous
DMBT1	-	chr10	124,390,666	4; 13; 6; 8	34; 18; 22; 15	9; 10; 15; 3	16; 13; 8; 6	CT	TT	TT	CT	exonic	nonsynonymous
DOCK1	-	chr10	129,250,514	3; 8; 4; 9				AG	GG	GG	GG	UTR3	
KDM5A	-	chr12	432,200	14; 1; 10; 0	9; 1; 13; 0	2; 0; 4; 0	4; 1; 5; 0	AA	AG	AA	AA	intronic	
CACNA1C	-	chr12	2,794,977	15; 20; 9; 20	11; 9; 10; 14		4; 7; 4; 8			GA	GA	exonic	synonymous
GJA3	5	chr13	20,717,217		29; 38; 29; 28	19; 18; 11; 17	15; 29; 13; 12	GG	GC	GG	GG	exonic	nonsynonymous
IFT88	-	chr13	21,218,984	13; 3; 14; 4	4; 1; 5; 2	3; 0; 5; 2	1; 2; 3; 2		GA			exonic	synonymous

Gene	DS	Chr	Start	IV-3*	V-1*	III-6*	IV-6*	IV-5*	III-7	II-7	III-8	location	Function
RNF17	2	chr13	25,341,452	8; 14; 11; 5	18; 10; 18; 16	12; 6; 11; 8	8; 12; 12; 5	TT	CT			exonic	nonsynonymous
NUPL1	-	chr13	25,901,527	10; 3; 6; 0	6; 0; 6; 0	2; 0; 4; 0	4; 0; 4; 0	AA	AA	AA	AA	intronic	
MTUS2	-	chr13	29,933,362			13; 5; 13; 4	25; 5; 14; 8	TT	TT	TT	TT	intronic	
SLC7A1	-	chr13	30,097,716	1; 5; 0; 5	0; 12; 0; 11	0; 6; 0; 7		CA	AA	CA	AA	intronic	
STOML3	-	chr13	39,541,282		0; 5; 0; 9	0; 7; 0; 5	0; 5; 0; 8	AC	CC	AC	CC	intronic	
RB1	-	chr13	48,936,851	12; 0; 8; 0	21; 0; 15; 0	9; 0; 16; 0			TT	TA	TT	intronic	
METTL17	1	chr14	21,458,181	3; 7; 2; 4	22; 0; 18; 4	16; 1; 13; 2	20; 4; 12; 2	GG	AG	GG	GG	exonic	nonsynonymous
PPP1R3E	-	chr14	23,768,174	0; 5; 0; 7	1; 33; 0; 31	0; 23; 2; 18	1; 22; 1; 23	AA	AG	AA		ncRNA_intronic	
SDR39U1	-	chr14	24,911,861	6; 0; 4; 0		5; 1; 7; 0	6; 3; 16; 2	GG	AG	GG	GG	intronic	
RYR3	-	chr15	34,048,618	4; 12; 2; 14	5; 20; 1; 21		2; 16; 0; 12	TC	CC	CC	CC	intronic	
SRP14	-	chr15	40,331,182	10; 9; 12; 7	68; 51; 82; 52	28; 20; 32; 19	47; 23; 27; 18	GG	AG	GG	AG	intronic	
PLCB2	6	chr15	40,590,557	6; 10; 2; 11	5; 32; 4; 27	8; 12; 3; 7	7; 13; 4; 15		GG			exonic	stopgain
MGA	-	chr15	42,032,211	4; 0; 5; 0	29; 27; 21; 24	14; 9; 12; 13	8; 22; 15; 11	AA	AG	AA	AG	intronic	
PRTG	-	chr15	55,964,577			5; 0; 6; 0	10; 1; 8; 0	AA	GA	AA	GA	intronic	
SLTM	-	chr15	59,180,748	2; 11; 0; 6	5; 18; 2; 17		2; 15; 0; 8	AG	AG	AG	AG	exonic	synonymous
TLN2	5	chr15	63,092,900	10; 6; 9; 14	21; 24; 24; 18	8; 12; 11; 10	8; 10; 12; 17	GC	GG	GC	GC	exonic	nonsynonymous
TRIP4	5	chr15	64,706,405	5; 9; 7; 7	19; 43; 18; 30	15; 24; 6; 17	8; 19; 10; 20	TC		TC		exonic	nonsynonymous
RAB11A	-	chr15	66,170,128	4; 2; 5; 2	40; 5; 28; 7	18; 1; 21; 3	18; 5; 20; 5	TC	TT	CT	CT	exonic	synonymous
PKM	-	chr15	72,491,888	1; 9; 1; 7	9; 13; 14; 19	2; 10; 5; 5		GG	GG	GG	GG	UTR3	
ADAMTS7	-	chr15	79,089,037	21; 12; 13; 6	5; 32; 7; 28		8; 12; 5; 11	GG				exonic	synonymous
SLC28A1	-	chr15	85,438,569		11; 6; 10; 5	5; 2; 5; 2	7; 2; 5; 4	TT	TT	TT	TT	intronic	
ZNF774	-	chr15	90,903,618	15; 12; 12; 8	33; 21; 31; 19	11; 8; 19; 18	18; 14; 17; 15	GG	GG	GG	GG	exonic	synonymous
MAN2A2	-	chr15	91,459,342	8; 7; 4; 5	23; 5; 21; 4	15; 5; 8; 0	15; 3; 10; 2	TT	TT	TT	TT	intronic	
LRRK1	1	chr15	101,606,029			17; 22; 23; 17	17; 24; 13; 18	CC	CA	CC	CA	exonic	nonsynonymous
PDILT	-	chr16	20,381,047	0; 0; 0; 7	0; 5; 0; 7	0; 3; 0; 6	1; 4; 0; 13	CC	CT	CC		intronic	
RLTPR	-	chr16	67,685,730	21; 6; 8; 2		6; 6; 9; 4	12; 7; 8; 10	AT	AT	AT	AA	exonic	nonsynonymous
PDXDC2P	-	chr16	70,076,801	10; 2; 6; 2	19; 6; 23; 6	10; 4; 12; 4	6; 1; 4; 2	GT		GT		ncRNA_intronic	
IST1	-	chr16	71,956,535		4; 13; 90; 112			AG	AG	AG	GG	exonic	nonframeshift deletion
CHST5	5	chr16	75,564,207	8; 18; 6; 14	23; 19; 16; 22	6; 12; 5; 7		AG	AG	AG	GG	exonic	nonsynonymous
SGSM2	-	chr17	2,267,976	8; 16; 10; 19	15; 37; 9; 34	4; 8; 6; 17			CC	CT	CC	intronic	
RAP1GAP2	-	chr17	2,929,448	2; 7; 1; 4	4; 14; 2; 10	0; 8; 2; 7	4; 9; 3; 12	GG				intronic	

Gene	DS	Chr	Start	IV-3*	V-1*	III-6*	IV-6*	IV-5*	III-7	II-7	III-8	location	Function
C17orf85	2	chr17	3,721,680	27; 12; 21; 6	33; 30; 21; 20	14; 9; 17; 9	15; 19; 12; 13	CC		CT		exonic	nonsynonymous
CHRNE	-	chr17	4,802,111	2; 14; 0; 9	17; 18; 26; 18	16; 8; 14; 5		CC	CG	CC	CC	exonic	nonsynonymous
RGS9	-	chr17	63,156,785			13; 19; 14; 15	17; 23; 13; 19	GA	GG	GA	GA	intronic	
ABCA10	-	chr17	67,148,723		0; 6; 0; 9	0; 5; 1; 6	1; 8; 2; 6	AT		AT		intronic	
MYO15B	-	chr17	73,608,847	8; 1; 6; 1	36; 8; 31; 9		28; 3; 21; 4	GA	GA	GA	GG	ncRNA_intronic	
FSCN2	-	chr17	79,495,897		7; 7; 7; 6	5; 3; 3; 4	2; 8; 6; 5	AA	AG	AA	AG	exonic	nonsynonymous
ACER1	-	chr19	6,333,333		25; 0; 28; 0		20; 1; 18; 1	CC	CC	CC	CT	intronic	
LRRC8E	-	chr19	7,960,423	5; 0; 4; 0	29; 0; 18; 6	12; 2; 13; 2	13; 1; 8; 3	no del		no del		intronic	
WIZ	-	chr19	15,550,580		35; 31; 19; 18	19; 22; 22; 16	10; 14; 21; 17	CC		CC		intronic	
GTPBP3	-	chr19	17,452,229			6; 2; 6; 2	5; 2; 1; 5	GG	AG	GG	GG	intronic	
RTN2	-	chr19	45,997,992			7; 11; 7; 10	21; 12; 14; 12	TC	TC	TT	TT	exonic	synonymous
SYMPK	-	chr19	46,366,389		2; 13; 1; 14	3; 8; 0; 6	2; 13; 2; 12	del	del			UTR5	
TMEM143	3	chr19	48,848,554			5; 11; 7; 10	5; 11; 10; 13	TC	TC		CC	exonic	nonsynonymous
NUCB1	2	chr19	49,409,026	4; 11; 8; 14	26; 25; 31; 22	15; 23; 14; 17		GG	GG	AG	GG	exonic	nonsynonymous

DS, Deleterious score, * indicate affected individuals.

The numbers in rows IV-3,V-1, III-3 and IV-6 present the number of reads identified from WES. The number of [reference forward, reference reverse, variant forward, variant reverse] are shown. Blank shows either no variant on that position or no reads are mapped on that position.

The genotype of IV-5, IV-8, II-7 and IV-7 were determined by dideoxy-sequencing.

Chapter 9 REFERENCES

Abu-Abed, S., Dolle, P., Metzger, D., Beckett, B., Chambon, P., and Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. Genes & development *15*, 226-240.
Adams, M.K., Belyaeva, O.V., Wu, L., and Kedishvili, N.Y. (2014). The retinaldehyde reductase activity of DHRS3 is reciprocally activated by retinol dehydrogenase 10 to control retinoid homeostasis. The Journal of biological chemistry *289*, 14868-14880.
Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics *Chapter 7*, Unit7 20.
Aviv, R.I., Rodger, E., and Hall, C.M. (2002). Craniosynostosis. Clinical radiology *57*, 93-102.
Bailey-Wilson, J.E., and Wilson, A.F. (2011). Linkage analysis in the next-generation sequencing era. Human heredity *72*, 228-236.
Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet *12*, 745-755.
Banday, A.R., Baumgartner, M., Al Seesi, S., Karunakaran, D.K., Venkatesh, A., Congdon, S., Lemoine, C., Kilcollins, A.M., Mandoiu, I., Punzo, C.*, et al.* (2014). Replication-dependent histone genes are actively transcribed in differentiating and aging retinal neurons. Cell Cycle *13*, 2526-2541.
Barroso, E., Berges-Soria, J., Benito-Sanz, S., Rivera-Pedroza, C.I., Ballesta-Martinez, M.J., Lopez-Gonzalez, V., Guillen-Navarro, E., and Heath, K.E. (2015). Identification of the fourth duplication of upstream IHH regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. American journal of medical genetics Part A *167A*, 902-906.
Becker, J., Semler, O., Gilissen, C., Li, Y., Bolz, H.J., Giunta, C., Bergmann, C., Rohrbach, M., Koerber, F., Zimmermann, K.*, et al.* (2011). Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. Am J Hum Genet *88*, 362-371.
Beederman, M., Farina, E.M., and Reid, R.R. (2014). Molecular basis of cranial suture biology and disease: Osteoblastic and osteoclastic perspectives. Genes & diseases *1*, 120-125.
Benito-Sanz, S., Royo, J.L., Barroso, E., Paumard-Hernandez, B., Barreda-Bonis, A.C., Liu, P., Gracia, R., Lupski, J.R., Campos-Barros, A., Gomez-Skarmeta, J.L.*, et al.* (2012). Identification of the first recurrent PAR1 deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. J Med Genet *49*, 442-450.
Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T.*, et al.* (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet *41*, 359-364.
Billings, S.E., Pierzchalski, K., Butler Tjaden, N.E., Pang, X.Y., Trainor, P.A., Kane, M.A., and Moise, A.R. (2013). The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *27*, 4877-4889.
Blomhoff, R., and Blomhoff, H.K. (2006). Overview of retinoid metabolism and function. Journal of neurobiology *66*, 606-630.
Bochukova, E.G., Roscioli, T., Hedges, D.J., Taylor, I.B., Johnson, D., David, D.J., Deininger, P.L., and Wilkie, A.O. (2009a). Rare mutations of FGFR2 causing apert syndrome: identification of the first partial gene deletion, and an Alu element insertion from a new subfamily. Hum Mutat *30*, 204-211.
Bochukova, E.G., Roscioli, T., Hedges, D.J., Taylor, I.B., Johnson, D., David, D.J., Deininger, P.L., and Wilkie, A.O.M. (2009b). Rare Mutations of FGFR2 Causing Apert Syndrome: Identification of the First Partial Gene Deletion, and an Alu Element Insertion From a New Subfamily. Human mutation *30*, 204-211.

Bonisch, C., and Hake, S.B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research *40*, 10719-10741.
Boulet, S.L., Rasmussen, S.A., and Honein, M.A. (2008). A population-based study of craniosynostosis in metropolitan Atlanta, 1989-2003. American journal of medical genetics Part A *146A*, 984-991.
Bowcock, A.M., and Krueger, J.G. (2005). Getting under the skin: the immunogenetics of psoriasis. Nature reviews Immunology *5*, 699-711.
Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., den Dunnen, J.T.*, et al.* (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. American journal of human genetics *100*, 695-705.
Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet *14*, 681-691.
Boyle, E.A., O'Roak, B.J., Martin, B.K., Kumar, A., and Shendure, J. (2014). MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics *30*, 2670-2672.
Bronner, M.E., and LeDouarin, N.M. (2012). Development and evolution of the neural crest: an overview. Developmental biology *366*, 2-9.
Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American journal of human genetics *81*, 1084-1097.
Buchanan, E.P., Xue, Y., Xue, A.S., Olshinka, A., and Lam, S. (2017). Multidisciplinary care of craniosynostosis. Journal of multidisciplinary healthcare *10*, 263-270.
Carlton, M.B., Colledge, W.H., and Evans, M.J. (1998). Crouzon-like craniofacial dysmorphology in the mouse is caused by an insertional mutation at the Fgf3/Fgf4 locus. Developmental dynamics : an official publication of the American Association of Anatomists *212*, 242-249.
Carmichael, S.L., Ma, C., Rasmussen, S.A., Honein, M.A., Lammer, E.J., and Shaw, G.M. (2008). Craniosynostosis and maternal smoking. Birth defects research Part A, Clinical and molecular teratology *82*, 78-85.
Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic acids research *31*, 3568-3571.
Carvalho, C.M., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nature reviews Genetics *17*, 224-238.
Chan, K.M., Wong, H.L., Jin, G., Liu, B., Cao, R., Cao, Y., Lehti, K., Tryggvason, K., and Zhou, Z. (2012). MT1-MMP inactivates ADAM9 to regulate FGFR2 signaling and calvarial osteogenesis. Developmental cell *22*, 1176-1190.
Chen, G.Q., Deng, C.X., and Li, Y.P. (2012). TGF-beta and BMP Signaling in Osteoblast Differentiation and Bone Formation. International journal of biological sciences *8*, 272-288.
Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome research *19*, 1553-1561.
Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B.*, et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422-1423.
Cohen, M.M., and MacLean, R.E. (2000). Craniosynostosis : diagnosis, evaluation, and management, 2nd edn (New York, Oxford University Press).
Cohlan, S.Q. (1954). Congenital anomalies in the rat produced by excessive intake of vitamin A during pregnancy. Pediatrics *13*, 556-567.
Collins, M.D., and Mao, G.E. (1999). Teratology of retinoids. Annu Rev Pharmacol Toxicol *39*, 399-430.
Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. American journal of medical genetics Part A *155A*, 270-279.

Cox, S.G., Kim, H., Garnett, A.T., Medeiros, D.M., An, W., and Crump, J.G. (2012). An essential role of variant histone H3.3 for ectomesenchyme potential of the cranial neural crest. PLoS genetics *8*, e1002938.
Cucinotta, C.E., Young, A.N., Klucevsek, K.M., and Arndt, K.M. (2015). The Nucleosome Acidic Patch Regulates the H2B K123 Monoubiquitylation Cascade and Transcription Elongation in Saccharomyces cerevisiae. PLoS genetics *11*, e1005420.
Cunningham, T.J., and Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. Nature reviews Molecular cell biology *16*, 110-123.
Daluiski, A., Engstrand, T., Bahamonde, M.E., Gamer, L.W., Agius, E., Stevenson, S.L., Cox, K., Rosen, V., and Lyons, K.M. (2001). Bone morphogenetic protein-3 is a negative regulator of bone density. Nature genetics *27*, 84-88.
Dathe, K., Kjaer, K.W., Brehm, A., Meinecke, P., Nurnberg, P., Neto, J.C., Brunoni, D., Tommerup, N., Ott, C.E., Klopocki, E.*, et al.* (2009). Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. Am J Hum Genet *84*, 483-492.
Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS computational biology *6*, e1001025.
Deckelbaum, R.A., Holmes, G., Zhao, Z., Tong, C., Basilico, C., and Loomis, C.A. (2012). Regulation of cranial morphogenesis and cell fate at the neural crest-mesoderm boundary by engrailed 1. Development *139*, 1346-1358.
Demirhan, O., Turkmen, S., Schwabe, G.C., Soyupak, S., Akgul, E., Tastemir, D., Karahan, D., Mundlos, S., and Lehmann, K. (2005). A homozygous BMPR1B mutation causes a new subtype of acromesomelic chondrodysplasia with genital anomalies. Journal of medical genetics *42*, 314-317.
Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research *37*, e67.
Dieterich, K., Quijano-Roy, S., Monnier, N., Zhou, J., Faure, J., Smirnow, D.A., Carlier, R., Laroche, C., Marcorelles, P., Mercier, S.*, et al.* (2013). The neuronal endopeptidase ECEL1 is associated with a distinct form of recessive distal arthrogryposis. Human molecular genetics *22*, 1483-1492.
Erlebacher, A., Filvaroff, E.H., Gitelman, S.E., and Derynck, R. (1995). Toward a molecular understanding of skeletal development. Cell *80*, 371-378.
Eswarakumar, V.P., Horowitz, M.C., Locklin, R., Morriss-Kay, G.M., and Lonai, P. (2004). A gain-of-function mutation of Fgfr2c demonstrates the roles of this receptor variant in osteogenesis. Proceedings of the National Academy of Sciences of the United States of America *101*, 12555-12560.
Fairbrother, W.G., Yeo, G.W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.A., and Burge, C.B. (2004). RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic acids research *32*, W187-190.
Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. Molecular cell *16*, 655-661.
Fenwick, A.L., Bowdin, S.C., Klatt, R.E.M., and Wilkie, A.O.M. (2011). A deletion of FGFR2 creating a chimeric IIIb/IIIc exon in a child with Apert syndrome. BMC medical genetics *12*.
Fenwick, A.L., Kliszczak, M., Cooper, F., Murray, J., Sanchez-Pulido, L., Twigg, S.R., Goriely, A., McGowan, S.J., Miller, K.A., Taylor, I.B.*, et al.* (2016). Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. Am J Hum Genet *99*, 125-138.
Florisson, J.M., Verkerk, A.J., Huigh, D., Hoogeboom, A.J., Swagemakers, S., Kremer, A., Heijsman, D., Lequin, M.H., Mathijssen, I.M., and van der Spek, P.J. (2013). Boston type craniosynostosis: report of a second mutation in MSX2. American journal of medical genetics Part A *161A*, 2626-2633.

Fluck, C.E., Tajima, T., Pandey, A.V., Arlt, W., Okuhara, K., Verge, C.F., Jabs, E.W., Mendonca, B.B., Fujieda, K., and Miller, W.L. (2004). Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. Nature genetics *36*, 228-230.
Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., Kraft, K., Kempfer, R., Jerkovic, I., Chan, W.L.*, et al.* (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265-269.
Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J.*, et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216-220.
Fukami, M., Nagai, T., Mochizuki, H., Muroya, K., Yamada, G., Takitani, K., and Ogata, T. (2010). Anorectal and urinary anomalies and aberrant retinoic acid metabolism in cytochrome P450 oxidoreductase deficiency. Molecular genetics and metabolism *100*, 269-273.
Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A.*, et al.* (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature *511*, 344-347.
Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. European journal of human genetics : EJHG *20*, 490-497.
Gonzaga-Jauregui, C., Lupski, J.R., and Gibbs, R.A. (2012). Human genome sequencing in health and disease. Annual review of medicine *63*, 35-61.
Goos, J.A., Fenwick, A.L., Swagemakers, S.M., McGowan, S.J., Knight, S.J., Twigg, S.R., Hoogeboom, A.J., van Dooren, M.F., Magielsen, F.J., Wall, S.A.*, et al.* (2016). Identification of Intragenic Exon Deletions and Duplication of TCF12 by Whole Genome or Targeted Sequencing as a Cause of TCF12-Related Craniosynostosis. Hum Mutat *37*, 732-736.
Graham, J.M., Jr., Badura, R.J., and Smith, D.W. (1980). Coronal craniostenosis: fetal head constraint as one possible cause. Pediatrics *65*, 995-999.
Graham, J.M., Jr., deSaxe, M., and Smith, D.W. (1979). Sagittal craniostenosis: fetal head constraint as one possible cause. The Journal of pediatrics *95*, 747-750.
Graul-Neumann, L.M., Deichsel, A., Wille, U., Kakar, N., Koll, R., Bassir, C., Ahmad, J., Cormier-Daire, V., Mundlos, S., Kubisch, C.*, et al.* (2014). Homozygous missense and nonsense mutations in BMPR1B cause acromesomelic chondrodysplasia-type Grebe. European journal of human genetics : EJHG *22*, 726-733.
Grillo, L., Greco, D., Pettinato, R., Avola, E., Potenza, N., Castiglia, L., Spalletta, A., Amata, S., Di Benedetto, D., Luciano, D.*, et al.* (2014). Increased FGF3 and FGF4 gene dosage is a risk factor for craniosynostosis. Gene *534*, 435-439.
Gudjonsson, J.E., Karason, A., Runarsdottir, E.H., Antonsdottir, A.A., Hauksson, V.B., Jonsson, H.H., Gulcher, J., Stefansson, K., and Valdimarsson, H. (2006). Distinct clinical differences between HLA-Cw*0602 positive and negative psoriasis patients--an analysis of 1019 HLA-C- and HLA-B-typed patients. J Invest Dermatol *126*, 740-745.
Hajihosseini, M.K., and Heath, J.K. (2002). Expression patterns of fibroblast growth factors-18 and -20 in mouse embryos is suggestive of novel roles in calvarial and limb development. Mechanisms of development *113*, 79-83.
Hebert, J.M., Rosenquist, T., Gotz, J., and Martin, G.R. (1994). FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. Cell *78*, 1017-1025.
Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J., and Shendure, J. (2013). Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome research *23*, 843-854.
Higginbottom, M.C., Jones, K.L., and James, H.E. (1980). Intrauterine constraint and craniosynostosis. Neurosurgery *6*, 39-44.
Hoepfner, S., Severin, F., Cabezas, A., Habermann, B., Runge, A., Gillooly, D., Stenmark, H., and Zerial, M. (2005). Modulation of receptor recycling and degradation by the endosomal kinesin KIF16B. Cell *121*, 437-450.

Hoischen, A., van Bon, B.W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G.*, et al.* (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nature genetics *42*, 483-485.
Hoischen, A., van Bon, B.W., Rodriguez-Santiago, B., Gilissen, C., Vissers, L.E., de Vries, P., Janssen, I., van Lier, B., Hastings, R., Smithson, S.F.*, et al.* (2011). De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nature genetics *43*, 729-731.
Housley, D.J., and Venta, P.J. (2006). The long and the short of it: evidence that FGF5 is a major determinant of canine 'hair'-itability. Animal genetics *37*, 309-315.
Huelke, D.F. (1998). An overview of anatomical considerations of infants and children in the adult world of automobile safety design. Annu Proc Assoc Adv Autmot Med *42*, 93-113.
Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet *46*, 205-212.
Hunter, A.G., and Rudd, N.L. (1976). Craniosynostosis. I. Sagittal synostosis: its genetics and associated clinical findings in 214 patients who lacked involvement of the coronal suture(s). Teratology *14*, 185-193.
Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nature genetics *36*, 949-951.
Jabs, E.W., Muller, U., Li, X., Ma, L., Luo, W., Haworth, I.S., Klisak, I., Sparkes, R., Warman, M.L., Mulliken, J.B.*, et al.* (1993). A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. Cell *75*, 443-450.
Janssen, A., Hosen, M.J., Jeannin, P., Coucke, P.J., De Paepe, A., and Vanakker, O.M. (2013). Second family with the Boston-type craniosynostosis syndrome: novel mutation and expansion of the clinical spectrum. American journal of medical genetics Part A *161A*, 2352-2357.
Jeradi, S., and Hammerschmidt, M. (2016). Retinoic acid-induced premature osteoblast-to-preosteocyte transitioning has multiple effects on calvarial development. Development *143*, 1205-1216.
Jiang, X., Iseki, S., Maxson, R.E., Sucov, H.M., and Morriss-Kay, G.M. (2002a). Tissue origins and interactions in the mammalian skull vault. Developmental biology *241*, 106-116.
Jiang, Y., Jahagirdar, B.N., Reinhardt, R.L., Schwartz, R.E., Keene, C.D., Ortiz-Gonzalez, X.R., Reyes, M., Lenvik, T., Lund, T., Blackstad, M.*, et al.* (2002b). Pluripotency of mesenchymal stem cells derived from adult marrow. Nature *418*, 41-49.
Johnson, D., Horsley, S.W., Moloney, D.M., Oldridge, M., Twigg, S.R., Walsh, S., Barrow, M., Njolstad, P.R., Kunz, J., Ashworth, G.J.*, et al.* (1998). A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. Am J Hum Genet *63*, 1282-1293.
Johnson, D., and Wilkie, A.O. (2011). Craniosynostosis. Eur J Hum Genet *19*, 369-376.
Justice, C.M., Yagnik, G., Kim, Y., Peter, I., Jabs, E.W., Erazo, M., Ye, X., Ainehsazan, E., Shi, L., Cunningham, M.L.*, et al.* (2012). A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. Nature genetics *44*, 1360-1364.
Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. Journal of the Royal Society, Interface / the Royal Society *10*, 20121022.
Kam, R.K., Shi, W., Chan, S.O., Chen, Y., Xu, G., Lau, C.B., Fung, K.P., Chan, W.Y., and Zhao, H. (2013). Dhrs3 protein attenuates retinoic acid signaling and is required for early embryonic patterning. The Journal of biological chemistry *288*, 31477-31487.
Kedishvili, N.Y. (2013). Enzymology of retinoic acid biosynthesis and degradation. Journal of lipid research *54*, 1744-1760.
Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K.*, et al.* (2008). Disruption of neurexin 1 associated with autism spectrum disorder. American journal of human genetics *82*, 199-207.

Kim, J., Ault, K.T., Chen, H.D., Xu, R.H., Roh, D.H., Lin, M.C., Park, M.J., and Kung, H.F. (1998). Transcriptional regulation of BMP-4 in the Xenopus embryo: analysis of genomic BMP-4 and its promoter. Biochemical and biophysical research communications *250*, 516-530.
Kimonis, V., Gold, J.A., Hoffman, T.L., Panchal, J., and Boyadjiev, S.A. (2007). Genetics of craniosynostosis. Seminars in pediatric neurology *14*, 150-161.
Kirmi, O., Lo, S.J., Johnson, D., and Anslow, P. (2009a). Craniosynostosis: A Radiological and Surgical Perspective. Semin Ultrasound Ct *30*, 492-512.
Kirmi, O., Lo, S.J., Johnson, D., and Anslow, P. (2009b). Craniosynostosis: a radiological and surgical perspective. Seminars in ultrasound, CT, and MR *30*, 492-512.
Klopocki, E., Lohan, S., Brancati, F., Koll, R., Brehm, A., Seemann, P., Dathe, K., Stricker, S., Hecht, J., Bosse, K.*, et al.* (2011). Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. Am J Hum Genet *88*, 70-75.
Koboldt, D.C., Larson, D.E., Sullivan, L.S., Bowne, S.J., Steinberg, K.M., Churchill, J.D., Buhr, A.C., Nutter, N., Pierce, E.A., Blanton, S.H.*, et al.* (2014). Exome-based mapping and variant prioritization for inherited Mendelian disorders. American journal of human genetics *94*, 373-384.
Koster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520-2522.
Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols *4*, 1073-1081.
Kurth, I., Klopocki, E., Stricker, S., van Oosterwijk, J., Vanek, S., Altmann, J., Santos, H.G., van Harssel, J.J., de Ravel, T., Wilkie, A.O.*, et al.* (2009). Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. Nature genetics *41*, 862-863.
Lajeunie, E., Barcik, U., Thorne, J.A., El Ghouzzi, V., Bourgeois, M., and Renier, D. (2001). Craniosynostosis and fetal exposure to sodium valproate. Journal of neurosurgery *95*, 778-782.
Lajeunie, E., Le Merrer, M., Bonaiti-Pellie, C., Marchac, D., and Renier, D. (1995). Genetic study of nonsyndromic coronal craniosynostosis. American journal of medical genetics *55*, 500-504.
Lajeunie, E., Le Merrer, M., Bonaiti-Pellie, C., Marchac, D., and Renier, D. (1996). Genetic study of scaphocephaly. American journal of medical genetics *62*, 282-285.
Lajeunie, E., Le Merrer, M., Marchac, D., and Renier, D. (1998). Syndromal and nonsyndromal primary trigonocephaly: analysis of a series of 237 patients. American journal of medical genetics *75*, 211-215.
Lammer, E.J., Chen, D.T., Hoar, R.M., Agnish, N.D., Benke, P.J., Braun, J.T., Curry, C.J., Fernhoff, P.M., Grix, A.W., Jr., Lott, I.T.*, et al.* (1985). Retinoic acid embryopathy. N Engl J Med *313*, 837-841.
Lana-Elola, E., Rice, R., Grigoriadis, A.E., and Rice, D.P. (2007). Cell fate specification during calvarial bone and suture development. Developmental biology *311*, 335-346.
Laue, K., Pogoda, H.M., Daniel, P.B., van Haeringen, A., Alanay, Y., von Ameln, S., Rachwalski, M., Morgan, T., Gray, M.J., Breuning, M.H.*, et al.* (2011). Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. Am J Hum Genet *89*, 595-606.
Le Tanno, P., Poreau, B., Devillard, F., Vieville, G., Amblard, F., Jouk, P.S., Satre, V., and Coutton, C. (2014). Maternal complex chromosomal rearrangement leads to TCF12 microdeletion in a patient presenting with coronal craniosynostosis and intellectual disability. Am J Med Genet A *164A*, 1530-1536.
Lehmann, K., Seemann, P., Stricker, S., Sammar, M., Meyer, B., Suring, K., Majewski, F., Tinschert, S., Grzeschik, K.H., Muller, D.*, et al.* (2003). Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2. Proceedings of the National Academy of Sciences of the United States of America *100*, 12277-12282.

Levi, B., Wan, D.C., Wong, V.W., Nelson, E., Hyun, J., and Longaker, M.T. (2012). Cranial suture biology: from pathways to patient care. The Journal of craniofacial surgery *23*, 13-19.
Li, D., Zhang, H., Ma, L., Han, Y., Xu, M., Wang, Z., Jiang, H., Zhang, W., Wang, L., and Pan, Y. (2016). Associations between microRNA binding site SNPs in FGFs and FGFRs and the risk of non-syndromic orofacial cleft. Scientific reports *6*, 31054.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Liu, X., Gao, L., Zhao, A., Zhang, R., Ji, B., Wang, L., Zheng, Y., Zeng, B., Valenzuela, R.K., He, L.*, et al.* (2014). Identification of duplication downstream of BMP2 in a Chinese family with brachydactyly type A2 (BDA2). PLoS One *9*, e94201.
Lowery, J.W., Lavigne, A.W., Kokabu, S., and Rosen, V. (2013). Comparative genomics identifies the mouse Bmp3 promoter and an upstream evolutionary conserved region (ECR) in mammals. PloS one *8*, e57840.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
Lundova, T., Zemanova, L., Malcekova, B., Skarka, A., Stambergova, H., Havrankova, J., Safr, M., and Wsol, V. (2015). Molecular and biochemical characterisation of human short-chain dehydrogenase/reductase member 3 (DHRS3). Chemico-biological interactions *234*, 178-187.
Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.
MacLean, G., Li, H., Metzger, D., Chambon, P., and Petkovich, M. (2007). Apoptotic extinction of germ cells in testes of Cyp26b1 knockout mice. Endocrinology *148*, 4560-4567.
Mammadova, A., Zhou, H., Carels, C.E., and Von den Hoff, J.W. (2016). Retinoic acid signalling in the development of the epidermis, the limbs and the secondary palate. Differentiation; research in biological diversity *92*, 326-335.
Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487-498.
Matsubara, K., Sano, N., Umehara, T., and Horikoshi, M. (2007). Global analysis of functional surfaces of core histones with comprehensive point mutants. Genes to cells : devoted to molecular & cellular mechanisms *12*, 13-33.
Mattiroli, F., Uckelmann, M., Sahtoe, D.D., van Dijk, W.J., and Sixma, T.K. (2014). The nucleosome acidic patch plays a critical role in RNF168-dependent ubiquitination of histone H2A. Nature communications *5*, 3291.
McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.
McMillin, M.J., Below, J.E., Shively, K.M., Beck, A.E., Gildersleeve, H.I., Pinner, J., Gogola, G.R., Hecht, J.T., Grange, D.K., Harris, D.J.*, et al.* (2013). Mutations in ECEL1 cause distal arthrogryposis type 5D. American journal of human genetics *92*, 150-156.
McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A.*, et al.* (2008). Runs of homozygosity in European populations. American journal of human genetics *83*, 359-372.
Miller, K.A., Twigg, S.R., McGowan, S.J., Phipps, J.M., Fenwick, A.L., Johnson, D., Wall, S.A., Noons, P., Rees, K.E., Tidey, E.A.*, et al.* (2017). Diagnostic value of exome and whole genome sequencing in craniosynostosis. J Med Genet *54*, 260-268.
Mishina, Y., and Snider, T.N. (2014). Neural crest cell signaling pathways critical to cranial bone development and pathology. Exp Cell Res *325*, 138-147.
Morriss-Kay, G.M., and Wilkie, A.O. (2005). Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. Journal of anatomy *207*, 637-653.
Morton, J.E., Frentz, S., Morgan, T., Sutherland-Smith, A.J., and Robertson, S.P. (2016). Biallelic mutations in CYP26B1: A differential diagnosis for Pfeiffer and Antley-Bixler syndromes. American journal of medical genetics Part A *170*, 2706-2710.

Nestle, F.O., Kaplan, D.H., and Barker, J. (2009). Psoriasis. The New England journal of medicine *361*, 496-509.
Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A.*, et al.* (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat Genet *42*, 30-35.
Nie, X., Luukko, K., and Kettunen, P. (2006a). BMP signalling in craniofacial development. The International journal of developmental biology *50*, 511-521.
Nie, X., Luukko, K., and Kettunen, P. (2006b). FGF signalling in craniofacial development and developmental disorders. Oral diseases *12*, 102-111.
Norton, N.S. (2016). Netter's head and neck anatomy for dentistry (ELSEVIER).
O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K.*, et al.* (2012a). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619-1622.
O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D.*, et al.* (2012b). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246-250.
Oldridge, M., Zackai, E.H., McDonald-McGinn, D.M., Iseki, S., Morriss-Kay, G.M., Twigg, S.R., Johnson, D., Wall, S.A., Jiang, W., Theda, C.*, et al.* (1999). De novo alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. American journal of human genetics *64*, 446-461.
Opperman, L.A. (2000). Cranial sutures as intramembranous bone growth sites. Developmental dynamics : an official publication of the American Association of Anatomists *219*, 472-485.
Ornitz, D.M., and Itoh, N. (2015). The Fibroblast Growth Factor signaling pathway. Wiley interdisciplinary reviews Developmental biology *4*, 215-266.
Ott, J., Wang, J., and Leal, S.M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. Nature reviews Genetics *16*, 275-284.
Otto, D.M., Henderson, C.J., Carrie, D., Davey, M., Gundersen, T.E., Blomhoff, R., Adams, R.H., Tickle, C., and Wolf, C.R. (2003). Identification of novel roles of the cytochrome p450 system in early embryogenesis: effects on vasculogenesis and retinoic Acid homeostasis. Molecular and cellular biology *23*, 6103-6116.
Pandey, A.V., and Fluck, C.E. (2013). NADPH P450 oxidoreductase: structure, function, and pathology of diseases. Pharmacol Ther *138*, 229-254.
Pang, A.W., MacDonald, J.R., Pinto, D., Wei, J., Rafiq, M.A., Conrad, D.F., Park, H., Hurles, M.E., Lee, C., Venter, J.C.*, et al.* (2010). Towards a comprehensive structural variation map of an individual human genome. Genome biology *11*, R52.
Pennimpede, T., Cameron, D.A., MacLean, G.A., Li, H., Abu-Abed, S., and Petkovich, M. (2010). The role of CYP26 enzymes in defining appropriate retinoic acid exposure during embryogenesis. Birth defects research Part A, Clinical and molecular teratology *88*, 883-894.
Persson, B., Kallberg, Y., Bray, J.E., Bruford, E., Dellaporta, S.L., Favia, A.D., Duarte, R.G., Jornvall, H., Kavanagh, K.L., Kedishvili, N.*, et al.* (2009). The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. Chemico-biological interactions *178*, 94-98.
Pertea, M., Lin, X., and Salzberg, S.L. (2001). GeneSplicer: a new computational method for splice site prediction. Nucleic acids research *29*, 1185-1190.
Pop, R., Conz, C., Lindenberg, K.S., Blesson, S., Schmalenberger, B., Briault, S., Pfeifer, D., and Scherer, G. (2004). Screening of the 1 Mb SOX9 5' control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. J Med Genet *41*, e47.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
Rasmussen, S.A., Yazdy, M.M., Carmichael, S.L., Jamieson, D.J., Canfield, M.A., and Honein, M.A. (2007). Maternal thyroid disease as a risk factor for craniosynostosis. Obstetrics and gynecology *110*, 369-377.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333-i339.
Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.A.*, et al.* (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nature genetics *49*, 36-45.
Reijntjes, S., Rodaway, A., and Maden, M. (2007). The retinoic acid metabolising gene, CYP26B1, patterns the cartilaginous cranial neural crest in zebrafish. The International journal of developmental biology *51*, 351-360.
Reissner, C., Klose, M., Fairless, R., and Missler, M. (2008). Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. Proceedings of the National Academy of Sciences of the United States of America *105*, 15124-15129.
Rhinn, M., and Dolle, P. (2012). Retinoic acid signalling during development. Development *139*, 843-858.
Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S.R.F., Wilkie, A.O.M., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nature genetics *46*, 912-918.
Rodriguez-Revenga, L., Mila, M., Rosenberg, C., Lamb, A., and Lee, C. (2007). Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. Genetics in medicine : official journal of the American College of Medical Genetics *9*, 600-606.
Rothman, K.J., Moore, L.L., Singer, M.R., Nguyen, U.S., Mannino, S., and Milunsky, A. (1995). Teratogenicity of high vitamin A intake. N Engl J Med *333*, 1369-1373.
Sadler, T.W., and Langman, J. (2006). Langman's medical embryology, 10th edn (Philadelphia, Lippincott Williams & Wilkins).
Sanchez-Lara, P.A., Carmichael, S.L., Graham, J.M., Jr., Lammer, E.J., Shaw, G.M., Ma, C., and Rasmussen, S.A. (2010). Fetal constraint as a potential risk factor for craniosynostosis. American journal of medical genetics Part A *152A*, 394-400.
Schaaf, C.P., Boone, P.M., Sampath, S., Williams, C., Bader, P.I., Mueller, J.M., Shchelochkov, O.A., Brown, C.W., Crawford, H.P., Phalen, J.A.*, et al.* (2012). Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions. European journal of human genetics : EJHG *20*, 1240-1247.
Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nature protocols *3*, 1101-1108.
Schoenebeck, J.J., Hutchinson, S.A., Byers, A., Beale, H.C., Carrington, B., Faden, D.L., Rimbault, M., Decker, B., Kidd, J.M., Sood, R.*, et al.* (2012). Variation of BMP3 contributes to dog breed skull diversity. PLoS Genet *8*, e1002849.
Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods *7*, 575-576.
Schwerd, T., Twigg, S.R.F., Aschenbrenner, D., Manrique, S., Miller, K.A., Taylor, I.B., Capitani, M., McGowan, S.J., Sweeney, E., Weber, A.*, et al.* (2017). A biallelic mutation in IL6ST encoding the GP130 co-receptor causes immunodeficiency and craniosynostosis. The Journal of experimental medicine *214*, 2547-2562.
Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. Cell *160*, 1049-1059.
Sharma, V.P., Fenwick, A.L., Brockop, M.S., McGowan, S.J., Goos, J.A., Hoogeboom, A.J., Brady, A.F., Jeelani, N.O., Lynch, S.A., Mulliken, J.B.*, et al.* (2013). Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. Nat Genet *45*, 304-307.
Shaytan, A.K., Landsman, D., and Panchenko, A.R. (2015). Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers. Current opinion in structural biology *32*, 48-57.
Shier D, B.J., Lewis R (2012). Hole's Human Anatomy &Amp; Physiology, 13 th Edition 13th edition

Siepel, A., Pollard, K.S., and Haussler, D. (2006). New Methods for Detecting Lineage-Specific Selection. Research in Computational Molecular Biology *Lecture Notes in Computer Science, 3909*.
Singh, R., Mortazavi, A., Telu, K.H., Nagarajan, P., Lucas, D.M., Thomas-Ahner, J.M., Clinton, S.K., Byrd, J.C., Freitas, M.A., and Parthun, M.R. (2013). Increasing the complexity of chromatin: functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. Nucleic acids research *41*, 9284-9295.
Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Human mutation *36*, 928-930.
Sommer, A., Katz, J., and Tarwotjo, I. (1984). Increased risk of respiratory disease and diarrhea in children with preexisting mild vitamin A deficiency. Am J Clin Nutr *40*, 1090-1095.
Stefan, C.P., Koehler, J.W., and Minogue, T.D. (2016). Targeted next-generation sequencing for the detection of ciprofloxacin resistance markers using molecular inversion probes. Scientific reports *6*, 25904.
Stuhlmiller, T.J., and Garcia-Castro, M.I. (2012). Current perspectives of the signaling pathways directing neural crest induction. Cellular and molecular life sciences : CMLS *69*, 3715-3737.
Sund, K.L., Zimmerman, S.L., Thomas, C., Mitchell, A.L., Prada, C.E., Grote, L., Bao, L., Martin, L.J., and Smolarek, T.A. (2013). Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. Genetics in medicine : official journal of the American College of Medical Genetics *15*, 70-78.
Takarada, T., Nakazato, R., Tsuchikane, A., Fujikawa, K., Iezaki, T., Yoneda, Y., and Hinoi, E. (2016). Genetic analysis of Runx2 function during intramembranous ossification. Development *143*, 211-218.
Tattini, L., D'Aurizio, R., and Magi, A. (2015). Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Front Bioeng Biotechnol *3*, 92.
Teare, M.D., and Santibanez Koref, M.F. (2014). Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. Briefings in functional genomics *13*, 378-383.
Tessadori, F., Giltay, J.C., Hurst, J.A., Massink, M.P., Duran, K., Vos, H.R., van Es, R.M., Scott, R.H., van Gassen, K.L.I., Bakkers, J.*, et al.* (2017). Germline mutations affecting the histone H4 core cause a developmental syndrome by altering DNA damage response and cell cycle control. Nature genetics.
Timberlake, A.T., Choi, J., Zaidi, S., Lu, Q., Nelson-Williams, C., Brooks, E.D., Bilguvar, K., Tikhonova, I., Mane, S., Yang, J.F.*, et al.* (2016). Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. eLife *5*.
Timberlake, A.T., Furey, C.G., Choi, J., Nelson-Williams, C., Loring, E., Galm, A., Kahle, K.T., Steinbacher, D.M., Larysz, D., Persing, J.A.*, et al.* (2017). De novo mutations in inhibitors of Wnt, BMP, and Ras/ERK signaling pathways in non-syndromic midline craniosynostosis. Proceedings of the National Academy of Sciences of the United States of America *114*, E7341-E7347.
Tomalik-Scharte, D., Maiter, D., Kirchheiner, J., Ivison, H.E., Fuhr, U., and Arlt, W. (2010). Impaired hepatic drug and steroid metabolism in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. European journal of endocrinology *163*, 919-924.
Twigg, S.R., Forecki, J., Goos, J.A., Richardson, I.C., Hoogeboom, A.J., van den Ouweland, A.M., Swagemakers, S.M., Lequin, M.H., Van Antwerp, D., McGowan, S.J.*, et al.* (2015). Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. Am J Hum Genet *97*, 378-388.
Twigg, S.R., Lloyd, D., Jenkins, D., Elcioglu, N.E., Cooper, C.D., Al-Sannaa, N., Annagur, A., Gillessen-Kaesbach, G., Huning, I., Knight, S.J.*, et al.* (2012). Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. Am J Hum Genet *91*, 897-905.

Twigg, S.R., Vorgia, E., McGowan, S.J., Peraki, I., Fenwick, A.L., Sharma, V.P., Allegra, M., Zaragkoulias, A., Sadighi Akha, E., Knight, S.J.*, et al.* (2013). Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis. Nat Genet *45*, 308-313.
Twigg, S.R., and Wilkie, A.O. (2015). A Genetic-Pathophysiological Framework for Craniosynostosis. Am J Hum Genet *97*, 359-377.
Twigg, S.R.F., Hufnagel, R.B., Miller, K.A., Zhou, Y., McGowan, S.J., Taylor, J., Craft, J., Taylor, J.C., Santoro, S.L., Huang, T.*, et al.* (2016). A Recurrent Mosaic Mutation in SMO, Encoding the Hedgehog Signal Transducer Smoothened, Is the Major Cause of Curry-Jones Syndrome. American journal of human genetics *98*, 1256-1265.
van der Meulen, J., van der Hulst, R., van Adrichem, L., Arnaud, E., Chin-Shong, D., Duncan, C., Habets, E., Hinojosa, J., Mathijssen, I., May, P.*, et al.* (2009). The increase of metopic synostosis: a pan-European observation. J Craniofac Surg *20*, 283-286.
Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M.*, et al.* (2010). A de novo paradigm for mental retardation. Nature genetics *42*, 1109-1112.
Vu, H.L., Panchal, J., Parker, E.E., Levine, N.S., and Francel, P. (2001). The timing of physiologic closure of the metopic suture: a review of 159 patients using reconstructed 3D CT scans of the craniofacial region. The Journal of craniofacial surgery *12*, 527-532.
Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K.B., King, M.C.*, et al.* (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. Am J Hum Genet *87*, 90-94.
Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research *38*, e164.
Wilkie, A.O. (1997). Craniosynostosis: genes and mechanisms. Hum Mol Genet *6*, 1647-1656.
Wilkie, A.O., Byren, J.C., Hurst, J.A., Jayamohan, J., Johnson, D., Knight, S.J., Lester, T., Richards, P.G., Twigg, S.R., and Wall, S.A. (2010). Prevalence and complications of single-gene and chromosomal disorders in craniosynostosis. Pediatrics *126*, e391-400.
Will, A.J., Cova, G., Osterwalder, M., Chan, W.L., Wittler, L., Brieske, N., Heinrich, V., de Villartay, J.P., Vingron, M., Klopocki, E.*, et al.* (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nature genetics *49*, 1539-1545.
Yashiro, K., Zhao, X., Uehara, M., Yamashita, K., Nishijima, M., Nishino, J., Saijoh, Y., Sakai, Y., and Hamada, H. (2004). Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. Developmental cell *6*, 411-422.
Yasuda, Y., Hashimoto, R., Yamamori, H., Ohi, K., Fukumoto, M., Umeda-Yano, S., Mohri, I., Ito, A., Taniike, M., and Takeda, M. (2011). Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder. Molecular autism *2*, 9.
Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Journal of computational biology : a journal of computational molecular cell biology *11*, 377-394.
Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.A., and Friedman, J.M. (2008). A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. Journal of medical genetics *45*, 239-243.
Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K.*, et al.* (2013). De novo mutations in histone-modifying genes in congenital heart disease. Nature *498*, 220-223.
Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. Hum Mol Genet *24*, R102-110.
Zhang, F., Seeman, P., Liu, P., Weterman, M.A., Gonzaga-Jauregui, C., Towne, C.F., Batish, S.D., De Vriendt, E., De Jonghe, P., Rautenstrauss, B.*, et al.* (2010). Mechanisms for

nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. Am J Hum Genet *86*, 892-903.
Zhang, M.Q. (1998). Statistical features of human exons and their flanking regions. Human molecular genetics *7*, 919-932.
Zhang, Z., Yang, W., Cao, Y., Shi, Y., Lei, C., Du, B., Li, X., and Zhang, Q. (2015). The Functions of BMP3 in Rabbit Articular Cartilage Repair. International journal of molecular sciences *16*, 25934-25946.
Zhou, J., Fan, J.Y., Rangasamy, D., and Tremethick, D.J. (2007). The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. Nature structural & molecular biology *14*, 1070-1076.