UNIVERSITY OF OXFORD

THE IDENTIFICATION AND CHARACTERISATION OF DISEASE GENES IN CRANIOSYNOSTOSIS

AIMEE LAURA FENWICK

LINACRE COLLEGE

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Trinity Term 2015

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

Statement of Originality

The *FGFR2* mutation c.1083A>G (section 3.2.1) was identified by the Genetics Laboratories at the Churchill Hospital (Oxford University Hospitals NHS Trust, Oxford, UK). The *FGFR2* mutation c.1083A>T (section 3.2.2) was identified by the Department of Clinical Genetics, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands.

The Fluidigm PCR and Ion Torrent sequencing experiments (sections 4.5.1 and 4.5.2) were designed and undertaken by Indira Taylor, Kerry Miller and Yan Zhou (Clinical Genetics Group, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK).

Exome sequencing of the individual with the *TWIST1* c.350A>T (p.Glu117Val) mutation, as well as the parental samples, was performed and analysed by Steve Twigg (section 6.2.1; Clinical Genetics Group, Weatherall Institute of Molecular Medicine). *TWIST1* sequencing and subsequent identification of the c.350A>G (p.Glu117Gly) mutation was by the Genetics Laboratories at the Churchill Hospital, Oxford (section 6.2.2).

ABSTRACT

The identification and characterisation of disease genes in craniosynostosis

A thesis submitted for the degree of D. Phil by Aimee L Fenwick, Linacre College, Oxford.
Trinity Term 2015

Current challenges to the understanding and clinical management of craniosynostosis (premature fusion of the cranial sutures) include the interpretation of changes in known disease genes, the identification of novel disease genes and the identification of pathogenic regulatory mutations. This thesis aims to address these issues and by doing so improve diagnosis, surgical prognosis and support for affected families.

Expanding the repertoire of known disease mechanisms, pathogenicity of an apparently synonymous substitution in FGFR2 is demonstrated, associated with a mild Crouzon syndrome phenotype. Furthermore, a new severe phenotype associated with localised mutations in *TWIST1* (usually associated with Saethre-Chotzen syndrome; SCS) is described.

In an attempt to better understand the underlying pathogenesis of SCS, six probands with a clinical diagnosis, but no *TWIST1* mutations, underwent whole-genome sequencing. Two individuals were found to have mutations in *TCF12*, a dimerization partner of *TWIST1*, and one patient had compound heterozygous mutations in *CDC45*. The latter, however, was not initially prioritised as one of the mutations led to a silent substitution, and this individual also had a *de novo* variant in a gene that interacts with a known craniosynostosis pathway.

To further explore the contribution of regulatory mutations to the pathogenesis of SCS, targeted sequencing of a 2.4 Mb region around *TWIST1*, and 640 kb around *TCF12*, was undertaken on DNA from 160 coronal synostosis patients. As well as coding mutations that had been missed, intronic splicing mutations and putative *TWIST1* promoter mutations were identified.

In summary, during the course of this work novel and sometimes unexpected pathogenic mechanisms of mutations both in previously known genes (*FGFR2*, *TWIST1*) and in newly defined genes (*TCF12*, *CDC45*) have been identified. The pathogenicity of apparently synonymous variants provides important lessons in the correct interpretation of patient sequence data, and the identification of novel disease genes expands the classification and range of pathophysiology in craniosynostosis.

ACKNOWLEDGEMENTS

This work was carried out at the Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford. It was funded by the WIMM and an EPA Cephalosporin scholarship from Linacre College.

My first and biggest thanks must go to my supervisors Andrew Wilkie and Steve Twigg, I have learned an incredible amount from being part of the Wilkie lab and I could not have completed this work without their guidance and advice. A huge thank you must also go to Simon McGowan (CBRG) for bioinformatics support - this would have been impossible without his help, and I also appreciated all the cycling chats!

I would also like to thank past and present members of the Wilkie lab for all their help and discussions (both scientific and otherwise) over the years, including Anne Goriely, Chris Babbs, Indira Taylor, Deborah Lloyd, Geoff Maher, Kerry Miller, Vikram Sharma, Akiko Hashimoto and Yan Zhou. This can also be extended to the Rotterdam group whose collaboration has been invaluable, in particular Irene Mahijssen and Jacqueline Goos.

My gratitude also goes to 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.

Finally I'd like to thank all my friends and family for their support, especially my husband Chris for being generally amazing and always making me smile.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

AMS	Ablepharon macrostomia syndrome
bHLH	Basic helix-loop-helix
BGS	Baller-Gerold syndrome
BSS	Barber-Say syndrome
CCD	Cleidocranial dysplasia
CFNS	Craniofrontonasal syndrome
ChIP-seq	ChIP coupled with high-throughput sequencing
CNC	Cranial neural crest
CNV	Copy number variant
E	Embryonic day
EMSA	Electrophoretic mobility shift assay
ESE	Exonic splice enhancer
ESS	Exonic splice silencer
FISH	Fluorescence in situ hybridisation
ISE	Intronic splice enhancer
ISS	Intronic splice silencer
MCM	Minichromosome maintenance proteins
MGS	Meier-Gorlin syndrome
MLPA	Multiplex-ligation-dependent probe amplification
MODY	Maturity-onset diabetes of the young
NAHR	Non-allelic homologous recombination
ORC	Origin recognition complex
ORIs	Origins of replication
P	Postnatal day
Pre-RC	Pre-replicative complex
qPCR	Quantitative PCR
RCPS	Richieri-Costa-Pereira syndrome
RTK	Receptor tyrosine kinase
RTS	Rothmund-Thomson syndrome
SCS	Saethre-Chotzen syndrome
SNP	Single nucleotide polymorphism
uORF	Upstream open reading frame

CHAPTER 1

INTRODUCTION

Chapter 1: Introduction

1.1 Overview

Craniosynostosis describes the premature fusion of one or more of the cranial sutures. It affects 1 in 2100 – 2500 children and is the second most common craniofacial malformation after cleft lip and/or cleft palate (Boulet, Rasmussen, and Honein 2008, Lajeunie, Le Merrer, et al. 1995). This severe congenital condition often requires multiple surgeries and can involve complications with raised intra-cranial pressure, breathing, dentition, hearing and learning difficulties. This thesis aims to identify new genetic causes, and further understand known causes of craniosynostosis with the goal of improving diagnosis, surgical prognosis, counselling for affected families and also increasing our understanding of how the human skull develops.

1.2 Development of the skull

The human skull can be divided into two compartments, comprising the viscerocranium which forms the face, and the neurocranium that surrounds and protects the brain. The neurocranium can be further divided into the base of the skull, and the calvaria (skull vault). Five separate bones make up the calvaria: the paired frontals and parietals, and the unpaired occipital (Morriss-Kay and Wilkie 2005, Bonaventure and El Ghouzzi 2003) Fig. 1-1.

At around embryonic day eight (E8) in the mouse, the cranial neural crest (CNC) begins to migrate away from the closing neural folds of the three-layered embryo. This migration is complete after two days. In humans, CNC migration takes place between E19 and E26 (Opperman 2000, Otis and Brent 1954). Thus, the calvarial mesenchyme is made up of cells from both the CNC and paraxial mesoderm; the parietal bones being of mesodermal origin and the frontal bones of neural crest (Jiang et al. 2002). The two cell populations migrate to defined locations overlying the cerebral hemispheres where they subsequently differentiate.

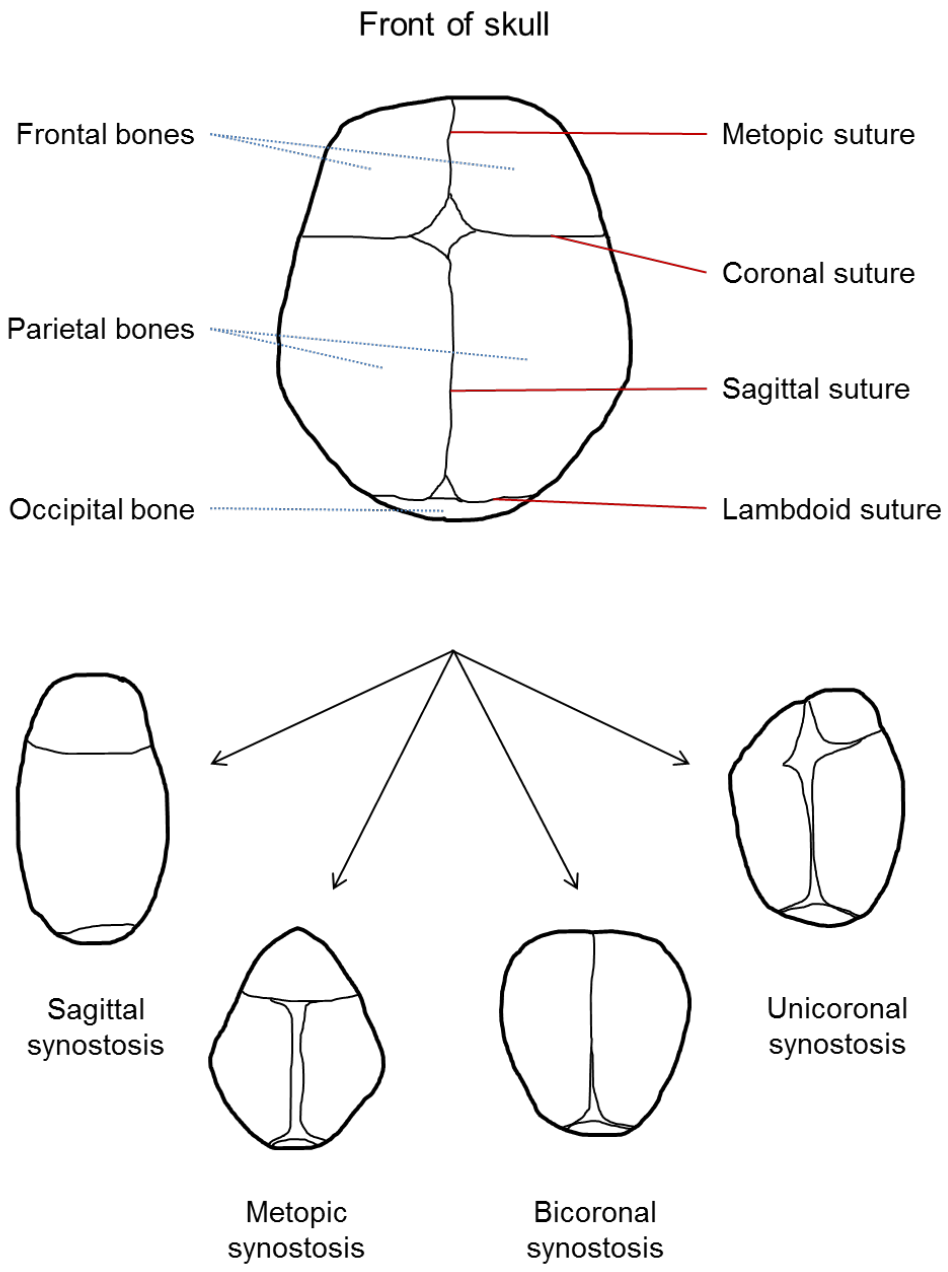


Figure 1-1

The human calvarial bones.

Schematic representation of the normal human skull is shown above, looking down at the top of the head. The five bones of the calvaria are indicated in blue, and the sutures are indicated in red. The consequence on skull shape from premature fusion of various sutures is depicted below.

In this way, skeletal development proceeds via patterning of skeletal elements, commitment of mesenchymal cells to osteogenic lineages, and terminal differentiation into osteoblasts (Deckelbaum et al. 2006). Osteoblasts are specialised cells that form bone, they express a number of bone-related extracellular matrix proteins, as well as alkaline phosphatase (Katagiri and Takahashi 2002). In mice, mandibular bone matrix appears at E13, followed by calvarial bone matrix at around E14 (Otis and Brent 1954). Once the calvarial bones have been generated they are remodeled by the resorptive activity of osteoclasts, so the bone undergoes constant formation and resorption by osteoblasts and osteoclasts respectively (Deckelbaum et al. 2006, Katagiri and Takahashi 2002).

Unlike most skeletal bones which are formed by endochondral ossification, the calvarial bones are formed by intramembranous ossification whereby mesenchymal cells cluster and differentiate directly into osteoblasts with no cartilage precursor. The osteoblasts start to secrete bone extracellular matrix (osteoid) forming bone spicules, which proceed radially from each of the clusters, or ossification centres (in humans the two ossification centres for each of the frontal and parietal bones appear in the eighth week of development (Rice 2008)). At the leading edges of this growth are the bone fronts, containing large numbers of osteoprogenitor cells and osteoblasts, which continue to secrete osteoid along the surface of the spicules, increasing their size and leading to the expansion of the bone (Morriss-Kay and Wilkie 2005, Opperman 2000).

1.2.1 Suture development

The five bones of the calvaria remain separated by fibrous tissue along their margins, indicated in Fig. 1-1, where growth is regulated by the proliferation/differentiation balance of the osteoprogenitor cells. These margins are the cranial sutures and provide the major growth centres, allowing rapid expansion of the developing skull to accommodate the growing brain, as well as passage through the birth canal (Morriss-Kay and Wilkie 2005).

In the mouse, the coronal suture is the first to be established as the ossification centres of the developing frontal and parietal bones are formed relatively closer together than those of the two

frontal or two parietal bones. As the frontal and parietal bones grow and come into apposition they overlap but remain separated by a population of undifferentiated mesenchymal cells of mesodermal origin (Jiang et al. 2002, Merrill et al. 2006). This boundary of CNC and mesoderm is established at E9 in the mouse. Lineage tracing experiments have shown that the coronal suture progenitors originate from a population of cephalic paraxial mesoderm cells that migrate to the supraorbital domain, which expand to form the coronal suture between E11-E13.5, maintaining the CNC/mesoderm boundary (Deckelbaum et al. 2012).

The interfrontal (equivalent to the human metopic), sagittal and lambdoid sutures are formed much later in development as butt joints when the bone fronts meet each other end-on (Rice 2008).

1.3 Craniosynostosis

Craniosynostosis defines the premature fusion of the cranial sutures, which usually becomes apparent between the last third of pregnancy and the end of the first year of life. Under normal circumstances, the cranial sutures will remain patent until the third decade of life except for the metopic suture which fuses within the first year (Weinzweig et al. 2003). If the sutures are obliterated by premature ossification, the skull no longer has the ability expand with the rapidly growing brain. Compensatory overgrowth at the remaining patent sutures occurs, leading to a dysmorphic head shape (Fig. 1-1) and the possibility of other associated problems such as raised intra-cranial pressure which can lead to cognitive impairment, or optic nerve damage causing impaired vision or blindness (Panigrahi 2011, Opperman 2000). A distorted skull shape can also lead to problems with breathing and dentition, as well as hearing loss. Syndromic cases, i.e. craniosynostosis associated with additional clinical features, usually require multidisciplinary management and make up approximately 20% of cases, the remainder being non-syndromic (Panigrahi 2011, Wilkie et al. 2010). Environmental factors such as external pressure on the fetal skull as a result of malposition can cause craniosynostosis (Wilkie 1997), however nearly one quarter of craniosynostosis has a genetic aetiology (Wilkie et al. 2010, Sharma et al. 2013). Of the genetic

causes, around a fifth can be attributed to mutations in several genes including *FGFR2*, *FGFR3*, *TWIST1* and *EFNB1* (see Table 1-1). In particular, the p.(Pro250Arg) substitution in FGFR3 is the single most frequent mutation associated with craniosynostosis, underlying Muenke syndrome (Nah et al. 2012).

Surgical intervention is the basis of treatment and management, directed towards maintenance of breathing, feeding, eye protection and treatment of raised intracranial pressure but also to correct visible facial differences and thereby improve quality of life. Several studies have documented the negative psychosocial impact of craniofacial dysmorphology on behavior, social interaction, creation of peer support network, and psychiatric disease. However, improvement of appearance by surgery can have positive effects in all of these spheres (Pope and Speltz 1997, Rumsey and Harcourt 2007). No medical interventions are known to prevent suture fusion (Johnson and Wilkie 2011).

A study of a 10-year birth cohort of children who required surgery because of craniosynostosis demonstrated that repeat craniofacial surgery was required for 58% of children with single-gene mutations but only 17% of those with chromosomal abnormalities, or those without a genetic diagnosis (Wilkie et al. 2010). Demonstrably, identifying the genetic causes of craniosynostosis is important not only in enabling families to have accurate counselling, but also in enabling patients to receive the appropriate surgical management.

Table 1-1

A summary of the major genetic causes of craniosynostosis.

Gene	Disorder	Inheritance pattern	Molecular pathogenesis	Typical suture fusion	Phenotype	Refs
EFNB1	Craniofrontonasal syndrome	XLD (male sparing)	Haploinsufficiency.	Coronal	Hypertelorism, notched nasal tip, chest anomalies, longitudinal splitting of nails. Heterozygous females more severely affected than hemizygous males due random X-inactivation and cellular interference.	(Twigg et al. 2004)
ERF	ERF-related craniosynostosis	AD	Haploinsufficiency.	Multisuture	Exorbitism, midface hypoplasia, Chiari type I malformation, postnatal onset of craniosynostosis.	(Twigg, Vorgia, et al. 2013)
FGFR1	Pfeiffer syndrome	AD	p.Pro252Arg gain-of-function mutation allows illegitimate ligand binding.	Coronal	Mild craniofacial features, broad medially deviated thumbs and halluces, cutaneous syndactyly.	(Muenke et al. 1994)
FGFR2	Apert syndrome	AD (n)	>98% cases caused by p.Ser252Trp or p.Pro253Arg; gain-of-function mutations allow illegitimate ligand binding.	Coronal, multisuture	Midface hypoplasia, dilated cerebral ventricles, complex syndactyly of the hands and feet.	(Wilkie, Slaney, et al. 1995)
"	Beare-Stevenson syndrome	AD (n)	>85% cases caused by p.Tyr375Cys, also p.Ser372Cys; introduction of cysteine alters disulphide bonding.	Multisuture	Choanal atresia, prominent umbilical stump, furrowed scalp/neck skin, acanthosis nigricans of flexures in survivors.	(Przylepa et al. 1996)
"	Crouzon syndrome	AD	Multiple missense mutations in IgIII domain or tyrosine kinase domains.	Multisuture, coronal, sagittal	Exorbitism, midface hypoplasia, beaked nose ("crouzonoid" facies), clinically normal hands and feet.	(Jabs et al. 1994, Reardon et al. 1994)
"	Pfeiffer syndrome	AD (n)	Overlap with above.	Multisuture	Crouzonoid facies with broad thumbs and halluces; in severe cases, cloverleaf skull, brain anomalies, abnormal tracheal sleeve, fused elbows.	(Lajeunie, Ma, et al. 1995, Rutland et al. 1995, Schell et al. 1995)
"	Bent bone dysplasia	AD (n)	Missense mutations to polar amino acids in the transmembrane domain.	Coronal	Generalised osteopenia with reduced mineralisation of the calvaria, bent long bones; perinatal lethal.	(Merrill et al. 2012)
FGFR3	Muenke syndrome	AD	p.Pro250Arg.	Coronal	May include sensorineural hearing loss, mild brachydactyly, cone-shaped epiphyses.	(Bellus et al. 1996)
POR	Antley-Bixler syndrome	AR	Cytochrome P450 oxidoreductase deficiency.	Bicoronal, multisuture	Midface hypoplasia, choanal stenosis/atresia, radiohumeral synostosis, multiple joint contractures, genital abnormalities; abnormal steroidogenesis.	(Fluck et al. 2004)
RAB23	Carpenter syndrome 1	AR	Loss-of-function, leading to dysregulation of hedgehog signaling.	Multisuture	Obesity, cardiac defects, polysyndactyly, brachydactyly, hypogenitalism, umbilical hernia, learning disability.	(Jenkins et al. 2007)
TWIST1	Saethre-Chotzen syndrome	AD	Haploinsufficiency.	Coronal	Low frontal hairline, hypertelorism, eyelid ptosis, downslanting palpebral fissures, blocked tear ducts, small ears with prominent crus helicis.	(el Ghouzzi et al. 1997, Howard et al. 1997)

AD, autosomal dominant; AR, autosomal recessive; XLD, X-linked dominant; (n) usually arises by new mutation.

1.3.1 Saethre-Chotzen syndrome

Saethre-Chotzen syndrome (SCS) is a form of syndromic craniosynostosis characterised by unilateral or bilateral coronal synostosis, ptosis, ocular hypertelorism, blocked tear ducts, a low frontal hairline, maxillary hypoplasia, a characteristic appearance of the ear (small pinna with a prominent crus; Fig. 1-2) and minor digit anomalies (el Ghouzzi et al. 1997). It was first defined by Saethre (1931), and then Chotzen (1932), who both described individuals exhibiting mild turricephaly, asymmetric skulls and partial soft tissue syndactyly of fingers and toes. It was subsequently shown to be caused by heterozygous mutations in *TWIST1* by two groups (el Ghouzzi et al. 1997, Howard et al. 1997). SCS is a relatively common syndromic form of craniosynostosis that has been estimated to have a prevalence of 1 in 25 000 – 1 in 50 000, however this figure was calculated before the molecular characterization of certain disorders such as Muenke syndrome, which show overlapping clinical features. A more up to date estimation, based on data produced by Wilkie et al (2010), would be around 1 in 60 000. Point mutations, large deletions and chromosomal rearrangements involving *TWIST1* have been described in patients indicating haploinsufficiency of the TWIST1 protein as the underlying mechanism (Gripp, Zackai, and Stolle 2000, Johnson et al. 1998, Wilkie, Yang, et al. 1995). Furthermore, trisomy at the *TWIST1* locus has been associated with cranium bifidum, characterised by a persistent calvarial foramen and open sutures, indicating a *TWIST1* dosage effect during calvarial development (Stankiewicz et al. 2001).

Not all patients manifest every feature associated with SCS and it can be challenging to make a clinical diagnosis due to this phenotypic heterogeneity, and overlap with other craniosynostosis syndromes (e.g. Muenke syndrome). However, a genetic diagnosis is particularly valuable in SCS as the disorder is associated with a high rate of recurrent intracranial hypertension, requiring further surgical intervention and conferring a greater risk than non-syndromic synostosis of the same sutures (Woods et al. 2009). Furthermore, a molecular diagnosis enables the affected family to have accurate genetic counselling, risk estimation and the possibility of prenatal and preimplantation genetic diagnosis.

Figure 1-2

Facial characteristics of Saethre-Chotzen syndrome.

Eight month old female patient demonstrating brachycephaly, right eyelid ptosis, and mild facial asymmetry (slight elevation of left eyebrow and deviation of nose and chin point to the right). Note indented nasal bridge and dysmorphic external ear with prominent transverse crus visible on lateral view (Woods et al. 2009).

1.3.2 *TWIST1*

Belonging to the basic helix-loop-helix (bHLH; see next section) family of transcription factors, *TWIST1* provides important transcriptional regulation in a variety of developmental processes, such as cellular differentiation and lineage commitment.

TWIST1 is highly conserved and crucial for normal development. It was originally identified in *Drosophila* as being required for gastrulation and mesoderm formation, the name “Twist” coming from the twisted appearance of mutant embryos, which die shortly after hatching (Thisse, el Messal, and Perrin-Schmitt 1987). During murine development *Twist1* expression can be seen in the mesenchyme of the head, branchial arches and limb buds. In the developing calvaria *Twist1* is highly expressed in the sutural mesenchyme, probably including early osteoprogenitors, but not in the functioning osteoblasts (Johnson, Iseki, et al. 2000, Rice et al. 2000). *Twist1*-null mice die at E11.5,

exhibiting a failure of cephalic neural tube closure as well as defects in head mesenchyme, branchial arches, somites and limb buds, indicating a regulatory role of *Twist1* in morphogenesis of the cranial neural tube and mesenchymal cells (Chen and Behringer 1995). Mice heterozygous for the *Twist1*-null allele (*Twist1*$^{+/-}$) demonstrate variable fusion of the coronal suture, both bilaterally and unilaterally and provide a model for SCS (Carver, Oram, and Gridley 2002, el Ghouzzi et al. 1997).

As well as interacting through its bHLH domain, murine Twist1 has been shown to interact with the Runt domain of Runx2 (runt-related transcription factor 2), via a C-terminal Twist box (Bialek et al. 2004). Expression of *Runx2* is essential for osteoblast differentiation, hence it is often referred to as the master regulator of osteogenesis. Twist1 was demonstrated to act as an inhibitor of Runx2 by binding to its Runt DNA-binding domain, thereby preventing it from binding to DNA (Bialek et al. 2004, Ishii et al. 2003).

1.3.3 Basic helix-loop-helix transcription factors

The bHLH motif is defined by a basic domain that binds DNA and an HLH domain comprising two amphipathic alpha helices separated by a loop region, enabling dimerization (Hamamori et al. 1997). The majority of bHLH proteins can be divided into three classes; class I proteins (also known as E-proteins) are widely expressed and can form homodimers as well as heterodimers with class II and III proteins, class II proteins are tissue-specific and most prefer to form heterodimers with E-proteins, class III proteins (also known as Id proteins; inhibitors of DNA binding) lack a DNA binding domain and therefore heterodimers that they form cannot bind DNA (Massari and Murre 2000, Zhuang, Cheng, and Weintraub 1996).

As a class II transcription factor, TWIST1 can heterodimerize with class I E-proteins or Id proteins. The E-proteins comprise E12 and E47 (alternatively spliced products of *E2A*, also known as *TCF3*), TCF4 and TCF12, and are thought to enhance the activity of their tissue-specific dimer partner. Functional dimers typically bind DNA at the E-box consensus sequence CANNTG, with different proteins demonstrating distinct sequence preferences for the central nucleotides, for example E47 binds tightly to CACCTG whereas MyoD/E47 heterodimers prefer CAGCTG (Ellenberger et al. 1994, Murre

et al. 1989). The crystal structure of E47 binding to DNA revealed that E47 interacts via its two basic region helices to each half of the binding site in the major groove of the DNA, with conserved hydrophobic residues at the interior of the domain providing van der Waals interactions. Specifically, the glutamic acid at position 555 in the human protein forms hydrogen bonds with the cytosine and adenine at positions one and two of the E-box sequence. This interaction is supported by hydrogen bonds from the nearby p.Arg558 which forms a bridge to the DNA backbone (Fig. 1-3). Both of these key residues (glutamic acid and arginine) are extremely highly conserved in all bHLH proteins (Ellenberger et al. 1994).

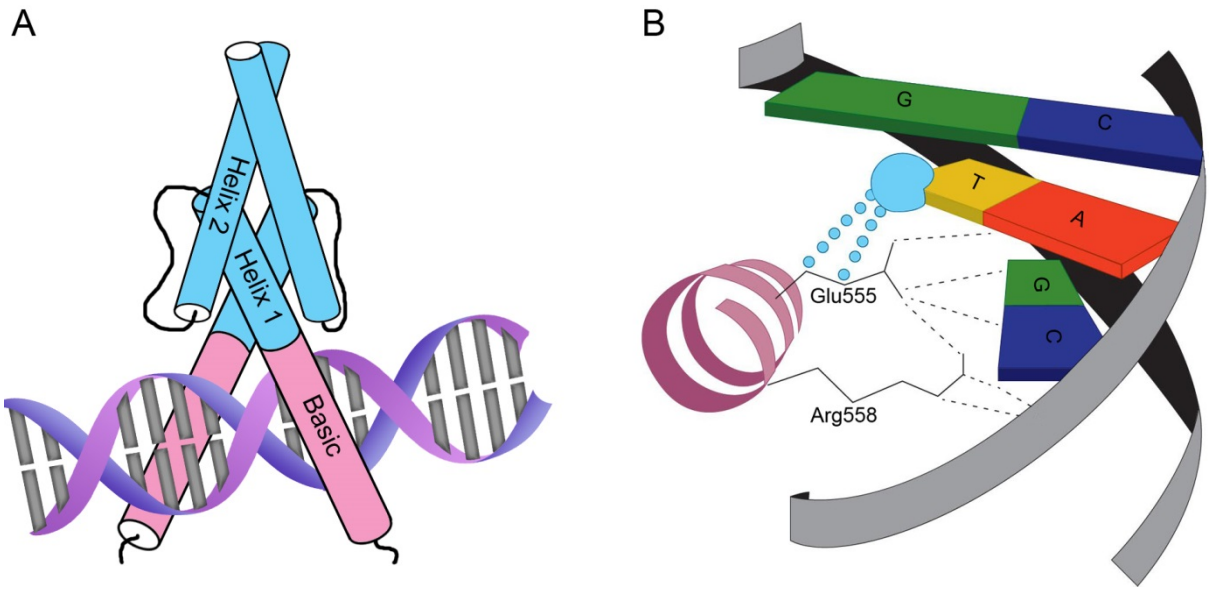


Figure 1-3

Representation of bHLH transcription factors interacting with DNA.

A. A bHLH dimer bound to DNA; each protein comprises two α-helices connected by a loop. Dimers form via interaction of hydrophobic residues on the corresponding faces of the two helices, the basic domain binds to DNA at E-box recognition motifs (-CANNTG-).
B. Detail of how the highly conserved p.Glu555 of E47 binds the E-box sequence; hydrogen bonds are indicated by dashed lines, van der Waals contacts between the β and γ carbons of p.Glu555 and the methyl group of thymine 2' are shown in blue.

1.4 *TWIST1*-negative SCS patients

Around 10% of patients with a clinical diagnosis of SCS have no identified mutations of *TWIST1*, despite extensive genetic testing. While it is possible that some of these may be misdiagnoses due to phenotypic heterogeneity, it is likely that there is another underlying genetic factor contributing to this disorder. A central hypothesis that is further explored in this thesis is that, in some cases, SCS is caused by the disruption of regulatory elements critical for the correct spatio-temporal expression of *TWIST1*. This is based on two main observations; firstly, there are several disorders that can be caused by non-coding mutations, and secondly, there are several reports of patients with a clinical diagnosis of SCS who have rearrangements involving the 7p21-22 region that do not disrupt the *TWIST1* locus.

1.4.1 Non-coding mutations in disease

As well as mutations of the coding sequence of a gene, human disease can be a result of disruption of its regulatory control. Chromosomal translocations can effect gene expression in several ways; for example the breakpoint may separate a gene from its promoter and associated regulatory elements, and/or place a gene under the control of illegitimate elements. Or, the rearrangement may result in a euchromatic gene being translocated to a region of heterochromatin, itself then becoming heterochromatinised and thus silenced (Kleinjan and van Heyningen 1998). This particular facet can be exploited to help identify regulatory elements as they reside in open euchromatin, to allow binding of transcription factors and other proteins, which is sensitive to DNaseI digestion. These so-called enhancer elements tend to be cell-type specific, and can reside upstream and/or downstream of their target gene.

Mutations involving non-coding elements have been implicated in several haploinsufficiency based disorders where no coding mutations could be identified in clinically affected patients, for example in aniridia (the absence of the iris). Aniridia is usually caused by heterozygous loss-of-function mutations in *PAX6*, however a number of cases have been described with no identifiable coding mutations and chromosomal rearrangements downstream of the gene. Closer analysis of this

downstream region, using DNaseI hypersensitivity and evolutionary sequence comparison, led to the identification of multiple regulatory elements spanning ~25 kb (Kleinjan et al. 2001). Pathogenicity was demonstrated using human/mouse somatic cell hybrids where *PAX6* was expressed in the presence of the wild-type allele, but not when they contained patient chromosomes with the downstream regulatory elements deleted (Lauderdale et al. 2000).

Point mutations in a highly conserved regulatory sequence 1 Mb upstream of *SHH* have been associated with pre-axial polydactyly (PPD) in several families (Lettice et al. 2003). The correct patterning of digits in the tetrapod limb is controlled by the zone of polarizing activity (ZPA), which expresses Sonic Hedgehog (*SHH*). The ZPA regulatory sequence (*ZRS*) appears to be responsible solely for the correct expression of *SHH* within the ZPA, as haploinsufficiency of *SHH* causes holoprosencephaly (perturbation of development of the midline structures), which at its most severe is incompatible with life (Roessler et al. 1996). The *ZRS* was identified by mapping a *de novo* translocation breakpoint in a patient with PPD, as well as analysis of the *Sasquatch* mutant mouse (a model for PPD), which highlighted a conserved 1.3 kb region within intron 5 of the neighbouring gene; *LMBR1*. Using a transgenic construct containing the *Zrs*, the authors drove expression of a *lacZ* reporter in the developing mouse embryo, which reproduced the endogenous *Shh* pattern in the limb bud at E10.5, and continued until E13.5 (Lettice et al. 2003).

Haploinsufficiency of the short stature homeobox protein (encoded by *SHOX*) is associated with isolated short stature and Leri-Weill syndrome (LWS; short stature, mesomelia, and wrist abnormalities). Sabherwal et al (2007) screened 122 LWS patients and as well as 17 intragenic mutations and 47 whole gene deletions, they identified four patients with an intact *SHOX* coding region, who harboured deletions 3' to the *SHOX* genomic locus (mapped by fluorescence *in situ* hybridization (FISH) and single nucleotide polymorphism (SNP) analysis). Although the deletions varied in size, they all shared a minimal overlap of ~200 kb. This common deleted region was analysed for the presence of evolutionary conserved genomic elements, and eight putative elements with the highest conservation were selected for further functional analysis. Of these eight elements,

three demonstrated enhancer activity in the chicken embryo, driving expression of GFP in the developing limb buds (Sabherwal et al. 2007). Subsequent analysis of the *SHOX* downstream regulatory region in zebrafish has led to the identification of a smaller and more deeply conserved sub-sequence, as well as additional target tissues (Kenyon et al. 2011). Deletions of this enhancer region, as well as the coding region, are now screened for routinely in patients presenting with short stature (Stuppia et al. 2012).

Mutations of highly conserved elements over 1 Mb from *SOX9* have been linked to Pierre Robin sequence (PRS); the combination of micrognathia, glossoptosis and cleft palate (Benko et al. 2009). Linkage was used to identify a region segregating with autosomal dominant, highly penetrant PRS, which was further refined by analysis of translocation breakpoints which clustered within 160 kb on chromosome 17, between *KCNJ2* and *SOX9*. A high-density tiling path array comparative genomic hybridization (CGH) extending 1.94 Mb centromeric and 1.76 Mb telomeric to *SOX9* identified heterozygous deletions in three unrelated PRS patients, all including highly conserved non-coding elements. Subsequent sequencing analysis revealed a point mutation in another PRS family, disrupting a non-coding element that shares 94% sequence similarity with the mouse, which was shown to increase binding affinity to the transcription factor MSX1 compared to the wild-type sequence. The authors also used a transgenic mouse to demonstrate strong activity of this putative enhancer in the craniofacial region at E11.5 (Benko et al. 2009).

In the examples listed above (summarised in Table 1-2), the disruption of enhancer elements either creates a distinct phenotype (such as with the ZRS and PPD, or *SOX9* regulation and PRS), or phenocopies a null mutation (like *PAX6* regulation and aniridia, or *SHOX* regulation and LWS/short stature). It is believed that genes have multiple enhancer (and silencer) elements which act in concert to regulate expression, therefore the balance of which elements are disrupted/deleted may explain this difference. Key to the identification of enhancer elements in these examples has been to narrow the region of interest by mapping chromosomal rearrangements in patients, cross-

referencing these observations with evolutionary conservation and other information such as DNaseI hypersensitivity, and screening further patients for similar mutations.

Table 1-2

Mutations of non-coding regulatory regions associated with human disease.

Phenotype	Gene affected	Aberration(s)	Distance from gene	Refs
Aniridia	*PAX6*	Translocation breakpoints Two *de novo* deletions	Up to 125 kb downstream	(Kleinjan et al. 2001, Lauderdale et al. 2000)
Pre-axial polydactyly	*SHH*	Breakpoint of *de novo* balanced translocation, point mutations	1 Mb upstream	(Lettice et al. 2003)
Leri-Weill syndrome	*SHOX*	Common minimal deletion of ~200 kb	Proximal downstream	(Sabherwal et al. 2007)
Pierre-Robin sequence	*SOX9*	Translocation breakpoints Microdeletions Point mutation	1.06–1.23 Mb upstream 1.58 Mb upstream/1.56 downstream 1.44 Mb upstream	(Benko et al. 2009)

1.4.2 Rearrangements of chromosome 7p21 – p22

There are several reports of patients with a clinical diagnosis of SCS who have rearrangements involving the 7p21 – p22 region that do not disrupt the *TWIST1* locus (summarised in Fig. 1-4 and Table 1-3). Such rearrangements can be used to narrow down the region of interest around *TWIST1* within which putative enhancer elements may lie.

The translocations described by Krebs et al (1997), Rose et al (1997) and Cai et al (2003) are all located 3´ to *TWIST1*. The closest breakpoint is just ~5 kb downstream, and three of the four translocations described by Rose et al (1997) were further refined to being 70-100 kb downstream in two cases, and 250 kb downstream in another (Patel et al. 1998). Unfortunately it is not clear which exact cases these refer to (as the information was presented as a poster but never formally published). However, unpublished data from the WIMM Clinical Genetics Group (Elena Bochukova), based on isolation of the chromosome 7 homologues into rodent somatic cell hybrids, demonstrated the breakpoint of the t(7;18)(p21.2;q23) translocation (family 1, Wilkie et al 1995) to lie within an intron of *HDAC9* (chr7:18773018-18782892), approximately 370 kb downstream of *TWIST1*. The

translocation was presumed to have arisen *de novo* in the proband's father, who had mild SCS features (unusual ears with a very short crus helicis and absence of the normal furrow between the antihelix and helix at the posterior margin of the ear, and a very high arched palate), and was transmitted to two daughters and a granddaughter who displayed a range in severity of SCS features and mild developmental delay. The proband had very marked brachycephaly (a radiograph confirmed coronal synostosis), as well as mild proptosis, a prominent nose, a short philtrum, a tented upper lip, unusual ears, and a very high arched palate with a bifid uvula. Her sister also had a prominent nose and tented upper lip (although these features were less marked than in the proband) a high arched palate with bifid uvula and unusual ears similar to other affected family members. Her skull radiograph showed parietal foramina bilaterally, and the coronal and sagittal sutures, although patent, had rather sclerotic margins. The proband's daughter had a normal skull shape, but had similar facial appearance to the other translocation carriers with bifid uvula, low-set unusual ears and mild developmental delay (Wilkie, Yang, et al. 1995). This translocation remains the most distal to *TWIST1* that is associated with the SCS phenotype.

More recently, the *de novo* balanced translocation t(7;12)(p21.2;p12.3) was reported in an individual presenting at birth with brachycephaly, down-slanting palpebral fissures, ocular hypertelorism, broad nose with low nasal bridge and low-set ears. Bicoronal synostosis was confirmed by MRI. The breakpoint was identified as being ~42 kb 3' of *TWIST1* (19112935), and the authors went on to show reduced expression of *TWIST1* in the proband's lymphocytes by quantitative PCR (qPCR), as well as a reduced expression of the other gene interrupted by the rearrangement; *PTPRO*. One of the transcripts of *PTPRO* encodes a protein involved in the regulation of osteoclast production and activity, and so the contribution of this gene to the phenotype cannot be ruled out (De Marco et al. 2011).

As well as the translocations described above, a patient with a clinical diagnosis of SCS was found to have a ~685 kb deletion by array CGH, ~230 kb 3' of *TWIST1* (chr7:18240186-18925857). Both the

patient and her affected daughter carried the deletion, which removed most of *HDAC9* (Dr Jenny Morton, unpublished data).

In terms of rearrangements 5' to *TWIST1*, the only information available is a comment by Scherer et al (2003) describing mutations up to 100 kb 5' of *TWIST1* associated with SCS as an example of position-effect mutations separating critical regulatory elements from their target genes, leading to their dysregulation during development (Scherer et al. 2003). However, this example is not referenced and the supplementary information only refers to the cases previously described by Krebs et al (1997) and Rose et al (1997), none of which (as reviewed above) had translocation breakpoints 5' of TWIST1. Hence the source of, and evidence for, such 5' rearrangements remains unclear.

Although the rearrangements 3' of *TWIST1* share a common region, they do not help to determine a downstream limit due to the fact that the entire region telomeric to the breakpoint is translocated. Helping to define this limit, there are also individuals who have rearrangements of the same region that do not have an SCS phenotype. Firstly, David et al (2003) described a family with Peters' anomaly (a congenital defect of the anterior chamber of the eye) who had the balanced translocation t(1;7) (q41;p21). The family did not have overt SCS characteristics, although they were noted to have telecanthus, a broad nasal bridge and midface hypoplasia. The chromosome 7 breakpoint was shown to occur at chr7:18747966, within an intron of *HDAC9*, and involve a 6 bp deletion (David et al. 2003). This breakpoint is telomeric to the t(7;18)(p21.2;q23) translocation.

A further study described a 7.8 Mb inversion with its breakpoint proximal to the rearrangement described by David et al (2003), also in an intron of *HDAC9*, in an individual with a range of dysmorphic features including developmental delay, slightly malformed ears, small down-slanting palpebral fissures, hypermetropia, a broad neck, short toes and short fingers with tapering of terminal phalanges and clinodactyly of fifth fingers (Duno et al. 2004).

The individuals described by David et al (2003) and Duno et al (2004) did not resemble an obvious SCS phenotype, although some features could be said to overlap. However, for the purposes of

defining a region for initial investigation, they do provide evidence for the location of putative *TWIST1* regulatory elements being centromeric to these rearrangements.

Taking these mappings at face value, the simplest interpretation would suggest the existence of a *TWIST1*-regulatory region in a narrow segment of DNA in *HDAC9*, bounded by the David et al (2003) translocation breakpoint and the limit of the t(7;18)(p21.2;q23) translocation. This leads to the specifically testable hypothesis that disruptions of this region, chr7:18747966-18782892, would lead to SCS.

Table 1-3

Summary of genomic rearrangements involving 7p21.

Patient	Features	Karyotype	*De novo*?	Genomic Rearrangement	Method	References
1	SCS	t(6;7)(q16.2;p15.3)	Yes	Translocation + 518 bp deletion 5 kb 3' to *TWIST1* (del:19149617-19150134)	FISH/PCR & sequencing	(Tsuji et al. 1994) (Tsuji et al. 1995) (Krebs et al. 1997)
2	SCS	t(7;10)(p21.2;21.2)	No	Breakpoint lies 70-250 kb 3' to *TWIST1*	FISH	(Reardon et al. 1993) (Rose et al. 1997) (Patel et a. 1998)
3	Normal	Father of patient 2	Yes			
4	SCS	t(7;18)(p21.2;q23)	No	Breakpoint lies within *HDAC9* intron ~370 kb 3' to *TWIST1* (18773018-18782892)	FISH, molecular characterization of somatic cell hybrids	(Wilkie, Yang, et al. 1995) (Rose et al. 1997) (E. Bochukova, unpublished)
5	Mild SCS	Father of patient 4	Yes			
6	Mild SCS	Sister of patient 4	No			
7	Mild SCS	Daughter of patient 4	No			
8	SCS, developmental delay	t(2;7)(q21.1;p21.2)	No	Breakpoint lies 70-250 kb 3' to *TWIST1*	FISH	(Wilkie, Yang, et al. 1995) (Rose et al. 1997) (Patel et al. 1998)
9	Mild SCS		Unknown			
10	SCS	t(5;7)(p15.3;p21.2)	Yes	Breakpoint lies 70-250 kb 3' to *TWIST1*	FISH	(Wilkie, Yang, et al. 1995) (Rose et al. 1997) (Patel et al. 1998)
11	SCS	inv(7)(p21.3q34)	Unknown	Breakpoint ~260 kb 3' of *TWIST1* (18895091)	FISH 10 kb probe	(Cai et al. 2003)
12	SCS, developmental delay, coarctation of aorta	t(2;7)(p23;p22)	No	Breakpoint >260 kb 3' of *TWIST1*	G band/FISH	(Reid et al. 1993) (Lewanda et al. 1994) (Cai et al. 2003)
13	Mild SCS	Mother of patient 12	Unknown			
14	SCS	t(7;12)(p21.2;p12.3)	Yes	Breakpoint ~42 KB 3' of *TWIST1* (19112935)	FISH/PCR & sequencing	(De Marco et al. 2011)
15	SCS	Del(chr7:18240186-18925857)	No	Deletion occurs ~230 kb 3' of *TWIST1*	Array CGH	(J. Morton, unpublished)
16	SCS	Mother of patient 15	Unknown			
17	Peters' anomaly	t(1;7) (q41;p21)	Yes	Breakpoint ~400 kb 3' of *TWIST1*, includes 6 bp deletion (18747984-18747989)	FISH/PCR & sequencing	(David et al. 2003)
18	Peters' anomaly	Daughter of patient 17	No			
19	Peters' anomaly	Son of patient 17	No			
20	Peters' anomaly	Son of patient 17	No			
21	Intellectual disability	inv(7)(p15.2p21.1)	Yes	Inversion breakpoint ~600 kb 3' of *TWIST*, includes 8 bp deletion (18556633-18556641)	Array-CGH/PCR & sequencing	(Duno et al. 2004)

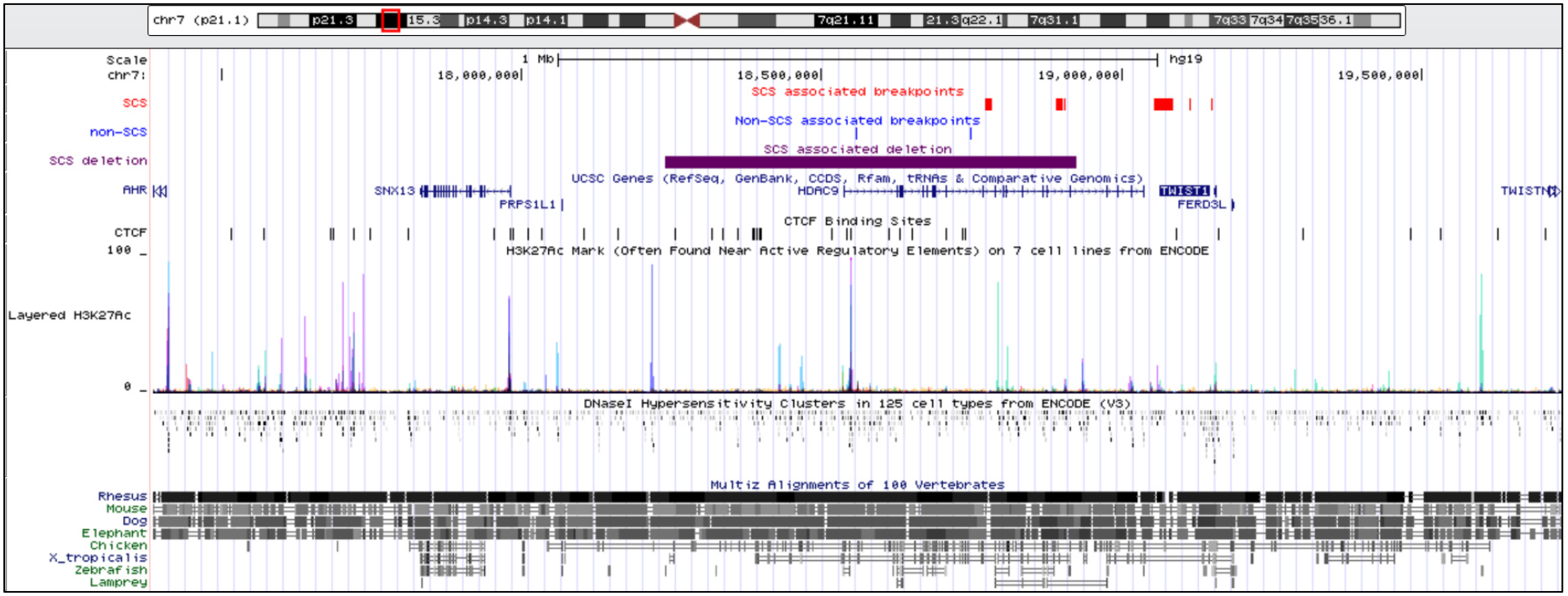


Figure 1-4

Scale map of 2.4 Mb region around *TWIST1*.

The whole region depicted is in synteny with mouse chromosome 12. Translocation breakpoints associated with SCS are indicated in red, those not associated with SCS are in blue; the size of the box represents the region within which the breakpoint is believed to occur. All translocations involved the 7p21.1 – p22.3 genomic region, i.e. downstream of *TWIST1* (which is transcribed in the reverse orientation). In the "SCS associated breakpoints" track, from telomere to centromere the breakpoints indicated are for patients 4, 11 and 12 (overlapping), 8 and 10 (unclear which is 70-100 kb and which is 250 kb from *TWIST1*), 14 and 1 in Table 1-3. The "Non-SCS associated breakpoint" track shows breakpoints for patients 21 and 17.
Tracks depicting the regulatory context of the region are also shown; CTCF binding sites, H3K27Ac, DNaseI hypersensitivity and conservation.
Generated using the UCSC genome browser (genome.ucsc.edu).

1.5 Molecular mechanisms of craniosynostosis

To date, the processes that control normal suture development, and conversely the mechanisms underlying craniosynostosis, are not well understood. Whereas fusion of the sagittal sutures is more likely to have an environmental cause, evidence points to coronal synostosis being associated with cases of monogenic etiology (Wilkie et al. 2010). This is likely to be due to the distinctive origin of the coronal suture, forming a functional boundary between two different embryonic tissue types (CNC and cephalic mesoderm, see section 1.2). Experimental evidence suggests that fusion of the coronal suture may occur as a result of faulty lineage commitment and migration to this boundary, or due to changes in the survival, proliferation or rate of differentiation of osteogenic cells within it (Deckelbaum et al. 2012, Merrill et al. 2006).

1.5.1 The coronal suture as a functional boundary

Boundaries between different cell populations often serve as tissue organisers, for example the vertebrate embryo hindbrain is subdivided into rhombomeres containing specific expression patterns of transcription factors controlling cell fate (Dahmann and Basler 1999). Similarly, the midbrain-hindbrain junction is regulated by Eph-ephrin signaling; interactions between the adjacent domains induce a boundary cell population which produces its own signaling to further regulate cell specification (Cheng et al. 2004). Eph receptors and their ligands, ephrins, represent the largest group of the receptor tyrosine kinase (RTK) family, mediating numerous developmental processes. The system has capacity for bidirectional signaling, both Ephs and ephrins can serve as receptors to transduce a signal into the cell in which they are expressed (Bush and Soriano 2012). In humans, mutations in the X-chromosomal *EFNB1* are associated with craniofrontonasal syndrome (CFNS; frontonasal dysplasia with coronal synostosis). Unusually, females are more severely affected than males, who typically present with only hypertelorism. This is due to cellular interference in heterozygous females, which cannot occur in males. Random X-inactivation is proposed to lead to patchwork loss of ephrin-B1, thereby disturbing tissue boundary formation at the developing coronal

suture, resulting in craniosynostosis. However, as the mouse model of CFNS does not have craniosynostosis, this hypothesis is based on observations of the limb (Twigg et al. 2004, Wieland et al. 2008).

To investigate the role of *Twist1* in the developing murine coronal suture, Merrill et al (2006) used *Wnt1-Cre*; *R26R* mice to observe the distribution of CNC-derived cells. Cell-type specific *lacZ* expression can be achieved by crossing the *R26R* strain of mice, which contain a *loxP*-flanked *lacZ* cassette at the ubiquitously expressed *Rosa26* locus, with mice expressing *Cre* recombinase (Soriano 1999). In *Wnt1-Cre*; *R26R* mice, *Cre* is expressed under the control of the *Wnt1* promoter. CNC cells are derived from *Wnt1*-expressing precursor cells, thus *Wnt1-Cre*; *R26R* mice express β-galactosidase in the population of neural crest initially emigrating away from the neural tube, copying the expression of the endogenous *Wnt1* gene (Chai et al. 2000). Using this technique, Merrill et al (2006) demonstrated that CNC-mesoderm cell mixing occurred at the points of coronal suture fusion in *Twist*$^{+/-}$ mutant mice. In wild-type embryos at E14.5, neural crest and non-neural crest cells were in distinct domains, however *Twist*$^{+/-}$ mutant embryos had irregular CNC-mesoderm boundaries with CNC-derived cells located ectopically in the presumptive parietal bone compartment (Merrill et al. 2006). Subsequently, the availability of a mesoderm lineage-specific reporter mouse (*Mesp1-Cre; R26R*) allowed confirmation of the mesodermal origin of the non-neural crest cells in the skulls of *Wnt1-Cre; R26R* embryos (Yoshida et al. 2008). Furthermore, Merrill et al (2006) demonstrated that ephrin-A2 and ephrin-A4 were expressed transiently in the mesoderm of the prospective parietal bone, with EphA4 being expressed in the adjacent CNC during boundary establishment. This expression was reduced in the *Twist1*$^{+/-}$ mutant mice, with the expression of ephrin-A2 and A4 being retracted anteriorly. Screening of the human orthologues *EFNA2*, *EFNA4* and *EPHA4* in patients with coronal synostosis led to the identification of distinct mutations in *EFNA4* in three unrelated individuals (Merrill et al. 2006).

Ting et al (2009) further investigated the relationship between *Twist1* and Eph-ephrin signaling, and craniosynostosis. *EphA4*$^{-/-}$ mice phenocopy the *Twist1*$^{+/-}$ mutant mice with partially closed coronal sutures at postnatal day (P) 21. Both strains of mutant mice also expressed the early osteoblast marker alkaline phosphatase (Alp) within the coronal suture at E14.5, leading to a disorganised appearance compared to the suture of the wild type which displayed a distinct layer of non-Alp-expressing cells (Ting et al. 2009). Reduced EphA4 expression in the sutures of *Twist1*$^{+/-}$ mutant mice, and increased penetrance and severity of coronal synostosis in *Twist1*$^{+/-}$; *EphA4*$^{+/-}$ compound heterozygous mice, indicate that *EphA4* is a downstream effector of *Twist1* in the murine coronal suture. The *Twist1*$^{+/-}$; *EphA4*$^{+/-}$ mutant mice also demonstrated ectopically located cells from both mesodermal and neural crest origin, visualised using the *Wnt1; Cre* and *Mesp1; Cre* reporter systems, further demonstrating disrupted integrity of the CNC/mesoderm osteogenic/non-osteogenic boundary (Ting et al. 2009).

In addition to its role in the regulation of Eph-ephrin signaling, *Twist1* has also been shown to regulate the Notch pathway in the prospective coronal suture, acting upstream of Notch2 and its ligand Jagged1. Notch signaling is critical for a variety of developmental processes, including developmental boundary formation and cell type specification (Lai 2004, Yen, Ting, and Maxson 2010). In humans, heterozygous mutations in *JAG1* lead to Alagille syndrome, which encompasses various features including impaired craniofacial development, and has been reported rarely to include coronal synostosis (Yilmaz et al. 2013). A pathophysiological link between Alagille syndrome and SCS has been described, based on evidence of the interaction of *Twist1* with *Jag1* in the mouse. In a similar experiment to that of Ting et al (2009), Yen et al (2010) generated *Mesp1-Cre; Jag1*$^{cko/+}$*; Twist1*$^{+/-}$ compound heterozygous mice which demonstrated increased penetrance and extent of suture fusion, as well as an increased number of ectopic CNC cells in the sutural mesenchyme, compared to *Twist1*$^{+/-}$ mutant mice (Yen, Ting, and Maxson 2010).

As detailed above, coronal suture fusion occurs as the CNC/mesoderm boundary loses its integrity, thus genes and pathways involved in cell migration, patterning and maintenance of this boundary represent (and have been demonstrated to be) good candidates for craniosynostosis disease genes.

1.5.2 Proliferation/differentiation at the coronal suture

Another important mechanism in cranial osteogenesis is the control of osteoprogenitor cell proliferation and differentiation at the suture. The role of *TWIST1* in the control of osteoblast differentiation is not fully understood, however studies in the mouse suggest *Twist1* to be a negative regulator of this process (Guenou et al. 2005, Rice et al. 2000).

As mentioned previously, the phenotype of SCS has significant overlap with that of other craniosynostosis syndromes, notably those arising from mutations in *FGFR2* and *FGFR3*. Fibroblast growth factor receptors (FGFRs) play major roles in skeletogenesis, and are clearly important in the development of the skull as indicated by several craniosynostosis syndromes that are caused by mutations in the *FGFR* genes (Table 1-1). Previous work in the mouse has shown that *Fgfr2* is expressed only in proliferating osteogenic progenitors, whereas *Fgfr1* expression is associated with differentiation. Moreover, an *in vivo* study which placed beads soaked with the ligand FGF2 onto the coronal suture of E15 mice demonstrated a decrease in *Fgfr2* expression and a reduction in cell proliferation, coupled with an increase in expression of *Fgfr1* and *Secreted phosphoprotein 1* (*Spp1*, also known as *Osteopontin*, a marker of mature osteoblasts; Iseki, Wilkie, and Morriss-Kay 1999).

To investigate whether *Twist1* and the Fgfr signaling pathway interact, Rice et al (2000) carried out expression studies on the developing mouse skull (E15-P1), in accordance with previous work the authors observed *Fgfr2* expression localised at the osteogenic fronts of the developing parietal bone, and *Twist1* and *Fgf2* to have overlapping expression in the mid-sutural mesenchyme between the bones (Rice et al. 2000). Furthermore, they demonstrated that FGF2 could stimulate the expression of *Twist1 in vitro* in E15 calvarial cultures. Analysis of Fgfr2 distribution in wild type and *Twist1*$^{+/-}$ mutant mice revealed altered expression in the mutants, with the protein localised more discretely in

the mid-sutural mesenchyme compared to its expression in osteoblasts at the osteogenic front in the wild type. Together, these data suggested an integration of *Twist1* and Fgfr signaling in osteogenesis (Rice et al. 2000).

In humans, activating mutations in *FGFR2* give rise to Apert syndrome (over 98% of cases are caused by the substitutions p.(Ser252Trp) or p.(Pro253Arg)), a severe form of craniosynostosis involving fusion of the coronal suture and bony or cutaneous syndactyly of the hands and feet (Wilkie, Slaney, et al. 1995, Bochukova et al. 2009). As with SCS there is a mouse model of Apert syndrome, which carries the p.(Ser252Trp) *Fgfr2* mutation. The $Fgfr2^{S252W/+}$ mice exhibit craniosynostosis and skull malformations that phenocopy the human patients, and like the $Twist1^{+/-}$ mice demonstrate Alp expressing cells within sutural mesenchyme at E14.5 (Chen et al. 2003). Further investigation of the Apert mouse undertaken by Holmes et al (2009) revealed that fusion of the coronal suture occurred very early in embryonic development, with developmental differences noticeable from E13.5. Detailed analysis of proliferation in the suture by BrdU assay revealed a modest increase at E12.5-14.5, followed by a significant decrease in the osteogenic fronts at E16.5 as osteoblasts were formed. Apoptosis was found to be a consequence rather than a cause of craniosynostosis in these mice (Holmes et al. 2009).

Previous studies in mice have demonstrated that *Twist1* both positively and negatively regulates mesenchymal cell specification and differentiation (Connerney et al. 2006, Spicer et al. 1996). The mechanism underlying these dual roles is speculated to be dependent on choice of dimerization partner. As discussed in section 1.3.3, *TWIST1* encodes a class II bHLH transcription factor that can heterodimerize with either class I E-proteins or Id proteins, and has also been shown in *Drosophila* to form functional homodimers (Castanon et al. 2001). *TWIST1* heterodimers, and *Drosophila twist* homodimers, interact with DNA at E-box sequence motifs (CANNTG) via their basic domains. Id proteins repress bHLH activation in a dominant-negative manner by forming non-functional heterodimers incapable of binding DNA, and expression of Ids can disrupt formation of class I/II

heterodimers (Massari and Murre 2000). In the mouse, Twist1 and Id1 are both expressed in the sutural mesenchyme, with Id1 being expressed at the osteogenic front rather than in the mid-suture, correlating with early osteogenesis. Expression of both proteins decreases as osteoblast differentiation proceeds (Rice et al. 2000).

To explore the hypothesis that *TWIST1* dimer partner choice may affect the proliferation/differentiation balance within the suture, Connerney et al (2006) performed transient transfections with various combinations of Twist1, Id1 and the class I E-protein E12. Analysis of cell lysates by Western blot demonstrated that increasing levels of Id1 protein could affect the Twist1/E-protein homodimer/heterodimer balance, with Id1 preferentially interacting with E12 (Connerney et al. 2006). Corroborating previous work by Rice et al (2000), the authors also observed that BMP7 could induce Id expression, as well as FGFR2; however this expression was inhibited by the presence of Twist1/E12 heterodimers. Additionally, they demonstrated that both Twist1 homodimers and heterodimers could bind DNA, and lead to differential gene expression *in vitro* (Connerney et al. 2006). The authors hypothesise that rather than acting as a negative regulator, Id expression promotes TWIST1 homodimer formation in the osteogenic fronts by preferentially dimerizing with E-proteins, therefore reducing the amount of E proteins available to form heterodimers with TWIST1. They go on to claim that in SCS, the reduced amount of TWIST1 would lead to more Id protein in relative terms, which would increase Id/E protein heterodimers thereby increasing TWIST1 homodimers (Connerney et al. 2006). Discernably, dimer choice affected transcriptional regulation and may play an important role in controlling osteogenic fate in the coronal suture, however it is not clear that haploinsufficiency of *TWIST1* leads to an increase of TWIST1 homodimers. Logically, one would expect that a lower concentration of TWIST1 would make it less likely for TWIST1 homodimers to form. Pathogenicity may, more simply, be due to an overall reduction in functional TWIST1/E-protein heterodimers, leading to reduced efficiency of the sutural mesenchyme to maintain an undifferentiated state. Skeletal muscle development is also regulated by bHLH transcription factors; in the mouse, widely expressed *E12* acts as a facilitator to muscle-specific factors (such as *Myod*),

amplifying their activity. In this system heterodimers have been demonstrated to have increased DNA binding compared to homodimers, and Id proteins act as negative regulators (Murray et al. 1992, Rice et al. 2000). An alternative theory to that of Connerney et al (2006) is that the function of *E12* is analogous to its role in skeletal muscle, i.e. to enable the activity of the tissue-specific *TWIST1*, which is negatively regulated by Ids such as *ID1*.

Evidently, the development of the coronal suture is a complex process involving intricate regulation of homodimers and heterodimers, and proliferation/differentiation in the sutural mesenchyme and osteogenic fronts. The interaction between FGFR signaling and Twist1 is key to this process; FGFR signaling can directly upregulate Twist1 (Rice et al. 2000), *Fgfr2* expression is correlated with proliferation, and Twist1 activates or represses Fgfr2 expression dependent on dimer partner (Connerney et al. 2008, Connerney et al. 2006, Iseki, Wilkie, and Morriss-Kay 1999). This complex feedback system renders the coronal suture extremely sensitive to dosage of key genes such as *TWIST1*, *FGFR*s and *EFNB1*, mutations of which lead to craniosynostosis.

1.6 Next generation sequencing

Next generation sequencing has enabled a surge in the identification of genes associated with rare diseases, no longer limiting success to large pedigrees with multiple affected family members. Since 2013, whole exome and genome sequencing has identified nearly three times as many genes as conventional approaches (Chong et al. 2015). In the seminal publication identifying a disease gene by exome sequencing, Ng et al (2010) demonstrated that sequencing a small number of unrelated, affected individuals can provide a powerful and efficient method for identifying the genetic cause of a rare monogenic disorder. In this case, Miller syndrome (a craniofacial disorder involving severe micrognathia, cleft lip and/or palate, hypoplasia or aplasia of the posterior elements of the limbs, coloboma of the eyelids and supernumerary nipples) was shown to be caused by biallelic mutations in *DHODH* (Ng, Buckingham, et al. 2010). Subsequently, whole genome and exome sequencing strategies have become mainstays in both clinical and research settings. Using these techniques, the

Clinical Genetics group in Oxford has identified several new disease genes for craniosynostosis syndromes, for example *MEGF8* in a Carpenter syndrome subtype (Twigg et al. 2012), and *ERF* in multi-suture synostosis (Twigg, Vorgia, et al. 2013). Following on from these successes, I aim to use whole genome sequencing to identify the molecular basis of SCS in a cohort of *TWIST1*-negative patients. Whole genome sequencing, rather than exome sequencing, will allow the identification of all non-coding variants to enable detection of putative regulatory elements, as well as sequencing the remainder of the exome to identify possible causative changes in other genes, if SCS is in fact genetically heterogeneous.

1.7 Gene identification strategies

Although next generation sequencing has been invaluable in identifying numerous disease genes, it still requires a well-designed experiment to prove fruitful. There are a number of approaches to whittle down the >3 million variants per individual (per genome) to the hundreds, or even tens, to follow up as candidates. Initially, all sequence data goes through a variation of the same pipeline; variants are filtered based on quality to reduce false-positives, exclusion of known variants from both in-house and public databases such as dbsnp, and in the first instance terminating (stop mutations and frameshifts), non-synonymous and splice-site variants are prioritised. These steps typically reduce the number of candidate variants to <500, how to interpret these depends on the initial experimental design as detailed below. Regardless of the strategy used, pathogenicity of any suspected disease gene needs to be corroborated independently, for example by identifying mutations in other patients, or by demonstrating a functional effect on the encoded protein.

1.7.1 Phenotype/overlap model

A number of unrelated individuals with the same phenotype are sequenced; genes with mutations present in all individuals are selected as candidates. This approach is highly dependent on patient selection, as genetic heterogeneity will reduce the likelihood of overlapping mutations being present. For example, exome sequencing of 10 individuals with Kabuki syndrome did not reveal a commonly

mutated gene in all individuals (Ng, Bigham, et al. 2010). When the authors reduced the stringency, a candidate gene was identified (*MLL2*) that carried mutations in seven of the patients. Follow-up sequencing of *MLL2* in a panel of 46 Kabuki patients only identified mutations in 23 individuals, indicating genetic heterogeneity and emphasizing the importance of accurate and consistent clinical phenotyping for successful disease gene identification (Gilissen et al. 2012, Ng, Bigham, et al. 2010).

1.7.2 *De novo* model

This typically involves sequencing a trio of unaffected parents and an affected child, where the disease in the child is believed to have occurred due to a *de novo* mutation, and removing all variants present in either parent. As the average exome contains <3 *de novo* mutations, this technique drastically reduces the number of candidate genes (Roach et al. 2010, Vissers et al. 2010). This has been proven to be a highly successful strategy by studies such as that by Vissers et al (2010) who identified *de novo* point mutations that could be linked to disease in 7 out of 10 patients with intellectual disability. More recently, a much larger study identified 12 new genes associated with developmental disorders by exome sequencing of parent-child trios (Fitzgerald et al. 2015).

1.7.3 Recessive model

If consanguinity is suspected, the assumption is made that the disease is caused by a homozygous variant inherited from each of the heterozygous parents. Large regions of homozygosity can be identified by carrying out SNP microarrays, or can be determined from the sequence data itself if family members are also sequenced. The power to discover the causative variant is greatly increased by sequencing individuals who are more distantly related. For example, first cousins will share on average 12.5% of their DNA, which reduces to ~3% in second cousins. This means significantly fewer variants will be identical by descent, and therefore there will be fewer variants to follow up with further testing.

Alternatively, if consanguinity is not suspected, then filtering the sequence data for only the homozygous or compound heterozygous non-synonymous variants in an individual will also

massively reduce the number of candidates. This strategy was used to successfully identify that mutations in *WDR35* cause Sensenbrenner syndrome, a condition involving sagittal craniosynostosis. By focusing on compound heterozygous mutations, the authors were able to reduce the number of candidate genes in two individuals from 139 and 158 to only three and four, respectively (Gilissen et al. 2010, Gilissen et al. 2012).

1.7.4 Candidate/pathway approach

If family members are not available for sequencing, then analysis of a single affected individual's genome (or exome) can still generate a molecular diagnosis, however variant prioritization may be more difficult than in other approaches. Predicted deleterious variants should be assessed starting with any premature truncating mutations such as frameshifts or the introduction of a stop codon, as well as mutations of the canonical splice donor/acceptor sites. Missense variants are more difficult to interpret, but conservation and predicted impact on protein structure are useful indicators of whether a particular variant can be tolerated or not. If the phenotype overlaps with other disorders then candidates can be prioritised based on molecular pathway, for example by looking for genes that interact with known disease genes. Haack et al (2010) successfully demonstrated that variants in *ACAD9* contribute to mitochondrial complex I deficiency by exome sequencing a single severely affected individual. By prioritizing genes encoding mitochondrial proteins, *ACAD9* became a strong candidate gene. To confirm pathogenicity, a patient panel was screened and a further two unrelated individuals with compound heterozygous *ACAD9* mutations were identified (Haack et al. 2010).

1.8 Regulation of gene splicing

During the course of this thesis, several variants were identified that affected gene splicing by mechanisms other than disrupting the canonical donor or acceptor sites. Failure to recognise exon/intron boundaries correctly can generate abnormal mRNA containing intronic sequence, or missing exonic sequence, which may render it unstable or give rise to a mutant protein product. In the normal situation, exon/intron boundaries are recognised via the intronic 3' splice acceptor site

(AG), the branchpoint and the 5' splice donor site (GT), by the spliceosome; an RNA-protein complex comprising small nuclear (sn) RNAs (U1, U2, U4, U5 and U6) and several auxiliary polypeptides. Recognition of the 5´ splice site happens by direct base-pairing between the site itself and the 5´ end of U1, whereas recognition of the 3´ splice site is more complex and requires binding of auxiliary proteins to the polypyrimidine tract located upstream and proximal to the 3' site. This enables U2 to bind the branchpoint, also upstream of the 3' acceptor site (usually 20-50 nucleotides), followed by U4, U5 and U6 which then catalyse the removal of the intron which is subsequently degraded (Cartegni, Chew, and Krainer 2002, Fu and Ares Jr 2014).

Many genes generate several different isoforms by alternative splicing, whereby pairs of donor and acceptor sites are in competition with at least one other set of sites. As well as the interaction of U1 and U2 snRNAs, several *cis*-acting elements are also involved in mediating this complex process, namely exonic splice enhancers (ESEs), exonic splice silencers (ESSs), intronic splice enhancers (ISEs) and intronic splice silencers (ISSs) that bind splicing regulatory proteins to either promote or inhibit tissue-specific splicing. One gene that has been extensively studied in this context is *FGFR2*, as it produces (amongst others) two mutually exclusive isoforms by alternatively splicing exons IIIb and IIIc resulting in FGFR2b and FGFR2c respectively. FGFR2b is expressed in the ectoderm and interacts with specific FGFs in the underlying mesenchyme, while FGFR2c is restricted to the mesenchyme and interacts with a broad range of FGFs expressed in the ectoderm (Zhang et al. 2006). In this system, FGFR2c is the default splice form due to the stronger 3' splice site of exon IIIc and the presence of ISS and ESS elements around exon IIIb. However, tissue-specific regulators in epithelial cells bind ISS and ISE elements to suppress use of the exon IIIc acceptor and activate splicing of exon IIIb (Takeuchi et al. 2010, Warzecha et al. 2009).

More recently, changes other than the well-characterised acceptor/donor site mutations have been implicated in disease. Rather than just the AG/GT being essential for exon/intron boundary recognition, there is a longer and highly conserved consensus sequence at each site; the 3' acceptor

splice site is a polypyrimidine tract <10 followed by cag/GK, the 5' donor splice site is AYAG/gtaagt (intronic sequence indicated by lower case and / representing exon/intron boundary). Mutations at any of these key residues can result in incorrect splicing. Furthermore, point mutations deep in the intron can create cryptic acceptors or donors which are recognised in preference to the wild-type sites. For example, Flanagan et al (2013) used targeted next generation sequencing to fully explore the introns of genes known to be implicated in hyperinsulinaemic hypoglycaemia (*ABCC8* and *HADH*) and identified two deep intronic mutations (c.1333-1013A>G in *ABCC8* and c.636+471G>T in *HADH*) that were predicted to create cryptic donor sites and an out-of-frame pseudoexon. Analysis of patients' cDNA confirmed the presence of the mutant transcripts (Flanagan et al. 2013). Similarly, Bach et al (2015) targeted the introns of *F8* in individuals with haemophilia A, who did not have mutations in the coding region or canonical splice donor/acceptor sites. The authors identified 15 patients with deep intronic mutations predicted to affect splicing, although they were not able to follow these up with any functional assays (Bach et al. 2015).

These types of intronic mutations can be hard to identify, especially if exome sequencing is carried out as most of the intronic DNA will not be sequenced. But even if intronic and intergenic variants are captured, their pathogenicity is more difficult to interpret than coding variants. This poses a challenge for the bioinformatic analysis of data, as most non-coding variants would not be automatically called as likely-pathogenic. Also, a synonymous variant would not necessarily be prioritised, but the nucleotide change may alter the exonic portion of a splice site consensus sequence, or the binding site for an ESE. Fortunately as more is understood about these mechanisms, models can be developed to help improve automated analysis of sequence data. However, these are not fool-proof and care must be taken to evaluate individual variants, for example if they occur in or near a gene of interest.

1.9 Research aims

In general terms, my goal is to expand our understanding of the molecular mechanisms of craniosynostosis. To do this, I aim to:

- Further characterise the mutation spectrum of known disease genes, such as *FGFR2*.
- Identify the molecular cause of craniosynostosis in *TWIST1*-negative SCS patients by next generation sequencing.
- Functionally characterise the effects of any mutations identified, for example by RNA analysis and cellular assays.

CHAPTER 2

MATERIALS & METHODS

Chapter: 2 Materials & Methods

2.1 Ethics approval

All families were recruited by Prof. A. Wilkie into a study of the genetics of craniofacial development: Genetic basis of craniofacial malformations (REC reference number: 09/H0706/20 (West London REC 3)).

2.2 Reagents

All chemicals, unless stated otherwise, were obtained from Life Technologies. Oligonucleotides were ordered from IDT, and the sequences listed in the relevant sections below.

2.3 General laboratory methods

NB All genomic coordinates in this thesis refer to GRCh37 (hg19).

2.3.1 DNA extraction

DNA was extracted from either venous blood collected into EDTA, lymphoblastoid cell lines or cultured fibroblasts established from skin biopsies obtained from the scalp incision at the time of surgical intervention. All DNA was extracted using the Nucleon Blood and Cell Culture (BACC) DNA extraction kit (Gen-Probe Inc.) according to the manufacturer's instructions.

2.3.2 RNA extraction

RNA was extracted from either venous blood collected into PAXgene Blood RNA tubes (Qiagen), lymphoblastoid cell lines or cultured fibroblasts established from skin biopsies as described above. Blood RNA was extracted according to the PAXgene blood RNA kit instructions (Qiagen), and RNA from fibroblasts was extracted using Trizol reagent (Invitrogen) and the RNeasy kit (Qiagen).

2.3.3 cDNA synthesis

cDNA was synthesised using the RevertAid First Strand Synthesis kit (Fermentas) with random hexamer primers according to the manufacturer's instructions.

2.3.4 Polymerase chain reaction

PCR amplification of DNA and cDNA was performed in a volume of 20 µl, containing 15 mM TrisHCl (pH 8.0), 50 mM KCl, 2.5 mM $MgCl_2$, 100 µM each dNTP, 0.4 µM primers, and 0.75 units of FastStart Taq DNA Polymerase (Roche). Cycling conditions consisted of a 5 min initial denaturation step at 94°C, followed by 35 cycles of 94°C for 30 s, annealing at 63˚C (unless specified) for 30 s and extension at 72°C for 30 s, with a final extension at 72°C for 7 min.

2.3.5 Product visualisation

PCR and digestion products were visualised by agarose gel electrophoresis, using 2-3% gels 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 prior to pouring to allow visualisation of the DNA by transillumination with UV light.

Samples were loaded in buffer (6x: 0.25% bromophenol blue, 30% w/v sucrose) and run simultaneously with the appropriate ladder (100 bp or 1 kb; NEB).

2.3.6 Dideoxy sequencing

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 denature the enzymes. Sequencing was then carried out using the BigDye Terminator v3.1 cycle sequencer system (Applied Biosystems).

2.3.7 DNA digestion

Digests were carried out in a total volume of 20 µl, containing 5 µl PCR product, 10 U of enzyme (NEB), 1x buffer solution and 1x bovine serum albumin (BSA) if necessary. Incubation was carried out at the recommended temperature for 2 h-overnight.

2.4 *FGFR2* analysis

2.4.1 Primer sequences for FGFR2 cDNA analysis

Annealing temperature: 61˚C

Variants	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
c.1083A>G c.1083A>T	Ex IIIa F/ Ex 11R	TCGGAGGAGACGTAGAGTTTGTCT GC	TGTTACCTGTCTCCGCAGGGGGATA	334

DNA bands were cut out and gel purified using the Q-Spin gel extraction kit (Geneflow). Dideoxy sequencing was carried out directly on the resulting DNA products.

2.5 Whole genome sequencing

Eight samples were selected for whole genome sequencing (Table 4-1). Genomic DNA (with an OD 260/280 ratio of between 1.8 and 2, and high molecular weight when run on a 1% gel) was diluted to a concentration of 50-100 ng/µl (measured on a Qubit fluorometer) and a minimum of 3-5 µg of DNA provided for library preparation and massively parallel sequencing at the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics), using the HiSeq 2000 platform (Illumina). The reads were mapped using Stampy (Lunter and Goodson 2011) to the human reference genome GRCh37, at an average depth of 25x (Table 4-2). Variant calling was by Platypus v0.2.4 (Rimmer et al. 2014), and annotation by ANNOVAR (Wang, Li, and Hakonarson 2010).

2.5.1 Primer sequences for variant confirmation (genomic DNA)

Variant	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
FLRT1 c.137C>T	1F/1R	CTGTCCTGTGCTCCTTGCAGGTATTC AG	CTTGGTCTTGAGGTCCTGGGGGATG C	384
CDC45 c.226A>C	4F/4R	TAGTGATGAAGGAAAAGGGGCCTC CTCG	CTGTTCCCAGTCCCACAGGGTAGTC AG	314
CDC45 c.318C>T				
CDC45 c.469C>T	5F/5R	ACCACGTATGGTGTAACTCTGGTGC CTCAC	CTTGGCCTGGCAGGCTTCAGGATGA C	436
CDC45 c.749+A>G	9F/9R	CATGAGCCTTAGACTTCTCTGCTTCC TTAC	ATCACACACATACCCAGAAAGGGG CTGCA	240
CDC45 c.773A>G				
CDC45 c.1583C>T	17F/17R	CATAGACAAAGAACCGGCGCTGCA AACTG	GACACTAGAGGCAAATACCACTCTA CTCAG	223

2.5.2 Primer sequences for *CDC45* cDNA analysis

Annealing temperature: 55˚C

Target	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
CDC45 c.226A>C *CDC45* c.318C>T	3F-2/ 5R-2	GGTGGCAAGAACTTGAAACTG	ATGTCTTCATAGGCGGGAAC	238/100 if ex 4 excluded
CDC45 Exons 1-3	Ex1F/ Ex3R	TCCGATTTCCGCAAAGAGTTCTACG	CAGCGTATATTGCACGTGGTCACA	138
CDC45 Exon 4	3F-2/ Ex4R	GGTGGCAAGAACTTGAAACTG	ACATTGACGACATTGACTGGCCTAT	166
CDC45 Exons 5-6	5F-2/6R	GTTCCCGCCTATGAAGACAT	CTCCTCCGCATGGTTTGCTC	137

2.5.3 Primer sequences for *RUNX2* genomic DNA analysis

Variant	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
RUNX2 c.1507T>G	7F/7R	CATGCACCACCACCTCGAATGGCAGCA	CCAGATACCACTGGGCCACTGCTGAG	183

2.6 *TCF12* analysis

2.6.1 Primer sequences for *TCF12* genomic DNA analysis

Mutation	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
c.1035+3 G>C	Ex12 F/Ex12 R	GAATAATGGCAGAGTGCTAAAATATAATAC	GATACTTATCACCTACAAAATCTAATCACTG	220
c.1283T>G	Ex16 F/Ex16 R	CACATATCACATAGTAGGAAACGTAGCAG	GACTAATACTTAGCCAATACTTTACTGC	460

2.6.2 Primer sequences for *TCF12* cDNA analysis

Annealing temperature: 61˚C

Mutation	Primers	Primer sequence 5′→3′ Forward	Reverse	Product size (bp)
c.526+1 G>A	Ex5F Ex9/10R	AGCCCTCATTACAGTGATCACTTGAATGAC	GGTCAGAAGAATTGTGGGTCCCATCTTGCAT	475

c.822C>G	Ex9F Ex11R	TATATGCACCATCCCCAAATTCAGA TGATTTCAAC	TGTGAGGCAGCAACGTAAGGTGAA CTGCTG	355
c.1035+3 G>C	Ex9/10F Ex14R	ATGCAAGATGGGACCCACAATTCTT CTGACC	GATGAAGGTGCTTGCCCTCCAGGTC T	479
c.1328 dupT	Ex13/14F Ex18R	AGTTGGATCACCTTCACCTCTCACA GGTAC	CAACAGTTCCAGACTGACTTTGCAA GCCA	449
c.1349delC	Ex13/14F Ex18R	AGTTGGATCACCTTCACCTCTCACA GGTAC	CAACAGTTCCAGACTGACTTTGCAA GCCA	449
c.1468-20 T>A	Ex16F/ Ex18/19R	ACCAGTTTGCCTGCTGGTCACAGTG	CTTCATTAGTACTGCTTGTTCTGCCT CTAG	421
c.1582+2 T>G	Ex16F/ Ex18/19R	ACCAGTTTGCCTGCTGGTCACAGTG	CTTCATTAGTACTGCTTGTTCTGCCT CTAG	421

2.6.3 Restriction digest analysis of *TCF12* mutations

Mutation	PCR Primers	Enzyme	Product size (bp)	Digest fragments (bp)	
				Wild type	Mutant
c.822C>G	Ex9F, Ex11R	*SmlI*	355	242, 113	355
c.1328dupT	Ex13/14F, Ex18R	*PflFI*	449	329, 120	449
c.1349delC	Ex13/14F, Ex18R	*Cac8I*	449	142, 132, 97, 57	228, 142, 57, 21

2.6.4 Primer sequences for *TCF12* cDNA screen for abnormal splice variants

Annealing temperature: 61˚C

Amplicon	Primers	Primer sequence 5'→3'		Product size (bp)
		Forward	Reverse	
1	5´UTR F / Ex6-7R	AAGCCCGTTCTCCCGGCCAAAGTGA AC	GACTAGATTGACAGCCTGGTAATCC AGTATC	484
2	Ex5F / Ex11R	AGCCCTCATTACAGTGATCACTTGA ATGAC	TGTGAGGCAGCAACGTAAGGTGAA CTGCTG	701
3	Ex9-10F / Ex 14R	ATGCAAGATGGGACCCACAATTCTT CTGACC	GATGAAGGTGCTTGCCCTCCAGGTC T	479
4	Ex13-14F/ Ex18-19R	AGTTGGATCACCTTCACCTCTCACA GGTAC	CTTCATTAGTACTGCTTGTTCTGCCT CTAG	671
5	Ex16 F/ 3´UTR R	ACCAGTTTGCCTGCTGGTCACAGTG	TTCCAAGTCCTTATGAGGAGAGGTC ACCT	833

2.6.5 Multiplex-ligation-dependent probe amplification

Multiplex-ligation-dependent probe amplification (MLPA) was performed using synthetic oligonucleotide probes designed to *TCF12* according to protocols available from MRC-Holland (http://www.mrc-holland.com/pages/indexpag.html). Fragments were analysed by capillary

electrophoresis using an ABI 3130 containing POP-7 polymer. Peaks were visualised using Gene Mapper v3.7 (Applied Biosystems). Common PCR primer annealing sequences are shown in lower case, hybridizing sequences are shown in upper case, and the 3' probe sequence was 5' phosphorylated.

Exon	5′ Probe	3′ Probe	Product size (bp)
2	gggttccctaagggttggaGGACCTGCTAGAAGTG GCCGAAGATG	AATCCCCAGCAACAACGCATGGCtctagattgga tcttgctggcac	91
3	gggttccctaagggttggaGGGTTATTTTGCAGATG TTTTCCCCACCTGTTA	ATAGTGGGAAAACTAGACCAACTACACTGGG AAGctctagattggatcttgctggcac	110
4	gggttccctaagggttggaCTTGGGGAACAAGTGG TCAACCAAGTCCTTCCTA	TGATTCATCTAGAGTAAGTTTGCTGATCAACC CTTGATTAAAGctctagattggatcttgctggcac	120
5	gggttccctaagggttggaCAGACAGCCCTCATTAC AGTGATCACTTGAA	TGACAGTCGATTAGGAGCCCATGAAGGtctag attggatcttgctggcac	100
6	gggttccctaagggttggaGCAATATAATTACCATG AATAGTCTAGCAGTTTGATTGTCTGTCAATAA ATGAAGGGTT	TTATATAAAGTTAATTTCTTTGTTTTATAGGAA AAACATCAGAGAGAGGCTCtctagattggatcttg ctggcac	153
7	gggttccctaagggttggaCAGCTATCTTCTTCAGG AAAACCTGGGACAGCATACTA	TTCATTCTCTGCTACAAGTTCCAGGAGGAGAC Ctctagattggatcttgctggcac	113
10	gggttccctaagggttggaCATCAAATGGGATGAG CCAGCCTGGTT	TTGGTGGAATTCTGGGGACCTCCACtctagattg gatcttgctggcac	94
13	gggttccctaagggttggaCACAGGTAGGCTTCTGT TTTATCTACTTCTAACTGGTGGGACTACTTGG AAAT	ATTTCTAGTTCATTTATTTCTCAACAAACTTTA ATAAAAATTTGTGAAAAATCATtctagattggatc ttgctggcac	150
16	gggttccctaagggttggaCCTTGTTGCAAGCAGTC GATCAGCTTCAATGGTAAA	ATCATGCTCATCTTTTTTGTAGTAAACCCTAA AGATTCTTGTCCTAAATgtctagattggatcttgctg gcac	128
17	gggttccctaagggttggaCTTCAAGCACAGACCTG AACCATAAAACACAAGAAAATT	ATAGAGGTAACTATATTGTTGGTTTTCAGAAA TAATGCAGACAgtctagattggatcttgctggcac	125
19	gggttccctaagggttggaCTTTATAGGAAACCTAA TAGACCTTGTTGTTACTTTATTTTCTAGCAGTA CTAA	TGAAGATGAGGATTTGAACCCTGAACAGAAG ATAGAAAGGtctagattggatcttgctggcac	136
20	gggttccctaagggttggaGGTCATATGTAAACATC AGCCAGGTAAGTACGGGTTTGAA	AAGAAACAGCAAGGAAATAACCTTAAGATTC ATCTTAAACTGGGTTAGTtctagattggatcttgctg gcac	131

2.6.6 Site directed mutagenesis

Point mutations were introduced to the full length *TCF12* cDNA clone using the QuikChange II XL kit (Agilent) using the following primers:

Mutation	Primer sequence 5′→3′ Forward	Reverse
p.Ser210Arg	CCGTGAATCTCCTAGGTATCCATCTCCT	AGGAGATGGATACCTAGGAGATTCACGG

p.Asn328Asp	GAACCAGAGGGGATGCTGCTGGAAG	CTTCCAGCAGCATCCCCTCTGGTTC
p.Leu624Pro	TGAAGCATTCAAAGAGCCTGGCCGAATGTGTCAG	CTGACACATTCGGCCAGGCTCTTTGAATGCTTCA
p.Gln638Glu	GAAGAGTGAAAAACCCGAAACAAAACTCCTTATTC	GAATAAGGAGTTTTGTTTCGGGTTTTTCACTCTTC
p.Ile643Asn	CCAAACAAAACTCCTTAATCTTCATCAAGCCG	CGGCTTGATGAAGATTAAGGAGTTTTGTTTGG
p.Gly42*	CTAGACCAACTACACTGTGAAGCAGTCAATTCAGT	ACTGAATTGACTGCTTCACAGTGTAGTTGGTCTAG
p.Arg135*	CTGTCAATCTAGTCTCCTGTGACAAGATCTGG	CCAGATCTTGTCACAGGAGACTAGATTGACAG
p.Ser241*	CTGACCTTTGGAGTTCATGAAATGGGATGAGC	GCTCATCCCATTTCATGAACTCCAAAGGTCAG
p.Leu428*	TCGAATGGAGGATCGTTGAGACAGACTGGATG	CATCCAGTCTGTCTCAACGATCCTCCATTCGA
p.Arg602fs*31	GATAGAAAGGGAGAAGGAGAGCGGATGGCTAA	TTAGCCATCCGCTCTCCTTCTCCCTTTCTATC
p.Glu655*	CAGTCATCCTTAGTCTATAACAGCAAGTCAGAGAG	CTCTCTGACTTGCTGTTATAGACTAAGGATGACTG

2.6.7 Transactivation analysis of *TCF12* mutations

A full length *TCF12* cDNA clone (pCMV6-XL5) was obtained from OriGene, and the *TWIST1* clone was a gift from Matasaka Nakamura (Tokyo Medical and Dental University, Japan). The pCaSpeR-Tinman::lacZ (pTinE1/E2/E3) construct was used as a bHLH protein-responsive reporter, containing three Tinman E-boxes (two 5'-CATGTG-3' and one 5'-CATATG-3'), and placZ (containing no E-boxes) as a control (both constructs were gifts from Ernst-Martin Füchtbauer, Aarhus University, Denmark).

Transfections were carried out in the HT1080 human fibrosarcoma cell line, cultured in high-glucose Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum. Cells were split into 12-well plates on the day prior to transfection, distributed at a density that would yield cells at ~50% confluence after overnight incubation. Cells were transfected in triplicate with 50 ng of reporter plasmid, 50 ng each of *TWIST1* and *TCF12* constructs or the corresponding empty expression vectors, using Promega FuGENE 6 transfection reagent (at a ratio of 3:1 transfection reagent:DNA) according to manufacturer's protocols. The cells were harvested after 24 h, and β-galactosidase assays were performed in triplicate on 5 µl of cellular extract using the Galacto-Star system (Applied Biosystems), measured on a Wallac VICTOR2 plate luminometer (Perkin Elmer).

2.7 TWIST1 p.Glu117 analysis

2.7.1 Site directed mutagenesis

As described in section 2.6.6, mutations were introduced to the full length *TWIST1* clone (Matasaka Nakamura, see section 2.6.7) using the following primers:

Mutation	Primer sequence 5'→3' Forward	Reverse
p.Arg116Trp	ATGGCCAACGTGTGGGAGCGCC	GGCGCTCCCACACGTTGGCCAT
p.Gln119Pro	GAGCGCCCGCGCACCCAGTC	GACTGGGTGCGCGGGCGCTC
p.Arg120Cys	GAGCGCCAGTGCACCCAGTC	GACTGGGTGCACTGGCGCTC
p.Thr121Ile	GAGCGCCAGCGCATCCAGTC	GACTGGATGCGCTGGCGCTC
p.Glu117Val	CAACGTGCGGGTGCGCCAGC	GCTGGCGCACCCGCACGTTG
p.Glu117Gly	CAACGTGCGGGGGCGCCAGC	GCTGGCGCCCCCGCACGTTG
p.Glu117Lys	CAACGTGCGGAAGCGCCAGC	GCTGGCGCTTCCGCACGTTG

2.7.2 Transactivation analysis of *TWIST1* mutations

As set out in section 2.6.7, with different combinations of wild type and/or mutant TWIST1 with wild type TCF12 as detailed in section 6.3.

2.8 Targeted sequencing

DNA samples from 160 craniosynostosis patients with coronal suture fusions (34 bicoronal, 111 unicoronal , 11 multisuture including bicoronal and 4 multisuture including unicoronal synostosis) were selected for targeted sequencing using the SeqCap EZ Choice Library system (Roche-Nimblegen). Probes were designed to the following coordinates, using the NimbleDesign software (https://design.nimblegen.com/):

Gene	Chr	From	To	Size
TWIST1	chr7	17385745	19735120	2349375
TCF12	chr15	57029979	57670037	640058
ERF	chr19	42746826	42789044	42218
				3031651

2.8.1 Sample preparation

Input DNA was diluted to a final concentration of 20 ng/µl in a total volume of 53 µl, transferred to a Covaris snap-cap tube (6 mm x 16 mm) and fragmented to an average size range of 180 – 220 bp using a Covaris S2 focused-ultrasonicator (intensity 5, duty cycle 10%, 200 cycles per burst, 120 s treatment time, maintained at 7˚C).

Sample libraries were prepared following the SeqCap EZ Library User's Guide. An additional 56 adapters were purchased (IDT) to allow two lots of 80 samples to be multiplexed. Universal blocking oligos were purchased from IDT (xGen® Universal Blocking Oligo -TS HT- i5 and i7).

Final pooled libraries were assessed for quality and concentration using the TapeStation system (Agilent), and 15 µl of each at ~3 ng/µl were supplied to the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics) for sequencing. Each multiplex pool of 80 samples was run on a single lane of the HiSeq2500 platform (Illumina, run in rapid mode).

The resulting data was aligned using Novoalign (http://www.novocraft.com/products/novoalign/) to the human reference genome GRCh37, and variants called with SAMtools v1.1 (Li et al. 2009), and Pindel v0.2.4 (Ye et al. 2009).

2.8.2 Primer sequences for variant confirmation

Variant	Primers	Primer sequence 5'→3' Forward	Reverse	Product size (bp)
TWIST1 c.374A>G	*TWIST1* F4/I1R*	CGGCAAGAAGTCTGCGGGCTG	GCAGCGGCGCGGTCCTTAC	449
TWIST1 c.481C>T				
TCF12 c.1115-2A>G	*TCF12* 14F/14R1	GTTATTGAGTATCTTTACCCTTTCCTTCACAAC	CTTTATTGTTATGCCAGATTGACCAAAAAC	241
TCF12 c.-59C>G	*TCF12* 2F/2R	GCGTAATCTTCCCCAGTACCTCTCTCTGTTC	CTGCCTAGTCTCCACGTTGCTTCCCCTCAC	293
ERF c.1201_1202delAA	*ERF* 4F2/4R2	GCTGAGCCCGATGTACCCCAGTGGTGG	CTGATGTCAGTCACCTCTACCTCCTC	487
Deletions				
Chr7:19140636-19143471	Sample 73 del F/R	TGTCGTAACTGGTCATTTATTTGATCC	GGTACCAGAACTACTCATTTTCTTCAGCTG	3612
Chr7:19037857-19037922	7:19037857-922 F/R	GTAGAGCCAAAGGCTTAACAAAAGACTC	CAATGGCATTCTGTTCTTACAACTACATGC	445

Chr7:18385479-18385523	7:18385479F/18385523R	GAGCCAAGGCCCTGAAGAATCTGAT	CTACCTTTAAACTCATGTCGTTTCTGCTG	278
Chr7:18570122-18570160	7:18570122-160 F/R	ATGGTGAGTTCACAATGCATGACCAATAGC	GTTCTGCCTGGGCAACAGAGTGAGATC	455
Chr7:18804498-18804519	7:18804498-519 F/R	ACTGAAATTTGGCTAGCCTCAAGACGACTG	AGTTCCTCCCAACACATATGGGAGATGTC	377
Chr7:18830885-18830905	7:18830885-905 F/R	AGTTTTCCCACCACTACAGAGCTACTAGTAC	GTCAAAGATACAGCACTCTCTGCTGCTG	294
Chr7:18272639-18272654	7:18272639-654 F/R	GCTCCCCTTTCTAATCTGCTATCCATG	CTGATGTATATGGGCAGCAATTATGAGCTG	295
Chr7:18663028-18663043	7:18663028-043 F/R	ATCCTGGTTCTTGGCTCACCTGGTTGTC	TGAGCAGAGGCAAAATACACATAGTAGGTGGTC	392
Chr7:17921226-17921237	7:17921226-237 F/R	CTTGGGTAGATTTTTGACAAGTAACTTTGAGTAG	TGAATTGACGTAGGGGGGATTGATTTAGTG	383
Chr15:57502663-57502673	15:57502663-673 F/R	GACTCAGACAAGGTATGAGCTACATACTAG	TCCTTGCCATAACACTGGTGATTTCAGACTG	442
Chr15:57384862-57384871	15:57384862871 F/R	GTGAAGTTGCACTTTCAAGAAGGGTATGCAC	GCTGTGTCTTAAAAGGACGATTCTAAGATCAC	460
Chr7:17545688-17545696	7:17545688-696 F/R	GCAATATGAATTGGCACTGGCACAGAAGTG	ACCTCTGTACCTCCTCAAACATGATTGATGC	255
Chr7:18326895-18326903	7:18326895-903 F/R	AGGCATATCCCTGTTCTGATTTGCAGACAG	GTCTTCAGGCTAACATGATTACTGGTTGCTC	474
Chr7:19018942-19018949	7:19018942-949 F/R	GGCAACCTTTCTCATGAACGTGATGGTTTCATGTC	GATGGGCTATTTGAGGAAAATCATGTTTCTAAGTAC	308
Single nucleotide variants				
Chr7:19157258 T>G	*TWIST1* Ex1F/5´R	GTCAGGCCAATGACACTGCTGCCCCCAAACT	CTGGAAACGGTGCCGGTGCTGCA	127
Chr7:17979676 G>C	7:17979676 F/R	CGAGCAACAACCCCAGGGTGCTGAC	CCAGCCCGTCAGAGAACATCATCTGTAC	400
Chr7:19146484 G>A	7:19146484 F/R	CTGACACCACCAGGGACCCAACG	CTGCCTGCCGGTCTTCCTAAACTGAC	293
Chr7:19152348 C>A	7:19152348 F/R	TGACACTTGGCACCGCCGCCTCCTAG	CTAGACCGGTATCAAGTTCTTCTCAGGCATATG	369
Chr7:19152403 T>A	7:19152403 F/R	CACGCAAACCGAAGGAGATACGGACATCG	GTCAGAACTGGTTATTTTCTCTAGACCGGTATC	333

NB Amplification of cDNA for analysis of *TCF12* c.-59C>G was carried out using primer pair 5´UTR F/Ex6-7R as described in section 2.6.4.

**TWIST1* F4/I1R PCR with 10% dimethyl sulfoxide added, annealing temperature: 68˚C.

CHAPTER 3

SYNONYMOUS SUBSTITUTIONS OF FGFR2 ARE ASSOCIATED WITH MILD CROUZON SYNDROME

Chapter 3: Synonymous substitutions of FGFR2 are associated with mild Crouzon syndrome

3.1 Introduction

As discussed in section 1.3, there is considerable genetic heterogeneity and frequent phenotypic overlap between different craniosynostosis syndromes. In particular, heterozygous mutations in *FGFR2* account for ~28% of genetic cases (Sharma et al. 2013) and cause Crouzon (Jabs et al. 1994, Reardon et al. 1994), Pfeiffer (Lajeunie, Ma, et al. 1995, Rutland et al. 1995, Schell et al. 1995), Apert (Wilkie, Slaney, et al. 1995), Beare-Stevenson (Przylepa et al. 1996) and bent bone dysplasia (Merrill et al. 2012) syndromes. All involve synostosis of the coronal and other cranial sutures, a distinctive "crouzonoid" craniofacial appearance (hypertelorism, exorbitism, prominent nose and midface hypoplasia), but differ in the presence and extent of abnormalities of the hands and feet, other skeletal manifestations and dermatological features (Agochukwu, Solomon, and Muenke 2012, Johnson and Wilkie 2011).

FGFR2, like the other members of the FGFR family, comprises an extracellular ligand-binding domain (composed of three immunoglobulin-like domains), a single transmembrane peptide and a cytoplasmic tyrosine kinase domain. Mutually exclusive alternative splicing of exons IIIb and IIIc gives rise to epithelial and mesenchymal isoforms respectively, which interact with different FGFs to regulate downstream processes (Eswarakumar, Lax, and Schlessinger 2005, Ibrahimi et al. 2005).

Presented here are two families with a mild Crouzon syndrome phenotype and reduced penetrance of craniosynostosis. Diagnostic screening of *FGFR2* due to the phenotype revealed different *FGFR2* mutations of the same nucleotide in each family; c.1083A>G and c.1083A>T. Both variants encode an apparently synonymous change at the p.Pro361 codon, however as they occurred close to the end of exon IIIc it was suspected that they may affect splicing.

3.2 Case presentations

3.2.1 Family 1

The female proband III-1 (Fig. 3-1A) was born after a normal pregnancy and was referred for craniofacial assessment at two years of age because prominent eyes and a head tilt. Her father II-1 (Fig. 3-1B), grandmother I-1 and two deceased uncles were all said to have a similar appearance with prominent staring eyes, but had not required any surgery. She had an occipitofrontal circumference (OFC) of 50 cm (+1.1 SD) and was noted to have a mildly crouzonoid appearance, with slight exorbitism and midface retrusion. In view of the mild crouzonoid features, sequencing of *FGFR2* exons IIIa and IIIc was requested. This demonstrated a heterozygous c.1083A>G variant within exon IIIc (Fig. 3-2A), which was also present in II-1 and I-1.

Computed tomography (CT) of the skull of individual III-1 at the age of 2.8 years demonstrated right lambdoid and occipitomastoid synostosis, all other major cranial sutures being patent. Ophthalmological review identified slightly reduced visual acuity and a latent divergent squint with slight left hypophoria. The patient is now four years old and has not undergone any surgical intervention, as she has a good overall head shape with no major midface retrusion, is making good developmental progress, and has no features to suggest significant intracranial restriction.

3.2.2 Family 2

The male proband (III-2 in Fig. 3-1C), born after an uneventful pregnancy, was referred for craniofacial assessment at the age of three months. Physical examination showed a mild cloverleaf skull with temporal bulging and reduced OFC (36 cm; -2.2 SD), hypertelorism, and severe exorbitism mainly at the infra-orbital level. Skull X-ray and CT showed pansynostosis and multiple craniolacunae, with no intracerebral anomalies.

Owing to the severe peri-orbital features and the absence of deformations of the upper and lower extremities a clinical diagnosis of Crouzon syndrome was suggested. The patient's mother,

grandmother and several cousins were reported to show mild facial features also suggestive of this diagnosis.

The proband underwent fronto-orbital advancement at the age of five months. Since the occiput was still severely flattened and both lambdoid sutures were fused, occipital craniotomy and remodelling was performed at the age of twelve months. Clumsiness and motor delay were first noted aged 18 months; psychological testing at the age of 12.8 years gave scores for non-verbal intelligence of 80 (SON-R) and visual-motor integration of 81. Clonidine was prescribed due to high distractibility and he underwent special education. During childhood, the exorbitism increased requiring further orbital advancement and cranial vault remodelling at the age of eight years. Several deciduous and permanent teeth were extracted because of Class III malocclusion and dental crowding.

Genetic testing of *FGFR2* was performed, identifying the heterozygous point mutation c.1083A>T (Fig. 3-2A). This variant was also present in his mother II-1.

Independently of these events, a second patient (III-1; Fig. 3-1D) was referred to the same clinic at the age of ten years with a scaphocephalic head shape. He had previously undergone vault remodelling at the age of 16 months owing to bicoronal synostosis. At the time of referral, he had an occipito-frontal circumference of 48.5 cm (-2.8 SD), hypertelorism and severe exorbitism. In addition, he had mild maxillary hypoplasia and both 2nd premolars of the lower jaw were absent. Ophthalmic examination showed myopia with divergent strabismus of the right eye associated with reduced visual acuity. A monobloc procedure without distraction (Le Fort III and an advancement of the forehead) was performed.

Due to the severe peri-orbital features a diagnosis of Crouzon syndrome was suggested. Analysis of *FGFR2* identified the heterozygous point mutation c.1083A>T. His father, grandmother and great-grandmother had a similar craniofacial appearance. Based on the pedigree analysis, it is evident that III-1 and III-2 are third cousins and that the *FGFR2* mutation present in these two branches is

identical by descent (Fig 3-1C). Apart from III-1 and III-2, none of the other affected family members had undergone craniofacial surgery.

3.3 Results

The synonymous variants c.1083A>G and c.1083A>T occur at the -2 position of the 5' (donor) splice site of *FGFR2* exon IIIc (Fig. 6-2A). The Neural Network splice site predictor (http://www.fruitfly.org/seq_tools/splice.html) generates a score for the wild type donor of 0.88, which is reduced to 0.37 by the A>G transition, and to 0.19 by the A>T transversion. In these circumstances, use of a cryptic splice site (score 0.84) within exon IIIc, 51 nucleotides upstream from the end of the exon, is expected based on analysis of a previous mutation c.1084+3A>G (Kan et al. 2004). This would lead to an in-frame deletion of 17 amino acids.

Amplification of cDNA from individuals heterozygous for either the *FGFR2* c.1083A>G or the c.1083A>T variants demonstrated the presence of two additional bands, not present in the wild type control, at approximately 430 bp and 330 bp (Fig. 3-2B). Sequencing of the normal 479 bp product from these individuals showed complete absence of the mutant allele (illustrated in Fig. 3-2C for II-1 from Family 1), indicating that both mutations abolish use of the normal exon IIIc donor splice site. Sequencing of the ~430 bp product confirmed that the cryptic splice donor within exon IIIc was preferred in the mutant allele (Fig. 3-2D), while the ~330 bp product demonstrated complete skipping of exon IIIc (Fig. 3-2E).

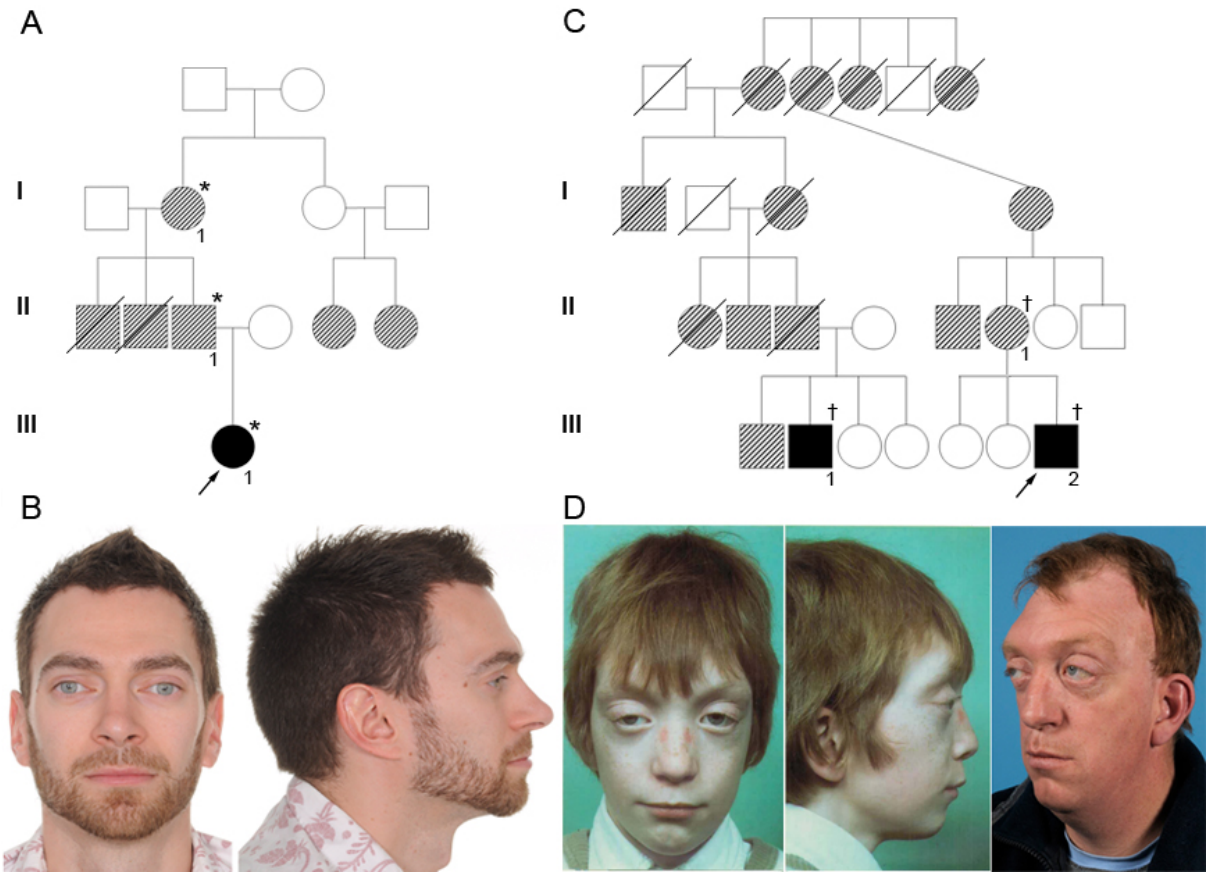


Figure 3-1

Pedigrees and facial features of individuals with *FGFR2* mutations.
A. Pedigree of Family 1; solid symbols represent clinically confirmed craniosynostosis and hatched symbols represent individuals with a similar crouzonoid appearance but without confirmed craniosynostosis. * confirmed heterozygosity for c.1083A>G.
B. Facial appearance of II-1 from Family 1 aged 33 years.
C. Abbreviated pedigree of Family 2; notation of symbols as in part A. † confirmed heterozygosity for c.1083A>T.
D. Facial appearance of III-1 from Family 2 aged 8.75 years (left) and 43 years (right).

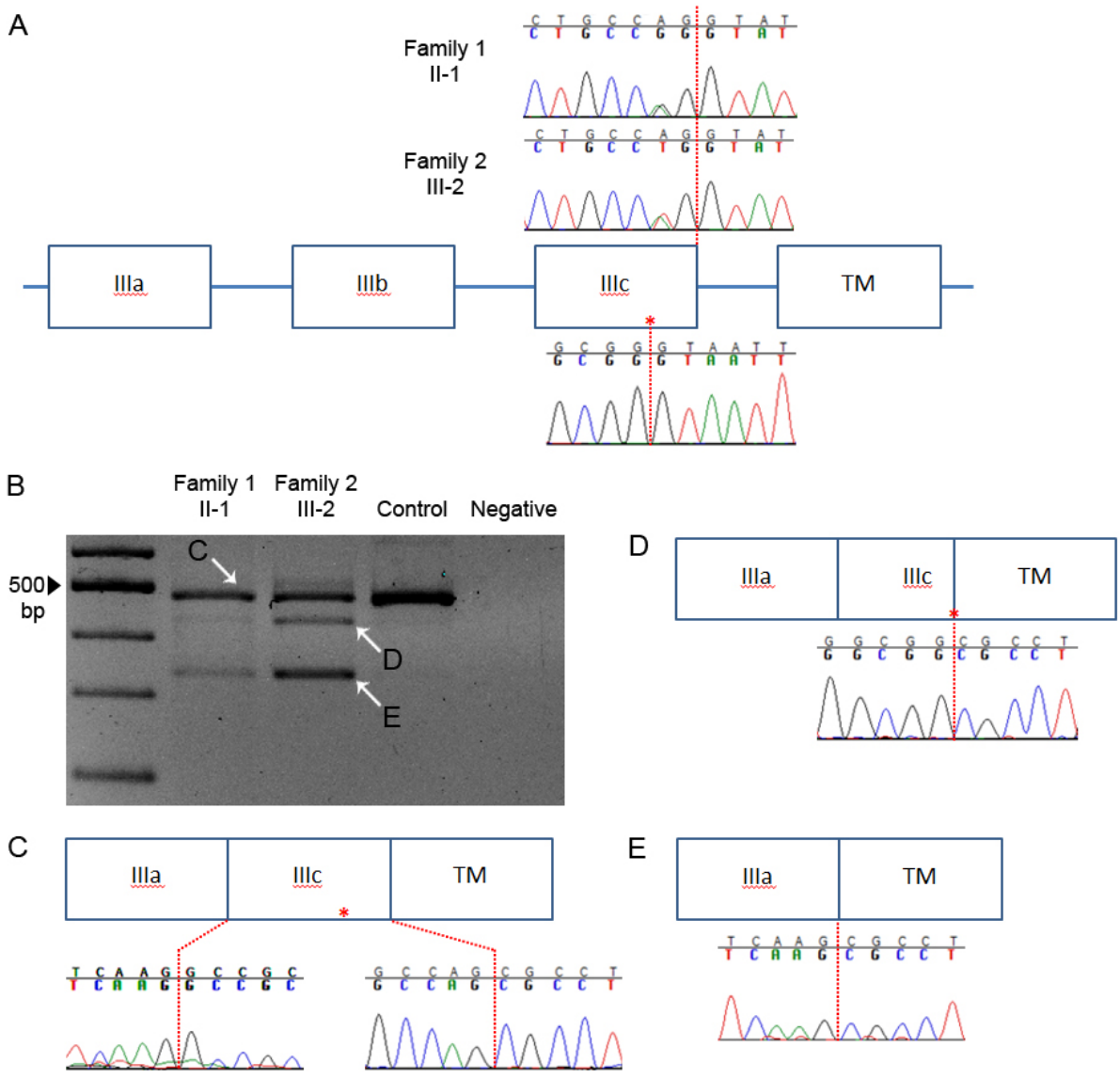


Figure 3-2

Genomic context and consequences of *FGFR2* mutations.

A. Schematic representation of genomic region affected by c.1083A>G and c.1083A>T mutations (not to scale). IIIa, IIIb and IIIc denote exons of *FGFR2* encoding the 3rd immunoglobulin-like domain; note that physiological skipping of exon IIIb normally occurs in blood mRNA.TM, exon encoding transmembrane domain. Sequencing of genomic DNA demonstrates heterozygosity for c.1083A>G (II-1, Family 1) or c.1083A>T (III-2, Family 2) mutation at the -2 position from the end of exon IIIc (indicated by dashed red line; the upper line in each trace shows the wild type sequence). Below the cartoon is shown the genomic sequence around the cryptic donor splice site within exon IIIc (marked with red asterisk).

B. Amplified cDNA corresponding to c.1083A>G and c.1083A>T mutations (same individuals as for genomic analyses), demonstrates two additional smaller products (indicated with white arrows **D** and **E**), in addition to the normal (**C**) product.

C. In the cDNA product coincident with the wild type band (II-1, Family 1) the mutant allele at the penultimate position of exon IIIc is not represented, indicating complete skipping of normal splicing. The consequences for splicing of the mutant allele (sequence traces illustrated are from III-1, Family 2) are either to activate the cryptic splice donor site within exon IIIc (**D**) or to skip exon IIIc completely (**E**).

3.4 Discussion

Mutations of the consensus A at the penultimate nucleotide of the exon have been associated previously with loss of splicing at the donor site (Podkrajsek et al. 2008, Twigg, Vorgia, et al. 2013). Other mutations near the *FGFR2* exon IIIc splice donor site have also been described (Table 3-1), all associated with a mild Crouzon phenotype. It is likely that the p.(Ala362Ser) substitution (Everett et al. 1999) in fact exerts its pathogenic effect via the G>T transversion at the -1 position in the exon, rather than as a missense substitution, but this was not tested experimentally.

Table 3-1

Summary of mutations affecting correct splicing of the *FGFR2* exon IIIc donor site.

Mutation	Protein	Proposed effect	Experimental demonstration	Reference
c.1032G>A	p.(Ala344Ala)	Activation of cryptic splice site	Yes	(Del Gatto and Breathnach 1995, Li et al. 1995, Reardon et al. 1994)
c.1083A>G	p.(Pro361Pro)	Loss of normal donor site with use of alternative cryptic splice site	Yes	This study
c.1083A>T	p.(Pro361Pro)	As above	Yes	This study
c.1084G>T	p.(Ala362Ser)	Annotated as missense but likely to affect splicing	No	(Everett et al. 1999)
c.1084+3A>G	-	Loss of normal donor site with use of alternative cryptic splice site	Yes	(Kan et al. 2004)
c.1084+3A>C	-	Loss of normal donor site with use of alternative cryptic splice site	No	(Cornejo-Roldan, Roessler, and Muenke 1999, Kress et al. 2000)

In the cases reported here, the wild type splice donor site is abolished by the A>G or A>T mutations at the -2 position from the intron, leading to the cryptic site being preferred. The apparently greater amount of mutant cDNA products associated with the A>T mutation appears to correlate both with the greater predicted disruption of the splice site and with the more severe phenotype in clinically

affected individuals from Family 2 compared with Family 1. However, since the A>G mutation did not support use of the normal exon IIIc donor splice site (Fig. 3-2C), other explanations for these differences are possible, such as differing proportions of cell types in the blood samples analysed, and/or differences in genetic background.

Using the cryptic donor would lead to an in-frame deletion of the last 17 amino acids of exon IIIc (p.Gly345_Pro361del), including four residues that form specific contacts with the ligand FGF2 (Plotnikov et al. 2000). The same result was also indicated for the mutation c.1032G>A, which creates a silent substitution at p.Ala344 and activates the same cryptic donor (Li et al. 1995). This deletion of the last 17 residues would shorten the distance from the disulphide bond in the Ig-III domain to the transmembrane domain, likely leading to a change of function. Craniosynostosis-causing FGFR mutations function in a constitutively active dominant manner, so it is likely that in these individuals a mutant protein is formed which is prone to forming covalently-linked dimers (Robertson et al. 1998), leading to variable features of Crouzon syndrome.

There were a range of phenotypic consequences associated with the mutations described here; whilst a mild crouzonoid phenotype was generally evident, only a minority of individuals developed overt craniosynostosis requiring calvarial surgery. The causes of the clinical variability are unknown, although one potential factor may be the extent of intrauterine fetal head constraint (Johnson, Wall, et al. 2000). Families carrying either of the mutations described here should be counselled accordingly, and all mutations near the *FGFR2* exon IIIc splice sites should be carefully evaluated as to whether they may be pathogenic, even if they are synonymous or outside the canonical AG/GT splice acceptor/donor sequences.

CHAPTER 4

WHOLE GENOME SEQUENCING OF *TWIST1*-NEGATIVE SAETHRE-CHOTZEN SYNDROME PATIENTS

Chapter 4: Whole genome sequencing of *TWIST1*-negative Saethre-Chotzen syndrome patients

4.1 Introduction

Whole genome sequencing was used to try and identify the molecular basis of SCS in a cohort of *TWIST1*-negative patients, based on the hypothesis that some cases of SCS are due to mutations of transcriptional elements regulating *TWIST1* expression. Previous work in the lab had been carried out to identify any such elements by designing a custom tiling path array spanning 2.4 Mb around the *TWIST1* locus, recapitulating work by Benko et al (2009) who used this method to identify micro-rearrangements affecting *SOX9* expression in Pierre Robin Sequence. However, in the Oxford study no overlapping copy number variants in two or more SCS patients were identified (of 14 analysed). Whole genome sequencing was selected as a strategy to allow identification of non-coding variants within a region of interest around *TWIST1*, as well as coding variants in the remainder of the exome.

The same region investigated by the tiling path array was also used as a focus for analysis of non-coding variants. To define this region, the positions of the rearrangements detailed in section 1.4.2 were taken into consideration as well as evolutionary conservation. An extended region spanning ~2.4 Mb was selected based on breaks in synteny with the mouse chromosome 12. This would ensure all putative regulatory regions would be included, based on the assumption that any conserved regulatory elements would remain in *cis* with their target gene, *TWIST1*.

4.2 Patient selection

Six individuals with a clinical diagnosis of SCS, but with normal *TWIST1* genotypes were selected for sequencing (including a case-parent trio with unaffected parents; Table 4-1). Affected individuals had undergone *FGFR1-3*, *EFNB1* and *TWIST1* mutation analysis, as well as genome-wide array CGH and the tiling path array around *TWIST1*.

Table 4-1

Clinical features of patients selected for whole genome sequencing.

Patient no.	Sex	Suspected *de novo* mutation?	Clinical features	Additional information
2930	Male	Yes	Right unicoronal synostosis Small mildly dysplastic ears Low frontal hairline Laterally deviated nasal septum 2/3 and 3/4 syndactyly	Unaffected parents (**1473**, **1474**) also sequenced.
4400	Female	No	Bicoronal synostosis Small posteriorly rotated ears High arched palate Down slanting palperable fissures Single palmar crease	Mother (not sequenced) has mild bilateral ptosis, low frontal hairline, small ears with prominent crease and bilateral fifth finger clinodacyly.
4437	Female	Yes	Bicoronal synostosis Small ears Low frontal hairline Myopia	
4454	Female	No	Bicoronal synostosis	Clinically affected aunt (not sequenced) with the same facial and skull appearance.
4580	Female	Yes	Right unicoronal synostosis Ears have prominent crura Low frontal hairline Mild exorbitism Mild lateral nasal deviation	
4809	Male	No	Brachycephalic	Clinically affected son (not sequenced) with bicoronal synostosis, low frontal hairline, single palmar crease and 3/4 syndactyly.

4.3 Whole genome sequence data

Samples were sequenced as part of the WGS500 sequencing project, which in 2011 was the first large scale whole genome sequencing project undertaken in the UK (Taylor et al. 2015). Sample preparation and data processing are described in section 2.5 (coverage is summarised in Table 4-2).

For each variant a predicted deleterious score was generated, which was initially set to zero, then incremented by one (to a maximum of six) dependent on the following predictions:

- SIFT (Kumar, Henikoff, and Ng 2009) Deleterious (sift<=0.05)
- PolyPhen2 (Adzhubei et al. 2010) Probably damaging (>=0.957)
- LRT (Chun and Fay 2009) Deleterious
- MutationTaster (Schwarz et al. 2010) Disease causing
- GERP++ (Davydov et al. 2010) ≥5
- PhyloP (Siepel, Pollard, and Haussler 2006) >1

In addition, the deleterious score was automatically set to six if the variant caused a stop-gain, a frameshift or was predicted to affect splicing. This "majority rule" approach was shown to be more accurate than any single method individually in a large study of 6515 exomes (Fu et al. 2013).

Table 4-2

Depth of exon coverage produced by whole genome sequencing.

Individual	% exons at 10x coverage	% exons at 30x coverage
2930	98.7	39.7
1473 (mother of 2930)	98.9	85.2
1474 (father of 2930)	98.5	32.8
4400	98.5	27.8
4437	98.3	24.6
4454	98.8	39.1
4580	98.6	27.9
4809	98.8	49.2

The sequencing data was analysed using two main approaches; analysis of coding variants, and analysis of non-coding variants in the region of interest around *TWIST1*. These are detailed in sections 4.6 and 4.7 respectively. Details of coding variant analysis strategies are described in section 1.7.

4.4 Identification of *TCF12* as a disease gene

At around the same time as receiving sequence data back for the SCS individuals, a concurrent exome sequencing project on patients with bicoronal synostosis identified *TCF12* as a candidate disease gene. However, in each of the first four unrelated subjects in whom a deleterious *TCF12* variant had been identified, a first degree relative (parent in three cases, offspring in the fourth case) was shown to carry the mutation, yet was clinically unaffected. Hence the causal relationship of *TCF12* mutations with the craniosynostosis phenotype remained uncertain. A direct search for *TCF12* variants in my data revealed a premature termination mutation in individual 4437 (p.Leu428*), and a splicing mutation in individual 4454 (c.1035+3G>C). Patient 4437 was the first identified *de novo* mutation (relationships confirmed via microsatellite analysis), and so provided strong early evidence that *TCF12* was indeed a gene of interest.

Of note, the variant in patient 4454 (c.1035+3G>C) was actually missed as pathogenic in the original calling of variants, lying just outside of the canonical two base-pair splice donor site (see section 1.8 for more details). Although the variant was present in an affected aunt it was also present in the proband's clinically unaffected mother and grandmother, further demonstrating non-penetrance. Subsequent iterations of variant calling were modified to include more intronic changes that could affect splicing.

The role of *TCF12* as a craniosynostosis disease gene is discussed fully in Chapter 5.

4.5 Analysis of coding variants

4.5.1 Trio analysis – *de novo*

Initially, sequence data from the trio was analysed assuming that the phenotype of the proband (patient 2930) was due to a *de novo* mutation and all variants present in either parent were removed. Variants occurring in either dbSNP (www.ncbi.nlm.nih.gov/SNP/) or the 1000 genomes project (www.1000genomes.org) were then filtered out, leaving 780 variants, of which 11 were exonic. These remaining 11 exonic variants were visualised in the Gbrowse genome browser to assess if they looked genuine, and were compared with both parental sequences to verify that they appeared *de novo* (Table 4-3).

Table 4-3

Predicted *de novo* exonic variants present in individual 2930 after filtering out common variants.

Gene	Nucleotide change	Amino acid change	*De novo*?
CARNS1	c.1063G>C	p.Gly355Arg	No - in 1473
FLRT1	c.137C>T	p.Thr46Met	Yes
GOLGA6L1	c.1783T>C	p.Trp595Arg	No - in 1473
HGC6.3	c.214A>G	p.Met72Val	No - in 1473
MUC5B	c.13751C>T	p.Thr4584Ile	No - in 1473
PDE4DIP	c.88C>T	p.Pro30Ser	No - in 1474
OTOP1	c.310_318delCTGCTGTGG	p.Leu104_Trp106del	No - in 1473 & 1474
FER1L5	c.1440G>T	p.(=)	Not real
IGFN1	c.4659A>G	p.(=)	No - in 1474
MUC5B	c.11397C>T	p.(=)	No - in 1473
ZNF814	c.996G>C	p.(=)	No - in 1473 & 1474

4.5.1.1 *FLRT1* as a candidate gene

As can be seen in Table 4-3, the single variant that appeared to be *de novo* after filtering was the nonsynonymous change c.137C>T (p.Thr46Met) in *FLRT1*, which had a deleterious score of 2. The variant was confirmed to be *de novo* by dideoxy sequencing (Fig. 4-1).

FLRT1 encodes a member of the fibronectin leucine rich transmembrane protein (FLRT) family, of which there are three members (*FLRT1-3*). All contain fibronectin and leucine rich repeat motifs in

their extracellular domains that are responsible for homotypic cell adhesion (Haines et al. 2006). Previous studies have shown that the FLRT proteins associate with members of the FGFR family and accentuate FGF signaling via the Ras/Raf/ERK pathway (Bottcher et al. 2004, Haines et al. 2006, Wheldon et al. 2010). Specifically, FGFR1 was shown to phosphorylate FLRT1 at three critical tyrosine residues as part of a futile cycle whereby FLRT-mediated activation of FGFR1 results in suppression of activation by feedback phosphorylation (Wheldon et al. 2010).

At E9.5-11.5 during mouse embryonic development, *Flrt1* is expressed in areas overlapping regions of *Fgfr1* expression; in the midbrain at the boundary with the hindbrain and in the dorsal diencephelon adjacent to the telencephalon, the eye, dorsal root and trigeminal ganglia and in cells adjacent to the urogenital ridge (Haines et al. 2006).

As an interactor of FGFR1 signaling, *FLRT1* makes an attractive candidate craniosynostosis gene, however the position of the variant in individual 2930 (p.Thr46) is not in a functional domain and although well conserved in its orthologues, the N-terminal of the protein before the first domain is not well conserved within the FLRT family (Fig. 4-1). In fact, there is some ambiguity about the position of the initiation methionine, with an alternative start indicated from a methionine 28 residues from the start site denoted in Fig. 4-1. Bearing this in mind, a Kozak consensus sequence (A/GNNATGG) prediction score was calculated for both the wild-type and variant sequences (http://atgpr.dbcls.jp/). The prediction tool in both cases selected the longer length sequence (as depicted in Fig. 4-1) to have the strongest score; 0.5 compared to 0.11 for the alternative. The c.137C>T variant produced an ATG with a score of only 0.04, making it unlikely to affect the protein by creating an illegitimate translation start site.

Given a per-generation mutation rate of 7.6×10^{-9} - 2.2×10^{-8}, we would expect on average 50-100 *de novo* mutations in any given individual, which equates to ~0.86 nonsynonymous variants per genome (Lynch 2010, Roach et al. 2010, Kong et al. 2012). In individual 2930, the observed *de novo* nonsynonymous mutation rate is as expected so we cannot infer pathogenicity simply by the

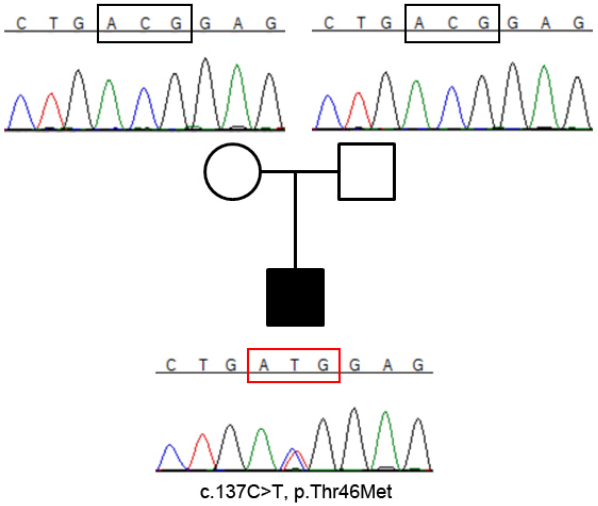
presence of the *FLRT1* p.Thr46Met variant. However, this variant remained suspicious given its interaction with *FGFR1* and was included in a screen of the Oxford craniosynostosis patient panel, using the Fluidigm PCR and Ion Torrent sequencing systems; of 230 patients screened, no suspected pathogenic variants were identified (I. Taylor, K. Miller, Y. Zhou).

4.5.2 Trio analysis – recessive

As well as the *de novo* strategy described above, sequence data for individual 2930 was analysed assuming a recessive inheritance pattern. Variants were filtered for homozygous, or compound heterozygous changes that were heterozygous in both parents. After removal of common variants, this time including variants that occurred in other non-craniosynostosis patients in the WGS500 sequencing project, 20 homozygous and 123 compound heterozygous variants remained. The remaining variants were then sorted for exonic changes, leaving five apparently homozygous variants (three synonymous and two nonsynonymous) and one gene with compound heterozygous variants (Table 4-4).

The apparent homozygous nonsynonymous variants after this filtering process were in fact hemizygous variants on the X chromosome in *OPHN1* and *GEMIN8*. *OPHN1* encodes oligophrenin-1, a Rho-GTPase-activating protein that promotes GTP hydrolysis of Rho subfamily members. The variant p.Thr520Ile has a high deleterious score (5), and is predicted to be possibly damaging by PolyPhen-2 with a HumVar score of 0.700. This position falls within the Rho-GAP domain and shows conservation of either threonine or serine residues at the equivalent position in a variety of mammals, birds and also in *X. tropicalis* and *X. laevis*. Mutations in *OPHN1* have been shown to cause X-linked mental retardation associated with cerebellar hypoplasia and facial dysmorphism, including prominent forehead, long flat facies, strabismus, large ears, long tubular nose and prominent chin (Al-Owain et al. 2011). However, these features do not resemble SCS, and individual 2930 was not noted to have any developmental delay.

A



B

```
                                                *
H.sapiens       1   MVVAHPTATA--TT-----TPTATVTATVVMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQ   100
P.troglodytes   1   MVVAHPTATA--TT-----TPTATVTATVVMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQ   100
M.mulatta       1   MVVAHPTATT--TT-----TTTATVTATVVMTTSTMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQ   100
C.lupus         1   MVVAHPATTA--TT-----TPTATVTATVMMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQ   100
B.taurus        1   MVVAHPTAAAATTA-----TPATTVTATVVMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQ   102
M.musculus      1   MVVAHSAATA--TT-----TPAATVTATVVMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPSDIPDDATTLYLQNNQINNAGIPQ   100
R.norvegicus    1   MVVAHPAATA--TT-----TPAATVTATVVMTTATMDLRDWLF---LCYGLIAFLTEVIDSTT-----------------CPSVCRCDNGFIYCNDRGLTSIPSDIPDDATTLYLQNNQINNAGIPQ   100
D.rerio (a)     1   ----------------------------MMAPEGIGELRDLLFLLLLCFTLLAELLEMTAASSQGLDG-------EDDIVCPSVCRCDEDFIYCNDRGLSAIPP-LPPSATVLYLQNNHIDNAGLPR    92
D.rerio (b)     1   ----------------------------MATESLVELRDWLFLLLLCLTLLVEVLEFAAATAAAAEAALGEAELQGEVPCPSACRCDDGFIYCNDRGLTLIPP-LPLTAAVLYLQNNRIDNAGLPT    98
X.tropicalis    1   MAITTTTTTTTSTTIAASAAATTATATTVAMTTATLDLRDWLF---LCYGLIAFLTEVIDSTN-----------------CPYVCRCDNGFIYCNDRGLTSIPPDIPDDATTLYLQNNQINNAGVPP   107

                                                *
FLRT1           1   MVVAHPTATATTTPTATVTATVVMTTATMDLRDWLFLCYGLIAFLTEVIDSTTCPSVCRCDNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQDLKTKVNVQVIYLYENDLDEFPINLPR   127
FLRT2           1   ------------MGLQTTKWP---SHGAFFLKSWLIISLGLYSQVSKLL---ACPSVCRCDRNFVYCNERSLTSVPLGIPEGVTVLYLHNNQINNAGFPAELHNVQSVHTVYLYGNQLDEFPMNLPK   108
FLRT3           1   --------------MISAAWSIFLIGTKIGL----FLQVAPLSVMAK-----SCPSVCRCDAGFIYCNDRFLTSIPTGIPEDATTLYLQNNQINNAGIPSDLKNLLKVERIYLYHNSLDEFPTNLPK   104
```

Figure 4-1

***FLRT1* mutation in individual 2930.**

A. *De novo* inheritance of c.137C>T *FLRT1* variant (pThr46Met) was confirmed by dideoxy sequencing

B. Alignment of N-terminal amino acids of FLRT1 orthologues (*H.sapiens* FLRT1, *P.troglodytes* FLRT1, *M.mulatta* FLRT1, *C.lupus* FLRT1, *B.taurus* FLRT1, *M.musculus* Flrt1, *R.norvegicus* Flrt1, *D.rerio* flrt1a and flrt1b, *X.tropicalis* flrt) and human paralogues (FLRT2 and FLRT3), p.Thr46 is highlighted in blue, LRRNT domain in yellow and LRR domains in green. The possible alternative initiation methionine is indicated by an asterisk.

Table 4-4

Homozygous and compound heterozygous variants present in individual 2930 after filtering out common variants.

Gene	Variant	Deleterious score	Type	Inheritance
OPHN1	c.1559C>T, p.Thr520Ile	5	Hemizygous	1473 heterozygous
GEMIN8	c.500G>A, p.Arg167His	2	Hemizygous	1473 heterozygous
CDC45	c.318C>T, p.(=) c.773A>G, p.Asp258Gly	0 4	Compound het	1474 heterozygous 1473 heterozygous

Gem-associated protein 8 (encoded by *GEMIN8*) is involved in the assembly of the spliceosome, forming part of the survival motor neuron complex (Carissimi et al. 2006). The variant p.Arg167His falls just inside of the coiled-coil domain (residues 135-168), however, it is predicted to be benign by PolyPhen-2 (HumVar model for Mendelian diseases) with a score of 0.067; the equivalent arginine residue is a histidine in the wallaby and the opossum. Also, although this particular substitution is not present in the Exome Aggregation Consortium dataset (ExAC; http://exac.broadinstitute.org/), there are several missense variants that fall within the coiled-coil domain.

Based on the available information on the functions of *GEMIN8* and *OPHN1*, it seemed unlikely that either mutation in the patient had contributed to his phenotype. The *CDC45* compound heterozygous variants were not initially prioritised, due to the fact that one was a synonymous change. However, another individual with potential compound heterozygous mutations in *CDC45* was identified in a concurrent exome sequencing project. The proband (5566), who had bicoronal synostosis, was heterozygous for the variants c.226A>C (p.Asn76His) and c.469C>T (p.Arg157Cys). Dideoxy sequencing of parental samples confirmed that the mutations occurred on different alleles, and that only one of the variants was present in an unaffected sibling (Fig. 4-3).

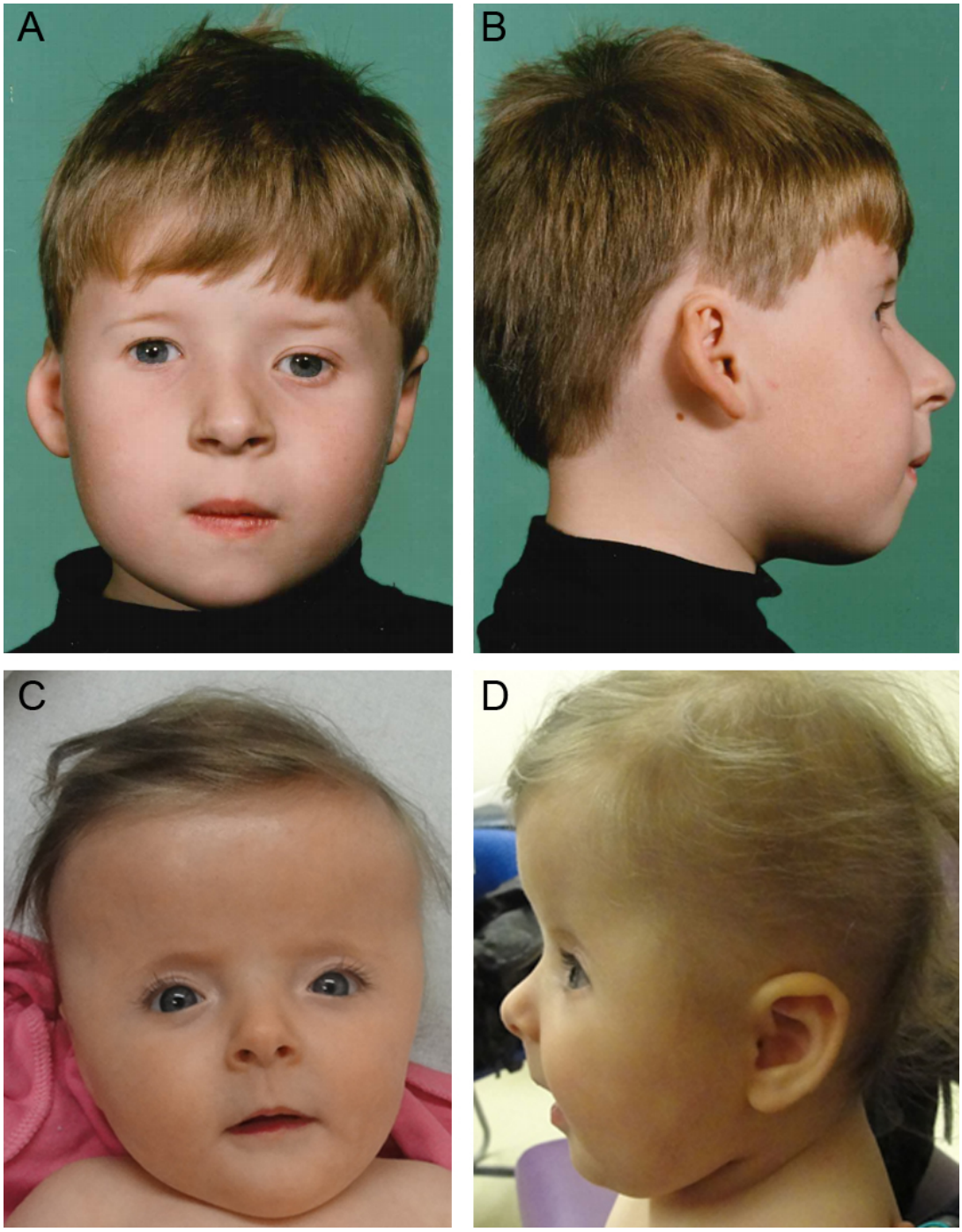


Figure 4-2

Facial appearance of individuals with compound heterozygous *CDC45* mutations.
A, B. Patient 2930 aged 7 years, after he had undergone surgical correction for right unicoronal synostosis. Note facial asymmetry with deviated nasal septum and mildly dysplastic right ear. Whole genome sequencing identified the heterozygous *CDC45* variants c.318C>T, p.(=) and c.773A>G (p.Asp258Gly).
C, D. Patient 5566 aged 6 months, pre-craniofacial corrective surgery, who presented with bicoronal synostosis and short stature (height and weight in 0.4th centile). Exome sequencing revealed the heterozygous *CDC45* variants c.226A>C (p.Asn76His) and c.469C>T (p.Arg157Cys).

4.5.2.1 *CDC45* as a candidate gene

CDC45 (cell division cycle 45) was originally identified as a cold-sensitive cell division cycle mutant in *Saccharomyces cerevisiae* (Moir et al. 1982). Since then, it has been shown to be critical for the initiation and elongation of DNA replication in eukaryotic organisms (Krastanova et al. 2012). During G1 phase of the cell cycle a pre-replicative complex (pre-RC) is formed, comprising origin recognition complex (ORC) proteins and the minichromosome maintenance complex (MCM2-7). The ORCs bind to DNA at the origins of replication (ORIs), and at the G1/S phase transition the helicase activity of the MCM proteins is activated, which unwinds the DNA from the ORIs to allow recruitment of replication proteins and polymerases (Bauerschmidt et al. 2007). CDC45 forms part of a large complex known as the CMG complex, which is the association of CDC45, MCM2-7 and the GINS proteins (Sld5, Psf1, Psf2 and Psf3). Together with phosphorylation events, the CMG is responsible for activation of the MCM helicases. The *D. melanogaster* CMG complex was shown to be able to unwind DNA *in vitro*, however the Mcm2-7 complex alone could not (Aparicio, Ibarra, and Mendez 2006, Krastanova et al. 2012, Moyer, Lewis, and Botchan 2006, Pollok et al. 2007).

Analysis of CDC45 expression in human cells demonstrated that it was much less abundant than the MCM proteins in proliferating cells, which supports the theory that recruitment of CDC45 to ORIs is rate limiting for replication initiation. It was also noted that the protein was absent from terminally differentiated cells, but present throughout the cell cycle of proliferating cells (Pollok et al. 2007).

In the developing mouse embryo, *Cdc45* is expressed in multiple tissues from E9.5-14.5 including the brain, pharyngeal arches, thymus, kidneys, liver, gut lungs, and pancreas (Shaikh et al. 1999). Homozygous mutant mice lacking a functional *Cdc45* die before E7.5, displaying impaired proliferation of the inner cell mass. No abnormalities were detected in the heterozygous mutant mice, which were normal size, fertile and did not display any obvious behavioral deficiencies (Yoshida et al. 2001).

Of note, mutations in the ORC family of proteins, affecting the pre-RC, are associated with Meier-Gorlin syndrome (MGS); an autosomal recessive disorder characterised by microtia, patellar aplasia/hypoplasia, and short stature. Cases have also been described where craniosynostosis is part of the phenotype (de Munnik et al. 2012, Hurst, Winter, and Baraitser 1988).

Furthermore, CDC45 directly interacts with RECQL4; a helicase which has been shown to play an essential role in the formation of the CMG complex. Mutations of *RECQL4* are associated with three overlapping conditions; Baller-Gerold (BGS), Rothmund-Thomson (RTS) and RAPADILINO syndromes. Of these, BGS is characterised by craniosynostosis (usually coronal), skeletal dysplasia with radial ray defects, and growth retardation (Piard, Aral, et al. 2015). Furthermore, there are several reports of patients diagnosed with BGS who have subsequently had *TWIST1* mutations identified (Gripp et al. 1999, Piard, Aral, et al. 2015, Seto et al. 2001), indicating a phenotypic overlap with SCS.

Based on the evidence described above of genes in the DNA replication pathway being associated with craniosynostosis, and also having two patients with coronal synostosis both with compound heterozygous mutations in *CDC45*, it seemed like a plausible candidate disease gene and was investigated further.

4.5.2.2 Determining pathogenicity of *CDC45* mutations

Human *CDC45* produces three transcript variants; a full length isoform of 598 amino acids, a 566 amino acid isoform (with exon 7 spliced out), and a 520 amino acid isoform (with exons 4 and 7 spliced out). For the purposes of this thesis, all numbering will refer to the full length isoform (NM_001178010).

Both of the mutations in individual 5566 were predicted to be pathogenic bioinformatically, with deleterious scores of 5 for the substitution p.Asn76His, and 6 for p.Arg157Cys. In particular the residue p.Asn76 is predicted to be critical in forming the active centre of the DHH phosphoesterase domain, a motif which is conserved throughout eukaryotic orthologues as well as in the archaeal and bacterial orthologue RecJ (Sanchez-Pulido and Ponting 2011). This asparagine is extremely highly

conserved, down to *D. melanogaster*, *C. elegans* and even *A. thaliana* (Fig. 4-3D). The residue p.Arg157 is also highly conserved in eukaryotes, and the substitution from a positively charged, amphipathic residue to the much smaller, neutral cysteine is not a conservative change. It is likely that this missense mutation would affect the protein function, as it occurs within a region unique to eukaryotic orthologues containing many charged residues (Krastanova et al. 2012). This is reflected in the maximum deleterious score of 6 for this variant.

Based on the above, and reports of other synonymous variants being associated with disease (such as in *FGFR2* as discussed in Chapter 3), I decided to investigate whether the synonymous variant in individual 2930 could in fact be pathogenic. The change c.318C>T (leading to a silent substitution at residue p.Val106) is located 25 nucleotides from the end of exon 4 and is not predicted to affect the strength of the wild type splice donor, or create a cryptic splice site (Neural Network). Nonetheless, as a lymphoblastoid cell-line was available for the patient, amplification of his cDNA was carried out to see if anything unusual could be seen, compared with normal controls.

As mentioned above, the variant c.318C>T occurs in exon 4, which is alternatively spliced in one of the three transcripts encoded by *CDC45*. By amplifying cDNA with primers in exons 3 and 5, we might expect two products to be amplified depending on the expression of the different isoforms in lymphoblastoid cells. This appeared to be the case in both the patient 2930, and an unaffected control cDNA sample (also synthesised from RNA extracted from a lymphoblastoid cell line). However, there seemed to be a skewing in the proportion of each of the products (Fig. 4-4A). Gel purification of the larger product (including exon 4), followed by dideoxy sequencing revealed the presence of only the wild type allele in individual 2930 (Fig. 4-4B), indicating that the c.318C>T mutation was somehow promoting exclusion of exon 4. As the variant p.Asn76His in patient 5566 also occurs in exon 4, cDNA from this patient was also sequenced and revealed the presence of both wild type and mutant alleles (Fig. 4-4C).

A qPCR assay was designed to compare the relative amount of products containing exon 4 in patients 2930 and 5566, and four unaffected control cDNA samples. This showed a 0.43 fold reduction in cDNA containing exon 4, compared with exons 1-3, in patient 2930 (Fig. 4-4D). A melt curve analysis was carried out on products amplified from primers in exons 3 and 5, which also indicated a difference in the composition of the cDNA product generated from patient 2930 (Fig. 4-4E).

To try and establish how the variant c.318C>T could be affecting splicing, the splicing regulation tool SROOGLE was used to identify any putative ESEs and ESSs present in exon 4 (http://sroogle.tau.ac.il). The software identified the sequence GTCGTCAA as a potential ESE, as predicted by the Zhang and Chasin method (Zhang and Chasin 2004), which would be removed by the C>T mutation at position 3 of this motif.

The second *CDC45* variant in patient 2930 (p.Asp258Gly) lies in a RecJ motif, as predicted by a multiple alignment of 15 eukaryotic CDC45 sequences, 16 archaeal sequences and 10 bacterial RecJ sequences (Krastanova et al. 2012). Although the aspartic acid residue itself is not conserved to this level, it is likely that substitution to a much smaller and uncharged glycine residue would affect the protein conformation. Likely pathogenicity is also indicated by this substitution being assigned a deleterious score of 4.

Fig 4-3

Confirmation of compound heterozygous *CDC45* mutations by dideoxy sequencing.
A. c.318C>T and c.773A>G (p.Asp258Gly) mutations in patient 2930 occur on different alleles, his unaffected siblings each inherited only one of the mutations.
B. c.226A>C (p.Asn76His) and c.496C>T (p.Arg157Cys) mutations in patient 5566, originally identified by exome sequencing, were confirmed to occur on different alleles. Her unaffected sister inherited only the c.226A>C (p.Asn76His) variant.
C. c.749+5G>A and c.1583C>T (p.Pro528Leu) mutations in individual 2916, identified by Fluidigm/Ion Torrent screen of the Oxford craniosynostosis patient panel, occur on different alleles.
D. Alignment of the N-terminal of CDC45 in various orthologues. The positions of the residues affected in individuals 2930 (p.Val106) and 5566 (p.Asn76) are highlighted in magenta, these occur within exon four, which is indicated by the blue dotted box. The black boxes indicate highly conserved RecJ motifs.

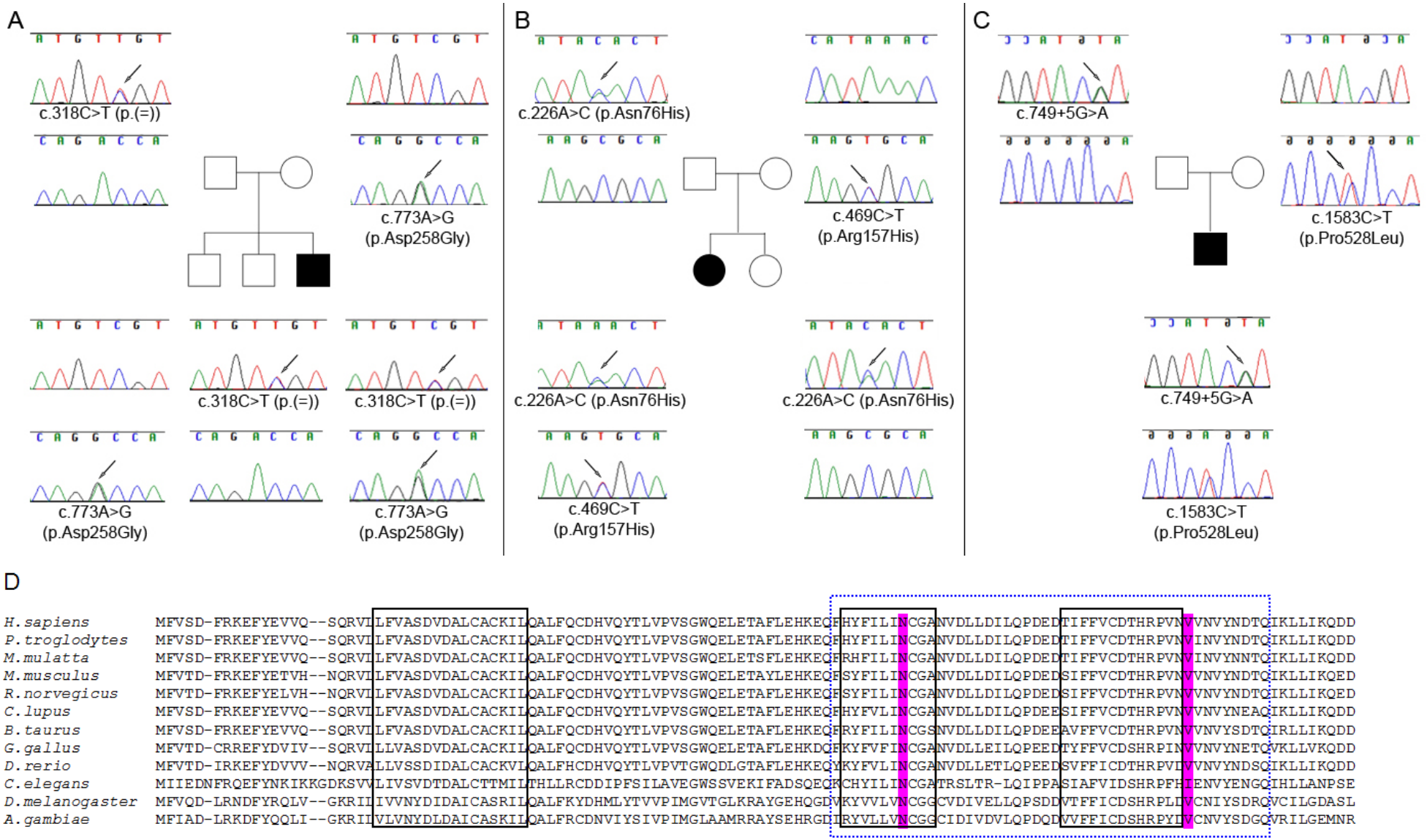

A
c.318C>T (p.(=))
c.773A>G
(p.Asp258Gly)
c.318C>T (p.(=))
c.318C>T (p.(=))
c.773A>G
(p.Asp258Gly)
c.773A>G
(p.Asp258Gly)
B
c.226A>C (p.Asn76His)
c.469C>T
(p.Arg157His)
c.226A>C (p.Asn76His)
c.226A>C (p.Asn76His)
c.469C>T
(p.Arg157His)
C
c.749+5G>A
c.1583C>T
(p.Pro528Leu)
c.749+5G>A
c.1583C>T
(p.Pro528Leu)
D
H.sapiens MFVSD-FRKEFYEVVQ--SQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKEQFHYFILINCGANVDLLDILQPDEDTIFFVCDTHRPVNVVNVYNDTQIKLLIKQDD
P.troglodytes MFVSD-FRKEFYEVVQ--SQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKEQFHYFILINCGANVDLLDILQPDEDTIFFVCDTHRPVNVINVYNDTQIKLLIKQDD
M.mulatta MFVSD-FRKEFYEVVQ--SQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETSFLEHKEQFRHFILINCGANVDLLDILQPDEDTIFFVCDTHRPVNVINVYNNTQIKLLIKQDD
M.musculus MFVTD-FRKEFYETVH--NQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAYLEHKEQFSYFILINCGANVDLLDILQPDEDSIFFVCDTHRPVNVVNVYNDTQIKLLIKQED
R.norvegicus MFVTD-FRKEFYELVH--NQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKEQFSYFILINCGANVDLLDILQPDEDSIFFVCDTHRPVNVVNVYNDTQIKLLIKQED
C.lupus MFVSD-FRKEFYEVVQ--SQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKEQFHYFVLINCGANVDLLDILQPDEESIFFVCDTHRPVNVVNVYNEAQIKLLIKQDD
B.taurus MFVSD-FRKEFYEVVQ--SQRVLLFVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKEQFRYFILINCGSNVDLLDILQPDEEAVFFVCDTHRPVNVVNVYSDTQIRLLIKQDD
G.gallus MFVTD-CRREFYDVIV--SQRVLLLVASDVDALCACKILQALFQCDHVQYTLVPVSGWQELETAFLEHKDQFKYFVFINCGANVDLLEILQPEEDTYFFVCDSHRPINVVNVYNETQVKLLVKQDD
D.rerio MFVTD-IRKEFYDVVV--NQRVALLVSSDIDALCACKVLQALFHCDHVQYTLVPVTGWQDLGTAFLEHKEQYKYFVLINCGANVDLLETLQPEEDSVFFICDTHRPVDVVNVYNDSQIKLLIKQDD
C.elegans MIIEDNFRQEFYNKIKKGDKSVVLIVSVDTDALCTTMILTHLLRCDDIPFSILAVEGWSSVEKIFADSQEQKCHYILINCGATRSLTR-LQIPPASIAFVIDSHRPFHIENVYENGQIHLLANPSE
D.melanogaster MFVQD-LRNDFYRQLV--GKRILIVVNYDIDAICASRILQALFKYDHMLYTVVPIMGVTGLKRAYGEHQGDVKYVVLVNCGGCVDIVELLQPSDDVTFFICDSHRPLIVCNIYSDRQVCILGDASL
A.gambiae MFIAD-LRKDFYQQLI--GKRIIVLVNYDLDAICASKILQALFRCDNVIYSIVPIMGLAAMRRAYSEHRGDIRYVLLVNCGGCIDIVDVLQPDQDVVFFICDSHRPYIVCNVYSDGQVRILGEMNR

Figure 4-3

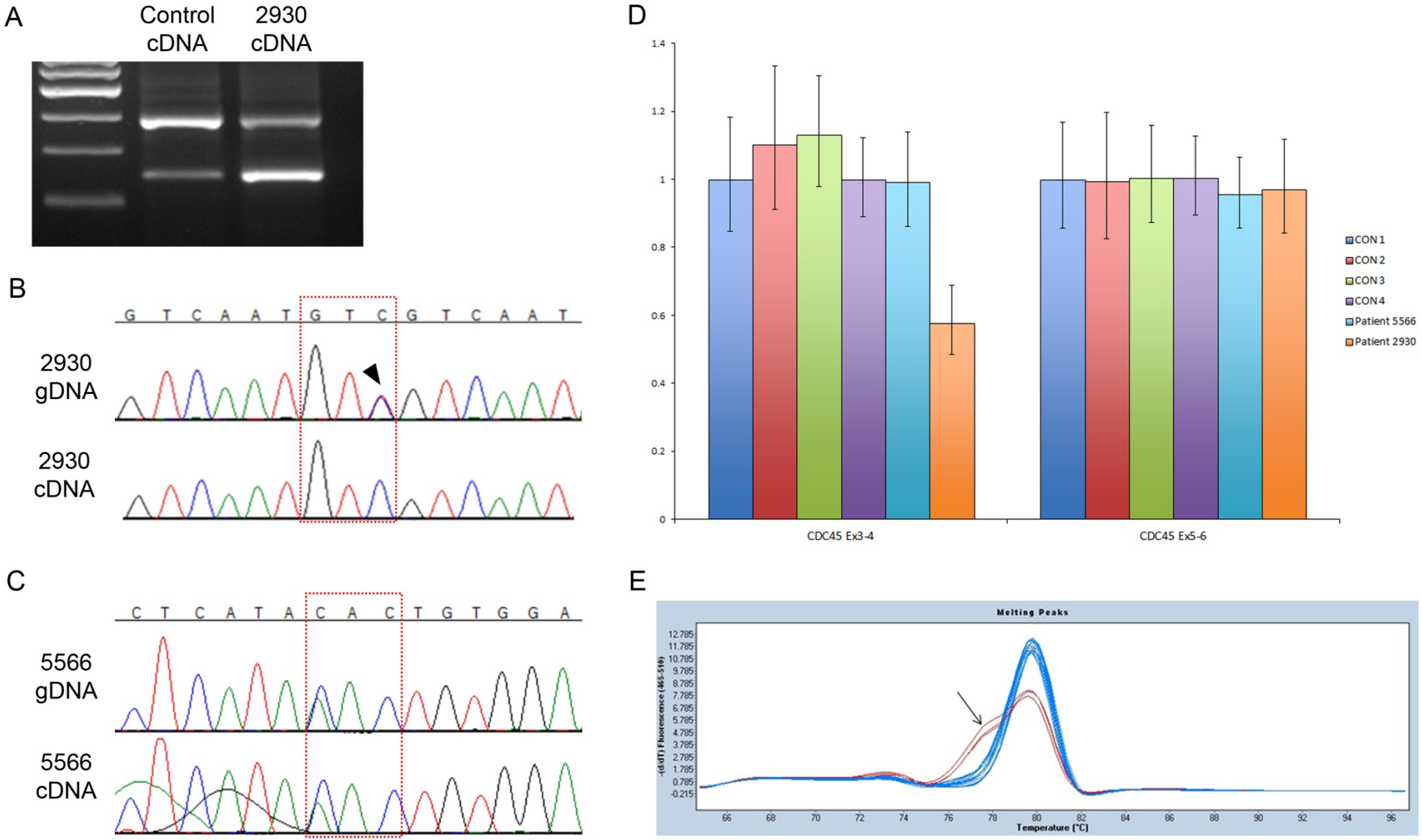

A
Control cDNA
2930 cDNA
B
G T C A A T G T C G T C A A T
2930 gDNA
2930 cDNA
C
C T C A T A C A C T G T G G A
5566 gDNA
5566 cDNA
D
1.4
1.2
1
0.8
0.6
0.4
0.2
0
CDC45 Ex3-4
CDC45 Ex5-6
CON 1
CON 2
CON 3
CON 4
Patient 5566
Patient 2930
E
Melting Peaks
-(d/dT) Fluorescence (465-510)
Temperature (°C)

Figure 4-4

Fig 4-4

Analysis to investigate possible effects of *CDC45* c.318C>T synonymous variant.
A. cDNA samples from a control and patient 2930 were amplified using primers in *CDC45* exons 3 and 5, note skewed ratio of products in 2930 compared to the control sample (larger product represents isoform containing exon four, smaller product is with exon four spliced out).
B. Sequence chromatogram showing position of c.318C>T mutation in patient 2930; the upper sequence trace is generated from genomic DNA (arrowhead indicates C>T transition), the lower trace is generated from cDNA extracted from the upper band of the gel shown in **A**. Note the absence of the mutant allele in the patient's cDNA.
C. Sequence chromatograms generated from genomic DNA (upper panel) and cDNA (lower panel) from patient 5566; the mutation c.226A>C (p.Asn76His) is present in both.
D. Results of qPCR to assay the relative amounts of exon four in patients 5566 and 2930, and four unaffected controls. Products were amplified with primers in exons 3 and 4, and compared with products generated from primers in exons 1 and 3. cDNA from patient 2930 showed a 0.43 fold reduction in exons 3-4 from exons 1-3. Relative amounts of exons 5-6 compared with exons 1-3 are also shown, no differences were observed. All experiments were carried out in triplicate; error bars show propagated standard deviation.
E. Melt curve of PCR products generated from primers in exons 3 and 5, product from patient 2930 is shown in red, control samples in blue. All samples were analysed in triplicate.

CDC45 was subsequently included in a screen of the Oxford craniosynostosis patient panel, using the Fluidigm PCR and Ion Torrent sequencing systems (I. Taylor, K. Miller, Y. Zhou). This led to the identification of a further patient sample (2916) with compound heterozygous variants in *CDC45*; c.749+5G>A and c.1583C>T (p.Pro528Leu). These were confirmed to be on different alleles by dideoxy sequencing of the parents' DNA (Fig. 4-3). The variant c.749+5G>A occurs five nucleotides after the end of exon 9, and is predicted to weaken the wild type splice donor from a score of 0.97 to 0.13 (Neural Network). Of note, there is a cryptic donor with a score of 0.98 135 nucleotides into the intron from the end of exon 9 that may be activated in this circumstance. Unfortunately, we were unable to obtain a sample from the patient from which to extract RNA and test this hypothesis. Individual 2916 was noted to have bicoronal synostosis, short stature, thin eyebrows and radioulnar synostosis.

4.5.2.3 Discussion

Due to the crucial role of *CDC45* in the cell cycle, it is likely that having no functional protein would be incompatible with life, as indicated by the *Cdc45* knockout mouse (Yoshida et al. 2001). As described in section 4.5.2.2, there is a naturally occurring isoform of *CDC45* that does not contain exon 4, which is where both the c.318C>T and p.Asn76His variants are located. This could explain how individuals 2930 and 5566 were not more severely affected, as the presence of some wild-type protein (i.e. from the shortest isoform) may have been enough to allow normal cell cycle progression except for in certain key processes. Although synonymous, the c.318C>T variant in patient 2930 has a more subtle effect by altering the ratio of isoforms towards splicing out exon 4. Similarly, one of the mutations in individual 2916 (c.749+5G>A) is suspected to affect splicing, but not being at the donor site itself it might be "leaky" in that a minority of transcripts may still use the wild type splice site. In this way, enough wild type protein could be produced to allow normal development except for in the cranial sutures. However, not having any cDNA from patient 2916 meant that this theory could not be tested.

As well as the overlapping phenotype of coronal synostosis, all patients had thin eyebrows and individuals 2916 and 5566 had short stature. Patient 2916 also had radioulnar synostosis. This is interesting, as although the main characteristic of RTS (caused by mutations in *RECQL4*) is poikiloderma (areas of hyper/hypopigmentation of the skin), features also include sparse scalp hair, eyelashes and/or eyebrows, small size, radial ray defects and radiographic bone abnormalities that include dysplasia, absent or malformed bones. Small size and radial ray defects are also common in BGS (Piard, Aral, et al. 2015). Furthermore, short stature is typically part of the MGS phenotype. In light of these data, a cohort of MGS patients were screened for mutations in *CDC45* leading to the identification of a further five compound heterozygous and two homozygous mutations (Louise, Bicknell, Andrew Jackson, MRC Human Genetics Unit Edinburgh, unpublished results).

These data now provide genetic corroboration for the role of *CDC45* as a *bona fide* disease gene. The patients originally identified with craniosynostosis appear to fall at the mild end of the phenotypic spectrum, which may reflect a range of severity of the combined effects of two hypomorphic mutations. Furthermore, these results indicate that the *de novo* mutation in *FLRT1* was indeed coincidental to the phenotype, which serves as a reminder that such findings need to be treated with caution.

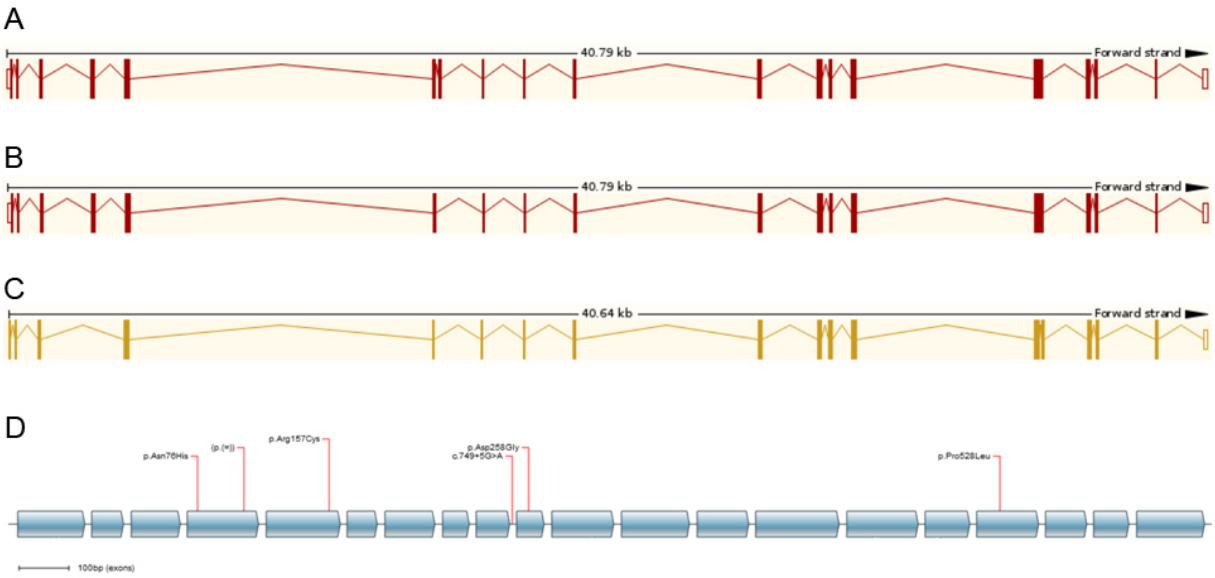


Figure 4-5

Structure of *CDC45*
A, B and C. *CDC45* transcript variants of 598 amino acids, 566 amino acids (exon 7 spliced out) and 520 amino acids (exons 4 and 7 spliced out) respectively.
D. Positions of mutations identified in this study are indicated, the exons are to scale and the introns have been removed.

4.5.3 Combined proband analysis

Owing to the identification of pathogenic variants in patients 4437 and 4454 (*TCF12*), and 2930 (*CDC45*), the number of proband samples remaining from whole genome sequencing in which to carry out combined analysis was reduced to three. Keeping filtering parameters relaxed (not excluding any common variants from eg dbSNP), and searching for variants occurring in the remaining probands 4400, 4580 and 4809 (and *not* in 2930, 4437 and 4454), did not result in any hits

in the same gene in all three samples. Similarly, there were no overlapping hits in any combination of two of the probands 4400, 4580 and 4809 (variant details listed in Tables 4-5, 4-6 and 4-7).

4.5.4 Patient 4809 individual analysis

Looking only at the variants in patient 4809, there were a total of 519 exonic calls. These were filtered to remove those occurring in the ExAC dataset, leaving 209 variants. Of these 209, 79 were nonsynonymous or resulted in a stop mutation, a frameshift or were predicted to effect splicing. At the time of this analysis our bioinformatic pipeline allowed filtering to remove any genes that have loss-of-function variants recorded in the Exome Variant Server (EVS) dataset, which left a list of nine variants (table 4-5). From this list of nine variants, one was immediately interesting; the nonsynonymous change p.Ser503Ala in *RUNX2* (NM_001024630.3). This variant was confirmed to segregate with the phenotype in the affected son of individual 4809 (Fig. 4-6A).

Table 4-5

Predicted deleterious variants present in individual 4809 after filtering.

Gene	Nucleotide change	Amino acid change	Deleterious score	Additional information
RUNX2	c.1507 T>G	p.Ser503Ala	5	Variant also present in affected son.
EGR2	c.769 C>T	p.Arg257Trp	5	Variant also present in affected son. Mutations associated with Charcot-Marie-Tooth disease type 1D and type 4E, and with Dejerine-Sottas syndrome.
MAT2B	c.814 C>G	p.Pro272Ala	5	p.Pro272Thr present in ExAC, allele frequency: 0.000008302
RPRML	c.148 G>C	p.Gly50Arg	5	Very little is known about the function of this gene.
CREBL2	c.281 G>A	p.Gly94Glu	4	*CREBL2* lies in a commonly deleted region on chromosome 12p13 frequently associated with various cancers.
PDAP1	c.526 T>G	p.Ser176Ala	4	Was found to be up-regulated in embryonic fibroblasts from mice lacking promyelocytic leukemia protein (Tang et al. 2013).
NKX2-4	c.835 G>A	p.Val279Met	3	A homeobox gene with a putative role of specifying diencephalic boundaries (Price et al. 1992).
TAGLN2	c.83 A>G	p.Gln28Arg	2	The specific function of this protein is not known, it is thought to be a tumor suppressor.
SPATA31A1	c.1193 T>G	p.Leu398Arg	1	Very little is known about the function of this gene.

4.5.4.1 *RUNX2* as a candidate gene

As mentioned in section 1.3.2, *RUNX2* is a master regulator of osteogenesis, responsible for the activation of osteoblast differentiation genes such as *BGLAP* (osteocalcin), *SPP1* (osteopontin) and *ALP* (alkaline phosphatase). Heterozygous loss-of-function mutations in *RUNX2* cause cleidocranial dysplasia (CCD), characterised by persistently open skull sutures with bulging calvaria, hypoplasia or aplasia of the clavicles and dental anomalies (Mundlos et al. 1997). Although the CCD phenotype is effectively the opposite of craniosynostosis, with wide open sutures, the involvement of *RUNX2* in calvarial development makes it an interesting candidate gene.

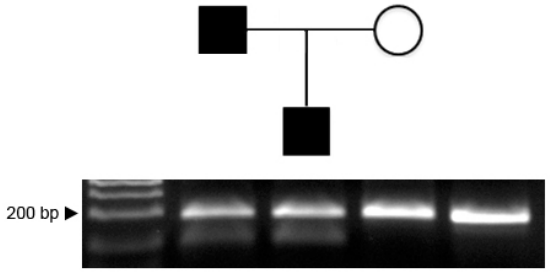
As loss-of-function mutations are associated with CCD, perhaps the *RUNX2* variant p.Ser503Ala results in craniosynostosis via another mechanism. The mutation lies in the C-terminal inhibition domain, within a Pro/Ser/Thr-rich region. Wee et al (2002) used serial alanine mutations of all serine residues in the C-terminal region of *RUNX2* to identify potential phosphorylation sites. This led to the identification of p.Ser465, p.Ser499 and p.Ser503 as phosphorylation targets. These specific serines were then substituted to alanine, separately and combined, in the context of the C-terminal region fused to a GAL4 DNA-binding domain, labelled with ^{32}P and visualised by Western blot. From this, the authors concluded that phosphorylation of p.Ser465 and p.Ser503 is likely to occur independently, but phosphorylation of p.Ser499 requires prior phosphorylation of p.Ser503. However, two-dimensional tryptic phosphopeptide analysis of full-length RUNX2 carrying the p.Ser503Ala mutation did not reveal any observable differences to the phosphopeptide pattern of the wild type protein (Wee et al. 2002).

Another study used mass spectroscopy to identify phosphoserine residues in murine Runx2, which also highlighted the equivalent serine to the human p.Ser503 (mouse p.Ser510) as being phosphorylated. Furthermore, pretreatment of cells with the MAPK inhibitor U0126 removed this phosphorylation, indicating that p.Ser510 is probably an ERK substrate (Ge et al. 2009).

It is difficult to predict if the variant p.Ser503Ala would definitely be pathogenic. The C-terminal region of RUNX2 shows a high level of conservation, as does the RUNX inhibition domain, and specifically the serine residue at the equivalent position (Fig. 4-6B and C). Also there is a lack of known variation at this position, out of 60706 unrelated individuals in the ExAC dataset. As loss-of-function mutations lead to CCD, the serine to alanine substitution would have to exert its effect via a different mechanism. For example, a release of inhibition as p.Ser503 lies within the RUNX inhibitory domain. Broadly speaking, as *RUNX2* is necessary for osteogenesis, a release of inhibition could lead to increased activation of target genes such as *BGLAP*, *SPP1* and *ALP*. Increased *Runx2* expression in mice was shown to correlate with increased osteoblast differentiation and lead to craniosynostosis (Maeno et al. 2011), and duplications including *RUNX2* in humans have also been associated with craniosynostosis: a duplication of a 1.1 Mb was identified in two affected family members with hypodontia and metopic craniosynostosis, and a 3.4 Mb duplication in an individual with right unicoronal synostosis and learning difficulties (Mefford et al. 2010, Wilkie et al. 2007, Varvagiannis et al. 2013). Together these would suggest that the cranial suture is sensitive to dosage of *RUNX2*, and an increase may be linked to premature fusion.

However, more work would be needed to assess the functional outcome of this particular variant. In the first instance, more family members could be genotyped to either rule out this variant as interesting, or indicate that it would be worth following up. The family in question has been invited to participate in further investigative studies, but is yet to provide any samples.

A



B

```
H.sapiens        416  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  521
P.troglodytes    486  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  591
M.mulatta        485  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  590
B.taurus         481  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPTQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  586
M.musculus       423  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSASYQFPMVPGGDRSPSRMVPPCTTTSNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  528
R.norvegicus     424  THYHTYLPPPYPGSSQSQSGPFQTSSTPYLYYGTSSASYQFPMVPGGDRSPSRMVPPCTTTSNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDESVWRPY  529
G.gallus         367  THYHTYLPPPYPGSSQNQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPNQSDGVEADGSHSSSPTVLNSSGRMDESVWRPY  472
D.rerio          362  THYHTYLPPPYPGSTQNQSGPFQTSSTPYLYYGASSGSYQFPMVPGGDRSPSRMLPPCTSASTGSTLINPNLPNQTDGGEADGSHSSSPTVLNSSGRMDESVWRPY  467
X.tropicalis     364  THYHTYLPPPYPGSSQNQSGPFQTSSTPYLYYGTSSGSYQFPMVPGGDRSPSRMLPPC-TTSNGGTLLNPNLANQNDGVDADGSHSSSPTVLNSTGRMDESVWRPY  468
```

C

```
H.sapiens        RUNX2   427 P-GSS-QSQ-----SGPFQTSSTPY-LYYGTSSGSYQFPMVP------GGDRSPSRMLPPC-TTTSNGS------TLLNPN-LPNQND---GVDADGSHSSSPTVLNSSGRMDESVWRPY 521
H.sapiens        RUNX1   359 P-GSS-QAQ-----GGPFQASSPSYHLYYGASAGSYQFSMV-------GGERSPPRILPPC-TNASTGS------ALLNPS-LPNQSD---VVEAEGSHSNSPTNMAPSARLEEAVWRPY 453
H.sapiens        RUNX3   313 P-GAP-QNQ-----SGPFQANPSPYHLYYGTSSGSYQFSMVAGSSS--GGDRSPTRMLASC-TSSAASV---AAGNLMNPS-LGGQSD---GVEADGSHSNSPTALSTPGRMDEAVWRPY 415

Consensus                1 P-GSp-QSQ-----SGPFQSSSSPYHLYYGTSSGSYQFSMVPGGgg-tGGERSPTRMLPPC-TNASTGS------SLMNPN-LPnQSDg--GVEADGSHSNSPTAMSASGRMDESVWRPY 102
R.norvegicus     Runx3   311 P-GAp-QSQ-----SGPFQANPAPYHLFYGTSSGSYQFSMAAAG----GGERSPTRMLTSCpTGASVSAg-----NLMNPS-LG-QAD---GVEADGSHSNSPTALSTPGRMDEAVWRPY 409
D.rerio          Runx1a  355 PaGSs-QAQ-----AGAFQASSSPYHLYYSSAAGSYQFSMMPSGga-aAGERSPPRILP-C-TNASTGS------ALLHPS-LPnQSEg--VVEAEGSHSSSPTSMSV-----EAVWRPY 451
D.rerio          Runx1b  318 P-GN--QNQ-----NSHFQTNSSPYHLYYGTGSGSYQFSMIPTGgsgsGGDRSPTRMLTSC-TAAGGAAgntgnnNLINAN-LGnQSD---GVDADGSHSNSPTAMSASTRIDESVWRPY 424
D.rerio          Runx2a  236 P-RQa-QSSppwsyEQPFQTSSTPY-LYYGASSGSYQFPMVP------GGDRSPSRMLPPC-TSASTGS------TLINPN-LPnQTD---GGEADGSHSSSPTVLNSSGRMDESVWRPY 335
D.rerio          Runx2b  353 P-GSt-QSQ-----SGPFQSSSTPY-LYYGASSGSYQFSMVP------GGERSPTRMMPPC-TSASTGT------SLVNPN-LPvQTDgsgGVEGDGSHSNSPTLLNPAGRMDEGVWRPY 450
T.rubripes       FrRunx1  91 PaNTphQAQ-----NGPFQSPSSAYHLYYSTATGSYQFPMMAGGgg-rGDSRSPPRILPPC-TNASTGS------PLLHPS-LPnQNDg-vGVEVESSHSSSPTNMAAA-----EAVWRPY 191
T.rubripes       FrRunx3 323 PpYTnnQTQ-------NFQSNS---YLYYGTGSGSYQFSMVAPGn--tGGERSPTRLLS-CsGATGTGGa----nSLMNPG-LNaQSE---GVEADGSHSNSPTAMNASGRLDESVWRPY 421
R.eglanteria     Runx3   256 P-GSa-QNQ-----SGPFQTSASPYHLYYGTSAGSYQFSMVA-------GERSPTRMLPSC-TSVSTGS------SLMNPN-LAhQND---VVDRQPSQSNSPTAMTTTGRMDESVWRPY 350
```

Figure 4-6

Possible pathogenicity of RUNX2 p.Ser503Ala variant in individual 4809.

A. The variant was confirmed to be present in patient 4809 and his affected son by restriction digest, *Bsp1286I* cuts the mutant allele into 95 and 88 bp fragments. An uncut control PCR product (283 bp) was also run for comparison, shown on the far right. The variant was not present in the unaffected mother.

B. Alignment of multiple RUNX2 orthologues, p.Ser503 is highlighted in blue; *H.sapiens* RUNX2, *M.mulatta* RUNX2, *B.taurus* RUNX2, *M.musculus* Runx2, *R.norvegicus* Runx2, *G.gallus* RUNX2, *D.rerio* runx2a, *X.troplicalis* runx2.

C. Alignment of human RUNX paralogues shown above the Runx inhibition domain alignment (pfam08504), red indicates highly conserved residues, blue less and grey the least conserved. Unaligned residues are in grey lower case.

4.5.5 Patient 4400 individual analysis

A total of 566 exonic variants were present in individual 4400, which reduced to 189 when variants present in the ExAC browser were removed. Of these, 106 resulted in a predicted splicing effect, stop, frameshift or nonsynonymous substitution. From this list of 106, only seven had no loss-of-function variants in EVS. (Table 4-6). This did not result in any strong candidates.

Table 4-6

Predicted deleterious variants present in individual 4400 after filtering.

Gene	Nucleotide change	Amino acid change	Deleterious score	Additional information
DNASE1L1 (X-chr)	c.1-3845C>T	-	5	Reduces the strength of the splice donor from an exon of the 5´UTR. Mutations are associated with acid maltase deficiency (Pompe disease).
OCRL (X-chr)	c.1775A>T	p.Asn592Ile	4	Mutations cause Lowe syndrome.
ZRANB2	c.722C>A	p.Ser241Tyr	4	Variant not present in affected mother.
GGT1	c.607C>T	p.Arg203Trp	3	p.Arg203Gln present in ExAC, allele freq 0.00008608, rs199703506.
P2RY11	c.952C>G	p.Pro318Ala	3	Belongs to the G-protein coupled receptor family. An *in vitro* assay indicated an involvement in IFN-γ-induced IL-6 production (Ishimaru et al. 2013).
MDP1	c.276A>T	p.Glu92Asp	-	Very little known about this gene – a magnesium-dependent phosphatase.
ELFN1	c.1513G>A	p.Glu505Lys	-	Expressed in the brain in the mouse, when mutated in mice causes seizures and hyperactivity (Dolan and Mitchell 2013).

Further details of genes that were followed up can be found in the Appendix.

4.5.6 Patient 4580 individual analysis

Patient 4580 had a total of 817 exonic variants, this reduced to 222 after removal of common variants. Filtering for frameshifts, stop gains, splicing and nonsynonymous variants resulted in 120 changes, with only seven having no loss-of-function variants in EVS. However, none of these mutations seemed strong craniosynostosis candidates (Table 4-7).

Table 4-7

Predicted deleterious variants present in individual 4580 after filtering.

Gene	Nucleotide change	Amino acid change	Deleterious score	Additional information
TNR	c.1400C>A	p.Thr467Lys	5	Encodes a member of the tenascin family of extracellular matrix glycoproteins.
STX7	c.440+1G>T	-	5	cDNA checked but no effect on splicing was observed.
DYNLL1	c.65T>C	p.Val22Ala	5	Encodes dynein light chain, modulates neuronal nitric oxide synthase activity.
UBQLN4	c.1166A>G	p.Gln389Arg	5	Encodes an adaptor protein that recruits Ubiquilin1 to the autophagy machinery (Lee, Arnott, and Brown 2013).
CRH	c.555C>G	p.His185Gln	4	Expressed in hypothalamus, T-lymphocytes and the placenta.
COG3	c.999G>T	p.Gln333His	3	Involved in ER-Golgi transport.
MNX1	c.2T>C	p.Met1?	-	Mutations in this gene result in Currarino syndrome.

Further details of genes that were followed up can be found in the Appendix.

4.6 Analysis of non-coding variants

Of the six initial probands sequenced, only three remained unsolved (one of whom may have a causative *RUNX2* variant); drastically reducing the power of this experiment to find regulatory region mutations. Nonetheless, all variants within the region chr7: 17385767-19735085 were selected for further analysis. Removing all variants present in the 1000 genomes dataset left 254, 332 and 373 variants in patients 4400, 4580 and 4809 respectively. Any that occurred in other non-craniosynostosis patients in the WGS500 dataset were discounted, and remaining variants visualised in Gbrowse to verify if they looked genuine or were due to sequencing artefacts. This left 15 variants in patient 4400, 31 in patient 4580 and 19 in patient 4809, none of which overlapped (Table 4-8). The nearest variants to each other in different patients, not present in dbSNP, were 2644 bp apart (19304948 in 4809 to 19307592 in 4580). The relatively higher number of variants in patient 4580 is likely to be due to her being of Asian ethnicity and therefore under-represented in the reference genomes.

All of the remaining variants were entered into the Ensembl variant effect predictor (VEP; www.ensembl.org/info/docs/tools/vep/), and also viewed alongside various tracks in the UCSC genome browser to observe conservation, and marks of potential regulatory activity (listed in Table 4-8). Positions of variants were also compared to published data on TWIST1 regulation, such as targets of the promoter indicated by Hi-C analysis (Jin et al. 2013).

Table 4-8

Non-coding variants present in remaining probands after removal of common variants.

Chr7 coordinate	Patient number			Position	Additional information
	4400	4580	4809		
17449943	-	C>T	-	intergenic	rs533757730
17470840	A>T	-	-	intergenic	rs555347601
17486034	-	-	G>A	intergenic	rs574836651
17511134	A>G	-	-	intergenic	rs529755531
17565126	-	-	A>G	intergenic	rs535710700
17565732	-	C>G	-	intergenic	rs548013412
17599829	-	-	C>T	intergenic	
17717623	C>A	-	-	intergenic	
17729292	-	C>T	-	intergenic	TF chip ENSR00001557659
17735688	-	G>T	-	intergenic	rs561339158
17747036	G>A	-	-	intergenic	
17751951	-	C>T	-	intergenic	rs527444671
17769240	-	C>T	-	intergenic	rs527899887
17786144	-	G>T	-	intergenic	rs573524287
17811700	G>A	-	-	intergenic	
17886647	-	T>C	-	*SNX13*: intronic	
17910547	T>C	-	-	*SNX13*: intronic	rs537966756
17911423	T>C	-	-	*SNX13*: intronic	
17943361	-	-	A>G	*SNX13*: intronic	
18059225	-	-	C>A	intergenic	DNaseI + TF chip + H3K27Ac ENSR00000622570
18129943	-	-	T>C	*HDAC9*: intronic	
18131350	-	-	C>T	*HDAC9*: intronic	rs552665326
18139283	G>A	-	-	*HDAC9*: intronic	
18140656	-	-	G>C	*HDAC9*: intronic	
18213199	-	-	T>G	*HDAC9*: intronic	rs538497843
18246217	T>C	-	-	*HDAC9*: intronic	Cons + DNaseI ENSR00000622587
18297020	-	T>C	-	*HDAC9*: intronic	
18307974	-	A>G	-	*HDAC9*: intronic	rs535930225
18316868	-	C>T	-	*HDAC9*: intronic	rs540809438
18337103	-	-	G>T	*HDAC9*: intronic	rs199912078
18345323	-	TGATCTCTG>T	-	*HDAC9*: intronic	
18376575	-	C>T	-	*HDAC9*: intronic	Cons
18422875	-	-	G>A	*HDAC9*: intronic	
18455863	-	G>A	-	*HDAC9*: intronic	rs551657409
18505201	-	A>C	-	*HDAC9*: intronic	rs565908869
18509991	-	C>G	-	*HDAC9*: intronic	
18574692	G>A	-	-	*HDAC9*: intronic	rs541498799
18669961	-	G>A	-	*HDAC9*: intronic	Active enhancer region (Jin et al. 2013)

18682217	-	A>G	-	*HDAC9*: intronic	rs527487944
18891032	-	C>T	-	*HDAC9*: intronic	rs540770807
18927662	-	-	C>T	*HDAC9*: intronic	rs528752322
18987487	-	T>C	-	*HDAC9*: intronic	
19019440	-	-	G>A	*HDAC9*: intronic	
19026597	-	-	G>C	*HDAC9*: intronic	rs529773248
19054942	C>T	-	-	intergenic	rs201926130 DNaseI ENSR00001557738
19113911	C>G	-	-	intergenic	
19129272	-	-	C>T	intergenic	
19133830	-	-	T>G	intergenic	
19154188	-	-	A>G	*TWIST1*: downstream	DNaseI ENSR00000622697
19158684	-	A>G	-	intergenic	rs551653212 DNase1 + TF chip ENSR00001557748
19186326	-	C>A	-	intergenic	
19249943	C>G	-	-	intergenic	
19255397	-	-	G>A	intergenic	Cons + DNaseI ENSR00000622707
19274870	-	G>A	-	intergenic	
19304948	-	-	AGG>A	intergenic	
19307592	-	G>A	-	intergenic	rs542973238
19391000	-	A>G	-	intergenic	
19432451	-	C>T	-	intergenic	
19489619	-	T>C	-	intergenic	
19498027	-	A>G	-	intergenic	
19573091	-	A>G	-	intergenic	
19573653	-	G>T	-	intergenic	
19648674	G>A	-	-	intergenic	
19656720	-	G>A	-	intergenic	rs113511073
19694069	A>G	-	-	intergenic	

Additional information

Ensembl regulatory regions are indicated by an ENSR accession number.

UCSC integrated regulation from ENCODE tracks: TF chip; Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs, DNaseI; DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE, H3K27Ac; H3K27Ac Mark on 7 cell lines from ENCODE.

UCSC conservation track: Cons; Vertebrate Multiz Alignment & Conservation (100 Species).

4.6.1 Further analysis of predicted regulatory region variants

Certain variants were subjected to further investigation, based on their position in relation to previous SCS-associated rearrangements (section 1.4.2) and whether they were in a predicted regulatory region, to see if they were present in family members. Both individuals 4400 and 4809 have affected family members, patient 4580 was assumed to have a *de novo* mutation but unfortunately DNA was not available from her mother for analysis (Table 4-9).

Table 4-9

Results of regulatory region variant family analysis.

Chr7 coordinate	Variant	Regulatory region	Patient	Inheritance pattern
18059225	C>A	ENSR00000622570	4809	Not present in affected son
18246217	T>C	ENSR00000622587	4400	Present in affected mother
19054942	C>T	ENSR00001557738	4400	Not present in affected mother
19154188	A>G	ENSR00000622697	4809	Not present in affected son
19158684	A>G	ENSR00001557748	4580	Not present in unaffected father
19255397	G>A	ENSR00000622707	4809	Present in affected son

4.6.2 Discussion

It is difficult to make any firm conclusions regarding the analysis of non-coding variants. By definition they are extremely difficult to interpret, and so one of the most compelling arguments for pathogenicity is evidence from multiple patients. However, in this case the number of patients was reduced to only three making this an extremely underpowered approach, with none of the variants clustering in a particular region, or overlapping in multiple individuals. Furthermore, there were no variants that coincided with the minimal region described in section 1.4.2, between the rearrangements described by David et al (2003) and Wilkie et al (1995).

To truly investigate the potential contribution of *TWIST1* regulatory mutations to SCS requires a much deeper dataset. Although the original experimental design made sense, technology has moved on such that next generation sequencing is now a much more routine practice. Rather than picking a few patients by strict phenotyping, the opposite approach can be taken; carrying out targeted sequencing of the region around *TWIST1* in as many coronal synostosis patients as possible. This approach is described fully in Chapter 7.

CHAPTER 5

MUTATIONS IN *TCF12* ARE ASSOCIATED WITH CORONAL CRANIOSYNOSTOSIS

Chapter 5: Mutations in *TCF12* are associated with coronal craniosynostosis

5.1 Introduction

During the course of this thesis, a concurrent exome sequencing project identified potential pathogenic mutations in *TCF12* (encoding transcription factor 12; also known as HEB, HTF4 and ALF1) in two out of seven patients with bicoronal craniosynostosis. However both mutations segregated to unaffected individuals in the respective families, so their pathogenic status remained uncertain. This prompted a specific search of my whole genome sequencing data and led to the identification of two different mutations in patients 4437 and 4454. As previously discussed in section 1.3.3, TWIST1 heterodimerizes with class I E-proteins such as TCF12 to bind DNA at E-box recognition sequences, and activate transcription.

5.2 Results

5.2.1 *TCF12* mutations identified through whole genome sequencing

In patient 4437 (Fig. 5-1A), a heterozygous stop mutation (c.1283T>G; p.Leu428*) was identified. This was revealed to be *de novo* by dideoxy sequencing of the parental samples (Fig. 5-1C; microsatellite analysis was used to confirm sample relationships). The second mutation in patient 4454 (Fig. 5-1B), was suspected to affect splicing (c.1035+3G>C, predicted to reduce the wild type 5' donor score from 0.55 to 0.01; Neural Network). Further testing revealed the mutation to be present in the proband's unaffected mother, clinically affected aunt and unaffected grandmother (Fig 5-1D). As this change lies outside the invariant splice donor sequence, cDNA from family members was used to ascertain the exact effect of the mutation. Amplification of the mutant cDNA resulted in a smaller product compared with the wild type. Gel purification and sequencing of this product demonstrated skipping of exon 12 (Fig 5-1F), which is predicted to create an out-of-frame translation product.

A summary of the clinical features exhibited by patients 4437 and 4454 can be viewed in Table 4-1.

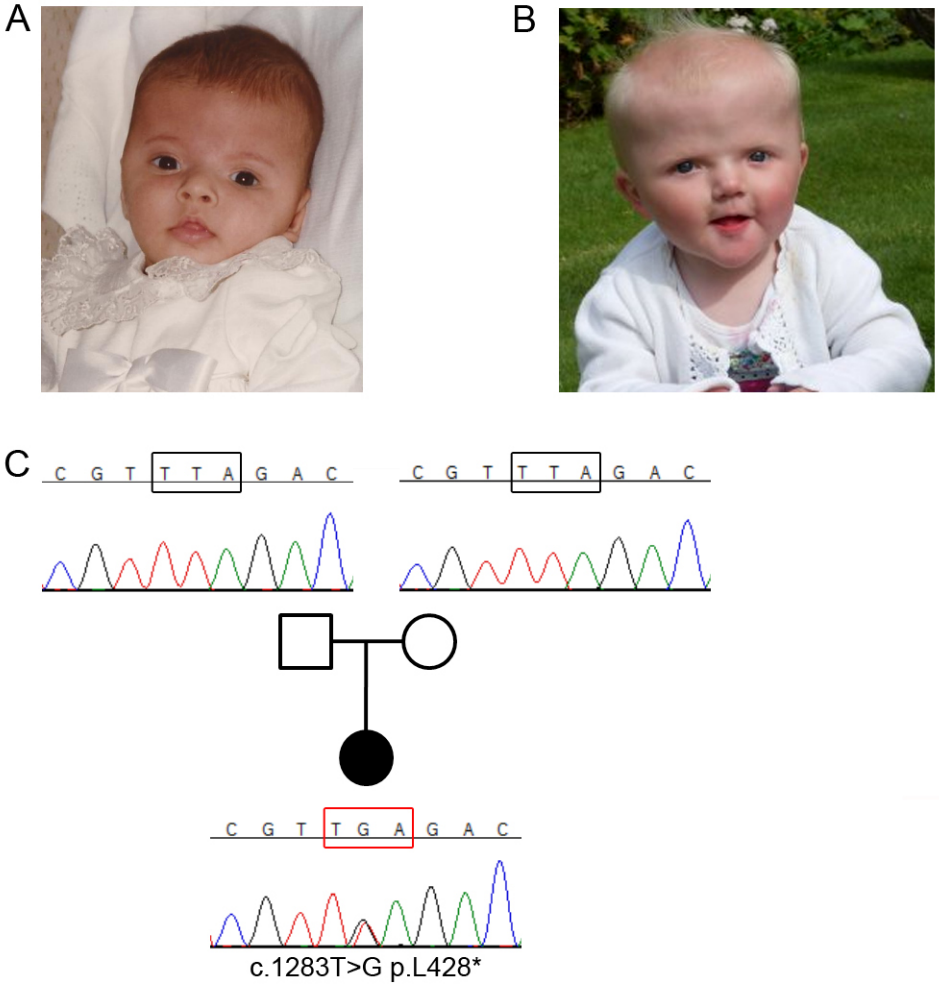

A
B
C
C G T T T A G A C
C G T T T A G A C
C G T T G A G A C
c.1283T>G p.L428*
D
I
II
III
1
2
1
2
3
1
Exon 12
Intron
C A T C T G T G A
c.1035+3G>C
E
I-1 I-2 II-2 II-3 II-1 N
500 bp
400 bp
F
Exon 11
Exon 13
A T T C T A G A T T T A T T

Figure 5-1

***TCF12* mutations identified through whole genome sequencing.**
A, B. Facial appearances of patients 4437 and 4454 respectively. Note brachycephalic head shape and low-set ears.
C. *De novo* heterozygous c.1283T>G mutation in individual 4437 causing a premature termination codon at amino acid position 428.
D. Whole genome sequencing of individual III-1 revealed the mutation c.1035+3G>C, dideoxy sequencing of the family identified the mutation as present in the proband's clinically unaffected mother and grandmother, as well as her affected aunt.
E. cDNA PCR amplification using primers Ex9/10F and Ex14R. Family members with the mutation consistently produce a second smaller amplification product. Sequencing of this smaller gel-purified DNA band, **F**, revealed skipping of exon 12.

5.2.2 Analysis of allelic expression

In a targeted screen, a total 36 pathogenic *TCF12* mutations were identified in 38 unrelated individuals, including 14 *de novo* mutations (Fig. 5-3). Intriguingly, over half (53%) of mutation positive individuals identified did not display any clinical features, demonstrating a high degree of non-penetrance (Sharma et al. 2013). For example, affected individuals with mutations c.1349delC (bicoronal) and c.1491dupT (right unicoronal) both have clinically unaffected, mutation positive mothers. To investigate whether under-expression of the wild type allele could be associated with the observed variability of clinical features, cDNA was amplified from clinically affected individuals and mutation positive relatives and the products subjected to restriction digest (Fig. 5-2). In all but one of the individuals analysed, the mutant allele appeared under-represented compared to the wild type indicating that the mutant product was subjected to nonsense-mediated decay. Direct comparison of mutant:wild type ratios in clinically affected and unaffected relatives did not however demonstrate a consistent pattern correlating with the observed phenotypes (Table 5-1).

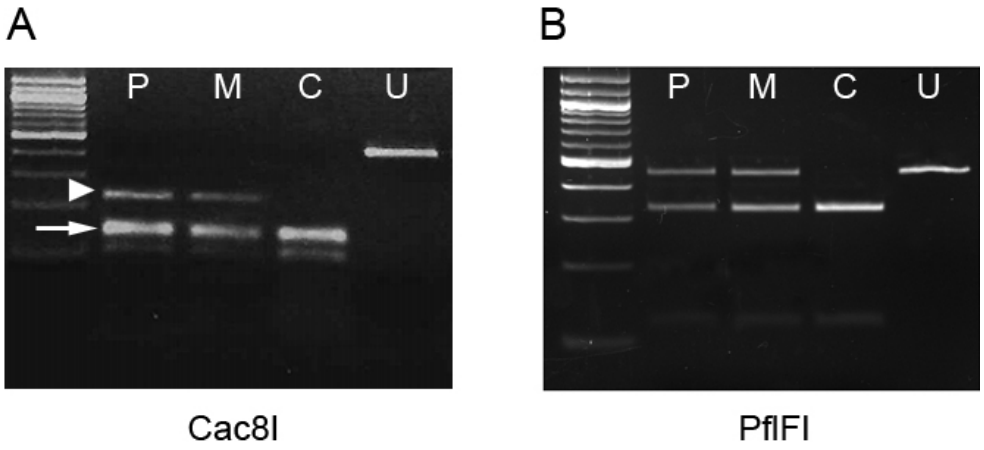


Figure 5-2

Analysis of *TCF12* allelic expression in two families showing variable penetrance.
A. c.1349delC: as above, but showing result of *Cac*8I digest, which cuts only the wt allele. Note lower amount of mutant (white arrow head) compared to wt fragment (white arrow) in both individuals (P, proband; M, clinically unaffected mother), suggesting that the mutant transcript is unstable.
B. c.1491dupT: gel showing result of *Pfl*FI digest of cDNA PCR product (primers Ex13/14F and Ex18R). *Pfl*FI cuts only the wt allele; note the variable expression of the mutant allele (upper band) between the proband (P) and his clinically unaffected mother (M). C, normal control cDNA; U, uncut PCR product.

Table 5-1

RNA studies of individuals heterozygous for *TCF12* mutations.

Mutation	Protein	Individuals analysed and disease status[a]	Cellular source of mRNA	Reverse transcriptase-PCR result	Mutant allele under-represented compared to wild type?	Similar mutant:wt ratio in affected and unaffected individuals?
c.526+1G>A	sp	Grandfather (UA) Mother (UA) Proband (RC)	blood fibroblasts	skipping of exon 7	Y	na (different tissue sources)
c.822C>G	sp	Proband (RC)	fibroblasts	abolishes exon 10 splice donor site	Y	na (*de novo* mutation)
c.825G>C	p.L275F(sp)	Proband (BC)	blood	abolishes exon 10 splice donor site, use of cryptic splice donor 23 bp into intron	Y	na (*de novo* mutation)
c.1035+3G>C	sp	Grandmother (UA) Mother (UA) Aunt (BC)	blood	skipping of exon 12	Y	Y
c.1349delC	p.P450Lfs*69	Mother (UA) Proband (BC)	blood	lower expression of mutant allele	Y	Y
c.1468-20T>A	p.V490Gfs*4 (sp)	Proband (BC) Proband's daughter (UA)	blood	use of cryptic acceptor splice site for exon 17	Y	Y
c.1491dupT	p.V498Cfs*12	Proband's father (UA)	blood	lower expression of mutant allele	Y	-
c.1491dupT	p.V498Cfs*12	Mother (UA) Proband (RC)	blood	lower expression of mutant allele in proband; similar expression of both alleles in proband's mother	Y/N	lower in affected
c.1582+2T>G	sp	Proband (BC)	fibroblasts	faint ~100 bp shorter fragment, consistent with skipping of exon 17	Y	-

[a]UA, unaffected; RC, right coronal; BC, bilateral coronal

5.2.3 Analysis of *TCF12* splicing mutations

Of the 36 mutations identified, there were six thought to affect correct splicing of *TCF12* (including c.1035+3G>C discussed in section 5.2.1). As well as two canonical splice donor site mutations (c.526+1G>A; c.1582+T>G), there were others of less obvious significance; *de novo* c.822C>G which results in a synonymous change four nucleotides before the end of exon 9, *de novo* c.825G>C causing the nonsynonymous substitution p.Leu275Phe, and c.1468-20T>A suspected of creating a cryptic intronic splice acceptor site (Fig. 5-4). These additional five mutations were subjected to further analysis to determine their pathogenicity (Table 5-1; details of primers in section 2.6.2).

As predicted, both canonical splice donor site mutations disrupted normal splicing. The c.526+1G>A mutation was of particular interest as it was the most 5´ mutation identified in *TCF12*, and also the only mutation at the time to be located upstream of the alternatively spliced first exon of a smaller isoform (HEBAlt). Therefore, demonstrating pathogenicity of this mutation would implicate the full length HEB isoform as being critical, rather than HEBAlt. Amplification of cDNA from three mutation positive individuals resulted in an additional smaller fragment, not seen in the control cDNA. Dideoxy sequencing of this smaller fragment revealed skipping of exon 7 (Fig. 5-4A). Similarly, amplification of cDNA from the individual with the c.1582+2T>G mutation also resulted in an extra fragment, approximately 100 bp smaller than the expected product. This size is consistent with skipping of exon 17 (Fig. 5-3E).

The apparently synonymous substitution c.822C>G was suspected of affecting splicing, lying only four nucleotides from the end of exon 9. The normal exon 9 donor site is unusually weak (score 0.09; Neural Network), and the score is little changed by the substitution (score 0.06). However, sequencing of amplified cDNA showed that the mutant allele is virtually absent from the normal cDNA, this was confirmed by restriction digestion (Fig. 5-4B).

Although the mutation c.825G>C may exert its effect by giving rise to the missense variant p.Leu275Phe, its proximity to the end of exon 10 merited further investigation. Sequencing of

amplified cDNA product from the proband demonstrated the presence of intronic sequence, indicating that the exon 10 splice donor site is abolished (Fig. 5-4C).

A text search of current exome sequencing projects led to the identification of the variant c.1468-20T>A, which was not called automatically due to its location within the intron. This mutation was predicted to create a cryptic acceptor with a higher score (0.99) than the normal acceptor (0.96). Use of this cryptic acceptor was confirmed by sequencing cDNA from the affected individual, demonstrating a low level of the neo-exon, which contains a premature stop codon (Fig. 5-4D).

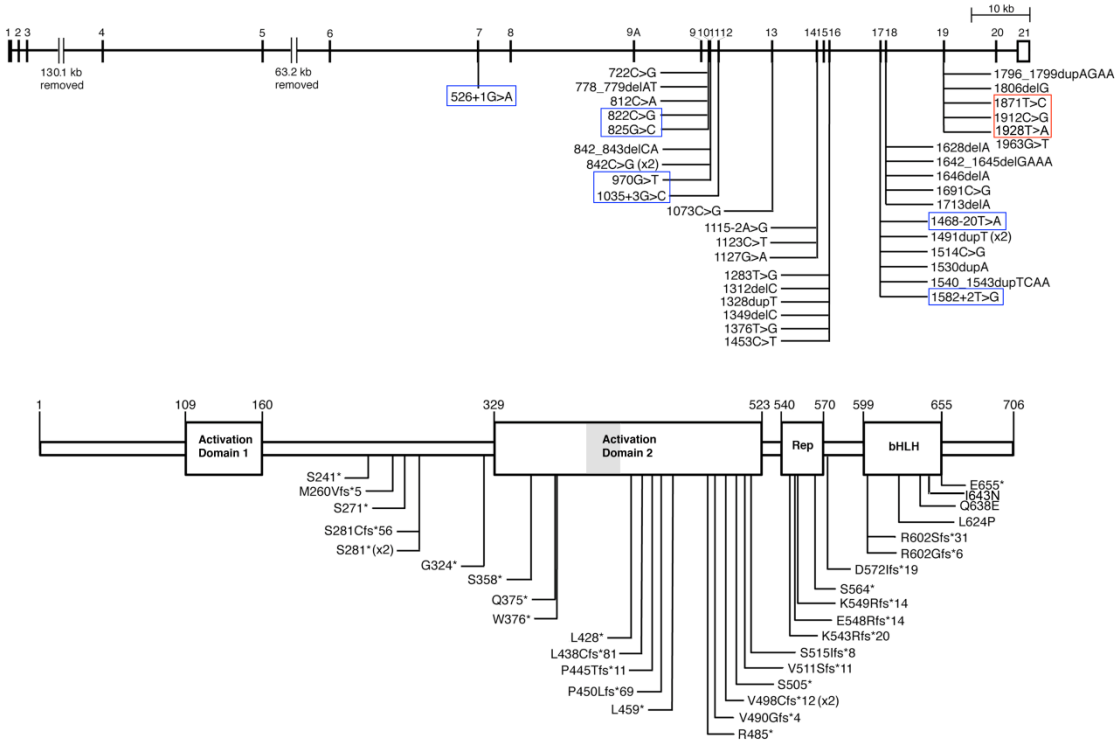


Figure 5-3

Structure of *TCF12* and the encoded protein.
Upper part shows TCF12 gene structure, the major transcript encodes 21 exons; alternative splicing of exon 15 occurs to generate proteins of 706 and 682 amino acids. Alternative transcription starting at exon 9A generates a different protein (HEBAlt) of 512 amino acids. Mutations affecting splicing are indicated in blue, and mutations causing nonsynonymous substitutions in the bHLH domain are indicated in red.
Lower part shows protein structure and identified domains. Identical mutations identified in independent families; x2.

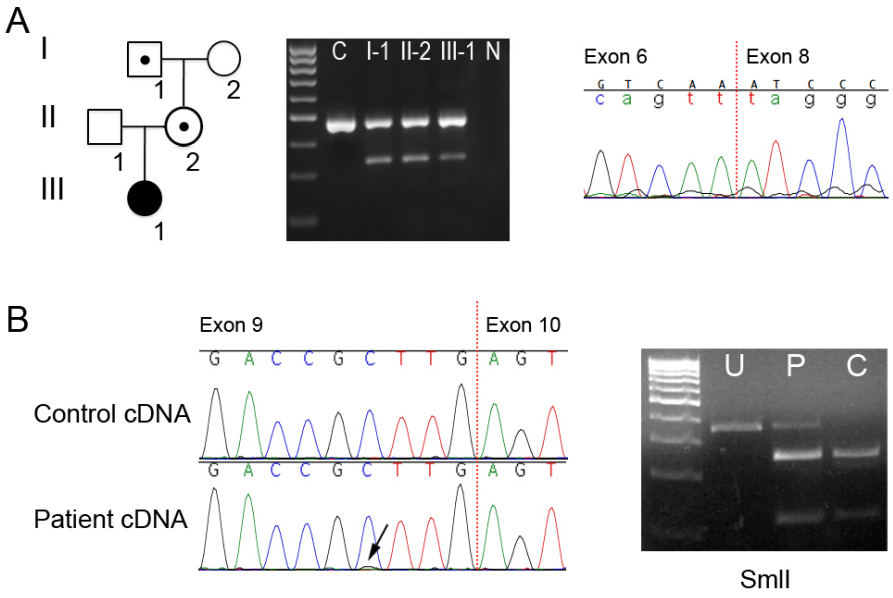

A
I
II
III
1
2
1
2
1
C I-1 II-2 III-1 N
Exon 6
Exon 8
B
Exon 9
Exon 10
Control cDNA
Patient cDNA
U P C
SmlI
C
Exon 9
Exon 10
I
II
1
2
1
II-1
I-1
I-2
G T A G G C T A T A
D
Exon 16
Exon 17
...AATG gt...
...tgggttttccttaactttag GTTG...
Control cDNA
Individual I-1 cDNA
...AATG gt...
...ag ggttttccttaactttag GTTG...
E
P C N
Ex16F / Ex18/19R PCR

Figure 5-4

Analysis of *TCF12* splicing mutations.

A. c.526+1G>A: this was the most 5' mutation identified in *TCF12*, occurring upstream of the alternatively transcribed exon 9A, its effect on splicing was investigated. cDNA PCR amplification using primers Ex5F and Ex9/10R produces an additional smaller fragment present (at a lower amount) in all 3 mutation carriers (I-1, II-2, III-1), but not in a normal fibroblast control (C); N, negative water control. Note that cDNA is from a different tissue source in the proband so mutant:wild type ratio cannot be compared with relatives. Sequencing of this smaller fragment demonstrated skipping of exon 7, predicted to create an out-of-frame translation product.

B. c.822C>G: this *de novo* substitution creates a synonymous change 4 nucleotides before the end of exon 9 (vertical red dotted line). Sequencing of cDNA with primers Ex9F and Ex11R showed that the mutant allele is virtually absent from the normal cDNA (arrow), this was confirmed by restriction digestion using *Sml*I, which cuts only the wild type allele, showing only a small proportion of undigested product in the proband's sample (P; U, uncut; C, normal control).

C. c.825G>C: this *de novo* variant may lead to the amino acid substitution p.Leu275Phe (indicated by an arrow), however sequencing of cDNA with primers Ex9F and Ex11R demonstrates the presence of intronic sequence in the proband (indicated in black text) indicating that the exon 10 splice donor site is abolished.

D. c.1468-20T>A: normal splice junctions between exons 16 and 17 are shown (exon sequence in upper case, intron sequence in lower case). The position of the T>A mutation is indicated in red in the lower sequence, lying 20 nucleotides upstream of the normal acceptor site, in the intron. This mutation is predicted to create a cryptic acceptor with a higher score (0.99) than the normal acceptor (0.96). cDNA sequencing using primers in exons 16 and 18/19 confirmed the use of this cryptic acceptor, which demonstrated a low level of the neo-exon (dotted box) and contains a premature stop codon (underlined in intron).

E. c.1582+2T>G: gel of cDNA PCR products amplified using primers Ex16F and Ex18/19R, note presence of fragment in proband (P) ~100 bp smaller than expected product (black arrow head), consistent with skipping of exon 17 (C, normal fibroblast control; N, negative water control).

5.2.4 Screen for intronic mutations

Due to the presence of the previously described mutation (c.1468-20T>A), and reports of deep intronic mutations in other genes being disease-causing (Flanagan et al. 2013), a screen of cDNA was undertaken to identify any possible abnormal splicing products in patients with bicoronal synostosis who were negative for *TCF12* coding mutations. A total of 15 patients had cell lines available for RNA extraction, from which cDNA was synthesised and amplified using overlapping primers covering the entire *TCF12* coding region (Fig. 5-5; primer details listed in section 2.6.4). No abnormally sized products indicative of aberrant splicing were observed when amplified cDNA products were visualised by agarose gel electrophoresis (Fig. 5-5).

Whole genome sequencing data were also closely scrutinised in the remaining unsolved SCS patients for any *TCF12* intronic mutations. Candidate variants were observed in Gbrowse to verify that they looked like true changes rather than sequencing artefacts and to rule out those that occurred in unaffected individuals. The remaining intronic variants were then assessed using Neural Network, which narrowed the results to two variants; chr15 57268541 G>C in patient 4809 (lying between exons three and four), and chr15 57546670 A>T in patient 4580 (between exons 16 and 17), both predicted to potentially create cryptic splice acceptor sites. Both individuals had EBV cell lines available for RNA extraction, and were included in the above screen.

5.2.5 Screen for *TCF12* deletions

Probes were designed to 12 of the 21 exons of *TCF12*, for a multiplex ligation-dependent probe amplification (MLPA) assay to identify mid-size (>100 bp) deletions (details in section 2.6.5). Exons 2 (containing the start methionine), 3, 4, 5, 6, 7, 10, 13, 16, 17, 19 and 20 (last coding exon) were assayed as these were distributed throughout the gene and likely to pick up deletions of closely adjacent neighbouring exons not directly targeted. No deletions were identified out of a total of 226 patients, however the assay was used to confirm and locate more precisely the extent of a deletion identified by array-CGH.

The proband was a 72-year-old female with short stature (146 cm; −3.5 SD), relative obesity (56.5 kg; +0.5 SD), and an occipitofrontal circumference at the lower limit of normal (52.5 cm; −2 SD). She had distinctive facial features including apparently small eyes, a thin upper lip, high palate, a prominent chin and facial asymmetry with left central facial nerve palsy. Array-CGH revealed a small (84–121 kb) microdeletion of 15q21.3 including the 5´ end of *TCF12*. As the resolution of the array did not show if exon 19 (encoding the bHLH domain) was included in the deletion, it was not clear if the effect on *TCF12* had contributed to the phenotype or not. MLPA analysis indicated a heterozygous deletion of exons 19 and 20, including the bHLH, and therefore the microdeletion is likely to be pathogenic (Fig. 5-6). The other gene included in the deleted interval is a long intergenic non-protein coding RNA (*LINC00926*), whose function is still unknown (Piard, Roze, et al. 2015).

Figure 5-5

Screen of *TCF12* cDNA PCR products for abnormal splice variants.
cDNA from *TWIST1*-negative bicoronal synostosis patients (including *TWIST1*-negative SCS patients 4580 and 4809), amplified with the following primer pairs:
A. 5'UTR F / Ex6/7 R
B. Ex5 F / Ex11 R, positive control (c.526+1G>A) loaded into lane 18, note presence of extra fragments indicative of abnormal splicing (NB gel picture is scanned and sellotape from original picture is covering last lane).
C. Ex9/10 F / Ex 14 R, positive control (c.1035+G>C) was loaded into lane 18; smaller fragment representing a mutant product lacking exon 12 is visible.
D. Ex13/14 F / Ex18/19 R, positive control (c.1582+2T>G) loaded into lane 18. Extremely weak mutant product, predicted to skip exon 17, just visible (arrowhead).
E. Ex16 F / 3'UTR R

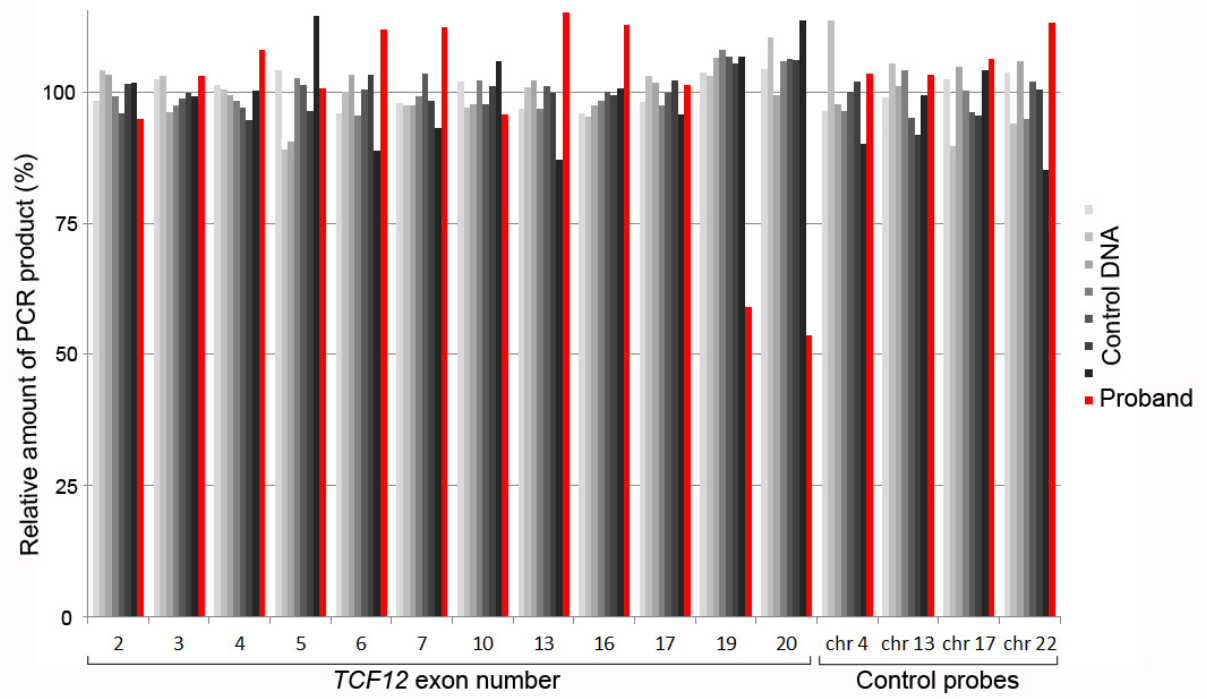


Figure 5-6

MLPA analysis of TCF12 in patient with 15q21.3 microdeletion.
Signals from the proband's DNA (red) were compared with seven normal control samples (shaded grey), twelve different exons of *TCF12* and four control loci were analysed. In the proband's sample the signals for exons 19 and 20 are 59% and 54% of normal values respectively, indicating a heterozygous deletion of these exons. Exon 17 and more 5' exons of *TCF12* exhibit a normal dosage (95-115%).

5.2.6 Investigation of *TCF12* missense mutations

Three predicted missense variants were identified in the highly conserved bHLH region of *TCF12* required for dimerization; p.Leu624Pro, p.Gln638Glu and p.Ile643Asn. Although these were predicted to be pathogenic based on the high conservation of this domain, and *TWIST1* bHLH mutations being pathogenic (Gripp et al. 1999), a functional assay was undertaken to more accurately determine their effects. Previous work demonstrated that transcriptional activation by Twist1 required the presence of the E-protein E12, and was E-box-dependent (Laursen et al. 2007). To determine if the *TCF12* bHLH missense variants affected dimerization with TWIST1, and were therefore pathogenic, a similar transactivation assay was carried out. Firstly, the mutations p.Leu624Pro, p.Gln638Glu and p.Ile643Asn were introduced into full length *TCF12* cDNA clones by site-directed mutagenesis (Fig. 5-7), as well as two missense changes believed to be non-pathogenic polymorphisms (p.Ser210Arg and p.Asn328Asp; mutagenesis primer sequences listed in section 2.6.6). All clones were fully sequence verified to ensure that only the desired point mutation had been introduced.

To determine the optimal amounts of each construct to transfect, wild type TCF12 and TWIST1 expression vectors were co-transfected into the HT1080 cell line with a lacZ reporter construct containing three E-box sequences, at a range of concentrations. Cells were harvested after 24 hours and their outputs compared (Fig. 5-8). Based on these results, 50 ng DNA was selected as the optimum amount to transfect as it elicited a measurable response above 0 ng, without being as sensitive to slight variations, for example due to pipetting error, as the range 0-10 ng. Accordingly, each missense variant was co-transfected with the wild type TWIST1 expression vector, and lacZ reporter construct containing three E-boxes, or a negative control reporter construct with no E-box motifs (full details in section 2.6.7).

Corroborating the findings of Laursen et al (2007), the combination of native TCF12 and TWIST1 proteins had a synergistic effect on activation relative to the activity of either protein individually.

This effect was 76% lower in the presence of the p.Leu624Pro mutation, and reduced 65% and 53% in the presence of the p.Gln638Glu and p.Ile643Asn mutations respectively (Fig. 5-9C). Transcriptional activation after transfection with the missense variants p.Ser210Arg and p.Asn328Asp did not differ from that of wild type TCF12 with TWIST1, indicating that they are unlikely to be the cause of craniosynostosis in these individuals.

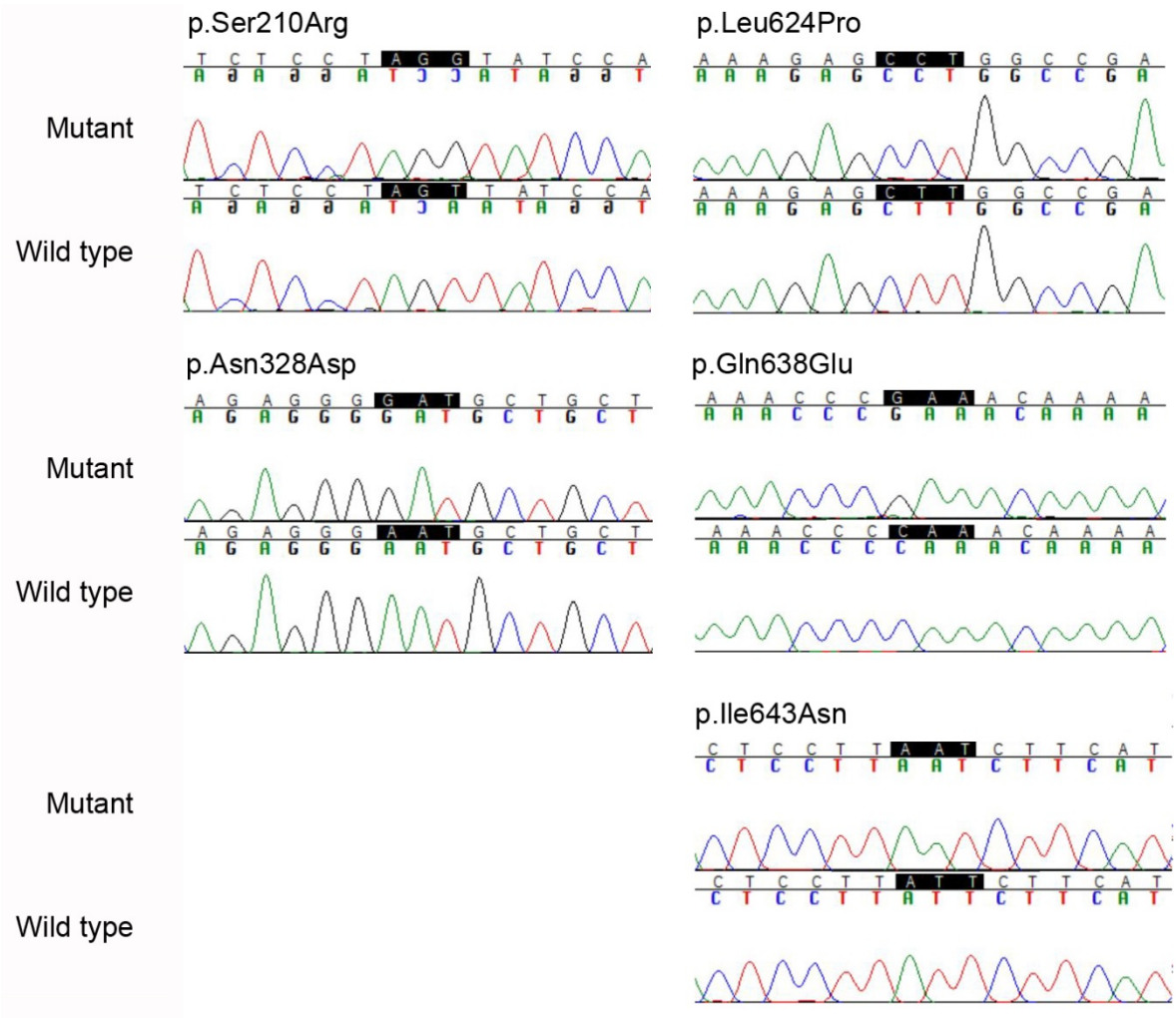


Figure 5-7

Site-directed mutagenesis of *TCF12* to introduce missense mutations.
Sequence chromatograms verifying successful introduction of five missense mutations to full length *TCF12* cDNA clones; p.Ser210Arg and p.Asn328Asp believed to be non-pathogenic polymorphisms (left hand side), and p.Leu624Pro, p.Gln638Glu and p.Ile643Asn all falling within the highly conserved bHLH domain and believed to be pathogenic (right hand side).

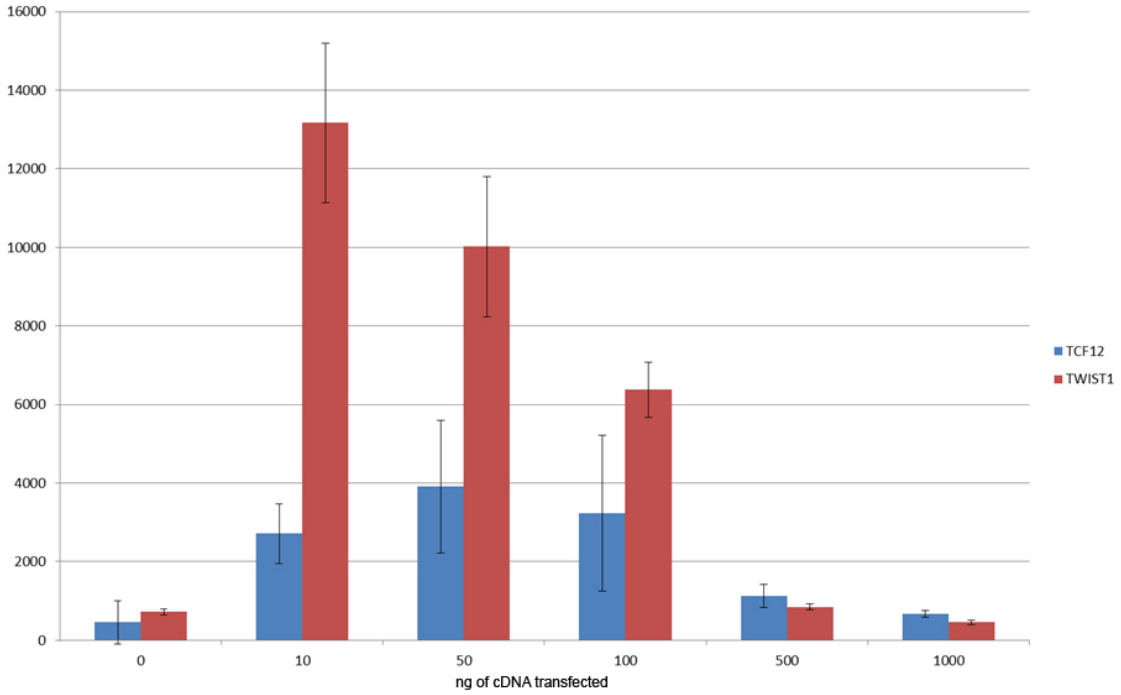


Figure 5-8

Optimisation of transfections with wild type TCF12 and TWIST1 expression vectors.
Analysis of transactivation with varying amounts of TCF12 and TWIST1 wild-type proteins as shown, using a reporter containing three E-box DNA binding motifs. Results are shown as mean±SEM of 3 experiments, output is measured in relative light units (RLU).

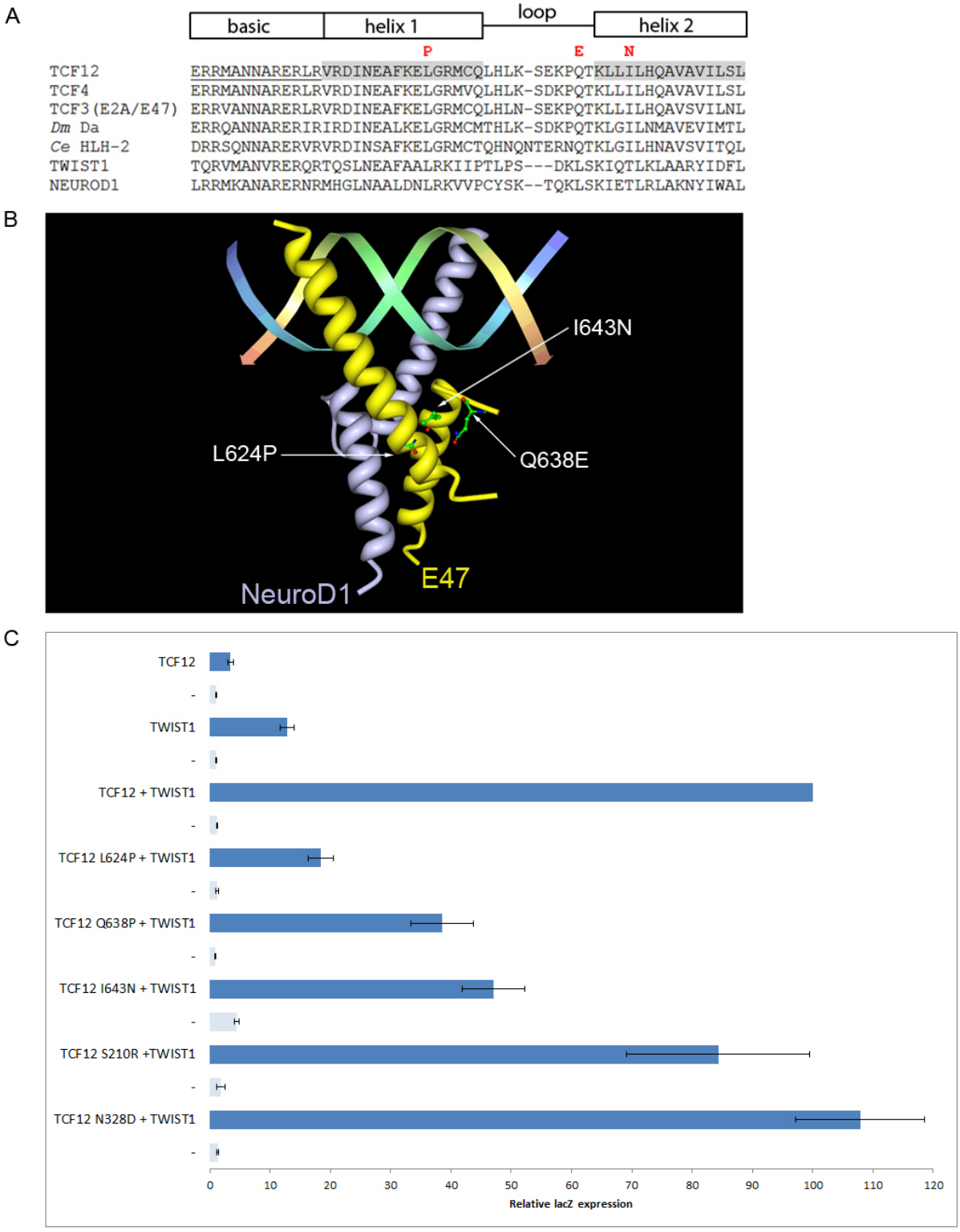


Figure 5-9

Structural and functional analysis of missense mutations in the bHLH domain of TCF12.
A. Comparative amino acid sequence analysis of the bHLH domains of the three human class I proteins (TCF12, TCF4 and TCF3; the latter is frequently termed E2A with E12/E47 alternative spliceforms), the unique homologs in Drosophila melanogaster (Da- daughterless) and Caenorhabditis elegans (HLH-2), and human class II proteins TWIST1 and NEUROD1. The missense mutations observed in TCF12 (p.Leu624Pro, p.Gln638Glu, p.Ile643Asn) are indicated in red.

B. Structure of a bHLH class I/II heterodimer (murine E47/NeuroD1) bound to DNA, visualised in Protein Workshop (PDB accession no. 2QL2). The side chains of the residues corresponding to p.Leu624, p.Gln638 and p.Ile643, which are conserved between TCF12 and TCF3, are highlighted. Substitutions at these buried residues are predicted to abrogate dimerization.
C. Analysis of transactivation with TCF12 and/or TWIST1 wild-type or mutant proteins as shown, using reporters either including (dark blue) or omitting (-, light blue) three E-box DNA binding motifs. Results are shown as mean±SEM of 3 experiments and normalised to the output from transfecting both TCF12 and TWIST1 with the E-box reporter.

5.2.7 Dominant negative screen

To investigate the possibility that premature termination mutations may result in a truncated TCF12 protein and act in a dominant negative manner, six different termination mutations were introduced by site-directed mutagenesis throughout the full length TCF12 cDNA clone: p.Gly42*, p.Arg135*, p.Ser241*, p.Leu428*, p.Arg602fs*31 and p.Glu655* (Fig. 5-10). The latter four were identified in patients with bicoronal synostosis, no termination mutations were identified 5´ to p.Ser241* so p.Gly42* and p.Arg135* were arbitrarily chosen to equally space out mutations throughout the gene. All clones were fully sequence verified to ensure that only the desired point mutation had been introduced. Constructs containing these six mutations were co-transfected with wild type TCF12 and wild type TWIST1 in a 1:1:2 ratio (mimicking a heterozygous state) along with the *lacZ* reporter constructs described in section 5.2.6. All of the terminating mutations demonstrated similar activity to a 50% reduction in the amount of wild type TCF12 (Fig. 5-11), indicating that they were acting in a haploinsufficient, rather than a dominant negative manner.

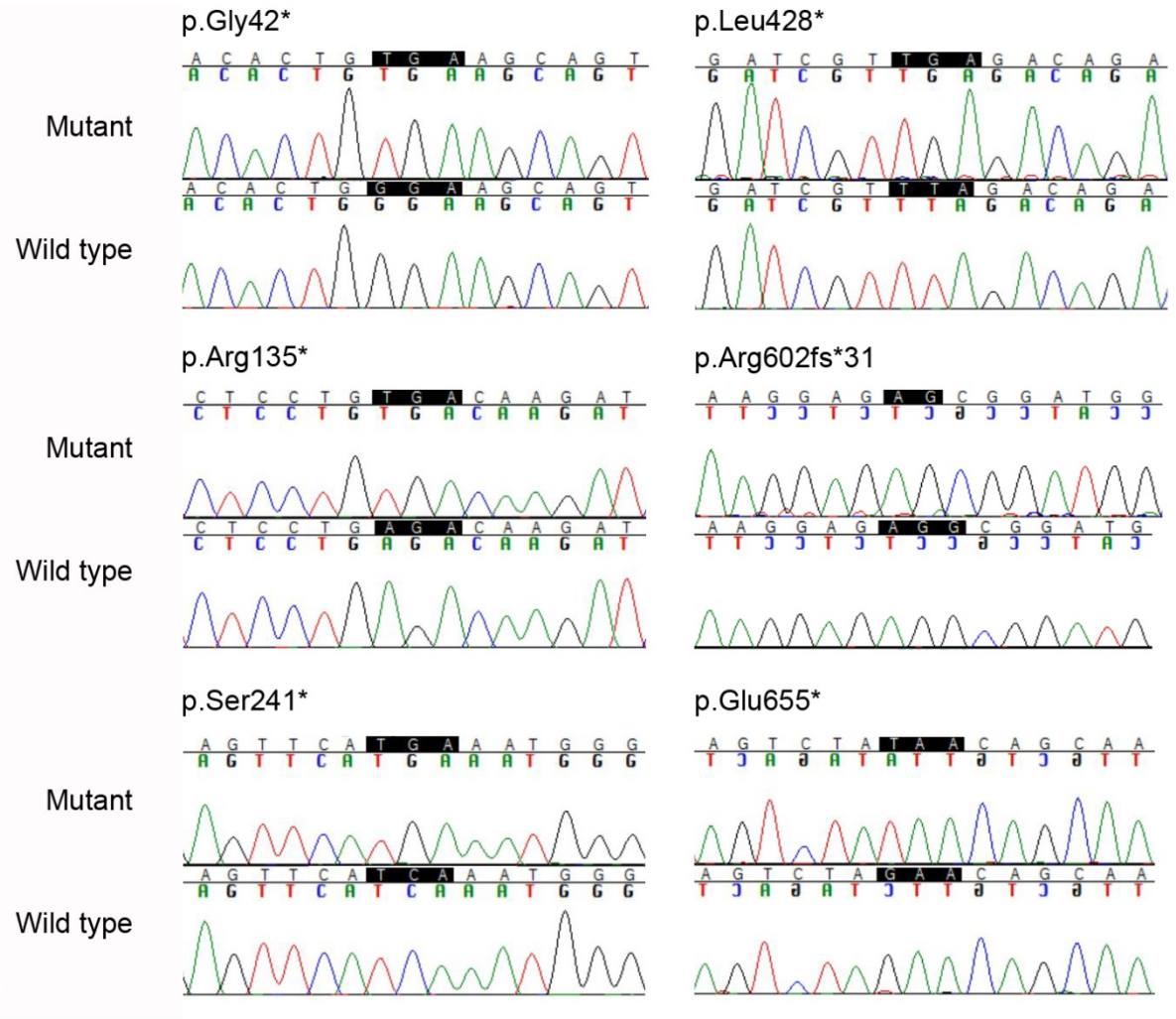


Figure 5-10

Site-directed mutagenesis of *TCF12* to introduce stop mutations.
Sequence chromatograms verifying successful introduction of six premature termination mutations into full length *TCF12* cDNA clones.

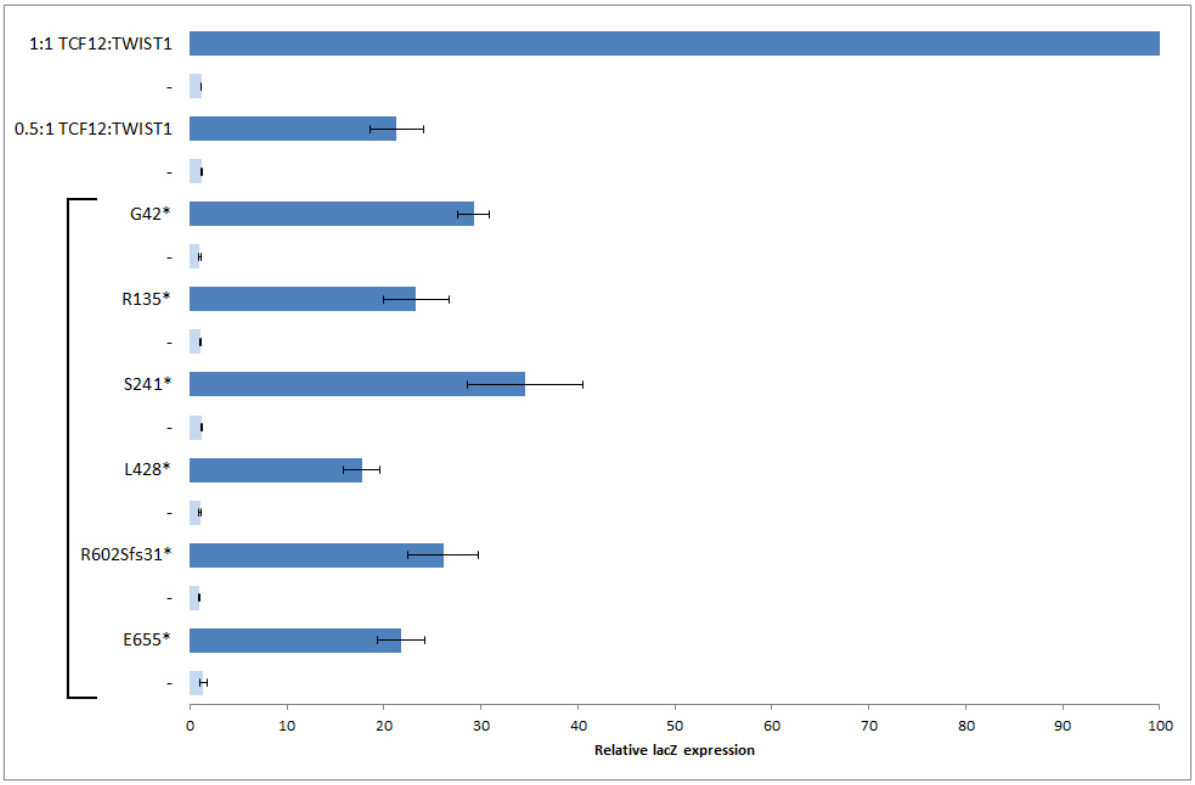


Figure 5-11

Assay for dominant negative activity of *TCF12* premature termination mutations.
Analysis of transactivation with TCF12 and TWIST1 wild-type or mutant proteins as shown, using reporters either including (dark blue) or omitting (-, light blue) three E-box DNA binding motifs. Results are shown as mean±SEM of 3 experiments and normalised to the output from transfecting a 1:1 ratio of both TCF12 and TWIST1 with the E-box reporter. The mutations enclosed by the bracket were all transfected at a ratio of 1:1:2 (mutant TCF12:wt TCF12:TWIST1).

5.3 Discussion

The clinical features associated with *TCF12*-related craniosynostosis share significant overlap with those of SCS, principal of which being unicoronal or bicoronal synostosis. This is illustrated by the fact that two of the *TWIST1*-negative SCS patients selected for whole-genome sequencing had *TCF12* mutations. In fact, in total seven probands with *TCF12* mutations had clinical features (including dysmorphic appearance of the face and external ears and minor limb anomalies) suggestive of SCS, which had been excluded by negative *TWIST1* mutation testing (Sharma et al. 2013). This phenotypic overlap is perhaps not surprising given the molecular interaction of the two proteins, supported by both the transactivation data detailed here and previous studies demonstrating that TCF12 and TWIST1 form functional heterodimers (Connerney et al. 2006).

Haploinsufficiency of *TCF12*, like *TWIST1* in SCS, is specifically associated with coronal synostosis indicating that the total quantity of TCF12-TWIST1 heterodimers is a critical factor in development of the coronal suture. Further supporting this, compound heterozygous mice (*Ella-Cre*; *Tcf12*$^{flox/+}$; *Twist1*$^{+/-}$) consistently showed complete fusion of the coronal suture and were more severely affected than either of the individual heterozygous mutant mice; the *Ella-Cre*; *Tcf12*$^{flox/+}$ animals did not display suture fusion by the age of three weeks, whereas *Twist1*$^{+/-}$ mice demonstrated variable coronal synostosis (Sharma et al. 2013). The importance of the TWIST1 heterodimer/homodimer balance in the developing coronal suture has been the subject of previous work investigating the role of *TWIST1* in craniosynostosis (Connerney et al. 2008, Connerney et al. 2006). As discussed in section 1.2 and 1.5.1, the coronal suture is established at the boundary of neural crest and cephalic mesoderm. TCF12-TWIST1 heterodimers may be involved in the specification of this boundary; previous work has shown that it is disrupted in *Twist*$^{+/-}$ mouse embryos (Merrill et al. 2006). By altering dimerization, mutant *TCF12* may affect the ability of *TWIST1* to function normally in mediating Eph-ephrin signaling. These heterodimers may also be important in inhibiting osteogenic differentiation in normal suture development. It has been reported that TWIST1 homodimers

increase *FGFR2* expression and thereby increase FGF responsiveness, conversely heterodimers decrease *FGFR2* expression (Connerney et al. 2008). The predicted expansion of *FGFR2* expression caused by a reduction in heterodimers would lead to increased BMP signaling, promoting osteoblast differentiation (Hayashi et al. 2007). Similarly, the ability of *TWIST1* to interact with *RUNX2* may also be affected, releasing inhibition of osteoblast differentiation (Bialek et al. 2004).

It is evident that the coronal suture is particularly sensitive to dosage of *TWIST1*, and fluctuations from the norm lead to disruption of proliferation and differentiation. The dosage is seemingly regulated by the homodimer/heterodimer balance, their differing activation potentials are illustrated by the transactivation assay whereby heterodimers are far more reactive (by roughly a factor of 10) than either *TWIST1* or *TCF12* homodimers. This reactivity was shown to be reduced by missense mutations at conserved residues in the bHLH domain (all three residues mutated in patients described here are conserved down to *C.elegans*) which presumably prevent the protein from being able to dimerize and therefore unable to bind DNA. Interestingly, one of the premature termination mutations fell 3´ to the bHLH domain (p.E655*), and another created a frameshift extending throughout the domain (R602Sfs31*), so either of these could conceivably exert a dominant negative effect if a truncated protein was produced that could still dimerize but was not functional. However, as demonstrated by further transactivation experiments these mutations had similar activity to all other stop mutations introduced, and that of a 50% reduction in the amount of wild type *TCF12*.

Although no deletions were identified in the MLPA screen of bicoronal synostosis patients in the original patient cohort, subsequent analysis by a collaborating group identified three intragenic deletions by whole genome sequencing that deleted exons 7-18, exon 19 and exon 20 respectively (in a cohort of 22 patients; J. Goos, I. Mathijssen, personal communication). This indicates that the genomic location is not protected in any way from such events.

Of note, *TCF12*-positive individuals seem to have a more benign surgical trajectory compared to *TWIST1*-positive individuals, Sharma et al. (2013) noted that they generally lacked requirement for

secondary surgery compared with most diagnostic categories and none was documented to have an interval onset of raised intracranial pressure after their initial surgical procedure. This provides important information for the clinical management of patients, and *TCF12* is now included in diagnostic screening of individuals presenting with bicoronal craniosynostosis. However, counseling of families is less clear-cut due to the substantial non-penetrance observed, which does not appear to be caused by relative under-representation of the wild type allele in the individuals looked at in this study. Another possibility is that the genetic background of the individual may have an effect, for example it has been reported that different haplotypes are associated with higher or lower gene expression (Cheung et al. 2010). With regards to *TCF12*, it may depend on the haplotype of the wild type allele as to whether an individual is clinically affected or not. Further investigation of clinically unaffected mutation positive individuals will need to be undertaken to confirm this hypothesis.

Several *TCF12* mutations may have been overlooked had other pathogenic variants not been discovered, particularly as the inheritance pattern was not immediately obvious due to the non-penetrance. Initially, the c.1035+3G>C was not flagged as potentially disease causing during the bioinformatic calling stage as it was outside the canonical GT splice donor site. Similarly, the c.1468-20T>A variant was identified only after a specific text search. As well as leading to more stringent parameters being implemented, these findings highlight the fact that a substantial proportion of disease causing mutations sought by whole exome or genome sequencing approaches may go unnoticed as at first-look they appear benign, especially if they don't fit in with the assumed inheritance pattern. Being able analyse data in the context of other concurrent sequencing projects was beneficial in helping to identify less obviously pathogenic mutations for further scrutiny.

CHAPTER 6

A NEW CRANIOFACIAL PHENOTYPE ASSOCIATED WITH LOCALISED MUTATIONS IN *TWIST1*

Chapter 6: A new craniofacial phenotype associated with localised mutations in *TWIST1*

6.1 Introduction

As outlined in section 1.3.1, SCS is caused by haploinsufficiency of *TWIST1*; both whole gene deletions and point mutations have been reported. Recently however, evidence has come to light that mutations at a specific residue of *TWIST1*, p.Glu117, lead to a radically different craniofacial phenotype, possibly through a dominant-negative mechanism. Of note, mutations at residue p.Glu117 have never been reported in association with SCS, although point mutations of the flanking residues have (Fig. 6-1A). The glutamic acid at position 117 falls within the basic DNA-binding domain and is highly conserved in all bHLH proteins (Fig. 6-1B). As discussed in section 1.3.3, the glutamic acid at this position in the DNA binding domain forms specific contacts with the cytosine and adenine of the CANNTG E-box motif, creating a bridge to the DNA backbone.

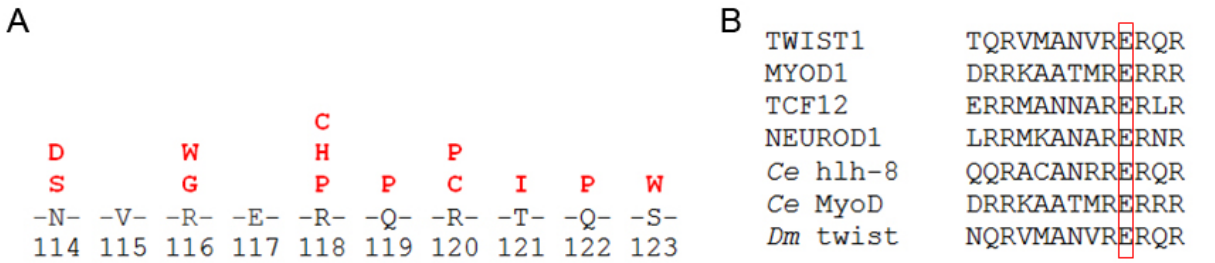


Figure 6-1

Substitutions and conservation of the TWIST1 basic domain.

A. Partial amino acid sequence of human TWIST1 bHLH domain, single amino acid substitutions that have been previously reported in association with SCS are indicated by the single-letter code of the substitution, above the sequence, in red.
B. Alignment of basic DNA binding domains from various bHLH proteins; the glutamic acid (outlined in red) is extremely highly conserved throughout human homologues, and down to *C.elegans* (*Ce*) and *D.melanogaster* (*Dm*) homologues.

Interestingly, previous experiments using *C. elegans* demonstrated a potential dominant-negative Twist mutant. The *C. elegans* Twist protein is encoded by *hlh-8*; animals homozygous for an *hlh-8* null allele were viable and fertile, but demonstrated mesodermal defects such as a lack of nonstriated muscle function necessary for egg-laying and defecation. These features were not observed in heterozygotes, which were indistinguishable from wild-type animals. This indicates that phenotypes associated with a null *hlh-8* allele show recessive inheritance. However, animals heterozygous for a point mutation (p.Glu29Lys; equivalent to human p.Glu117) were egg-laying defective, but not visibly constipated leading the authors to describe this allele as semi-dominant (Corsi et al. 2002). Furthermore, in an *in vitro* DNA-binding assay the authors demonstrated that the p.Glu29Lys mutant was still able to homodimerize and heterodimerize, and bind DNA, whereas Twist containing substitutions equivalent to the human SCS mutations p.Arg116Trp, p.Arg118His and p.Gln119Pro could not.

6.2 Case presentations

6.2.1 Case 1

Parent/child trio-based exome sequencing was carried out to try and identify the genetic basis of the phenotype observed in the proband (Fig. 6-3A and B), who had clinically unaffected parents. He was a 3-year old male with hypertelorism, a wide anterior fontanelle, upper eyelid colobomas, deficient bony orbits with pseudoproptosis, small low-set dysplastic cupped ears, syndactyly of fingers, bilateral talipes, bilateral undescended testes, imperforate anus and hypertrichosis. The exome sequencing revealed a *de novo* heterozygous c.350A>T, p.Glu117Val substitution in *TWIST1* (S. Twigg; verified by dideoxy sequencing, Fig. 6-2). As the patient did not resemble typical SCS, the remaining exome was analysed but no convincing pathogenic changes, other than p.Glu117Val, were identified.

6.2.2 Case 2

Presentation of Case 1 at the UK Dysmorphology Group prompted *TWIST1* sequencing of a female subject with similar, but more severe, features (Fig. 6-3C and D). Remarkably, this revealed a

different *de novo* heterozygous mutation at the identical nucleotide; c.350A>G, p.Glu117Gly (verified by dideoxy sequencing at the Genetics Laboratories, Churchill Hospital; Fig. 6-2). The female patient had bilateral upper eyelid colobomas, a widely patent anterior fontanelle and metopic suture extending down to a very broad nasal bridge with a broad and flattened nasal tip and a low columella. A midline cleft palate was present and choanal atresia was managed with a tracheostomy. The pinnae were small, thickened and dysplastic. Skin syndactyly of fingers and toes (2-4) and camptodactyly of the fingers caused limitation of hand movement.

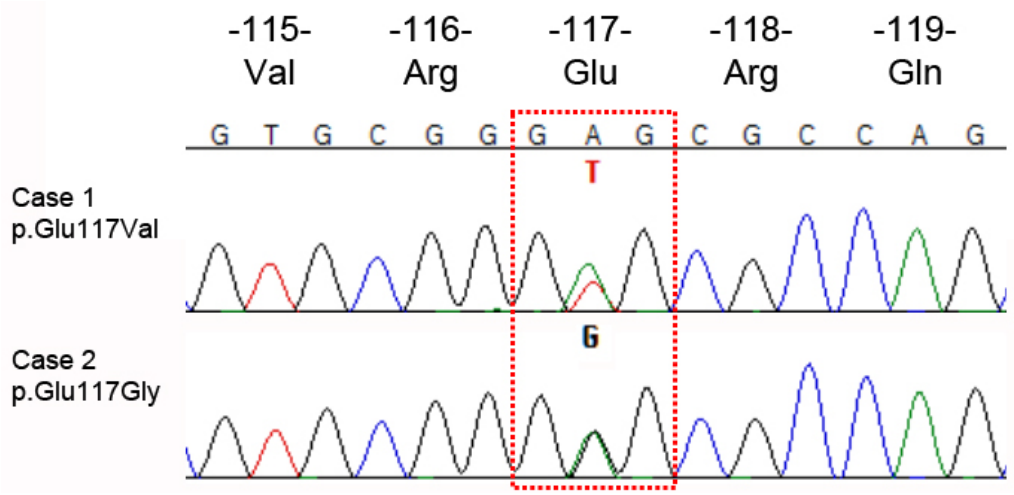


Figure 6-2

Sequence chromatograms confirming *TWIST1* mutations.
Upper panel shows c.350A>T (p.Glu117Val) identified in Case 1, lower panel shows c.350A>G (p.Glu117Gly) identified in Case 2.

Figure 6-3

Facial appearance of patients with TWIST1 p.Glu117 substitutions.
A, B. Male proband with *de novo TWIST1* c.350A>T variant, leading to the substitution p.Glu117Val. Note hypertelorism, upper eyelid colobomas, deficient bony orbits with pseudoproptosis and small low-set dysplastic cupped ears.
C, D. More severely affected female patient with upper eyelid colobomas, a widely patent anterior fontanelle and metopic suture, a very broad nasal bridge with a broad and flattened nasal tip and a low columella. This individual has a *de novo TWIST1* c.350A>G variant, leading to the substitution p.Glu117Gly.

6.3 Transactivation analysis of *TWIST1* missense mutations

To explore the effect of substitutions at p.Glu117 further, the transactivation assay developed to analyse *TCF12* missense mutations (section 5.2.6-7) was used, this time combining wild type TCF12 with mutant TWIST1. Site-directed mutagenesis was used to introduce four different point mutations previously reported as causing typical SCS (p.Arg116Trp, p.Gln119Pro, p.Arg120Cys, p.Thr121Ile), as well as the two mutations identified in the patients described here (p.Glu117Val and p.Glu117Gly), and the equivalent of the *C.elegans* p.Gly29Lys mutant (p.Glu117Lys). All mutations were sequence verified (Fig. 6-4), primer sequences are listed in section 2.7.1.

To first investigate the ability of TWIST1 homodimers to activate transcription, the mutant proteins alone, and in combination with wild type TWIST1, were transfected (Fig. 6-5A and B). As in previous experiments, expression was normalised to the output from transfecting both wild type TCF12 and TWIST1 together. This clearly showed that mutant proteins alone, or in combination with wild type TWIST1, could not activate transcription above levels seen when just wild type TWIST1 was transfected.

Next, the ability of the mutant proteins to heterodimerize with TCF12 was assayed. Mutant TWIST1 was transfected at a 1:1 ratio with wild type TCF12, and normalised as above (Fig. 6-5C). Only the p.Glu117Gly substitution was able to activate transcription in combination with TCF12, although at 55% of the level of wild type TWIST1.

To assess the potential of the mutant proteins to act in a dominant negative manner, cells were transfected with a 2:1:1 ratio of wild type TCF12:wild type TWIST1:mutant TWIST1 (mimicking the heterozygous state; Fig. 6-5D). In this situation, the presence of mutant TWIST1 reduced the output by approximately 50%, except for p.Glu117Lys which demonstrated a much more reduced level of *lacZ* expression.

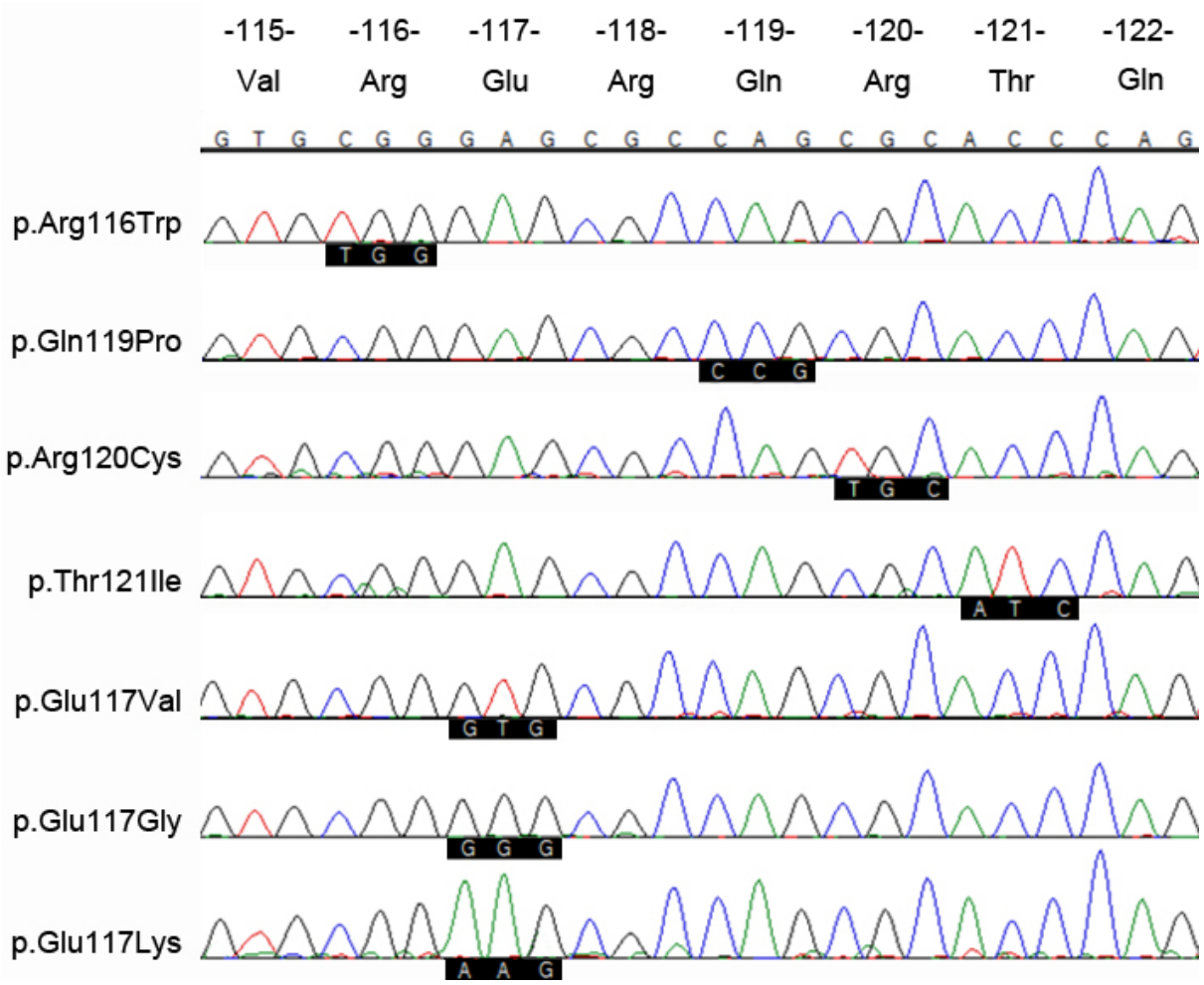
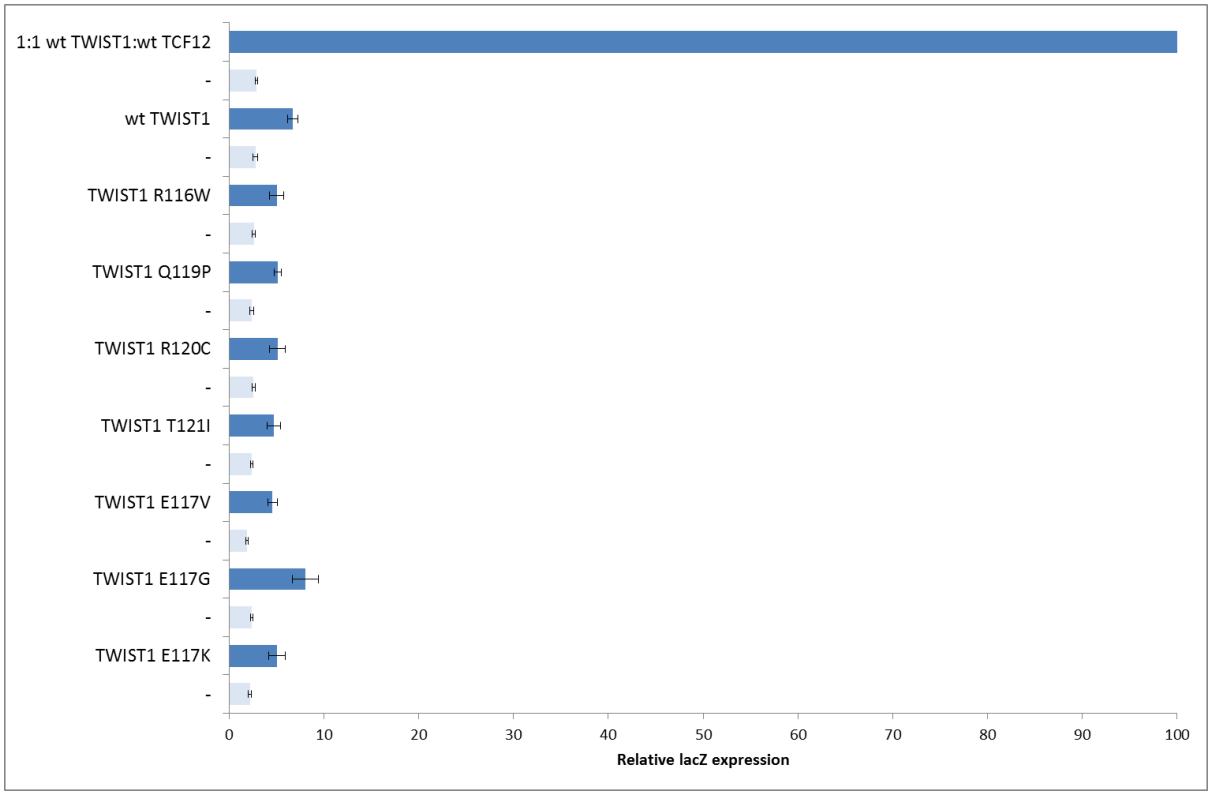


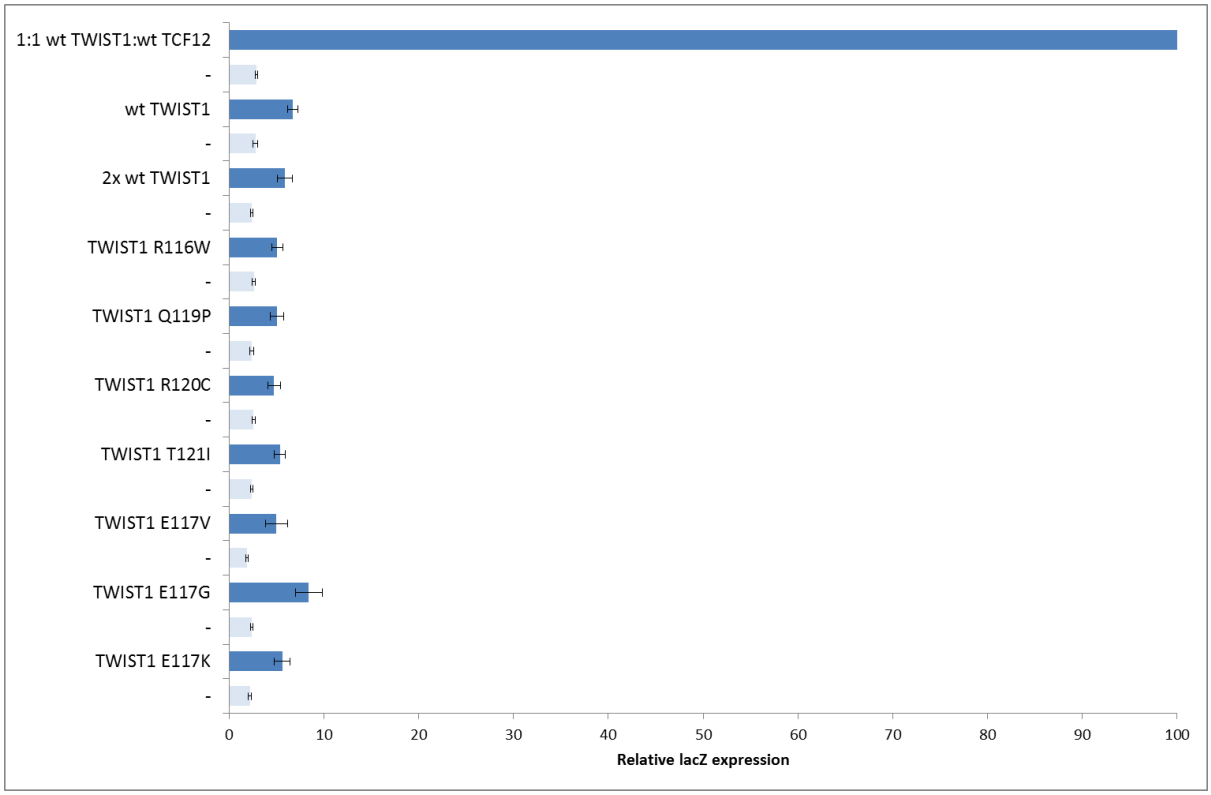
Figure 6-4

Site directed mutagenesis of *TWIST1*.
Sequence chromatograms verifying successful introduction of missense mutations to full length *TWIST1* cDNA clones; p.Arg116Trp, p.Gln119Pro, p.Arg120Cys and p.Thr121Ile have all been previously reported in association with SCS. Mutations affecting p.Glu117; p.Glu117Val and p.Glu117Gly identified in patients with severe non-SCS phenotype, and p.Glu117Lys recreating the dominant-negative mutant identified in *C.elegans*.
The wild type *TWIST1* sequence is shown across the top, with the amino acid sequence indicated above (residues 115-120 fall within the basic DNA binding domain, residues 121-122 are in helix 1).

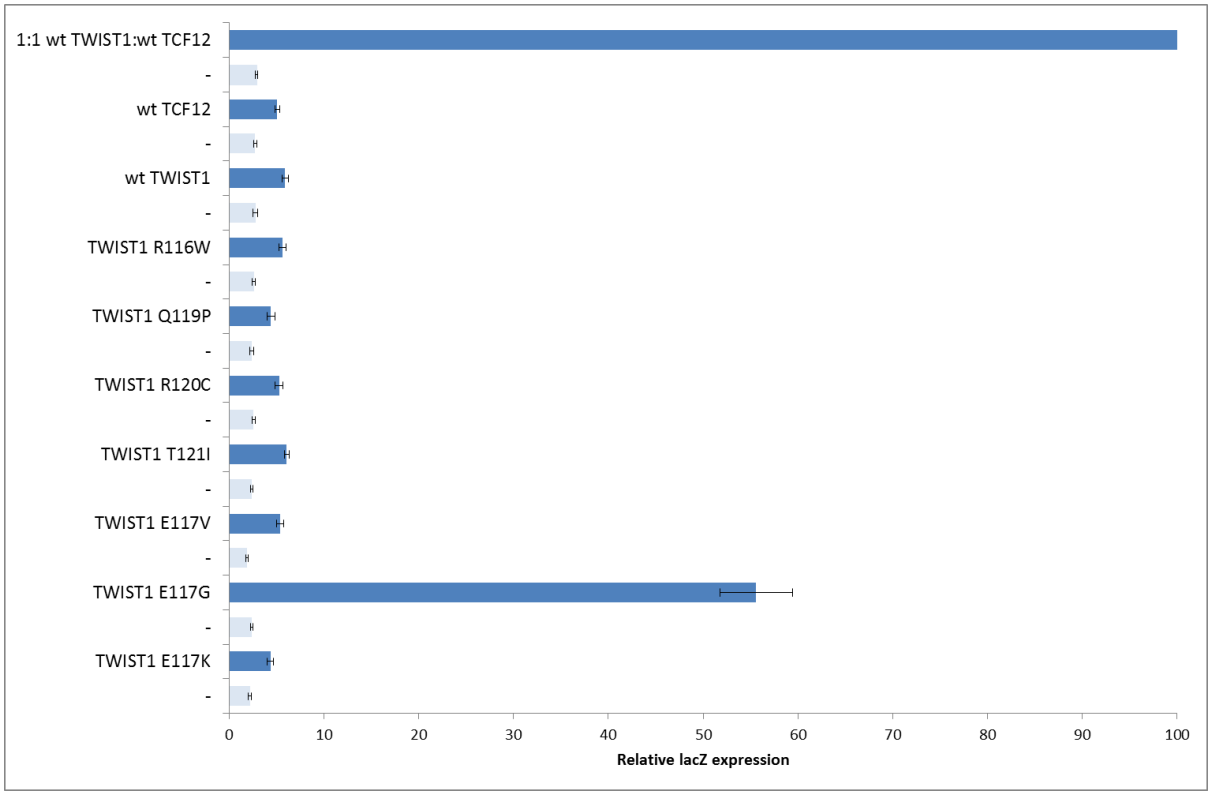
A


1:1 wt TWIST1:wt TCF12
-
wt TWIST1
-
TWIST1 R116W
-
TWIST1 Q119P
-
TWIST1 R120C
-
TWIST1 T121I
-
TWIST1 E117V
-
TWIST1 E117G
-
TWIST1 E117K
-
0
10
20
30
40
50
60
70
80
90
100
Relative lacZ expression

B


1:1 wt TWIST1:wt TCF12
-
wt TWIST1
-
2x wt TWIST1
-
TWIST1 R116W
-
TWIST1 Q119P
-
TWIST1 R120C
-
TWIST1 T121I
-
TWIST1 E117V
-
TWIST1 E117G
-
TWIST1 E117K
-
0
10
20
30
40
50
60
70
80
90
100
Relative lacZ expression

C


1:1 wt TWIST1:wt TCF12
-
wt TCF12
-
wt TWIST1
-
TWIST1 R116W
-
TWIST1 Q119P
-
TWIST1 R120C
-
TWIST1 T121I
-
TWIST1 E117V
-
TWIST1 E117G
-
TWIST1 E117K
-
0
10
20
30
40
50
60
70
80
90
100
Relative lacZ expression

D

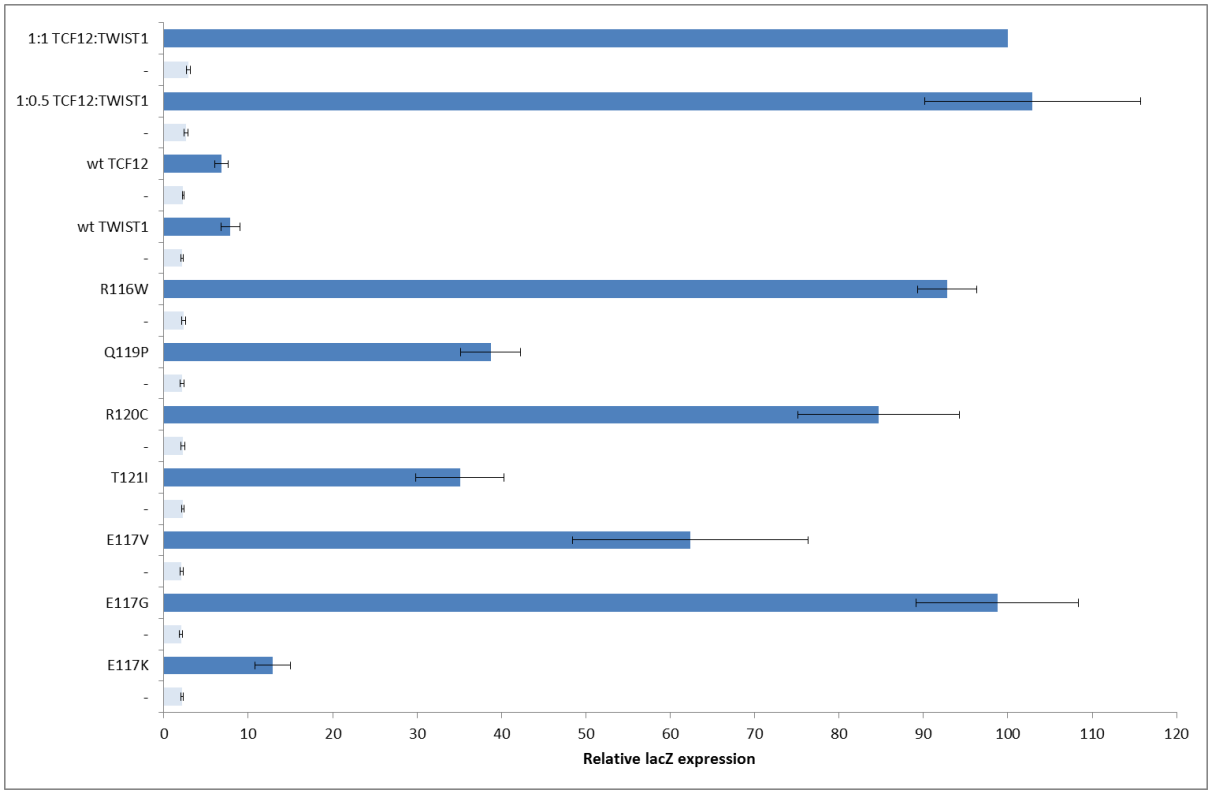

1:1 TCF12:TWIST1
-
1:0.5 TCF12:TWIST1
-
wt TCF12
-
wt TWIST1
-
R116W
-
Q119P
-
R120C
-
T121I
-
E117V
-
E117G
-
E117K
-
0
10
20
30
40
50
60
70
80
90
100
110
120
Relative lacZ expression

Figure 6-5

Transactivation analysis of TWIST1 mutations.
A. Output from transfecting wild type or mutant TWIST1 alone.
B. Output from transfecting a 1:1 ratio of wild type or mutant TWIST1 with wild type TWIST1. Mutant proteins alone, or in combination with wt TWIST1, could not activate transcription above levels seen when just wt TWIST1 was transfected.
C. Mutant TWIST1 was transfected at a 1:1 ratio with wild type TCF12. Only the p.Glu117Gly substitution was able to activate transcription in combination with TCF12, at 55% of the level of wild type TWIST1. Wild type TWIST1 and TCF12 alone were also transfected for comparison.
D. To mimic the heterozygous state, mutant proteins were transfected at a 2:1:1 ratio with wild type TCF12, wild type TWIST1 and mutant TWIST1. The presence of mutant TWIST1 reduced the output by approximately 50%, except for p.Glu117Lys which demonstrated a much more reduced level of *lacZ* expression. Wild type TWIST1 and TCF12 alone were also transfected for comparison.
Output from wild type TWIST1 with wild type TCF12 was set at 100 and other results normalised to this. Outputs from negative control expression vectors with no E-box motifs are represented by pale blue bars. All results are shown as mean±SEM of 3 experiments.

6.4 Discussion

It would appear that there is an exclusive genotype/phenotype correlation associated with mutations at p.Glu117 of TWIST1; no mutations at this position have ever been reported as causing SCS, and mutations of p.Glu117 are associated with an entirely different, novel phenotype. These findings indicate a different pathophysiological mechanism from the haploinsufficient loss-of-function of *TWIST1* that leads to SCS. Mutations at p.Glu117 could exert their effect through either a gain-of-function (for example by widening the repertoire of DNA binding sites, leading to ectopic activity) or dominant negative mechanism (such as sequestering wild type TWIST1 in non-productive heterodimers on DNA). As I had already established a transactivation system to assay bHLH dimers (section 5.2.6-7), I decided to use this to see if it would provide any insight into the activity of the p.Glu117 substitutions.

Only p.Glu117Gly could activate expression when transfected together with wild type TCF12 (Fig. 6-5C), both the p.Glu117Val and p.Glu117Lys substitutions were indistinguishable from the SCS substitutions, and generated a similar output to that of either wild type TWIST1 or TCF12 alone. In

terms of dominant negative activity, the result seen with p.Glu117Lys (Fig. 6-5D), which strongly reduced *lacZ* expression, seems to corroborate the findings of Corsi et al (2002) and further indicates that the choice of amino acid at position 117 has specific effects on function. Both p.Glu117Gly and p.Glu117Val produced outputs that were within the range generated by the SCS substitutions. Interestingly, distinct phenotypic and functional consequences have been observed from substitutions of the equivalent glutamic acid residue in other bHLH proteins; E47, NEUROD1 and TWIST2 (summarised in Table 6-1).

In a recent study, Boisson et al (2013) described four unrelated patients, two male and two female, with agammaglobulinemia (lack of gamma globulin in the blood plasma) and lacking B cell antigen receptors (BCRs). This atypical phenotype was investigated by parent/child trio-based exome sequencing, which led to the identification of the *de novo* variant c.1663G>A in *TCF3* (*E2A*), in an exon specific to the E47 protein isoform. The identical heterozygous mutation, causing the substitution p.Glu555Lys, was present in the three other patients and also appeared to be *de novo* in each of them. It was not present in 40 patients with B cell defects but without the BCR deficiency, or 1500 control samples. The authors demonstrated by Western blot that the mutant E47 localised normally to the nucleus, and by an electrophoretic mobility shift assay (EMSA) that it showed a strong dominant-negative effect when bound to DNA as a homodimer with wild type E47. Moreover, E47 containing p.Glu555Lys was able to bind DNA as a heterodimer with MyoD (Boisson et al. 2013).

Of note, acquired heterozygous loss of *TCF3* is frequently associated with Sézary syndrome, which is a subtype of T cell lymphoma (Steininger et al. 2011). However, various somatically acquired mutations affecting the bHLH domain of E47 were identified in individuals with Burkitt's lymphoma, including the variant p.Glu555Gln (Schmitz et al. 2012).

Furthermore, the substitution p.Glu110Lys in NEUROD1, equivalent to TWIST1 p.Glu117, was reported in 12 family members with maturity-onset diabetes of the young (MODY; Kristinsson et al. 2001). Mutations in NEUROD1 had previously been identified in patients with type II diabetes;

p.Arg111Leu (immediately adjacent to the p.Glu110 residue) and p.His206Profs36* (Malecki et al. 1999). Interestingly, the missense variant p.Arg111Leu was associated with a more typical phenotype than that of the premature truncating mutation, which was more reminiscent of MODY. NEUROD1 containing p.His206Profs36* was shown to be able to bind an insulin E-box element with high affinity, whereas the protein with the p.Arg111Leu substitution could not (Malecki et al. 1999). These results seem indicative of dominant-negative activity of the p.Glu110Lys and p.His206Profs36* variants, which were associated with a severe MODY phenotype requiring early intervention (Kristinsson et al. 2001).

Ablepharon macrostomia and Barber-Say syndromes (AMS and BSS) are both forms of congenital ectodermal dysplasia with overlapping features; macrostomia, hypertelorism, abnormal nose and ears, hypoplasia of nipples and redundancy of the skin. However, AMS is distinct in that features also include ablepharon, no hair at birth and late development of sparse, thin hair. Characteristic features of BSS are bilateral ectropion and generalised hypertrichosis from birth (Mazzanti et al. 1998). Although it had been speculated that these syndromes were caused by mutations in the same gene, it was only proven very recently that this was indeed the case. Marchegiani et al (2015) found that both syndromes were caused by heterozygous mutations in *TWIST2*, and that there was a clear genotype/phenotype correlation associated with particular amino acid substitutions. Exome sequencing of six family members with AMS revealed the *TWIST2* variant c.223G>A, causing the substitution p.Glu75Lys (the equivalent residue to TWIST1 p.Glu117). Subsequent targeted sequencing of *TWIST2* identified the identical variant in 10 AMS patients from seven families, whereas of 12 BSS patients, nine had p.Glu75Gln substitutions, two had p.Glu75Ala substitutions and one had the duplication p.Gln77_Arg78dup (Marchegiani et al. 2015). Owing to their position in the basic DNA binding domain, it was suspected that the *TWIST2* mutations would alter DNA binding activity. To further investigate, ChIP coupled with high-throughput sequencing (ChIP-seq) was carried out on HeLa cells overexpressing recombinant FLAG-HA-tagged TWIST2 proteins; 630 binding peaks associated with wild type TWIST2 were identified (using the consensus binding motif –CATCTGG-).

This number was reduced by 60-80% in the presence of the mutant proteins, furthermore, the mutants bound to many sites not shared with wild type TWIST2 (Marchegiani et al. 2015). To assay the effects *in vivo*, human TWIST2 containing p.Glu75Lys or p.Glu75Gln was injected into zebrafish embryos; the resulting fish displayed more severe developmental defects than those caused by overexpression of wild type TWIST2. In particular the mutant protein containing p.Glu75Lys produced the most severely affected animals (Marchegiani et al. 2015). These results are suggestive of a dominant-negative mechanism, as homozygous loss-of-function mutations of *TWIST2* lead to the overlapping but less severe ectodermal dysplasia Setleis syndrome, however as the ChIP-seq indicated a broadened range of DNA binding sites, a gain-of-function effect cannot be ruled out.

Although the results from the transactivation assays carried out with TWIST1 were not clear-cut; there does seem to be evidence that a glutamic acid to lysine substitution at this critical position in the basic DNA binding domain of bHLH proteins has a dominant negative effect. It is reasonable to assume that this reversal of charge, from negative to positive, would have a strong effect on the function of the protein. The effects of any mutation at this position may be dependent on individual properties of the amino acid substituted, hence the differing results observed in TWIST1 (and other bHLH proteins). This could be due to the retention of E-box binding affinity, or a reduction in specificity with increased illegitimate DNA binding.

To explore these possibilities more fully, an EMSA could be designed with differing DNA probes containing variations of the E-box motif to determine the binding capabilities of dimers containing the various p.Glu117 substitutions, using a TWIST1 antibody to confirm presence of the target protein. Furthermore a reporter system with non-consensus E-box motifs could show if the mutations led to gain-of-function properties, demonstrating if the dimers could still drive expression.

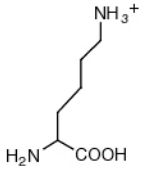
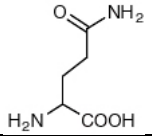
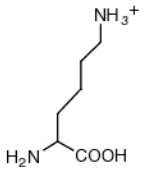
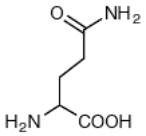
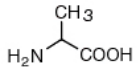
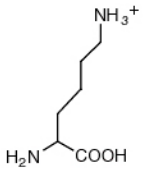
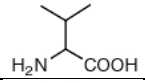
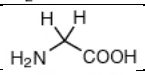
A recent study looked in detail into TWIST1 DNA binding, using ChIP-seq for TWIST1-binding DNA elements in human mammary epithelial cells (Chang et al. 2015). The canonical E-box motif made up over 75% of TWIST1-binding peaks, and intriguingly ~40% of the total TWIST1-binding peaks were at

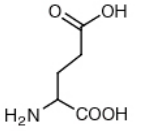
a double E-box motif; two E-boxes separated by five nucleotides. This double motif was not recognised by other bHLH proteins, and the authors showed in an EMSA that TWIST1/E47 complexes showed a higher affinity toward the double E-box motif than E47 alone. Modelling based on the crystal structure of a NeuroD1/E47 heterodimer indicated that two TWIST1/E47 heterodimers could bind to the double E-box together to form a stable ternary complex; the five nucleotide spacing between the two E-boxes allowing a full turn of the DNA double helix between the two E-boxes, and aligning the two heterodimers in the same orientation allowing stabilizing protein-protein interactions (Chang et al. 2015). It is possible that dependent on the amino acid substitution in the basic DNA-binding domain, the structure of the protein is sufficiently altered such that preferential binding to double E-box motifs cannot occur due to steric hindrance.

In conclusion, on the basis of the similarity of the two patients' clinical features, which are not typical for SCS, it seems that heterozygous substitutions at p.Glu117 of TWIST1 are associated with a unique, severe phenotype. In SCS haploinsufficency of *TWIST1* in the coronal suture leads to craniosynostosis, whereas in the cases presented here, a different mechanism may act at an earlier stage of development to disrupt facial tissues derived from the cranial neural crest. As discussed in section 1.3.2, *Twist1*-null mice exhibit failure of cephalic neural tube closure and die embryonically, demonstrating a role for Twist1 in neural crest cell patterning and development. There are also some similarities between the features in the patients with the p.Glu117 mutations and those seen in Treacher Collins syndrome, which results from defective neural crest cell development during head and facial morphogenesis, and involves hypoplasia of the facial bones, particularly the maxilla, mandible and zygomatic complex, microtia associated with conductive hearing loss, cleft palate, downslanting palpebral fissures and lower eyelid colobomas (Trainor 2010).

Table 6-1

Summary of substitutions occurring in bHLH proteins at the equivalent residue to TWIST1 p.Glu117.

Protein	Substitution	Phenotype	Substituted amino acid structure	Refs
E47	p.Glu555Lys	Agammaglobulin + lack of BCRs		(Boisson et al. 2013)
E47	p.Glu555Gln	Burkitt's lymphoma		(Schmitz et al. 2012)
TWIST2	p.Glu75Lys	AMS		(Marchegiani et al. 2015)
TWIST2	p.Glu75Gln	BSS		(Marchegiani et al. 2015)
TWIST2	p.Glu75Ala	BSS		(Marchegiani et al. 2015)
NEUROD1	p.Glu117Lys	MODY		(Kristinsson et al. 2001)
TWIST1	p.Glu117Val	Non-SCS severe phenotype		This study.
TWIST1	p.Glu117Gly	Non-SCS severe phenotype		This study.

Structure of glutamic acid for comparison: 

CHAPTER 7

TARGETED SEQUENCING OF *TWIST1*

Chapter 7: Targeted sequencing of *TWIST1*

7.1 Introduction

A targeted sequencing approach was used to further test the hypothesis that regulatory mutations lead to SCS. As well as targeting the 2.4 Mb region around *TWIST1* (chr7: 17385767-19735085), the *TCF12* locus was included to try and identify any intronic variants that may be pathogenic by affecting splicing, or indeed any that were missed by conventional sequencing. Similarly, the recently identified craniosynostosis gene *ERF* was also covered (Twigg, Vorgia, et al. 2013).

Massively parallel sequencing of the selected genomic regions was carried out on a total of 160 DNA samples from: 34 bicoronal, 111 unicoronal, 11 multisuture including bicoronal, and 4 multisuture including unicoronal synostosis patients (experimental details in section 2.8). Sequencing was carried out in two separate multiplexes of 80 samples each, depth of coverage is summarised in Table 7-1.

Table 7-1

Read depth statistics from targeted sequencing.

		ERF	***TCF12***	***TWIST1***
Round 1 (80 samples)	Min.	23	34	34
	Median	121	176	180
	Mean	133	190	198
	Max.	357	464	478
Round 2 (80 samples)	Min.	34	51	52
	Median	79	121	128
	Mean	101	150	157
	Max.	415	534	536

7.2 Analysis of coding variants

Initially the dataset was analysed for variants in the coding regions of the included genes, which led to the identification of six pathogenic mutations (Table 7-2). Two of these mutations (*TWIST1* c.193G>T (p.Glu65*) and *TCF12* c.1098dupA (p.Pro367Thrfs22*)) were identified by the Genetics Laboratory (Churchill hospital, Oxford) during the course of this experiment, but a further two (*TWIST1* c.374A>G (p.Asn125Ser) and c.481C>T (p.Gln161*)) had originally been missed during diagnostic screening. A further *TCF12* mutation (c.1115-2A>G), as well as an *ERF* mutation (c.1201_1202delAA; p.Lys401Glufs10*) were also identified.

The patient with the *TCF12* c.1115-2A>G variant (5939) had a much more severe phenotype than is typically associated with *TCF12* mutations (detailed in Table 7-2). The female proband inherited the variant from her father who was clinically unaffected, further demonstrating the non-penetrance observed with *TCF12*-associated craniosynostosis. This mutation has been reported previously, where both the proband and her mother had bicoronal synostosis without additional syndromic features (Sharma et al. 2013). With this in mind, the remaining sequence data for individual 5939 was carefully checked for any additional variants that could be contributing to the phenotype. The only other potentially damaging variant was the intronic transversion *TCF12* c.1114+372T>A, which introduced a putative splice acceptor. However, this variant was also inherited from the unaffected father (not shown).

The *ERF* variant c.1201_1202delAA (p.Lys401Glufs10*) was also interesting as the patient (5243) had only unicoronal synostosis and the majority of previously reported *ERF* mutations are associated with multi-suture synostosis, in fact the identical variant was previously identified in an individual with pan-synostosis. The only reported single suture fusions have twice involved the sagittal, and once the lambdoid sutures (Twigg, Vorgia, et al. 2013).

Table 7-2

Summary of coding variants identified by targeted sequencing.

Individual	Gene	Variant	Phenotype	Additional information
5896	*TWIST1*	c.193G>T p.Glu65*	Bicoronal synostosis, slightly deviated thumbs & halluces, slight 2/3 finger syndactyly, ext ear crus.	Also identified by diagnostic lab during the course of this experiment.
2505	*TWIST1*	c.374A>G p.Asn125Ser	Left unicoronal synostosis.	Inherited from father. Originally missed by diagnostic lab.
5563	*TWIST1*	c.481C>T p.Gln161*	Right unicoronal synostosis.	Originally missed by diagnostic lab.
5538	*TCF12*	c.892C>T p.His298Asp	Bicoronal synostosis, developmental delay, short stature.	Predicted to be benign, in ExAC (freq 0.00001650).
6988	*TCF12*	c.1098dupA p.Pro367Thrfs22*	Bicoronal synostosis.	Also identified by diagnostic lab during the course of this experiment.
5939	*TCF12*	c.1115-2A>G (exon 14-2)	Bicoronal, sagittal, bilambdoid synostosis. High narrow palate, slender fingers, broad thumbs, 5th finger clinodactyly, laterally deviated 2nd toes, long narrow feet, broad great toes. Prune belly.	Inherited from clinically unaffected father. Mutation also reported in unrelated family in Sharma et al (2013).
5243	*ERF*	c.1201_1202delAA p.Lys401Glufs10*	Right unicoronal synostosis. Large eyes, tiny nose and mouth. Large broad great toes, no syndactyly. Poor motor development.	Inherited from father who was noted to have a prognathic mandible. Mutation also reported in unrelated family in Twigg et al (2013).
5285	*ERF*	c.1207G>A p.Gly403Ser	Bicoronal synostosis.	Predicted to be benign, in ExAC (freq 0.0001323).

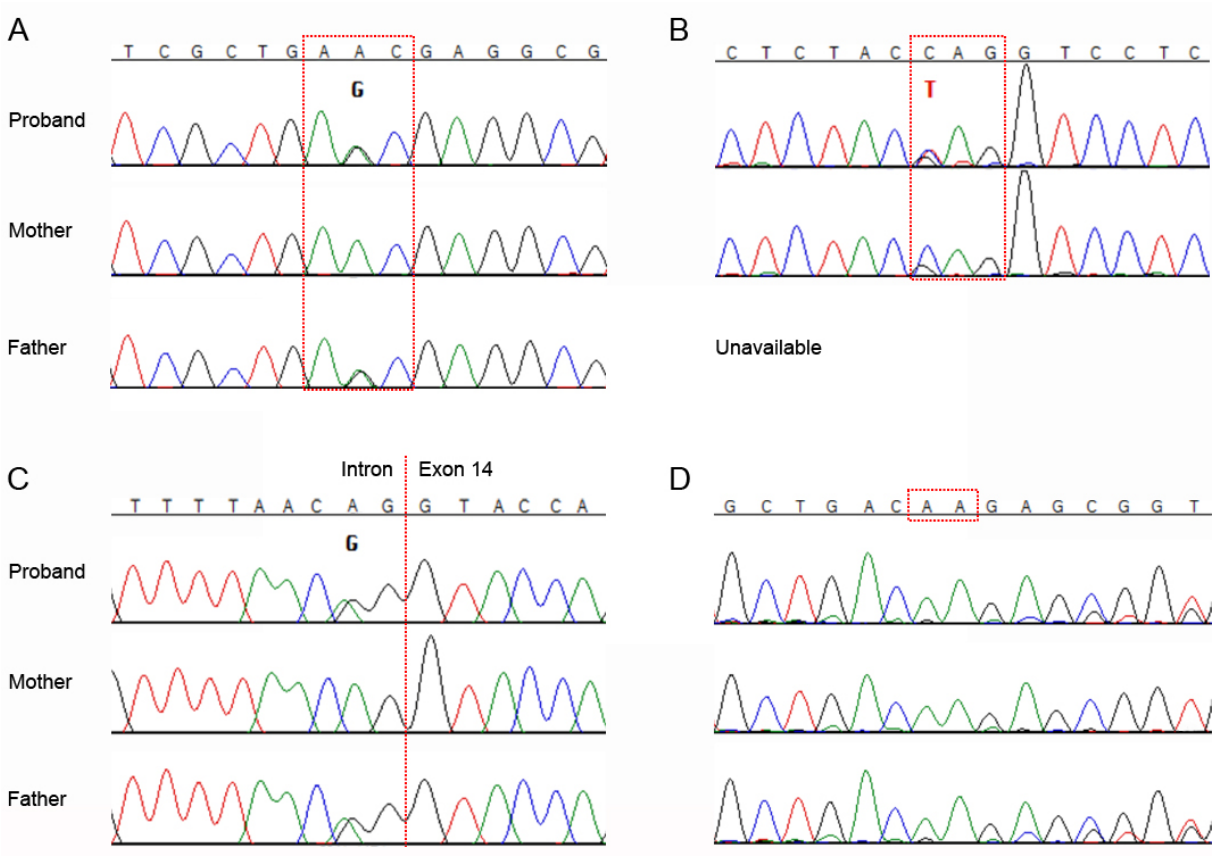


Figure 7-1

Confirmation of pathogenic mutations identified by targeted sequencing.
A. *TWIST1* c.374A>G (p.Asn125Ser) in patient 2505, inherited from her father.
B. *TWIST1* c.481C>T (p.Gln161*) in patient 5563, DNA from the father was unavailable for testing.
C. *TCF12* c.1115-2A>G in patient 5939, inherited from her clinically unaffected father.
D. *ERF* c.1201_1202delAA (p.Lys401Glufs10*) in patient 5243, inherited from her father who had a mild phenotype without craniosynostosis.

7.3 Analysis of splice-altering variants

To assess possible splicing effects, the *in silico* prediction tool MaxEntScan was used (Yeo and Burge 2004). This method, which scores acceptor and donor sites, has been shown to be one of the best in predicting splice-altering variations (Jian, Boerwinkle, and Liu 2014). Scores were generated for reference and variant sequences and then compared, all variants with a score change of >1 were looked into further (listed in Table 7-3). The only likely pathogenic result was the *TCF12* c.1115-2A>G mutation that had already been identified in patient 5939.

Table 7-3

Summary of results generated by MaxEntScan with a score change >1.

Gene	Position	Variant	MaxEnt ref score	MaxEnt var score	Additional information
SNX13	chr7:17908130	T>A	9.19	10.72	rs2389858
HDAC9	chr7:18875084	C>G	12.21	9.99	rs41273078, 1000g (freq 0.01)
TWIST1	chr7:19155769	delAG	11.38	10.24	rs138390369
TCF12	chr15:57511641	T>A	-20	-21.4	rs188382628
TCF12	chr15:57523338	T>C	9.91	8.71	rs117914532, 1000g (freq 0.01)
TCF12	chr15:57543546	A>G	9.41	1.45	Patient 5939 described in section 7.2

7.3.1 A potentially pathogenic mutation upstream of *TCF12*

The 3' and 5' UTRs, as well as the upstream and downstream regions of *TWIST1*, *TCF12* and *ERF* were also manually analysed, which led to the identification of a C>G variant that occurred upstream of *TCF12* (chr15:57212041; patient 2490). The position of the variant was -59 from the start of the first coding exon (-71 from the ATG), and was predicted to introduce a cryptic acceptor with a score of 0.99, whereas the wild type acceptor had a score of 0.90 (Neural Network). The variant was confirmed by dideoxy sequencing and was not present in the proband's unaffected mother (Fig. 7-2A), unfortunately no DNA from the father (also unaffected) was available to determine if the mutation had occurred *de novo*. cDNA was generated from a fibroblast cell line from the patient and assayed for use of the cryptic acceptor; the amplified cDNA generated an extra, slightly larger, band

than the expected product size (484 bp; Fig. 7-2B). Dideoxy sequencing of both gel purified products revealed the presence of a mutant allele, generated by using the cryptic splice acceptor, 58 nucleotides larger in size than the wild type product (Fig. 7-2C and D).

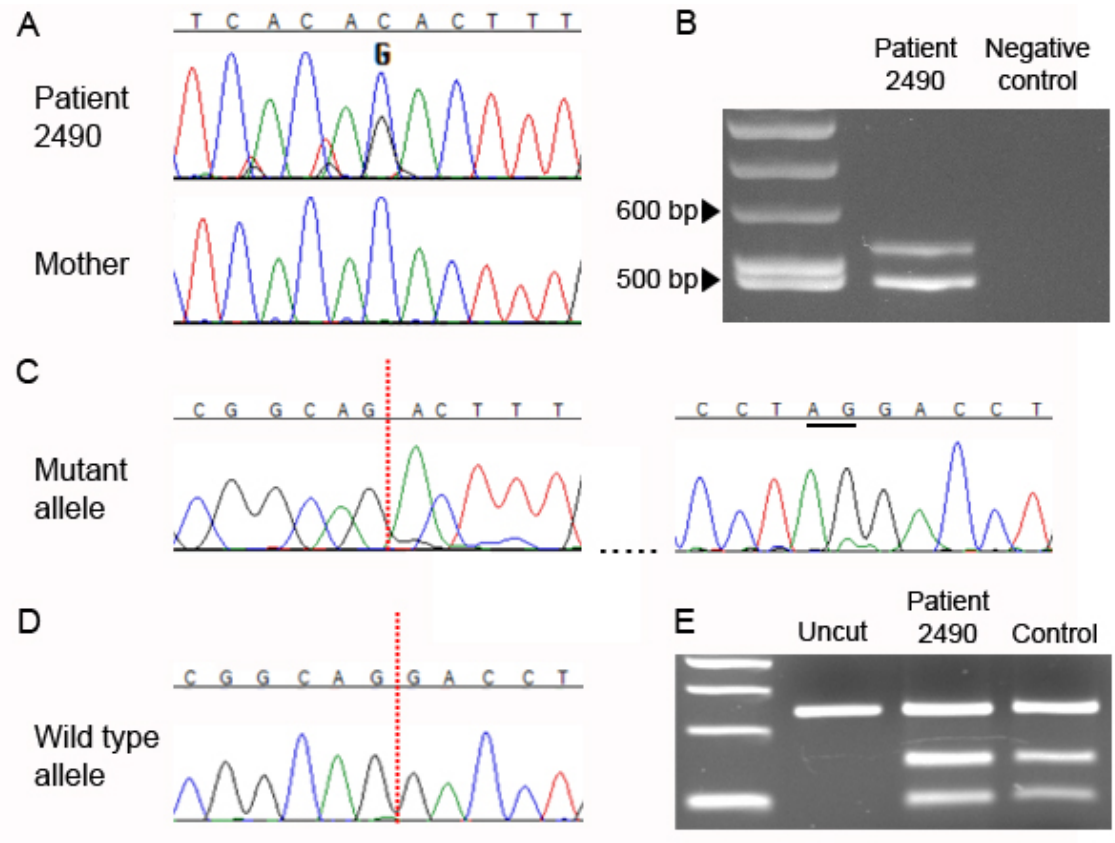


Figure 7-2

Analysis of *TCF12* upstream variant (chr15:57212041 C>G) in patient 2490.
A. Confirmation by dideoxy sequencing of the variant, which was not present in the mother's DNA. No DNA from the father was available for testing.
B. Amplification of the patient's cDNA revealed an extra product, larger than the 484 bp expected fragment size.
C. Sequencing of the upper band from B shows splicing using the cryptic acceptor generated by the mutation, and inclusion of intronic sequence. The wild type acceptor (underlined) is not used.
D. Sequencing of the smaller product generated from the patient's cDNA (shown in B) demonstrates normal splicing using the wild type acceptor. Red dotted lines represent exon-exon boundaries.
E. *HphI* digest of cDNA from patient 2490 and a normal control, the enzyme cuts the SNP rs3803452 (located in the 3' UTR) into fragments of 141 and 93 bp.

Although the *TCF12* variant chr15:57212041 C>G was shown to disrupt normal splicing, it is not clear-cut as to whether this would be pathogenic or not. The included intronic sequence occurs before the initiation codon and does not introduce an alternative ATG, so it may be that translation of the protein is unaffected. Of note, the variant is not present in the 1000 genomes dataset. To further investigate, a digest was designed to assay potential allelic imbalance by exploiting a heterozygous SNP (rs3803452) in the 3' UTR, which introduces an *HphI* recognition site. Equal amounts of starting cDNA from the patient and a control were amplified, and the resulting products digested. When compared on an agarose gel there did seem to be a difference in the ratios of cut and uncut product between the two samples, with patient 2490 seemingly having a higher proportion of digested product (Fig. 7-2E).The reduction in the amount of one allele may be indicative of nonsense-mediated decay, however, as it is not known if the SNP rs3803452 is in cis or trans with the chr15:57212041 C>G variant, this is difficult to interpret. Determining this would allow the stability of the mutant mRNA to be assessed, or to quantify more accurately, a method such as qPCR could be carried out.

7.3.2 Intronic splice variants

Unique intragenic single nucleotide substitutions (on the basis that these would be more likely to be genuine variants and not alignment errors) in *TWIST1* and *TCF12* were filtered against dbSNP and 1000 genomes to remove previously reported variants. The remaining substitutions were entered into Alamut Visual (http://www.interactive-biosoftware.com), which integrates predictions from MaxEntScan, Neural Network, Human Splice Finder (http://www.umd.be/HSF/) and GeneSplicer (http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml) to produce scores for putative splice sites. Of note there were no previously unreported variants in the intron of *TWIST1*. The *TCF12* intronic variants that were predicted to introduce a novel cryptic splice acceptor or donor site are listed in Table 7-4.

Table 7-4

Summary of splice sites introduced by *TCF12* intronic variants.

Patient	Variant	cDNA position	Predicted to create	Neural network score	
				Variant	Wild type*
1599	Chr15:57354622 A>G	c.149-1326	Donor	0.69	0.97
1329	Chr15:57446343 A>G	c.326-12257	Donor	0.92	0.98
6346	Chr15:57511296 A>G	c.580-12054	Acceptor	0.86	0.70
5897	Chr15:57554588 A>G	c.1582+182	Donor	0.93	0.97
2623	Chr15:57556049 A>G	c.1745+577	Donor	0.23	0.86
3175	Chr15:57556761 A>G	c.1745+1289	Acceptor	0.70	0.88

*The wild type score shown is for either the downstream wild type acceptor if an acceptor is predicted or the upstream wild type donor if a donor is predicted.

Of these putative cryptic splice sites, the predicted donor in individual 5897 and acceptor in individual 3175 seem the most interesting as both generate high scores. The donor site introduced in patient 5897 has a comparable score to the wild type donor, which is only 182 bp away. It is possible that in this context some splicing may occur using the cryptic site, generating a mutant product containing an extra 181 nucleotides at the end of exon 17. If this did occur then a stop codon would be introduced after the addition of nine amino acids. Follow up analysis revealed the variant had been inherited from the proband's unaffected father. In patient 3175, the introduction of a reasonably strong acceptor 1289 bases from the end of exon 18 could trigger the inclusion of a neo-exon. Follow up testing would need to be carried out on cDNA from these patients to assay if either variant alters normal splicing, as well as checking the parental samples of patient 3175 to determine the origin of this variant.

7.4 Analysis of deletions

The bioinformatic tool Pindel was used to detect simple deletions within the sequence data. This uses a pattern growth approach to identify the breakpoints of deletions (1 bp – 10 kb) from paired-end short reads (Ye et al. 2009). Using single unmapped reads from a read-pair, Pindel tests the possibility that they might contain a breakpoint, separating the unmapped reads into two fragments and aligning them individually to the reference sequence. Multiple reads containing the same information increase the confidence that the event is real (Ye et al. 2009).

The results generated after running all sequencing results through Pindel were filtered to remove deletions with <10 supporting reads. The remaining variants were checked against the 1000 genomes dataset and the database of genomic variants (DGV), and previously reported variation removed from further analysis. Similarly, deletion events that were called in multiple samples due to misalignment in repetitive elements were removed. Deletion sizes down to 6 bp were analysed, as this size and smaller would have been identified by the initial variant calling (verified by cross checking the two methods). No exonic deletions were identified using this method.

The largest deletion that was not a known copy number variant (CNV) was of approximately 3 kb in patient 4642 (left unicoronal synostosis). This was confirmed by a long-range PCR, which also revealed that it had been inherited from the proband's unaffected father (Fig. 7-3A). Looking at the sequence context revealed two *Alu Y* elements flanking the deleted region, indicating non-allelic homologous recombination (NAHR) as the underlying mechanism. This was confirmed by dideoxy sequencing of the gel-purified smaller fragment, which when aligned with sequences of the *Alu Y* elements showed nucleotides specific to each element between which the rearrangement must have occurred (Fig. 7-3A).

A second deletion (chr7:18570122-18570160), identified in patient 4482, was shown in an initial experiment to be present in the proband's affected mother. In this family only the proband had craniosynostosis, however, a short thumb phenotype segregated more extensively through the family. Expanded analysis on the basis that these two features were causally linked, showed that the variant did not segregate, and had been passed down from the proband's unaffected grandmother (Fig. 7-3B).

None of the deletions identified by Pindel, after filtering, were found to have occurred *de novo* (where both parental samples were available; summary of findings in Table 7-5), and none occurred in the minimal *TWIST1* region ch7:18747966-18782892.

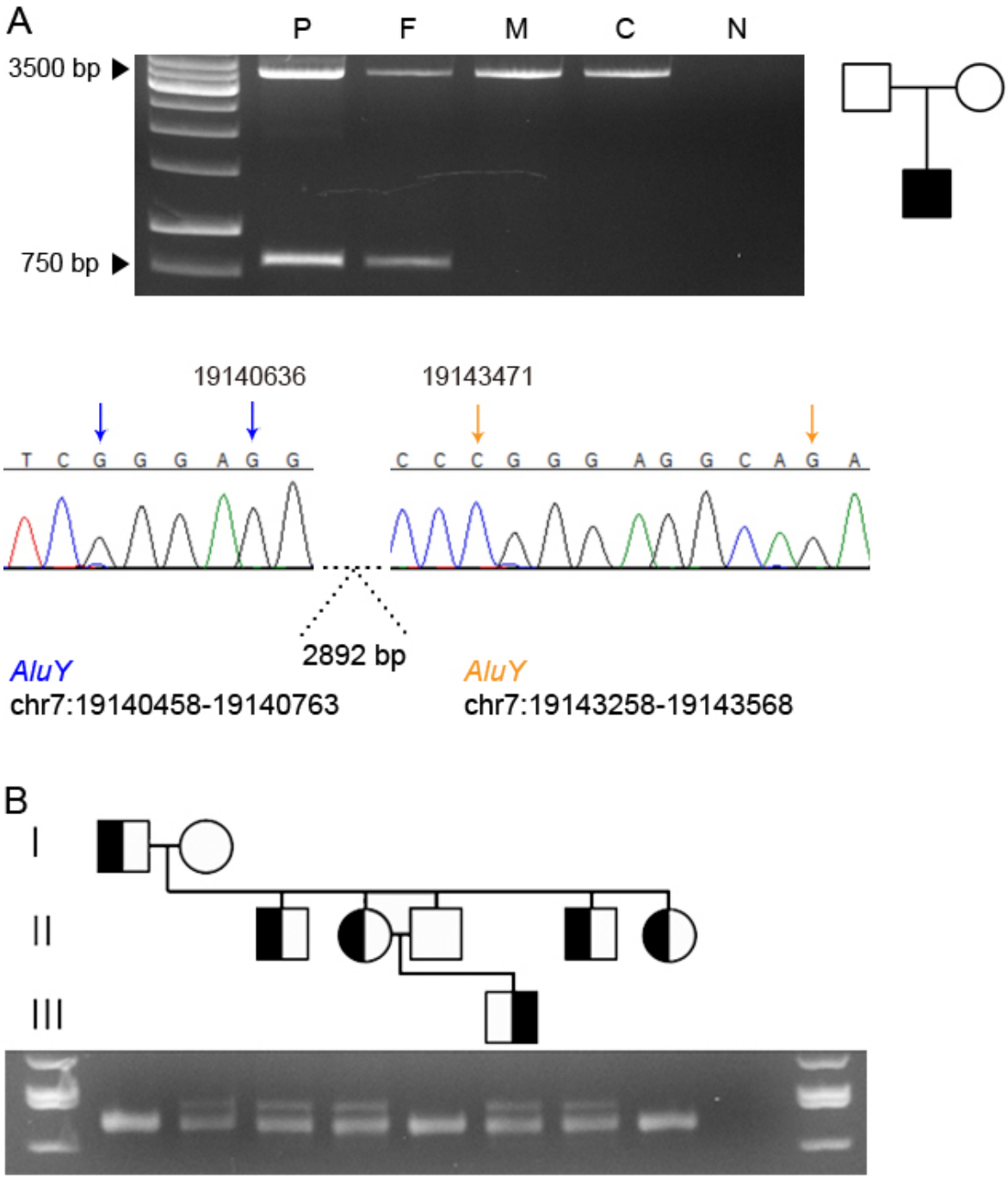


Figure 7-3

Further analysis of deletions in patients 4642 and 4482.

A. Long-range PCR confirmed presence of the deletion in patient 4642, which was probably caused by non-allelic homologous recombination between two flanking *Alu* elements. Sequencing of the smaller product narrowed down the position of the breakpoint to within a 26 bp homologous stretch; nucleotides specific to each element are indicated by blue or orange downward arrows above the chromatograms. P, proband; F, father; M, mother; C, control DNA; N, negative control.

B. Segregation of deletion chr7:18570122-18570160 in family of patient 4482; individuals with the short thumb phenotype represented by left-filled sectors, craniosynostosis by right-filled sector. The variant is inherited from the proband's unaffected grandmother (I-2), and is not present in an affected aunt (II-5).

Table 7-5

Summary of novel deletions identified by Pindel.

Patient	Deletion coordinates	Size (bp)	Inheritance
4642	Within region chr7: 19140636-19143471	~2895	From unaffected father
6168	Chr7:19037857-19037922	64	From unaffected father
6168	Chr7:18385479-18385523	43	From unaffected father
4482	Chr7:18570122-18570160	37	Does not segregate with phenotype
1535 5583 2479 3048	Chr7:18804498-18804519	20	From unaffected mother From unaffected mother From unaffected mother Not in unaffected mother, no DNA from unaffected father
4049	Chr7:18830885-18830905	19	From unaffected father
6058	Chr7:18272639-18272654	14	From unaffected mother
5776	Chr7:18663028-18663043	14	From unaffected father
2818 2479	Chr7:17921226-17921237	10	From unaffected mother From unaffected mother
5679	Chr15:57502663-57502673	9	From unaffected mother
5073	Chr15:57384862-57384871	8	From unaffected mother
6656	Chr7:17545688-17545696	7	From unaffected father
1342	Chr7:18326895-18326903	7	From unaffected father
1270	Chr7:19018942-19018949	6	Not in unaffected mother, no DNA from unaffected father

7.5 *TWIST1* variant analysis

7.5.1 *TWIST1* gene locus variants

As mentioned in section 7.3.1, variants occurring in the 5' and 3' UTRs and the upstream and downstream regions of *TWIST1* were all closely looked at to determine if they were genuine (rather than sequencing or alignment artefacts), and absent from the ExAC and 1000 genomes datasets. This led to the identification of five variants in the 5' UTR, within 70 bp of each other, in four patients (listed in Table 7-6). Of these individuals, only patient 4049 had parental samples available for follow-up testing, which demonstrated that the variant was inherited from her unaffected mother. All of the nucleotides, except for the cytosine at position chr7:19157213, are conserved in the mouse and

occur proximal to the transcription start site; chr7:19157258 is 33 bp downstream of the TATAA box and this particular variant alters the second nucleotide from the transcription start site (Fig. 7-4). Owing to its position in the 5' UTR, the stretch of DNA chr7:19157188-19157258 contains various regulatory features that can be visualised in the UCSC genome browser, such as interaction with multiple transcription factors (ChIP-seq from ENCODE track), DNaseI hypersensitivity in 102 cell types (out of 125 tested; ENCODE) and histone H3K27Ac marks (indicative of active enhancers and promoters; ENCODE). Additionally, there was a variant with low read depth that would need to be confirmed by dieoxy sequencing (c.-82G>A), and three variants within ~500 bp of the 5' UTR, upstream of *TWIST1*, spanning a region containing the same gamut of regulatory features.

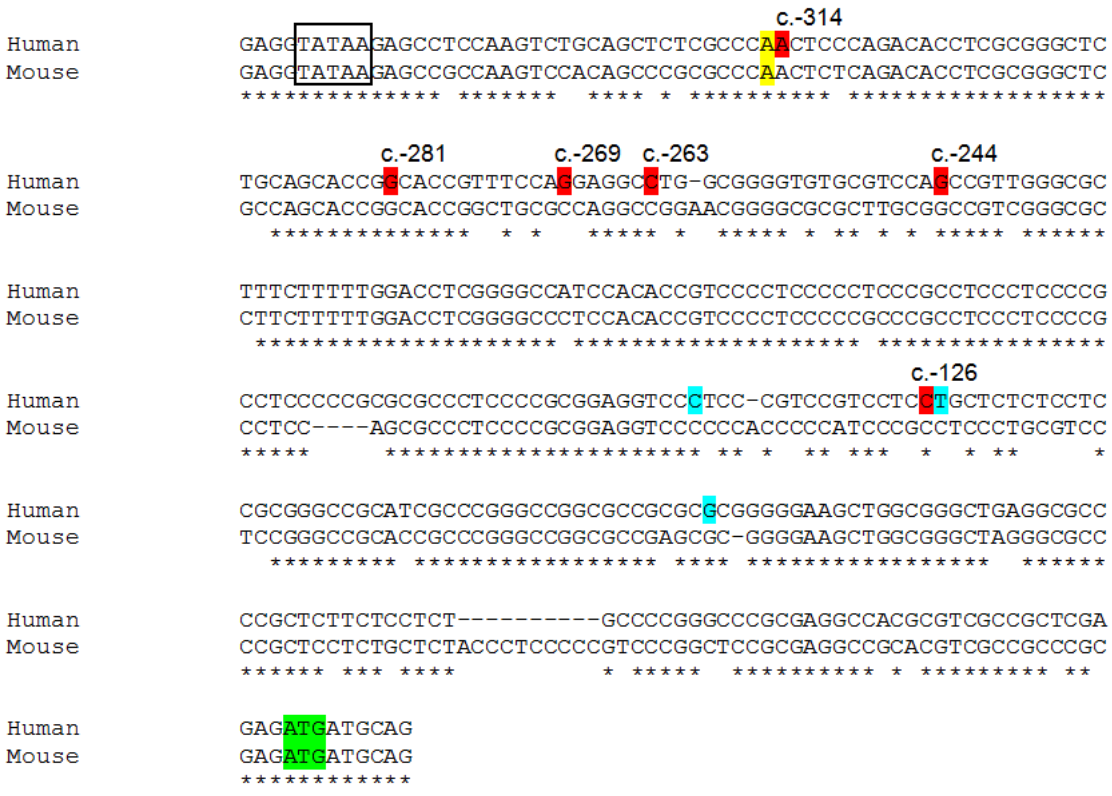


Figure 7-4

Alignment of human *TWIST1* and mouse *Twist1* 5' UTRs.
The black box indicates the functional TATAA box, the transcription start site is marked in yellow, locations of the variants identified in this study are highlighted red (cDNA position written above), SNPs identified in individuals in this study are in blue (L-R; rs4647961, rs377717796 and rs2522206) and the start ATG in green. Conserved nucleotides are shown with an asterisk.

Table 7-6

***TWIST1* 5' UTR and upstream variants not present in 1000 genomes dataset.**

Individual	Coordinates	Position	Variant	Phenotype
4856	Chr7: 19157070	5´ UTR c.-126	G>A	SCS-like (NB only 3 mutant and 3 variant reads)
1599	Chr7:19157188	5´ UTR c.-244	C>G	Bicoronal synostosis, positive family history with affected mother and brother.
4479	Chr7:19157207 Chr7:19157225*	5´ UTR c.-263 5´ UTR c.-281	G>T C>A	Right unicoronal synostosis with SCS features, similar features in mother.
4989	Chr7:19157213	5´ UTR c.-269	C>T	Right unicoronal synostosis
4049	Chr7:19157258	5´ UTR c.-314	T>G	Left unicoronal synostosis, variant inherited from unaffected mother.
5807	Chr7:19157509	Upstream† – 565 bp	C>G	Right unicoronal synostosis
6155	Chr7:19157640	Upstream – 696 bp	C>G	Left unicoronal synostosis
3527	Chr7:19157737	Upstream – 793 bp	G>T	Right unicoronal synostosis

*both variants are on the same allele.
†distance from position of c.1 adenine.

Looking at all the previously undescribed variants identified on chromosome 7 in this study (filtering for a minimum read depth of 20 and 30-70% variant reads) gives a total of 3927 variants, with a mean separation of 597 bp. These were then analysed for groups of three or more variants falling within 200 bp of each other, resulting in 15 groups of three variants (mean separation 93 bp), one group of four variants (across 165 bp) and the group of five variants described above (across 70 bp).

The wild type and translated 5' UTR containing the five variants were entered into SNPfold (http://ribosnitch.bio.unc.edu/snpfold/SNPfold.html), an algorithm developed to predict the potential consequences of single nucleotide variants on RNA. No significant differences were predicted for any of the variants from the wild type. However, cell lines are available from four of the patients (1599, 4049, 4856 and 4989), so RNA can be extracted for further analysis of the possible effects of these variants.

7.5.2 *TWIST1* minimal region variants

As well as the genomic regions proximal to *TWIST1*, the minimal region defined in section 1.4.2 (chr7:18747966-18782892) was specifically interrogated in all samples. All variants within this region

were checked to see if they were present in the 1000 genomes dataset, and were not included if they occurred in a repetitive element, as coverage was typically low in these regions and included a lot of false-positives due to alignment errors. Of note, only one of the 17 remaining variants occurred in a region enriched for transcription factor binding sites, DNaseI hypersensitivity or histone H3K27Ac marks (using the UCSC genome browser). This variant (chr7: 18780607 G>C) was in a region of DNaseI hypersensitivity in one of the 125 cell lines analysed for this track (the Jurkat cell line established from peripheral blood). None of the variants clustered together in a particular region (Table 7-7).

Table 7-7

Chr7:18747966-18782892 variants not present in the 1000 genomes dataset.

Individual	Coordinates	Variant	Distance from previous (bp)
1342	Chr7:18750274	C>T	1323
5887	Chr7:18751597	A>G	794
4566	Chr7:18752391	G>A	1881
806	Chr7:18754272	T>G	2449
4376	Chr7:18756721	A>G	4875
5776	Chr7:18761596	C>A	511
3650	Chr7:18762107	A>C	987
5116	Chr7:18763094	G>T	1862
6122	Chr7:18764956	C>T	4054
1599	Chr7:18769010	A>T	535
4566	Chr7:18769545	G>C	516
6904	Chr7:18770061	T>A	1662
4554	Chr7:18771723	T>C	1514
2492	Chr7:18773237	delTTA	383
1822	Chr7:18773620	A>T	6987
5073	Chr7:18780607	G>C	160
4554	Chr7:18780767	G>A	1323

7.5.3 Extended chromosome 7 variant analysis

The sequencing dataset was compared to recent results from a pilot study using ChIP-seq to identify H3K27Ac marks in human fetal sutural material (unpublished; Y Zhou). This highlighted five variants, which were verified by dideoxy sequencing and parental samples checked (where available, all individuals had clinically unaffected parents; Table 7-8).

Table 7-8

Summary of variants overlapping regions enriched for H3K27Ac marks.

Individual	Coordinates	Variant	Inheritance
4124	Chr7:17979676	G>C	From unaffected father
4392	Chr7:17980508	C>T	Parental samples not available
4870	Chr7:19146484	G>A	From unaffected father
4566	Chr7:19152348	C>A	From unaffected mother
6041	Chr7:19152403	T>A	From unaffected mother

Results were also compared to regions specified as *TWIST1* enhancers by the Vista enhancer browser (enhancer.lbl.gov*)*: hs644 chr7:18885971-18887230, hs2306 chr7:18868716-18870405 and hs2307 chr7:18874272-18875369. These regions were identified by extreme evolutionary sequence conservation, and defined as positive enhancers after showing reproducible expression in at least three independent transgenic embryos. To do this, putative enhancers were cloned upstream of a *lacZ* reporter gene and introduced into murine embryos via pronuclear injection, embryos were then harvested at E11.5 and stained to view the expression pattern.

Looking specifically within these three positive enhancer regions (and disregarding those present in the 1000 genomes dataset), there were two individuals with variants in hs644 (chr7:18886108 A>C and chr7:18887000 G>A; 892 bp apart), two with variants in hs2306 (chr7:18869708 A>T and chr7:18870201 T>C; 493 bp apart) and one with a variant in hs2307 (chr7:18875025 G>C). Three of these individuals were of Asian ethnicity.

7.6 Discussion

The targeted sequencing approach has provided results that have expanded the phenotypes associated with *ERF* and *TCF12*-related craniosynostosis. Patient 5243 is the first *ERF* case with unicoronal synostosis (mutations commonly cause multisuture synostosis) and as she did not resemble the typical phenotype, *ERF* screening had not previously been requested clinically. Her father, who also carries the mutation, was noted to have a prognathic mandible and had had some orthodontic problems, which further highlights the phenotypic heterogeneity associated with *ERF* mutations.

Conversely, *TCF12* mutations are not typically associated with multisuture synostosis as seen in patient 5939; for this reason, clinical *TCF12* testing had not previously been requested. This individual had a fairly severe range of phenotypic features (listed in Table 7-2) and although it is likely that the prune belly had a separate cause, the other features fall within the spectrum of those observed with craniosynostosis syndromes. The reason for the range of penetrance seen with *TCF12* mutations is as yet unknown. As discussed in section 1.5, the development of the coronal suture is an extremely complex process and the extent to which it is affected by a particular mutation may depend on several other factors, which may vary depending on a person's genetic background as well as environmental factors. For example, the allelic backgrounds of both non-mutant and mutant copies of *TCF12* could contribute to the phenotypic heterogeneity; gene expression is known to vary among individuals and there is a genetic contribution to this variation associated with different haplotypes (Cheung et al. 2003, Cheung et al. 2010, Sharma et al. 2013). To explore this further, the expression levels linked to haplotypes containing *TCF12* mutations could be compared between clinically affected and unaffected individuals, as well as unaffected controls with two normal copies of *TCF12*. It also remains a possibility that another mutation in a different gene, in combination with the *TCF12* mutation, led to the range of features in patient 5939. To know for certain, exome sequencing would need to be carried out, perhaps on the parent-child trio, but even this might not yield a conclusive answer.

This experiment has also highlighted that molecular diagnoses may have been missed. Both such cases described here (TWIST1 p.Asn125Ser and p.Gln161*) were originally tested over 10 years ago, and screening methods have improved drastically in that timeframe. It is unlikely that with modern sequencing analysis these mutations would be missed, as identification by the human eye alone is no longer relied upon. However, this does indicate that it may be worth re-checking older "negative" results as two such findings out of 160 samples is not insignificant.

It is perhaps surprising that no deletions of *TCF12* were identified, as mentioned in section 5.3, three different intragenic deletions have been identified in craniosynostosis patients, indicating that such events are viable. However, the rate at which deletions occur is dependent on the genomic context. For example, in deletions caused by NAHR the breakpoints have been shown to cluster in defined hot spots within duplicated sequences, although rare, non-recurrent NAHRs can also be mediated by highly homologous repetitive sequences, such as that described in section 7.4 between two *Alu Y* elements (Turner et al. 2008). Unlike genomic locations such as chromosome 17p11.2, which is recurrently deleted in Smith-Magenis syndrome and duplicated in Potocki-Lupski syndrome (both developmental disorders), chromosome 15q21.3 (containing *TCF12*) does not seem to be frequently associated with these types of rearrangement. The *de novo* rate of deletions is hard to gauge accurately, and studies looking specifically at recombination hotspots have reported a highly variable range (7-fold across individuals tested). However, these events are unlike point mutations, which occur more frequently on the paternal allele, and are not correlated with increased paternal age (MacArthur et al. 2014, Turner et al. 2008).

An interesting feature highlighted by this targeted sequencing strategy is the occurrence of mutations in the 5′ UTRs of both *TCF12* and *TWIST1*. The 5′ UTR contains a lot of important features necessary for the regulation of transcription; secondary structure (such as hairpins), binding sites for RNA binding proteins and upstream open reading frames (uORFs). Experimental evidence suggests that uORFs can affect efficient translation of the downstream coding sequence by disrupting ribosomes, reducing protein levels (Calvo, Pagliarini, and Mootha 2009). Disruption of uORFs has been linked to several human diseases including the craniofacial disorder CFNS (Twigg, Babbs, et al. 2013). In this example a point mutation in the 5′ UTR of *EFNB1* (c.-95T>C) removes the stop codon of a four codon uORF, resulting in a 44 codon predicted translation product that overlaps (in a different frame) the regular *EFNB1* coding sequence (Twigg, Babbs, et al. 2013). Interestingly, the 5′ UTR of *TCF12* contains an uORF, which is extended from 10 to 16 codons with the inclusion of the extra nucleotides in the mutant product, but still retains an in-frame stop codon. Whether this small

difference would be enough to significantly impact translation is not clear, to determine this an assay such as that used by Twigg et al (2013) would need to be carried out. In their publication, the authors used a luciferase reporter construct to measure translational output of both the wild type and mutant *EFNB1* with the c.-95T>C variant, confirming that the mutation had an inhibitory effect.

If the chr15:57212041 C>G variant does not alter *TCF12* expression via effects on the uORF, the inclusion of an extra 58 bp could still interfere with the correct processing of *TCF12*. As well as physically disturbing normal transcription by altering the size of the 5' UTR and distance from the promoter the variant may also disrupt the binding of regulatory proteins, and/or affect secondary structure, affecting the rate of transcription. Non-coding expansions have been linked to many diseases, for example in Richieri-Costa-Pereira syndrome (RCPS); a craniofacial disorder characterised by features overlapping PRS, a midline cleft mandible and limb defects. RCPS affected individuals were found to have 14-16 repeats of a 18-20 nucleotide motif, compared with 3-12 repeats observed in normal controls, and had a 30-40% reduction in gene expression as indicated by qPCR (Favaro et al. 2014). Further expression analysis using qPCR would help to show if *TCF12* levels are indeed reduced in individual 2490. If the variant had occurred *de novo* it would be more evidence in support of pathogenicity, so obtaining DNA from the father for testing would worthwhile. However, due to non-penetrance, it is equally likely that the father could carry the variant and be clinically unaffected.

Predicting the pathogenicity of the *TWIST1* 5' UTR variants (Table 7-5) is difficult, however, the apparent clustering of variants within the promoter region is intriguing. Experimental evidence showed that the TATAA box indicated in Fig. 7-4 was indeed functional and that transcription started from the adenine 32 bp downstream, furthermore, the *TWIST1* promoter region shares 71% sequence similarity with that of the mouse (Wang et al. 1996). The substitution of the adjacent nucleotide to the transcription start site (in patient 4049; c.-314T>G) could conceivably affect the binding of the RNA polymerase II complex. Equally, the other variants within this region may affect the progression of the transcription machinery. The fact that the c.-314T>G variant was inherited

from an unaffected parent points towards non-pathogenicity, however given that substantial non-penetrance is seen with *TCF12* (and *ERF*) mutations, it doesn't necessarily rule it out.

No effect on secondary structure was predicted by the introduction of the 5' UTR substitutions, which suggests that if they are indeed pathogenic, it would be via another mechanism like the alteration of recognition sites for RNA binding proteins. The c.-263G>T variant actually introduces an ATG (*TWIST1* is transcribed in the reverse orientation) and so may initiate an uORF, which would not encounter a stop codon until that of the normally transcribed ORF. This patient in particular (4479) has a phenotype indicative of SCS, which is likely to be inherited from his mother so obtaining further family DNA samples to check segregation of the variant would be informative.

Promoter mutations have been linked to disease previously, for example, Wieczorek et al (2014) identified two different deletions in the promoter region of *TXNL4A* in patients with Burn-McKeown syndrome (BMKS; the combination of bilateral choanal atresia, hearing loss, cleft lip and/or palate and other craniofacial features). To test if these deletions could affect gene expression, the authors carried out luciferase reporter assays using constructs containing either the wild type or mutated promoter regions, the 5' UTR and the first 21 codons of *TXNL4A* upstream of a luciferase reporter gene. The mutant promoters reduced activity by 59% and 72% compared to the wild type, indicating that the deleted regions are essential for normal expression of *TXNL4A* (Wieczorek et al. 2014). In eight cases, individuals with BMKS were compound heterozygous for a coding mutation as well as a common 34 bp promoter deletion. However, one family was homozygous for a second (overlapping) promoter deletion, suggesting that this promoter mutation alone was sufficient to cause the phenotype (Wieczorek et al. 2014). Similarly, making constructs containing the different *TWIST1* 5' UTR variants fused to a reporter gene would allow their effects here to be assayed, and compared with the wild type.

CHAPTER 8

CONCLUSIONS

Chapter 8: Conclusions

The overall aim of this thesis was to increase the current understanding of craniosynostosis, by gaining more of an insight into the underlying biology and also by improving diagnoses. More specifically, I wanted to use next generation sequencing to identify the molecular cause of craniosynostosis in *TWIST1*-negative SCS patients.

8.1 The pathophysiology of craniosynostosis

Both the identification of new disease genes (*TCF12* and *CDC45*) and the extension of pathogenic mechanisms associated with known disease genes (*FGFR2* and *TWIST1*) have helped to advance the understanding of the pathophysiology of craniosynostosis.

The identification of *CDC45* as a disease gene (section 4.5.2.1-3) expands the phenotype of MGS and strengthens the link between the DNA replication pathway and craniosynostosis. Screening of other replication pathway members in craniosynostosis patients is currently being undertaken, which could potentially identify more disease genes and reveal the extent of the contribution of this pathway to suture formation. In addition, work exploring the specific effects of *CDC45* mutations on DNA replication and cell proliferation is ongoing.

The discovery of a new craniofacial phenotype associated with substitutions of p.Glu117 in TWIST1 (Chapter 6) is fascinating in that it demonstrates a completely different pathophysiological mechanism to that of SCS. Although rare, this new syndrome provides further evidence of the specific role of this highly conserved residue in the basic DNA-binding domain of bHLH proteins.

A second novel genotype-phenotype correlation is perhaps demonstrated by the *RUNX2* variant in patient 4809 (c.1507 T>G, p.Ser503Ala; section 4.5.4.1). As haploinsufficiency of *RUNX2* leads to a distinct phenotype (CCD), a different mechanism, such as gain-of-function, may lead to craniosynostosis. Obtaining DNA samples from more family members to ascertain the origin of the

mutation would help to determine if it could be causative, and then perhaps targeted sequencing of the phosphorylated residues in the C-terminal region in other patients could be undertaken.

8.2 Clinical applications

It is now clear that some patients with a diagnosis of SCS have mutations in *TCF12*, and that there is also a phenotypic overlap with *CDC45* mutations. Diagnostic screening of *TCF12* has been approved within the NHS, and the segregation from diagnoses of SCS is important due to the different surgical trajectories (*TCF12* patients typically having a lower reoperation rate). However, the substantial non-penetrance seen with *TCF12* mutations makes genetic counselling challenging; for example, the c.1115-2A>G mutation in patient 5939 (section 7.2), if it is solely responsible for the severe phenotype exhibited, was inherited from the proband's clinically unaffected father. As yet, there is no explanation for this phenomenon and further work will need to be undertaken, for example to see if there is any correlation with differing levels of background allele expression. Interestingly, there are various *TCF12* loss-of-function alleles listed in ExAC, which are presumably non-penetrant. This perhaps points towards *TCF12* being more of a susceptibility locus (albeit with low allele frequency) rather than an out-and-out causative one.

The discovery of *CDC45* mutations has an important bearing on genetic counselling due to the recessive inheritance pattern, compared with the majority of craniosynostosis syndromes (including SCS) that are dominantly inherited. As patient 2930 (Fig. 4-2) was assumed to have a sporadic mutation, he was advised that he had 50% risk of having an affected child whereas in reality that risk was ~0. The timing of the identification of *CDC45* meant that the parents of patient 5566 (Fig. 4-2) were able to have a revised sibling recurrence risk of 25%, rather than ~5%, and in fact decided not have more children based on this.

8.3 The interpretation of next generation sequence data

The analysis undertaken during the course of this thesis has provided valuable lessons in the interpretation of next generation sequence data. It has demonstrated that clinical phenotyping may be misleading; as patient 2930 (Fig. 4-2) was thought to strongly resemble an SCS phenotype the focus was directed towards *TWIST1*. Also, care should be taken if assuming a *de novo* cause for a particular disease, as in this case the *FLRT1* variant was incidental to the phenotype. Furthermore, the underlying pathogenic variant may not be obvious; *CDC45* was originally overlooked due to the synonymous variant, which was subsequently demonstrated to affect splicing. This was also shown to be true in the case of the *FGFR2* mutations discussed in Chapter 3.

Collectively, the discoveries of coding region variants made during the course of this thesis demonstrate the various ways by which intragenic mutations can be missed, and indicates that all variants occurring within known disease genes should be carefully investigated before being discounted.

8.4 The non-coding genome

One of the main drivers from early on in this project was to explore the non-coding regions of the genome for mutations; first through whole genome sequencing then through targeted sequencing. Strikingly, nearly all of the positive findings actually occurred in the coding part of the genome. Although recent research has started to focus on non-coding DNA, there is still not the same strength of understanding that exists for the coding genome. Discoveries can come through either functional disruption or statistical association, for example in the novel *TWIST1* phenotype there was prior knowledge of the glutamic acid having a specialised role, as well as two separate *de novo* mutation events. This statistical association is usually only possible in coding regions, outside genes it is very difficult to interpret variation; during this work I found a large number of novel rearrangements but their significance is unknown. Even within genes, far less is understood about the functions of the UTRs, which can be highly conserved, than the coding regions. There is certainly evidence pointing

towards the 5' UTR of *TWIST1* having an important role in regulating its expression, which merits further investigation. Three of the individuals who have variants in this region are more SCS-like than non-syndromic, and also this cluster of variants seems to be somewhat unusual in their proximity to each other in relation to the distribution of variants throughout the rest of the 2.4 Mb region investigated. Statistical analysis would need to be carried out to determine if this is in fact significant, such as running a Monte Carlo simulation to estimate the probability distribution of variation.

It is probable that non-coding pathogenic variants (such as UTR regulatory mutations and intronic variants that affect normal splicing) have been missed for many human diseases. Now that expanded exome sequencing (to incorporate both 5' and 3' UTRs) and whole genome sequencing are becoming more routinely carried out we can expect to see an increase in the identification of non-coding variants associated with disease. This may be particularly significant for genes like *TWIST1*, which due to high GC content and repetitive sequence has been technically difficult to work with.

In terms of *TWIST1* enhancers, it may be that mutations would not lead to an SCS phenotype as the pathogenic mechanism could be more subtle than that of a coding variant. It is believed that genes have multiple enhancers driving similar patterns of expression, which provide a combinatorial effect towards expression and may also result in redundancy (Attanasio et al. 2013, Cotney et al. 2012). In this way, if a single enhancer is affected by a mutation it may not be pathogenic, however when large rearrangements occur (like those described in section 1.4.2) gene expression is sufficiently disrupted to give rise to the full phenotype as seen when the gene itself is mutated. As well as this, regulation via enhancer elements is very cell-type specific; to identify which elements interact with *TWIST1* during development of the cranial suture, the undifferentiated sutural mesenchyme would need to be assayed. With the correct cell type, experiments such as ChIP-seq and chromatin conformation capture could be carried out. The limiting factor is the amount of starting material with which to perform these experiments, one option would be to try in the mouse as this has been shown to be a good model of various craniosynostosis syndromes, and then look for homology in human DNA.

Another option, which is currently being undertaken, is to look in human fetal sutural material although this is technically demanding and options are limited by cell number and the fact that the tissue has been frozen.

As already mentioned, this study has highlighted the difficulties in analysing and interpreting patient sequence data; next generation sequencing technology alone cannot provide all the answers. The growth in freely available resources, like ExAC and the 1000 genomes project, continues to be incredibly useful for understanding common variants, however, studies suggest that on average we each have ~3.2 million single nucleotide variants per genome, of which ~5000 will not be present in such datasets (The UK10K Consortium, 2015). Therefore, the ability to analyse data in the context of other craniosynostosis sequencing projects has been invaluable in helping to identify less obviously pathogenic mutations, although two of the individuals who were included in this project (4400 and 4580) are still without any strong candidates. At the time, these were among the first whole genomes sequenced in the UK and compared with more recent sequencing results the quality is obviously lower, for example around *TWIST1* there are areas that are only covered by one or two reads, probably caused by the high GC content and repetitive segments. In this way, it is possible that some pathogenic variants may have been missed. Equally it may be the case that these individuals have a polygenic, rather than a monogenic, cause. Gene-environment interactions have also been shown to be important, for example susceptibility loci have been identified for sagittal synostosis (Justice et al. 2012). Hopefully with continued analysis and as more craniosynostosis genes are discovered we will be able to provide these patients with an answer.

APPENDIX

APPENDIX

Whole genome sequencing: further details on selected genes

Patient 4580		
Gene	**Del score**	**Details**
ZRANB3	Compound heterozygous	Both variants from unaffected father
NOV	6	Present in unaffected father
APAF1	6	Not present in unaffected father, in ExAC allele freq: 0.0001
RANBP17	5	Present in unaffected father
TGFBR3	4	Present in unaffected father
FLII	4	Present in unaffected father
MSX1	2	Variant present in FaceBase dataset
CYP26B1	1	Not present in unaffected father, in ExAC allele freq: 0.0000171
TECPR1	-	Not present in unaffected father
EPHA2	Synonymous	Nothing abnormal detected in cDNA PCR

Patient 4400		
Gene	**Del score**	**Details**
ARNTL	6	Present in affected mother, one splicing mutation in EVS.
EPHB4	6	Present in affected mother, other individual with *EPHB4* variant does not segregate. In ExAC allele freq: 0.0001 (rs200549250).
ORC1	6	Not present in affected mother
TULP3	6	Present in affected mother, numerous stop and frameshift variants in EVS.
KLHDC1	6	Present in affected mother, abolishes splice donor. Other stop and frameshift variants in EVS.
CNN1	5	Not present in affected mother
HS6ST1	5	Not present in affected mother (rs373944145)
ARHGEF17	4	Not present in affected mother
CHD3	4	Not present in affected mother
SIM2	4	Not present in affected mother
ZRANB2	4	Not present in affected mother

Patient 4809		
Gene	**Del score**	**Details**
ARHGEF17	6	Not present in affected son
EFTUD2	?splice	Present in affected son
INSR	4	Present in affected son

PUBLICATIONS

Publications

Sharma VP, **Fenwick AL**, Brockop MS, McGowan SJ, Goos JA, Hoogeboom AJ, Brady AF, Jeelani NO, Lynch SA, Mulliken JB, Murray DJ, Phipps JM, Sweeney E, Tomkins SE, Wilson LC, Bennett S, Cornall RJ, Broxholme J, Kanapin A; 500 Whole-Genome Sequences (WGS500) Consortium, Johnson D, Wall SA, van der Spek PJ, Mathijssen IM, Maxson RE, Twigg SR, Wilkie AO. 2013. "Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis." *Nat Genet* 45(3):304-7.

Fenwick AL, Goos JA, Rankin J, Lord H, Lester T, Hoogeboom AJ, van den Ouweland AM, Wall SA, Mathijssen IM, Wilkie AO. 2014. "Apparently synonymous substitutions in FGFR2 affect splicing and result in mild Crouzon syndrome." *BMC Med Genet* 15:95.

Piard J, Rozé V, Czorny A, Lenoir M, Valduga M, **Fenwick AL**, Wilkie AO, Maldergem LV. (2015) "TCF12 microdeletion in a 72-year-old woman with intellectual disability." *Am J Med Genet A.* 167A(8):1897-901.

Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, **Fenwick AL**, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lamble S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SR, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AO, Bentley D, Donnelly P, McVean G. 2015. "Factors influencing success of clinical genome sequencing across a broad spectrum of disorders." *Nat Genet* 47(7):717-26.

REFERENCES

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S and Sunyaev S.R. 2010. "A method and server for predicting damaging missense mutations." *Nat Meth* 7 (4):248-249.

Agochukwu, N. B., B. D. Solomon, and M. Muenke. 2012. "Impact of genetics on the diagnosis and clinical management of syndromic craniosynostoses." *Childs Nerv Syst* 28 (9):1447-63.

Al-Owain, M., N. Kaya, H. Al-Zaidan, N. Al-Hashmi, A. Al-Bakheet, M. Al-Muhaizea, A. Chedrawi, R. K. Basran, and A. Milunsky. 2011. "Novel intragenic deletion in OPHN1 in a family causing XLMR with cerebellar hypoplasia and distinctive facial appearance." *Clin Genet* 79 (4):363-70.

Aparicio, T., A. Ibarra, and J. Mendez. 2006. "Cdc45-MCM-GINS, a new power player for DNA replication." *Cell Div* 1:18.

Attanasio, C., A. S. Nord, Y. Zhu, M. J. Blow, Z. Li, D. K. Liberton, H. Morrison, I. Plajzer-Frick, A. Holt, R. Hosseini, S. Phouanenavong, J. A. Akiyama, M. Shoukry, V. Afzal, E. M. Rubin, D. R. FitzPatrick, B. Ren, B. Hallgrimsson, L. A. Pennacchio, and A. Visel. 2013. "Fine tuning of craniofacial morphology by distant-acting enhancers." *Science* 342 (6157):1241006.

Bach, J. E., B. Wolf, J. Oldenburg, C. R. Muller, and S. Rost. 2015. "Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene." *Thromb Haemost* 114 (3).

Bauerschmidt, C., S. Pollok, E. Kremmer, H. P. Nasheuer, and F. Grosse. 2007. "Interactions of human Cdc45 with the Mcm2-7 complex, the GINS complex, and DNA polymerases delta and epsilon during S phase." *Genes Cells* 12 (6):745-58.

Bellus, G. A., K. Gaudenz, E. H. Zackai, L. A. Clarke, J. Szabo, C. A. Francomano, and M. Muenke. 1996. "Identical mutations in three different fibroblast growth factor receptor genes in autosomal dominant craniosynostosis syndromes." *Nat Genet* 14 (2):174-6.

Benko, S., J. A. Fantes, J. Amiel, D. J. Kleinjan, S. Thomas, J. Ramsay, N. Jamshidi, A. Essafi, S. Heaney, C. T. Gordon, D. McBride, C. Golzio, M. Fisher, P. Perry, V. Abadie, C. Ayuso, M. Holder-Espinasse, N. Kilpatrick, M. M. Lees, A. Picard, I. K. Temple, P. Thomas, M. P. Vazquez, M. Vekemans, H. Roest Crollius, N. D. Hastie, A. Munnich, H. C. Etchevers, A. Pelet, P. G. Farlie, D. R. Fitzpatrick, and S. Lyonnet. 2009. "Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence." *Nature genetics* 41 (3):359-64.

Bialek, P., B. Kern, X. Yang, M. Schrock, D. Sosic, N. Hong, H. Wu, K. Yu, D. M. Ornitz, E. N. Olson, M. J. Justice, and G. Karsenty. 2004. "A twist code determines the onset of osteoblast differentiation." *Dev Cell* 6 (3):423-35.

Bochukova, E. G., T. Roscioli, D. J. Hedges, I. B. Taylor, D. Johnson, D. J. David, P. L. Deininger, and A. O. Wilkie. 2009. "Rare mutations of FGFR2 causing apert syndrome: identification of the first partial gene deletion, and an Alu element insertion from a new subfamily." *Hum Mut* 30 (2):204-11.

Boisson, B., Y. D. Wang, A. Bosompem, C. S. Ma, A. Lim, T. Kochetkov, S. G. Tangye, J. L. Casanova, and M. E. Conley. 2013. "A recurrent dominant negative E47 mutation causes agammaglobulinemia and BCR(-) B cells." *J Clin Invest* 123 (11):4781-5.

Bonaventure, J., and V. El Ghouzzi. 2003. "Molecular and cellular bases of syndromic craniosynostoses." *Expert Rev Mol Med* 5 (4):1-17.

Bottcher, R. T., N. Pollet, H. Delius, and C. Niehrs. 2004. "The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling." *Nat Cell Biol* 6 (1):38-44.

Boulet, S. L., S. A. Rasmussen, and M. A. Honein. 2008. "A population-based study of craniosynostosis in metropolitan Atlanta, 1989-2003." *Am J Med Genet A* 146A (8):984-91.

Bush, J. O., and P. Soriano. 2012. "Eph/ephrin signaling: genetic, phosphoproteomic, and transcriptomic approaches." *Semin Cell Dev Biol* 23 (1):26-34.

Cai, J., B. K. Goodman, A. S. Patel, J. B. Mulliken, L. Van Maldergem, G. E. Hoganson, W. A. Paznekas, Z. Ben-Neriah, R. Sheffer, M. L. Cunningham, D. L. Daentl, and E. W. Jabs. 2003. "Increased

risk for developmental delay in Saethre-Chotzen syndrome is associated with TWIST deletions: an improved strategy for TWIST mutation screening." *Hum Genet* 114 (1):68-76.

Calvo, S. E., D. J. Pagliarini, and V. K. Mootha. 2009. "Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans." *Proc Natl Acad Sci U S A* 106 (18):7507-12.

Carissimi, C., L. Saieva, J. Baccon, P. Chiarella, A. Maiolica, A. Sawyer, J. Rappsilber, and L. Pellizzoni. 2006. "Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly." *J Biol Chem* 281 (12):8126-34.

Cartegni, L., Chew, S.L., and Krainer, A.R. 2002. "Listening to silence and understanding nonsense: exonic mutations that affect splicing." *Nat Rev Genet* 3 (4):285-298.

Carver, E. A., K. F. Oram, and T. Gridley. 2002. "Craniosynostosis in Twist heterozygous mice: a model for Saethre-Chotzen syndrome." *Anatomical Rec* 268 (2):90-2.

Castanon, I., S. Von Stetina, J. Kass, and M. K. Baylies. 2001. "Dimerization partners determine the activity of the Twist bHLH protein during Drosophila mesoderm development." *Development* 128 (16):3145-59.

Chai, Y., X. Jiang, Y. Ito, P. Bringas, Jr., J. Han, D. H. Rowitch, P. Soriano, A. P. McMahon, and H. M. Sucov. 2000. "Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis." *Development* 127 (8):1671-9.

Chang, A. T., Y. Liu, K. Ayyanathan, C. Benner, Y. Jiang, J. W. Prokop, H. Paz, D. Wang, H. R. Li, X. D. Fu, F. J. Rauscher, 3rd, and J. Yang. 2015. "An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors." *Genes Dev* 29 (6):603-16.

Chen, L., Li, D., Li, C., Engel, A., and C.X. Deng. 2003. "A Ser250Trp substitution in mouse fibroblast growth factor receptor 2 (Fgfr2) results in craniosynostosis." *Bone* 33 (2):169-178.

Chen, Z. F., and R. R. Behringer. 1995. "twist is required in head mesenchyme for cranial neural tube morphogenesis." *Genes Dev* 9 (6):686-99.

Cheng, Y. C., M. Amoyel, X. Qiu, Y. J. Jiang, Q. Xu, and D. G. Wilkinson. 2004. "Notch activation regulates the segregation and differentiation of rhombomere boundary cells in the zebrafish hindbrain." *Dev Cell* 6 (4):539-50.

Cheung, V. G., L. K. Conlin, T. M. Weber, M. Arcaro, K. Y. Jen, M. Morley, and R. S. Spielman. 2003. "Natural variation in human gene expression assessed in lymphoblastoid cells." *Nat Genet* 33 (3):422-5.

Cheung, V. G., R. R. Nayak, I. X. Wang, S. Elwyn, S. M. Cousins, M. Morley, and R. S. Spielman. 2010. "Polymorphic cis- and trans-regulation of human gene expression." *PLoS biology* 8 (9).

Chong, J. X., K. J. Buckingham, S. N. Jhangiani, C. Boehm, N. Sobreira, J. D. Smith, T. M. Harrell, M. J. McMillin, W. Wiszniewski, T. Gambin, Z. H. Coban Akdemir, K. Doheny, A. F. Scott, D. Avramopoulos, A. Chakravarti, J. Hoover-Fong, D. Mathews, P. D. Witmer, H. Ling, K. Hetrick, L. Watkins, K. E. Patterson, F. Reinier, E. Blue, D. Muzny, M. Kircher, K. Bilguvar, F. Lopez-Giraldez, V. R. Sutton, H. K. Tabor, S. M. Leal, M. Gunel, S. Mane, R. A. Gibbs, E. Boerwinkle, A. Hamosh, J. Shendure, J. R. Lupski, R. P. Lifton, D. Valle, D. A. Nickerson, and M. J. Bamshad. 2015. "The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities." *Am J Hum Genet* 97 (2):199-215.

Chun, S., and J. C. Fay. 2009. "Identification of deleterious mutations within three human genomes." *Genome Res* 19 (9):1553-61.

Connerney, J., V. Andreeva, Y. Leshem, M. A. Mercado, K. Dowell, X. Yang, V. Lindner, R. E. Friesel, and D. B. Spicer. 2008. "Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure." *Dev Biol* 318 (2):323-34.

Connerney, J., V. Andreeva, Y. Leshem, C. Muentener, M. A. Mercado, and D. B. Spicer. 2006. "Twist1 dimer selection regulates cranial suture patterning and fusion." *Dev Dyn* 235 (5):1345-57.

Cornejo-Roldan, L. R., E. Roessler, and M. Muenke. 1999. "Analysis of the mutational spectrum of the FGFR2 gene in Pfeiffer syndrome." *Hum Genet* 104 (5):425-31.

Corsi, A. K., T. M. Brodigan, E. M. Jorgensen, and M. Krause. 2002. "Characterization of a dominant negative C. elegans Twist mutant protein with implications for human Saethre-Chotzen syndrome." *Development* 129 (11):2761-72.

Cotney, J., J. Leng, S. Oh, L. E. DeMare, S. K. Reilly, M. B. Gerstein, and J. P. Noonan. 2012. "Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb." *Genome Res* 22 (6):1069-80.

Dahmann, C., and K. Basler. 1999. "Compartment boundaries: at the edge of development." *Trends Genet* 15 (8):320-6.

David, D., J. Cardoso, B. Marques, R. Marques, E. D. Silva, H. Santos, and M. G. Boavida. 2003. "Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFbeta2 in the pathogenesis of Peters' anomaly." *Genomics* 81 (5):489-503.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, and S. Batzoglou. 2010. "Identifying a high fraction of the human genome to be under selective constraint using GERP++." *PLoS Comput Biol* 6 (12):e1001025.

De Marco, P., A. Raso, S. Beri, S. Gimelli, E. Merello, S. Mascelli, M. Baldi, A. M. Baffico, M. Pavanello, A. Cama, V. Capra, R. Giorda, and G. Gimelli. 2011. "A de novo balanced translocation t(7;12)(p21.2;p12.3) in a patient with Saethre-Chotzen-like phenotype downregulates TWIST and an osteoclastic protein-tyrosine phosphatase, PTP-oc." *Eur J Med Genet* 54 (5):e478-83.

de Munnik, S. A., L. S. Bicknell, S. Aftimos, J. Y. Al-Aama, Y. van Bever, M. B. Bober, J. Clayton-Smith, A. Y. Edrees, M. Feingold, A. Fryer, J. M. van Hagen, R. C. Hennekam, M. C. Jansweijer, D. Johnson, S. G. Kant, J. M. Opitz, A. R. Ramadevi, W. Reardon, A. Ross, P. Sarda, C. T. Schrander-Stumpel, J. Schoots, I. K. Temple, P. A. Terhal, A. Toutain, C. A. Wise, M. Wright, D. L. Skidmore, M. E. Samuels, L. H. Hoefsloot, N. V. Knoers, H. G. Brunner, A. P. Jackson, and E. M. Bongers. 2012. "Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis." *Eur J Hum Genet* 20 (6):598-606.

Deckelbaum, R. A., G. Holmes, Z. Zhao, C. Tong, C. Basilico, and C. A. Loomis. 2012. "Regulation of cranial morphogenesis and cell fate at the neural crest-mesoderm boundary by engrailed 1." *Development* 139 (7):1346-58.

Deckelbaum, R. A., A. Majithia, T. Booker, J. E. Henderson, and C. A. Loomis. 2006. "The homeoprotein engrailed 1 has pleiotropic functions in calvarial intramembranous bone formation and remodeling." *Development* 133 (1):63-74.

Del Gatto, F., and R. Breathnach. 1995. "A Crouzon syndrome synonymous mutation activates a 5' splice site within the IIIc exon of the FGFR2 gene." *Genomics* 27 (3):558-9.

Dolan, J., and K. J. Mitchell. 2013. "Mutation of Elfn1 in mice causes seizures and hyperactivity." *PLoS One* 8 (11):e80491.

Duno, M., H. Hove, M. Kirchhoff, K. Devriendt, and M. Schwartz. 2004. "Mapping genomic deletions down to the base: a quantitative copy number scanning approach used to characterise and clone the breakpoints of a recurrent 7p14.2p15.3 deletion." *Human Genet* 115 (6):459-67.

el Ghouzzi, V., M. Le Merrer, F. Perrin-Schmitt, E. Lajeunie, P. Benit, D. Renier, P. Bourgeois, A. L. Bolcato-Bellemin, A. Munnich, and J. Bonaventure. 1997. "Mutations of the TWIST gene in the Saethre-Chotzen syndrome." *Nature Genet* 15 (1):42-6.

Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison. 1994. "Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer." *Genes Dev* 8 (8):970-80.

Eswarakumar, V. P., I. Lax, and J. Schlessinger. 2005. "Cellular signaling by fibroblast growth factor receptors." *Cytokine Growth Factor Rev* 16 (2):139-49.

Everett, E. T., D. A. Britto, R. E. Ward, and J. K. Hartsfield, Jr. 1999. "A novel FGFR2 gene mutation in Crouzon syndrome associated with apparent nonpenetrance." *Cleft Palate Craniofac J* 36 (6):533-41.

Favaro, F. P., L. Alvizi, R. M. Zechi-Ceide, D. Bertola, T. M. Felix, J. de Souza, S. Raskin, S. R. Twigg, A. M. Weiner, P. Armas, E. Margarit, N. B. Calcaterra, G. R. Andersen, S. J. McGowan, A. O. Wilkie, A. Richieri-Costa, M. L. de Almeida, and M. R. Passos-Bueno. 2014. "A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects." *Am J Hum Genet* 94 (1):120-8.

Fitzgerald, T. W. , S. S. Gerety, W. D. Jones, and M. van Kogelenberg et al. 2015. "Large-scale discovery of novel genetic causes of developmental disorders." *Nature* 519 (7542):223-8.

Flanagan, S. E., W. Xie, R. Caswell, A. Damhuis, C. Vianey-Saban, T. Akcay, F. Darendeliler, F. Bas, A. Guven, Z. Siklar, G. Ocal, M. Berberoglu, N. Murphy, M. O'Sullivan, A. Green, P. E. Clayton, I. Banerjee, P. T. Clayton, K. Hussain, M. N. Weedon, and S. Ellard. 2013. "Next-generation sequencing reveals deep intronic cryptic ABCC8 and HADH splicing founder mutations causing hyperinsulinism by pseudoexon activation." *Am J Hum Genet* 92 (1):131-6.

Fluck, C. E., T. Tajima, A. V. Pandey, W. Arlt, K. Okuhara, C. F. Verge, E. W. Jabs, B. B. Mendonca, K. Fujieda, and W. L. Miller. 2004. "Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome." *Nat Genet* 36 (3):228-30.

Fu, W., T. D. O'Connor, G. Jun, H. M. Kang, G. Abecasis, S. M. Leal, S. Gabriel, M. J. Rieder, D. Altshuler, J. Shendure, D. A. Nickerson, M. J. Bamshad, and J. M. Akey. 2013. "Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants." *Nature* 493 (7431):216-20.

Fu, X.D., and M. Ares Jr. 2014. "Context-dependent control of alternative splicing by RNA-binding proteins." *Nat Rev Genet* 15 (10):689-701.

Ge, C., G. Xiao, D. Jiang, Q. Yang, N. E. Hatch, H. Roca, and R. T. Franceschi. 2009. "Identification and functional characterization of ERK/MAPK phosphorylation sites in the Runx2 transcription factor." *J Biol Chem* 284 (47):32533-43.

Gilissen, C., H. H. Arts, A. Hoischen, L. Spruijt, D. A. Mans, P. Arts, B. van Lier, M. Steehouwer, J. van Reeuwijk, S. G. Kant, R. Roepman, N. V. Knoers, J. A. Veltman, and H. G. Brunner. 2010. "Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome." *Am J Hum Genet* 87 (3):418-23.

Gilissen, C., A. Hoischen, H. G. Brunner, and J. A. Veltman. 2012. "Disease gene identification strategies for exome sequencing." *Eur J Hum Genet* 20 (5):490-7.

Gripp, K. W., C. A. Stolle, L. Celle, D. M. McDonald-McGinn, L. A. Whitaker, and E. H. Zackai. 1999. "TWIST gene mutation in a patient with radial aplasia and craniosynostosis: further evidence for heterogeneity of Baller-Gerold syndrome." *Am J Med Genet* 82 (2):170-6.

Gripp, K. W., E. H. Zackai, and C. A. Stolle. 2000. "Mutations in the human TWIST gene." *Hum Mut* 15 (2):150-5.

Guenou, H., K. Kaabeche, S. L. Mee, and P. J. Marie. 2005. "A role for fibroblast growth factor receptor-2 in the altered osteoblast phenotype induced by Twist haploinsufficiency in the Saethre-Chotzen syndrome." *Hum Mol Genet* 14 (11):1429-39.

Haack, T. B., K. Danhauser, B. Haberberger, J. Hoser, V. Strecker, D. Boehm, G. Uziel, E. Lamantea, F. Invernizzi, J. Poulton, B. Rolinski, A. Iuso, S. Biskup, T. Schmidt, H. W. Mewes, I. Wittig, T. Meitinger, M. Zeviani, and H. Prokisch. 2010. "Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency." *Nat Genet* 42 (12):1131-4.

Haines, B. P., L. M. Wheldon, D. Summerbell, J. K. Heath, and P. W. Rigby. 2006. "Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development." *Dev Biol* 297 (1):14-25.

Hamamori, Y., H. Y. Wu, V. Sartorelli, and L. Kedes. 1997. "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist." *Mol Cell Biol* 17 (11):6563-73.

Hayashi, M., K. Nimura, K. Kashiwagi, T. Harada, K. Takaoka, H. Kato, K. Tamai, and Y. Kaneda. 2007. "Comparative roles of Twist-1 and Id1 in transcriptional regulation by BMP signaling." *J Cell Sci* 120 (Pt 8):1350-7.

Holmes, G., Rothschild, G., Roy, U.B., Deng, C.X., Mansukhani, A., and Basilico, C. 2009. "Early onset of craniosynostosis in an Apert mouse model reveals critical features of this pathology." *Dev Biol* 328 (2):273-284.

Howard, T. D., W. A. Paznekas, E. D. Green, L. C. Chiang, N. Ma, R. I. Ortiz de Luna, C. Garcia Delgado, M. Gonzalez-Ramos, A. D. Kline, and E. W. Jabs. 1997. "Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome." *Nat Genet* 15 (1):36-41.

Hurst, J. A., R. M. Winter, and M. Baraitser. 1988. "Distinctive syndrome of short stature, craniosynostosis, skeletal changes, and malformed ears." *Am J Med Genet* 29 (1):107-15.

Ibrahimi, O. A., B. K. Yeh, A. V. Eliseenkova, F. Zhang, S. K. Olsen, M. Igarashi, S. A. Aaronson, R. J. Linhardt, and M. Mohammadi. 2005. "Analysis of mutations in fibroblast growth factor (FGF) and a pathogenic mutation in FGF receptor (FGFR) provides direct evidence for the symmetric two-end model for FGFR dimerization." *Mol Cell Biol* 25 (2):671-84.

Iseki, S., A. O. Wilkie, and G. M. Morriss-Kay. 1999. "Fgfr1 and Fgfr2 have distinct differentiation- and proliferation-related roles in the developing mouse skull vault." *Development* 126 (24):5611-20.

Ishii, M., A. E. Merrill, Y. S. Chan, I. Gitelman, D. P. Rice, H. M. Sucov, and R. E. Maxson, Jr. 2003. "Msx2 and Twist cooperatively control the development of the neural crest-derived skeletogenic mesenchyme of the murine skull vault." *Development* 130 (24):6131-42.

Ishimaru, M., M. Tsukimoto, H. Harada, and S. Kojima. 2013. "Involvement of P2Y(1)(1) receptor in IFN-gamma-induced IL-6 production in human keratinocytes." *Eur J Pharmacol* 703 (1-3):67-73.

Jabs, E. W., X. Li, A. F. Scott, G. Meyers, W. Chen, M. Eccles, J. I. Mao, L. R. Charnas, C. E. Jackson, and M. Jaye. 1994. "Jackson-Weiss and Crouzon syndromes are allelic with mutations in fibroblast growth factor receptor 2." *Nat Genet* 8 (3):275-9.

Jenkins, D., D. Seelow, F. S. Jehee, C. A. Perlyn, L. G. Alonso, D. F. Bueno, D. Donnai, D. Josifova, I. M. Mathijssen, J. E. Morton, K. H. Orstavik, E. Sweeney, S. A. Wall, J. L. Marsh, P. Nurnberg, M. R. Passos-Bueno, and A. O. Wilkie. 2007. "RAB23 mutations in Carpenter syndrome imply an unexpected role for hedgehog signaling in cranial-suture development and obesity." *Am J Hum Genet* 80 (6):1162-70.

Jian, X., Boerwinkle, E. and Liu, X. 2014. "In silico prediction of splice-altering single nucleotide variants in the human genome." *Nucleic Acids Res* 42(22):13534-44.

Jiang, X., S. Iseki, R. E. Maxson, H. M. Sucov, and G. M. Morriss-Kay. 2002. "Tissue origins and interactions in the mammalian skull vault." *Dev Biol* 241 (1):106-16.

Jin, F., Yan L., Dixon J.R., Selvaraj S., Ye Z., Lee A.Y., Yen C.A., Schmitt A.D., Espinoza C.A. and Ren B. 2013. "A high-resolution map of the three-dimensional chromatin interactome in human cells." *Nature* 503(7475):290-4

Johnson, D., S. W. Horsley, D. M. Moloney, M. Oldridge, S. R. Twigg, S. Walsh, M. Barrow, P. R. Njolstad, J. Kunz, G. J. Ashworth, S. A. Wall, L. Kearney, and A. O. Wilkie. 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 (5):1282-93.

Johnson, D., S. Iseki, A. O. Wilkie, and G. M. Morriss-Kay. 2000. "Expression patterns of Twist and Fgfr1, -2 and -3 in the developing mouse coronal suture suggest a key role for twist in suture initiation and biogenesis." *Mech Dev* 91 (1-2):341-5.

Johnson, D., S. A. Wall, S. Mann, and A. O. Wilkie. 2000. "A novel mutation, Ala315Ser, in FGFR2: a gene-environment interaction leading to craniosynostosis?" *Eur J Hum Genet* 8 (8):571-7.

Johnson, D., and A. O. Wilkie. 2011. "Craniosynostosis." *Eur J Hum Genet* 19 (4):369-76.

Justice, C. M., G. Yagnik, Y. Kim, I. Peter, E. W. Jabs, M. Erazo, X. Ye, E. Ainehsazan, L. Shi, M. L. Cunningham, V. Kimonis, T. Roscioli, S. A. Wall, A. O. Wilkie, J. Stoler, J. T. Richtsmeier, Y. Heuze, P. A. Sanchez-Lara, M. F. Buckley, C. M. Druschel, J. L. Mills, M. Caggana, P. A. Romitti, D. M. Kay, C. Senders, P. J. Taub, O. D. Klein, J. Boggan, M. Zwienenberg-Lee, C. Naydenov, J. Kim, A. F. Wilson, and S. A. Boyadjiev. 2012. "A genome-wide association study identifies

susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9." *Nat Genet* 44 (12):1360-4.

Kan, R., S. R. Twigg, J. Berg, L. Wang, F. Jin, and A. O. Wilkie. 2004. "Expression analysis of an FGFR2 IIIc 5' splice site mutation (1084+3A->G)." *J Med Genet* 41 (8):e108.

Katagiri, T., and N. Takahashi. 2002. "Regulatory mechanisms of osteoblast and osteoclast differentiation." *Oral Dis* 8 (3):147-59.

Kenyon, E. J., G. K. McEwen, H. Callaway, and G. Elgar. 2011. "Functional analysis of conserved non-coding regions around the short stature hox gene (shox) in whole zebrafish embryos." *PLoS One* 6 (6):e21498.

Kleinjan, D. A., A. Seawright, A. Schedl, R. A. Quinlan, S. Danes, and V. van Heyningen. 2001. "Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6." *Hum Mol Genet* 10 (19):2049-59.

Kleinjan, D. J., and V. van Heyningen. 1998. "Position effect in human genetic disease." *Hum Mol Genet* 7 (10):1611-8.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, W. S. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir, and K. Stefansson. 2012. "Rate of de novo mutations and the importance of father's age to disease risk." *Nature* 488 (7412):471-5.

Krastanova, I., V. Sannino, H. Amenitsch, O. Gileadi, F. M. Pisani, and S. Onesti. 2012. "Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases." *J Biol Chem* 287 (6):4121-8.

Krebs, I., I. Weis, M. Hudler, J. M. Rommens, H. Roth, S. W. Scherer, L. C. Tsui, E. M. Fuchtbauer, K. H. Grzeschik, K. Tsuji, and J. Kunz. 1997. "Translocation breakpoint maps 5 kb 3' from TWIST in a patient affected with Saethre-Chotzen syndrome." *Hum Mol Genet* 6 (7):1079-86.

Kress, W., H. Collmann, M. Busse, B. Halliger-Keller, and C. R. Mueller. 2000. "Clustering of FGFR2 gene mutations inpatients with Pfeiffer and Crouzon syndromes (FGFR2-associated craniosynostoses)." *Cytogenet Cell Genet* 91 (1-4):134-7.

Kristinsson, S. Y., E. T. Thorolfsdottir, B. Talseth, E. Steingrimsson, A. V. Thorsson, T. Helgason, A. B. Hreidarsson, and R. Arngrimsson. 2001. "MODY in Iceland is associated with mutations in HNF-1alpha and a novel mutation in NeuroD1." *Diabetologia* 44 (11):2098-103.

Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. 2009. "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." *Nat Protocols* 4 (8):1073-1081.

Lai, E. C. 2004. "Notch signaling: control of cell communication and cell fate." *Development* 131 (5):965-73.

Lajeunie, E., M. Le Merrer, C. Bonaiti-Pellie, D. Marchac, and D. Renier. 1995. "Genetic study of nonsyndromic coronal craniosynostosis." *Am J Med Genet* 55 (4):500-4.

Lajeunie, E., H. W. Ma, J. Bonaventure, A. Munnich, M. Le Merrer, and D. Renier. 1995. "FGFR2 mutations in Pfeiffer syndrome." *Nat Genet* 9 (2):108.

Lauderdale, J. D., J. S. Wilensky, E. R. Oliver, D. S. Walton, and T. Glaser. 2000. "3' deletions cause aniridia by preventing PAX6 gene expression." *Proc Natl Acad Sci U S A* 97 (25):13755-9.

Laursen, K. B., E. Mielke, P. Iannaccone, and E. M. Fuchtbauer. 2007. "Mechanism of transcriptional activation by the proto-oncogene Twist1." *J Biol Chem* 282 (48):34623-33.

Lee, D. Y., D. Arnott, and E. J. Brown. 2013. "Ubiquilin4 is an adaptor protein that recruits Ubiquilin1 to the autophagy machinery." *EMBO Rep* 14 (4):373-81.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. 2003. "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." *Hum Mol Genet* 12 (14):1725-35.

Lewanda, A. F., E. D. Green, J. Weissenbach, H. Jerald, E. Taylor, M. L. Summar, J. A. Phillips, 3rd, M. Cohen, M. Feingold, W. Mouradian, and et al. 1994. "Evidence that the Saethre-Chotzen syndrome locus lies between D7S664 and D7S507, by genetic analysis and detection of a microdeletion in a patient." *Am J Hum Genet* 55 (6):1195-201.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25 (16):2078-9.

Li, X., W. J. Park, R. E. Pyeritz, and E. W. Jabs. 1995. "Effect on splicing of a silent FGFR2 mutation in Crouzon syndrome." *Nat Genet* 9 (3):232-3.

Lunter, G., and M. Goodson. 2011. "Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads." *Genome Res* 21 (6):936-9.

Lynch, M. 2010. "Rate, molecular spectrum, and consequences of human mutation." *Proc Natl Acad Sci U S A* 107 (3):961-8.

MacArthur, J. A., T. D. Spector, S. J. Lindsay, M. Mangino, R. Gill, K. S. Small, and M. E. Hurles. 2014. "The rate of nonallelic homologous recombination in males is highly variable, correlated between monozygotic twins and independent of age." *PLoS Genet* 10 (3):e1004195.

Maeno, T., T. Moriishi, C. A. Yoshida, H. Komori, N. Kanatani, S. Izumi, K. Takaoka, and T. Komori. 2011. "Early onset of Runx2 expression caused craniosynostosis, ectopic bone formation, and limb defects." *Bone* 49 (4):673-82.

Malecki, M. T., U. S. Jhala, A. Antonellis, L. Fields, A. Doria, T. Orban, M. Saad, J. H. Warram, M. Montminy, and A. S. Krolewski. 1999. "Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus." *Nat Genet* 23 (3):323-8.

Marchegiani, S., T. Davis, F. Tessadori, G. van Haaften, F. Brancati, A. Hoischen, H. Huang, E. Valkanas, B. Pusey, D. Schanze, H. Venselaar, A. T. Vulto-van Silfhout, L. A. Wolfe, C. J. Tifft, P. M. Zerfas, G. Zambruno, A. Kariminejad, F. Sabbagh-Kermani, J. Lee, M. G. Tsokos, C. C. Lee, V. Ferraz, E. M. da Silva, C. A. Stevens, N. Roche, O. Bartsch, P. Farndon, E. Bermejo-Sanchez, B. P. Brooks, V. Maduro, B. Dallapiccola, F. J. Ramos, H. Y. Chung, C. Le Caignec, F. Martins, W. K. Jacyk, L. Mazzanti, H. G. Brunner, J. Bakkers, S. Lin, M. C. Malicdan, C. F. Boerkoel, W. A. Gahl, B. B. de Vries, M. M. van Haelst, M. Zenker, and T. C. Markello. 2015. "Recurrent Mutations in the Basic Domain of TWIST2 Cause Ablepharon Macrostomia and Barber-Say Syndromes." *Am J Hum Genet* 97 (1):99-110.

Massari, M. E., and C. Murre. 2000. "Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms." *Mol Cell Biol* 20 (2):429-40.

Mazzanti, L., R. Bergamaschi, I. Neri, A. Perri, A. Patrizi, E. Cacciari, and A. Forabosco. 1998. "Barber-Say Syndrome: report of a new case." *Am J Med Genet* 78 (2):188-91.

Mefford, H. C., N. Shafer, F. Antonacci, J. M. Tsai, S. S. Park, A. V. Hing, M. J. Rieder, M. D. Smyth, M. L. Speltz, E. E. Eichler, and M. L. Cunningham. 2010. "Copy number variation analysis in single-suture craniosynostosis: multiple rare variants including RUNX2 duplication in two cousins with metopic craniosynostosis." *Am J Med Genet A* 152A (9):2203-10.

Merrill, A. E., E. G. Bochukova, S. M. Brugger, M. Ishii, D. T. Pilz, S. A. Wall, K. M. Lyons, A. O. Wilkie, and R. E. Maxson, Jr. 2006. "Cell mixing at a neural crest-mesoderm boundary and deficient ephrin-Eph signaling in the pathogenesis of craniosynostosis." *Hum Mol Genet* 15 (8):1319-28.

Merrill, A. E., A. Sarukhanov, P. Krejci, B. Idoni, N. Camacho, K. D. Estrada, K. M. Lyons, H. Deixler, H. Robinson, D. Chitayat, C. J. Curry, R. S. Lachman, W. R. Wilcox, and D. Krakow. 2012. "Bent bone dysplasia-FGFR2 type, a distinct skeletal disorder, has deficient canonical FGF signaling." *Am J Hum Genet* 90 (3):550-7.

Moir, D., S. E. Stewart, B. C. Osmond, and D. Botstein. 1982. "Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies." *Genetics* 100 (4):547-63.

Morriss-Kay, G. M., and A. O. Wilkie. 2005. "Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies." *J Anat* 207 (5):637-53.

Moyer, S. E., P. W. Lewis, and M. R. Botchan. 2006. "Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase." *Proc Natl Acad Sci U S A* 103 (27):10236-41.

Muenke, M., U. Schell, A. Hehr, N. H. Robin, H. W. Losken, A. Schinzel, L. J. Pulleyn, P. Rutland, W. Reardon, S. Malcolm, and et al. 1994. "A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome." *Nat Genet* 8 (3):269-74.

Mundlos, S., F. Otto, C. Mundlos, J. B. Mulliken, A. S. Aylsworth, S. Albright, D. Lindhout, W. G. Cole, W. Henn, J. H. Knoll, M. J. Owen, R. Mertelsmann, B. U. Zabel, and B. R. Olsen. 1997. "Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia." *Cell* 89 (5):773-9.

Murray, S. S., C. A. Glackin, K. A. Winters, D. Gazit, A. J. Kahn, and E. J. Murray. 1992. "Expression of helix-loop-helix regulatory genes during differentiation of mouse osteoblastic cells." *J Bone Miner Res* 7 (10):1131-8.

Murre, Cornelis, Patrick Schonleber McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, Carlos V. Cabrera, Jean N. Buskin, Stephen D. Hauschka, Andrew B. Lassar, Harold Weintraub, and David Baltimore. 1989. "Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence." *Cell* 58 (3):537-544.

Nah, H. D., E. Koyama, N. B. Agochukwu, S. P. Bartlett, and M. Muenke. 2012. "Phenotype profile of a genetic mouse model for Muenke syndrome." *Childs Nerv Syst* 28 (9):1483-93.

Ng, S. B., A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad, and J. Shendure. 2010. "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." *Nat Genet* 42 (9):790-3.

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure, and M. J. Bamshad. 2010. "Exome sequencing identifies the cause of a mendelian disorder." *Nat Genet* 42 (1):30-5.

Opperman, L. A. 2000. "Cranial sutures as intramembranous bone growth sites." *Dev Dyn* 219 (4):472-85.

Otis, Eileen M., and Robert Brent. 1954. "Equivalent ages in mouse and human embryos." *Anat Rec* 120 (1):33-63.

Panigrahi, I. 2011. "Craniosynostosis genetics: The mystery unfolds." *Indian J Hum Genet* 17 (2):48-53.

Patel, P., Reardon, W., Malcolm, S., and Winter, R.M. 1998. "Translocation breakpoints mapped approximately 70-250 Kb from the TWIST gene in three patients with Saethre-Chotzen Syndrome" J Med Genet Supplement 1 35:12.16

Piard, J., B. Aral, P. Vabres, M. Holder-Espinasse, A. Megarbane, S. Gauthier, V. Capra, G. Pierquin, P. Callier, C. Baumann, L. Pasquier, G. Baujat, L. Martorell, A. Rodriguez, A. F. Brady, F. Boralevi, M. A. Gonzalez-Ensenat, M. Rio, C. Bodemer, N. Philip, M. P. Cordier, A. Goldenberg, B. Demeer, M. Wright, E. Blair, E. Puzenat, P. Parent, Y. Sznajer, C. Francannet, N. DiDonato, O. Boute, V. Barlogis, O. Moldovan, D. Bessis, C. Coubes, M. Tardieu, V. Cormier-Daire, A. B. Sousa, J. Franques, A. Toutain, M. Tajir, S. C. Elalaoui, D. Genevieve, J. Thevenon, J. B. Courcet, J. B. Riviere, C. Collet, N. Gigot, L. Faivre, and C. Thauvin-Robinet. 2015. "Search for ReCQL4 mutations in 39 patients genotyped for suspected Rothmund-Thomson/Baller-Gerold syndromes." *Clin Genet* 87 (3):244-51.

Piard, J., V. Roze, A. Czorny, M. Lenoir, M. Valduga, A. L. Fenwick, A. O. Wilkie, and L. V. Maldergem. 2015. "TCF12 microdeletion in a 72-year-old woman with intellectual disability." *Am J Med Genet A* 167 (8):1897-901.

Plotnikov, A. N., S. R. Hubbard, J. Schlessinger, and M. Mohammadi. 2000. "Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity." *Cell* 101 (4):413-24.

Podkrajsek, K. T., T. Milenkovic, R. J. Odink, H. L. Claasen-van der Grinten, N. Bratanic, T. Hovnik, and T. Battelino. 2008. "Detection of a complete autoimmune regulator gene deletion and two additional novel mutations in a cohort of patients with atypical phenotypic variants of autoimmune polyglandular syndrome type 1." *Eur J Endocrinol* 159 (5):633-9.

Pollok, S., C. Bauerschmidt, J. Sanger, H. P. Nasheuer, and F. Grosse. 2007. "Human Cdc45 is a proliferation-associated antigen." *FEBS J* 274 (14):3669-84.

Pope, A. W., and M. L. Speltz. 1997. "Research of psychosocial issues of children with craniofacial anomalies: progress and challenges." *Cleft Palate Craniofac J* 34 (5):371-3.

Price, M., D. Lazzaro, T. Pohl, M. G. Mattei, U. Ruther, J. C. Olivo, D. Duboule, and R. Di Lauro. 1992. "Regional expression of the homeobox gene Nkx-2.2 in the developing mammalian forebrain." *Neuron* 8 (2):241-55.

Przylepa, K. A., W. Paznekas, M. Zhang, M. Golabi, W. Bias, M. J. Bamshad, J. C. Carey, B. D. Hall, R. Stevenson, S. Orlow, M. M. Cohen, Jr., and E. W. Jabs. 1996. "Fibroblast growth factor receptor 2 mutations in Beare-Stevenson cutis gyrata syndrome." *Nat Genet* 13 (4):492-4.

Reardon, W., S. P. McManus, D. Summers, and R. M. Winter. 1993. "Cytogenetic evidence that the Saethre-Chotzen gene maps to 7p21.2." *Am J Med Genet* 47 (5):633-6.

Reardon, W., R. M. Winter, P. Rutland, L. J. Pulleyn, B. M. Jones, and S. Malcolm. 1994. "Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome." *Nat Genet* 8 (1):98-103.

Reid, C. S., L. E. McMorrow, D. M. McDonald-McGinn, K. J. Grace, F. J. Ramos, E. H. Zackai, M. M. Cohen, Jr., and E. W. Jabs. 1993. "Saethre-Chotzen syndrome with familial translocation at chromosome 7p22." *Am J Med Genet* 47 (5):637-9.

Rice, D.P. 2008. " Craniofacial sutures development, disease and treatment." In *Frontiers of Oral Biology*, edited by Paul Sharpe, 1-21. Karger.

Rice, D. P., T. Aberg, Y. Chan, Z. Tang, P. J. Kettunen, L. Pakarinen, R. E. Maxson, and I. Thesleff. 2000. "Integration of FGF and TWIST in calvarial bone and suture development." *Development* 127 (9):1845-55.

Rimmer, Andy, Hang Phan, Iain Mathieson, Zamin Iqbal, Stephen R. F. Twigg, W. G. S. Consortium, Andrew O. M. Wilkie, Gil McVean, and Gerton Lunter. 2014. "Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications." *Nat Genet* 46 (8):912-918.

Roach, J. C., G. Glusman, A. F. Smit, C. D. Huff, R. Hubley, P. T. Shannon, L. Rowen, K. P. Pant, N. Goodman, M. Bamshad, J. Shendure, R. Drmanac, L. B. Jorde, L. Hood, and D. J. Galas. 2010. "Analysis of genetic inheritance in a family quartet by whole-genome sequencing." *Science* 328 (5978):636-9.

Robertson, S. C., A. N. Meyer, K. C. Hart, B. D. Galvin, M. K. Webster, and D. J. Donoghue. 1998. "Activating mutations in the extracellular domain of the fibroblast growth factor receptor 2 function by disruption of the disulfide bond in the third immunoglobulin-like domain." *Proc Nat Acad Sci U S A* 95 (8):4567-72.

Roessler, E., E. Belloni, K. Gaudenz, P. Jay, P. Berta, S. W. Scherer, L. C. Tsui, and M. Muenke. 1996. "Mutations in the human Sonic Hedgehog gene cause holoprosencephaly." *Nat Genet* 14 (3):357-60.

Rose, C. S., P. Patel, W. Reardon, S. Malcolm, and R. M. Winter. 1997. "The TWIST gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases." *Hum Mol Genet* 6 (8):1369-73.

Rumsey, N., and D. Harcourt. 2007. "Visible difference amongst children and adolescents: issues and interventions." *Dev Neurorehabil* 10 (2):113-23.

Rutland, P., L. J. Pulleyn, W. Reardon, M. Baraitser, R. Hayward, B. Jones, S. Malcolm, R. M. Winter, M. Oldridge, S. F. Slaney, and et al. 1995. "Identical mutations in the FGFR2 gene cause both Pfeiffer and Crouzon syndrome phenotypes." *Nature genetics* 9 (2):173-6.

Sabherwal, N., F. Bangs, R. Roth, B. Weiss, K. Jantz, E. Tiecke, G. K. Hinkel, C. Spaich, B. P. Hauffa, H. van der Kamp, J. Kapeller, C. Tickle, and G. Rappold. 2007. "Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients." *Hum Mol Genet* 16 (2):210-22.

Sanchez-Pulido, L., and C. P. Ponting. 2011. "Cdc45: the missing RecJ ortholog in eukaryotes?" *Bioinformatics* 27 (14):1885-8.

Schell, U., A. Hehr, G. J. Feldman, N. H. Robin, E. H. Zackai, C. de Die-Smulders, D. H. Viskochil, J. M. Stewart, G. Wolff, H. Ohashi, and et al. 1995. "Mutations in FGFR1 and FGFR2 cause familial and sporadic Pfeiffer syndrome." *Hum Mol Genet* 4 (3):323-8.

Scherer, S. W., J. Cheung, J. R. MacDonald, L. R. Osborne, K. Nakabayashi, J. A. Herbrick, A. R. Carson, L. Parker-Katiraee, J. Skaug, R. Khaja, J. Zhang, A. K. Hudek, M. Li, M. Haddad, G. E. Duggan, B. A. Fernandez, E. Kanematsu, S. Gentles, C. C. Christopoulos, S. Choufani, D. Kwasnicka, X. H. Zheng, Z. Lai, D. Nusskern, Q. Zhang, Z. Gu, F. Lu, S. Zeesman, M. J. Nowaczyk, I. Teshima, D. Chitayat, C. Shuman, R. Weksberg, E. H. Zackai, T. A. Grebe, S. R. Cox, S. J. Kirkpatrick, N. Rahman, J. M. Friedman, H. H. Heng, P. G. Pelicci, F. Lo-Coco, E. Belloni, L. G. Shaffer, B. Pober, C. C. Morton, J. F. Gusella, G. A. Bruns, B. R. Korf, B. J. Quade, A. H. Ligon, H. Ferguson, A. W. Higgins, N. T. Leach, S. R. Herrick, E. Lemyre, C. G. Farra, H. G. Kim, A. M. Summers, K. W. Gripp, W. Roberts, P. Szatmari, E. J. Winsor, K. H. Grzeschik, A. Teebi, B. A. Minassian, J. Kere, L. Armengol, M. A. Pujana, X. Estivill, M. D. Wilson, B. F. Koop, S. Tosi, G. E. Moore, A. P. Boright, E. Zlotorynski, B. Kerem, P. M. Kroisel, E. Petek, D. G. Oscier, S. J. Mould, H. Dohner, K. Dohner, J. M. Rommens, J. B. Vincent, J. C. Venter, P. W. Li, R. J. Mural, M. D. Adams, and L. C. Tsui. 2003. "Human chromosome 7: DNA sequence and biology." *Science* 300 (5620):767-72.

Schmitz, R., Young R.M., Ceribelli M., Jhavar S., Xiao W., Zhang M., Wright G., Shaffer A.L., Hodson D.J., Buras E., Liu X., Powell J, Yang Y., Xu W., Zhao H., Kohlhammer H., Rosenwald A., Kluin P., Müller-Hermelink H.K., Ott G., Gascoyne R.D., Connors J.M., Rimsza L.M., Campo E., Jaffe E.S., Delabie J., Smeland E.B., Ogwang M.D., Reynolds S.J., Fisher R.I., Braziel R.M., Tubbs R.R., Cook J.R., Weisenburger D.D., Chan W.C., Pittaluga S., Wilson W., Waldmann T.A., Rowe M., Mbulaiteye S.M., Rickinson A.B., and L.M. Staudt.. 2012. "Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics." *Nature* 490 (7418):116-120.

Schwarz, J.M., Rodelsperger, C., Schuelke, M., and D. Seelow. 2010. "MutationTaster evaluates disease-causing potential of sequence alterations." *Nat Meth* 7 (8):575-576.

Seto, M. L., S. J. Lee, R. W. Sze, and M. L. Cunningham. 2001. "Another TWIST on Baller-Gerold syndrome." *Am J Med Genet* 104 (4):323-30.

Shaikh, T. H., S. Gottlieb, B. Sellinger, F. Chen, B. A. Roe, R. J. Oakey, B. S. Emanuel, and M. L. Budarf. 1999. "Characterization of CDC45L: a gene in the 22q11.2 deletion region expressed during murine and human development." *Mamm Genome* 10 (3):322-6.

Sharma, V. P., A. L. Fenwick, M. S. Brockop, S. J. McGowan, J. A. Goos, A. J. Hoogeboom, A. F. Brady, N. O. Jeelani, S. A. Lynch, J. B. Mulliken, D. J. Murray, J. M. Phipps, E. Sweeney, S. E. Tomkins, L. C. Wilson, S. Bennett, R. J. Cornall, J. Broxholme, A. Kanapin, D. Johnson, S. A. Wall, P. J. van der Spek, I. M. Mathijssen, R. E. Maxson, S. R. Twigg, and A. O. Wilkie. 2013. "Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis." *Nat Genet* 45 (3):304-7.

Siepel, A., Pollard, K.S., and Haussler, D. 2006. "New Methods for Detecting Lineage-Specific Selection." In *Research in Computational Molecular Biology*, edited by Alberto Apostolico, Concettina Guerra, Sorin Istrail, PavelA Pevzner and Michael Waterman, 190-205. Springer Berlin Heidelberg.

Soriano, P. 1999. "Generalized lacZ expression with the ROSA26 Cre reporter strain." *Nat Genet* 21 (1):70-1.

Spicer, D. B., J. Rhee, W. L. Cheung, and A. B. Lassar. 1996. "Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist." *Science* 272 (5267):1476-80.

Stankiewicz, P., H. Thiele, C. Baldermann, A. Kruger, I. Giannakudis, S. Dorr, N. Werner, J. Kunz, G. A. Rappold, and I. Hansmann. 2001. "Phenotypic findings due to trisomy 7p15.3-pter including the TWIST locus." *Am J Med Genet* 103 (1):56-62.

Steininger, A., M. Mobs, R. Ullmann, K. Kochert, S. Kreher, B. Lamprecht, I. Anagnostopoulos, M. Hummel, J. Richter, M. Beyer, M. Janz, C. D. Klemke, H. Stein, B. Dorken, W. Sterry, E. Schrock, S. Mathas, and C. Assaf. 2011. "Genomic loss of the putative tumor suppressor gene E2A in human lymphoma." *J Exp Med* 208 (8):1585-93.

Stuppia, L., I. Antonucci, G. Palka, and V. Gatta. 2012. "Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases." *Int J Mol Sci* 13 (3):3245-76.

Takeuchi, A., M. Hosokawa, T. Nojima, and M. Hagiwara. 2010. "Splicing reporter mice revealed the evolutionally conserved switching mechanism of tissue-specific alternative exon selection." *PLoS One* 5 (6):e10946.

Tang, M.K., Y.J. Liang, J.Y. Chan, S. W. Wong, E. Chen, Y. Yao, J. Gan, L. Xiao, H. C. Leung, H. F. Kung, H. Wang, and K. K. Lee. 2013. "Promyelocytic leukemia (PML) protein plays important roles in regulating cell adhesion, morphology, proliferation and migration." *PLoS One* 8 (3):e59477.

Taylor, J.C., Martin H.C., Lise S., Broxholme J., Cazier J.B., Rimmer A., Kanapin A., Lunter G., Fiddy S., Allan C., Aricescu A.R., Attar M., Babbs C., Becq J., Beeson D., Bento C., Bignell P., Blair E., Buckle V.J., Bull K., Cais O., Cario H., Chapel H., Copley R.R., Cornall R., Craft J., Dahan K., Davenport E.E., Dendrou C., Devuyst O., Fenwick A.L., Flint J., Fugger L., Gilbert R.D., Goriely A., Green A., Greger I.H., Grocock R., Gruszczyk A.V., Hastings R., Hatton E., Higgs D., Hill A., Holmes C., Howard M., Hughes L., Humburg P., Johnson D., Karpe F., Kingsbury Z., Kini U., Knight J.C., Krohn J., Lamble S., Langman C., Lonie L., Luck J., McCarthy D., McGowan S.J., McMullin M.F., Miller K.A., Murray L., Németh A.H., Nesbit M.A., Nutt D., Ormondroyd E., Oturai A.B., Pagnamenta A., Patel S.Y., Percy M., Petousi N., Piazza P., Piret S.E., Polanco-Echeverry G., Popitsch N., Powrie F., Pugh C., Quek L., Robbins P.A., Robson K., Russo A., Sahgal N., van Schouwenburg P.A., Schuh A., Silverman E., Simmons A., Sørensen P.S., Sweeney E., Taylor J., Thakker R.V., Tomlinson I., Trebes A., Twigg S.R., Uhlig H.H., Vyas P., Vyse T., Wall S.A., Watkins H., Whyte M.P., Witty L., Wright B., Yau C., Buck D., Humphray S., Ratcliffe P.J., Bell J.I., Wilkie A.O., Bentley D., Donnelly P. and McVean G. 2015. “Factors influencing success of clinical genome sequencing across a broad spectrum of disorders.” *Nat Genet* 47(7):717-26.

The UK10K Consortium. 2015. "The UK10K project identifies rare variants in health and disease." *Nature* 526(7571):82-90.

Thisse, B., M. el Messal, and F. Perrin-Schmitt. 1987. "The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern." *Nucleic Acids Res* 15 (8):3439-53.

Ting, M. C., N. L. Wu, P. G. Roybal, J. Sun, L. Liu, Y. Yen, and R. E. Maxson, Jr. 2009. "EphA4 as an effector of Twist1 in the guidance of osteogenic precursor cells during calvarial bone growth and in craniosynostosis." *Development* 136 (5):855-64.

Trainor, P. A. 2010. "Craniofacial birth defects: The role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention." *Am J Med Genet A* 152a (12):2984-94.

Tsuji, K., K. Narahara, K. Kikkawa, M. Murakami, Y. Yokoyama, S. Ninomiya, and Y. Seino. 1994. "Craniosynostosis and hemizygosity for D7S135 caused by a de novo and apparently balanced t(6;7) translocation." *Am J Med Genet* 49 (1):98-102.

Tsuji, K., K. Narahara, Y. Yokoyama, K. H. Grzeschik, and J. Kunz. 1995. "The breakpoint on 7p in a patient with t(6;7) and craniosynostosis is spanned by a YAC clone containing the D7S503 locus." *Hum Genet* 95 (3):303-7.

Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S., and M.E. Hurles. 2008. "Germline rates of de novo meiotic deletions and duplications causing several genomic disorders." *Nat Genet* 40 (1):90-95.

Twigg, S. R., C. Babbs, M. E. van den Elzen, A. Goriely, S. Taylor, S. J. McGowan, E. Giannoulatou, L. Lonie, J. Ragoussis, E. Sadighi Akha, S. J. Knight, R. M. Zechi-Ceide, J. A. Hoogeboom, B. R. Pober, H. V. Toriello, S. A. Wall, M. Rita Passos-Bueno, H. G. Brunner, I. M. Mathijssen, and A. O. Wilkie. 2013. "Cellular interference in craniofrontonasal syndrome: males mosaic for mutations in the X-linked EFNB1 gene are more severely affected than true hemizygotes." *Hum Mol Genet* 22 (8):1654-62.

Twigg, S. R., R. Kan, C. Babbs, E. G. Bochukova, S. P. Robertson, S. A. Wall, G. M. Morriss-Kay, and A. O. Wilkie. 2004. "Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation, cause craniofrontonasal syndrome." *Proc Nat Acad Sci U S A* 101 (23):8652-7.

Twigg, S. R., D. Lloyd, D. Jenkins, N. E. Elcioglu, C. D. Cooper, N. Al-Sannaa, A. Annagur, G. Gillessen-Kaesbach, I. Huning, S. J. Knight, J. A. Goodship, B. D. Keavney, P. L. Beales, O. Gileadi, S. J. McGowan, and A. O. Wilkie. 2012. "Mutations in Multidomain Protein MEGF8 Identify a Carpenter Syndrome Subtype Associated with Defective Lateralization." *Am J Hum Genet*.

Twigg, S. R., E. Vorgia, S. J. McGowan, I. Peraki, A. L. Fenwick, V. P. Sharma, M. Allegra, A. Zaragkoulias, E. Sadighi Akha, S. J. Knight, H. Lord, T. Lester, L. Izatt, A. K. Lampe, S. N. Mohammed, F. J. Stewart, A. Verloes, L. C. Wilson, C. Healy, P. T. Sharpe, P. Hammond, J. Hughes, S. Taylor, D. Johnson, S. A. Wall, G. Mavrothalassitis, and A. O. Wilkie. 2013. "Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis." *Nat Genet* 45 (3):308-13.

Varvagiannis, K., A. Stefanidou, Y. Gyftodimou, H. Lord, L. Williams, C. Sarri, E. Pandelia, E. Bazopoulou-Kyrkanidou, C. Noakes, T. Lester, A. O. Wilkie, and M. B. Petersen. 2013. "Pure de novo partial trisomy 6p in a girl with craniosynostosis." *Am J Med Genet A* 161a (2):343-51.

Vissers, L. E., J. de Ligt, C. Gilissen, I. Janssen, M. Steehouwer, P. de Vries, B. van Lier, P. Arts, N. Wieskamp, M. del Rosario, B. W. van Bon, A. Hoischen, B. B. de Vries, H. G. Brunner, and J. A. Veltman. 2010. "A de novo paradigm for mental retardation." *Nat Genet* 42 (12):1109-12.

Wang, K., M. Li, and H. Hakonarson. 2010. "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." *Nucleic Acids Res* 38 (16):e164. doi: 10.1093/nar/gkq603.

Wang, S. M., P. D. Phillips, F. Sierra, and V. J. Cristofalo. 1996. "Altered expression of the twist gene in young versus senescent human diploid fibroblasts." *Exp Cell Res* 228 (1):138-45.

Warzecha, C. C., T. K. Sato, B. Nabet, J. B. Hogenesch, and R. P. Carstens. 2009. "ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing." *Mol Cell* 33 (5):591-601.

Wee, H. J., G. Huang, K. Shigesada, and Y. Ito. 2002. "Serine phosphorylation of RUNX2 with novel potential functions as negative regulatory mechanisms." *EMBO Rep* 3 (10):967-74.

Weinzweig, J., R. E. Kirschner, A. Farley, P. Reiss, J. Hunter, L. A. Whitaker, and S. P. Bartlett. 2003. "Metopic synostosis: Defining the temporal sequence of normal suture fusion and differentiating it from synostosis on the basis of computed tomography images." *Plast Reconstr Surg* 112 (5):1211-8.

Wheldon, L. M., B. P. Haines, R. Rajappa, I. Mason, P. W. Rigby, and J. K. Heath. 2010. "Critical role of FLRT1 phosphorylation in the interdependent regulation of FLRT1 function and FGF receptor signalling." *PloS one* 5 (4):e10264.

Wieczorek, D., W. G. Newman, T. Wieland, T. Berulava, M. Kaffe, D. Falkenstein, C. Beetz, E. Graf, T. Schwarzmayr, S. Douzgou, J. Clayton-Smith, S. B. Daly, S. G. Williams, S. S. Bhaskar, J. E. Urquhart, B. Anderson, J. O'Sullivan, O. Boute, J. Gundlach, J. C. Czeschik, A. J. van Essen, F. Hazan, S. Park, A. Hing, A. Kuechler, D. R. Lohmann, K. U. Ludwig, E. Mangold, L. Steenpass,

M. Zeschnigk, J. R. Lemke, C. M. Lourenco, U. Hehr, E. C. Prott, M. Waldenberger, A. C. Bohmer, B. Horsthemke, R. T. O'Keefe, T. Meitinger, J. Burn, H. J. Ludecke, and T. M. Strom. 2014. "Compound heterozygosity of low-frequency promoter deletions and rare loss-of-function mutations in TXNL4A causes Burn-McKeown syndrome." *Am J Hum Genet* 95 (6):698-707.

Wieland, I., R. Makarov, W. Reardon, S. Tinschert, A. Goldenberg, P. Thierry, and P. Wieacker. 2008. "Dissecting the molecular mechanisms in craniofrontonasal syndrome: differential mRNA expression of mutant EFNB1 and the cellular mosaic." *Eur J Hum Genet* 16 (2):184-91.

Wilkie, A. O. 1997. "Craniosynostosis: genes and mechanisms." *Hum Mol Genet* 6 (10):1647-56.

Wilkie, A. O., E. G. Bochukova, R. M. Hansen, I. B. Taylor, S. V. Rannan-Eliya, J. C. Byren, S. A. Wall, L. Ramos, M. Venancio, J. A. Hurst, W. O'Rourke A, L. J. Williams, A. Seller, and T. Lester. 2007. "Clinical dividends from the molecular genetic diagnosis of craniosynostosis." *Am J Med Genet A* 143A (16):1941-9.

Wilkie, A. O., J. C. Byren, J. A. Hurst, J. Jayamohan, D. Johnson, S. J. Knight, T. Lester, P. G. Richards, S. R. Twigg, and S. A. Wall. 2010. "Prevalence and complications of single-gene and chromosomal disorders in craniosynostosis." *Pediatrics* 126 (2):e391-400.

Wilkie, A. O., S. F. Slaney, M. Oldridge, M. D. Poole, G. J. Ashworth, A. D. Hockley, R. D. Hayward, D. J. David, L. J. Pulleyn, P. Rutland, and et al. 1995. "Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome." *Nat Genet* 9 (2):165-72.

Wilkie, A. O., S. P. Yang, D. Summers, M. D. Poole, W. Reardon, and R. M. Winter. 1995. "Saethre-Chotzen syndrome associated with balanced translocations involving 7p21: three further families." *J Med Genet* 32 (3):174-80.

Woods, R. H., E. Ul-Haq, A. O. Wilkie, J. Jayamohan, P. G. Richards, D. Johnson, T. Lester, and S. A. Wall. 2009. "Reoperation for intracranial hypertension in TWIST1-confirmed Saethre-Chotzen syndrome: a 15-year review." *Plast Reconstr Surg* 123 (6):1801-10.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler, and Z. Ning. 2009. "Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads." *Bioinformatics* 25 (21):2865-71.

Yen, H. Y., M. C. Ting, and R. E. Maxson. 2010. "Jagged1 functions downstream of Twist1 in the specification of the coronal suture and the formation of a boundary between osteogenic and non-osteogenic cells." *Dev Biol* 347 (2):258-70.

Yeo, G., and C. B. Burge. 2004. "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." *J Comput Biol* 11 (2-3):377-94.

Yilmaz, S., T. Turhan, S. Mutluer, and S. Aydogdu. 2013. "The association of Alagille syndrome and craniosynostosis." *Pediatr Neurol* 48 (2):146-8. doi: 10.1016/j.pediatrneurol.2012.10.014.

Yoshida, K., F. Kuo, E. L. George, A. H. Sharpe, and A. Dutta. 2001. "Requirement of CDC45 for postimplantation mouse development." *Mol Cell Biol* 21 (14):4598-603. d

Yoshida, T., P. Vivatbutsiri, G. Morriss-Kay, Y. Saga, and S. Iseki. 2008. "Cell lineage in mammalian craniofacial mesenchyme." *Mech Dev* 125 (9-10):797-808.

Zhang, X. H., and L. A. Chasin. 2004. "Computational definition of sequence motifs governing constitutive exon splicing." *Genes Dev* 18 (11):1241-50.

Zhang, X., O. A. Ibrahimi, S. K. Olsen, H. Umemori, M. Mohammadi, and D. M. Ornitz. 2006. "Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family." *J Biol Chem* 281 (23):15694-700.

Zhuang, Y., P. Cheng, and H. Weintraub. 1996. "B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB." *Mol Cell Biol* 16 (6):2898-905.