Genetics and Pathophysiology of Coronal Craniosynostosis revealed by Next-Generation DNA Sequencing

Vikram Pramod Sharma

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2015

Thesis Title: **Genetics and Pathophysiology of Coronal Craniosynostosis revealed by Next-Generation DNA Sequencing**

Candidate: **Mr Vikram Pramod Sharma**

College: **Green Templeton**

Degree submitted for: **Doctor of Philosophy (DPhil) in Clinical Laboratory Sciences**

Term submitted: **Trinity**

Year submitted: **2015**

Thesis Abstract

This thesis further delineates the molecular genetic basis of a relatively common craniofacial condition, coronal craniosynostosis. It used whole-exome sequencing to identify novel disease genes in patients with non-syndromic coronal synostosis and negative genetic testing.

Initially, 2 patients were identified with damaging, frameshift mutations in a gene not previously linked with craniosynostosis – Transcription Factor 12 (*TCF12*). A further intronic mutation was identified in a third patient. This gene encodes a transcription factor that dimerises with TWIST1, mutations of which cause Saethre-Chotzen syndrome, also associated with coronal synostosis. Screening 344 undiagnosed patients identified 35 further mutations, all with coronal synostosis with 14 cases arising *de novo*. This work was published and testing for *TCF12*-related craniosynostosis was translated clinically.

Significant non-penetrance (60%) was identified in mutation-positive relatives and the genetic background was investigated. Firstly, analysis of parental origins of *de novo* mutations identified 6 of paternal origin and helped refine haplotype assignment. Secondly, haplotype analysis of *TCF12*-mutation carriers revealed modest correlation with phenotypic status, but this was insufficient to be useful in clinical testing. Thirdly, *TCF12* haplotypes were analysed for association with non-syndromic coronal synostosis, but no significant association was found.

Further exome sequencing revealed a *de novo* frameshift mutation in Transcription Factor 20 (*TCF20)* in a patient with coronal synostosis and autism, although the mutation only correlated with the latter phenotype. Analysis of 5 trios revealed a novel variant in myosin heavy chain 4 (*MYH4*) in 1 family, although its role in suture development is uncertain. Reviewing pooled exome data from 19 mutation-negative patients revealed no further disease genes.

In summary, this thesis describes novel gene discovery, defines a new clinical entity and investigates genetic background of penetrant and non-penetrant individuals. Further exome sequencing identified another disease gene, a *de novo* mutation and compiled lists of damaging variants to allow future work.

ACKNOWLEDGEMENTS

The work contained within this thesis would not have been possible without an All-star team to help me achieve my ambitions.

Firstly, my sincere thanks and eternal gratitude go to my primary supervisor, Prof Andrew Wilkie. He is an outstanding Clinician-Scientist and a true world leader in the field. He has provided me with exceptional guidance, teaching and mentorship throughout my time in his laboratory. Without his close support, none of this would have been possible - thank you Andrew! I am also indebted and very thankful to Dr Stephen Twigg, who has trained me (from the very beginning) in all manner of wet and dry lab techniques and shared many pearls of molecular biological wisdom with me on a daily basis - thanks Steve!

I have also received fantastic help from the other scientists in the lab, and I am very grateful to Prof Anne Goriely, Dr Geoffrey Maher, Dr Deborah Lloyd, Ms Aimee Fenwick, Mrs Indira Taylor and Dr Akiko Hashimoto. Thanks also go to my close colleagues, Dr Chris Babbs, Mr Nigel Roberts, Dr Veronica Buckle, Ms Jill Brown and Mr Kevin Clark. A special thanks also to Ms Julie Phipps, our wonderful research nurse, for all her help with samples and research home visits – I couldn't have done it without you Julie!

My sincere thanks also go to all the members of the Oxford Craniofacial Unit, especially Mr Steven Wall, who has provided excellent additional supervision and great insights from his vast craniofacial experience to help guide my scientific investigations. My further gratitude goes to Mr David Johnson, Dr Deidre Cilliers, Dr Jo Byren, Mrs Helen Williams, Mrs Sue

Greenwood and Mrs Kate Burden for all their help with my research patients. I am also thankful for help from collaborating units in Birmingham (Mr Hiroshi Nishikawa, Mr Nicholas White, Mr Peter Noons, Dr Jenny Mortimer, London (Mr Jonathan Britto, Mr David Dunaway, Mr Owase Jeelani, Dr Louise Wilson) Liverpool (Mr Christian Duncan, Dr Elizabeth Sweeney) and Rotterdam (Prof Irene Mathijssen and Dr Jacqueline Goos).

I am very grateful to all the patients and families that have participated enthusiastically participated in this research.

This work was kindly supported by: Children's Craniofacial Charity/John Radcliffe Hospital, Oxford, The Clinical Genetics Group, Weatherall Institute of Molecular Medicine, the NIHR Biomedical Research Centre Oxford, Oxford University Clinical Academic Graduate School and Oxfordshire Health Services Research Committee, Medical Research Fund, Medical Sciences Division, University of Oxford and the Royal College of Surgeons of England Research Fellowship (sponsored by The Rosetrees Trust and The Blonde-McIndoe Research Foundation).

And last but certainly not least, I would like to thank my family and friends for their unconditional support and encouragement throughout these years. My wife Sheena and son Aarav, my parents Sushama and Pramod, my brother Gautam, my mother and father-in-law, Ansulekha and Ashoka and my sister-in-law Selena. This thesis would not have been possible without their continued love and patience!

TABLE OF CONTENTS

TITLE ii

ABSTRACT ii

ACKNOWLEDGEMENTS iv

LIST OF FIGURES xv

LIST OF TABLES xix

ABBREVIATIONS xxiii

CHAPTER 1: INTRODUCTION ……..1

General Introduction…….……...2

Part 1 Vertebrate Skull Development…….…....3

1.1.1 Embryology of skull and suture development …….…3

1.1.2 Patterning and boundary formation in brain and skull…….…..6

1.1.3 Molecular Mechanisms of Coronal Suture Development….……...8

1.1.3.1 Mesodermal Lineage Determination…….…….8

1.1.3.2 Cell Migration…….……12

1.1.3.3 Boundary Formation…….…….13

1.1.3.4 Osteogenic proliferation-differentiation balance……14

Part 2 Aetiology and genetics of human craniosynostosis……..18

1.2.1 Classification of Craniosynostosis……...18

1.2.2 Suture type…….18

1.2.3 Epidemiology of craniosynostosis…….21

1.2.4 Aetiology of craniosynostosis……....23

1.2.4.1 Saethre-Chotzen Syndrome……...23

1.2.4.2 Craniofrontonasal Syndrome……....25

1.2.4.3 FGFR-related craniosynostosis syndromes……..26

1.2.5 Clinical Testing and Genetic Counselling……....43

1.2.6 Surgical management……45

Part 3 Important Molecules and Signalling Pathways in Skull and Coronal Suture Development……47

1.3.1 bHLH transcription factors……..47

1.3.1.1 TWIST1 – a key molecule in the coronal suture and craniosynostosis……..49

1.3.2 *MSX2*……...50

1.3.3 *Jagged1*……...51

1.3.4 Other molecules implicated in craniosynostosis…….53

1.3.5 Molecular pathways implicated in craniosynostosis – overview of *Twist1*, *Fgfr* and *Efnb1* relationships ……..53

Part 4 Aims…….61

1.4. Overall aims of this study……...61

Part 5 High-Throughput Sequencing……...62

1.5. Overview of next-generation sequencing with rationale for using whole-exome sequencing……...62

Part 6 General Strategies for identifying disease genes…….64

1.6.1 Introduction……...64

1.6.2 Cohort/Overlap (multiple affected with dominant disorder)…...65

1.6.3 *De novo*……...66

1.6.4 Double hit (single affected patient with recessive disorder)……..68

1.6.5 Homozygosity (single affected individual with consanguineous parents)……...68

1.6.6 Candidate (dominant disorders)……..69

1.6.7 Choice of strategy to be used in this study……..71

CHAPTER 2: MATERIALS AND METHODS…….72

2.1 Ethics approval…….73

2.2 Patient selection and sample acquisition……73

2.2.1 Selection and stratification of patient cohort with apparently non-syndromic bilateral coronal synostosis……73

2.3 Reagents……83

2.4 Centrifugation……..83

2.5 Extraction of DNA……..……...83

2.5.1 DNA extraction from venous blood…….83

2.5.2 DNA extraction from saliva……..84

2.5.3 DNA quantification ..……86

2.6 Polymerase Chain Reaction (PCR) ……...86

2.7 PCR product visualisation …….87

2.8 Screening for mutations …….88

2.8.1 Restriction digest of PCR products using endonucleases …….88

2.8.2 Dideoxy DNA sequencing ……88

2.9 Sample preparation for whole-exome sequencing ……...90

2.9.1 Overview ……...90

2.9.2 Sample Preparation …….92

2.9.2.1 Shearing the DNA …….92

2.9.2.2 Purification of sample with magnetic beads ……...93

2.9.2.3 Assess the quality of the sample ……...93

2.9.2.4 Repair the ends ……..94

2.9.2.5 Purify the sample with magnetic beads……...95

2.9.2.6 Add the 'A' bases to the 3' end of the DNA fragments …….95

2.9.2.7 Purify the sample with magnetic beads ……..96

2.9.2.8 Ligate the indexing-specific paired-end adaptor ……...96

2.9.2.9 Purify the sample with magnetic beads ……..96

2.9.2.10 Amplify the adaptor-ligated library ……..96

2.9.2.11 Purify the sample with magnetic beads ……97

2.9.2.12 Assess the quality of the sample ……97

2.9.3 Hybridisation ……98

2.9.3.1 Hybridise the library ……98

2.9.3.2 Preparation of magnetic beads ……...100

2.9.3.3 Hybrid capture with SureSelect ……101

2.9.4 Post-Hybridisation Amplification and Cluster Station Preparation ……101

2.9.4.1 Amplify captured library and add index tags ……..102

2.9.4.2 Purify the sample with magnetic beads…….102

2.9.4.3 Assess the quality of the sample ……103

2.9.4.4 Pool samples for Multiplexed Sequencing ……103

2.9.4.5 Prepare samples for cluster amplification ……104

2.10 Microsatellite analysis of apparent de novo mutations to confirm correct sample relationships/paternity……...105

2.11 Allele-specific PCR for identification of parental origin of mutations ……106

2.12 Selection of tag-SNPs to allow genotyping of a population of interest using International HapMap data ……...108

CHAPTER 3: A NOVEL DISEASE GENE ASSOCIATED WITH CORONAL SYNOSTOSIS IS REVEALED BY WHOLE-EXOME SEQUENCING ……109

3.1 Introduction ……..……..110

3.2 Patients, materials and methods ……..……..111

3.3 Gene lists and shortlisted candidates ……..114

3.4 Identification of damaging, loss-of-function mutations in a novel transcriptional regulator in 3 patients ……...118

3.5 Discussion.……...124

CHAPTER 4: TRANSCRIPTION FACTOR 12 (*TCF12*): SCREENING A MIXED DISEASE PANEL, DEFINING A NEW CLINICAL ENTITY AND EXTENDING THE CLINICAL AND MUTATIONAL SPECTRUM ……....128

4.1 Introduction ……...129

4.2 Patients, Materials and Methods ……129

4.3 Screening of a mixed craniosynostosis panel identifies further cases of *TCF12*-mutation positive patients ……130

4.4 Pedigrees of mutation positive families ……...131

4.5 Summary of clinical features in *TCF12* mutation-positive individuals delineates a new clinical entity ……....141

4.6 Discussion ……...142

CHAPTER 5: NON-PENETRANCE IN *TCF12*-MUTATION POSITIVE INDIVIDUALS AND INVESTIGATION INTO PARENTAL OF ORIGINS OF *DE NOVO* MUTATIONS ……145

5.1 Introduction. ……...146

5.2 Patients, materials and methods..……..146

5.3 Identification of significant non-penetrance in mutation-positive relatives with normal cephalometric measurements ……...147

5.4 Hypothesis that non-penetrance may be related to haplotype and relative levels of gene expression ……...148

5.5 Investigation into parental origins of *TCF12* mutations to elicit phase information reveals a number of paternal origin mutations and refines future haplotype assignment ……..156

5.6 Maternally-derived *TCF12* mutations……..163

5.7 Discussion ……...165

CHAPTER 6: HAPLOTYPE ANALYSIS OF *TCF12*-MUTATION POSITIVE INDIVIDUALS TO ASSESS INFLUENCE ON CRANIOSYNOSTOSIS PHENOTYPE ……....167

6.1 Haplotypes and The HapMap database ……..168

6.2 Utilisation of tag-SNPs to genotype a population of interest and identification of validated haplotypes ……....176

6.3 Genotyping of controls to establish allele frequencies of each tag SNP in an unaffected, control population …….178

6.4 Construction of *TCF12* haplotype pedigree reveals presence of transmitted partner alleles and non-transmitted bystander alleles ...181

6.5 Haplotyping of samples from mutation-positive individuals and family members ……..…….186

6.6 Utilising a transmission disequilibrium test (TDT) to look for linkage between *TCF12* mutations and tag-SNPs ……194

6.7 Investigation into association of *TCF12* haplotype in mutation-negative patients with non-syndromic coronal synostosis ……...196

6.8 Discussion ……...198

CHAPTER 7: SEARCH FOR FURTHER CANDIDATE GENES BY EXOME SEQUENCING OF TRIOS AND FURTHER PROBANDS……..202

7.1 Introduction ……...203

7.2 Patients, materials and methods ……...203

7.3 Gene lists and candidate shortlisting identifies a frameshift mutation in a novel gene accounting for autism but not coronal synostosis observed in one patient ……..207

7.4 Whole-exome sequencing of 5 proband-parent trios identifies a de novo mutation in a gene encoding a muscle protein of unknown significance in craniofacial development ……..210

7.5 Analysis of gene lists of putative disease-causing variants from combined analysis of 19 mutation-negative patients utilising cohort/overlap strategy reveals several potential candidates ……..220

7.6 Discussion ……...232

CHAPTER 8: DISCUSSION……...236

8.1 Introduction and general discussion…….237

8.2 A combined clinical and molecular approach to cases of mutation-negative coronal synostosis identifies a new clinical entity: *TCF12*-related craniosynostosis……238

8.3 Investigation into non-penetrance in mutation-positive carriers, parental origins of de novo mutations and *TCF12* haplotype……..245

8.4 New insights into cranial suture biology from double mutant mice demonstrates importance of dosage sensitivity in coronal suture boundary maintenance ……...250

8.5 Investigation into the immune phenotype of patients with *TCF12*-related craniosynostosis implies the immune system is not affected in individuals with heterozygous mutations ……..252

8.6 Evaluation of candidate disease genes and putative disease variants using further exome sequencing with existing and alternate strategies……253

8.7 Remaining questions in coronal craniosynostosis……256

8.8 Conclusion ……..262

8.9 Future Work ……...263

LIST OF FIGURES

CHAPTER 1

Figure 1.1 The newborn skull illustrating major bones and sutures….4

Figure 1.2 Summary of skull deformities resulting from different patterns of suture fusion ……...5

Figure 1.3 Origin of tissues comprising the skull vault in the mouse (at E17.5) based on Wnt1-Cre/R26R fate mapping studies……..11

Figure 1.4 Saethre-Chotzen syndrome…….....25

Figure 1.5 Summary of relative proportions major types of craniosynostosis ……..43

Figure 1.6 Representation of possible structure of TWIST1 ……....48

Figure 1.7.1 Summary of molecular pathways implicated in the pathogenesis of craniosynostosis……...56

Figure 1.7.2 Summary of interplay between key molecules implicated in craniosynostosis at the cellular level …….....60

CHAPTER 2

Figure 2.1 Custom clinical proforma used to stratify patients for clinical assessment……75

Figure 2.2 Summary of sample preparation for whole-exome sequencing using SureSelect Target Enrichment kit (Agilent Technologies) ……. 91

Figure 2.3 Summary of steps to hybridisation of the library .…….100

CHAPTER 3

Figure 3.1 The 3 patients each with bilateral coronal synostosis and mutations in *TCF12* are shown……119

Figure 3.2 The predicted effects of the *TCF12* splicing mutation (c.1468-20T>A) identified in patient OX2636……..120

Figure 3.3 Identification of the first *TCF12 de novo* mutation by whole-genome and di-deoxy sequencing……122

CHAPTER 4

Figure 4.1 Summary of the structure of *TCF12* in mutation-positive patients with coronal synostosis …….132

Figure 4.2 Pedigrees of *TCF12*-mutation positive patients and their family members ……...139

CHAPTER 5

Figure 5.1 Craniofacial appearance of non-penetrant carriers of heterozygous *TCF12*-mutations …….148

Figure 5.2 Summary of the effect of a recombination event on the association between SNPs in strong LD and a causal variant……..153

Figure 5.3 Approaches used to identify causal variants or haplotypes associated with disease ……155

Figure 5.4 Strategies to amplify specific alleles to inform parent of origin for *TCF12* mutations ……158

Figure 5.5 Investigation into parental origins of de novo *TCF12* mutations ……..162

Figure 5.6 Discovery of maternal-origin *TCF12* mutations……..164

CHAPTER 6

Figure 6.1 Relationship of SNPs, Haplotypes and tag SNPs…….....172

Figure 6.2 Plots of *TCF12* demonstrating strong LD across the region occupied by the gene ……....174

Figure 6.3 Differences in levels of wild-type *TCF12* expression in individuals heterozygous for *TCF12* mutations may account for the phenotypic differences seen in penetrant and non-penetrant individuals ……175

Figure 6.4 Phased haplotype display of *TCF12* for CEU population and locations of 5 tag-SNPs…….177

Figure 6.5 Genotyping results for 5 individuals from a UK control population for each of the 5 tag-SNPs…….179

Figure 6.6 Construction of haplotypes in families where the *TCF12* mutation is observed in 2 or more generations …….181

Figure 6.7 Haplotyping results of all 48 family pedigrees…….186

Figure 6.8 Frequencies of each *TCF12*-allele plotted according to the ancestral haplotype ('TTTAT') in 46 families (159 individuals)……..190

Figure 6.9 Frequencies of each *TCF12*-allele plotted according to the ancestral haplotype ('TTTAT') in 46 families (159 individuals) with 50:50 apportioning of mutant and partner alleles in 6 uninformative trios ……....191

Figure 6.10 *TCF12* expression vs genotype……193

CHAPTER 7

Figure 7.1 Identification of a frameshift mutation in *TCF20* (c.3518delA; p.K1173Rfs*5) by whole-exome sequencing in patient OX2438 …….209

Figure 7.2 Discovery of a de novo variant in *MYH4* following exome sequencing of a proband-parent trio …….212

Figure 7.3 Use of an in-house craniosynostosis variant database ('CranioVAR') to identify possible de novo mutations in the 5 exome sequenced trios ……214

Figure 7.4 Summary of analysis of 5 top candidate genes identified following overlap analysis of 19 mutation-negative patient exomes who all have coronal synostosis ……..230

Figure 7.5 The predicted effects of the slicing mutation in PPP2R3A (c.3223-7_3223-4dupCAAT) identified in patient OX4481 …….231

LIST OF TABLES

CHAPTER 1

Table 1 Summary of craniosynostosis syndromes with a defined molecular basis ……..42

CHAPTER 2

Table 2.1 Summary of Clinical Characteristics of 22 patients with Coronal Synostosis and negative genetic testing deemed suitable for whole-exome sequencing ……..80

Table 2.2 Summary of Clinical Characteristics for the parents of 5 mutation-negative patients analysed by exome sequencing of patient-parent trios …….82

CHAPTER 3

Table 3.1 Summary of statistics for exome sequencing of the 7 patients with bilateral coronal synostosis (Round 1) ……..113

Table 3.2 Summary of the top 20 candidates identified in Round 1 of exome sequencing demonstrating non-synonymous variants shared between 2 or 3 patients ……...117

CHAPTER 4

Table 4.1 Spectrum of *TCF12* mutations identified in 48 patients with coronal synostosis ……134

Table 4.2 *TCF12* mutations based on a per-suture basis across different craniosynostosis types ……..134

Table 4.3 Summary of clinical features of *TCF12*-related craniosynostosis ……...140

CHAPTER 5

Table 5.1 Families where the *TCF12* mutation arose de novo were screened by dideoxy sequencing to identify informative SNPs……160

CHAPTER 6

Table 6.1 Summary of validated haplotypes obtained from The International HapMap Project website …….178

Table 6.2 Proportions of each SNP allele following genotyping of a UK control population…….180

Table 6.3 Transmission Disequilibrium Test for 46 families……....195

Table 6.4 Testing association between *TCF12* haplotype and non-syndromic coronal and other suture synostosis in 227 mutation-negative patients……...197

CHAPTER 7

Table 7.1 Summary of statistics for exome sequencing of 7 further patients with bilateral coronal synostosis (Round 2)…….206

Table 7.2 Metrics of exome sequencing 10 parents of 5 mutation negative patients (proband data also shown italicised for comparison) ……215

Table 7.3 Metrics of exome sequencing 6 additional mutation-negative patients with coronal synostosis……..216

Table 7.4 Summary of exome patient-parent trio analysis…..……..219

Table 7.5 Overlap analysis of 19 mutation-negative patients with coronal synostosis with intersection of variants with genes implicated in craniosynostosis disease pathways……..228

BIBLIOGRAPHY ……....265

PUBLICATIONS ARISING FROM THIS THESIS …….……...305

PRESENTATIONS ARISING FROM THIS THESIS ……....306

APPENDIX ……...309

ABBREVIATIONS

1000G	1000 Genomes (http://www.1000genomes.org/)
5´	Five prime
3´	Three prime
A_{260}/A_{280}	Absorbance at 260nm/Absorbance at 280 nm
AKT1	V-Akt Murine Thymoma Viral Oncogene Homolog 1
ALP	Alkaline Phosphatase
ALPL	Alkaline Phosphatase, Liver/Bone/Kidney
ANNOVAR	Annotation of variants computer programme
ARMS	amplification- refractory mutation system
ASXL1	Additional Sex Combs Like Transcriptional Regulator 1
BCH	Birmingham Children's Hospital NHS Foundation Trust patient
BCS	Bilateral Coronal Synostosis
bHLH	basic Helix-Loop-Helix
Bp	Base pair
cDNA	Complementary DNA
CEPH	Centre d' Étude du Polymorphisme Humain
CEU	Utah residents with Northern and Western European ancestry from the CEPH
Cre	Cre recombinase
CSF	Cerebrospinal Fluid
CT	Computerised axial tomography
CTSK	Cathepsin K
CYP26B1	Cytochrome P450, Family 26, Subfamily B, Polypeptide 1
dbSNP	The Single Nucleotide Polymorphism Database
ddNTPs	Dideoxynucleotide
DECIPHER	The DECIPHER database (https://decipher.sanger.ac.uk/)
Dermo1-Cre	Dermo1-CRE recombinase.
dHAND	Deciduum, Heart, Autonomic Nervous System And Neural Crest Derivatives-Expressed Protein 2
DiI	(1,1'-Dioctadecyl-3,3,3',3'-Tetramethylindocarbocyanine Perchlorate ('DiI'; DiIC18(3)))
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
E2-2	E2-2 protein
E2A	E2A protein
E6-8	Embryonic day 6-8 (mouse)
E-box	Enhancer box
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalography
EFNA4	Ephrin-A4
EFNB1	Ephrin-B1
eHAND	Extraembryonic Tissues, Heart, Autonomic Nervous System And Neural Crest Derivatives-Expressed Protein 1
EIIa-Cre	EIIa-Cre recombinase transgene
EN1	Engrailed Homeobox 1

ERF	Ets2 Repressor Factor
ERK1/2	Extracellular-signal-regulated kinase 1/2
ESCO2	Establishment Of Sister Chromatid Cohesion N-Acetyltransferase 2
EVS	Exome Variant Server (http://evs.gs.washington.edu/EVS/)
ExAC	Exome Aggregation Consortium (http://exac.broadinstitute.org/)
FAM20C	Family With Sequence Similarity 20, Member C
FBN1	Fibrillin 1
FGF	Fibrolast Growth Factor
FGFR	Fibrolast Growth Factor Receptor
FHOD3	Formin Homology 2 Domain Containing 3
FOAR	Fronto-orbital advancement and remodelling
FREM1	FRAS1 Related Extracellular Matrix 1
g	Grams
g	Relative centrifugal force
Gb	Gigabase
GERP	Genomic Evolutionary Rate Profiling
GLI1-3	Glioma-Associated Oncogene Family Zinc Finger 1-3
GOS	Great Ormond Street Hospital for Children NHS Foundation Trust patient
HEB	HEB protein
HEBAlt	HEB Alternative
HEBCan	HEB Canonical
HELQ	Helicase, POLQ-Like
Hh	Hedghog
HSD174B4	17β-hydroxysteroid dehydrogenase type 4
ICPM	Intra-cranial pressure monitoring
ID	Inhibitor Of DNA Binding 1, Dominant Negative Helix-Loop-Helix Protein
IHH	Indian Hedgehog
IL11RA	Interleukin 11 Receptor, Alpha
JAG1	Jagged 1
KCl	Potassium chloride
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LacZ	Encodes β-galactosidase
LAH	Liverpool Alder Hey Children's NHS Foundation Trust patient
LD	Linkage disequilibrium
loxP	Locus of Crossover in P1
LRT	Likelihood ratio test
LUCS	Left Unilateral Coronal Synostosis
M	Metopic
MAPK1	Mitogen-Activated Protein Kinase 1
Max	MYC Associated Factor X
Mb	Megabase
MCS	Multi-species Conserved Sequences
MEGF8	Multiple EGF-Like-Domains 8
Mesp1	Mesoderm Posterior Basic Helix-Loop-Helix Transcription Factor 1

$MgCl_2$	Magnesium Chloride
MIM #	Mendelian Inheritance in Man number
min	Minute (s)
miRNA	Micro Ribonucleic acid
ml	Millilitres
MLL2	Myeloid/Lymphoid Or Mixed-Lineage Leukemia 2
MLPA	Multiplex ligation-dependent probe amplification
mM	mili molar
MRI	Magnetic Resonance Imaging
MSX2	Msh Homeobox 2
Myc	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog
MYH4	Myosin Heavy Chain 4
MyoD	MyoD Family Inhibitor
ncRNA	Non-coding ribonucleic acid
NELL-1	NEL-Like 1 (Chicken)
NeuroD	Neuronal Differentiation 1
ng	Nanograms
NGS	Next-generation sequencing
^oC	Degrees Celcius
OMIM	Online Mendelian Inheritance in Man (http://www.omim.org/)
Osx	Osterix
OX	Oxford University Hospitals NHS Foundation Trust patient
P1	Post-natal day 1 (mouse)
P21	Post-natal day 21 (mouse)
PCR	Polymerase Chain Reaction
PEDF	Pigment epithelium-derived factor
pg	Pico grams
PhyloP	Phylogenetic p value program
P_i	inorganic phosphate
PI3K	phosphatidylinositol-3-kinase
PolyPhen2	Polymorphism Phenotyping v2
POR	P450 (Cytochrome) Oxidoreductase
PPP2R3A	Protein Phosphatase 2, Regulatory Subunit B", Alpha
PR	Posterior Release/Posterior Calvarial Remodelling
RAB23	RAB Family Small GTP Binding Protein RAB 23
RAF1	Raf-1 Proto-Oncogene, Serine/Threonine Kinase
RANKL	Receptor activator of nuclear factor kappa-β ligand
RAS	Rat Sarcoma Viral Oncogene
RBM8A	RNA Binding Motif Protein 8A
RECQL4	RecQ Protein-Like 4
REEL 3	The Receptive-Expressive Emergent Language Test - Third Edition
RET	Ret Proto-Oncogene
rs	reference SNP cluster ID
RUCS	Right Unilateral Coronal Synostosis
RUNX2	Runt-Related Transcription Factor 2
S	Sagittal
s	Second (s)

SAMtools	The Sequence Alignment/Map format tool computer program
SIFT	Sorting Intolerant From Tolerant computer program
SKI	SKI Proto-Oncogene
SNP	Single Nucleotide Polymorphism
SOX10	SRY (Sex Determining Region Y)-Box 10
SSR	site-specific recombination
TAR	thrombocytopenia with absent radii
TCF3	Transcription Factor 3
TCF4	Transcription Factor 4
TCF12	Transcription Factor 12
TCF20	Transcription Factor 20
TDT	transmission disequilibrium test
TGFBR1/2	Transforming Growth Factor, Beta Receptor 1/2
TWIST1	Twist Homolog 1 (Drosophila)
UBQLN4	Ubiquilin 4
UCS	Unilateral Coronal Synostosis (side unspecified)
V	Volts
WDR35	WD Repeat Domain 35
WIPPSI-R/3	Wechsler Preschool and Primary Scale of Intelligence third edition
WISC-III	Wechsler Intelligence Scale for Children third edition
Wnt1	Wingless-Type MMTV Integration Site Family, Member 1
X-gal	BCIG or 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Xq	Long arm of chromosome X
ZNF16	Zinc Finger Protein 16
µg	Micro grams
µl	Micro litres

CHAPTER 1

INTRODUCTION

General Introduction

The work described in this thesis aims to identify the molecular genetic basis of a relatively common craniofacial condition - coronal craniosynostosis. In the introductory section, I will outline key information relevant to human skull development, review current knowledge on the aetiology and genetics of craniosynostosis and highlight molecular mechanisms in suture development. This will focus on the role of *TWIST1*, and demonstrate the rationale for searching for additional genes responsible for craniosynostosis. The advent, maturation and refinement of high-throughput DNA sequencing technology and bioinformatic techniques during the lifetime of this project has represented a significant asset in the search for novel disease genes in craniosynostosis. Translation of findings into a clinical setting has significant benefits for genetic counselling as it improves the accuracy of advice offered, aids surgical risk stratification, and will improve diagnostic outcomes for coronal synostosis. Identification of new disease genes will increase our understanding of skull development and also clearly define a precise molecular target for the development of novel therapies in the future.

Part 1 Vertebrate Skull Development

1.1.1 Embryology of skull and suture development

The human skull vault is a complex composite structure, consisting primarily of five flat bones – two paired frontals and parietals and an unpaired occipital bone. They all form by intramembranous ossification (a process by which bone is formed without a cartilaginous precursor), inside a skeletogenic mesenchymal membrane, sandwiched between the meninges that surrounds the brain and the dermal mesenchyme (Morriss-Kay & Wilkie, 2005). The vault and base of the skull constitute the neurocranium and the jaws and other facial derivatives constitute the viscerocranium. The skull vault is derived from two separate embryological cell populations – the neural crest and cephalic mesoderm. While the frontal bones are neural crest in origin, the parietal bones are derived from the cephalic mesoderm. The occipital bone has an upper squamous part (equivalent to the murine interparietal bone), which has two ossification centres, either side of the midline and is seen at the 8th week of development (Rice, 2008). The lower squamous part of the occipital bone, whose murine counterpart is the supraoccipital bone, also has two ossification centres, observed in the 7th week. The membranous interparietal bone in mice fuses with the ossified supraoccipital bone (which is derived from sclerotomal components of the occipital somites), and forms the posterior of the skull vault (Morriss-Kay & Wilkie, 2005).

Between the bony plates of the skull are fibrous joints known as sutures that unite them and form major sites of bone growth at the leading edges of the cranial bones (Opperman, 2000). The major sutures of the skull include: metopic or frontal (between the frontal bones), sagittal (between the parietal

bones), coronal (between the frontal and parietal bones) and lambdoid (between parietal and occipital bones). The sutures allow the skull to co-ordinate and accommodate growth with that of the rapidly expanding brain from before birth until maturity of the craniofacial skeleton. The midline (metopic and sagittal) and lambdoid sutures overlie spaces between cerebral or cerebellar hemispheres, but the coronal suture does not form over any distinct anatomical landmark, lying at the juxtaposition of frontal and parietal bones (see Figure 1.1).

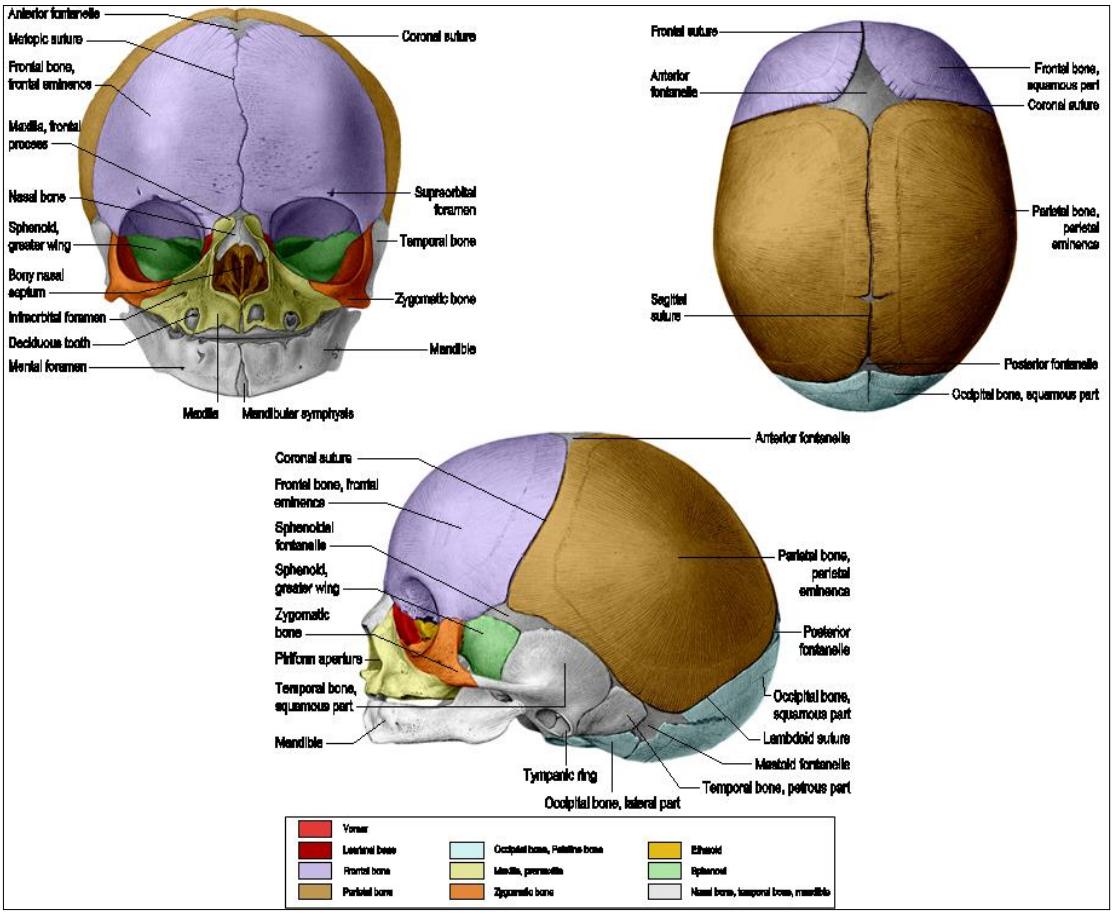


Figure 1.1: The newborn skull illustrating major bones and sutures (From *Gray's Anatomy, 40th edition,* used with kind permission from Prof S Standring and Elsevier, Standring, 2009).

To fulfil their role as sites of bone growth, the sutures must remain open. Their premature fusion by osseous obliteration – craniosynostosis, prevents

this and results in the loss of a growth centre, restricting skull growth at that point. Skull growth proceeds perpendicularly to suture orientation and synostosis of one or more suture will result in compensatory overgrowth at other sutures, remodelling of the skull and cause distinct patterns of calvarial deformity (see Figure 1.2). Other functions of sutures are as absorbers of mechanical stress that protect osteogenic tissue and underlying brain, and as points of articulation that hold various parts of the skull together. This allows for its controlled deformation as the skull passes through the birth canal (Rice, 2008). This is also mediated by the fontanelles which are found in the skull at a point where 3 or more bones converge (Rice, 2008). The posterior and anterior fontanelles are unossified areas in the developing skull that diminish in size as the calvarial bones grow postnatally. They close at approximately 2-3 months (posterior fontanelle) and 1-2 years (anterior fontanelle) after birth (Scheuer and Black, 2004).

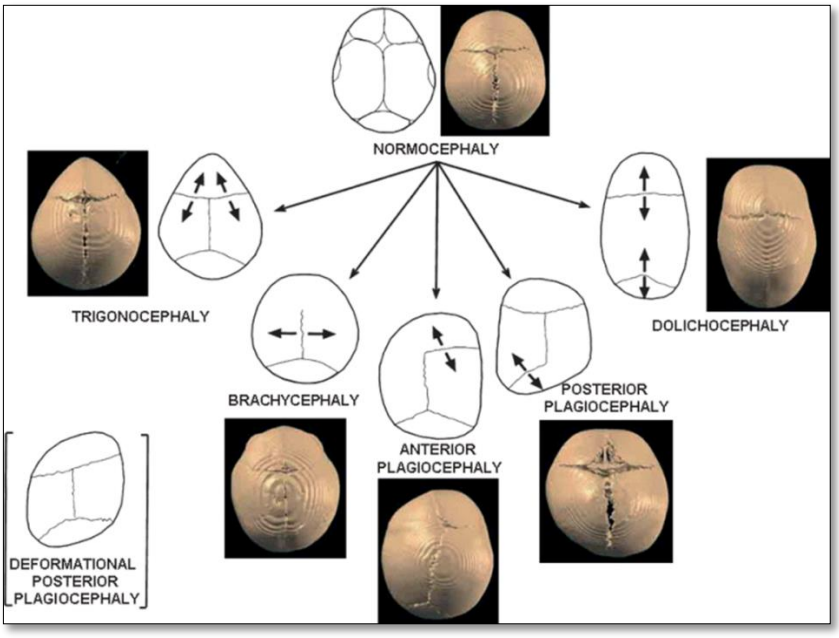


Figure 1.2 Summary of skull deformities resulting from different patterns of suture fusion. Skull growth proceeds in the direction denoted by the black arrows. From Boyadjiev (2007).

The process of physiological suture growth involves expansion of the bone fronts that occupy and recruit mesenchymal tissue to form the leading edges of the skull bones. Although the cranial bones are initially separate, the bone fronts abut or overlap as ossification progresses. The tips of each bone front contain many osteoblasts and osteoprogenitor cells. The sagittal and metopic sutures lie in the midline and do not overlap, whereas the transversely orientated coronal and lambdoid sutures do overlap. As the neurocranium expands, the coronal suture is highly cellular and in the mouse, there is a reduction in the number of cells lining the bone fronts 21 days postnatally (P21, when mice are typically weaned), mimicking development seen in the human skull (Opperman, 2000).

For suture growth to continue until that of the skull is complete, a pool of osteogenic stem cells must be maintained as they provide the raw material for calvarial growth (Johnson et al. 2000; Doll 2004; Rice and Rice 2013). These stem cells reside primarily at the periphery of each membranous bone and also on the endo- and ectocranial surfaces, in smaller populations. The continued growth of the skull depends on maintaining a balance between their proliferation and differentiation and is discussed later (see 1.1.3.4).

1.1.2 Patterning and boundary formation in brain and skull

The formation and maintenance of tissue boundaries is a key process in the correct development of many body systems, preventing illegitimate mixing of cells destined to form different parts of the embryo and giving spatial information for surrounding cells (Kiecker & Lumsden, 2005). In complex structures such as the brain and skull, correct maintenance of boundaries is even more critical for normal growth and patterning. It has long been

established that calvarial bones of the mouse skull grow from initial primordia of cells located basolaterally. They then proceed in an apical manner, towards the skull vertex (Iseki et al., 1997), eventually forming a bony layer between established meningeal and mesenchymal layers. This growth by expansion of the calvarial primordia occurs by *Runx2/Fgfr2*-expressing cells almost exclusively concerned with vertical growth and without recruitment from adjacent mesenchyme (Yoshida et al., 2008).

The narrowing of the gaps between the bony plates of the skull forms the sagittal, metopic and lambdoid sutures (Morriss-Kay & Wilkie, 2005). The condensations of tissue destined to form the frontal and parietal bones arise close to the skull base, in between the brain and eye. The paired bones then confront each other at the sagittal suture as osteogenesis progresses in an apical direction. Initially, the sagittal suture resides within the sulcus between cerebral hemispheres, with osteogenic fronts facing inwards toward the meningeal membranes. Postnatally, butt joints form by virtue of the osteogenic fronts that now appose each other directly, and later the sagittal suture changes morphologically to a joint with numerous projections that interdigitate (Rice, 2008).

The coronal suture does not form in the same way and the frontal bone is overlapped by the parietal bone from the beginning of its development (Morriss-Kay & Wilkie, 2005). The overlap between the bones is due partly to the fact that condensations of osteogenic tissue forming the frontal and parietal bones are initially closer together than those forming the separate frontal or parietal bones. The important junction at which the coronal sutures lie, between neural crest and mesoderm-derived tissues, is well established

by embryonic day 9 (E9) in mice (*Mus musculus*) and the latter cell population forming the parietal bone stays external to the former cell population forming the frontal bone. Following its establishment, coronal suture morphogenesis then progresses from this basolateral position to the apex in the midline, like a zip fastener. In contrast to the end-on butt joint formed by the sagittal suture, the coronal suture forms a bevelled joint (Rice, 2008).

Due to its importance in this study, the molecular mechanisms underlying coronal suture development will be considered in greater depth in the next section.

1.1.3 Molecular Mechanisms of Coronal Suture Development

1.1.3.1 Mesodermal Lineage Determination

Marking the origin of a cell and following it to its ultimate differentiated state or fate, requires a precise mapping technique. An invasive method of performing this is by labelling with DiI – a fluorescent dye used for staining purposes (1, 1'-Dioctadecyl-3, 3, 3', 3'-Tetramethylindocarbocyanine Per-chlorate; Invitrogen, USA). This can be injected into the mouse head at specific embryonic stages to investigate skull vault growth (Yoshida et al., 2008). A non-invasive means of determining cell lineage within a mixed population uses two genetically altered alleles that express a reporter allele and an enzyme that is a site-specific recombinase (SSR) which can express a marker of interest in a persistent manner (Anastassiadis et al., 2010).

A recombinase commonly used in studies in mice is Cre (which causes recombination in the P1 bacteriophage) whose action on *loxP* sites (two 13

bp inverted repeat sequences surrounding an asymmetric core of 8 bp), allow recombination at certain DNA sequences with high fidelity. This will allow manipulation of the genome and incorporation of deletions, insertions or inversions of specific genes (Branda et al., 2004).

Removal of the stop codon upstream of the *R26R* reporter gene by Cre, allows continuous expression of β-galactosidase. Cre recombinase expression is under the control of the *Wnt1* promoter, expressed in the dorsal neural tube. Therefore, skull bones derived from this specific cell population will express β-galactosidase on histological tissue sections. This is then easily visualised by staining with X-gal, an organic compound comprised of a galactose molecule linked to a substituted indole, that stains blue when hydrolysed by β-galactosidase (Jiang et al., 2002). Use of this method allows permanent marking of neural crest cells as the offspring will possess the *Wnt1-cre/R26R* construct and *LacZ* will be expressed in all cells that originate from the neural crest, dorsal neural tube or the cranial neural plate. Staining with X-gal will then allow of cells of neural crest origin to be distinguished from those derived of mesoderm (Chai et al., 2000).

The coronal suture is derived from gastrulating mesoderm at E6.5-7.5 with direct labelling of supraorbital precursors within the cephalic mesenchyme revealing the pattern of cell origin and migration (Deckelbaum et al., 2012). Gli1-expressing cells in the mesoderm detected at E7.5-E8.0 have been fate mapped using *Gli1-CreER*^{T2} recombination and these precursor cells migrate rapidly from their initial axial position to the supraorbital region by E8.5-9.0, forming a regulatory centre that is subsequently established at E10.5-E11.5 (Deckelbaum et al., 2012).

Mesoderm-derived cells can also be traced by following expression and migration of the bHLH transcription factor, *Mesp1* [Mesoderm Posterior 1 Homolog (mouse)]. This is initially expressed at E6.5 in early mesodermal cells, with additional contributions from small numbers of mesodermal cells in the frontonasal mesenchyme, before neural crest cells begin their migration (Yoshida et al., 2008).

Fate mapping has also demonstrated that a significant proportion of the calvarial mesenchyme is derived from hedgehog (Hh)-responsive *Gli1*$^{+}$ cells in the cephalic–paraxial mesoderm, located next to the neural tube. Importantly, it is these cells that will later form the coronal suture. They also transiently express Engrailed-1 (*En1*, encoding a homeodomain transcription factor) and have an origin in the supraorbital domain of the developing skull (Deckelbaum et al., 2012).

The coronal suture is later observed at E14, as a fusiform swelling of skeletogenic membrane containing Twist1-expressing cells at the base of the skull, proceeding to the apex as development continues (Johnson et al., 2000). Although the skull is not yet mineralised, by E15 the coronal suture lies between neural crest and mesodermal mesenchyme of the developing skull and remains as a distinct functional boundary.

Analysis of whole mount skull vaults at E17.5 stained with X-gal by Yoshida et al. (2008) confirmed a mesodermal origin of the parietal and lateral interparietal bones in the mouse and neural crest origin of the frontal and medial parts of the interparietal bone. Additionally, the mesenchyme of the coronal suture that lies between the overlapping frontal and parietal bones is of mesodermal origin and the dermal connective tissue that overlies it is

negative on X-gal staining. Furthermore, the boundary of neural crest and mesoderm-derived tissue lies rostral to the coronal suture.

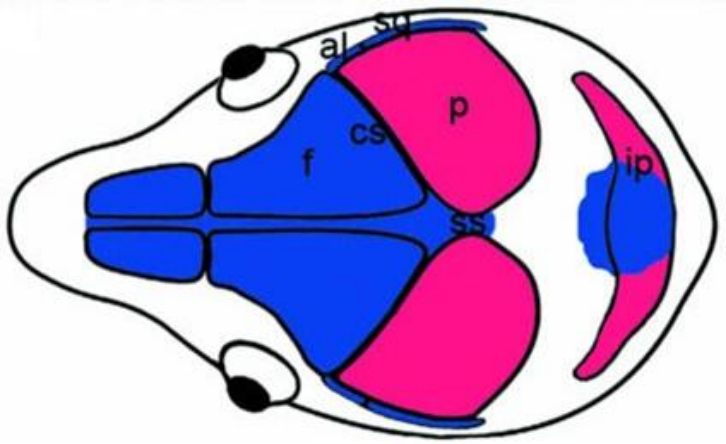


Figure 1.3 Origin of tissues comprising the skull vault in the mouse (at E17.5) based on Wnt1-Cre/R26R fate mapping studies. The frontal bone (f) is derived from neural crest and stains blue when treated with X-gal. It lies rostral to the the parietal bone (p) which has a different origin, being derived from mesoderm (pink). The interparietal bone (ip) has a mixed tissue origin derived from neural crest medially and mesoderm laterally. Other landmarks shown include the coronal suture (cs), sagittal suture (ss) and alisphenoid (al) and squamosal (sq) membrane bones. From Morriss-Kay & Wilkie (2005).

It is by clonal expansion and differentiation that this osteogenic precursor population not only forms the coronal suture, but is key in specifying its position and integrity. The precise regulation of *En1* expression is vital for ensuring cells of mesoderm lineage form the coronal suture and lack of this marker in neural crest cells results in illegitimate entry into its territory, causing abnormalities of suture morphology. During intramembranous ossification (occurring between E14.5-P1), *En1* is found in all developing calvarial bones and sutures. While *En1* heterozygotes have no obvious phenotype, homozygous null mice usually die in the early post-natal period due to mid-hindbrain defects. Of the rare survivors, widened coronal and

lambdoid sutures were consistently observed in addition to incomplete closure of the interfrontal suture (Deckelbaum et al., 2006).

1.1.3.2 Cell Migration

As a further refinement, activity of SSRs can be temporally controlled by fusion with specific protein domains such as those that bind hormones. Without an appropriate ligand being present, the SSR fusion protein will remain restricted to the cytoplasm by virtue of its hormone-binding domain and bound to a heat shock chaperone protein (Legué & Joyner, 2010). The hormone-binding domain can be modified to have a decreased affinity for endogenous hormones (e.g. oestrogen for the CreER fusion protein) and preferential binding to antagonistic molecules such as 4-hydroxy-tamoxifen (denoted as ER^T). Binding of the ligand will induce conformational change in the hormone-binding domain of the fusion protein and release from its cytoplasmic chaperone. The use of *–CreER^{T1};R26R* male mice with various drivers (*En1*, *Wnt1* and *Gli1*) and timed injection of tamoxifen has been implemented in performing fate-mapping studies to precisely delineate precursor cells that contribute to the formation of the coronal suture (Deckelbaum et al., 2012).

The coronal suture had previously been shown to lie between the overlap of frontal and parietal bones by utilisation of a *Wnt1-Cre/R26R* system (Jiang et al., 2002). However, due to a small but significant degree of mixing of *Gli1*-derived mesodermal cells with neural crest-derived cells superiorly to the developing eye, the frontal bone is not purely neural crest in origin

(Deckelbaum et al., 2012). Fate mapping studies with the validated *Mesp1-cre/R26R* mouse demonstrate the mesodermal origin of the parietal bone (Yoshida et al., 2008).

1.1.3.3 Boundary Formation

The coronal suture is a key tissue boundary that separates the frontal and parietal bones derived from neural crest and mesodermal mesenchyme respectively. The mesenchymal tissue within this suture is normally not osteogenic in nature, but is prone to undergoing osseous changes resulting in craniosynostosis. Under normal physiological conditions, the coronal suture is a stable boundary, separating neural crest and paraxial mesoderm cell populations. Central to its integrity are the Twist1 and Msx2 molecules that interact directly, and ephrin-Eph signalling that acts downstream. Impaired function of these molecules by mutations in their respective genes will lead to a destabilisation of the coronal suture boundary, illegitimate cell mixing and fusion of frontal and parietal bones (Merrill et al., 2006). This defect can be visualised histologically as a cloud of disorganised alkaline phosphatase (ALP) expressing cells in *Twist1*$^{+/-}$ mice at the site of the future coronal suture at E14.5. By comparison, no change in ALP expression is seen in wild-type mice, with a thin layer of non-ALP expressing cells in-between ALP-positive cells in the prospective frontal and parietal bones.

The neural crest-mesoderm boundary that denotes the coronal suture loses alignment with the established dermal mesenchyme boundary as vertically-orientated skull growth occurs, so that during later stages of development, the superior part of the coronal suture lies caudal to this mesenchyme boundary (Yoshida et al., 2008). Furthermore, the pattern of radial mineralisation that

is observed during skull growth occurs only after the vertical phase of growth is complete.

During physiological closure, opposing bone fronts within a layer of mesenchyme will approach each other due to deposition of osteoid by osteoblasts. The overlap between fronts causes a sutural 'blastema' to form (Opperman, 2000). Once the activity preventing osteoblast differentiation within the suture is no longer inhibited, mineralisation and fusion of the suture will occur (Richtsmeier & Flaherty, 2013).

The coronal suture is not only a physical boundary, but also a functional one. In *Wnt1-Cre/R26R/Twist1*$^{+/-}$ embryos, ectopically placed cells of neural-crest origin were found in the parietal bone compartment and the normally sharply demarcated boundary in wild-type mice, was observed to be irregular in mutant mice at E14.5 but not earlier embryonic stages (Merrill et al., 2006). Immediately anterior to the coronal suture on the ectocranial side of the frontal bone is a single layer of cells containing Ephrin-A2 and A4 proteins. This layer of cells will be later overlapped by the parietal bone and it has been shown that this layer of cells is mesodermal in origin (Merrill et al., 2006).

1.1.3.4 Osteogenic proliferation-differentiation balance

Specific fibroblast growth factors (FGFs) have been found to activate specific isoforms of their receptors, FGFR1, 2 and 3 (Nie et al., 2006). FGF2 is an important ligand that regulates the balance of osteogenic cell proliferation (via FGFR2 signalling) and differentiation (via FGFR1 signalling) within the coronal suture (Iseki et al. 1999; Kim et al. 1998; Johnson et al. 2000). Application of an exogenous source via FGF2-soaked beads, prevents cell division of osteogenic cells by up-regulating *Fgfr1* expression and down-

regulating *Fgfr2* expression (Iseki et al., 1999). This temporary paracrine effect was designed to recreate the effect of gain-of-function mutations in *FGFR1-3* that result in human craniosynostosis syndromes (see Table 1). However, mineralisation at the site of the implanted bead is delayed, owing possibly to the effects of high levels of *osteopontin* expression.

By comparing the expression patterns of *Twist1* with *Fgfr1-3*, Johnson et al. (2000) demonstrated that in the mouse coronal suture at days E14-18, expression of *Twist* occurs before that of *Fgfr*s. Additionally, while *Twist* is expressed in the mid-sutural mesenchyme, *Fgfr* expression is found in and around developing bony domains: *Fgfr1* in osteoid attached to differentiating cells, *Fgfr2* was seen on cells peripherally located in bony plates and *Fgfr3* has an intermediate position between zones of *Fgfr1* and *Fgfr2* expression as described previously by Iseki et al. (1999) very close to, but distinct from the coronal suture. This suggests a supportive role in the maintenance of this layer as a growth centre for this suture. It may also have a role in the positioning of the coronal suture as it has been observed at E13.5, lying caudally to the eye and eventually forming the mesoderm-derived greater wing of the sphenoid bone The expression domains of midsutural *Twist* overlap with *Fgfr2*-expressing cells that extend in a fan shape around the sutural end of the frontal and parietal bones.

Other important Fgf ligands include Fgf10 and Fgf7. Fgf10 has a role in craniofacial development in 3 locations, by virtue of its interaction with the IIIb splice variant of *Fgfr2* - the calvaria, the palate and the teeth (Veistinen et al., 2009). Alternate splicing of the carboxyl half of the Ig-III domain will generate the IIIb (activated by Fgf3, -7 and -10) and IIIc isoforms (activated

by Fgf2, -4, -8, -9, -18 and -20) (Hajihosseini et al., 2009). Expression of *Fgf3* that binds Fgfr2IIIb with high affinity in the developing tooth, is not seen at this embryonic stage of skull development.

Fgfr1 negatively regulates the rate at which osteoprogenitor cells mature through various stages of development to eventually form osteocytes and mature osteoblasts, the latter of which then secrete matrix. Once differentiation is complete, expression of *Fgfr1* is down-regulated. Co-expression of *Osteonectin* suggests it has a role in the differentiation of osteoblasts as opposed to the function of differentiated osteoblasts, not being expressed until after osteogenic cells have commenced bone matrix secretion (Iseki et al., 1999).

The Fgfr2IIIb and Fgfr2IIIc isoforms are both expressed in the calvarial sutures and noted in the frontal bone primordium at E13.5 (Veistinen et al., 2009). *Fgf10* and *Fgf7* transcripts are detected in the developing calvaria, with *Fgf10* mRNA expression by osteoprogenitors observed in the frontal bone condensation, whereas *Fgf7* is expressed strongly in the undifferentiated mesenchyme adjacent to the developing frontal bone (Veistinen et al., 2009).

Gain-of-function mutations of *Fgfr2* due to a heterozygous deletion of *Fgfr2IIIc*, will result in a change in the isoform that is expressed due to exon switching. This increases the amount of *Fgfr2IIIb* and results in synostosis of coronal sutures, zygomatic arches and sternebrae of mice (Hajihosseini, 2001). Mechanistically, the splicing switch means that cells which originally had Fgfr2IIIc on their surface will now respond to a wider repertoire of ligands, such as Fgf10, by expression of Fgfr2IIIb. This change in specificity will cause inappropriate Fgfr-signalling in the wrong tissue.

Bochukova et al. (2010) performed an analysis of mRNA expression profiles in scalp fibroblasts obtained from patients undergoing craniofacial surgery for correction of syndromic forms of craniosynostosis. In their patient group, coronal suture fusion predominated and a monogenic cause had been identified. Comparison was also made with patients with non-syndromic sagittal synostosis also undergoing surgical correction. A shared expression profile was demonstrated for scalp fibroblasts derived from patients with Apert (*FGFR2*), Muenke (*FGFR3 Pro250Arg*) and Saethre-Chotzen syndrome (*TWIST1*), distinct from those derived from non-syndromic sagittal synostosis, implying that this latter condition has a separate origin. This study further highlights the interplay of signal transduction pathways that contain *FGFR2/3* and *TWIST1* in the cranial sutures.

Part 2 Aetiology and genetics of human craniosynostosis

1.2.1 Classification of Craniosynostosis

Craniosynostosis may be broadly classified along the lines of whether there are additional clinical features that cluster with the presence of a fused suture (syndromic) or if suture fusion is an isolated occurrence with only the expected craniofacial deformity (non-syndromic). The non-syndromic cases can be further subdivided according to the pattern of suture that is fused.

1.2.2 Suture type

The commonest type of craniosynostosis affects the sagittal suture and accounts for up to 58% of cases, with a birth prevalence of 1.9-2.3 per 10, 000 live births. It has a strong male preponderance with over 3 times as many males as females affected. Premature sagittal suture fusion gives a long and narrow head shape, termed scaphocephaly, with frontal and occipital bossing and temporal hollowing. The ratio of maximum skull width to length (cephalic index) is less than 0.75 (normal range is 0.75-0.85) (Boyadjiev, 2007). A significant risk factor is been intrauterine constraint of the foetal head presenting during pregnancy as a result of early descent into the pelvis of the mother. This manifests as symptoms of lower pelvic pressure and discomfort due to cephalopelvic disproportion, and a sensation of 'lightening' as the foetus had 'dropped' lower into the pelvis as originally described in a series of 16 cases of sagittal synostosis (Graham et al., 1979).

More recently, data from a multi-centre, case-control study of birth defects of 670 infants with craniosynostosis was evaluated for associations with parity, gestational age, birth weight and multiple pregnancy (Sanchez-Lara et al.,

2010). The control population comprised 5928 infants. Findings included increased risk of sagittal synostosis in multiparous pregnancy only with a history of fertility treatments. Low birth weight and prematurity were also found to be associated with craniosynostosis, reflecting findings made 10 years prior in a population-based study in Western Australia (Singer et al., 1999).

Coronal suture fusion is the second commonest pattern making up 20-30% of non-syndromic cases. There is a strong female predisposition this time, accounting for up to 75% of cases (Boyadjiev, 2007). Due to a higher proportion of familial cases, a stronger genetic influence is to be suspected (Lajeunie et al., 1995) and targeted molecular testing is recommended (Johnson & Wilkie, 2011). It results in frontal plagiocephaly that is manifested by a flat forehead on the affected side, elevation of the superior orbital rim and compensatory bulge on the contralateral forehead. There is also deviation of nasal root towards the fused suture with chin point to the contralateral side and movement of the ear in an antero-superior direction (Bruneteau and Mulliken, 1992). Bilateral coronal synostosis results in a broad and flattened skull known as brachycephaly.

Metopic suture fusion is the third commonest, accounting for 14% of cases and more males than females generally affected (3.3:1) (Boyadjiev, 2007). A triangular forehead (trigonocephaly) results from this type of fusion and this particular head shape is often associated with many dysmorphic syndromes. Foetal head constraint has also been implicated in its pathogenesis. In 2 cases of metopic synostosis, one had a presumed aetiology related to growth restriction due to a bicornuate uterus, with the foetal head lodged in the left

uterine horn, alongside the placenta. In a second case, the affected patient was one of 3 live born triplets, who had head constraint due to a combination of breech presentation and wedging of the forehead between the hips of her siblings, who were both in vertex presentation (Graham & Smith, 1980). Multiple pregnancy and nulliparity has also been associated with a 2.2-fold or 2.0-fold risk of developing metopic synostosis, respectively. Low birth weight or pre-term delivery also shows an increased risk of 3.6-fold and 2.8-fold, respectively (Sanchez-Lara et al., 2010).

A 9-year pan-European study of 7 craniofacial units showed that the number of cases of metopic suture fusion had significantly increased between 1997-2006, especially in 2001 (van der Meulen et al., 2009). This was in terms of absolute numbers of metopic synostosis and total number of craniosynostosis cases, although no explanation was given for this phenomenon. Although the number of cases of sagittal synostosis were used for comparison, the proportions of coronal synostosis were not surveyed in this study, missing out the second commonest pattern of suture fusion.

Lambdoid synostosis is the rarest type of non-syndromic craniosynostosis, affecting just over 3% of all cases (Boyadjiev, 2007). The posterior plagiocephaly that results from either unilateral or bilateral fusion must be differentiated from a positional or deformational plagiocephaly. This results from an infant sleeping on their back which is not due to suture fusion and resolves spontaneously as the child gets older. Finally multi-suture fusion accounts for 5% of cases and is associated with a higher re-operation rate, increased incidence of neurocognitive impairment and raised intracranial pressure. A diagnosis of multisuture synostosis should also warrant genetic

testing due to 6/17 syndromic patients (excluding those that were clinically obvious such as Apert syndrome) and 1/9 nonsyndromic cases yielding a formal diagnosis in a craniosynostosis cohort study that assessed the impact of genetic disorders (Wilkie et al., 2010).

1.2.3 Epidemiology of craniosynostosis

Estimates of the general prevalence of craniosynostosis include 6.7 per 10,000 live births (Kweldam et al. 2011) in patients managed at 6 craniofacial units across the whole of The Netherlands for all types and 1 in 10,000 live births for coronal synostosis specifically (Slater et al. 2008). A review of 522 patients presenting to a tertiary referral centre in Melbourne, Australia estimated the prevalence of non-syndromic craniosynostosis to be 3.1 per 10,000 live births (Lee et al., 2012), with subtypes including 47.1% sagittal, 21.5% metopic and 17.1% unilateral coronal fusion. Multisuture synostosis accounted for 13% of cases with bilateral coronal suture fusion observed in 3.3% of cases.

A higher birth prevalence was reported for a study over a 14 year period from Atlanta, as 4.3 per 10,000 births (Boulet et al., 2008), although this figure included syndromic patients which may account for the difference with the Melbourne study. Of the 281 infants with craniosynostosis, 84% had isolated craniosynostosis, 9% were syndromic with an underlying genetic defect, and 7% had multiple defects involving cardiovascular, central nervous, musculoskeletal or gastrointestinal systems. Sagittal synostosis was again commonest out of the non-syndromic cases (39%), followed by metopic (19%) and coronal and lambdoid (17% each). The figure for lambdoid synostosis is likely to be an overestimate as deformational posterior

plagiocephaly was misdiagnosed as lambdoid synostosis during the first 5 years of the study. Birth weight was shown to be correlated with risk of craniosynostosis, with high birth weight (>4kg) associated with fusion of the sagittal suture and low (<2.4kg) or very low (<1.5kg) birth weight having a greater risk for metopic and lambdoid synostosis.

Lajeunie and colleagues (Lajeunie et al. 1995) performed a survey specifically looking at the incidence of coronal synostosis that was present across the whole of France. By performing segregation analysis, this group stated that coronal synostosis is a dominantly transmitted disorder with a 60% penetrance and 61% of cases occurring sporadically or as *de novo* mutations. Of the 260 patient cohort with apparently non-syndromic coronal synostosis, 67.8% were female and an overall ratio of males to females of 1:2.11 was observed.

In a study looking at family histories of 660 mutation-negative patients with non-syndromic craniosynostosis, coronal synostosis was found to have the lowest incidence rate in relatives of patients with coronal synostosis, compared to any other suture (Greenwood et al. 2013). This observation is likely to be influenced by the exclusion criteria of the study that did not include monozygotic or unknown zygosity twins to the proband, families with 3 generations of affected individuals, documented mutations in *FGFR1-3* or *TWIST1* and any patients with a genetic syndrome.

However, in comparison with estimates from other groups that cite positive family histories being present in 8-15% of coronal synostosis patients (Mulliken et al. 2004; Lajeunie et al. 1995), this more recent survey (Greenwood et al., 2013) states a much lower rate of only 2% of patients

having affected relatives. This may be a more accurate estimate as those with a proven molecular diagnosis were excluded and labelled as syndromic, but may not have been in earlier studies performed before the advent of targeted molecular testing.

1.2.4 Aetiology of craniosynostosis

In terms of the syndromic craniosynostoses, many eponymous syndromes exist, but they are perhaps best classified by the underlying genetic lesion due to diverse and often overlapping phenotypic features that can create diagnostic confusion. Many of the syndromes where the molecular basis has been delineated are summarised in the Table 1 below. Due to the phenotypic overlap of Saethre-Chotzen syndrome with cases of apparently non-syndromic coronal synostosis, this will be considered separately.

1.2.4.1 Saethre-Chotzen Syndrome

This syndrome was first described in 1931 by Haakon Saethre, a Norwegian psychiatrist, in a mother, two daughters and multiple relatives who presented with an asymmetric skull, simple soft tissue syndactyly of fingers (2/3) and toes (3/4) (Saethre, 1931). A year later, and independently to Saethre's description, Fritz Chotzen a German psychiatrist, described the same types of cranial and limb malformations in a father and his two sons (Chotzen, 1932). Classically, this autosomal dominant syndrome (MIM #101400) has a typical craniofacial gestalt (see Figure 1.3) with coronal synostosis, low frontal hairline, ptosis of the eyelids, blocked tear ducts and small ears that are

posteriorly rotated with prominent crura (Johnson et al., 1998; Johnson, 2003). Extra-cranial features include brachydactyly and simple soft tissue syndactyly of second and third digits and hearing loss secondary to Eustachian tube dysfunction that tends to resolve with age (Gallagher et al. 2012; Rosen et al. 2011).

The discovery by two groups in 1997 (el Ghouzzi, et al. 1997; Howard, et al. 1997) that described mutations of *TWIST1* in patients with Saethre-Chotzen syndrome, identified its molecular genetic basis, with descriptions of a variety of heterozygous loss-of-function mutations in this gene. A more recent survey describes pathogenic mutations in 71% of patients with this diagnosis, nearly 50% of cases having an intragenic mutation and 21% having complete gene deletions (de Heer et al., 2005). Other studies with smaller patient numbers have described *TWIST1* mutations in 80% or more of patients with typical phenotypic characteristics (Chun et al., 2002; Johnson et al., 1998). Although intellectual development is usually normal, those patients with documented *TWIST1* whole-gene deletions have an 8-fold greater risk of developmental delay compared to those with intragenic mutations (Cai et al., 2003).

Figure 1.4 Saethre-Chotzen syndrome - summary of mutations identified in *TWIST1* (left), with the majority of intragenic mutations denoted by green triangles or squares. Typical phenotypic appearances shown in antero-posterior and lateral images (right), illustrate coronal synostosis, ptosis of the right eye, deviated nose and chin point to the right, low frontal hairline and low-set, posteriorly rotated ears with prominent transverse crus. Adapted from Johnson and Wilkie (2010) and Woods et al (2005).

1.2.4.2 Craniofrontonasal Syndrome

Mutations of Ephrin-B1 (*EFNB1*), located on chromosome Xq13.1 that cause loss-of-function of the protein, are responsible for craniofrontonasal syndrome (CFNS; MIM #304110) (Twigg et al., 2004; Wieland et al., 2004). Patients with this gene mutation consistently have coronal synostosis, hypertelorism, longitudinal ridging or splitting of nails on at least one digit, webbing of the neck and clinodactyly of one or more toes. Paradoxically for an X-linked disorder, CFNS females are more severely affected (particularly with severe hypertelorism), whereas males inheriting the mutation generally do not have craniosynostosis. The unusual mechanism underlying this is known as cellular interference and is due to a combination of random X-inactivation in females making them mosaic, resulting in abnormal cellular interactions on the cell surface, and presumed functional redundancy of *EFNB1* in hemizygous males who are not mosaics. However, Twigg et al. (2013) described 6 males with CFNS which arose sporadically, all with

mosaic *EFNB1* mutations, both formally confirming a molecular diagnosis and further substantiating the cellular interference phenomenon.

1.2.4.3 FGFR-related craniosynostosis syndromes

There are 7 well described craniosynostosis syndromes with an *FGFR* mutation as an underlying defect, usually named eponymously with a proven molecular diagnosis coming many years after the initial clinical description. As a group, they account for the majority of known monogenic causes, acting by gain-of-function mechanisms (see 1.2.5).

Pfeiffer syndrome (MIM #101600) is genetically heterogeneous as in 5% of cases it can be caused by mutations of *FGFR1* due to the Pro252Arg mutation when it is associated with a milder phenotype (Lajeunie et al., 2006); *FGFR2* mutations, with a more severe phenotype, account for the majority. There are a variety of mutations described in this syndrome that cause more severe craniofacial deformities, developmental delay, hydrocephalus and early death resulting from respiratory distress. Typical clinical features include multi-suture synostosis that can be very severe and result in a 'cloverleaf' skull deformity, in addition to thumbs and toes that are broad and deviated radially, as well as syndactyly. The Jackson-Weiss syndrome consists of craniosynostosis and deviated 1st metatarsals that are shorter and wider than others, but with hands that are normal. Mutations in *FGFR2* (Ala344Gly) and *FGFR1* (Pro252Arg) have been reported by Jabs et al. (1994) and Roscioli et al. (2000), respectively. Therefore, this can be included within the Pfeiffer spectrum (Johnson & Wilkie, 2011).

Apert syndrome (MIM #101200) presents with bilateral coronal synostosis, midface hypoplasia and bilateral complex syndactyly of hands and feet, with varying degrees of limb fusion. There is also a conductive hearing loss that is observed in the majority of patients (Agochukwu et al., 2012). Approximately two-thirds of cases result from the Ser252Trp mutation in *FGFR2* associated with a milder limb phenotype and degree of developmental delay, but a stronger association with cleft palate. By contrast, the one-third of cases that result from the Pro253Arg mutation in *FGFR2* have a more severe limb phenotype (Slaney et al. 1996; Lajeunie et al. 1999). Either mutation will increase not only the affinity of the receptor for FGF ligands, but also the range of ligands that it will bind. An exclusive paternal origin has been demonstrated for Apert mutations with nearly all arising as *de novo* events (Moloney et al. 1996).

Crouzon syndrome (MIM #123500) is typically the mildest of the *FGFR2*-related craniosynostoses and presents with bilateral coronal or multisuture synostosis and a typical 'Crouzonoid' facial appearance that includes prominent eyes, shallow orbits, hypertelorism and maxillary hypoplasia (Reardon et al. 1994). Again, a conductive hearing loss is commonest, but over a quarter will have sensorineural hearing loss and 14% have a mixed pattern (Agochukwu et al., 2012). Over 95% of mutations occur in the immunoglobulin-like III domain (IgIII), with Cys342Tyr seen mostly commonly in a series of 57 Crouzon patients (Lajeunie et al. 1999; Wilkie 2008). However, mutations in the IgI domain have also been reported in patients with atypical Crouzon syndrome. Pulleyn et al. (1996) describe the first in a child who had Crouzonoid features, sagittal synostosis, proptosis and choanal stenosis with a documented Tyr105Cys mutation. More subtle

clinical features were noted by Sharma et al. (2012) in a patient with sagittal synostosis and mild Crouzonoid features seen peri-obitally as a result of a Cys62Arg substitution in *FGFR2* encoding the IgI domain.

In Beare-Stevenson cutis gyrata syndrome (MIM #123790), a severe pattern of craniosynostosis has been observed with over two-thirds of published cases (4/6 patients) having a cloverleaf skull, in addition to ear defects and cutis gyrata of the skin of the head and neck, trunk and limbs with deep furrows and a corrugated appearance.

Muenke syndrome (MIM #602849) is reported as the commonest genetic cause of craniosynostosis, accounting for 5% of all cases (Wilkie et al., 2010) and results from a single mutation- Pro250Arg in *FGFR3*. Typical features include coronal synostosis, brachydactyly with carpal and tarsal fusions seen on limb radiographs (Bellus et al. 1996) and low frequency hearing loss. Homologous mutations are seen in *FGFR1* in Pfeiffer syndrome (Pro252Arg) and in *FGFR2* in Apert syndromes (Pro253Arg).

Crouzon with acanthosis nigricans (CAN) or crouzonodermoskeletal syndrome (MIM #612247) presents with classic features of Crouzon syndrome, but with the addition of verrucous hyperplasia and skin hypertrophy and hyperpigmentation in skin flexures. This variant of Crouzon syndrome results from an Ala391Glu mutation in the transmembrane region of *FGFR3* (Meyers et al., 1995). In an analysis surveying 35 patients with CAN, additional neurological associations included hydrocephalus in 25.6-38% of cases and Chiari I malformations in approximately 75% of patients (Arnaud-López et al., 2007). Choanal atresia or stenosis was observed in 41-

43% of cases, related to structural rearrangements of the cranial base and if noted in the presence of Crouzon syndrome, is highly suggestive of CAN.

Syndrome	MIM #	Incidence	Cytogenetic locus	Gene	Protein function	Mutations described	Major Clinical Features	Key references
AUTOSOMAL DOMINANT								
Muenke	602849	1:30,000	4p16.3	*FGFR3* (Fibroblast Growth Factor Receptor 3)	Receptor (Tyrosine-protein kinase cell surface receptor for FGFs)	Pro250Arg (gain-of-function)	Uni- or bilateral coronal synostosis, mild digit anomalies (brachydactyly), sensorineural hearing loss	Bellus et al. (1996); Muenke et al. (1997)
Crouzon with Acanthosis Nigricans (CAN)	612247	1:1,000,000	4p16.3	*FGFR3*	Receptor (see above)	Ala391Glu	Multisuture synostosis, verrucous hyperplasia and skin hypertrophy, hyperpigmentation and accentuation of skin markings, especially in flexural areas; normal extremities and intellect	Meyers et al. (1995); Arnaud-López et al., (2007)

Crouzon	123500	1:60,000	10q26.13	*FGFR2* (Fibroblast Growth Factor Receptor 2)	Receptor (As above)	94% occur in exons IIIa or IIIc (gain-of-function)	Bilateral coronal or late-onset pansynostosis, hypertelorism , strabismus, proptosis ('Crouzonoid facies') - normal hands and feet; mandibular prognathism	Reardon et al (1994)
Apert	101200	1:65,000	10q26.13	*FGFR2*	Receptor (As above)	Ser252Trp (66%), Pro253Arg (33%) (gain-of-function)	Bilateral coronal synostosis with bilateral symmetrical complex syndactyly of hands and feet	Wilkie et al. (1995)
Saethre-Chotzen	101400	1:50,000	7p21.1	*TWIST1* (twist family bHLH transcription factor 1)	Transcription factor (forms homo- and heterodimers to regulate osteoblast differentiation)	Variety including missense, frameshift, nonsense and whole gene deletions (loss-of-function)	Uni- or bilateral coronal synostosis, 2/3 simple syndactyly and brachydactyly, broad radially deviated 1st metacarpals /metatarsals	el Ghouzzi et al. (1997); Howard et al. (1997)
Pfeiffer	101600	1:100,000	10q26.13/ 8p11.22	*FGFR2/ FGFR1* (Fibroblast Growth Factor Receptor 1)	Receptor (As above)	94% occur in *FGFR2* exons 8 and 10 (Ig-III domain); also rarely Pro252Arg (*FGFR1*) (gain-of-function); Mutations with severe	Multi-suture craniosynostosis, exorbitism, and midface hypoplasia; radially deviated broad thumbs and great toes are typical. Less frequently, partial	Muenke et al. (1994); Johnson and Wilkie (2011)

						phenotype include: Trp290Cys, Tyr340Cys, Cys342Arg or Ser351Cys	syndactyly in the hands and feet is observed	
Beare-Stevenson Cutis Gyrata Syndrome	123790	Unknown	10q26.13	*FGFR2*	Receptor (As above)	Tyr375Cys Ser372Cys Also 63bp deletion (1506del63) believed to affect splicing.	Multi-suture synostosis (cloverleaf skull), cutis gyrata (furrowed skin with a corrugated appearance), hydrocephalus, hypertelorism, proptosis, choanal atresia and palatal abnormalities.	Przylepa et al. (1996); Slavotinek et al. (2009)
Boston-type	604757	3 families	5q35.2	*MSX2* (Msh Homeobox 2)	Transcription factor (transcription regulator in bone development	Pro148His or Pro148Leu; gene duplications also associated with craniosynostosis (see 1.3.3)	Variable patterns of fusion ranging from single to multi-suture synostosis including cloverleaf skull; additionally hypotelorism, midface retrusion and hand anomalies described in case reports	Jabs et al. (1993)
Greig	175700	1-9: 1,000,000	7p14.1	*GLI3* (GLI Family Zinc Finger 3)	Transcription factor	Intragenic frameshift and nonsense mutations, several whole	Midline synostoses predominate (metopic and sagittal); pre-	Vortkamp et al. (1991); Hurst et al. (2011)

					(dual function transcriptional activator and repressor of the sonic hedgehog pathway)	gene deletions described	axial and post-axial polydactyly with broad or duplicated thumbs, broad or duplicated halluces, post-axial polydactyly of hands and feet and cutaneous syndactyly of the fingers or toes	
Kabuki	147920	1:32,000	12q13.12	*MLL2* (Myeloid/ Lymphoid Or Mixed-Lineage Leukaemia 2)	Enzyme (histone methyltransferase, involved in epigenetic transcriptional activation)	Numerous frameshift and nonsense mutations (loss-of-function)	Distinctive facial gestalt, skeletal anomalies, short stature, developmental delay, abnormal dermatoglyphics ; few cases with proven craniosynostosis, including single suture synostosis resulting in scaphocephaly (1 case, Gillis et al. (1990) or trigonocephaly and 1 case of bilateral coronal and 1 case of unilateral coronal synostosis (Martinez-Lage et al. (2010)	Ng et al. (2010b)

Loeys-Dietz	609192 (LDS 1) /610168 (LDS 2)	Unknown	9q22.33/ 3p24.1	*TGFBR1/2* (Transforming Growth Factor, Beta Receptor 1/2)	Receptor (transmembrane serine/threonine kinase involved in wide variety of cellular processes)	Missense mutations affecting conserved residues of tyrosine kinase domain of receptor	Triad of arterial tortuosity and aneurysms, hypertelorism, and bifid uvula or cleft palate, variable craniosynostosis affecting sagittal suture	Loeys et al. (2005)
Alagille	118450	>1:70,000	20p12.1	*JAG1* (Jagged 1)	Ligand (binds multiple Notch receptors, mediates signalling)	1 frameshift (c.2405delGAAAG) and 1 nonsense mutation (E553X) described	Coronal synostosis in 3 published cases; paucity of intrahepatic bile ducts, cholestasis, cardiac disease, skeletal abnormalities, ocular abnormalities, characteristic facial phenotype	Kamath et al. (2002)
***EFNA4*-related**	601380	Unknown	1q21-q22	*EFNA4* (ephrin-A4)	Ligand (GPI-bound ligand, binds cell surface Eph receptors promiscuously, involved in bidirectional signalling	2 missense (H60Y and P117T), 1 frameshift (471_472delCCinsA) described	Non-syndromic coronal synostosis	Merrill et al. (2006)
***FBN1*- related**	134979	Unknown	15q21.1	*FBN1* (Fibrillin 1)	Structural protein	Cys1223Tyr and Pro1148Ala	2 cases described having 'scaphocephaly with	Sood et al. (1996)

					(calcium-binding microfibrils and regulates osteoblast maturation)		craniosynostosis' (presumed sagittal suture fusion)	
Shprintzen-Goldberg	182212	Unknown	1p36.33	*SKI* (v-ski avian sarcoma viral oncogene homolog)	Transcription factor (repressor of TGF-β signalling)	10 mutations in 10 individuals with substitution or deletion of amino acids with high sequence conservation	57-100% of mutation positive cases have craniosynostosis with 79%-94% of cases with scaphocephaly.	Doyle et al. (2012)
***ERF*-related craniosynostosis**	600775	Unknown	19q13.2	*ERF* (Ets2 repressor factor)	Transcriptional repressor (regulate osteoblast proliferation and differentiation)	Heterozygous Loss-of-function mutations described in 12 families	Late-presenting multisuture craniosynostosis, Crouzonoid features with exorbitism and midface hypoplasia, Chiari I malformation, developmental problems (concentration/language)	Twigg et al. (2013)
Noonan 3	609942	Unknown	12p12.1	*KRAS* (Kirsten Rat Sarcoma Viral Oncogene Homolog)	Enzyme (bind GDP/GTP and possess intrinsic GTPase activity)	2 heterozygous missense changes described in exon 2	Sagittal and partial bilateral coronal and unilateral lambdoid synostosis described; also hypertelorism with low-set	Kratz et al. (2009)

							ears, short nose, anteverted nares webbed neck or cystic hygroma	
Craniosynostosis Philadelphia type and Syndactyly type 1	185900	Unknown	2q34-q36	*IHH locus* (Indian Hedgehog Homolog)	Ligand (intracellular molecule, binds patched receptor, role in endochondral ossification)	Various duplications around the *IHH* locus, overlapping in a ~9.1kb region that was 40kb 5' of *IHH*	Craniosynostosis of varying severity ranging from sagittal synostosis to cloverleaf skull and re-synostosis post-surgery; also varying degrees of cutaneous and osseous syndactyly	Klopocki et al. (2011)
Trigonocephaly 2	614485	Unknown	9p22.3	*FREM1* (FRAS1 related extracellular matrix 1)	Structural protein (extracellular matrix protein with role in epidermal differentiation and epidermal adhesion at embryonic development)	5 genomic deletions and 3 point mutations	Metopic synostosis, midface hypoplasia and speech and language delay noted in series of 8 patients	Vissers et al. (2011)
Bohring-Opitz	605039	Unknown	20q11.21	*ASXL1* [Additional Sex Combs Like 1 (Drosophila)]	Transcription factor (transcriptional regulator of retinoic acid receptors)	Nonsense and frameshift mutations (loss-of-function)	Metopic synostosis, exophthalmos, facial nevus flammeus , up-slanting palpebral fissures, hirsutism,	Hoischen et al. (2011)

							flexion of wrists and elbows, ulnar deviation of the wrists and metacarpophalangeal joints	
AUTOSOMAL RECESSIVE								
Antley-Bixler	201750, 613571	Unknown	7q11.23	*POR* [(P450 (cytochrome) oxidoreductase] if abnormal steroidogenesis	Enzyme (electron transfer from NADP to cytochrome P450 in microsomes)	Variety of POR missense mutations with certain variants commoner in specific ethnic populations but Arg457His likely to be a global founder mutation	Craniosynostosis (usually involving the coronal suture as brachycephaly commonly seen), radio-ulnar or radio-humeral synostosis, and femoral bowing; femoral fractures, midface hypoplasia, proptosis choanal atresia, arachnodactyly, camptodactyly, and rocker-bottom feet	Flück et al. (2004); Huang et al. (2005)
Carpenter	201000	1:1,000,000	6p11.2	*RAB23* (RAB23, member RAS oncogene family)	Enzyme (small GTPase that regulates intracellular	Biallelic frameshift and nonsense mutations, founder mutation (L145X) in	Fusion of midline sutures common (i.e. metopic and sagittal), cloverleaf skull in severe cases;	Jenkins et al. (2007); Jenkins et al. (2011)

					membrane trafficking)	patients of northern European descent; further missense and splice site mutations also described	obesity, brachydactyly, congenital heart defects, umbilical hernia, hypogenitalism, genu valgum, molar agenesis, learning disability	Twigg et al., (2012)
	614976	Unknown	19q13.2	*MEGF8* (Multiple Epidermal Growth Factor-Like Domains 8)	Membrane protein (possible role in trafficking of membrane-bound receptor molecules)	5 individuals reported with homozygous or compound heterozygous mutations (parental consanguinity reported in 4 cases)	3/5 with isolate metopic suture fusion, 1/5 with metopic and sagittal, 1/5 with coronal synostosis; also laterality defects (dextrocardia, transposition of great vessels) and genital anomalies (bilateral cryptorchidism)	
Baller-Gerold	218600	<1:1,000,000	8q24.3	*RECQL4* (RecQ Protein-Like 4)	Enzyme (DNA-dependent ATPase/helicase and may modulate chromosome segregation)	Compound heterozygous or homozygous missense, frameshift and splicing mutations	Coronal synostosis and radial aplasia	Van Maldergem et al. (2006)

Roberts	268300	Unknown	8p21.1	*ESCO2* (establishment of sister chromatid cohesion N-acetyltransferase 2)	Enzyme (Acetyltransferase required for cohesion of sister chromatids at S phase of mitosis)	Mainly homozygous frameshift mutations; nonsense and missense also described	Craniosynostosis (not specified) found in 4% (2/49) patients; other features include microcephaly, exophthalmos, hypertelorism, ear malformation, cleft lip and palate, brachydactyly and clinodactyly affecting upper and lower limbs and developmental delay	Vega et al. (2005); Vega et al. (2010)
Craniosynostosis and dental anomalies (CRSDA)	614188	Unknown	9p13.3	*IL11RA* (interleukin 11 receptor, alpha)	Receptor (receptor for interleukin-11, part of a signalling complex involved in control of proliferation/ differentiation of progenitor cells controlling development of craniofacial bones and teeth)	Variety of homozygous frameshift, missense and nonsense mutations	Craniosynostosis, delayed tooth eruption, maxillary hypoplasia, supernumerary teeth, digital abnormalities	Nieminen et al. (2011)

Raine	259775	Unknown	7p22.3	*FAM20C* (Family With Sequence Similarity 20, Member C)	Enzyme (calcium-binding kinase that phosphorylates caseins and other proteins concerned with mineralisation of tissues)	Homozygous splice site, compound heterozygous missense and chromosome 7 rearrangement and microdeletion and whole gene deletion described	Increased ossification of skull and facial bones, narrow prominent forehead, midface hypoplasia, proptosis, depressed nasal bridge; poor survival beyond 10 days postnatally due to thoracic hypoplasia	Simpson et al. (2007); Ababneh et al. (2013)
Sensenbrenner/ cranioectodermal dysplasia	218330	Unknown	2p24.1	*WDR35* (WD Repeat Domain 35)	Ciliary intraflagellar transport protein (part of the intraflagellar transport complex for retrograde ciliary transport)	Compound heterozygous mutations in 2 patients (splice site and missense and frameshift and missense)	Sagittal synostosis, short limbs, thin hair, dental anomalies; also medullary cystic kidney disease, hepatic fibrosis, retinitis pigmentosa and brain anomalies	Gilissen et al. (2010)
Pycnodysostosis	265800	1.7:1,000,000	1q21.3	*CTSK* (Cathepsin K)	Enzyme (involved in osteoclastic bone resorption and bone remodelling)	>40 differing loss-of-function mutations described in either homozygous or compound heterozygous state	Wide sutures with delayed closure, cases of craniosynostosis, acroosteolysis of distal phalanges, osteosclerosis and bone fragility and short stature	Gelb et al. (1996); Arman et al. (2014)

Craniosynostosis with radiohumeral fusions and other skeletal and craniofacial anomalies	614416	1 published family (3 siblings)	2p13.2	*CYP26B1* (Cytochrome P450, Family 26, Subfamily B, Polypeptide 1)	Enzyme (role in metabolism of retinoic acid, causing its oxidation and inactivation)	Homozygous missense changes (p.Ser146Pro and p.Arg363Leu variants identified)	Severe craniofacial malformations including craniosynostosis, occipital encephalocele, radiohumeral fusions, oligodactyly and advanced osseous maturation.	Laue et al. (2011)
Hypophosphatasia,infantile	241500	1:6730 (moderate HP) and 1:297,000 (severe HP)	1p36.12	*ALPL* (alkaline phosphatase, liver/bone/ kidney)	Enzyme (possible role in skeletal mineralisation)	Homozygous (A162T) and multiple compound heterozygous variants described	6 clinical forms recognised: perinatal (lethal), perinatal benign, infantile, childhood, adult, odontohypophosphatasia; infantile and childhood forms associated with craniosynostosis; also rachitic ribs, hypercalciuria, and premature loss of deciduous teeth	Weiss et al. (1988); Mornet (2007); Mornet et al. (2011)
X-LINKED								
craniofrontonasal	304110	1:120,000	Xq13.1	*EFNB1* (ephrin-B1)	Ligand (binds receptor tyrosine kinases EPHB1 and EPHA1)	Wide array of mutations including: whole gene deletions, nonsense, frameshift and	~80% have craniosynostosis: 48% bilateral coronal, 4% Right coronal, 22% Left coronal, 4%	Twigg et al. (2004); Wieland et al. (2004); van den Elzen et al. (2013)

		splice site, and missense substitutions (loss-of-function)	bilateral coronal and sagittal synostosis; also: marked hypertelorism, bifid nasal tip, indentation of columella, broad nasal bridge low implant of breasts (90%), rounded, sloping and narrow shoulders, clinodactyly ≥ 1 digit, longitudinal nail ridges/splits, aberrant shaped eyebrows - females usually more severely affected than males

Table 1 Summary of craniosynostosis syndromes with a defined molecular basis

1.2.5 Clinical Testing and Genetic Counselling

A cohort study from Oxford analysed the prevalence of single gene disorders in patients with craniosynostosis and reported a proven genetic cause in 21% of cases (Wilkie et al., 2010). Of the proportion of formal genetic diagnoses that were made, 86% were monogenic causes and 15% were due to chromosomal abnormalities.

The monogenic causes comprised mutations of *FGFR2* (32%) followed by *FGFR3* (25%), *TWIST1* (19%) and *EFNB1* (7%). A summary is given below (see Figure 1.4):

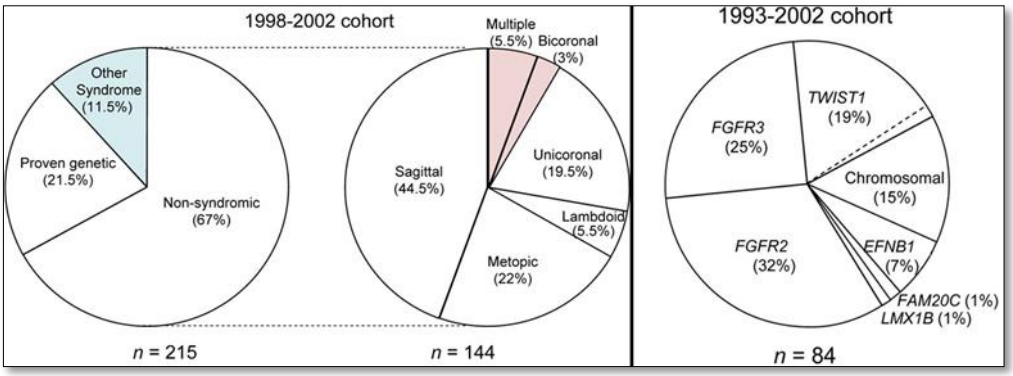


Figure 1.5 Summary of relative proportions major types of craniosynostosis Non-syndromic and syndromic craniosynostosis (left) with analysis on a per suture basis (middle). Of those with a proven genetic basis (21.5%), relative proportions of types of genetic diagnosis are shown (right). From Wilkie et al (2010).

In some cases, the diagnosis is very obvious such as Apert syndrome or typical Crouzon syndrome, and hence direct targeted testing should be undertaken. In apparently non-syndromic cases however, choosing the appropriate test may be more difficult. The underlying genetic lesion was identified in 37.5% of bilateral coronal, 17.5% of unilateral coronal and 11% of multi-suture synostosis cases, with mutations identified in *FGFR2* (exons

IIIa and IIIc), *FGFR3* (Pro250Arg) and *TWIST1*. Therefore, in light of these data, the minimum level of genetic testing that should be offered in non-syndromic cases should target these mutational hotspots. The observed enrichment of mutations that are detected in patients with a fused coronal suture may well be related to its unique embryonic origin and function as a tissue boundary between frontal and parietal bone territories (see 1.1.3.3). The sagittal suture has a different tissue origin (see 1.1.2) and its premature fusion may depend more on non-genetic factors, possibly explaining the poor diagnostic yield from routine genetic testing.

In terms of offering accurate genetic counselling, if genetic testing and family history are negative, a 10% recurrence sibling risk is quoted for bilateral coronal and multisuture synostosis, 5% for unilateral coronal cases and 2% for metopic and sagittal synostosis. Risks to offspring in apparently non-syndromic cases is again highest for bilateral coronal and multi-suture synostosis (30-50%), and of the order of approximately 5% in unilateral coronal, metopic and sagittal synostosis (Johnson & Wilkie, 2011; Wilkie et al., 2010). When the monogenic cause is thought to arise *de novo*, a 1% recurrence risk is quoted for *FGFR* mutations. The possibility of mosaicism must be borne in mind, especially in cases of craniofrontonasal syndrome as 19% of cases with *EFNB1* mutations are reported to be mosaic (Twigg et al., 2006). Therefore, a 10% recurrence risk for siblings of females with apparently *de novo* mutations is appropriate with this syndromic diagnosis.

1.2.6 Surgical management

Surgical correction of the calvarial deformity usually occurs on an elective basis, although exact timing of surgery varies between units and obviously depends on the age at which the patient initially presents. Regardless of local policy, all cases of craniosynostosis must be managed in a tertiary centre with a multidisciplinary team comprised of surgeons from plastic and reconstructive, neurosurgical, maxillofacial, otolaryngological and dental backgrounds, in addition to geneticists, clinical psychology, speech and language therapists and eye care professionals.

Generally, most non-syndromic cases are operated on when the patient is 1 year of age, with syndromic cases usually requiring earlier operative intervention (Forrest & Hopper, 2013). Correction of the calvarial deformity by remodelling the skull will help to obviate functional problems such as raised intracranial pressure and relieving the cranio-cerebral disproportion. Raised pressure may also result from hydrocephalus (necessitating shunt placement, usually from the ventricles to the abdomen, to drain off excess cerebrospinal fluid) or partial airway obstruction causing obstructive sleep apnoea. Depending on the underlying anatomical problem this may be rectified by tracheostomy, removal of tonsils and adenoids, or need advancement of the midface with osteotomies and distraction osteogenesis with a ridged external distractor. Associated malformations affecting the sensory functioning of the patient can range from sensorineural hearing loss in Muenke syndrome, dental malocclusion associated with *FGFR2* mutations and strabismus in syndromes that affect the coronal suture (Johnson & Wilkie, 2011).

There is a growing body of evidence that highlights the importance of a formal molecular diagnosis and careful phenotypic evaluation to aid surgical management. Specific genotypes (*FGFR3* Pro250Arg) are associated with an increased number of surgical procedures, with a 5 times higher re-operation rate than for patients without this mutation (Thomas et al., 2005). In a 15 year retrospective review of Saethre-Chotzen syndrome patients with a proven heterozygous *TWIST1* mutation, there was an increased re-operation rate observed due to raised intracranial pressure of 35% at 12 months post-operatively, 42% at 5 years and an approximate 50% lifetime risk (Woods et al., 2009). Unfavourable post-operative outcomes following strip craniectomy for correction of sagittal synostosis have been observed in 7 out of 89 patients treated by this method, who were noted to have a vertex bulge after the procedure. Of these, 2 patients had an identifiable genetic mutation in *FGFR2* (Ala344Ala) or *FGFR3* (Asn540Lys) and all 7 patients required secondary, more significant skull remodelling procedures to address issues of recurrent raised intracranial pressure, deteriorating head shape or both (Marucci et al., 2008).

Part 3 Important Molecules and Signalling Pathways in Skull and Coronal Suture Development

1.3.1 bHLH transcription factors

Basic helix-loop-helix (bHLH) transcription factors are a large family of over 240 proteins found in many organisms ranging from humans to yeast and are involved in a variety of key developmental processes (Atchley & Williams, 1997). TWIST1 is a key bHLH transcription factor in developing head mesenchyme and due to its critical role, it will be considered separately below. A hexanucleotide Enhancer box (E-box) mediates gene transcription that is cell-type specific (Massari & Murre, 2000). The basic region of a bHLH protein makes contact with the major groove in DNA, as do the residues in the loop and second helix. Conserved hydrophobic residues add stability to the HLH domain by van der Waals forces acting between them. Highly conserved glutamate and arginine residues are found across most bHLH proteins, having a key role in stabilizing its binding to DNA. The glutamate residue in the basic region makes contact with cytosine and adenosine bases in the E-box half-site. The adjacent arginine residue is responsible for stabilising its neighbour by interaction not only with these nucleotides but also the phosphodiester backbone (Atchley and Fitch, 1997). The x-ray crystal structure is illustrated in Figure 1.6 below.

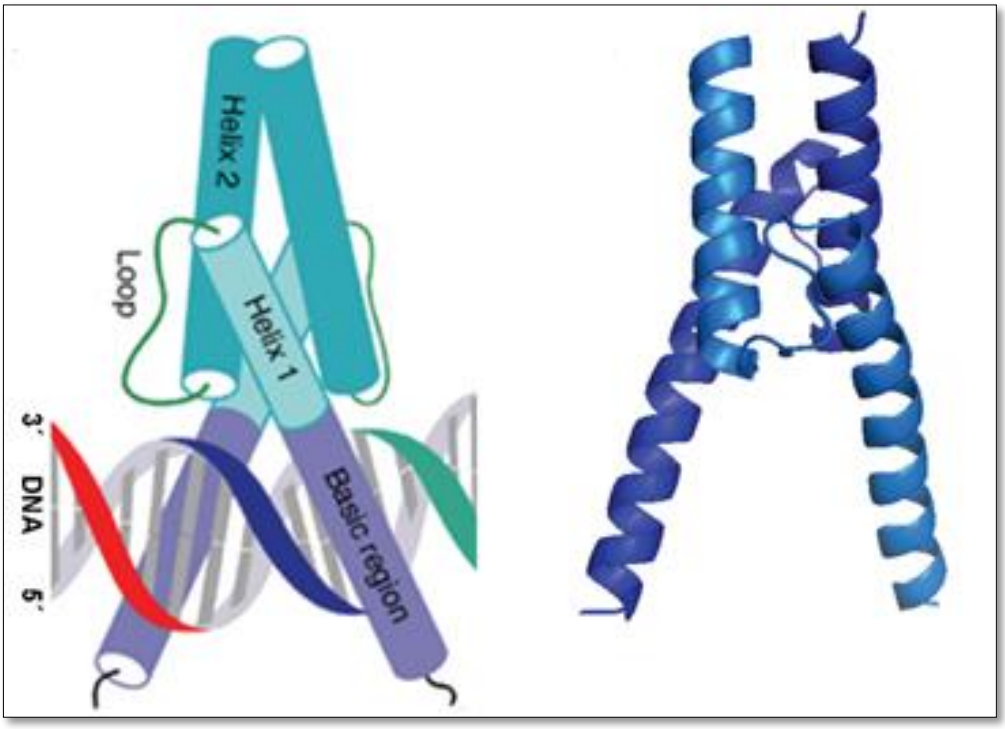


Figure 1.6 Representation of possible structure of TWIST1 The homodimer is shown complexed with DNA (left) and computer modelled representation (right). From Maia et al. (2012).

Seven major classes of bHLH proteins were originally proposed on the basis of DNA-binding specificity, distribution in tissues and potential to dimerise (Massari and Murre, 2000). Class I bHLH proteins include E2A, HEB and E2-2, collectively known as E-proteins and have wide ranging tissue specificities, forming homo- and heterodimers. Those in Class II include eHAND, dHAND, TWIST1, NeuroD and MyoD. Class III and IV bHLH proteins include Myc and Max respectively that also form homo- and heterodimers. The lack of a basic region in HLH proteins such as Id is a hallmark of the Class V group and Class VI HLH proteins including Hairy, have a proline in the basic region. The bHLH-PAS domain found in proteins such as Single-minded and Period are found in the final Class VII bHLH proteins. An alternate classification system has also been suggested, based on

evolutionary relationships, E-box binding, residue conservation, existence of additional domains and sequencing of new genomes, dividing bHLH proteins into 6 groups, A-F (Atchley and Fitch 1997; Ledent et al. 2002).

As discussed previously, Twist proteins contain highly conserved bHLH domains and their antagonist Id proteins lack the DNA-binding domain. Direct Twist/Id protein-protein interaction will inhibit the formation of heterodimers.

Twist action can vary between species, having an activating role in myogenesis in *Drosophila* and an inhibitory role in the mouse and more generally in inhibiting mesodermal differentiation (Hebrok et al., 1994).

1.3.1.1 TWIST1 – a key molecule in the coronal suture and craniosynostosis

First identified as a zygotic gene important in patterning of the dorsoventral axis of *Drosophila* during embryogenesis (Simpson, 1983), TWIST1 has a well described role in developing head mesenchyme. In vertebrates there are 2 *Twist* genes, *Twist-1* and *-2*, with the former being crucial for neural tube closure in mice (Chen & Behringer, 1995). Human *TWIST1* has an open reading frame that encodes a 202 amino acid protein. It has a 45 bp untranslated segment, followed by a 536 bp intron and finally a second exon that is untranslated (Gripp et al., 2000).

Conventionally, class II bHLH proteins do not form stable homodimers, with the notable exception of Twist1, where in mesoderm tissue it can dimerise with itself (Castanon et al., 2001). Inhibitors of dimerization or Id proteins, lack the basic domain necessary for DNA binding and therefore disrupt the normal action of E-proteins to form heterodimers, which they do

preferentially (Kee, 2009). *Twist1* is known to be expressed throughout the intervening mesenchyme that lies between osteogenic fronts of the bony plates of the skull (Connerney et al., 2006).

As its expression is down-regulated in osteoprogenitor cells within mouse calvarial cultures, it has been suggested that Twist1 either prevents differentiation or promotes proliferation (Hebrok et al., 1994; Johnson et al., 2000). This differential role may be related to Twist1 forming homo or heterodimers, and the relative proportions of each (Connerney et al., 2006). Twist1 contains an anti-osteogenic domain known as a Twist box that interacts with and inhibits Runx2 function, preventing differentiation of osteoblasts and premature ossification of suture mesenchyme (Bialek et al., 2004).

Although complete loss of *Twist1* in homozygous mice is embryonically lethal with null mice demonstrating neural tube closure failure and exencephaly (Chen & Behringer, 1995), *Twist1*$^{+/-}$ heterozygous mice are viable, and regularly exhibit coronal synostosis (Carver et al., 2002). This indicates that a full dose of this gene is required for normal suture development and stable boundary function (see also 1.1.3.3).

1.3.2 *MSX2*

In humans, mutation of *MSX2* (Msh homeobox 2, a transcriptional regulator and a member of the muscle segment homeobox gene family) was the first demonstrated cause of craniosynostosis and was identified by study of a large kindred from Boston (Jabs et al., 1993). Affected individuals had a gain-of-function mutation in the gene (c.443C>T; p.Pro148His) resulting in a protein with increased DNA binding affinity. The second reported instance was of a

large Bosnian family presenting to the Dutch Craniofacial Centre with 8 affected family members, all with the same missense mutation (Pro148Leu) (Florisson et al., 2013). Concurrently, the same mutation was reported by Janssen et al. (2013) in a 4-generation Belgian family with sagittal and metopic synostosis of varying severity, along with 3/4 digit syndactyly, phalangeal agenesis and non-penetrance.

The importance of this gene for correct craniofacial development was further underscored by description of loss-of-function mutations in 3 families, which reduced the binding affinity of the MSX2 protein for DNA, resulting in large and persistent defects in the parietal bones of affected individuals (Wilkie et al., 2000). The presence of both activating and inactivating mutations of *MSX2* thereby underscore the importance of this gene for correct craniofacial development.

1.3.3 *Jagged1*

Due to the low frequency association of coronal synostosis with Alagille syndrome (see Table 1) caused by mutations in the Notch ligand *JAGGED1*, Yen et al. (2010) used conditional targeting in mutant mice to inactivate *Jagged1* to recreate the observed human phenotype. Since $Jagged1^{-/-}$ is embryonically lethal and $Jagged1^{+/-}$ mice do not have craniosynostosis, this group used conditional *Jagged1* alleles to analyse the effect of its inactivation in skull vault development. It was observed that the Jagged1 protein was present in mesoderm-derived cells within the early coronal suture as well as a layer of cells outside the prospective bony layer into which migratory osteogenic precursors enter. By using *Mesp1-Cre* and *Dermo1-Cre* (a twist-related bHLH protein expressed in the developing dermis) to instigate

recombination in the part of the coronal suture that is of mesoderm origin, these mice were crossed with those carrying R26R alleles. *Wnt1-Cre* drivers were also used. LacZ staining was punctate rather than uniform in tissue of mesoderm origin including the prospective coronal suture, parietal bone and dermis. This indicated that recombination was caused in only a proportion of mesoderm-derived cells consistent with the boundary defect that is observed in craniosynostosis.

Jagged1 conditional knock-outs were then used with all 3 types of drivers, with antibody staining for Jagged1 revealing patterns of expression in mutant sutures at E14.5. *Wnt1-Cre* did not affect *Jagged1* expression, confirming it is only expressed in mesoderm-derived cells in the coronal suture. *Dermo1*-Cre resulted in a loss of Jagged1 expression in sutural and ectocranial cells, and *Mesp1-Cre* reduced expression in sutural cells more than ectocranial cells. This latter group also had early signs of coronal suture fusion as indicated by narrowing at its prospective site. The *Mesp1-Cre;Jagged1*$^{cko/cko}$ mutant was crossed with an *R26R* reporter to analyse the distribution of lacZ-expressing cells within the coronal suture domain. Significant numbers of these cells were found to cross from the mesoderm compartment to the neural crest compartment, thus indicating a role for *Jagged1* in maintenance of boundary integrity of the coronal suture. *Jagged1* was also shown to affect Wnt/β-catenin signalling as well as Bmp (bone morphogenic protein) and RTK (receptor tyrosine kinase), but not ephrinA-EphA signalling in the coronal suture, suggesting a parallel and downstream action of the latter. An interaction of Twist1 and Jagged1 was suggested by the increase in observed penetrance of coronal synostosis in *Twist1*$^{+/-}$; *Jagged1*$^{+/-}$ compound heterozygotes, from 44% to 77%, as well as increased fusion of squamosal

and lambdoid sutures. Furthermore, in the double mutant mice, the total number of LacZ-positive cells that crossed over the whole length of the coronal suture boundary into the mid-suture mesenchyme was increased significantly compared to the Twist$^{+/-}$ and Jagged$^{+/-}$ single mutants (Yen et al., 2010).

1.3.4 Other molecules implicated in craniosynostosis

Another transcription factor that is important in ossification in the mouse is Osterix (*Osx)* that has been shown to act downstream of Runx2 and is required for progression from preosteoblast to a mature, functional osteoblast (Nakashima et al., 2002). Specifically, homozygous null mutant mice demonstrated a near total lack of mineralisation of facial and skull bones formed by intramembranous ossification and certain bones formed by endochondral ossification such as ribs, limbs and vertebrae were abnormally bent and hypoplastic. *NELL-1* (NEL-Like 1 (Chicken)) has been shown to be up regulated in the human synostotic coronal suture (Ting et al., 1999) and evidence from mice shows this is due to an increase in osteoblast differentiation causing overgrowth of bone and premature suture fusion. If overexpressed to an even higher level, *Nell-1* induces early osteoblast apoptosis (Zhang et al., 2003).

1.3.5 Molecular pathways implicated in craniosynostosis – overview of *Twist1*, *Fgfr* and *Efnb1* relationships

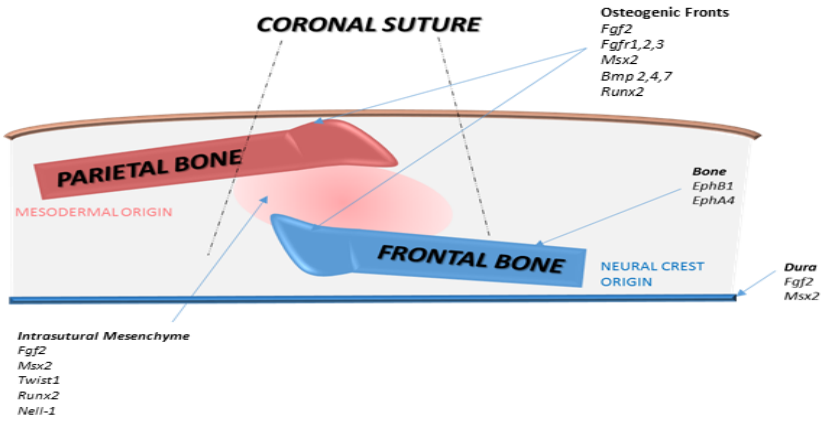
A complex signalling system exists between cells that mediate skull shape, the overall pattern of sutures and individual morphology of cranial bones. A disruption to this system by a specific mutation in the 5 key genes involved in this signalling process results in early fusion of the sutures, i.e. gain-of-

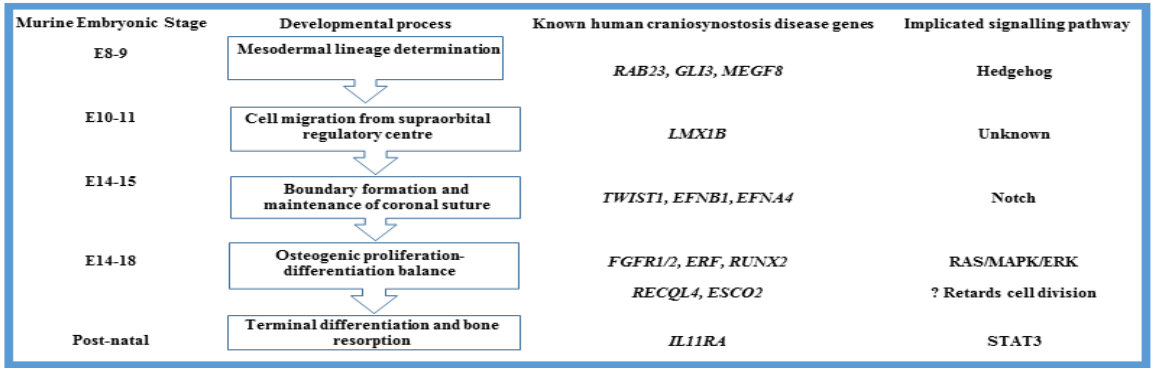
function mutations in *FGFR1-3* and loss-of-function mutations in *TWIST1* and *EFNB1* (Richtsmeier & Flaherty, 2013). This section will focus on molecular interactions between this important group (see also 1.1.3.4).

The 5 fibroblast growth factor receptors (FGFRs) will bind to one or more of 18 polypeptide fibroblast growth factors (FGFs) with varying affinities to control cellular differentiation, proliferation and survival with alternate splicing of the receptor controlling tissue specificity (Beenken & Mohammadi, 2009; Laestander & Engström, 2014). The effects of FGFs on cells with osteogenic potential depends on their stage of differentiation as although mature osteoblasts are not affected by FGF1, 2, 4 or 18, foetal calvarial osteoblasts are responsive (Ornitz & Marie, 2002). Endogenous FGF2 regulates normal development of the cranial vault, due to a positive feedback loop with itself and when applied exogenously, along with FGF4, influences apoptosis and coronal suture fusion (Mathijssen et al., 2001). FGF18 also stimulates osteoclast activity (which act on the endocranial surface of the skull) by formation of cyclooxygenase-2 and RANKL (Receptor activator of nuclear factor kappa-β ligand) (Shimoaka et al., 2002) and null embryos have delayed suture closure.

Rice et al. (2000) performed detailed gene expression studies using *in situ* hybridisation to detect splicing variants of *Fgfr1-3, Fgfr4*, *Twist1*, *Id1* to test if they occupy common molecular pathways controlling osteoblast differentiation in the skull. Fgfr1 was observed in the mid-sutural mesenchyme, calvarial bone osteoblasts and meninges. Fgfr2 was detected in the central area of the mid-sutural mesenchyme consisting of undifferentiated cells, destined to form osteoprogenitors and osteoblasts. *Twist1* is intensely expressed in the head mesenchyme with highest expression in

osteoprogenitor cells in the calvaria. When the dose of *Twist* is reduced in haploinsufficient *Twist*$^{+/-}$ mutant mice, *Fgfr* expression is perturbed from its normal location in differentiating osteoblasts at osteogenic fronts to an ectopic location in the mid-sutural mesenchyme (Rice et al., 2000). This supports the idea of an inhibitory role of Twist in preventing the final differentiation of osteogenic precursors into osteoblasts, especially as mutations in *TWIST1* or *FGFRs1-3* cause craniosynostosis in humans by loss and gain-of-function mechanisms respectively (Johnson & Wilkie, 2011). The integration of these molecules is shown in Figure 1.7.1 below.





Murine Embryonic Stage	Developmental process	Known human craniosynostosis disease genes	Implicated signalling pathway
E8-9	Mesodermal lineage determination	*RAB23, GLI3, MEGF8*	Hedgehog
E10-11	Cell migration from supraorbital regulatory centre	*LMX1B*	Unknown
E14-15	Boundary formation and maintenance of coronal suture	*TWIST1, EFNB1, EFNA4*	Notch
E14-18	Osteogenic proliferation-differentiation balance	*FGFR1/2, ERF, RUNX2*	RAS/MAPK/ERK
		RECQL4, ESCO2	? Retards cell division
Post-natal	Terminal differentiation and bone resorption	*IL11RA*	STAT3

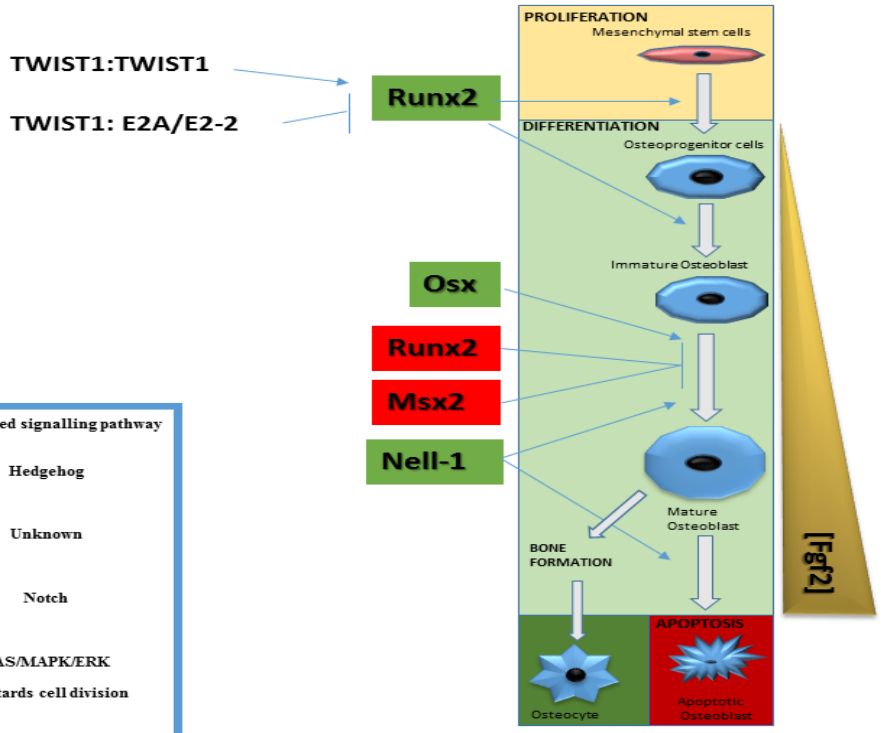


Figure 1.7.1 Summary of molecular pathways implicated in the pathogenesis of craniosynostosis. *Top* – Schematic of a patent coronal suture at E16.5 in sagittal section. The mesoderm-derived parietal bone overlaps the neural-crest derived frontal bone. Important molecules identified from evidence in mouse models implicated in craniosynostosis and are active in the osteogenic fronts, bone, intrasutural mesenchyme and dura are shown (Rice et al., 2000; Rice et al., 2003; Richtsmeier & Flaherty, 2013). *Bottom* – Summary of key events in coronal suture development in the mouse in parallel with evidence from human craniosynostosis syndromes where the molecular basis has been delineated as well as the signalling pathway where these molecules operate (personal communication from Prof A Wilkie). *Right* – Main stages of osteoblastogenesis in the coronal suture which is itself a balance between proliferation of

mesenchymal stem cells and differentiation of osteoprogenitor cells under the control of Runx2 and other regulatory molecules which can have a positive (green) and negative (red) effect. Osterix (*Osx)* is required for osteoblast differentiation and in its absence, there is no bone formation by either intramembranous or endochondral processes. Twist1 can form homo- or heterodimers that can positively or negatively affect Runx2 activity. Nell-1 drives differentiation of osteoblasts and can induce premature apoptosis depending on levels of expression. This proceeds along an Fgf2 concentration gradient (Iseki et al., 1999; Ornitz & Marie, 2002; Rice et al., 2000; Su et al., 2014).

Ephrins are membrane-bound ligands that bind to their cognate tyrosine kinase receptors (Ephs) and are categorised as class-A (membrane-tethered by GPI linkage, binding EphA) or class-B (transmembrane, mainly binding EphB). This latter class contain 3 members (B1, B2, B3) all possessing a PDZ domain and intracellular region with multiple tyrosine residues (Holder & Klein, 1999). The receptor EphA4 can bind ligands from both classes. Coronal synostosis is present in approximately 80% of patients with *EFNB1* mutations (van den Elzen et al., 2013).

Efnb1 and its associated receptor act as a molecular signal responsible for repulsive effects between different cell types, specifying the formation of the coronal suture and responsible for maintenance of a tissue boundary that prevents mixing of neural crest-derived frontal bone and mesoderm-derived parietal bone. These proteins are involved in cellular adhesion, repulsion and migration. The patchy distribution in females (as the condition is X-linked) of cells either expressing or not expressing this protein, causes abnormalities in signalling and cell sorting, forming ectopic tissue boundaries – a process referred to as ‘cellular interference’ (Wieacker & Wieland, 2005). This

consequently disrupts the sharp tissue boundary normally demarcated by the coronal suture, thus causing craniosynostosis.

Ephrin-A2 and –A4 are present on the frontal bone ectocranial surface, in a single cellular layer that is anterior to the coronal suture and of mesodermal origin in the murine skull (Merrill et al., 2006). The cognate receptor of these ligands, EphA4, is also present on the ectocranial side of the frontal bone in an upper layer of mesodermal origin and a lower layer of neural crest origin. This dual layer of expression forms a boundary between cells of neural crest and mesodermal origin and in *Twist*$^{+/-}$ mice, is retracted anteriorly and merges into a single layer of cells. In corollary to its role in hindbrain segmentation, deficiency of ephrin-Eph signalling at the osteogenic interface of neural crest and mesodermal cells, leads to a breakdown in the border that would normally separate osteogenic mesenchyme and undifferentiated sutural cells. The discovery of 3 patients with non-syndromic coronal synostosis each with loss-of-function mutations in *EFNA4* suggested a functional role in the development of the coronal suture. Further evidence from *Twist*$^{+/-}$ mice with the coronal boundary defect and studies showing the expression of ephrin-A2 and –A4 is closely tied with that of their receptor (Merrill et al., 2006). This demonstrates that any deficiencies in the interface between neural crest and mesoderm cells will allow the former to invade the latter and induce differentiation of that osteogenic mesenchyme into osteoblasts.

Furthermore, Ting et al. (2009) demonstrated *EphA4* interacts with *Twist1* to regulate development of the coronal suture. In homozygous mutants,

EphA4$^{-/-}$ mice had partial coronal synostosis, similar to *Twist1*$^{+/-}$ mice. By comparison, *Efna2*$^{-/-}$ mice had normal coronal sutures indicating *EphA4* has a more significant role in its correct development. The appearance of the ALP-expressing layer of cells in the prospective coronal suture was also more disorganised, with a broader expression domain in *EphA4*$^{-/-}$. These null mutants also demonstrated a boundary defect between neural-crest and mesoderm when a *Wnt1-Cre;R26R* reporter system was used to track cell fate, with a substantial number of *lacZ*-positive cells observed to reside ectopically in neural crest territories. *EphA4* was also shown to be an effector of *Twist1* as *Twist1*$^{+/-}$; *EphA4*$^{+/-}$ double mutants had a greater degree of penetrance of craniosynostosis versus *Twist1*$^{+/-}$ alone, in addition to expanded ALP-expressing domains with an increased number of osteogenic cells in their sutures. By analysing the behaviour and contributions of migratory osteogenic cells to frontal and parietal bones using DiI-labelled embryos, this group also showed that in *Twist1*$^{+/-}$; *EphA4*$^{+/-}$ mutants, these cells were present in significant numbers in the sutural space, a phenomenon not observed in wild-type animals at E16.5.

Additionally, *Twist1-EphA4* interactions were shown to be important in partitioning within the coronal suture, keeping osteogenic and non-osteogenic areas separate in wild-type animals, with migratory osteogenic cells staying restricted to an EphA4-expressing layer, on the ectocranial side of the suture. Mutant embryos by comparison had labelled migratory cells seen diffusely in and around the osteogenic layer of the parietal bone, in addition to a breakdown of the coronal suture boundary in these animals as

osteogenic cells of both neural crest and mesoderm origin were found in the coronal suture. The mistargeting of osteogenic cells in mutant mice provides a possible mechanism for premature suture fusion due to the sequential and combinatorial action of aberrant migration and disruption to the normal *Runx2*-dependent proliferation-differentiation balance. A summary of the interactions of the key molecules thought to be important in craniosynostosis at a cellular level are shown in figure 1.7.2

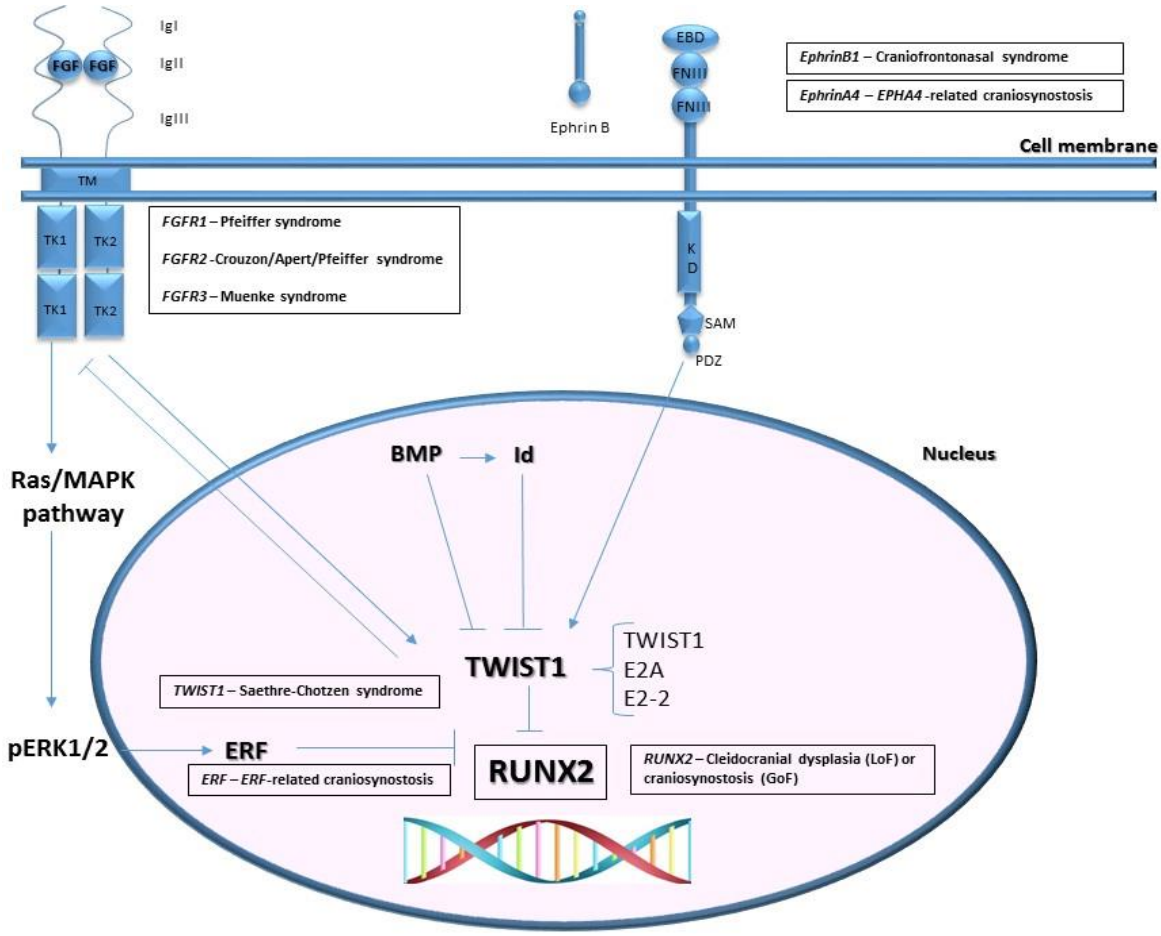


Figure 1.7.2 Summary of interplay between key molecules implicated in craniosynostosis at the cellular level. Interaction between FGF/FGFR and TWIST1/Id pathways affect a final common pathway that regulates osteogenesis – inhibition of RUNX2

in osteoprogenitor cells and the intrasutural mesenchyme, preventing final commitment to osteoblast identity. Loss of this inhibition due to gain-of-function (GoF) mutations of *FGFR1/2/3* or *RUNX2* or loss-of-function mutations (LoF) of *TWIST1, EpH/EPHA ERF or RUNX2* will dysregulate osteoblast differentiation and result in craniosynostosis (names of associated syndromes given in boxes). KEY: IgI/II/III = Immunoglobulin-like domains I/II/III; TK1/2 = tyrosine kinase domains; EBD = ephrin-binding domain; FNIII = fibronectin III; KD = kinase domain; SAM = sterile alpha motif; PDZ = PDZ domain; pERK1/2 = phosphorylated ERK (O'Rourke & Tam 2002; Beenken & Mohammadi 2009; Fitzpatrick 2013).

Part 4 Aims

1.4 Overall aims of this study

The aim of this study is to therefore search for additional genes that are mutated in bilateral coronal synostosis as existing knowledge has indicated patients with this particular pattern of suture fusion would be a good target and an excess of known genetic causes affect the coronal or multiple sutures. In addition the unique embryonic origin and function of the coronal suture as a boundary between the neural-crest derived frontal bone and mesoderm-derived parietal bone and the known genetic mutations that disrupt it, mean it is highly likely to harbour further pathogenic variants that abrogate its normal development. To do this, I decided to employ next-generation/high-throughput sequencing of the protein-coding fraction ('exome') of the genome for a carefully stratified cohort of patients as this is likely to harbour coding mutations of large effect responsible for a disorder such as craniosynostosis. There are a variety of strategies that will be used to search for novel disease genes and these will be outlined in the next section.

Part 5 High-Throughput Sequencing

1.5. Overview of next-generation sequencing with rationale for using whole-exome sequencing

The gold standard technique for determining the order of bases in a DNA sequence was devised by double Nobel Laureate, Fred Sanger and colleagues in 1977. This was classically based on a chain-termination method using a template of single-stranded DNA, a primer, DNA polymerase, standard deoxynucleotide-triphosphates (dNTPs) and modified di-deoxynucleotidetriphosphates (ddNTPs) (Sanger & Nicklen, 1977). By use of linkage mapping and the resequencing of candidate genes, the genetic basis of approximately 33%-50% of Mendelian or single gene disorders has been identified (Bamshad et al., 2011). The disease genes underlying the remainder have been elusive to identification by standard techniques described above due to factors such as the small number of affected cases, reduced reproductive fitness, reduced or non-penetrance and locus heterogeneity whereby similar phenotypes result from mutations at different genetic loci.

With the development of the 'next-generation' of DNA sequencing (NGS) platforms in 2005, it has been possible to capture most of the coding variants that are found in a person's genome. The 'first-generation' of Sanger sequencing was based on separation by electrophoresis of chain termination products from single sequencing reactions. By comparison, NGS comprises clonally amplified templates on which DNA is sequenced by synthesis and nucleotides are added to the complementary strand instead of chain

termination, to allow determination of the DNA sequence. Additionally the amplified DNA templates that are spatially separated on a flow cell are sequenced simultaneously without a physical separation step, in a massively parallel manner (Anderson & Schrijver, 2010).

The process of exome sequencing has revolutionised gene discovery as it applies this new technology to identify variants in the coding fraction of the genome, covering over 95% of exons that are thought to contain more than 85% of disease causing mutations in Mendelian conditions (Rabbani., 2014). The first published application of these techniques was in 2010 that identified the genetic basis of 2 monogenic disorders, Miller syndrome (Ng et al. 2010a) and Kabuki syndrome (Ng et al. 2010b). Although sequencing only the exome of a patient will miss out a large proportion of non-coding variants, it represents a very important fraction of the genome that is enriched for variants with large effect sizes.

Contained within the exome will be deleterious missense or small insertion-deletions which will have adverse functional consequences and many monogenic disorders are caused by mutations that disrupt protein-coding sequences. It is estimated that overall, approximately 20% of missense mutations are loss-of-function with almost one-third of further missense variants having neutral effects and the remainder and majority having effects that are mildly deleterious (Kryukov et al., 2007).

Part 6 General Strategies for identifying disease genes

1.6.1 Introduction

Of more than 23,000 entries listed on OMIM (as of September 2015) most describe genes, but over 3500 rare diseases have an unknown molecular basis or are suspected to have a single gene defect as their cause (OMIM.org; Hamosh et al. 2005). The obvious benefits of having a firm molecular genetic diagnosis include providing accurate genetic counselling, testing for carrier status and prenatal testing. Additionally, by identifying a defined molecular target as a starting point, various pathways associated with the encoded protein, both in health and disease, can be better understood. If the disease locus is relatively large, then a candidate gene approach can be used. However, for rare disorders from small families which often occur sporadically such as craniosynostosis, the underlying genetic basis can remain elusive and an alternative, unbiased means of identifying novel disease genes is required (Gilissen et al., 2012).

It is preferable to start the search for novel disease genes underlying Mendelian disorder in the 1-3% of the genome that is protein coding as around 85% of known causes are found in that subset (Botstein & Risch, 2003). Mendelian diseases are caused by mutations of large effect and subsequently identified variants must be unique to that patient group, reside in the exonic component of the genome and significantly impair the protein encoded by that gene.

Of the 20,000 variants identified by exome sequencing, following quality control filtering such as setting a minimum read depth for independently

called variants, proportion of reads with that variant and sequence quality around the variant, the number of putative variants responsible for a disorder is reduced significantly. Filtering against known public databases such as dbSNP or 1000 Genomes will further reduce the number of candidate variants by up to 95%. This will leave a more manageable list of a few hundred non-synonymous variants that can be further ranked in terms of their deleterious effect (i.e. frameshift, nonsense, or affecting canonical splice sites). Various methods for selection of pathogenic variants will now be described.

1.6.2 Cohort/Overlap (multiple affected with dominant disorder)

By searching for a single gene that is mutated in multiple patients that have been stratified to share the same phenotypic characteristics, the number of possible candidates reduces significantly, and it may be possible to identify a single gene that is shared between all members of a small patient cohort. This was demonstrated by the identification of mutations in *ASXL1* in Bohring-Opitz syndrome, typified by a metopic synostosis, exophthalmos, flexed and ulnar deviated fingers and wrists and severe intellectual disability (Hoischen et al., 2011).

The overlap approach was also key in identifying the molecular genetic basis of Kabuki syndrome (Ng, et al. 2010b) after modification of filtering criteria that allowed for a less stringent analysis of shared candidate disease genes between patients, eventually identifying *MLL2* as the cause in 7 out of 10 cases. This study also demonstrated the genetic and clinical heterogeneity of

this disorder as cascade testing of a cohort of 43 further patients, only identified mutations in *MLL2* in 26 cases. This underscores the importance of careful stratification of patients to give a cohort with a consistent phenotype and aid identification of novel disease genes using the overlap principle. In cranioectodermal dysplasia/Sensenbrenner syndrome, 2 patients with very similar phenotypes and damaging compound heterozygous changes in *WDR35* were found using an overlap strategy.

1.6.3 *De novo*

In the case of common disorders such as autism or intellectual disability that have a high level of genetic heterogeneity, the genomic space occupied by potentially mutated genes (i.e. the proportion of the genome that contains a number of genes, mutations of which result in a disease) is too large to allow discovery of 2 patients with the same gene mutation by the overlap approach. Conversely, this large mutational target does increase the chances of a *de novo* mutation arising during meiosis, especially if it causes a reduction in reproductive fitness and occurs sporadically (Gilissen et al., 2012). The general principle in this approach is to exome sequence patient-parent trios, then search for *de novo* mutations by excluding all inherited variants, yielding on average 0-3 potential pathogenic variants per trio analysed (Vissers et al., 2010; O'Roak et al., 2011) .

Findings from common neurodevelopmental disorder studies looking for novel genes in autistic spectrum disorder and schizophrenia have surprisingly and successfully identified novel *de* novo variants that may account for a

proportion of these disorders. Exome sequencing of 14 trios where schizophrenia was clinically diagnosed in the proband, revealed 15 *de novo* mutations in 8 patients, with 4/15 being nonsense mutations producing a truncated protein (Girard et al., 2011). Analysis of 20 trios with idiopathic autism spectrum disorder by exome sequencing revealed 21 *de novo* mutations with 11/21 predicted to alter protein function (O'Roak et al., 2011). However, additional cases must be sought to provide further evidence of pathogenicity. A caveat of this approach is that it potentially increases the number of mapping artefacts detected as there is a focus on variants that are only present in the child. Therefore, a high degree of sequencing coverage that is the same for all 3 samples that are exome sequenced will be required. In a diagnostic setting, whole-exome sequencing is proving to be successful in achieving a molecular diagnosis in 26% of 814 cases with rare Mendelian disorders, with analysis of trios having a significantly higher yield of 31% due to *de novo* and compound heterozygous mutations (Lee et al. 2014). A similar yield was obtained for 500 probands tested with diagnostic exome sequencing, with trios having a higher rate of diagnosis compared to singletons of 37% and 21% respectively (Farwell et al., 2014). Of this population, 40.4% had specific craniofacial features without a genetic diagnosis that warranted testing by diagnostic exome sequencing, achieving a positive result in 30.7% of cases.

1.6.4 Double hit (single affected patient with recessive disorder)

If only a single patient is available and in the absence of parental consanguinity, if a disease is suspected to have a recessive basis, exome sequencing with analysis that focuses on homozygous or compound heterozygous variants may be successful due to the limited number of private non-synonymous changes seen in such an individual. This has been demonstrated previously in the identification of the genetic basis of Perrault syndrome in 1 patient who was a compound heterozygote for a missense and nonsense mutation in *HSD174B4* (encoding 17β-hydroxysteroid dehydrogenase type 4, involved in fatty acid β-oxidation and steroid metabolism) (Pierce et al. 2010). Drawbacks of this method include dependence on having an ethnically-matched control population with reliable and high quality data.

1.6.5 Homozygosity (single affected individual with consanguineous parents)

In the presence of consanguinity however, it can be assumed that any disease causing homozygous variant is inherited from both parents and is likely to be found in a large run of homozygosity within the patient's genome Conventionally, this method uses SNP microarrays to reliably map the region of interest, but there is precedence for exome sequencing to be used in a similar manner if the data contains a high enough number of informative SNPs to identify homozygous regions on a genome-wide basis (Becker et al., 2011). This has been particularly useful in identifying a homozygous

truncating mutation in a gene encoding pigment epithelium-derived factor (PEDF), a multifunction protein that inhibits angiogenesis very strongly, in a single patient that was severely affected by osteogenesis imperfecta (Becker et al., 2011).

1.6.6 Candidate (dominant disorders)

In this approach, the patient selected is presumed to be the only individual affected by a dominant disorder without additional family members exhibiting the same phenotype. Prioritisation of variants in this situation will focus on the likely deleterious effect on protein structure and functioning and will therefore highlight frame-shifting, nonsense or splice site mutations. For missense sequence changes, *in silico* programmes can be used and contrasted with predicted effects of more benign variants and an aggregate scoring method can be used to rank a list of variants by their predicted deleterious nature.

Such programs include PhyloP to assess the degree of sequence conservation (Pollard et al., 2010) and SIFT ('Sorting Intolerant From Tolerant') to look at whether an amino acid change does indeed affect protein function, with scores of less than 0.05 as damaging (Ng & Henikoff, 2003; Kumar et al., 2009). PolyPhen2 (Polymorphism Phenotyping v2) also considers the functional effects of amino acid substitutions (with scores predicting that a substitution may be damaging or benign) (Adzhubei et al., 2010). The LRT (likelihood ratio test) looks at the deleterious effect of mutations in highly conserved amino acids with protein-coding sequences (with lower p-values

predicted to be more damaging) (Chun & Fay, 2009). Mutation Taster will combine information from a variety of existing databases and analysis tools to interrogate data containing sequence altering variants for their effects on evolutionary conservation, changes affecting splicing, loss of protein effects and alteration in the amount of mRNA (Schwarz et al., 2010). A high 'security' of a prediction is indicated by a score close to 1. Finally, GERP (Genomic Evolutionary Rate Profiling) creates multi-species alignment to analyse the nucleotide conservation within it and measures substitution deficits, i.e. substitutions that would have occurred if the variant was in neutral, but did not occur as it is under a functional constraint (Cooper et al., 2005). Further iterations of this program (GERP++) will estimate constraint for each column of the alignment and identify constrained elements from this output. This provides scores for selective constraint at the nucleotide and element level (Davydov et al., 2010) and ranges from -12.3 to 6.17, from least to most conserved.

Caution must be used with these computer programmes as the predictions obtained by their algorithms do have precedents for missing variants, which later turn out to be responsible for a disease (Ng et al. 2010a). A combination of prediction tools can be used on a bioinformatic basis to prioritise variants most likely to be pathogenic, with validation by Sanger sequencing and evidence from similar mutations in other patients. Other drawbacks include an inherent bias of this approach as this method depends on existing information from established disease pathways. Use of a majority rule approach that combines the functional prediction and conservation-based methods described above can overcome variations in benefits and drawbacks

that a single prediction method affords. It has been used on a large-scale sequencing project on 15336 genes in 6515 individuals to estimate the age at which deleterious single nucleotide variants arose (Fu et al., 2013).

1.6.7 Choice of strategy to be used in this study

The main strategies that will be used are a cohort approach in the first instance, looking at overlapping deleterious, non-synonymous variants present in two or more patients in addition to a candidate gene strategy, looking for the types of genes summarised in Table 1. Secondly, there will be a search for *de novo* mutations by exome sequencing of patient-parent trios. It is felt that by selecting an accurately stratified cohort with only the coronal suture affected, absence of syndromic features and negative family history, this will maximise the chances of success of identifying new genes, mutations of which cause coronal synostosis.

CHAPTER 2

MATERIALS AND METHODS

2.1 Ethics approval

This work in this thesis has been approved by the Riverside Research Ethics Committee, under the study title ‘Genetic basis of craniofacial malformation’ (REC reference number: 09/H0706/20).

2.2 Patient selection and sample acquisition

All patients whose DNA was used for exome sequencing were managed under the care of the Oxford Craniofacial Unit, John Radcliffe Hospital, Oxford and later craniofacial units in Birmingham and London. Informed consent was obtained by their parent or guardian in most cases, on behalf of the patient if under 16 or they lacked capacity. Whole blood was collected in plastic vacutainer bottles containing EDTA (ethylenediaminetetraacetic acid) preservative as an anticoagulant. Further samples were obtained from both parents if possible, and any relatives with a history of craniofacial malformation

2.2.1 Selection and stratification of patient cohort with apparently non-syndromic bilateral coronal synostosis

A thorough clinical evaluation was performed of all patients at the Oxford Craniofacial Unit with bilateral coronal synostosis in whom all gene tests were negative. The coronal suture fusion was confirmed radiologically with computed tomography (CT) with three-dimension reconstruction in all cases. Clinical assessment of limbs had to be normal for inclusion as were

intra-cranial findings on CT or magnetic resonance imaging (MRI). This included brain parenchyma, cerebrospinal fluid and spaces including ventricles. A custom clinical proforma was designed to capture the maximum amount of relevant clinical information and patients stratified accordingly (see below).

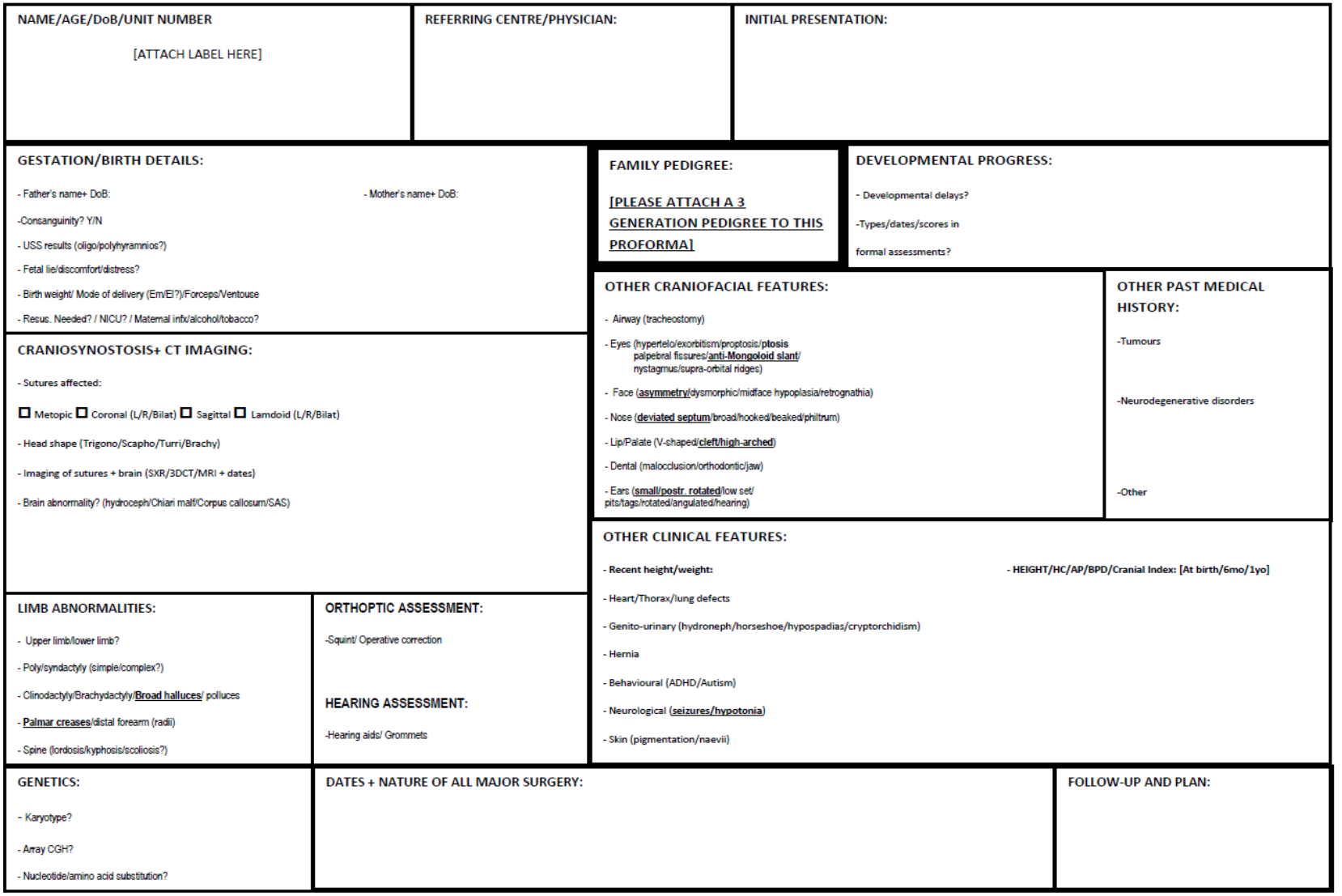


Figure 2.1 Custom clinical proforma used to stratify patients for clinical assessment

PATIENT NUMBER	GENDER	AGE AT ASSESSMENT (YR/M)	CLINICAL PRESENTATION	SUTURE FUSION ON 3DCT	HANDS	FEET	BRAIN	GENETIC TESTING	NUMBER OF MAJOR CRANIOFACIAL PROCEDURES	AGE OF 1ST SURGERY	NEUROCOGNITIVE	ADDITIONAL FINDINGS
1 OX484	M	23	Brachycephaly, proptosis, hypertelorism High flat forehead, low hairline, prominent ear crura, prominent nose with deviated nasal septum	BCS	N	N	N	Karyotype N L1 L2 MLPA	1	0.5	WISC-III: low range/low average	Numerous absence seizures until 17yo, controlled with Valproate/ Ethosuxamide (EEG: abnormal spike and wave discharges)
2 OX1662	F	15	Marked turricephaly, flat occiput, no exorbitism, marked anterior bossing	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	0.83	No concerns	Born 42/40, Ventouse and forceps
3 OX1560	F	15	Marked brachycephaly, early turricephaly, no anterior/posterior fontanelles present, very mild exorbitism,	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	2	N	Born 7 weeks prematurely (33/40). Family history of myotonic dystrophy – father, paternal uncle, paternal grandfather (proband tested – negative)
4 OX4581	F	5	Brachycephaly, low hairline, no ptosis,	BCS	N	N	N	Karyotype N L1 L2 MLPA	1 (PR→ FOAR)	1	Mild language delay	Father has flat forehead, mother adopted

5 OX4481	M	10	Mild turri-brachycephaly, slight ptosis, high arched palate	BCS (2 Wormian bones)	N	N		Karyotype N L1 L2 MLPA *TCF12*-negative	1	1.5	N	Brachydactyly in proband, and 4 relatives; mother has very broad thumbs ? partial agenesis of corpus callosum; 2x ICPM, borderline raised pressure
6 OX797	F	18	Mild brachycephaly, brow posterior, mid-face normal, infra-temporal bulge	BCS (Copper beating on CT)	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1 (PR→ FOAR)	1	1	N
7 OX2636	M	42	Significant brachycephaly and forehead recession	BCS	N	N	N	Karyotype N L1 L2 MLPA	None – presented in adulthood with other co-morbidities, not for operative intervention	**0**	**0**	**N**(Significant copper beating seen, ballooning of middle cranial fossa, consistent with chronic raised intracranial pressure)
8 OX3245	F	11mo	Frank BCS with ridging of coronal sutures, turricephaly, temporal bulging	**BCS**	**N**	**N**	**N**	Karyotype N L1 L2 MLPA *TCF12*-negative	1	1yr2m	Functioning at average range all areas, no concerns (Griffiths MDS /Bayley score(1996)	Midface and maxillary hypoplasia

9 OX3245	F	3.5	High flat and prominent forehead, BCS with temporal bulging; mild brachycephaly, distinct turricephaly	BCS +S (copper-beaten cranium suggestive of raised intra-cranial pressure)	**N**	**N**	**N**	Karyotype N L1 L2 MLPA *TCF12*-negative	**1**	4yr3m	Functioning within average range (WIPPSI-R)	Abnormal head shape noted at 1yo, delayed referral, aggressive behaviour 2-6mths pre-consultation, raised intracranial pressure, no issues post-op
10 OX5098	F	5.5	Mild turri-brachycephaly, Slightly prominent forehead, turricephaly mild ptosis,	BCS + S (copper beaten appearance)	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	6yr2m	Above average functioning in WPPSI-3	
11 OX1584	M	5mo	Frank BCS, brachycephaly, mild turricephaly, ridged coronal sutures, no ptosis	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	10m	Extra support at school for hand writing; occasional inconsistent speech errors	Class II malocclusion
12 OX4462	F	6	Turri-brachycephaly	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	1yr	Normal developmen t	
13 OX1799	F	56	Brachycephalic	RUCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	0	n/a	Familial CRS, daughter has RUCS	
14 OX2438	F	9m	Flat forehead, enlarged right orbit	RUCS	N	N	N	Karyotype N L1 L2	1	1yr1m	Autistic spectrum disorder	Mother had LUCS

							MLPA *TCF12*-negative					
15 BCH5656	F	7	Brachycephalic with palpable bicoronal ridges	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	2	7m (PR)	No developmental concerns	Nil of note
16 OX6041	F	3	Brachycephalic, some brow recession, mild temporal bossing	BCS (partial)	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	11m	No Speech/ language concerns; bilingual (English/ Tamil) REEL 3 Receptive = 97%/ Expressive = 100%	No parental consanguinity reported
17 OX5569	F	1.5	Ridged coronal sutures, exorbitism (but protecting eyes), small anterior fontanelle, large, soft non-bulging posterior fontanelle, mild turricephaly	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	1yr9m	No Speech/ language concerns	Older brother with cerebral palsy and microcephaly
18 BCH6136	F	6	Clinically non-syndromic	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	PR at 4.5mo FOAR at 1yr7m	No Speech/ language concerns	Nil
19 OX4809	M	10	Mild Saethre-Chotzen phenotype	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	1yr	No Speech/ language concerns	Possibly affected first-degree relative. Orchidopexy x2

20 OX4580	F	30	Mild Saethre-Chotzen phenotype	RUCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	8mo	No Speech/ language concerns	Father unaffected, mother deceased
21 OX4400	F	8	Mild Saethre-Chotzen phenotype	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	2	1yr (PR) 1yr11m (FOAR)	No Speech/ language concerns	Mother has mild bilateral blepharoptosis, bilateral small finger clinodactyly, low anterior hairline (-ve testing)
22 GOS6568	F	2	Flat forehead and occiput, wide temples, narrow face	BCS	N	N	N	Karyotype N L1 L2 MLPA *TCF12*-negative	1	1yr	No Speech/ language concerns	Nil

Table 2.1 Summary of Clinical Characteristics of 22 patients with Coronal Synostosis and negative genetic testing deemed suitable for whole-exome sequencing. Level 1 and 2 (L1 and L2 respectively) refers to screening for mutational hotspots in craniosynostosis including as a minimum *FGFR1*, *FGFR2* (exons IIIa and IIIc), *FGFR3* and *TWIST1* (for full screening algorithm see Johnson & Wilkie, 2011). FOAR = fronto-orbital advancement and remodelling; PR = posterior release/posterior calvarial distraction.

PATIENT NUMBER	GENDER	AGE AT ASSESSMENT (YR/M)	CLINICAL PRESENTATION	SUTURE FUSION ON 3DCT	HANDS	FEET	BRAIN	NEUROCOGNITIVE	ADDITIONAL FINDINGS
OX1664 (MOTHER OF 1662)	F	28	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
OX1663 (FATHER OF 1662)	M	30	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
OX1590 (MOTHER OF 1560)	F	28	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
OX1591 (FATHER OF 1560)	M	31	Clinically normal phenotype	n/a	N	N	n/a	Nil	Myotonic dystrophy in father, paternal uncle, paternal grandfather (child tested negative)
OX3374 (MOTHER OF OX3245)	F	32	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil

OX3375 (FATHER OF OX3245)	M	34	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
BCH4972 (MOTHER OF BCH5656)	F	31	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
BCH4973 (FATHER OF BCH5656)	M	42	Clinically normal phenotype	n/a	N	N	n/a	Nil	Nil
OX6040 (MOTHER OF OX6041)	F	32	Clinically normal phenotype	n/a	N	N	n/a	Nil	South Asian ancestry
OX6039 (FATHER OF OX6039)	M	35	Clinically normal phenotype	n/a	N	N	n/a	Nil	South Asian ancestry

Table 2.2 Summary of Clinical Characteristics for the parents of 5 mutation-negative patients analysed by exome sequencing of patient-parent trios. Level 1 and 2 (L1 and L2 respectively) refers to screening for mutational hotspots in craniosynostosis including as a minimum *FGFR1*, *FGFR2* (exons IIIa and IIIc), *FGFR3* and *TWIST1* (for full screening algorithm see Johnson & Wilkie, 2011).

2.3 Reagents

Chemical reagents were obtained from Agilent, Sigma-Aldrich and IDT. Enzymes used for PCR and restriction digests were obtained from Roche and New England Biolabs, respectively.

2.4 Centrifugation

Samples and prepared reactions were spun on an Eppendorf Centrifuge 5810R or Eppendorf 5415D Microfuge.

2.5 Extraction of DNA

2.5.1 DNA extraction from venous blood

1.3-4.0ml of whole venous blood collected from a peripheral vein of a patient into a plastic collection tube with EDTA preservative. The NUCLEON® Blood Non Cholorform DNA Extraction kit (Gen-Probe Life Sciences Ltd, UK) was used to extract genomic DNA. Cell preparation proceeded with the whole blood being transferred to a 15ml polypropylene centrifuge tube and 2 volumes of Reagent A added. The tube was then inverted for 4 min at room temperature to mix the contents, then centrifuged at 3500g for 5 minutes. The supernatant was discarded and a further 2.5ml of Reagent A was added and vortexed to resuspend the pellet for 1 minute. The tube was then centrifuged at 3500g for 5 minutes and the supernatant discarded without disturbing the pellet. For cell lysis, 350µl of Reagent B is added to the pellet and the sample vortexed to resuspend it for 1 min. To deproteinise the sample, 125 µl of Reagent C is added and the tube is inverted 7 times or more to mix. Then, 100 µl of Nucleon Resin is added

drop-wise at the top of the sample and the tube centrifuged at 3500g for 4 minutes. To precipitate the DNA, the supernatant is decanted to a clean 15 ml polypropylene centrifuge tube without disturbing the pellet. To the new tube, 1 volume of 100% Propan-2-ol is added and the tube inverted several times until the DNA precipitate is visible. The pellet is then centrifuged at 4000g for 5 min, and the supernatant discarded. Finally, the DNA is washed by adding 500 µl of 70% ethanol to the tube contents, ensuring the pellet is dislodged from the bottom of the tube. It is then centrifuged at 4000g for 2 min and the supernatant discarded. The pellet is air dried for 10 min at room temperature, then resuspended in TE buffer ($T_{10}E_1$ Buffer: 1M Tris-HCl and 0.5M EDTA) overnight before use. The DNA is quantified and 1-1.2 µl of a 20ng/ml dilution is used per PCR reaction.

2.5.2 DNA extraction from saliva

If venous blood could not be obtained from a patient or relative, 1-2ml of saliva was collected using the Oragene DISCOVER (OGR-250) system (DNA Genotek, USA). The saliva sample is mixed in the DNA Genotek kit by inversion and gentle shaking for a few seconds before incubation in a water bath at 50 ^{o}C for 1 hr. This incubation step is crucial to permanently inactivated nucleases and maximise the yield of DNA obtained. The whole sample is then transferred to a 15 ml centrifuge tube by pouring or pipetting. Then, 1/25th volume of PT-L2P is added and the sample mixed by vortexing for a few seconds. This reagent will cause impurities and inhibitors to be precipitated, turning the mixture cloudy. The tube is then incubated on ice for 10 min to remove further impurities and then centrifuged at 3500 *g* for 10 min at room temperature. This will reduce the

amount of turbid material that is carried over into the purified DNA. The clear supernatant is transferred to a new 15 ml centrifuge tube by pipetting and the pellet discarded. A small volume of supernatant should be left behind to avoid disturbing the pellet as it contains impurities. Then, 1.2x volume of 95% ethanol (room temperature) is added to the supernatant, which is then mixed by inverting 10 times. This will cause the DNA to precipitate and may appear as a clot of DNA fibres. The DNA is allowed to fully precipitate by allowing the sample to stand at room temperature for 10 min, before further centrifugation at 3500 *g* for 10 min. Precipitated DNA will be present as pellet a in the bottom of the tube and the supernatant is carefully discarded by pipetting. Then, 1 ml of 70% ethanol is added to the tube without disturbance of the pellet and the tube gently swirled before the ethanol is removed completely. This step can be aided by brief centrifugation for 1 minute at 3500 *g*. The DNA is rehydrated by adding 0.2-1 ml of TE solution and the sample vortexed for 30 sec to ensure any DNA smeared on the side of the tube is recovered. Incubation at room temperature overnight and vortexing will ensure complete rehydration of the DNA pellet and prevent inaccurate estimation of DNA concentration. Alternatively, the tube may be incubated at 50 ^{o}C for 1 hr, with intermittent vortexing. The rehydrated DNA is transferred to a 1.5 ml microcentrifuge tube and stored at -20^{o}C. The DNA is then quantified by UV absorbance and diluted to 20 ng/ml before use in PCR reactions.

2.5.3 DNA quantification

The concentration of genomic DNA that was extracted using the methods described above was quantified using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific, USA) by measuring UV absorbance at 260 nm and 280 nm. This measure figure at 260 nm, denoted by A_{260}, equates to approximately 50 µg/ml of double stranded DNA. A measure of the purity of DNA is given by the A_{260}/A_{280} ratio which should be as close to 1.8 as possible for pure DNA and will be decreased in the presence of impurities and contaminants such as protein or phenol from the extraction process. A drawback of this methods is that it is unable to distinguish between DNA and RNA and at low concentrations of DNA, can significantly overestimate the concentration. The sensitivity and range is typically 2ng/ ml - 15 µg/ml. A more accurate method for nucleic acid quantification is by using fluorescence-based dyes that specifically bind DNA, RNA or protein and quantifying the concentration using the Qubit® 2.0 Fluorimeter (Life Technologies, Thermo Fisher Scientific, USA). It has the advantage of being able to measure DNA and RNA in the same sample and has better accuracy and precision at low DNA concentrations, down to 10pg/µl. Its sensitivity and range is also greater covering a DNA concentration of 10pg/µl to 1 µg/ml.

2.6 Polymerase Chain Reaction (PCR)

PCR reactions were conducted using 1-1.2 µl of genomic DNA diluted to a working concentration of 20 ng/ul in a total volume of 20 µl. Reactions were performed in skirted 96-well low profile plates and sealed with

adhesive film. A master mix was prepared and multiplied up as required. For a 1x reaction, it comprised of 1 µl of 10 mM forward and reverse oligonucleotides (from Sigma-Aldrich or IDT), 1 µl of 2mM deoxynucleotide triphosphate (dNTPs),0.15 µl of FastTaq DNA polymerase (Roche), 2.5 µl of buffer (containing 15 mM TrisHCl (pH 8.0), 50 mM KCl, 2.5 mM $MgCl_2$) and 13.85 µl of MiliQ ultrapure water, with or without 10% DMSO (Dimethyl sulfoxide). The reaction was incubated on a BioRad DNA Engine Dyad Peltier Thermal Cycler. Cycling conditions consisted of an 8 min denaturation step at 94ºC, followed by 35 cycles of 94ºC for 30 s, annealing at 60-65ºC for 30 s and extension at 72ºC for 30 s, with a final extension at 72ºC for 10 min. A heated lid at 100ºC was used to prevent evaporative loss of PCR product. The primers used are given in Table 2, Appendix.

2.7 PCR product visualisation

Products of the PCR reaction were run out and imaged on 1-5% agarose gel (UltraPure™ Agarose, Life Technologies, Thermo Fisher Scientific, USA) depending on the size of the amplicon. As per product size, 3-15g of agarose was added to 300 ml of 1x TBE (89 mM Tris-base, 89 mM boric acid, 2 mM EDTA). The mixture is heated in a microwave until it turns clear, then 10 µl of Ethidium bromide is added before the cooled gel is poured. It will intercalate itself in the DNA, inserting between the base pairs. This staining will allow the amplified DNA to be visualised when transilluminated under UV light. The PCR product is combined with 6 µl of 3x loading buffer (diluted from a 6x mix of 0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll 400). The agarose gel is placed in a

tank containing 1x TBE buffer and a 100 bp DNA ladder (GeneRuler 100bp, Thermo Scientific) added to the first well to allow the amplicons to be sized. PCR product mixed with loading buffer is then added to the other wells along with a negative water control, and electrophoresis carried out at 90-150V.

2.8 Screening for mutations

2.8.1 Restriction digest of PCR products using endonucleases

Depending on the strength of PCR product visualised on gel electrophoresis, 8-12 µl was digested in a total reaction volume of 20 µl containing 1 µl of restriction enzyme, 2 µl of 10x buffer (as specified by the manufacturer), and 10-12 µl of MiliQ water. Incubation in a thermal cycler follows, with temperature and duration as per manufacturer guidelines. Digested product is then run out on an agarose gel with uncut PCR product for comparison, to look for presence of mutant bands, verifying the presence of a mutation. Suitable restriction enzymes are chosen by use of an in-house bioinformatic analysis tool ('Weapon of Mutation Detection', Computational Biology Research Group, University of Oxford) that then suggests a list of possible enzymes to use.

2.8.2 Dideoxy DNA sequencing

For direct sequencing of DNA, 1.2 µl of PCR product was purified by incubation with 5 µl of a diluted mixture of Exonuclease-1 (New England BioLabs, Ipswich, MA) and Shrimp Alkaline Phosphatase (Affymetrix, San

Diego, CA). This will clean up the PCR product by removing excess primers and dNTPs respectively. This mixture was heated to 37 ^{o}C for 15 minutes, then to 80 ^{o}C for 15 minutes to inactivate the enzymes.

Next, a 5 µl of mix of Big Dye (containing Taq DNA polymerase, standard dNTPs, labelled ddNTPs and buffer), one primer and water were combined with the amplified product and incubated on a thermal cycler. Cycling conditions included a 5 min denaturation step at 95^{o}C, followed by 24 cycles of annealing at 55^{o}C for 4 min. The product is then precipitated by adding 50 µl of 80% ethanol (at room temperature) to the 96-well plate and allowing it to stand at room temperature for 30-60 min. The plate is then centrifuged at 3600 rpm for 30 min, before turning the plate over onto tissue to remove the ethanol, and 45-50 µl of 70% ethanol at room temperature is added. A further centrifugation step at 3600 rpm for 5-10 min is performed, before the plate is inverted again to remove the ethanol, and the plate left to open to the atmosphere to dry for at least 60 min. Then, 15 µl per well of Hi-Di™ Formamide (Applied Biosystems®, Thermo Fisher Scientific, UK) is added to resuspend the sample prior to electrokinetic injection on a capillary electrophoresis system for sequencing (Applied Biosystems 3730 DNA Analyser, Thermo Fisher Scientific, USA).

2.9 Sample preparation for whole-exome sequencing

2.9.1 Overview

Broadly, the process of massively paralleled sequencing follows 3 main steps: capture, sequencing and analysis. Firstly, the genes of interest are isolated from a sample. Secondly, they are put through a high throughput sequencer. Thirdly, there is a special application of computing to the genetic data using a variety of bioinformatic techniques.

For each proband in the stratified cohort, 3 µg of DNA (extracted from whole blood) was used with the Agilent SureSelect Human All Exon Kit (v.2; 44 Mb) to capture exonic DNA from a library prepared from the genomic DNA. This enriched DNA was sequenced on an Illumina GAIIx platform with 51 bp paired-end reads. As a result, 3.98- 4.56 Gb of sequence for each sample was produced and fed into a custom bioinformatic pipeline. After mapping with Novoalign (Novocraft Technologies) to the hg19 genome and removing artefacts by utilising using custom perl scripts, an average coverage of 33-44 fold per exon was obtained. Variants calls were made by SAMtools (Li et al, 2009) with annotation using the latest ANNOVAR (Wang et al, 2010) update, in addition to PolyPhen-2 (Adzhubei et al, 2010). Variants found in dbSNP134 were removed, followed by manual examination of gene lists for variants present in 2 or more samples. This will be discussed in more detail in the sections that will follow and is adapted from the SureSelect Target Enrichment System for Illumina Paired-End Sequencing – Human All Exon (Agilent Technologies). The workflow is summarised in below:

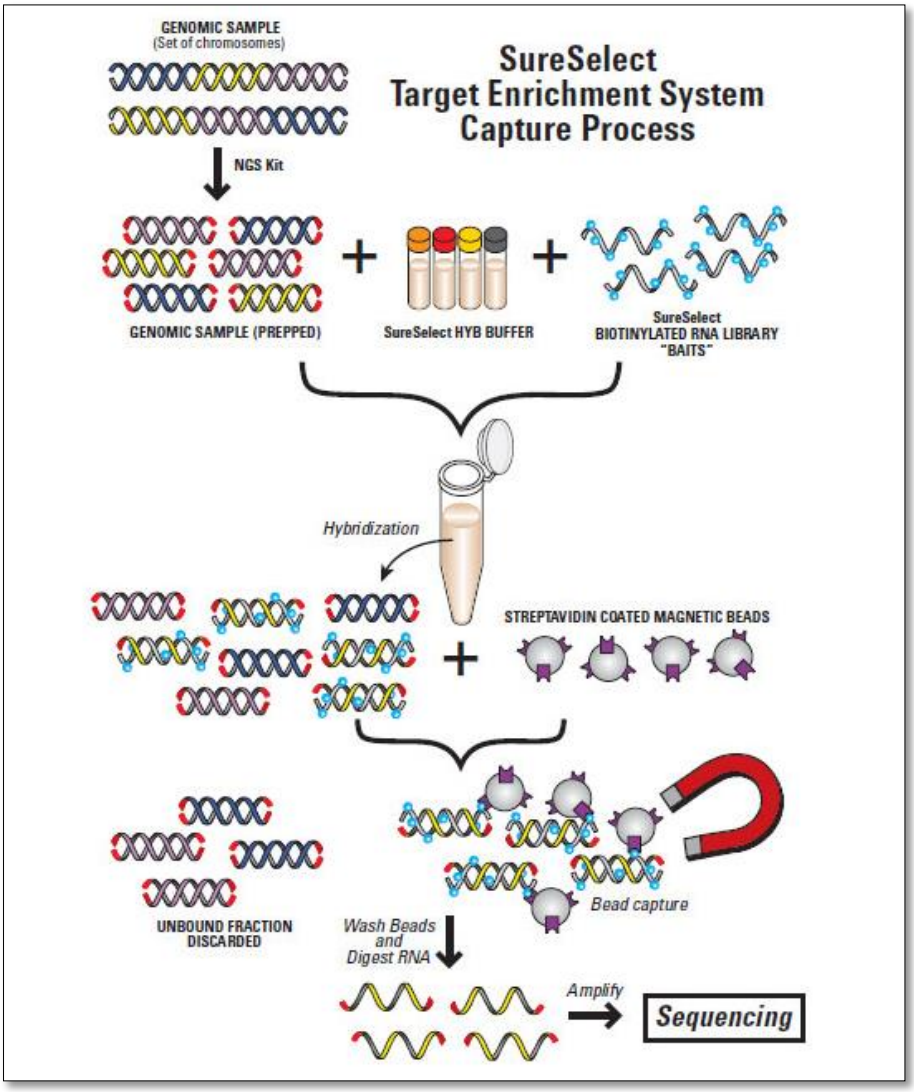


Figure 2.2: Summary of sample preparation for whole-exome sequencing using SureSelect Target Enrichment kit (Agilent Technologies). Genomic DNA is sheared by sonication to produce small fragments that are then repaired and ligated with adaptors. It is then hybridised with probes to all exons, miRNA and ncRNA (approximately 1-2% of the whole genome). The exonic sequences are then pulled down by bead capture before further amplification and quality assessment steps. Finally, the prepared samples are sent for DNA sequencing and data analysis. From Agilent Technologies product website [(Available at: http://www.genomics.agilent .com/article.jsp?pageId=3083&_requestid=502210 [Accessed 16/7/15)].

2.9.2 Sample Preparation

Each patient sample for genomic DNA is quantified using the Qubit 2.0 fluorimeter as described in 2.5.3. A minimum of 3 µg of high quality genomic DNA is required (A_{260}/A_{280} ratio of 1.8-2.0). The DNA is sheared using the Covaris S2 Focused-ultrasonicator and Covaris SonoLab 7™ software, in borosilicate glass Covaris microTUBE vials designed to have low impedance and improved acoustic energy transmission (Covaris, Wobrun, MA, USA).

2.9.2.1 Shearing the DNA

The water tank on the Covaris device is filled with fresh deionised water up to fill level 12 and the chiller temperature set to 2-5^{o}C. This is left for at least 60 min to allow the temperature to reach 4^{o}C. Then, 3 µg of genomic DNA is added to a 1.5 ml LoBind tube (Eppendorf AG, Hamburg) and diluted with 1x Low TE Buffer to a total volume of 130 µl. The mixture is then transferred to a Covaris microTUBE slowly through the pre-split septa on top, with extra care taken not to introduce an air bubble into the tube. The microTUBE is placed in the tube holder and the DNA sheared with the following settings: Duty Cycle = 10%; Intensity = 5; Cycles per burst = 200, Time = 6 cycles of 60 seconds each; Mode = Frequency sweeping; Temperature = 4^{o}C-7^{o}C. The sheared DNA is removed slowly from the microTUBE and transferred to a fresh 1.5 ml LoBind tube. The process of shearing the DNA creates smaller fragments of 150-200bp to allow further downstream processing and facilitate pair-end sequencing.

2.9.2.2 Purification of sample with magnetic beads

The sheared DNA is then purified using magnetic Agencourt AMPure XP beads (Beckman Coulter, Beverly, MA). The beads are allowed to stand at room temperature for 30 min, before mixing to ensure the reagent is homogenous. Then, 180 µl of the beads are added to 1.5 ml LoBind tubes. The sheared DNA is then added and each tube vortexed and incubated in a magnetic stand for 5 min until the solution clears. The cleared solution is then carefully discarded without disruption to the beads. They are then washed with 500 µl of fresh 70% ethanol. The tubes are left for 1 min before the ethanol is discarded. The ethanol wash is then repeated once more. The samples are then dried on a heat block at 37^{o}C for 2 min until the residual ethanol evaporates completely. Then 50 µl of nuclease-free water is added to each tube and each sample vortexed to mix, then incubated at 2 min at room temperature in the magnetic stand. Then, 50 µl of supernatant is removed and transferred to fresh LoBind tubes and the beads discarded.

2.9.2.3 Assess the quality of the sample

The quality of each sample is assessed using the 2100 Bioanalyzer device, Agilent 2100 Expert Software (vB.02.02 or higher), Bioanalyzer DNA 1000 chip and reagent kit (Agilent Technologies, USA). The analyser electrodes are cleaned prior to use and the DNA chip is prepared as per guidelines with the samples, ladder and other reagents. With the chip loaded, the run on the 2100 Bioanalyzer must commence within 5 min, choosing the appropriate assay from the drop down menu on the software program. The

run is commenced and the samples labelled in the data and Assay context dialogue box. Ensure the subsequent electrophenogram shows a distribution across all samples analysed with a peak in the size range of 150– 200 bp. Alternatively, the samples may be run out on a 3% agarose gel.

2.9.2.4 Repair the ends

The DNA fragments are repaired as follows. The reaction mix is prepared on ice. The volumes for the preparation of library are as follows: 48 µl DNA sample, 35.2 µl nuclease-free water, 10 µl 10x End repair buffer, 1.6 µl dNTP Mix, 1 µl T4 DNA polymerase , 2 µl Klenow DNA polymerase, 2.2 µl T4 Polynucleotide Kinase. The mix is incubated in a thermal cycler for 30 min at 20^{o}C without a heated lid. The T4 DNA polymerase is used to trim or fill in the 3' overhangs and catalyses 3' → 5' synthesis from primed single-stranded DNA. It possesses 3'→5' exonuclease activity but lacks 5'→3' exonuclease activity. The Klenow DNA polymerase retains its polymerisation and 3'→5' exonuclease activity, but has not retained its 5'→ 3' exonuclease activity. It fills in the 3' ends of the DNA that has receded to make the 5' overhang blunt and will digest protruding 3' overhangs. It keeps its polymerisation fidelity, without degrading the 5' termini, therefore repairing 5' overhangs to generate blunt ends. Finally, the T4 Polynucleotide Kinase is a dual function enzyme with 5'-kinase and 3'-phospate activity used to repair DNA. Its kinase activity causes the transfer of inorganic phosphate (P_i) from ATP to the 5' OH terminus of DNA. It will also catalyses the removal of 3´-phosphoryl groups from 3´-phosphoryl

polynucleotides, deoxynucleoside 3´-monophosphates and deoxynucleoside 3´-diphosphates.

2.9.2.5 Purify the sample with magnetic beads

The reaction mix is purified as per 2.10.2.2 except this time for the addition of 32 µl of nuclease-free water after the second drying step.

2.9.2.6 Add the 'A' bases to the 3' end of the DNA fragments

Addition of A bases to the reaction mix then performed over ice. For 1 library, the following volumes are used, mixing well by vortexing: 11 µl of nuclease-free water, 5 µl of 10X Klenow Polymerase Buffer, 1 µl of dATP, and 3 µl of Exo(-) Klenow. 20 µl of the reaction mix is added to each well, and then 30 µl of each DNA sample. The combined mixture is incubated on a thermal cycler for 30 min at 37ºC without a heated lid. The purpose of this step is to add 'A' bases to the 3' end of the blunt phosphorylated DNA fragments and prevent the formation of adaptor dimers and concatamers. The Exo(-) Klenow enzyme is a DNA-dependent DNA polymerase that lacks both the 5´→3´ nick-translation and 3´→5´ proof reading exonuclease activities from its originally derived DNA Polymerase I. It has an N-terminal truncation with 2 mutations (D355A and E357A) that abolishes its 3´→ 5´ exonuclease activity.

2.9.2.7 Purify the sample with magnetic beads

The reaction mix is purified as per 2.10.2.2 except this time for the addition of 15 µl of nuclease-free water after the second drying step.

2.9.2.8 Ligate the indexing-specific paired-end adaptor

During this step, a 10:1 molar ratio of adaptor to genomic DNA is used based on a 3 µg starting concentration. For 1 library, the following ligation master mix is prepared over ice: 15.5 µl of nuclease-free water, 10 µl 5X T4 DNA Ligase buffer, 10 µl SureSelect Adaptor Oligo Mix, 1.5 µl T4 DNA Ligase. To this, 13 µl of DNA sample is added. This step is responsible for adding distinct 3' and 5' sequences to the DNA sample to prepare it prior to the PCR library amplification step.

2.9.2.9 Purify the sample with magnetic beads

The reaction mix is purified as per 2.10.2.5 with the addition of 32 µl of nuclease-free water after the second drying step.

2.9.2.10 Amplify the adaptor-ligated library

This step will selectively enrich the prepared DNA sample for those molecules with adaptor molecules on each end. It will add additional sequences to the end of the adaptors for hybridisation and amplify the amount of DNA in the library. One-third of the adaptor-ligated fragments

are used for amplification, with the remainder stored at -20ºC for future use. The reaction mix is prepared on ice and mixed well by vortexing. The following volumes are used for the preparation of 1 library: 15 µl of Indexing Adaptor-ligated library, 21 µl nuclease-free water, 1.25 µl of SureSelect Primer, 1.25 µl of SureSelect ILM Indexing Pre Capture PCR Reverse Primer, 10 µl of 5X Herculase II Rxn Buffer, 0.5 µl of 100mM dNTP Mix, 1 µl of Herculase II Fusion DNA Polymerase. 35 µl of reaction mix is added to each tube followed by 15 µl of each DNA sample. The reaction mix is then placed in a thermal cycler as per the following program: 98ºC for 2 min, 98ºC for 30 sec, 65ºC for 30 sec, 72 ºC for 1 min, repeat previous 4 steps, 4-6 times, then 72ºC for 10 min and a final 4ºC hold step. The polymerase used in this step provides high-fidelity polymerisation of the complex DNA template that is being amplified, especially if GC-content is high, along with fast cycling times.

2.9.2.11 Purify the sample with magnetic beads

The amplified DNA sample is purified as described above, this time with 90 µl of magnetic beads, and 30 µl of nuclease-free water.

2.9.2.12 Assess the quality of the sample

The quality of the sample is assessed as per 2.10.2.3 and ensure the results on the electrophenogram demonstrate bands with a peak in the size range of 250-275 bp.

2.9.3 Hybridisation

The prepped library is then hybridised with biotinylated RNA library baits.

2.9.3.1 Hybridise the library

One hybridisation and capture is performed per library sample without pooling at this stage. The reaction requires 500 750 ng of DNA upto a maximum volume of 3.4 µl. If the concentration is below 221 ng/ µl, then the sample may be further concentrated by utilisation of a vacuum concentrator at ≤ 45 ^{o}C and reconstituted with nuclease free water so that the final concentration is ≥ 221 ng/ µl. The hybridisation buffer is prepared at room temperature as per the following volumes for 1 capture that accounts for excesses, making a total volume of 49 µl of which 40 µl is actually required: 25 µl SureSelect Hyb#1, 1 µl SureSelect Hyb#2, 10 ul SureSelect Hyb#3, 13 µl SureSelect Hyb#4. This capture library mix is then placed on ice, ready for target enrichment. Depending on the capture size, either 2 µl (<3.0Mb) or 5 µl (>3.0Mb) of library is added to the hybridisation buffer. Using nuclease free water, a 1:9 (10%) or 1:3 (25%) RNase Block dilution is prepared and 5 µl or 2 µl of RNase block dilution is added. The RNase Block is added to each capture library with thorough mixing by pipetting. The block mix is prepared as follows, to a total of 5.6 µl for 1 reaction: 2.5 µl of SureSelect Indexing Block#1, 2.5 µl of SureSelect Block #2, 0.6 µl of SureSelect Indexing Block #3.

In a separate PCR plate, the prepped library is prepared for target enrichment. To the plate, 3.4 µl of 221ng/ µl prepped library is added to row B, with each sample placed in a separate well. Then, 5.6 µl of the Block mix is added to each library sample in row B and mixed well by pipetting, before sealing with strip caps and placing the whole PCR plate on a thermal cycler as per the following program: 5 min at 95ºC, then hold at 65ºC (with a heated lid at 105ºC). While keeping the plate in the thermal cycler at 65ºC, 40 µl of hybridisation buffer is added to row A of the plate, equivalent to the number of libraries that have been made. The plate is then kept at this temperature for a minimum of 5 min.

The capture library mix described previously, is added to row C, with 7 µl per well added, keeping the plate at 65 ºC. The wells in this row are sealed with strip caps and the plate incubated for 2min. After this, 13 µl of hybridisation buffer from row A is added to the capture library mix in row C and mixed thoroughly by pipetting. The entire contents from row B (prepped library mix) is added to row C (hybridisation solution), and then mixed well by slowly pipetting up and down 10 times. The total volume of hybridisation mixture should be 27-29 µl dependent on evaporative losses during incubation. The wells are sealed with new strip caps and double adhesive film before the hybridisation mixture is incubated for 16-24 h at 65 ºC in the thermal cycler with a heated lid at 105 ºC. This is summarised in the diagram below.

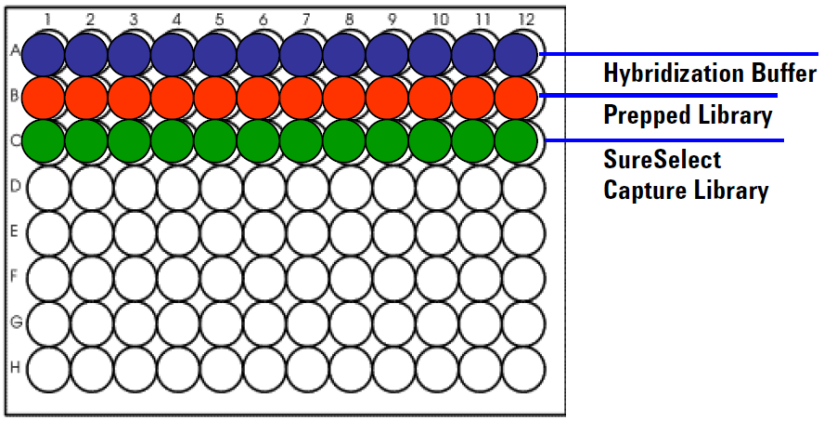


Figure 2.3: Summary of steps to hybridisation of the library. This uses prepped library (red), hybridisation buffer (blue) and SureSelect Capture Library (green).

2.9.3.2 Preparation of magnetic beads

This step requires SureSelect Binding Buffer and SureSelect Wash 2 from the SureSelect Target Enrichment Kit Box#1 (Agilent Technologies, USA) the latter of which is pre-warmed in a water bath at 65 ^{o}C. Then, the Dynabeads® MyOne Streptavidin T1 (Invitrogen, USA) magnetic beads are re-suspended by vigorous mixing on a vortex mixer. These beads are used to separate biotinylated nucleic acids and have a very high streptavidin-biotin binding affinity. To fresh 1.5 ml microfuge tubes, 50 µl of beads are added. The beads are then washed with 200 µl of SureSelect Binding Buffer, mixed on a vortex for 5 sec, then the tubes placed in a magnetic separator. The supernatant is then removed and the washing step repeated 3 times further. The beads are then resuspended in 200 µl of the buffer.

2.9.3.3 Hybrid capture with SureSelect

Following the overnight incubation, the volume of hybridisation mixture remaining is estimated and more than 20 µl should remain if capture performance has been optimal. The PCR plate is kept on the thermal cycler at 65 ^{o}C while the hybridisation mixture is added directly to the bead solution, after which the tube is inverted 5 times to mix. The tubes are then incubated on a nutating or rotating mixer at room temperature for 30 min. After brief centrifugation, the tubes containing the hybrid-capture bead solution are placed in a magnetic stand and the supernatant discarded. The beads are then resuspended in 500 µl of SureSelct Wash 1 and vortexed for 5 sec to ensure proper mixing. The tube is briefly centrifuged and the placed back on the magnetic stand and supernatant discarded. A final washing step follows whereby the beads are resuspended in 500 µl of SureSelect Wash 2 that has been pre-warmed to 65 ^{o}C and mixed for 5 sec. The samples are incubated in a water bath for 10 min at 65 ^{o}C with occasional mixing by vortexing, briefly centrifuged, then placed in a magnetic stand to separate beads and buffer. This washing step is repeated 3 times and care taken to ensure all the wash buffer has been removed. Finally, the beads are mixed with 30 µl of nuclease-free water on a vortex mixer for 5 sec to resuspend the beads.

2.9.4 Post-Hybridisation Amplification and Cluster Station Preparation

In the next 5 steps, there will be addition of index tags by amplification, followed by purification and assessments of library quality and quantity,

and then sample dilution for cluster amplification and pooling of the samples for multiplexed sequencing.

2.9.4.1 Amplify captured library and add index tags

The reaction mix is prepared over ice according as per the following for 1 reaction, up to a total volume of 50 µl: 14 µl of captured on-bead DNA, 22.5 µl of nuclease-free water, 10 µl of 5x Herculase II Rxn Buffer, 0.5 µl of 100 mM dNTP mix, 1 µl Herculase II Fusion DNA polymerase, 1 µl SureSelect ILM Indexing Post Capture Forward PCR primer, 1 µl of PCR Primer Index 1-16 (depending on number of samples). To 35 µl of reaction mix, 1 µl of index primer is added as appropriate, with a different one used for each sample. A pipette is used to add 14 µl of each DNA sample to each well, and the PCR plate placed on a thermal cycler and the following program is run: 98^{o}C for 2 min, 98^{o}C for 30 sec, 57^{o}C for 30 sec, 72^{o}C for 1 min, repeat previous 3 steps 10-12 times (for All Exon kit), 72^{o}C for 10 min, hold at 4^{o}C.

2.9.4.2 Purify the sample with magnetic beads

The reaction mix is purified as described in 2.10.2.11.

2.9.4.3 Assess the quality of the sample

The quality of the sample is assessed as per 2.10.2.3 to ensure the results on the electropherogram demonstrate bands with a peak in the size range of 300-400 bp.

2.9.4.4 Pool samples for Multiplexed Sequencing

The libraries are combined so that each sample that has been index-tagged will be pooled together in equimolar amounts. The volume of index sample to use is calculated as per formula below:

$$\text{Volume of Index to use} = \frac{V(f) \times C(f)}{\# \times C(i)}$$

V (*f*) = final desired volume of pool

C (*f*) = final desired concentration of all DNA in the pool

= number of index

C (*i*) = initial concentration of each index sample

The final volume of the pooled library is adjusted to the final concentration. TE is added if the final volume combined index-tagged samples is lower than the desired volume. If it is greater than the final volume, the sample is lyophilised and reconstituted as necessary.

2.9.4.5 Prepare samples for cluster amplification

This and subsequent steps were performed at the Wellcome Trust Centre for Human Genetics (University of Oxford). During cluster generation, libraries will be turned into clonal clusters on a flow cell. Each sample is treated identically, denatured and hybridised to a flow cell and captured DNA will form a template for synthesis of a second strand which is then amplified to form a clonal cluster. Addition of a sequencing primer forms a locus for sequencing-by-synthesis. This process will result in double stranded DNA. Clusters are generated in a contained environment on a flow cell that is coated with a lawn of oligonucleotide pairs. Sequencing is performed on the clusters on the flow cell. Single DNA molecules are clonally amplified into clusters using a cluster station which automates the process and aspirates DNA samples into the flow cell. The library is initially denatured so that only a single stranded molecule is sequenced and these will bind randomly to the surface of the flow cell. Each library contains flow cell binding sites (P5 and P7) in their adapter region that allow the DNA library molecule to anneal to complimentary oligonucleotides on the surface of the flow cell, which itself will act as a primers for strand synthesis. After washing away the original strand, only those library fragments that are covalently bonded to the surface of the flow cell remain. Then, by process of bridge amplification, 1000 copies of each fragment are generated, making clusters. Cleavage of the P5 region will results in clusters that are adhered only by their P7 region, ensuring all will be sequenced in the same direction. A sequencing primer anneals to the P5

end of the molecule and sequencing-by-synthesis begins by addition of 4 labelled reversible terminators, primers and DNA polymerase.

The first base is imaged by laser excitation and the fluorescence that is emitted is captured and the first base determined. After Read 1 has finished, all associated reagents are removed and an index primer is added that anneals to the P7 site and the barcode is sequenced. For Read 2, everything is removed from the template, which then forms further clusters by bridge amplification leaving fragment copies that are bonded to the surface of the flow cell in different orientations. Now, the P7 site is cleaved, leaving clusters attached by their P5 site, and so all copies are subsequently sequenced in the same direction. The P7 region is then attached by the sequencing primer and the other end of the template is sequenced. The cycles of sequencing are repeated multiple times to determine the sequences of bases in the fragment by capturing images of sequenced DNA therefore allowing the sequence to be determined. The data is then aligned and compared to a reference sequence and any differences identified. During data analysis, the raw data generated from the images or base calls is turned into raw sequence alignments.

2.10 Microsatellite analysis of apparent de novo mutations to confirm correct sample relationships/paternity

Genomic DNA for the probands and parents was obtained and diluted to 20 ng/µl working concentration. A PCR amplification across 13 microsatellites specifically chosen for their location and high heterozygosity. Cycling

conditions consisted of an 8 min denaturation step at 94°C, followed by 35 cycles of 94°C for 30 s, annealing at 63°C for 30 s, extension at 72°C for 30 s, with a final extension at 72°C for 10 min. Depending on strength of PCR product, 2-3 µl was then transferred to a new plate and sent for fragment analysis on an Applied Biosystems 3730 DNA Analyser. Sample relationships were demonstrated with correct segregation of at least 10 microsatellites. The following microsatellites were used: D1S286, D3S1311, D4S403, D5S2027, D6S1610, D7S519, D9S158, D10S548, D11S898, D13S1265, D14S280, D16S415 and D18S474. Using the ALFRED database (http://alfred.med.yale.edu/alfred/index.asp), the two commonest allele frequencies are identified and added together for an ethnic population most similar to the patient of interest. This is repeated for each of the other 12 microsatellites, then all allele frequencies are multiplied together and a final value calculated, giving a probability of >99% of correct paternity if all 13 segregate appropriately.

2.11 Allele-specific PCR for identification of parental origin of mutations

In order to ascertain the parent of origin for a *de novo* mutation, a standard approach is to use natural sequence variants or SNPs that are closely linked to the mutation and allow phase to be established, and a modified amplification refractory mutations system (ARMS) is used as previously described (Moloney et al., 1996). Firstly, a 500bp -1kB region surrounding each exon that contains the mutation is screened by dideoxy sequencing to

identify SNPs that are closely linked and might be informative in a patient-parent trio. An informative trio is one were the child is heterozygous for a SNP (e.g. [A/G]), one parent is also heterozygous (i.e. [A/G]) and another parent is homozygous (i.e. [G/G]). Alternatively, the parents can be oppositely homozygous (i.e. [A/A] and [G/G]) – see Table 7, Appendix.

The selective PCR of a single allele to detect a specific SNP occurs by selective amplification utilising a primer match/mismatch of one of the alleles at the 3'-end of the primer. This 3'-terminal mismatch will destabilise the intermediate primer-template complex and so will be refractory to primer extension. An allele-specific assay is designed using 3 primers: one common forward primer and two separate reverse primers, one of which with a specific mismatch to the penultimate base or two different forward primers (one with the mismatch) and one common reverse primer. A free to use web-based program is used to select and design suitable combinations of forward and reverse primers (WASP - a Web-based Allele Specific Primer designing tool, Genome Institute, BIOTEC; http://bioinfo.biotec.or.th/WASP). An appropriate restriction digest is then performed on the PCR product to determine if the wild-type or mutant allele has been selectively amplified and on which allele (maternal or paternal), the mutation has arisen. PCR and digest products are analysed on a 3% agarose gel, on a voltage of 90V. Sets of primers, conditions for selective amplification and restriction enzymes utilised are given in Table 8, Appendix.

2.12 Selection of tag-SNPs to allow genotyping of a population of interest using International HapMap data

The International HapMap Project database (http://hapmap.ncbi.nlm.nih.gov) was used as a resource to identify common patterns of variation in human DNA, (single nucleotide polymorphisms – SNPs), across a matched ethnic populations. An individual carrying a particular SNP at one locus may carry other certain alleles at nearby sites and this non-random association is a phenomenon referred to as linkage disequilibrium (LD) (International HapMap Consortium, 2005) and *TCF12* is located in a region of strong LD. Sets of adjacent SNPs along the same chromosome are inherited in blocks and the pattern of SNPs on a block is referred to as a haplotype. Although each block contains a large number of SNPs, specific ones can be used to identify haplotypes – tag SNPs.

The following window settings were used: chr15:54983065..55183064 landmark, HapMap Data Phase III/Rel#2, Feb09. Using the Tagger web browser (http://www.broadinstitute.org/mpg/tagger), 5 tag-SNPs were identified that capture 100% of alleles with a mean r^2 of 0.917 (rs477426, rs2585082, rs1628067, rs12901270, rs2440979). By using pairwise tagging, all 5 tag-SNPs selected act as a direct proxy for untyped SNPs due to a high correlation with each other.

CHAPTER 3

A NOVEL DISEASE GENE ASSOCIATED WITH CORONAL SYNOSTOSIS IS REVEALED BY WHOLE-EXOME SEQUENCING

3.1. Introduction

As described previously (see 1.3.1-1.3.5 and Table 1) many of the genes implicated in craniosynostosis syndromes with a defined monogenic basis encode transcription factors and growth factor receptors. Therefore, when considering a search to identify new disease genes, any variants subsequently identified in these two groups should be prioritised as likely candidates. Mutations of *TWIST1* cause Saethre-Chotzen syndrome (el Ghouzzi et al. 1997; Howard et al. 1997) the main clinical features of which include coronal synostosis, eyelid ptosis, low frontal hairline and small, posteriorly rotated ears with prominent crura. TWIST1 is a helix-loop-helix protein that forms either homodimers with itself or heterodimers with other E-proteins. These include E2A, E2-2 and HEB encoded by genes *TCF3, TCF4* and *TCF12* respectively (Kee, 2009). As a dimer, this molecule will bind to a consensus sequence on DNA (denoted by 5′-CANNTG-3′) and activate or repress transcription (Atchley and Fitch, 1997). Evidence from mouse models suggests a key role for *Twist1* in the developing coronal sutures that impacts on suture patency in both the sutural and ectocranial mesenchyme (Ting et al., 2009). It is at the top of a molecular regulatory hierarchy and by virtue of its antiosteogenic domain (Twist box) has an inhibitory effect on Runx2. It achieves this by interacting with its DNA binding domain, preventing osteoblast differentiation and ossification of the suture mesenchyme (Bialek et al., 2004).

3.2 Patients, materials and methods

Following extensive review of clinical records, 7 patients were shortlisted on the basis of key common features: coronal synostosis, normal limbs, brain, neurocognitive development and negative genetic testing (see Table 2.1). Genomic DNA was prepared according to the Agilent protocol using the SureSelect Human All Exon Kit (v.2; 44 Mb) (see Methods 2.9) and whole-exome sequencing of the 7 samples was performed. The metrics are shown in Table 3.1. The use and further refinement of an existing bioinformatic pipeline was crucial in processing the raw data into a useable format with visualisation of the data on a custom web browser (Stein et al., 2002). Scripts were written by Dr S McGowan (Computational Biology Research Group, Weatherall Institute of Molecular Medicine, University of Oxford) with the overall aim of identifying pathological changes that were absent from public databases of variants.

Firstly, the raw data (Step 1) were aligned to the human genome (GRCh37/hg19 assembly) using Novoalign (Novocraft Technologies) with all hits recorded (Step 2). Reads that did not map due to poor quality or mapping to repeat regions were discarded (Step 3). Artefacts were removed using custom perl scripts. The predicted depth of coverage was calculated by multiplying the number of reads that came through the pipeline (e.g. 62, 004, 714 for patient OX484) by 51 to convert it to number of bases. This figure is then divided by the exome size of 38MB used in this particular version of the SureSelect Human All Exon Kit kit, to give the avearge read depth stated in Table 3.1 (i.e. 83). In practice however, due to significant variations in coverage, the actual average coverage obtained was 33-44 fold

per exon. PCR artefacts were then removed (Step 4) so that variant SNPs were not overcalled. Pseudogenes were removed subsequently (Step 5) and a variant filter identified regions not in reference sequences and highlighted candidate variants. Variants calls were made by SAMtools (Li et al., 2009) with annotation using the latest ANNOVAR update (Wang et al., 2010) in addition to PolyPhen-2 (Adzhubei et al., 2010). During the subsequent filter pile-up, all variants that were found in dbSNP134/1000 Genomes databases were excluded and the remainder compared with other in-house exomes of previously analysed craniosynostosis patients. Finally, the data was translated to a web-based browser (Stein et al., 2002) to allow visualisation of the exome data and the number of patients sharing particular gene variants identified.

			Number of Reads						
Sequence data processing in Round 1 of Exome Sequencing			**OX484**	**OX1662**	**OX1560**	**OX4581**	**OX4481**	**OX797**	**OX2636**
1	**Input**	**Raw data obtained from 51 bp paired-end sequencing**	53144400 (x2)	48544560 (x2)	47763120 (x2)	52822320 (x2)	44967960 (x2)	45374040 (x2)	56051280 (x2)
2	**MAP_SAM_OUTPUT: bowtie -a**	**Map to whole genome – all hits recorded**	116062026	108450438	107088788	114853288	101667618	102278508	117515982
3	**GET_HITS: get_mapped.pl**	**Remove unmapped sequences**	108303834	101619726	99630886	106622626	92668534	95972490	107193860
4	**GET_UNIQUE_PAIRS: parse_sam5.pl**	**Remove PCR artefacts (only paired-reads with highest total Phred score are retained)**	84104248	88423380	70592851	78811644	67602873	85689325	65943076
5	**FIND_BEST_HITS: find_best_hits.pl**	**Remove pseudogenes (if a read pair maps to >1 location only those which have the lowest number of mismatches over both ends and which map on or near an exome capture bait are retained)**	62004714	64069498	51982420	58013352	48588386	62659182	48717538
Variant analysis									
Actual read depth at 10x (% of exons)			67.6	68.7	66.6	67.9	65.7	67.1	66.1
Actual read depth at 30x (% of exons)			41.3	44.1	35.7	40.7	33.7	38.9	35.8
Total number of variants			104829	103455	101206	101684	99343	108364	90382
Number of exonic variants			37835	38533	38549	38603	37761	38093	36819
dbSNP134/1000 Genomes filtering			1081	1000	1233	1127	1163	1304	979
Shared gene variants									
Number of patients sharing gene variants			7	6	5	4	3	2	
Number of genes			0	0	0	0	15	65	

Table 3.1 Summary of statistics for exome sequencing of the 7 patients with bilateral coronal synostosis (Round 1).

3.3 Gene lists and shortlisted candidates

Following exome sequencing and after exclusion of unlikely variants present in all 7, 6, 5 or 4 patients, a list of 80 candidate genes was obtained. This included 15 genes with shared variants in 3 patients and 65 genes with shared variants in 2 patients. These were manually analysed for read depth and quality of the sequence surrounding a change. Likely candidate genes were further evaluated for roles in known craniosynostosis disease pathways. Variants in single genes were also shortlisted. The majority of these were then excluded and the top 20 candidate genes identified are shown in Table 3.2 below (see Table 1, Appendix for full gene list)

Patients affected	Gene	Cytogenetic Locus	SNV/Nucleotide change/Amino acid change	Polyphen2 Prediction	High quality forward and reverse reads (wild-type | mutant)	Gene Function (from www.genecards.com)
OX1560 OX2636	*CCDC41* - coiled-coil domain-containing protein 41	12q22	C > T; c.G76A:p.G26S G > T; c.C1009A:p.L337I		15,21 | 10,24 36,10 | 18,11	Component of distal appendage region of the centriole involved in the initiation of primary cilium assembly. Thought to collaborate with IFT20 to support the vesicle-centriole association at the onset of ciliogenesis.
OX4581 OX2636	*PRKDC* - DNA-dependent protein kinase catalytic subunit	8q11.21	C > T; c.G1526A:p.R509H T > C; c.A6341G:p.K2114R	Benign	12,17 | 8,8 4,2 | 9,6	Serine/threonine-protein kinase that acts as a molecular sensor for DNA damage
OX1560 OX4581	*ZNF474* - zinc finger protein 474	5q23.2	T > -; c.953delT:p.L318X T > -; c.953delT:p.L318X		8,10 | 6,6 14,6 | 11,2	Zinc finger transcription factor
OX1560 OX2636	*ASMTL* - acetylserotonin O-methyltransferase-like	Xp22.3	G > A; c.C61T:p.R21W A > -; c.1690delT:p.X564E	Probably damaging	1,6 | 4,8 6,1 | 4,0	Unknown function. The presence of the putative catalytic domain of S-adenosyl-L-methionine binding argues for a methyltransferase activity
OX4581 OX2636	*PCNXL2* - pecanex-like 2 (Drosophila)	1q42.2	C > T; c.G4349A:p.R1450Q G > A; c.C3526T:p.H1176Y	Probably damaging Probably damaging	7,0 | 5,0 2,2 | 4,2	May play a role in tumorigenesis of colorectal carcinomas with high microsatellite instability (MSI-H)
OX1560 OX2636	*GPR111* - G-protein coupled receptor 111	6p12.3	T > G; c.T1350G:p.C450W T > C; c.T1850C:p.L617P	Probably damaging Probably damaging	6,11 | 8,18 7,7 | 2,7	Orphan receptor
OX484 OX797	*SYCE2* - synaptonemal complex central element protein 2	19p13.2	G > A; c.C265T:p.H89Y G > A; c.C265T:p.H89Y	Possibly damaging Possibly damaging	28,11 | 28,4 21,2 | 16,12	The protein encoded by this gene is part of the synaptonemal complex formed between homologous chromosomes during meiotic prophase. The encoded protein associates with SYCP1 and SYCE1 and is found only where chromosome cores are synapsed
OX1662	*SSH2* - Slingshot protein phosphatase 2	17q11.2	C > G; c.G1516C:p.V506L	Benign	15,15 | 18,11	A protein tyrosine phosphatase that plays a key role in the regulation of actin filaments. The encoded protein

OX4581			G > C; c.C1736G:p.P579R	Possibly damaging	17,9 | 18,9	dephosphorylates and activates cofilin, which promotes actin filament depolymerisation.
OX484 OX797	*SNORD50A* - unknown	6q14.3	AA > -; c.52_0del:p.18_0del AA > -; c.52_0del:p.18_0del		17,4 | 12,3 12,4 | 9,3	Belongs to the C/D box class of small nucleolar RNAs (snoRNAs), which are thought to function as guide RNAs in the site-specific ribose methylation of preribosomal RNA
OX484 OX2636	*ZBTB38* - zinc finger and BTB domain-containing protein	3q23	A > G; c.A1849G:p.N617D A > G; c.A1849G:p.N617D	Benign Benign	19,18 | 17,22 17,13 | 16,8	A zinc finger transcriptional activator that binds methylated DNA. The encoded protein can form homodimers or heterodimers through the zinc finger domains. In mouse, inhibition of this protein has been associated with apoptosis in some cell types.
OX484 OX1560	*MDP1,NEDD8-MDP1* - NEDD8-MDP1 Readthrough	14q12	A > -; c.352delT:p.F118fs A > -; c.352delT:p.F118fs		29,7 | 16,1 7,7 | 13,6	This locus represents naturally occurring read-through transcription between the neighbouring NEDD8 (neural precursor cell expressed, developmentally down-regulated 8) and MDP1 (magnesium-dependent phosphatase 1) genes on chromosome 14. The read-through transcript encodes a fusion protein that shares sequence identity with the products of each individual gene.
OX1560 OX2636	*ERBB2IP* - erbb2 interacting protein	5q12.3	G > A; c.G2857A:p.E953K G > A; c.G867A:p.M289I	Possibly damaging Benign	14,23 | 4,11 11,22 | 13,9	Acts as an adapter for the receptor ERBB2, in epithelia. By binding the unphosphorylated 'Tyr-1248' of receptor ERBB2, it may contribute to stabilize this unphosphorylated state; affects the Ras signalling pathway by disrupting Ras-Raf interaction.
OX484 OX4581	*CTAG2* - cancer/testis antigen 2	Xq28	G > -; c.476delC:p.S159fs G > T; c.C247A:p.P83T	 Benign	0,0 | 17,5 1,5 | 0,3	An autoimmunogenic tumour antigen that belongs to the ESO/LAGE family of cancer-testis antigens. This protein is expressed in a wide array of cancers including melanoma, breast cancer, bladder cancer and prostate cancer. This protein is also expressed in normal testis tissue.
OX4581 OX2636	*ZSCAN29* - zinc finger and SCAN domain-containing protein	15q15.3	C > A; c.G861T:p.R287S C > T; c.G2456A:p.G819E	Possibly damaging Benign	12,12 | 16,29 20,28 | 16,23	May be involved in transcriptional regulation
OX797 OX2636	*DIDO1* - death-inducer obliterator 1	20q13.33	G > A; c.C2255T:p.A752V G > C; c.C6169G:p.Q2057E	Benign Benign	21,44 | 20,40 0,11 | 3,14	Putative transcription factor, weakly pro-apoptotic when overexpressed
OX484 OX2636	*H6PD* - GDH/6PGL endoplasmic bifunctional protein	1p36.22	CCCAGGCA > -; H6PD:NM_004285:wholegene		14,2 | 5,3 0,0 | 4,0	Oxidizes glucose-6-phosphate and glucose, as well as other hexose-6-phosphates

			CCCAGGCA > -; H6PD:NM_004285:wholegene			
OX484 OX2636	*NT5C3* - 5'-nucleotidase, cytosolic III	7p14.3	G > -; c.243delC:p.N81fs G > -; c.243delC:p.N81fs		5,27 | 1,28 9,16 | 1,14	Can act both as nucleotidase and as phosphotransferase. Mutations in this gene are a cause of haemolytic anaemia due to uridine 5-prime monophosphate hydrolase deficiency
OX484 OX4581	*TCF12* - Transcription factor 12	15q21.1	C > -; c.1349delC:p.P450fs GAAA > -; c.1642_1645del:p.548_549del		7,19 | 4,13 9,5 | 6,4	Member of the basic helix-loop-helix (bHLH) E-protein family, recognises consensus binding site (E-box) CANNTG; expressed in many tissues (skeletal muscle, thymus, B/T-cells), and may participate in regulating lineage-specific gene expression through the formation of heterodimers with other bHLH E-proteins.
OX484 OX4581 OX797	*PRSS48* – protease serine protease 48	4q31.3	- > GTCAG; c.131_132insGTCAG:p.S44fs - > GTCAG; c.131_132insGTCAG:p.S44fs - > GTCAG; c.131_132insGTCAG:p.S44fs		17,5 | 16,4 16,3 | 7,4 10,3 | 14,2	A protein with serine-type endopeptidase activity.
OX484 OX1662 OX4481	*MARK1* - MAP/microtubule affinity-regulating kinase 1	1q41	T > G; c.T2386G:p.X796E T > G; c.T2386G:p.X796E T > G; c.T2386G:p.X796E		0,13 | 0,7 0,18 | 0,10 0,6 | 0,4	Serine/threonine-protein kinase involved in cell polarity and microtubule dynamics regulation. Involved in the regulation of neuronal migration and a positive regulator of the Wnt signalling pathway, probably by mediating phosphorylation of dishevelled proteins

Table 3.2 Summary of the top 20 candidates identified in Round 1 of exome sequencing demonstrating non-synonymous variants shared between 2 or 3 patients

3.4 Identification of damaging, loss-of-function mutations in a novel transcriptional regulator in 3 patients

Of the genes shortlisted in Table 3.2, the most outstanding candidate was Transcription Factor 12 (*TCF12*). This had different frameshift mutations in 2 patients, both of which cause loss-of-function and a truncated protein. The two mutation-positive samples had differing heterozygous mutations - patient OX484 had a single nucleotide deletion in exon 16 (c.1349delC; p.P450Lfs*69) and patient OX4581 had a 4 nucleotide deletion in exon 18 (c. 1642_1645delGAAA; p.E548Rfs*14). A text search of the exome sequence data was also performed manually in the remaining samples, revealing a novel intronic T>A variant (c.1468-20T>A) in patient OX2636 (see Figure 3.1). This was predicted to create a cryptic splice acceptor site *in silico* with a higher predicted score than the wild-type site (see Figure 3.2). All 3 mutations were validated by dideoxy sequencing of the DNA samples initially used for exome sequencing.

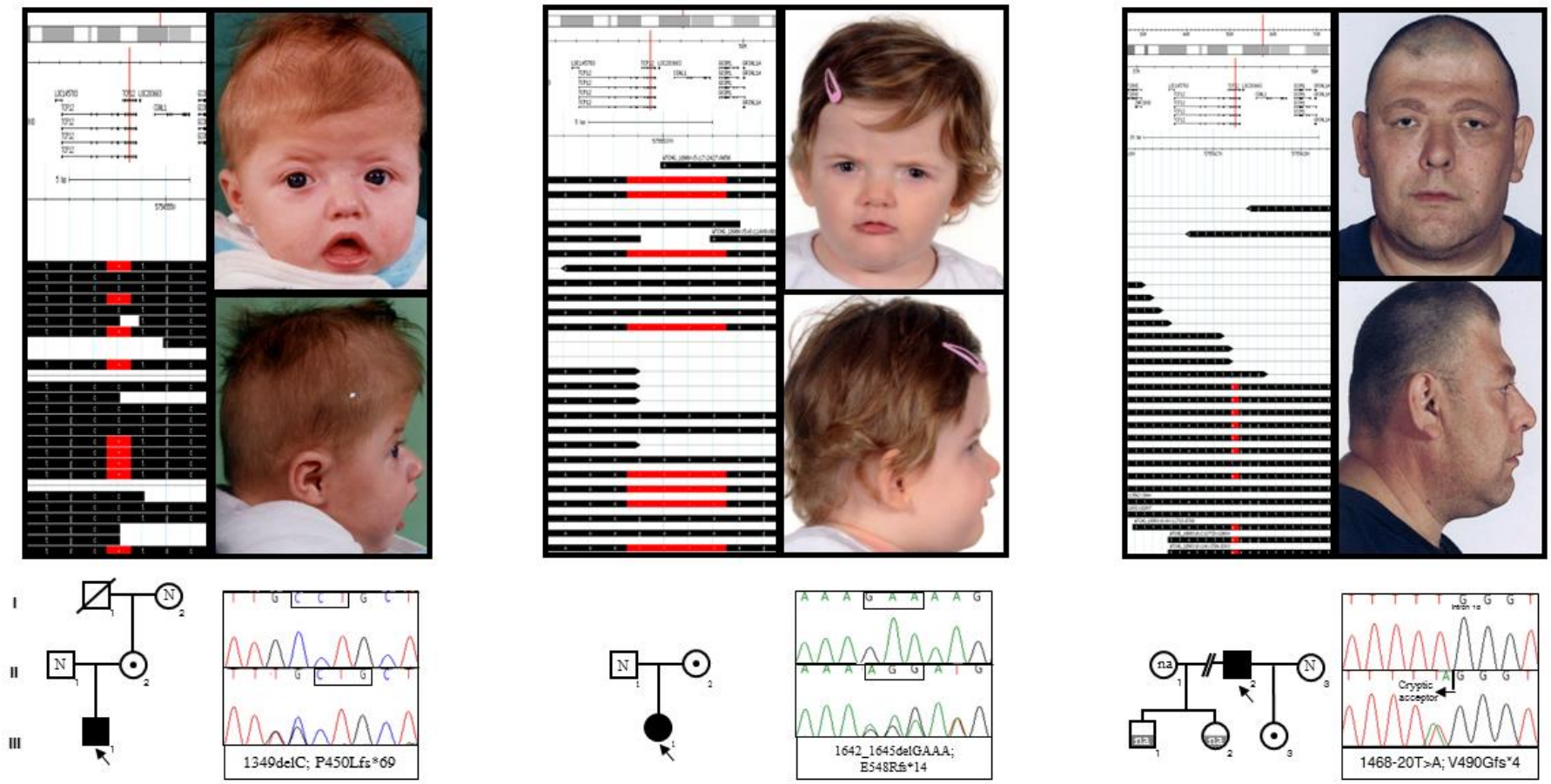


Figure 3.1 The 3 patients each with bilateral coronal synostosis and mutations in *TCF12* are shown. Patient OX484 has a single nucleotide deletion (Left); OX4581 has a 4 nucleotide deletion (Middle); OX2636 has an intronic nucleotide substitution (Right). Reads mapped normally to the human genome are in black, with variants indicated in red. Family pedigrees identifying further mutation carriers along with validation of each mutation by dideoxy sequencing are also shown.

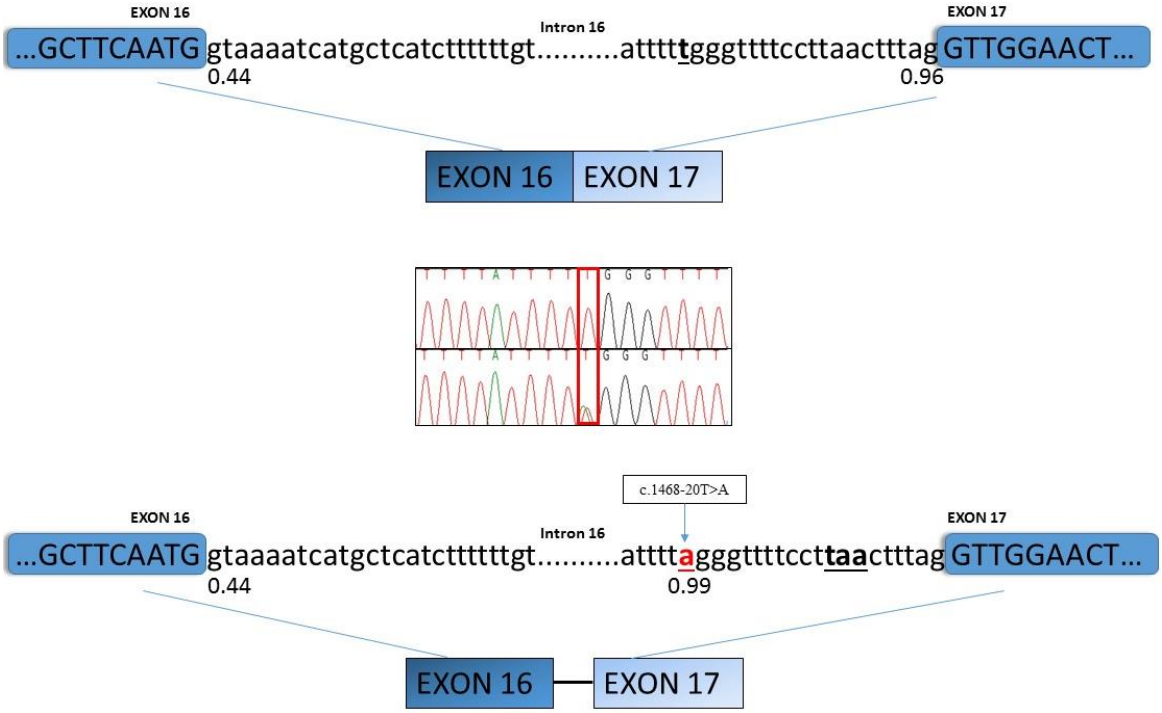


Figure 3.2 The predicted effects of the *TCF12* splicing mutation (c.1468-20T>A) identified in patient OX2636. The intronic region between exons 16 and 17 with normal donor and acceptor sites (score 0.44 and 0.96 respectively) is shown (Top) along with wild-type and mutant dideoxy sequencing traces (Middle). The mutation (shown in red) is found 20 nucleotides before the normal acceptor splice site (Bottom). It is predicted to create a stronger (score 0.99) cryptic splice acceptor site. Predicted scores are from the Splice Site Prediction Neural Network program (www.fruitfly.org/seq_tools/splice.html). Analysis of the complementary DNA (cDNA) of this patient by dideoxy sequencing confirmed use of the cryptic splice site and the presence of low levels of the neo-exon which contained a premature termination codon (taa, underlined above) – performed by A Fenwick (see Sharma et al. 2013).

Where available, DNA from other family members was tested for the presence of a *TCF12* mutation. In each family subsequently analysed, carriers of *TCF12* mutations had no clinically demonstrable craniofacial abnormalities indicating the presence of non-penetrance. The issue of non-

penetrance in carriers of *TCF12* mutations and its further investigation is described in Chapter 5. This finding seemed initially to negate the candidacy of *TCF12* as a disease gene as the mutation did not segregate with any clinically discernible craniofacial phenotype in carriers (see Figure 3.1).

However, due to the discovery of 3 patients with mutations in *TCF12*, other members of the Clinical Genetics Group working on parallel next-generation sequencing projects on undiagnosed craniosynostosis patients focussed their search for mutations in this gene. As a result, a *de novo* mutation in *TCF12* (c.1283T>G; L428*) was discovered in a patient-parent trio in which the child (GOS4437) had bilateral coronal synostosis, but both parents were clinically unaffected and mutation negative (see Figure 3.3). Correct sample relationships were confirmed by a panel of 13 microsatellites (see Table 4, Appendix). This strongly reinforced the case to support *TCF12* as a possible disease gene.

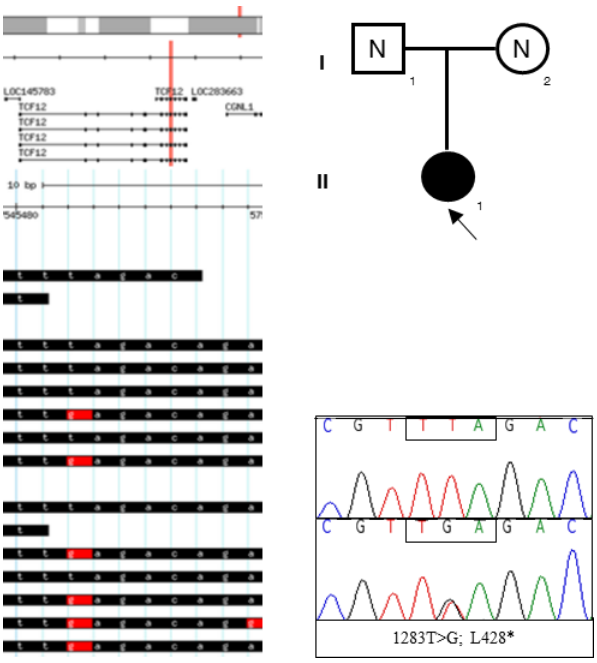


Figure 3.3 Identification of the first *TCF12 de novo* mutation by whole-genome and di-deoxy sequencing. This discovery was in patient GOS4437 [made by A Fenwick, see Sharma et al., (2013)]. This strongly supported the candidacy of *TCF12* as a novel disease gene.

To ensure that any subsequently identified variants were indeed novel mutations and not low frequency benign variants, each was compared with over 6500 samples on the Exome Variant Server (http://evs.gs.washington.edu/EVS/). This publically available database contains exome data from patients with diseases of the heart, lungs and blood, but no reported craniofacial malformations and was deemed to be a large and suitable control population. Specifically, there were no loss-of-function mutations of *TCF12* (splice, nonsense or frameshift) reported in this database, adding further evidence that mutations identified in the

patients described above were more likely to be the cause of their coronal synostosis.

To assess the impact of this novel gene discovery, a mixed craniosynostosis disease panel containing patients with different patterns of premature suture fusion was screened to identify further cases of *TCF12* mutations. This is described in the next chapter (see 4.1 and 4.2) and allowed a mutation spectrum to be established in addition to assessment of the prevalence of further mutations and the delineation of a new clinical entity.

3.5 Discussion

The results presented in this chapter demonstrate the power of a combined clinical and molecular strategy that has exploited whole-exome sequencing to identify a novel cause of coronal synostosis in a carefully stratified patient cohort.

The overlap strategy applied here had previously been used by other groups to identify the molecular genetic basis of syndromes with distinct craniofacial phenotypes (Hoischen et al., 2010; Ng et al., 2010). Furthermore, once the 3 patients who were *TCF12*-mutation positive were removed from the cohort of 7, no further candidate genes remained in the other 4 patients when the overlap strategy was used. The full lists of shared gene variants illustrating this iterative process that narrowed down the search are given in Table 1, Appendix.

Some shortcomings of the bioinformatic tools used to process and analyse the exome data became apparent which necessitated further refinement of the existing pipeline. Initially, the software package Bowtie was used as a variant caller for short read alignment of the exome data (Langmead et al., 2009). However, a drawback of the original version of this program that was initially used was its inability to report indel mutations. Due to the algorithm used in this program, it was unable to align small nucleotide insertions and deletions ('indels'). This meant any read containing this type of variant would be reported as unmapped and lack any information regarding alignment (Ruffalo et al., 2011). Therefore, frameshift mutations such as those identified in *TCF12* were not initially called until the exome

data was re-processed with a new variant caller (Novoalign) that could detect indels in short-read data (Krawitz et al., 2010). This was further combined with software (ANNOVAR) that functionally annotates genetic variants (Wang et al., 2010).

Another drawback of the automated pipeline used for gene identification was that it initially missed the novel splice site mutation found in patient OX2636. It was later discovered by manually searching lists of single gene variants and highlights another way in which exome sequencing may overlook potential disease causing mutations. The practice of focussing primarily on target regions is commonplace when using various capture kits for exome sequencing studies, but will inevitably miss a significant proportion of data on intronic or intergenic variants that may be of functional importance (Guo et al., 2012).

TCF12 encodes HEB/HTF4 (HeLa E-box binding protein/Helix-Loop-Helix Transcription Factor 4) and is located on chromosome 15q21 (Zhang et al., 1995). It is a member of the basic helix-loop-helix (bHLH) E-protein family that recognises a consensus binding site on DNA (i.e. the E-box, 5'-CANNTG-3') (Massari and Murre, 2000). The encoded protein activates transcription by binding this E-box. HEB exists in 2 forms, a longer canonical (HEBCan) and shorter alternative (HEBAlt) form due to alternative transcription start sites. These are known to affect cell lineage in neuronal and immunological tissue (Wang et al., 2006; Kee, 2009; Braunstein & Anderson, 2012).

TCF12 is a class I bHLH transcription factor that along with its paralogous genes, forms dimerisation partner proteins for class II bHLH transcription factors, most notably TWIST1 (Connerney et al., 2006). As mutations of *TWIST1* are known to cause Saethre-Chotzen syndrome (Howard et al., 1997; el Ghouzzi et al., 1997) that presents primarily with coronal synostosis, the discovery of a new molecule that fitted well into established craniosynostosis disease pathways was highly significant.

The mutations in patients OX484, OX4581 and OX2636 were all validated by dideoxy sequencing, confirming the changes identified by whole-exome sequencing were genuine. Of note, testing for *TCF12* mutations in other family members revealed non-penetrance of coronal synostosis as none of the carriers had any obvious craniofacial abnormalities that were clinically detectable. While this alone may have gone against the initial candidacy of *TCF12* as a novel disease gene, the discovery of a *de novo* mutation in patient GOS4437 argued strongly in favour of its pathogenicity. This is because *de novo* mutations are relatively rare events that lie at the furthest end of the spectrum of genetic variation. They tend to be more deleterious as they have been under less severe selection pressure and hence are more likely to be causative for sporadically occurring diseases such as craniosynostosis (Veltman & Brunner, 2012).

Although *TCF12* had never previously been linked to coronal synostosis, it had been shown to have a key role in initiating neuronal differentiation as *Tcf12*-null mice demonstrate a consistent exencephaly phenotype of low penetrance, stunted growth and die within 2 weeks of birth although the exact cause of death is not specified. Immunologically, these null mice also

demonstrated a slight reduction in the numbers of B cells. By contrast, heterozygous animals survive into adulthood and *Tcf12* is thought to have a crucial role in the maintenance of undifferentiated neural stem cells (Zhuang et al. 1996; Uittenbogaard and Chiaramello 2002). The gene also has a role in regulating B- and T-lymphocytes development, with a significant reduction in the number of pro-B cells noted in homozygotes (Wojciechowski et al., 2007). HEBAlt has specifically been shown to be required to efficiently drive a pool of early T-cell precursor cells that become specified into T-cell lineage (Wang et al., 2006). HEBCan negatively regulates natural killer (NK) cell fate in the thymus (Braunstein & Anderson, 2012). During myogenesis, HEB forms a functional heterodimer that complexes with MyoD, a master regulator in myoblasts, to control specification and differentiation in skeletal muscle (Hu et al., 1992; Yang et al., 2009).

Furthermore, there had been published clinical case reports of interstitial deletions at the 15q locus in patients with coronal synostosis, although the exact causative gene had not been delineated (Fukushima et al., 1990; Shur et al. 2003). In summary, *TCF12* was considered to be an excellent candidate that fitted well within known pathophysiological mechanisms and its further delineation as a new craniosynostosis syndrome is described in Chapter 4.

CHAPTER 4

TRANSCRIPTION FACTOR 12 (*TCF12*): SCREENING A MIXED DISEASE PANEL, DEFINING A NEW CLINICAL ENTITY AND EXTENDING THE CLINICAL AND MUTATIONAL SPECTRUM

4.1 Introduction

To further assess the significance of the discovery of a novel disease gene, it is necessary to screen an appropriate and representative population of patients with that disease. To that end, all coding exons of *TCF12* were screened using dideoxy sequencing in a mixed craniosynostosis disease panel.

4.2 Patients, Materials and Methods

Dideoxy sequencing was utilised to screen exons 2-20 of *TCF12* for damaging mutations likely to impact on protein function. Initially, 337 unrelated craniosynostosis patients were screened by this method, the principle of which has been previously described in section 2.8.2. The primers used are summarised in Table 2, Appendix. Whenever a new mutation was identified, it was re-confirmed by repeating the sequencing or performing a suitable restriction digest (see Table 3, Appendix) on the proband and DNA samples of family members where available. Newly identified *TCF12* mutations were cross-referenced with the Exome Variant Server (http://evs.gs.washington.edu/EVS/) to confirm each was restricted to craniosynostosis patients only.

4.3 Screening of a mixed craniosynostosis panel identifies further cases of *TCF12*-mutation positive patients

Dideoxy sequencing of a total of 337 craniosynostosis patients identified further mutations in 35 unrelated cases. This included 287 patients mostly from the UK and 50 patients from a Dutch cohort. Where available, the DNA of further family members was obtained and tested. Of the 48 families described in Table 4.1, the results for families #1- #38 were as a direct result of the work described in this thesis.

There are 10 further families (#39 - #48) that were referred from other centres with an established *TCF12*-mutation positive diagnosis following the initial publication of results (Sharma et al., 2013). They are included in Table 4.1 (italicised) and Figure 4.1 to indicate the spectrum of identified mutations and to further aid the delineation of this clinical syndrome. They are omitted from the original population and subsequent analysis in Table 4.2 to prevent bias. In the original screening population (n=347) mutations of *TCF12* were found in patients with coronal synostosis in 22 of 69 bilateral cases (32%) and 14 of 141 (10%) unilateral cases. There were 2 further cases where sagittal suture fusion was identified. In 93 patients with isolated single suture synostosis, no *TCF12* mutations was found. These findings are summarised in Table 4.1 and Figure 4.1

The combined results of 48 *TCF12* mutations comprised: 19 frameshift, 18 nonsense, 8 splicing and 3 missense mutations, with the majority located at the 3' end of the gene between exons 10-19 except for 2 mutations at exon 5 and 7. The missense mutations altered highly conserved residues within

the functionally important bHLH domain, crucial for protein dimerisation. Similar mutations in *TWIST1* and *TCF4* have been shown to be damaging in Saethre-Chotzen (Johnson et al., 1998) and Pitt-Hopkins syndromes (Amiel et al., 2007) respectively. Functional work by colleagues (A Fenwick, S Twigg) confirmed the damaging effect of these mutations (Sharma et al., 2013). Of the 38 families originally identified by the work described in Table 4.1, 13 cases arose *de novo* and correct biological relationships between patient-parent trios were confirmed using a panel of microsatellites (as described in 2.10, results in Table 4, Appendix). This provided further evidence of pathogenicity that mutations of *TCF12* were associated with coronal synostosis.

4.4 Pedigrees of mutation positive families

Cascade testing of the original 38 families identified 34 further individuals that were mutation-positive which are summarised in Figure 4.2. Of these individuals, 16 had craniosynostosis or relevant craniofacial abnormalities. This implied a significant degree of non-penetrance in 23 individuals (53%) whereby there is a failure of an individual with a particular genotype to manifest a given phenotype (Knight, 2009). Due to a lack of discernible clinical features in carriers of *TCF12* mutations, the genetic background of these individuals and others with craniosynostosis was investigated in an attempt to explain this variability and is described in Chapter 5.

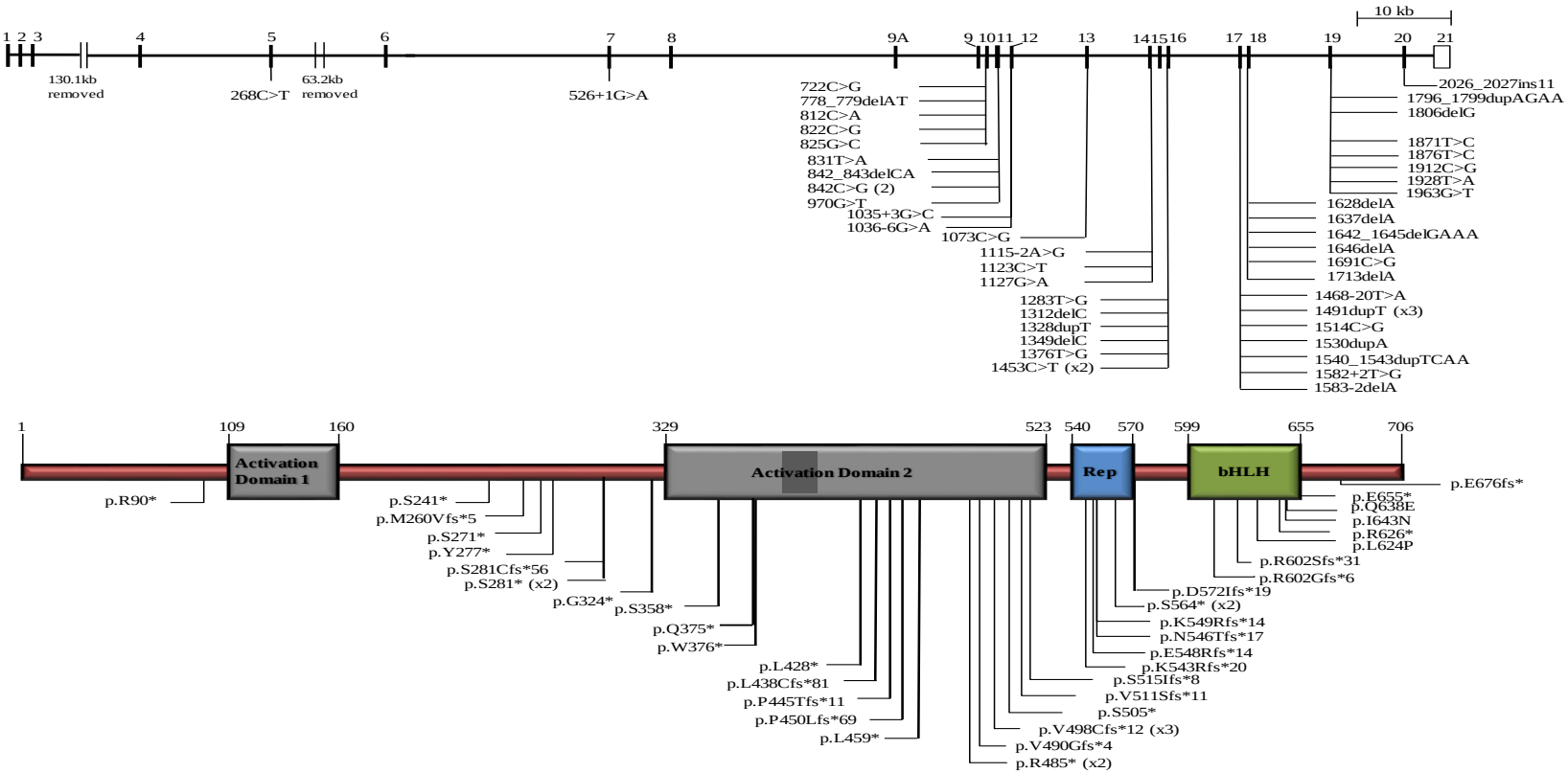


Figure 4.1 Summary of the structure of *TCF12* in mutation-positive patients with coronal synostosis. The major transcript of the gene (top) has 21 exons with alternate splicing of exon 15 creating a 682 or 706 amino acid protein (HEBα and HEBβ – domain in grey) respectively. All mutations identified in 48 unrelated individuals with coronal synostosis are shown. Transcription at the alternate 9A exon creates a 512 amino acid protein. The encoded protein (bottom) contains 2 activation domains, and transcriptional repression domain (Rep) that modulates transcriptional activity of the protein [Markus et al. (2002)]. Splicing mutations are not shown and recurrent mutations are indicated as (x2)/ (x3).

Family #	Proband ID	Gender	Suture pattern	FHx?	Suspected SCS?	Clinically non-synd.	Exon	Mutation	Protein	*De novo*	Clinically affected	Clinically unaffected (carrier)
1	953	F	RC			Y	7	526+1G>A	(sp)		1	2
2	07D6887	F	RC+S	Y			10	722C>G	S241*		2	2
3	06D7856	F	BC	Y	Y		10	778_779delAT	M260Vfs*5		3	
4	3496	F	RC	Y			10	812C>A	S271*		1	
5	2065	F	RC			Y	10	822C>G	(sp)	Y	1	
6	98D4142	F	BC		Y		10	825G>C	L275F (sp)	Y	1	
7	838	F	BC			Y	11	842_843delCA	S281Cfs*56	Y	1	
8	06D1742	M	RC	Y	Y		11	842C>G	S281*		1	
9	01D3258	M	BC		Y		11	842C>G	S281*		1	
10	01D3723	F	BC	Y			11	970G>T	G324*		1	
11	4438	F	BC	Y			12	1035+3G>C	(sp)		2	2
12	06D4640	F	LC			Y	13	1073C>G	S358*		1	
13	07D2829	M	BC	Y	Y		14	1115-2A>G	(sp)		2	
14	09D7966	F	BC	Y			14	1123C>T	Q375*		2	
15	3955	F	RC			Y	14	1127G>A	W376*		1	1
16	4437	F	BC		Y		16	1283T>G	L428*	Y	1	
17	3956	M	BC			Y	16	1312delC	L438Cfs*81	Y	1	
18	09D4333	F	RC				16	1328dupT	P445Tfs*11		1	
19	484	M	BC			Y	16	1349delC	P450Lfs*69		1	1
20	4397	F	BC	Y			16	1376T>G	L459*		2	1
21	5189	F	BC	*	Y		16	1453C>T	R485*	Y	1	
22	2636	M	BC	Y			17	1468-20T>A	V490Gfs*4		1	1
23	5137	M	BC				17	1491dupT	V498Cfs*12		1	1
24	1669	M	RC			Y	17	1491dupT	V498Cfs*12		1	1
25	01D4012	F	BC		Y		17	1514C>G	S505*	Y	1	
26	4855	F	RC			Y	17	1530dupA	V511Sfs*11	Y	1	
27	4502	F	RC	Y	Y		17	1540_1543dupTCAA	S515Ifs*8		3	
28	3664	F	BC		Y		17	1582+2T>G	(sp)	Y[m]	1	1
29	4618	M	RC	*		Y	18	1628delA	K543Rfs*20	Y	1	
30	4581	F	BC			Y	18	1642_1645delGAAA	E548Rfs*14		1	1
31	00D0716	M	RC	Y			18	1646delA	K549Rfs*14		1	
32	4951	F	BC			Y	18	1691C>G	S564*	Y	1	
33	08D7786	M	RC			Y	18	1713delA	D572Ifs*19	Y	1	
34	4881	M	BC	Y			19	1796_1799dupAGAA	R602Gfs*6		3	

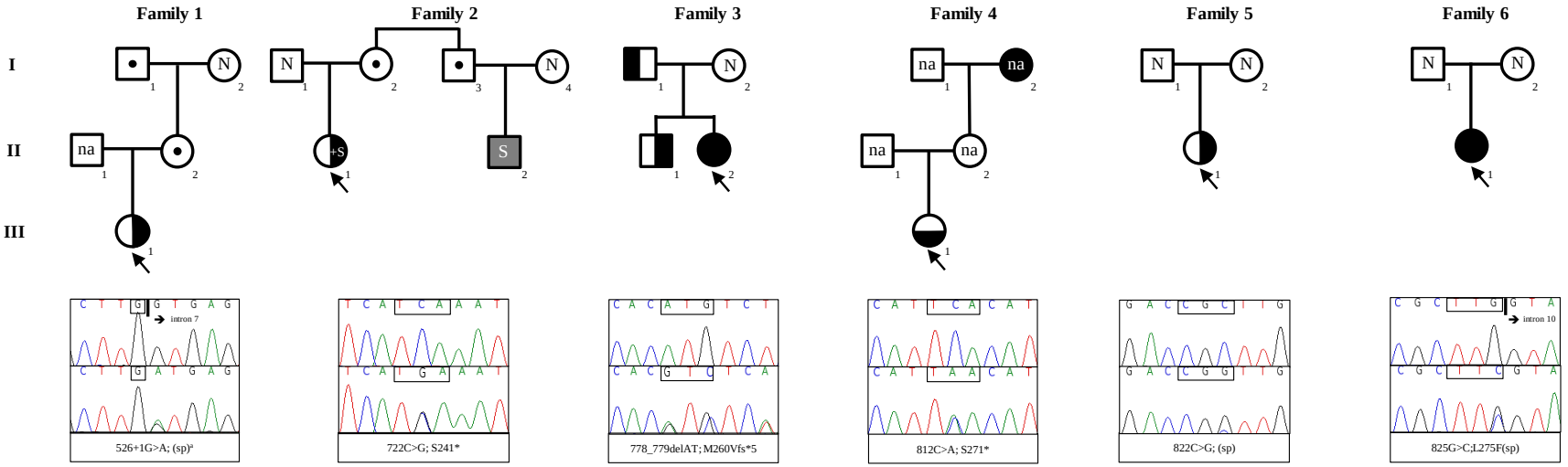
35	4594	M	RC	Y	Y		19	1806delG	R602Sfs*31	Y	2	1
36	1520	M	LC			Y	19	1871T>C	L624P		1	1
37	99D0117	F	BC	Y	Y		19	1912C>G	Q638E		1	
38	4379	M	BC+S	*		Y	19	1963G>T	E655*	Y	1	
39	*12D1921*	*F*	*RC*		*Y*	*N*	*5*	*268C>T*	*R90**		*1*	*1*
40	*6594*	*M*	*BC*		*N*	*Y*	*11*	*831T>A*	*Y277**		*1*	
41	*5463*	*F*	*BC*			*Y*	*12*	*c.1036-6G>A*	*(sp)*	*Y*	*1*	
42	*5547*	*M*	*BC*			*Y*	*16*	*1453C>T*	*R485**	*Y*	*1*	
43	*6580*	*M*	*BC*	*Y*		*Y*	*17*	*1491dupT*	*V498Cfs*12*		*1*	
44	*6559*	*F*	*RC*			*Y*	*17*	*1583-2delA*	*(sp)*		*1*	*1*
45	*6563*	*M*	*RC+S*			*Y*	*18*	*1637delA*	*N546Tfs*17*		*1*	*2*
46	*6672*	*F*	*LC*			*Y*	*19*	*1876C>T*	*R626**		*1*	
47	*6201*	*M*	*RC*			*Y*	*19*	*1928T>A*	*I643N*		*1*	*1*
48	*6221*	*F*	*RC*			*Y*	*20*	*2026_2027ins11*	*E676Gfs*49*		*1*	*1*

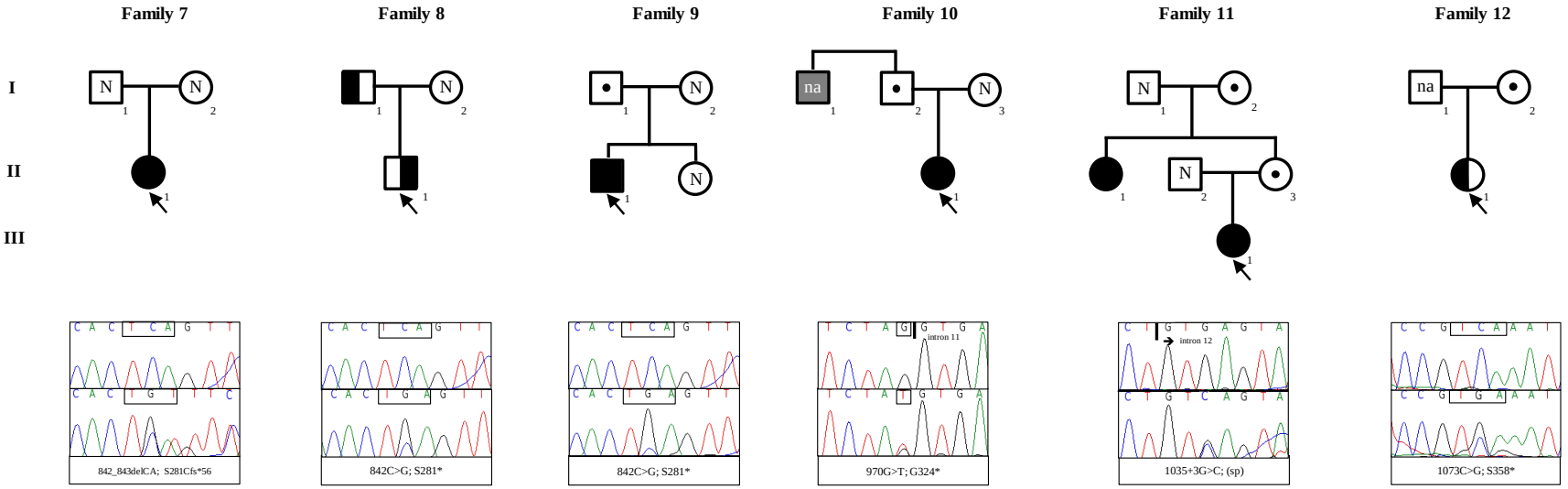
Table 4.1 Spectrum of *TCF12* mutations identified in 48 patients with coronal synostosis

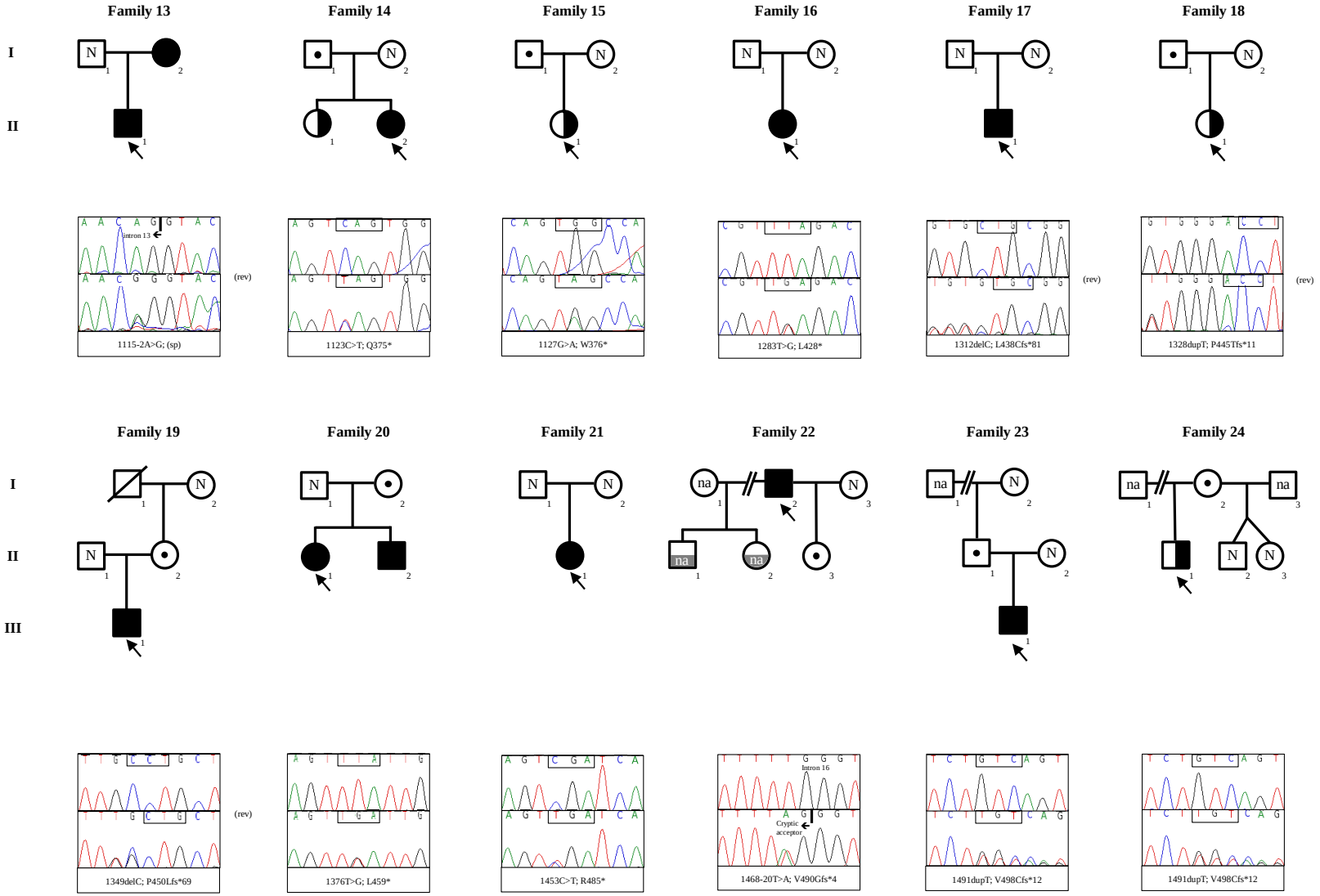
KEY: RC = Right coronal, LC=Left Coronal; S = Sagittal; + = feature is present, FHx = Family History, (sp) = mutation affecting splicing, * = Coincidental craniofacial signs in parent, but *TCF12* negative. m = *de novo* mutation due to mosaic mutation in maternal grandmother.

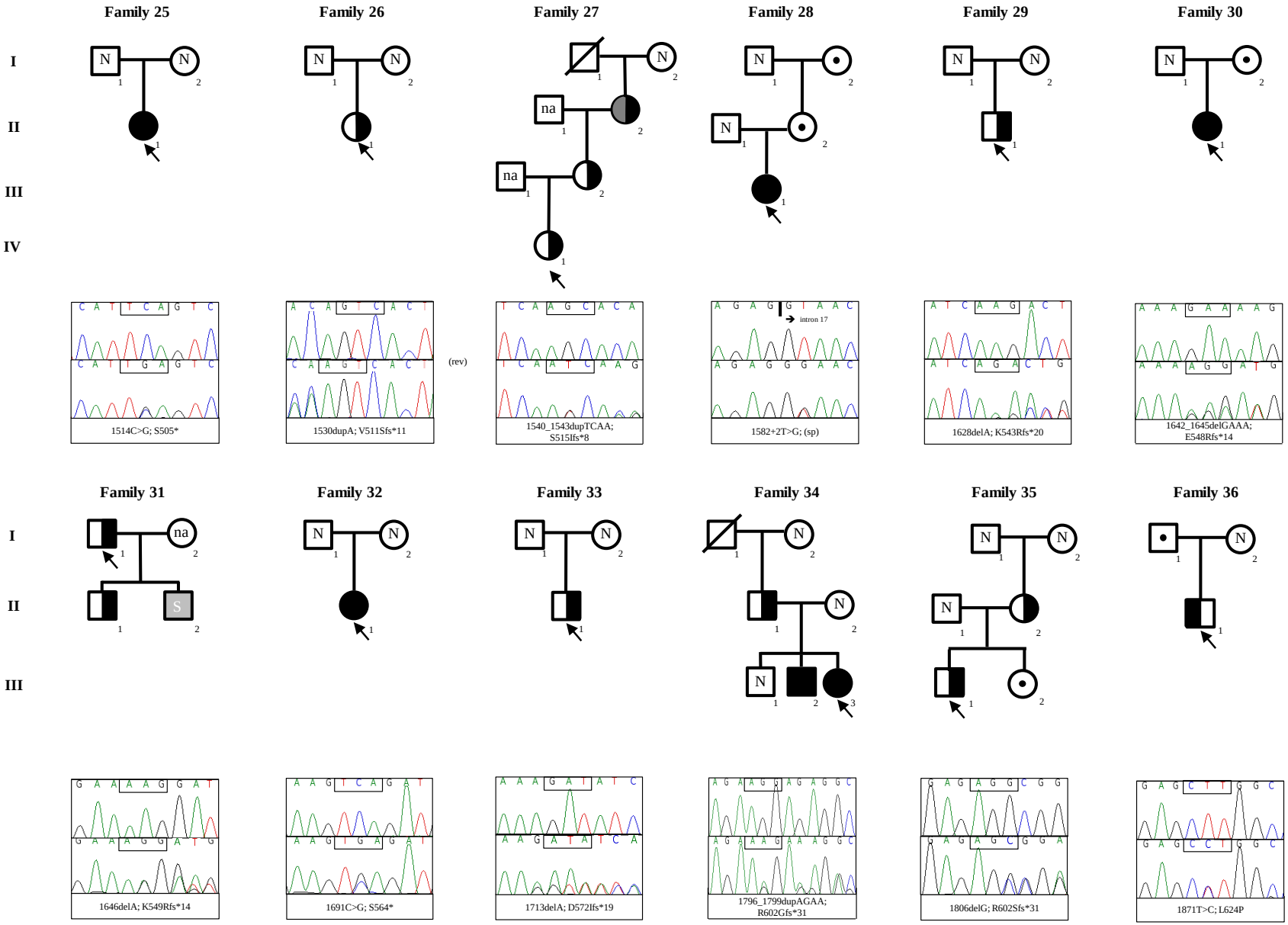
		Syndromic		Non-syndromic		Both
Suture(s) fused	Total	*TCF12*-mutation positive	Total	*TCF12*-mutation positive	Total	*TCF12*-mutation positive
Metopic	14	0	32	0	46	0
Sagittal	7	0	32	0	39	0
Bilateral coronal	41	17	28	5	69	22
Unilateral coronal	26	5	115	9	141	14
Uni/bilateral lambdoid	0	0	8	0	8	0
Multisuture	19	1	19	1	38	2
Craniosynostosis unspecified	4	0	2	0	6	0
Combined	111	23	236	15	347	38

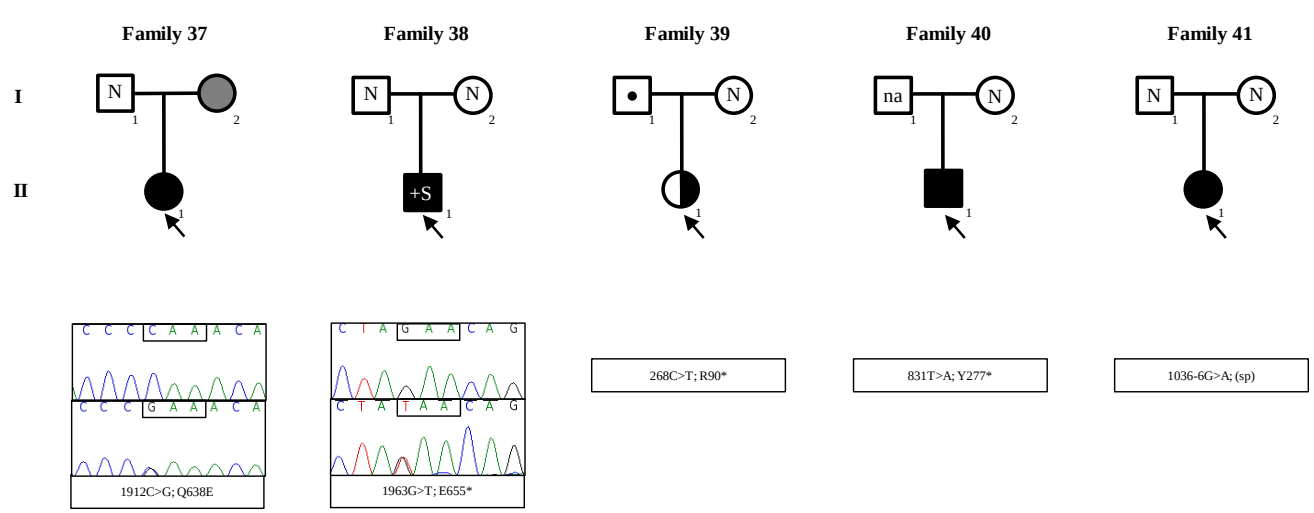
Table 4.2: *TCF12* mutations based on a per-suture basis across different craniosynostoses

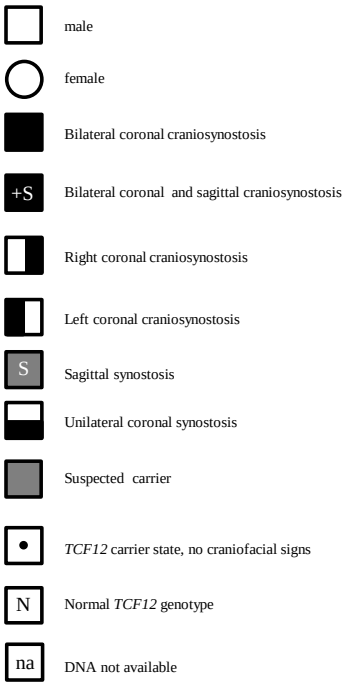

Family 1
Family 2
Family 3
Family 4
Family 5
Family 6
I
II
III
N
na
+S
S
intron 7
526+1G>A; (sp)a
722C>G; S241*
778_779delAT; M260Vfs*5
812C>A; S271*
822C>G; (sp)
intron 10
825G>C;L275F(sp)

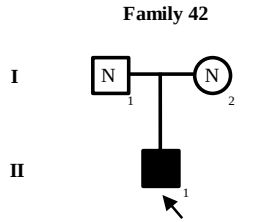

Family 7
Family 8
Family 9
Family 10
Family 11
Family 12
I
II
III
N
na
intron 11
intron 12
842_843delCA; S281Cfs*56
842C>G; S281*
842C>G; S281*
970G>T; G324*
1035+3G>C; (sp)
1073C>G; S358*


Family 13
Family 14
Family 15
Family 16
Family 17
Family 18
I
II
N
1
2
intron 13
(rev)
1115-2A>G; (sp)
1123C>T; Q375*
1127G>A; W376*
1283T>G; L428*
1312delC; L438Cfs*81
1328dupT; P445Tfs*11
Family 19
Family 20
Family 21
Family 22
Family 23
Family 24
III
na
3
1349delC; P450Lfs*69
1376T>G; L459*
1453C>T; R485*
Intron 16
Cryptic acceptor
1468-20T>A; V490Gfs*4
1491dupT; V498Cfs*12


Family 25
Family 26
Family 27
Family 28
Family 29
Family 30
I
II
III
IV
na
N
S
1514C>G; S505*
1530dupA; V511Sfs*11
(rev)
1540_1543dupTCAA; S515Ifs*8
intron 17
1582+2T>G; (sp)
1628delA; K543Rfs*20
1642_1645delGAAA; E548Rfs*14
Family 31
Family 32
Family 33
Family 34
Family 35
Family 36
1646delA; K549Rfs*14
1691C>G; S564*
1713delA; D572Ifs*19
1796_1799dupAGAA; R602Gfs*31
1806delG; R602Sfs*31
1871T>C; L624P







For each family the individuals analysed in the pedigree (see key to symbols) are shown together with a representative sequence chromatogram for a normal control (above) and an affected individual (below); (rev) indicates sequence shown is in reverse direction. Families 39-48 were ascertained from other diagnostic centres following the publication of the initial work (Sharma et al., 2013), hence sequence chromatograms are not given as a *TCF12* mutation was already identified by other laboratories.

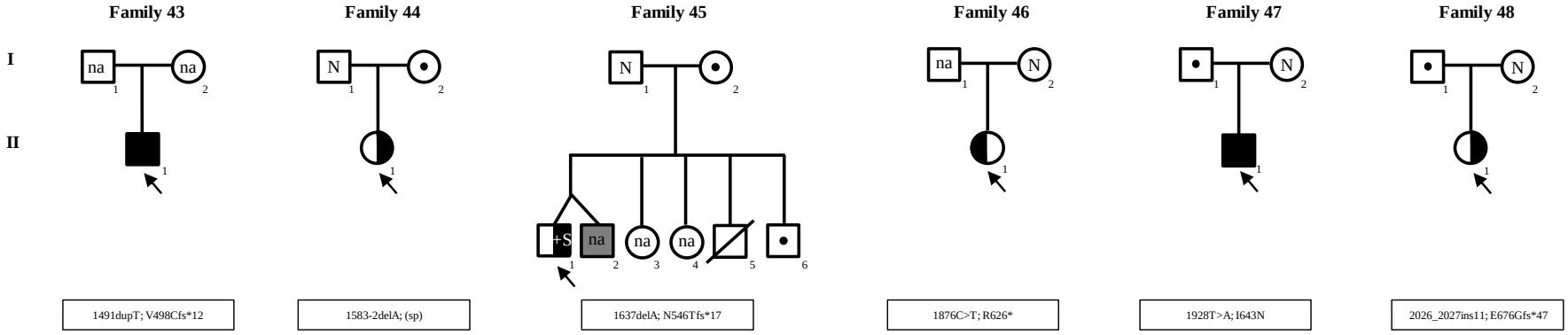


Figure 4.2 Pedigrees of *TCF12*-mutation positive patients and their family members (key to symbols shown in last panel). Corresponding sequence chromatograms below each pedigree indicate the mutation (lower box) in comparison to the normal trace of an unaffected, mutation-negative individual (upper box). The first codon of a non-synonymous exonic mutation is indicated within a small rectangular box. (rev) implies reverse strand sequence has been shown.

Clinical Characteristic	Number of individuals	Details
Coronal synostosis	62	Bilateral = 30; Right-sided = 21; Right-sided but no sample available to test = 1; Left-sided = 6; Coronal + Sagittal = 4; Sagittal alone =2; No craniosynostosis but mutation positive = 28
Craniofacial signs overlapping with Saethre-Chotzen Syndrome	21	Minor ear anomalies (prominent transverse crus/deepened concha causing ear prominence) = 10; Low anterior hairline = 9; Blepharoptosis = 3
Limb anomalies	21	Transverse palmar crease = 9; Hallux valgus = 7; 2 toe syndactyly = 3; Brachydactyly = 4
Brain abnormalities on radiological imaging	15 of 45	Prominent ventricles +/- CSF spaces = 7 (2 cases associated with Chiari malformations) ; Right cerebellar hypoplasia = 1 (LCS); Partial/complete agenesis of corpus callosum = 2; Miscellaneous findings: paraventricular cyst (2), pineal mass (1), thinning of optic chiasm (1), non-specific periventricular white matter lesion (1 - no focal neurological deficits reported with any of these findings)
Ocular anomalies	12	Strabismus = 11 (always in presence of unilateral coronal synostosis); Ocular torticollis = 1
Dental anomalies	7	Class I/II malocclusion or crowding of teeth in 7 cases; 1 case required formal orthognathic surgery (bimaxillary osteotomy, good result, no further interventions)
Developmental delay	13	5 adults with mild intellectual disability (1 case with severe autism); 2 adults with speech delay /childhood dyslexia; 2 children with mild global developmental delay; 1 child with significant speech and language delay pre-op (becoming a mild expressive delay post-op); 2 children with mild speech delay; 1 child with verbal/social skills delay; 1 child with Asperger's syndrome diagnosis; 1 with autism; 1 with complex neurodevelopmental delay; 1 with unspecified but significant behaviour problems. 1 individual had a concurrent diagnosis of mucopolysaccharidosis type III, associated with significant behavioural disturbances and was therefore excluded from this analysis.
Other abnormalities	1	1 case of anal atresia (not observed in any other mutation-positive individuals)

Table 4.3 Summary of clinical features of *TCF12*-related craniosynostosis. 90 mutation positive individuals were assessed (39 male, 51 female)

4.5 Summary of clinical features in *TCF12* mutation-positive individuals delineates a new clinical entity

Cascade testing and acquisition of mutation positive samples from other centres revealed common features shared between affected individuals. Full clinical details that were available for affected individuals and their family members are given in Table 5, Appendix for all 48 families. Coronal synostosis was seen in all index cases with a bilateral pattern of fusion being observed in 32% of bilateral cases and 10% had unilateral coronal fusion. Bilateral fusion was also seen most commonly, followed by right coronal, left coronal and combinations of coronal and sagittal synostosis. Of those cases affected with coronal synostosis, a non-syndromic presentation was commonest but 21 individuals had craniofacial signs consistent with Saethre-Chotzen syndrome, namely minor ear anomalies, low anterior hairline and blepharoptosis.

The minor limb anomalies that were observed were not diagnostic and were not associated with any functional impairment in hand function. Descent of part of the cerebellum below the foramen magnum (Chiari malformation) was seen on brain imaging in 2 cases, but not associated with any neurological deficit. Other findings on CT or MRI were also asymptomatic. Ocular problems consisted of strabismus that was always present with unilateral coronal suture fusion, and in one case, ocular torticollis. In terms of dental issues, Class I and II malocclusion was seen most commonly and usually corrected with simple orthodontic

appliances as required. In the former, the bite is normal, but there is a slight overlap of the upper teeth over the lower; in the latter, the upper teeth and jaw significantly overlap the bottom teeth and jaw. One case required formal orthognathic surgery under general anaesthetic to correct maxillary prominence, but had an uneventful recovery and good post-operative outcome. Varying degrees of developmental delay or learning difficulties were seen in 20% of cases. This subset of patients will require further investigation by specialists in clinical psychology and speech and language therapy to further elucidate patterns of developmental delay. In turn, this would better stratify these patients as they may require further support during their educational development.

4.6 Discussion

The screening of a mixed craniosynostosis disease cohort for mutations of *TCF12* revealed numerous further cases, accounting for a significant proportion of new diagnoses in cases of coronal synostosis negative for mutations in known disease genes. Inclusion of additional cases from other referring centres has been crucial in expanding the spectrum of this disease from both clinical and molecular perspectives. Analysis of the pattern of premature suture fusion, revealed mutations of *TCF12* are strongly associated with bilateral or unilateral coronal synostosis, accounting for 32% and 10% of cases respectively (see Table 4.2).

While the commonest syndromic presentation was apparent Saethre-Chotzen syndrome, the majority of DNA samples analysed were from

non-syndromic patients, with dysmorphic features that were consistent only with the associated calvarial deformity. Furthermore, these patients were negative for mutations in *TWIST1*. Limb anomalies observed in *TCF12*-related craniosynostosis were very mild with no functional impairment on hand function, balance or ambulation. The brain abnormalities that were seen on radiological imaging did not cause any focal neurological deficit or impairment and tended to be incidental findings. Strabismus was only seen in the presence of unilateral coronal suture fusion. An abnormal bite or malocclusion was observed in 10% of cases and was usually corrected with minor orthodontic intervention. Only one case required formal corrective orthognathic surgery.

Although the majority of individuals with proven *TCF12* mutations had normal attainment of developmental targets, an important proportion had significant intellectual disability that included speech delay, Asperger syndrome and autistic spectrum disorder. This smaller subset of patients will require further investigation and re-evaluation of psychometric testing to delineate patterns of subtle differences in development that may be quantifiable and possibly related to genotype. Finally, one case of anal atresia was seen in a mutation-positive patient with right coronal synostosis. As this was not seen in any of the other 89 individuals with a *TCF12* mutation, this is likely to represent an incidental finding.

From a molecular standpoint, the mutations identified were loss-of-function with the majority being frameshift or nonsense mutations, implying a haploinsufficiency disease mechanism. This is where in a diploid organism, possession of only one normal copy of a gene (due to a

mutation at a specific locus) is insufficient to maintain a normal phenotype (Knight, 2009). A high number of *de novo* mutations that were identified added to the burden of proof supporting the pathogenicity of *TCF12* mutations, as they arose in individuals with coronal synostosis. However, no genotype-phenotype correlation was observed with the type of mutation and severity of disease. A significant proportion of mutation-positive individuals were non-penetrant for coronal synostosis with a normal clinical appearance. This is a potential pitfall in genetic counselling as the parents of affected children who appear phenotypically normal, may be given incorrect estimates of recurrence risks for future offspring. To investigate if the genetic background of carriers of *TCF12* mutations was relevant to the phenotypic variability that was observed, an analysis of haplotypes was performed and is described in the next two chapters.

CHAPTER 5

NON-PENETRANCE IN *TCF12*-MUTATION POSITIVE INDIVIDUALS AND INVESTIGATION INTO PARENTAL ORIGINS OF *DE NOVO* MUTATIONS

5.1 Introduction

During the process of identifying further mutation-positive individuals in the family of an affected proband, it became apparent that a significant proportion did not have any discernible craniofacial features consistent with coronal synostosis. This non-penetrance, whereby an individual who possesses a particular mutation or genotype fails to exhibit clinical signs of the associated disorder within a defined time period (Cooper et al, 2013) confounds clinical assessment. The genotype is an individual's hereditary or genetic constitution at either a specific locus or their whole genome. Their phenotype is a product of their genotype and is an observable characteristic (Knight, 2009). It can lead to difficulties in quoting accurate penetrance rates in dominant conditions by genetic counsellors (Strachan & Read, 2010). Due to the lack of a clinically detectable phenotype in carriers of *TCF12* mutations, the genetic background of these individuals was investigated by analysis of haplotypes. The overall aim was to assess the effect of a combination of alleles in a specified region of *TCF12* that are passed down a pedigree and affect phenotypic variability.

5.2 Patients, materials and methods

In the first instance, the parental origins of cases of *de novo* mutations were analysed to identify which parental allele the mutation arose on and to refine the assignment of haplotypes during the wider analysis.

Secondly, haplotypes of *TCF12*-mutation carriers that were non-penetrant for coronal synostosis were analysed within the context of the whole family. The combined dataset was then used to identify any specific haplotypes that may account for the phenotypic variability observed in individuals with *TCF12* mutations. Thirdly, *TCF12* haplotypes were analysed in cases of non-syndromic coronal synostosis to look for association. These latter 2 stages are described in Chapter 6.

5.3 Identification of significant non-penetrance in mutation-positive relatives with normal cephalometric measurements

Of the 48 families described in Chapter 4, it was not possible to perform cascade testing on 2 families. Removing 14 cases where the *TCF12* mutation arose *de novo*, left 32 families whose members had coronal synostosis. This allowed subsequent testing of 76 individuals, the majority of whom were phenotypically normal. Within the 32 families, 40 additional mutation-positive individuals were found. 16 had craniosynostosis or craniofacial features consistent with premature suture fusion, whereas 24 appeared clinically normal. This demonstrates significant non-penetrance (60%). Where available, appropriate cephalometric measurements of head circumference, interpupillary distance, inner and outer canthal distances were recorded and compared with normative data for individuals within the UK. Comparison was made with average values and a specified range of the 5^{th}-95^{th} centile for this normal population. Although the sample size was small (n=12), the

carriers analysed were found to have cephalometric measurements within the normal range of adults in the UK. Typical examples of *TCF12*-mutation positive non-penetrant carriers are shown in Figure 5.1 and their cephalometric data given in Table 6, Appendix.

Figure 5.1 Craniofacial appearance of non-penetrant carriers of heterozygous *TCF12*-mutations. Cephalometric measurements (head circumference, interpupillary, inner and outer canthal distances) all fell within normal limits.

5.4 Hypothesis that non-penetrance may be related to haplotype and relative levels of gene expression

In the absence of any discernible craniofacial features that would indicate a carrier state, the genetic background of these individuals was analysed. It is hypothesised that expression levels of different alleles may account for the phenotypic variability that is seen and an investigation into the

genetic background or haplotype of individuals represents an indirect method of acquiring such information.

In congenital intestinal aganglionosis (Hirschsprung disease), mutations of the *RET* proto-oncogene account for more than 80% of known mutations. Those that affect splicing or are coding show a 50 -70% penetrance, with mutations in *RET* accounting for 20% of sporadic and 50% of familial cases, but overall accounting only 0.1% of heritability (Emison et al., 2010). This important body of work will now be considered in more detail.

Emison et al., (2005) performed transmission disequilibrium tests (TDT) to look for linkage of SNPs transmitted from non-carrier parents to offspring in 126 Hirschsprung disease proband-parent trios, finding the most significant association within intron 1 of *RET*. A TDT aims to measure the over-transmission of a SNP allele from a heterozygous parent to their child (Spielman et al., 1993). An analysis of multi-conserved sequences in 9 mammalian species across a defined 27.6kb genomic region with highest association, identified RET+3 (or rs2435357) as the SNP with the highest transmission distortion and statistical significance. Two further SNPs were found either side of RET+3 and found to have complete non-random association with each other and RET+3 (i.e. perfect linkage disequilibrium). The ancestral allele (i.e. an approximation of the allele present in our most recent common ancestor, e.g. chimpanzee) was the allele associated with these

additional SNPs and was confirmed by analysis of conservation in 9 mammalian species – RET+3:C. The derived allele (i.e. that which has arisen through evolution due to a mutation) was polymorphic and over transmitted and hypothesised to be the locus of disease variation – RET+3:T.

By genotyping 51 individuals of various ethnicities, this group found the frequencies of the mutant RET+3:T varied significantly from 0.01 in Africa, 0.25 in Europe to 0.45 in Asia correlating with the high incidence of the disease in Asian newborn children. The presence of this non-coding mutation in *RET* provides an additional disease loci that impacts nearly all cases of Hirschsprung disease, explaining ~ 2% of cases. This underscores the importance of looking for mutations outside of coding exons that account for 1-3% of the human genome.

By performing additional molecular analysis of SNPs across the entire locus of this gene in addition to sequencing the protein coding sequence in 882 probands and 1478 first-degree relatives from 3 ethnic backgrounds, Emison et al. (2010) then elegantly demonstrated that disease prevalence, recurrence and severity was significantly influenced by both rare and common variants within the *RET* locus.

They demonstrated that in Hirschsprung disease and control populations from USA, Europe and Hong Kong, the mutant derived T-allele was also shown to be commoner in disease subjects in comparison to the ancestral

C-allele when *rs2435357* [C/T] was analysed (Emison et al., 2010). For cases versus controls, the T allele was commoner among Hirschsprung's patients in American (49% to 26%), European (54%-62% to 20-30%) and Chinese (88% to 47%) populations. Furthermore, the relative risk of being affected varied on a per-genotype basis with TT homozygotes having the greatest risk in all subgroups measured (i.e. gender, ethnicity, length of disease segment of bowel). By comparison, in CT heterozygotes and CC homozygotes, intermediate and low relative risks were observed respectively (Emison et al., 2010). A *trans*-activation assay in HeLa cells provided functional *in vitro* evidence of the effects of the mutant T allele. A reduction in SOX10 binding by the intronic *RET* enhancer MCS+9.7 was shown in cases of T-bearing versus C-bearing alleles. SOX10 is known to directly activate *RET* transcription. The T allele is therefore a susceptibility allele present in a non-coding enhancer element in the first exon of *RET*.

Furthermore, investigation for genetic association between 14 *RET* SNPs and Hirschsprung disease by a transmission disequilibrium test demonstrated highly significant results, especially for *rs2435357* and a neighbouring SNP *rs2435362*. In the presence of an association, this test demonstrates linkage between a disease locus and a marker transmitted from a non-carrier heterozygous parent (Spielman et al. 1993). The former SNP was shown *in vitro* to disrupt binding of SOX10 and RET transactivation and possession of *rs2435362T* is a susceptibility factor for disease development. By constructing SNP haplotypes, Emison et al.

(2010) demonstrated differences in those that were transmitted and not transmitted on chromosomes through a pedigree. The mutant allele was shown to exist on 2 separate disease haplotypes – the ancestral (i.e. the haplotype of the most recent common ancestor) that was present at higher frequencies, and the derived haplotype. It is thought to have arisen as a single mutation on the ancestral haplotype that later recombined to form the derived haplotype.

Due to the recombination event that gave rise to these two separate haplotypes, it must also be borne in mind that any SNP allele in non-random association or linkage disequilibrium (LD, see Section 6.1 for full description) with a possible causal variant would also switch between populations. This additional factor must therefore also be considered when analysing allelic associations. This is demonstrated in Figure 5.2 below. Higher T-allele frequencies were also seen in specific ethnic populations, thereby explaining the difference in numbers of cases with the disease in European and Chinese populations. Possession of the *rs2435362T* SNP was shown to be a susceptibility factor in all sub-types of Hirschsprung disease with the penetrance of the T-allele higher in those with common subtypes with lower recurrence risks observed in relatives.

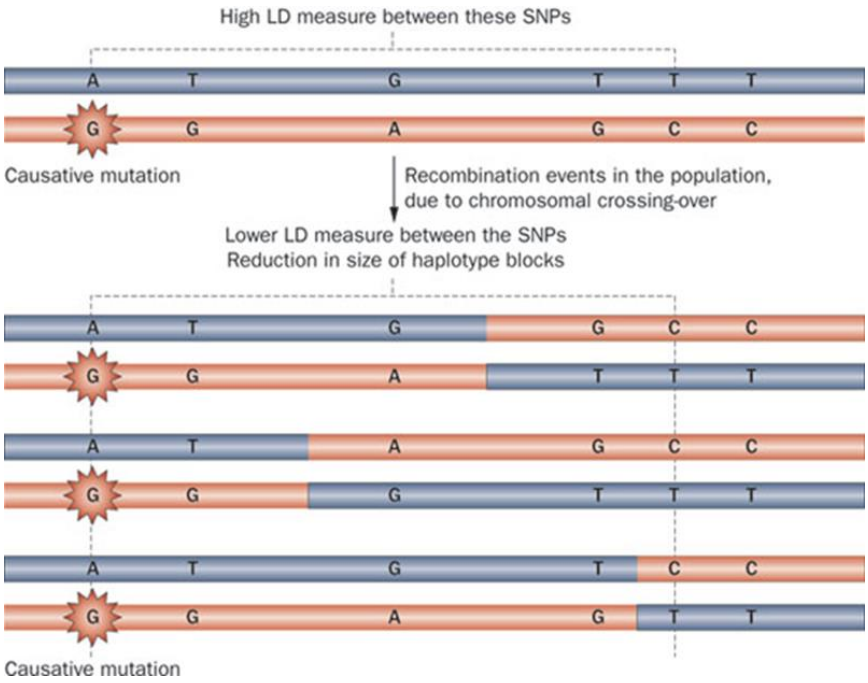


Figure 5.2 Summary of the effect of a recombination event on the association between SNPs in strong LD and a causal variant. The consequence of recombination is that the LD between SNPs breaks down over generations as the mutation causing variant spreads within a population. Not only will this create new haplotypes, but it will also shrink the size of the original haplotype block. The switch from the original ancestral population to the recombined derived population would complicate any subsequent study of allelic association (From Rosset et al., 2011).

Conservation of SNPs across the genome that are then inherited through successive generations are known as haplotype blocks (Knight, 2009). In association genetics, it may be possible to correlate a SNP with a disease or susceptibility to a disease. SNPs that occur in specific regulatory regions of genes may also have significant effects on coding exons. In the case of thrombocytopenia with absent radii (TAR), it was shown that compound inheritance of one of two low-frequency non-coding SNPs in the 5′ UTR of *RBM8A* (RNA Binding Motif Protein 8A) in addition to a null allele (containing a loss-of-function mutation) caused this disease

(Albers et al., 2012). The additional causative allele was identified by exome sequencing 5 individuals with the disease who also had a 1q21.1 deletion. Although no coding mutations were found, the minor allele of a low frequency SNP (rs139428292) SNP was detected in the 5′ UTR of *RBM8A* of 4 individuals and an unknown SNP in the first intron of the fifth individual. Further dideoxy sequencing of 48 cases of TAR of the same ethnicity and possessing the 1q21.1 deletion, demonstrated the presence of the 5′ UTR SNP in 35 samples, and the intronic SNP in 11 samples. This group also investigated if there were other variants in strong LD with either SNP. Although none were found, the minor allele of the 5′ UTR SNP was observed on 2 distinct haplotype backgrounds. This provided further evidence that this allele is causative in TAR as it was found on either genetic background and therefore in association with ancestral or derived alleles.

It is therefore possible that possession of a particular variant or SNP on a certain *TCF12* haplotype or alternatively a particular haplotype itself that is in LD with a specific variant may be a susceptibility or protective factor for the development of coronal synostosis. This in turn may account for the non-penetrance that has been observed and thus warrants further investigation. A summary of approaches to haplotyping is shown in Figure 5.3.

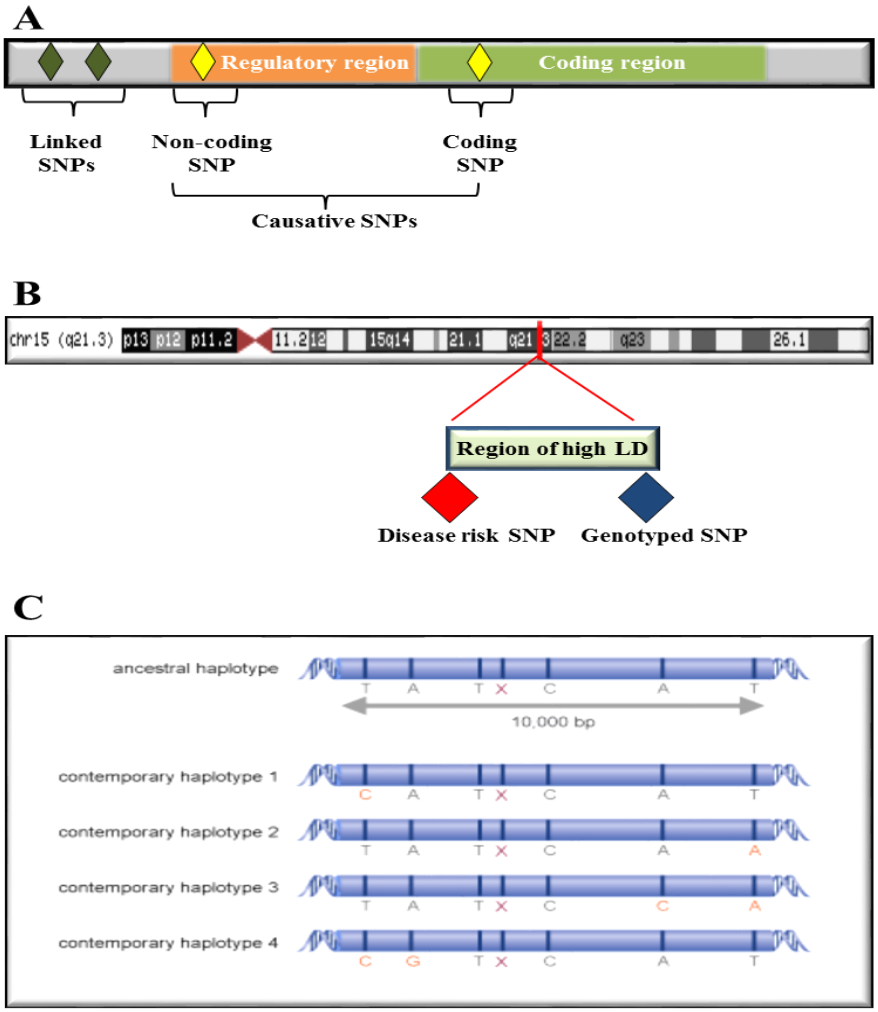


Figure 5.3 Approaches used to identify causal variants or haplotypes associated with disease. **A:** Linked or indicative SNPs located outside of the gene of interest do not affect protein production but can be linked with the risk of developing a certain disease. Causative SNPs will influence the function of the encoded protein. Those that are non-coding are found in the regulatory sequences of a gene and will alter gene expression in terms of level, location and timing. Coding SNPs will directly alter the amino acid sequence of the polypeptide (adapted from Genetic Science Learning Center, 2014) **B:** SNPs that are genotyped on a chromosome are commonly in LD with an influential SNP that is associated with disease. The genotyped SNP is a surrogate marker that is therefore statistically associated with disease via an indirect association but is not itself a causal variant (adapted from Bush & Moore, 2012) **C:** Use of haplotypes to identify a variant that alters the risk of developing a disease . A new variant associated with a disease ('X') will enter a population on an existing ancestral

haplotype, ‘TATCAT’. Through successive generations, recombination will erode this haplotype signature, but the disease allele and its closest SNPs (‘T’ and ‘C’) are generally inherited together as a unit. Elucidation of this haplotype in a disease cohort and typing of those alleles contained therein, allows a conserved region to be identified and a disease causing variant to be found (from Twyman, 2003).

5.5 Investigation into parental origins of *TCF12* mutations to elicit phase information reveals a number of paternal origin mutations and refines future haplotype assignment

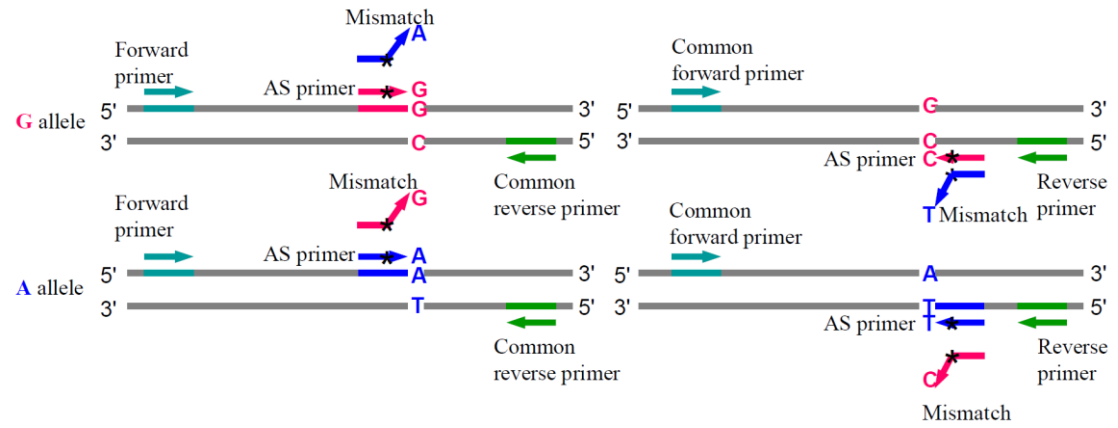
Of the 16 *de novo* mutations identified in Chapter 4, it was necessary to identify the parental origin of the mutation to allow correct phase information to be assigned. Without this, it would not be possible to delineate mutant from partner alleles. Analysis of the haplotype structure around *TCF12* by use of HapMap (see Figure 6.5) revealed limited allelic diversity surrounding this gene. Therefore, it would be reasonable to conduct a parent of origin study to look at association with penetrance or expressivity. By doing so, the data that would be subsequently generated could also be used in the wider haplotype analysis to increase sample size and the overall informativeness of the study.

It has previously been demonstrated that in sporadic cases of craniosynostosis such as Apert syndrome, the parent of origin for the mutation can be identified by use of natural sequence variants close to the mutation. Moloney et al. (1996) used a modified amplification-

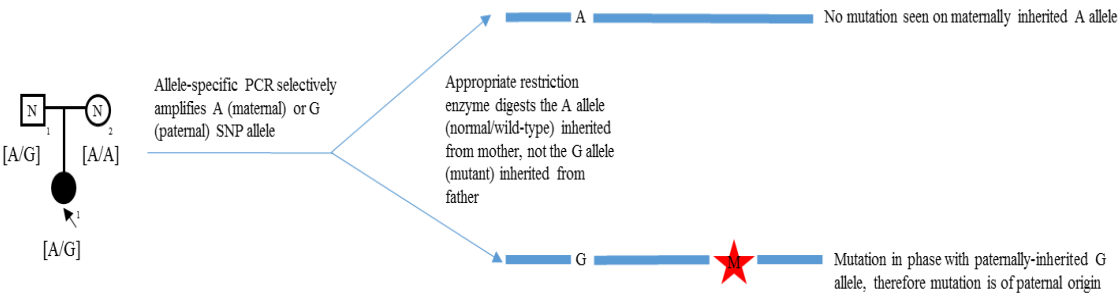
refractory mutation system (ARMS) to distinguish parental alleles in children affected by this syndrome in cases where the analysis of a specific polymorphism was informative (i.e. child is heterozygous for a SNP, as is one parent with the other parent homozygous or both parents oppositely homozygous). I therefore designed an analogous strategy to determine the parental origin of *de novo* mutations in *TCF12*.

All *de novo* cases of *TCF12* mutations identified (see Figure 4.1) arose between exons 10-19. Patient-parent trios were screened by dideoxy sequencing up to 1 kb both 5′ and 3′ of the mutation (See Table 7, Appendix for full list of primers and conditions). This would identify a combination of SNPs to make further analysis informative. In such cases, an allele-specific PCR was designed to selectively amplify one of the SNP-alleles and also to confirm the result of the screening (see Section 2.11). An appropriate restriction digest or dideoxy sequencing of the allele-specific products from the child was then performed, to determine on which allele the mutation was present and hence identify the parent of origin (see Table 7, Appendix). This is summarised in Table 5.1 and Figures 5.5 below. Of the 16 *de novo* cases of *TCF12* mutations identified (see Chapter 4 and Table 5.1), in 6 families it was demonstrated that the mutation arose on the paternally-inherited allele implying a paternal origin. In the case of family #7, a mutation-specific PCR was designed and the DNA of the proband sequenced to identify a paternally-derived SNP in phase with the mutant allele. The strategies used for each type of assay is shown below in Figure 5.4.

A



B



C

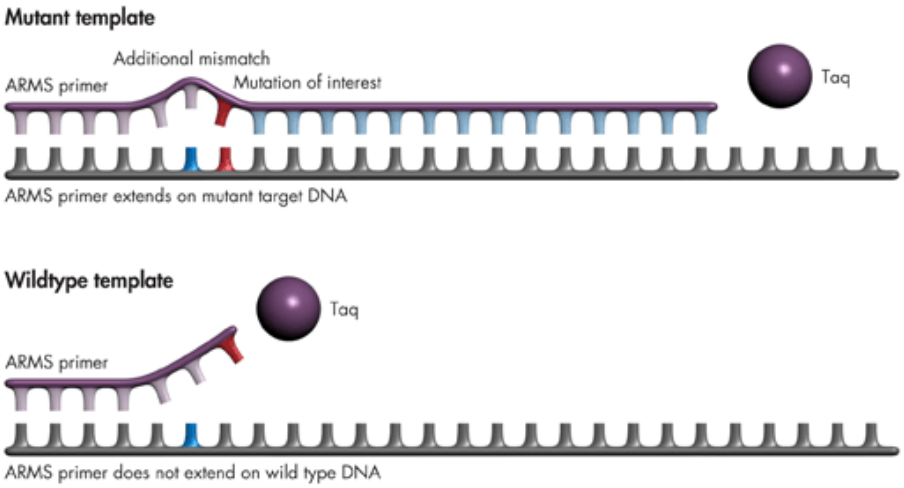


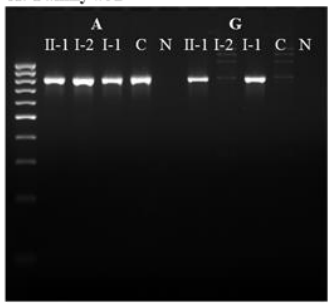
Figure 5.4 Strategies to amplify specific alleles to inform parent of origin for *TCF12* mutations. A: Allele-specific PCR uses alternate forward primers with a common reverse primer or a common forward with alternate reverse primers to

differentially amplify a specific allele for the hypothetical SNP [G/A]. A deliberate mismatch at the penultimate base at the 3' end destabilises the primer-DNA template complex so the primer with this mismatch is refractory to primer extension (From You et al., 2008)**. B:** Utilising allele-specific PCR then digestion of specific PCR products with an appropriate restriction enzyme demonstrates a mutant allele in phase with the paternally-inherited G allele thereby implying a paternal origin to the mutation in question (adapted from Moloney et al., 1996). **C:** Mutation-specific PCR utilises a common forward primer and reverse primers specific to either the mutant or the wild-type allele. This will give an amplicon that will include the SNP of interest in phase with either the wild-type or mutant allele. The mismatch at the 3' end of the primer will allow the Taq polymerase to discriminate between matched and mismatched template. Knowledge of SNP genotypes in an informative proband-parent trio will allow the phase to be established and the parent of origin for the mutation to be correctly assigned (From SABiosciences, 2010).

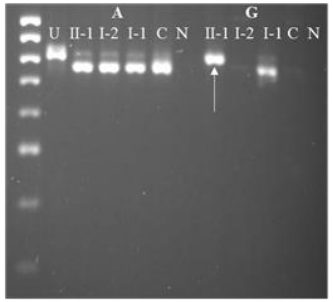
Family #	Proband ID	Exon	Mutation	Protein	Heterozygous SNPs ≤1 kb of mutation	MAF	Proband genotype	Mother genotype	Father genotype	Parent of Origin	Father's age at birth of proband
5	2065	10	822C>G	(sp)	**rs2290939 [T>C]** rs12439017 [(T>C]	**0.169** 0.045	**[T/C]** [T>C]	**[C/C]** [T/T]	**[T/T]** [T/C]	**Father**	**39.3**
6	98D4142	10	825G>C	L275F (sp)	None	-	-	-	-	-	
7	838	11	842_843delCA	S281Cfs*56	**rs2290938 [T>C]** rs2248620 [G>A]	**0.36** **0.171**	**[T/C]** [A/A]	**[T/C]** [A/A]	**[C/C]** [A/A]	**Father**	**30.4**
16	4437	16	1283T>G	L428*	None	-	-	-	-	-	
17	3956	16	1312delC	L438Cfs*81	None	-	-	-	-	-	
21	5189	16	1453C>T	R485*	**rs2615227 [C>G]**	**0.40**	**[C/G]**	**[C/C]**	**[G/G]**	**Father**	**26.8**
25	01D4012	17	1514C>G	S505*	None	-	-	-	-	-	
26	4855	17	1530dupA	V511Sfs*11	None	-	-	-	-	-	
28	3664	17	1582+2T>G	(sp)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
29	4618	18	1628delA	K543Rfs*20	**rs11631084 [G>A]** rs11636340 [T>A] rs7182038 [C>A]	**0.356** 0.356 0.195	**[A/G]** [T/A] [C/A]	**[A/G]** [T/A] [C/A]	**[G/G]** [T/A] [C/A]	**Father**	**31.8**
32	4951	18	1691C>G	S564*	**rs56163695 [A>G]**	**0.012**	**[A/G]**	**[A/A]**	**[A/G]**	**Father**	**36.4**
33	08D7786	18	1713delA	D572Ifs*19	None	-	-	-	-	-	
35	4594	19	1806delG	R602Sfs*31	None	-	-	-	-	(Mother)	
38	4379	19	1963G>T	E655*	None	-	-	-	-	-	
41	*5463*	*12*	*c.1036-6G>A*	*(sp)*	***rs12904933 [C/T]***	***0.36***	***[C/T]***	***[T/T]***	***[C/T]***	***Father***	***31.2***
42	*5547*	*16*	*1453C>T*	*R485**	*None*	-	-	-	-	-	

Table 5.1 Families where the *TCF12* mutation arose *de novo* were screened by dideoxy sequencing to identify informative SNPs. This allowed further investigation of the parental origin of the mutation. Results in bold indicate informative SNPs used for analysis. Other heterozygous SNPs also identified are listed. Numerous homozygous SNPs were identified during the screening process but are not shown for clarity. Italicised results refer to samples referred from other diagnostic centres for further investigation.

A: Family #32

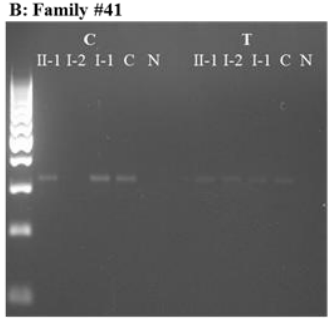


Genotype:
II-1 (P) = A/G, I-2 (M) = A/A,
I-1 (F) = A/G

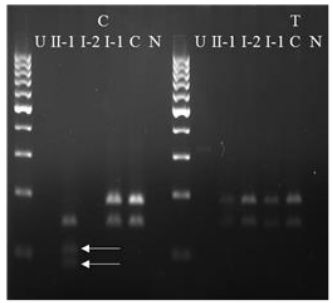


AhdI digest:
WT = 768 bp, 48 bp MT = 816 bp

B: Family #41

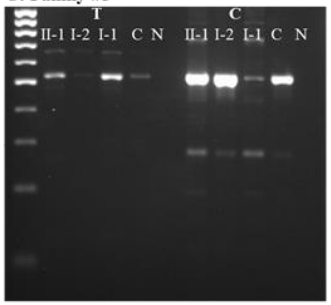


Genotype:
II-1 (P) = C/T, I-2 (M) = T/T,
I-1 (F) = C/T

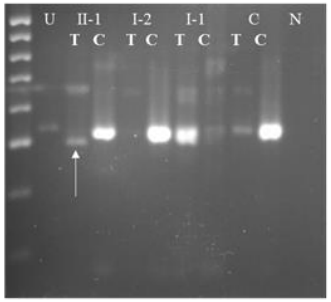


BsaI digest:
WT = 183 bp,140 bp
MT = 140 bp, 101 bp, 82 bp

C: Family #5

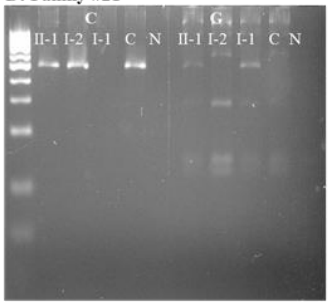


Genotype:
II-1 (P) = T/C, I-2 (M) = C/C, I-1
(F) = T/T

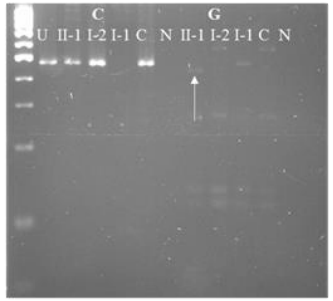


HpaII digest:
WT = 525 bp MT = 495 bp, 30 bp

D: Family #21

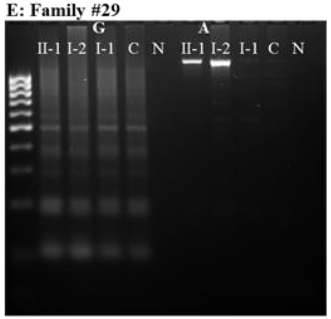


Genotype:
II-1 (P) = C/G, I-2 (M) = C/C, I-1
(F) = G/G



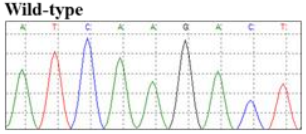
BclI digest:
WT = 468 bp MT = 431 bp, 37 bp

E: Family #29

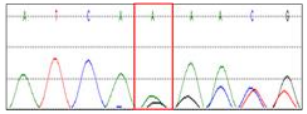


Genotype: II-1 (P) = G/A, I-2 (M) = G/A,
I-1 (F) = G/G

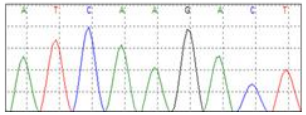
Wild-type



Proband

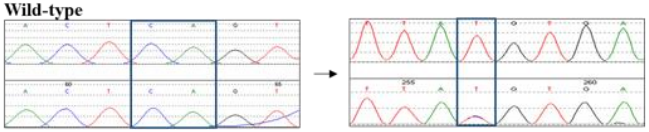


A-specific product

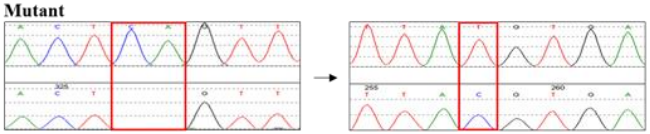


F: Family #7

Wild-type



Mutant



Genotype: II-1 (P) = T/C, I-2 (M) = T/C,
I-1 (F) = C/C

Figure 5.5. Investigation into parental origins of *de novo TCF12* mutations. Genotypes identified by dideoxy sequencing as shown in Table 5.1 are given below each PCR result. **A:** (Left) allele-specific PCR confirming genotypes for rs56163695 in Family #32 (amplicon = 816 bp). (Right) restriction digest of samples, loaded in the same order, with AhdI demonstrates no digestion of the paternally-derived G allele in the sample from II-1, which carries the mutation (white arrow). All other samples, including the maternally-derived A allele in the sample from II-1, digest normally. **B:** Allele-specific PCR confirming genotypes for rs12904933 in Family #41 (amplicon = 323 bp). (Right) Restriction digest of samples in same order with BsaI demonstrates digestion of the paternally-derived C allele that carries the mutation (white arrows). All other samples, including the maternally-derived T allele in the sample from II-1, digest normally **C:** (Left) Allele-specific PCR confirming genotypes for rs2290939 for Family #5 (amplicon = 525 bp). (Right) Restriction digest with HpaII demonstrates digestion of the paternally-derived T allele that carries the mutation (white arrow). Note, gel on right side is loaded adjacently with each SNP allele (i.e. T or C) to more clearly delineate the 30 bp difference between wild-type and mutant bands. **D:** (Left) Allele-specific PCR confirming genotypes for rs2615227 for family #21 (amplicon = 468 bp). (Right) Restriction digest with BclI demonstrates digestion of the paternally-derived G allele that carries the mutation (white arrow). All other samples, including the maternally-derived C allele sample from II-1, digest normally **E:** (Left) Allele-specific PCR for rs11631084 amplifies maternally derived A-allele only in family #29 (amplicon = 1463 bp) (amplification of G-specific product failed). (Right) Sequencing the A-allele specific product in the proband reveals no c.1628delA mutation, implying the mutation has arisen on the other, paternally-derived G allele. The proband's mutant sequence (middle) is shown for reference **F:** Sequencing results of a mutation-specific PCR using primers specific to wild-type (top - blue boxes) and mutant alleles (bottom – red boxes) from proband DNA. This assay shows a paternally-derived C-allele in phase with the c.842_843delCA mutation. All samples were run on 3% -3.5% Agarose gel with a 1 kB DNA ladder. Key: II-1 = proband, I-2 = mother, I-1 = father, U = uncut proband DNA, C = control DNA (of prior unknown genotype), N = negative water control.

5.6 Maternally-derived *TCF12* mutations

Although the analyses above only revealed paternally originating mutations, in two other families, different approaches demonstrated a maternal origin in each case. Firstly, in Family #28, a mosaic state for the c.1583+2T>G mutation was identified in DNA isolated from a mouth swab obtained from the maternal grandmother (Individual I-2) of the proband by a diagnostic laboratory. Due to this new finding, the sequencing for this sample was repeated during this study and visualised using a different computer programme to make the mutation more obvious (see Figure 5.4). This phenomenon of maternal mosaic mutations has been seen previously in craniosynostosis in the context of craniofrontonasal syndrome due to mutations in *EFNB1* (Twigg et al., 2006). A second case of maternal origin was identified in another maternal grandmother (individual II-2) in Family #35 by analysis of haplotype segregation. This second instance of maternal origin did not have any evidence of mosaicism.

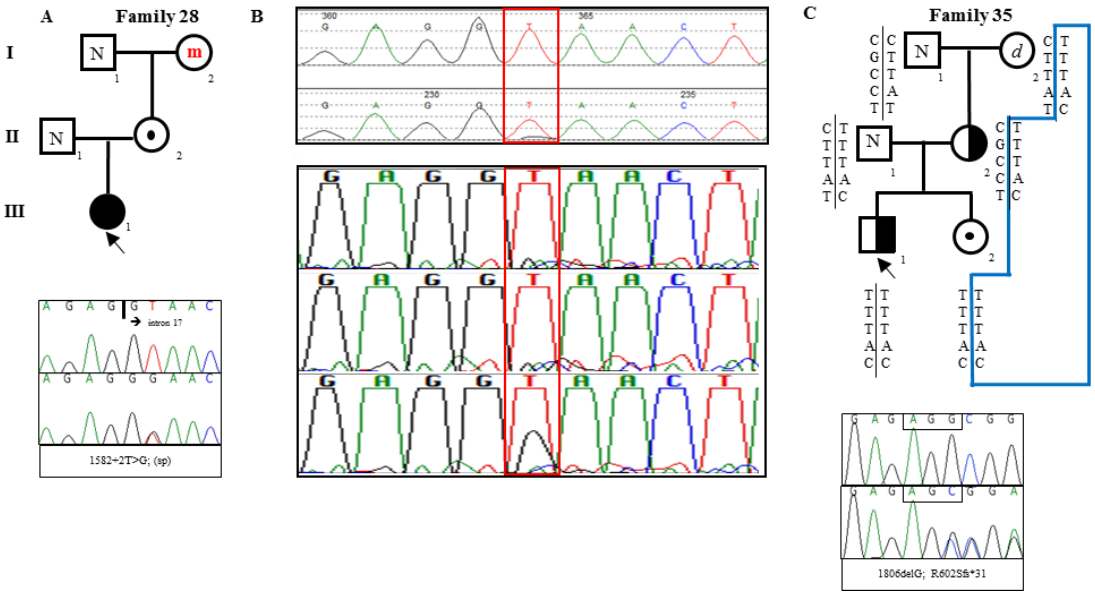


Figure 5.6 Discovery of maternal-origin *TCF12* mutations.

A: Pedigree where maternal grandmother (I-2) was found to be mosaic for the c.1583+2T>G mutation (denoted by **m**). **B:** Original chromatogram where the mosaic mutation was present at very low levels and missed initially (i.e. discrete black horizontal line within highlighted red box). **C:** Visualisation of sequencing data using Sequencher® computer programme. Top trace is grandfather (I-1), middle is unaffected control, and lower is grandmother (I-2). With the 'gain' turned up on the chromatogram viewer, the mosaic mutation is more apparent. **C:** Analysis of genetic background (see Chapter 6) reveals segregation of a particular haplotype ('TTTAC', denoted by blue outline) originating in the maternal grandmother (I-2, denoted by *d*) on which the mutation arose in the mother (II-2).

5.7 Discussion

A non-penetrance rate of 60% was observed in carriers of *TCF12* mutations who had normal phenotypic appearances and cephalometric measurements. In Hirschsprung disease and thrombocytopenia with absent radii, haplotype has been shown to be important in disease susceptibly and it appears logical that the genetic background of penetrant and non-penetrant *TCF12*-mutation carriers should also be investigated further (see Chapter 6).

Analysis of the parent of origin in *de novo* cases involving proband-parent trios has revealed 6 instances of the mutation arising on the paternally-inherited allele. The average age of these fathers was 32.7 years old. This is 0.2 years younger than the 2013 standardised mean age for fathers in England and Wales of 32.9 years old according to the Office for National Statistics (Vital Statistics Outputs Branch, 2014). The standardised mean age for mothers in 2013 was 30.0 years old. In comparison to other craniosynostosis syndromes due to mutations in *FGFR2* or *FGFR3* that demonstrate a paternal-age effect (with fathers on average 2 years older than the national average) fathers with *TCF12* mutation-positive offspring were also younger (Goriely & Wilkie, 2012).

Beyond craniosynostosis, a wide variety of disorders ranging from achondroplasia and Costello syndrome (short stature, intellectual disability, joint laxity, distinctive facies) to Ehlers-Danlos (inherited

connective tissue disorders) and Waardenburg syndromes (congenital deafness and unusual pigmentation of hair, skin and irides) have associations with advanced paternal age (Momand et al., 2013). The unifying pathological mechanism is due to increased frequency of *de novo* point mutations in germs cells of older fathers in addition to men transmitting a higher number of mutations to their offspring than women. A large whole-genome sequencing study of 79 Icelandic trios has shown an exponential doubling in paternal mutations every 16.5 years with significant increases in risk of schizophrenia and autistic spectrum disorder (Kong et al., 2012)

There were also 2 maternally-derived mutations observed during this study (see Section 5.6), with one individual demonstrating a case of a mosaic mutation, which has been more commonly described in craniofrontonasal syndrome (Twigg et al., 2006). Furthermore, all previously described paternal age effect mutations tend to result in gain-of-function from a biochemical standpoint whereas all *TCF12* mutations thus far identified, cause loss-of-function. Therefore, despite a small sample size of 6 fathers as the parent of origin, *TCF12*-related craniosynostosis is unlikely to exhibit a paternal age effect. This is consistent with a wide variety of disorders including Treacher-Collins (*TCOF1*), Retinoblastoma (*RB*) and tuberous sclerosis (*TSC2*) (Splendore et al., 2003). Knowledge of the origin of the mutant allele will also be used to refine the assignment of *TCF12* haplotypes and analysis of allele frequencies (see Chapter 6).

CHAPTER 6

HAPLOTYPE ANALYSIS OF *TCF12*-MUTATION POSITIVE INDIVIDUALS TO ASSESS INFLUENCE ON CRANIOSYNOSTOSIS PHENOTYPE

6.1 Haplotypes and The HapMap database

In 2002, a landmark international collaboration between researchers in 6 countries began to establish the International HapMap Project with the aim of developing a haplotype map of the genome to delineate common patterns of genetic variation in humans. Datasets were released in 3 phases, with the last in 2009. It is thought that within the 0.1% difference that exists between the DNA sequences of 2 individuals, lie variants that have important implications in their risk of disease susceptibility. There is on average one variant every 17-21 bases pairs (Nelson et al., 2012) with the commonest pattern of variation occurring at single bases known as single nucleotide polymorphisms (SNPs), occurring once every 300 nucleotides (Harel et al, 2014). Each form of a SNP is called an allele and the set of alleles an individual has is a genotype. Sets of adjacent SNPs along the same chromosome are inherited in blocks and the pattern of SNPs on a block is referred to as a haplotype. Although each block may contain a large number of SNPs, specific ones can be used to identify haplotypes – tag-SNPs (see below). Alternatively, a haplotype can be considered as a linear arrangement of alleles along a chromosome that is inherited *en bloc* from parent to offspring in the absence of recombination (The International HapMap Consortium, 2005).

An individual carrying a particular SNP at one locus may predictably carry other certain alleles at nearby sites - this non-random association is

a phenomenon referred to as linkage disequilibrium (LD) (International HapMap Consortium, 2005). Recombination over many generations will break down LD and the ancestral haplotype, therefore analysis of certain polymorphisms (e.g. ApoE4 associated with Alzheimer's disease) necessitate typing of SNPs that are very close to or including the polymorphism itself, in order to get relevant information (Roses, 2000).

Of approximately 48 million SNPs [(global minor allele frequency, (MAF) > 1%] that exist in humans (Koboldt, 2012), HapMap has identified 250,000 - 500,000 'tag' SNPs that provide almost as much genetic information. From a research standpoint, genotyping tag-SNPs at a particular chromosomal locus can be used to identify most of the genetic variation at that region. The principle is summarised in Figure 6.1 below. The LD of alleles at two or more loci is a measure of statistical association (Knight, 2009). Analysis of areas of high LD allows for easier variant mapping as it is not necessary to type every SNP. By carefully choosing and typing specific flanking markers that are highly statistically correlated, a disease gene can be mapped. LD can be measured by its coefficient, *D*, which is the deviation in observed frequency of haplotype from expected if the genotypes occurred independently. Comparison between pairs of SNPs uses a normalised value, *D´*, which is a ratio of observed LD to the maximum possible value in deviation of given allele frequencies. The square of the correlation coefficient is r^2 and denotes the statistical correlation between 2 SNPs. A perfect correlation between 2 SNPs is denoted by $r^2 = 1$ and

implies the SNPs give exactly the same information. By contrast, no correlation is indicated by $r^2 = 0$ (perfect equilibrium). Both r^2 and D', are 1 in the situation of complete LD and 0 when there is no LD. Where they differ is if $r^2 = 1$, 2 haplotypes are present (from 4 that are possible). By comparison, if $D' = 1$, it can correspond to 2 or 3 haplotypes, thus creating more uncertainty (Knight, 2009).

The overarching aim of the International HapMap Project was to provide a description of haplotype blocks and the SNPs that tag them. Therefore, in a specified ethnic population, a reduced number of SNPs will need to be genotyped - from 48 million to 200, 000 - 300, 000 (International HapMap Consortium, 2005). This significantly reduces the burden of genotyping while increasing the efficiency of the whole process. This is especially true in areas of moderate to high LD. Therefore, tag-SNPs can tag conserved haplotypes with relatively little loss of information (Johnson et al., 2001).

To determine haplotype, the most straightforward method is to collect DNA from family members and genotype them for a particular SNP. This generally requires 2 or 3 generations of family members to track phase unambiguously. Determining the correct orientation of alleles on a particular chromosome is a potential pitfall in haplotyping. This is especially true in an individual who is a double heterozygote and therefore their association of a particular allele with a chromosome cannot be certain. This problem is exacerbated as more SNPs are

attempted to be analysed. If an individual is doubly homozygous, phase and therefore haplotype is straightforward and unambiguous. Similarly, if the person is homozygous for one SNP, and heterozygous for another, then again haplotypes can be assigned unambiguously.

Other drawbacks of this method can include significant effort and cost to acquire samples and extended family members may not always be available. This is especially true if the disease occurs in later years of life such as certain neurodegenerative conditions. Otherwise, molecular techniques such as allele-specific PCR or long insert cloning can be used although automation is difficult and throughput is usually low. If family members are not available or molecular haplotyping is not desired, statistical estimation can be used to identify haplotype probability values that optimise probability of observed data by utilising likelihood-based estimation maximisation algorithms. This process aims to identify the most likely haplotype for individuals with ambiguous phase by virtue of an iterative process which continues until the estimates between 2 consecutive iterations is minimal (Fallin & Schork, 2000) and the ambiguous haplotype is inferred from homozygotes and single heterozygotes. This will provide the best estimate of haplotype frequencies that are consistent with observed genotypes in a specified data set (Palmer, 2007).

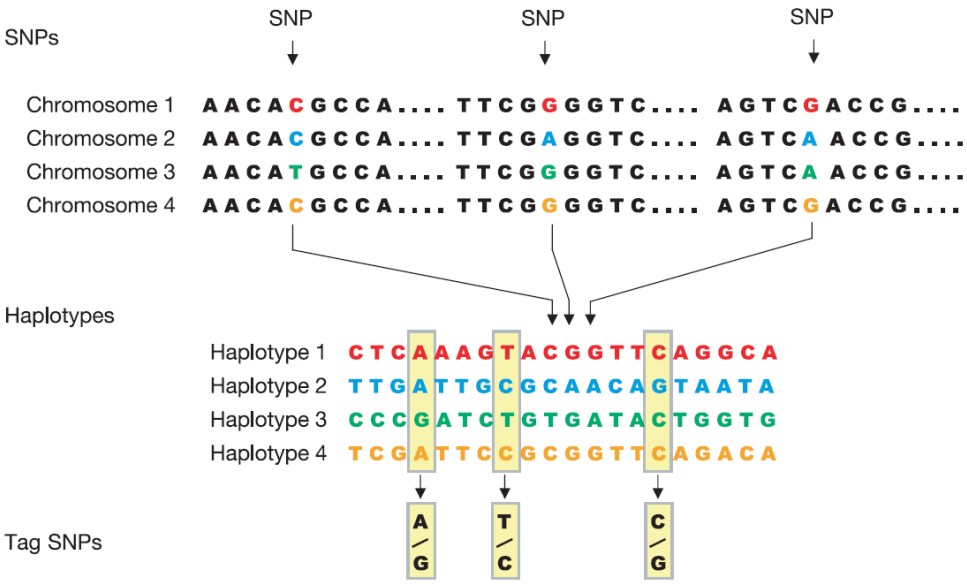


Figure 6.1 Relationship of SNPs, Haplotypes and tag SNPs. *Top:* a short segment of DNA with 4 versions of the same chromosomal locus. Variation occurs at 3 locations and each SNP has 2 possible alleles. *Middle:* the combination of alleles at nearby SNPs makes 4 possible haplotypes. The genotypes for 20 SNPs found over 6000 nucleotide bases is shown. *Bottom:* genotyping 3 tag SNPs out of 20 SNPs is enough to correctly identify the 4 haplotypes. *From* The International HapMap Consortium (2005).

TCF12 is located in a region of strong LD in a relatively variable region of the human genome (see Figure 6.2). By contrast, its partner protein TWIST1 is much less polymorphic when analysed on the HapMap Project website, although its contiguous genomic regions are themselves variable. Additionally, *TWIST1* has 84 listed variants compared to 720 for *TCF12* listed on the Exome Aggregation Consortium website containing information on over 60,000 unrelated individuals. The most frequent variant in *TWIST1* is observed 117 times out of 117058 alleles from the large cohort of individuals listed on that website. Therefore, an investigation into *TCF12* haplotype background and its role in disease

susceptibility (i.e. coronal craniosynostosis) was chosen. Furthermore, the differences in phenotype that is observed in heterozygous mutation-positive individuals may be related to relative expression levels of wild-type alleles (see Figure 6.3).

As described in Section 5.4, work by Emison et al. (2010) demonstrated that the mutant allele associated with Hirschsprung disease was observed on both ancestral and derived haplotype backgrounds. Although present at higher frequencies on the former, a recombination event will have caused many SNPs in linkage disequilibrium with a causal variant to switch between populations. This would complicate any interpretations of allelic association. Any investigation of *TCF12* haplotypes must also be mindful of this fact. Nevertheless, a specific *TCF12* haplotype may be a useful indirect marker of disease, particularly if possessed by a non-penetrant individual as it may be in LD with a causal variant or other such susceptibility locus. This would be consistent with a complex additive model of inheritance whereby possession of a *TCF12* coding mutation or specific haplotype would be needed to give a coronal synostosis phenotype (Ruiz-Ferrer et al., 2006).

Figure 6.2 Plots of *TCF12* demonstrating strong LD across the region occupied by the gene. Image at 200 kb. The markers in red are inherited through generations as a single unit or haplotype block. The red colour indicates strong linkage disequilibrium (*D'* > 0.8). White areas indicate uninformative gaps between SNP markers. Typically, genes that are involved in core biological processes are preferentially located in regions of high LD whereas those that interact with the environment, immune system or neurophysiological processes reside in regions of low LD.

Figure 6.3 Differences in levels of wild-type *TCF12* expression in individuals heterozygous for *TCF12* mutations may account for the phenotypic differences seen in penetrant and non-penetrant individuals. In the case of family #30 (c.1642_1645delGAAA; p.E548Rfs*14), the proband (left) is penetrant for bilateral coronal synostosis and demonstrates typical phenotypic features. Her mother (right) has normal clinical features with no discernible craniofacial phenotype. Both individuals are heterozygous for the same *TCF12* mutation.

6.2 Utilisation of tag-SNPs to genotype a population of interest and identification of validated haplotypes

The International HapMap Project website (http://hapmap.ncbi.nlm.nih.gov/) was used to obtain phased haplotype data (see Section 2.12 for window settings). The Tagger web browser was then used to utilise a pairwise tagging algorithm that picked 5 tag-SNPs from the CEU HapMap population (http://www.broadinstitute.org/mpg/tagger). This method worked very well, capturing 79 of 79 alleles observed over this region with 100% of these with a mean r^2 of 0.917. The set of 5 tag-SNPs selected (rs477426, rs2585082, rs1628067, rs12901270, rs2440979) acted as a direct proxy for untyped SNPs. They were located in a region of the genome with high LD, thus allowing identification of genetic variation without the need to genotype every SNP. This is illustrated in Figure 6.4 below.

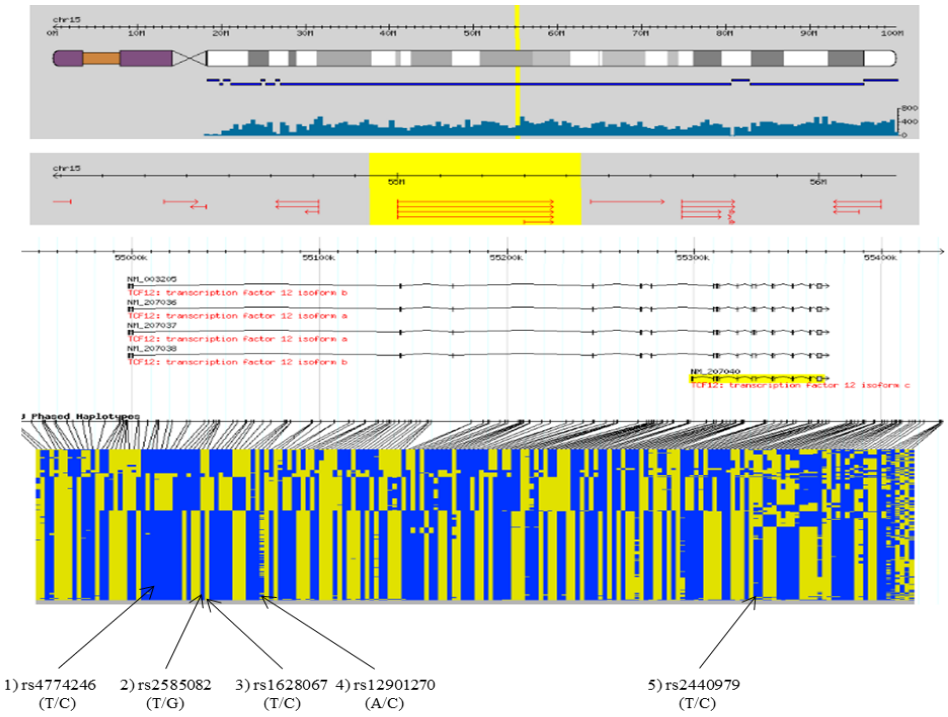


Figure 6.4: Phased haplotype display of *TCF12* for CEU* population and locations of 5 tag-SNPs. This set of tag-SNPs was used for subsequent haplotyping of mutation positive individuals and their family members. * CEU - Utah residents with Northern and Western European ancestry from the CEPH collection.

Phased haplotype data from the HapMap website was downloaded to obtain validated haplotypes, initially from an ethnically matched control CEU population [Utah residents with Northern and Western European ancestry from the CEPH (Centre d' Étude du Polymorphisme Humain) collection]. This identified 10 known haplotypes from a total of 234 individuals. Data from 10 other ethnic populations comprising 2022 individuals were also obtained, adding 6 further known haplotypes (i.e. a total of 16 possible *TCF12* haplotypes). This is shown in Table 6.1 below and the full data obtained in Tables 11-14, Appendix.

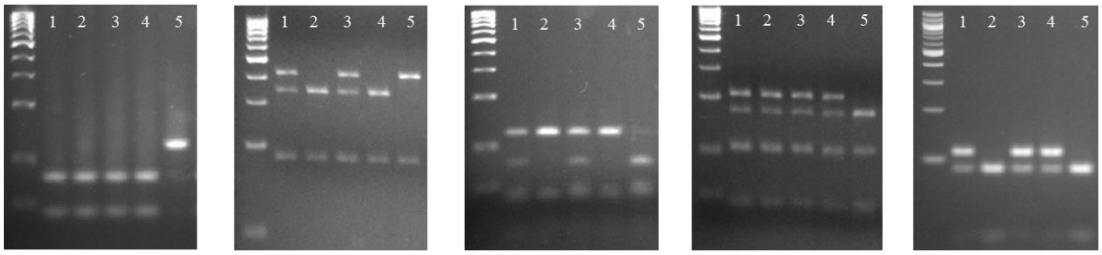
CEU Population		All HapMap populations	
NNNNN	8	**NNNNN**	36
CGCAC	64	**CGCAC**	339
CGCAT	4	**CGCAT**	29
CGCCC	47	**CGCCC**	147
CGCCT	24	**CGCCT**	56
CTTAC	8	**CTTAC**	86
CTTAT	29	**CTTAT**	822
TTCAC	2	**TTCAC**	130
TTCAT	4	**TTCAT**	53
TTTAC	8	**TTTAC**	64
TTTAT	44	**TTTAT**	244
Total	242	**TGTAT**	40
Total known	**234**	**CGTAT**	2
		CTCAC	4

	CTCAT	4
	CTTCC	1
	TGCAT	1
	Total	2058
	Total known	**2022**

Table 6.1 Summary of validated haplotypes obtained from The International HapMap Project website. Phased haplotypes are shown for CEU (left column) and all ethnic populations (right column) to highlight the full range of possible haplotypes.

6.3 Genotyping of controls to establish allele frequencies of each tag SNP in an unaffected, control population

To obtain allele frequencies for the 5 tag-SNPs in a control population, 2 control plates consisting of 156-192 individuals were genotyped (full genotyping results are given in Tables 15-18, Appendix). This provided an additional control population with a genetic background that more closely reflects that of the UK population rather than continental Europe (as per CEU HapMap population). It also provided a means of internal quality control as the allele frequencies subsequently obtained were comparable to the HapMap data, demonstrating a low rate of genotyping error. An example of the genotyping results for 5 individuals and proportions of each SNP-allele for all 5 tag-SNPs is shown in Figure 6.5 below.



rs4774246 [T/C]

Amplicon = 204 bp

RsaI digest:

T allele = 142 bp , 62 bp

C allele = 76 bp, 66 bp, 62 bp

Genotypes:

1 [C/C]
2 [C/C]
3 [C/C]
4 [C/C]
5 [T/T]

rs2585082 [T/G]

Amplicon = 614 bp

BfaI digest:

T allele = 401 bp, 179 bp, 22 bp, 12 bp

G allele = 332 bp, 179 bp, 69 bp, 22 bp, 12 bp

Genotypes:

1 [G/T]
2 [G/G]
3 [G/T]
4 [G/G]
5 [T/T]

rs1628067 [T/C]

Amplicon = 254 bp

MseI digest:

T allele =76 bp, 69 bp, 66 bp, 43 bp

C allele = 142 bp, 69 bp, 43 bp

Genotypes:

1 [C/T]
2 [C/C]
3 [C/T]
4 [C/C]
5 [T/T]

rs12901270 [A/C]

Amplicon = 351 bp

BfaI digest:

A allele = 169 bp, 104 bp, 39 bp, 39 bp

C allele = 208 bp, 104 bp, 39 bp

Genotypes:

1 [A/C]
2 [A/C]
3 [A/C]
4 [A/C]
5 [A/A]

rs2440979 [T/C]

Amplicon = 112 bp

BstxI digest:

T allele = 112 bp

C allele = 89 bp, 23 bp

Genotypes:

1 [C/T]
2 [C/C]
3 [C/T]
4 [C/T]
5 [C/C]

Figure 6.5: Genotyping results for 5 individuals from a UK control population for each of the 5 tag-SNPs. For each SNP, the size of the PCR product (amplicon) is given as well as the appropriate restriction digest used with associated SNP alleles/ fragment sizes. A 1 kb DNA ladder and 3-5% agarose gels were used.

SNP		Genotype 1	Genotype 2	Genotype 3	χ^2	*P value*	Numbers/proportions of SNP allele 1	Numbers/proportions of SNP allele 2	Total number of individuals genotyped
		CC	CT	TT			T	C	
rs447426 [T/C]	Obs Exp	93 83.2	76 73.5	15 16.2	0.196	0.658	106 0.288	262 0.712	184
		GG	GT	TT			T	G	
rs2585082 [T/G]	Obs Exp	57 55.5	92 94.9	42 40.5	0.18	0.671	176 0.461	206 0.539	191
		CC	CT	TT			T	C	
rs1628067 [T/C]	Obs Exp	61 59.7	91 93.6	38 36.7	0.147	0.701	167 0.438	214 0.562	190
		AA	AC	CC			A	C	
rs12901270 [A/C]	Obs Exp	105 101.5	64 71.1	16 12.5	1.841	0.174	274 0.741	96 0.259	185
		TT	CT	CC			C	T	
rs2440979 [T/C]	Obs Exp	48 48.5	97 96	47 47.5	0.021	0.884	193 0.503	191 0.497	192

Table 6.2 Proportions of each SNP allele following genotyping of a UK control population. Genotyping this population (156-192 individuals) provided data on an additional control population whose genetic background would more closely resemble the UK compared to continental Europe as in the HapMap database. It would also help to detect any genotyping errors prior to investigation of *TCF12*-muation positive individuals and their family member. If P value for the $\chi 2$ statistic was <0.05 (with 1 degree of freedom), then the allele in question would not be in Hardy-Weinberg Equilibrium indicating a genotyping error.

6.4 Construction of *TCF12* haplotype pedigree reveals presence of transmitted partner alleles and non-transmitted bystander alleles

The analysis of *TCF12* haplotypes of mutation positive individuals and their family members would aim to establish whether or not there were any discernible differences that may be important in disease susceptibility (see Figure 6.3). Haplotypes were constructed using data from family members of at least 2 generations in familial cases of transmission (see Figure 6.6 below) and in cases of *de novo* mutations, using information from the parent of origin investigation to assign the mutant allele (see Section 5.5). The mutant *TCF12* allele (M) was observed to be transmitted with certain partner alleles (P). Additionally, a number of bystander alleles (B) were seen in parents who do not carry the mutation and were not transmitted through the pedigree. The results for the 48 families that were haplotyped are shown in Figure 6.7 and a summary of haplotypes obtained given in Table 19, Appendix.

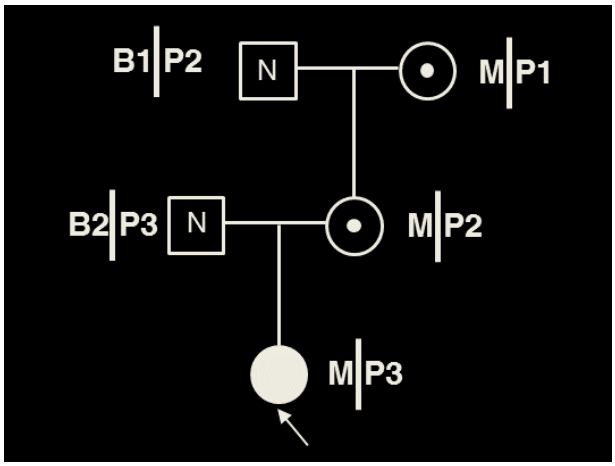
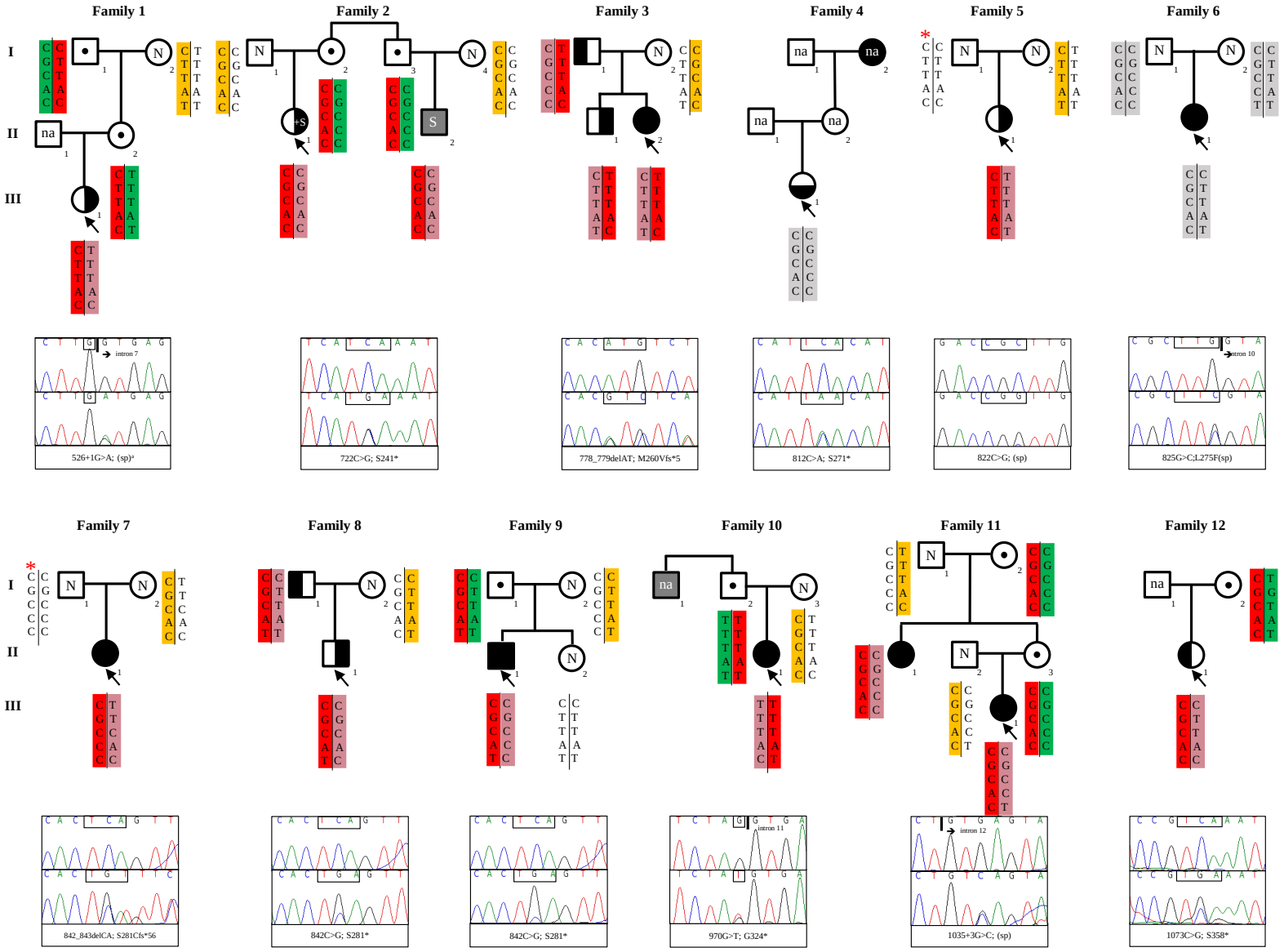
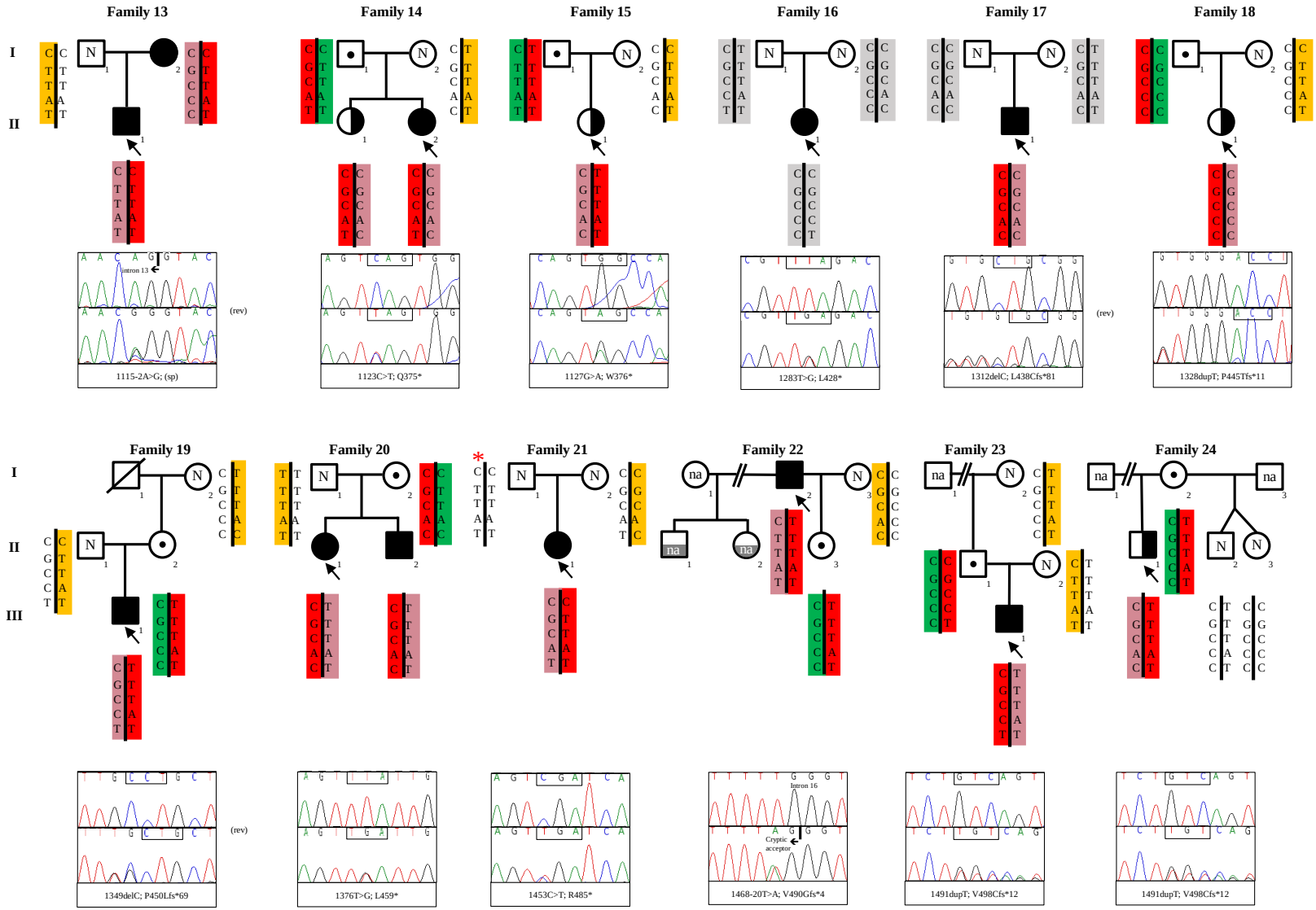
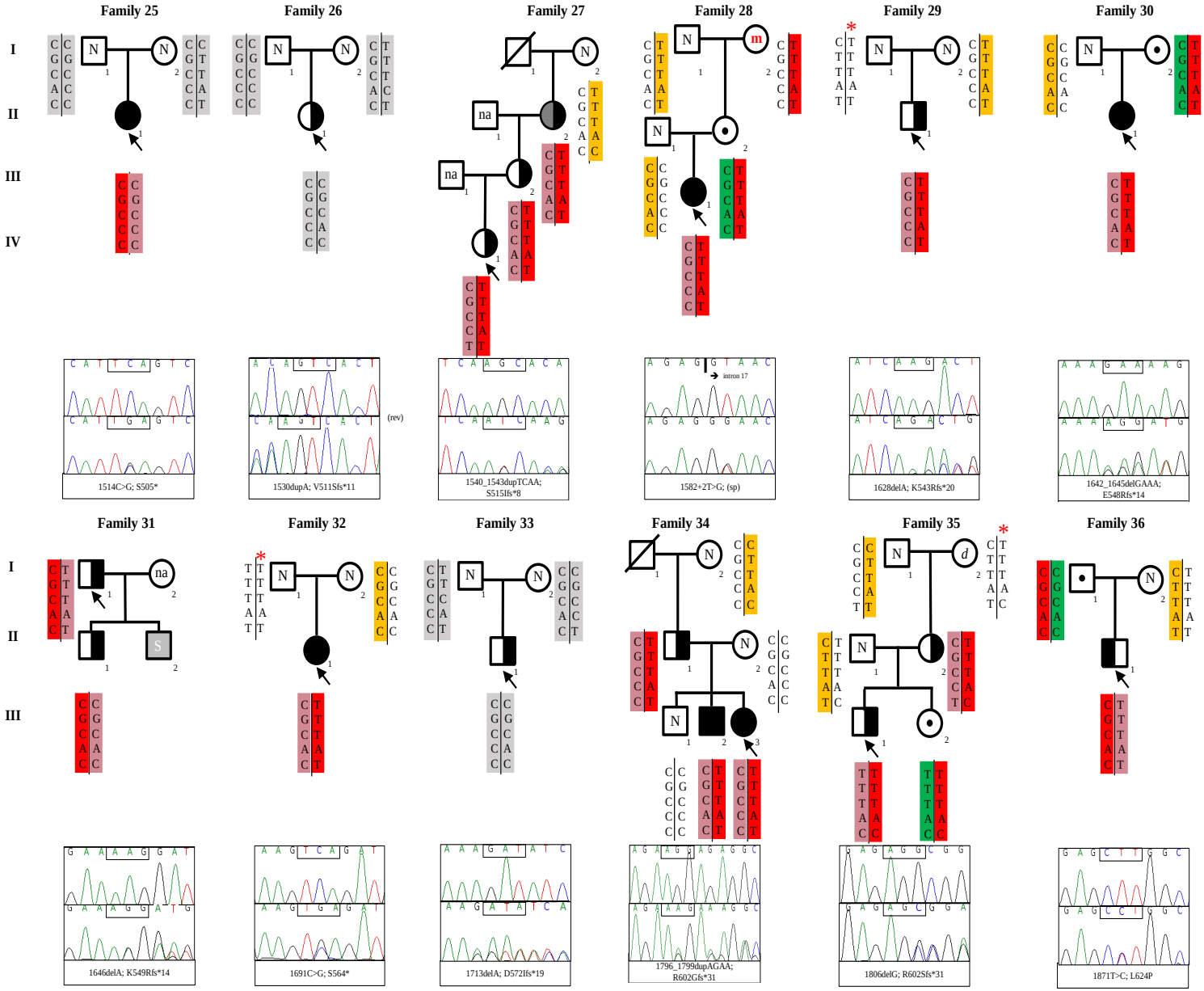


Figure 6.6: Construction of haplotypes in families where the *TCF12* mutation is observed in 2 or more generations. Key: M = mutant *TCF12* allele; P1/P2/P3 = partner *TCF12* alleles; B1/B2 = bystander *TCF12* alleles.


Family 1
Family 2
Family 3
Family 4
Family 5
Family 6
I
II
III
na
+S
S
526+1G>A; (sp)a
722C>G; S241*
778_779delAT; M260Vfs*5
812C>A; S271*
822C>G; (sp)
825G>C;L275F(sp)
Family 7
Family 8
Family 9
Family 10
Family 11
Family 12
842_843delCA; S281Cfs*56
842C>G; S281*
842C>G; S281*
970G>T; G324*
1035+3G>C; (sp)
1073C>G; S358*


Family 13
Family 14
Family 15
Family 16
Family 17
Family 18
I
II
1115-2A>G; (sp)
1123C>T; Q375*
1127G>A; W376*
1283T>G; L428*
1312delC; L438Cfs*81
1328dupT; P445Tfs*11
(rev)
Family 19
Family 20
Family 21
Family 22
Family 23
Family 24
III
1349delC; P450Lfs*69
1376T>G; L459*
1453C>T; R485*
1468-20T>A; V490Gfs*4
1491dupT; V498Cfs*12
1491dupT; V498Cfs*12


Family 25
Family 26
Family 27
Family 28
Family 29
Family 30
I
II
III
IV
1514C>G; S505*
1530dupA; V511Sfs*11
(rev)
1540_1543dupTCAA; S515Ifs*8
1582+2T>G; (sp)
intron 17
1628delA; K543Rfs*20
1642_1645delGAAA; E548Rfs*14
Family 31
Family 32
Family 33
Family 34
Family 35
Family 36
1646delA; K549Rfs*14
1691C>G; S564*
1713delA; D572Ifs*19
1796_1799dupAGAA; R602Gfs*31
1806delG; R602Sfs*31
1871T>C; L624P

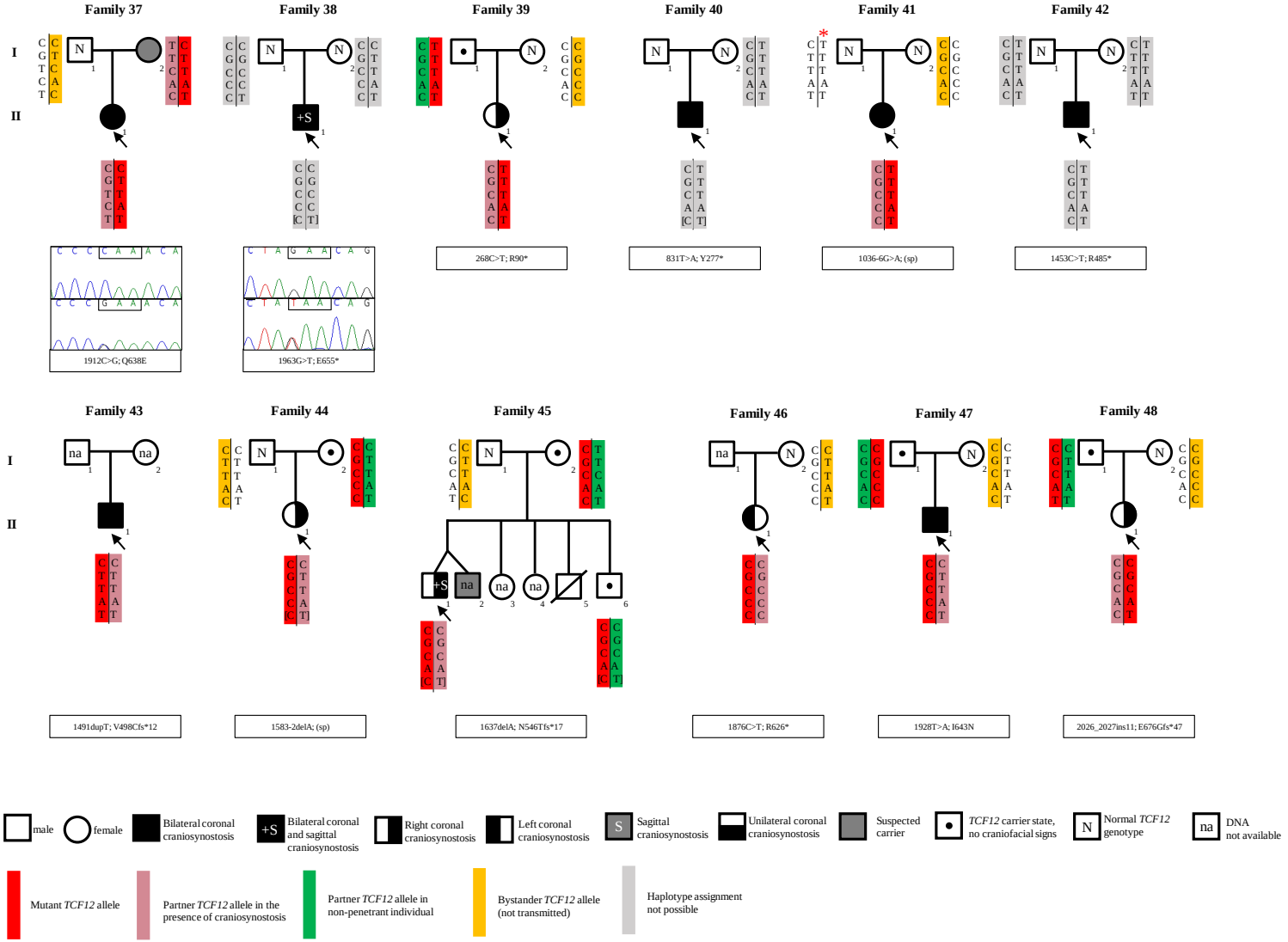

Family 37
Family 38
Family 39
Family 40
Family 41
Family 42
I
II
+S
1912C>G; Q638E
1963G>T; E655*
268C>T; R90*
831T>A; Y277*
1036-6G>A; (sp)
1453C>T; R485*
Family 43
Family 44
Family 45
Family 46
Family 47
Family 48
na
1491dupT; V498Cfs*12
1583-2delA; (sp)
1637delA; N546Tfs*17
1876C>T; R626*
1928T>A; I643N
2026_2027ins11; E676Gfs*47
male
female
Bilateral coronal craniosynostosis
Bilateral coronal and sagittal craniosynostosis
Right coronal craniosynostosis
Left coronal craniosynostosis
Sagittal craniosynostosis
Unilateral coronal craniosynostosis
Suspected carrier
TCF12 carrier state, no craniofacial signs
Normal TCF12 genotype
DNA not available
Mutant TCF12 allele
Partner TCF12 allele in the presence of craniosynostosis
Partner TCF12 allele in non-penetrant individual
Bystander TCF12 allele (not transmitted)
Haplotype assignment not possible

Figure 6.7 Haplotyping results of all 48 family pedigrees (key to symbols shown in last panel). Mutant *TCF12* alleles (red) were found to be present with a partner alleles either in the presence (pink) or absence (green) of craniosynostosis. Haplotypes in grey could not be assigned correctly due to lack of other family members to make analysis informative. A maternal mosaic mutation in Family #28 is denoted by **m** in individual I-2. Segregation of haplotypes in Family #35 indicate the mutation arose *de novo* (*d*) in the maternal grandmother (I-2). Of note, individuals in Families #11 (II-1 and II-3) #35 (III-1 and III-2) and #45 (II-1 and II-6) share the same haplotype but show phenotypic variability, being either penetrant or non-penetrant for coronal synostosis. In Family #34, individuals with the same mutant *TCF12* allele inherited from the father (TTTAT), inherit opposite partner alleles from their unaffected mother (III-2, CGCAC; III-3, CGCCC), but both have coronal synostosis. The paternal origin of the mutant allele in informative proband-parent trios where the *TCF12* mutation arose *de novo* is denoted by *. Square brackets [ ], indicate assignment of haplotype is uncertain as genotyping individuals for this SNP did not enable further refinement of haplotype.

6.5 Haplotyping of samples from mutation-positive individuals and family members

The frequencies for each allele was plotted according to the ancestral haplotype (i.e. that of the most recent common ancestor, 'TTTAT') and tested for statistical significance using Fisher's exact test (see figure 6.8 below). The ancestral haplotype is determined by comparison with primate (chimpanzee) DNA using multiple sequence alignments (Paten et al., 2008; Park, 2015). Proportions of each SNP that was genotyped and statistically analysed are given in Tables 20-26.2, Appendix.

Initially, family trios in which the *TCF12* mutation had arisen as a *de novo* mutation were haplotyped. Information from the parental origin investigation described in Section 5.5 was then used to refine the assignment of haplotypes and to delineate mutant and partner *TCF12* alleles (in presence of craniosynostosis) in the proband only to acquire further data points (Families #5, #7, #21, #29, #32, #41). Haplotyping of further families acquired from other diagnostic centres increased the sample size further (see Table 5.1). The combined results are shown in Figure 6.8 below. Those trios that were uninformative, were not included in the initial analysis of haplotypes, but were considered separately (see Figure 6.9).

Due to the phenotypic variability seen in individuals with mutant *TCF12* alleles, being either penetrant or non-penetrant for coronal synostosis, it is hypothesised that possession of certain wild-type *TCF12* alleles may be protective against craniosynostosis. This may also be consistent with a derived haplotype ('CGCCC'). Plotting the allele frequencies for the HapMap CEU population (continental Europe) and control plates (UK population) according to the ancestral haplotype ('TTTAT') gave similar results (see Figure 6.8). Adding in frequencies of the partner allele observed in the presence of craniosynostosis was shown to be similar to the control background.

The mutant *TCF12* allele was shown to have a bias away from any potential protective partner allele, toward the ancestral haplotype and

was consistent with the hypothesis. There could be a possible dominant negative effect with the mutant *TCF12* allele reducing the expression of the normal protein. Unexpectedly, the only significant bias was seen for the non-transmitted bystander *TCF12* allele with statistically significant results for 1 tag-SNP. Although this allele was included initially to account for subtle differences in ethnicity of cases versus controls, thereby acting as an internal control, this particular allele showed the biggest difference against either control population analysed.

There were no observed instances of a mutant allele and a bystander allele possibly because this may be non-viable. To investigate the possibility of a non-viable haplotype further (i.e. mutant *TCF12* allele with non-protective bystander allele), clinical records for all probands and their families were re-analysed specifically to identify a history of recurrent miscarriages (see Table 2.1). Additionally, the parents of children with *TCF12* mutations were questioned at scheduled outpatient appointments following receipt of diagnostic results and appropriate genetic counselling on an *ad hoc* basis. However, no such instances were reported.

The proportions of *TCF12* alleles present in non-penetrant individuals were similar to background and as no results were statistically significant and so their possession may not be biologically relevant. Importantly, there were three instances where siblings from the same family who share the same haplotype were either penetrant or non-penetrant for

coronal synostosis (individuals II-1 and II-3, Family #11, individuals III-1 and III-2, Family #35 and individuals II-1 and II-6, Family #45). Additionally, there was one instance (Family #34) where individuals with the same paternally-inherited mutant *TCF12* allele, inherited opposite partner alleles from their unaffected mother (individuals III-2 and III-3), but both had bilateral coronal synostosis.

In a second analysis, the mutant and partner alleles for 6 probands from uninformative families (Families #6, #16, #26, #33, #38 and #42) were apportioned 50:50 on each parental chromosome as it could not be assumed they were all of paternal origin. However, the results for mutant and partner alleles in the presence of craniosynostosis were not statistically significant (see Figure 6.9).

Therefore, modifiers in the environment, other genes influencing *TCF12* expression or specific epigenetic regulators may have a more significant role in influencing the phenotypic variability seen in non-penetrant individuals.

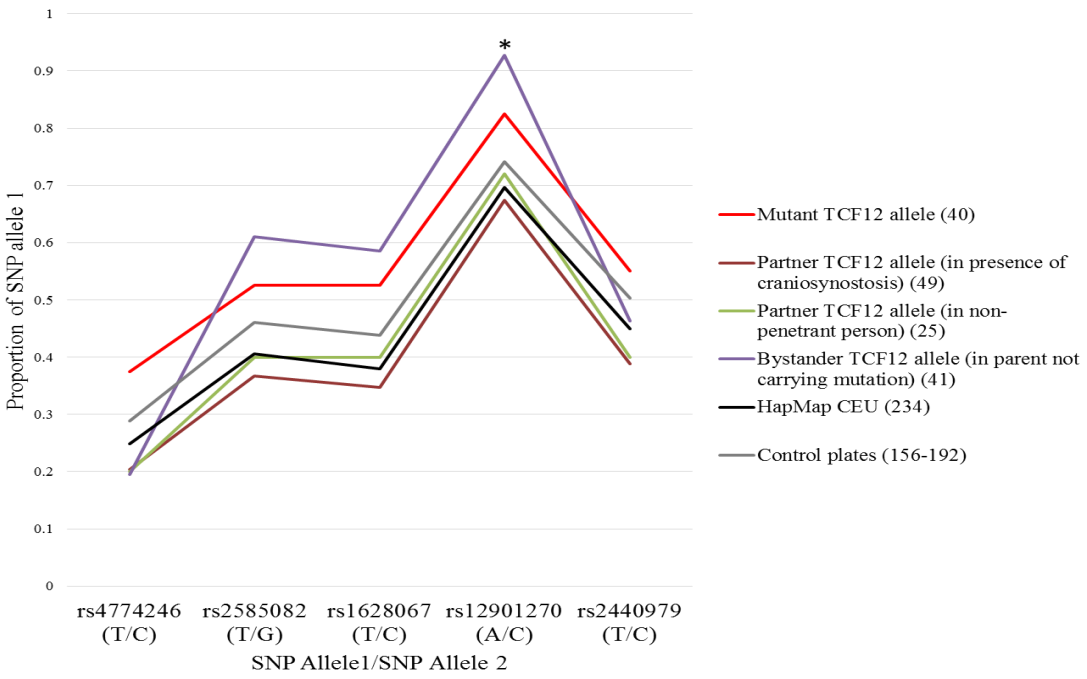


Figure 6.8 Frequencies of each *TCF12*-allele plotted according to the ancestral haplotype ('TTTAT') in 46 families (159 individuals). This analysis only includes informative *de novo* trios. Two control populations (grey and black lines) demonstrate relative proportions of each SNP in healthy individuals without coronal synostosis. Proportions of partner alleles in the presence of craniosynostosis show a bias toward the derived haplotype (brown line), going against the protective hypothesis. The partner allele in non-penetrant individuals (green line) is similar to background and no results attain statistical significance for any tag-SNP. The mutant *TCF12* allele (red line) shows a bias towards the ancestral haplotype, away from the protective partner allele, again consistent with the pre-disposition to disease hypothesis. Most striking, but also unexpected, is the bias in bystander allele frequency (purple line), which also attains statistical significance for rs12901270 (*$p<0.05$).

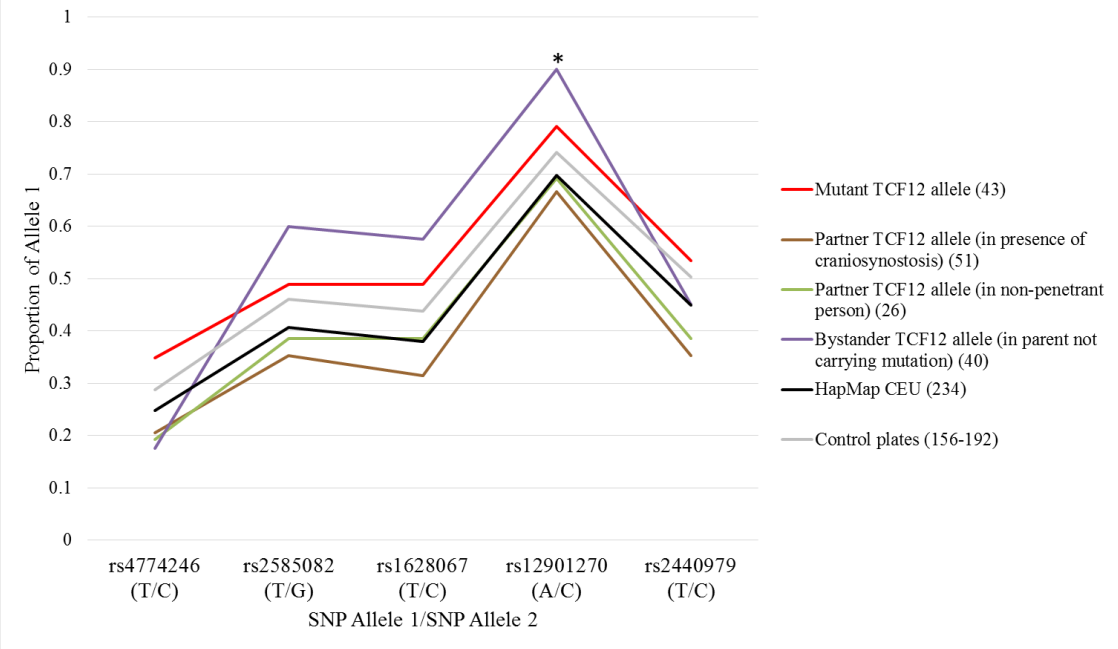


Figure 6.9 Frequencies of each *TCF12*-allele plotted according to the ancestral haplotype ('TTTAT') in 46 families (159 individuals) with 50:50 apportioning of mutant and partner alleles in 6 uninformative trios. No statistically significant difference in proportions of mutant *TCF12* alleles and those present in disease penetrant individuals was noted in the above analysis in uninformative *de novo* trios when each mutant allele was apportioned a 50% chance of being present on either chromosome. All other frequencies and proportions for other alleles are as per Figure 6.8.

As differences in gene expression may causally account for and correlate with differences in allele frequency observed for mutant, partner and bystander alleles, an existing dataset by Cheung et al (2010) was mined for information. This group utilised SNP genotyping information derived from 45 CEPH pedigrees from The HapMap database and measured gene expression in B-cells using microarrays on a genome-wide basis. Data of gene expression for 4 out of 5 tag-SNPs was obtained and is shown below in Figure 6.10. Comparison of low, intermediate and high

expressing genotypes revealed statistically significant differences for 2 tag-SNPs (rs2585082 and rs12901270) for high expressing alleles consistent with the derived haplotype (-GC-) found on protective partner chromosomes (see Tables 33.1-33.4, Appendix). By comparison, mutant *TCF12* alleles have a weak preference for lower expressing alleles consistent with ancestral haplotype (-TTA-). There is also the possibility that specific combinations of alleles, such as bystander (with a significant bias in frequencies as seen in figure 6.8) with mutant *TCF12* allele may be non-viable and were not observed in the 46 families that were haplotyped.

Comparing highest and lowest expressing genotypes (G/G and T/T respectively) at SNP rs2585082 using a two-tailed *t*-test gave a statistically significant P-value of 0.012 ($p < 0.05$). For rs12901270, comparing lowest (A/A) and highest (C/C) expressing genotypes also gave significant results (*P*-value = 0.014; $p<0.05$). For rs4774246 and rs1628067, comparison of lowest and highest expression genotypes was not statistically significant at the 95% confidence interval (*P*-values of 0.052 and 0.051 respectively). This is summarised in Figure 6.10.

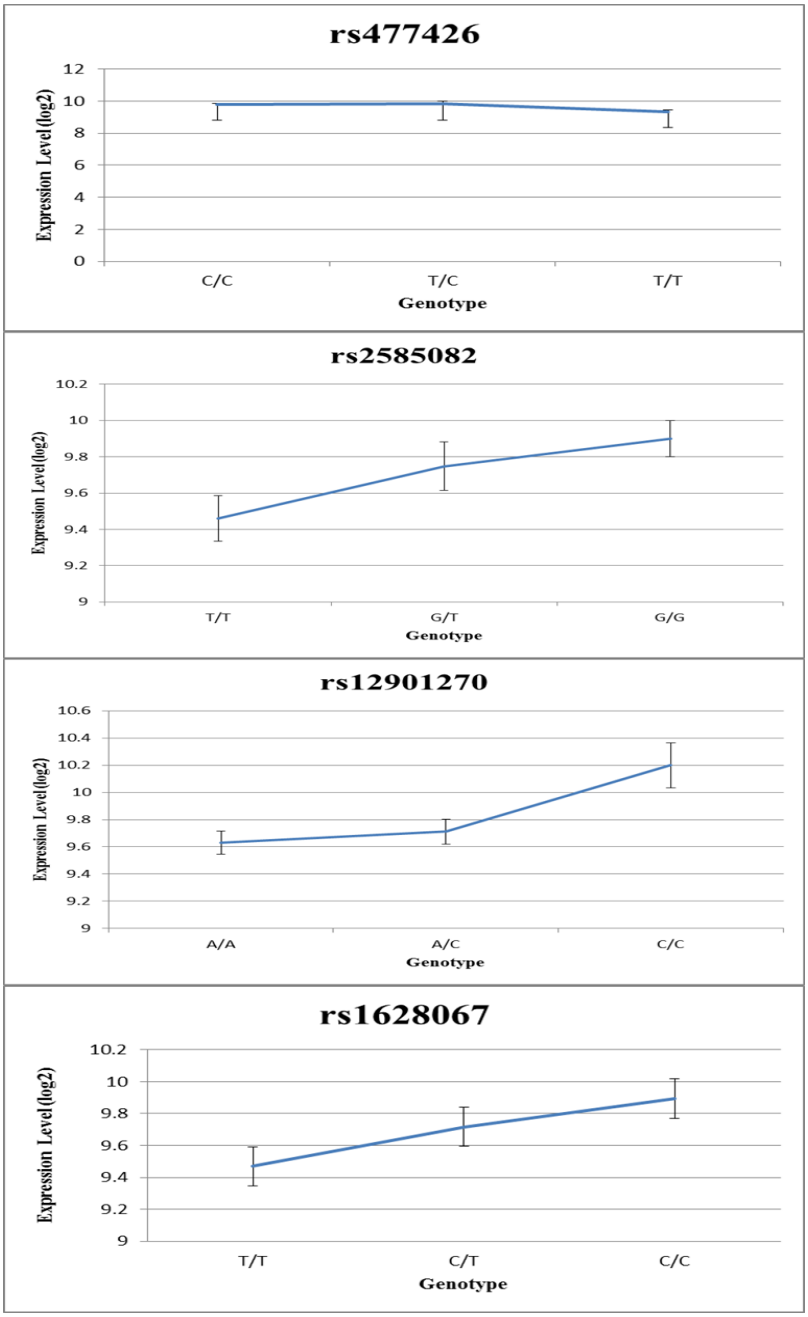


Figure 6.10 *TCF12* expression vs genotype. Analysis on a per-genotype basis revealed statistically significant differences (mean ±SEM) between high and low expressing genotypes for 2 tag-SNPs (i.e. G/G vs T/T for rs2585082 and C/C vs A/A for rs12901270) using a *t-test*. Data were derived from measurements of gene expression on genome-wide scale in B cells performed by Cheung et al (2000) using a CEPH control population of 45 individuals**.**

6.6 Utilising a transmission disequilibrium test (TDT) to look for linkage between *TCF12* mutations and tag-SNPs

A transmission disequilibrium test was used to demonstrate linkage (in the presence of an association) between a disease locus (i.e. a *TCF12* mutation) and a marker (a tag-SNP) from a non-carrier heterozygous parent as originally demonstrated by Spielman et al. (1993). This revealed statistically significant results with over transmission of the G-allele of rs2585082 [T/**G**], the C-allele of rs1628067 [T/**C**] and the C-allele of rs12901270 [A/**C**] (see Table 6.4).

SNP	rs4774246	[T/C]	rs2585082	[T/G]	rs1628067	[T/C]	rs12901270 [A/C]		rs2440979 [T/C]	
Family (#) where each SNP allele is inherited from heterozygous non-carrier parent	**T**	**C**	**T**	**G**	**T**	**C**	**A**	**C**	**T**	**C**
	1	11 (x2)	3 (x2)	8	3(x2)	8	9	9	3(x2)	8
	5	14 (x2)	7	9	9	9	39	11(x2)	9	9
	7	19	9	11 (x2)	10	11(x2)	48	18	11	14 (x2)
	10	23	10	14 (x2)	34	14 (x2)		19 (x2)	21	15
	23	27	47	15	37	15		22	37	18
	35 (x2)	28		18	47	18		23	44	23
	36	29		19 (x2)		19 (x2)		28	45 (x2)	28
				23		23		29	47	29
				27		27		34		35 (x2)
				28		28		35		
				29		29		37		
				34		34		41		
				35		35		46		
				37		45 (x2)				
				45 (x2)		46				
				46						
Total instances of transmission	8	9	6	20	7	19	3	15	10	11
Chi-squared		0.0588		7.538		5.538		8		0.0476
p-value (1df)		0.8084		0.0060		0.0186		0.0047		0.8273
Statistically Significant? (p<0.05)		No		**Yes**		**Yes**		**Yes**		No

Table 6.3 Transmission Disequilibrium Test for 46 families. Testing for linkage of tag-SNPs inherited from heterozygous non-carrier parents of *TCF12* mutation-positive individuals. There were statistically significant results for rs2585082, rs1628067 and rs12901270 as well as for the differences in high and low expressing genotypes (Figure 6.10). This demonstrates a significant association between *TCF12* haplotype and specific alleles.

6.7 Investigation into association of *TCF12* haplotype in mutation-negative patients with non-syndromic coronal synostosis

In order to analyse the effects of a possible association between *TCF12* haplotype and craniosynostosis, a panel of 244 patients with mixed patterns of suture fusion who were mutation-negative was genotyped with tag-SNP rs12901270 [A/C]. This particular SNP demonstrated the greatest proportions that were consistent with the ancestral haplotype and therefore will have been subject to less recombination than others in the set. In particular, the presence of an association between *TCF12* genotype and non-syndromic coronal synostosis was sought. Of this population, 17 patients did not have a clear diagnosis (i.e. referred for analysis with 'craniosynostosis unspecified'), therefore were excluded from further analysis. This left 227 patients eligible for analysis. Fisher's exact test was performed to test for the presence of a statistical association between *TCF12* genotype and non-syndromic coronal synostosis, both as isolated patterns, or combinations with other sutures. Comparison was made with genotyping results from the UK control population described in Section 6.3. Additionally, all cases of midline, lambdoid and multisuture fusion were analysed for association with rs12901270. However, no statistically significant association was seen in any case (see Table 6.4).

	Mutation-negative patients				Total number of each SNP allele [A/C]		UK Control population			
Pattern of suture fusion	AA	AC	CC	TOTAL	TOTAL A	TOTAL C	TOTAL A	TOTAL C	Fisher's exact test	Statistically sig?
BC (n=23)	12	9	2	23	33	13	274	96	0.725	No
RC (n=46)	24	19	3	46	67	25	274	96	0.793	No
LC (n=36)	20	14	2	36	54	18	274	96	1	No
UC (n=2)	1	1	0	2	3	1	274	96	1	No
Coronal – all other combinations (n=16)	9	7	0	16	25	7	274	96	1	No
S (n=32)	14	17	1	32	45	19	274	96	0.541	No
M (n=44)	22	17	5	44	61	27	274	96	0.422	No
Midline – all other combinations (n=5)	3	1	1	5	7	3	274	96	0.341	No
Lambdoid – all combinations (n=13)	9	3	1	13	21	5	274	96	0.6416	No
Multisuture (n=10)	8	2	0	10	18	2	274	96	0.182	No
TOTAL	122	90	15	227	597	223	274	96	0.672	No

Table 6.4 Testing association between *TCF12* genotype and non-syndromic coronal and other suture synostosis in 227 mutation-negative patients. No statistically significant relationship was found between *TCF12* genotype (using rs12901270) and the craniosynostosis patients in this cohort. 'Coronal – all other combinations' is the total number of isolated coronal synostosis cases as enumerated and combinations of each coronal pattern of fusion plus sagittal, metopic or lambdoid. 'Midline – all combinations' includes isolated sagittal or metopic suture fusion in addition to cases of sagittal and metopic suture fusion. 'Lambdoid – all combinations' includes isolated patterns of lambdoid suture fusion plus combinations with sagittal suture fusion. These results have been combined as individual numbers were very low. 'Multisuture' refers to patients with 3 or more fused sutures. KEY: BC = bilateral coronal; RC = right coronal; LC = left coronal; BC, S = bilateral coronal and sagittal; RC, S = right coronal and sagittal; LC, S = left coronal and sagittal; S = sagittal; UC = unilateral coronal (side not specified). Note: The following patterns of suture fusion were not observed in this cohort: RC, M; LC, M; UC, M; BC, RL; RC, BL; RC, LL; LC, BL; LC, RL; LC, LL; UC, BL; UC, RL; UC, LL; M, S; S, RL; S, LL; M, BL; M, RL; M, LL.

6.8 Discussion

The lack of a definitive phenotype in individuals positive for *TCF12* mutations presents a significant pitfall in clinical assessment, particularly from the genetic counselling standpoint. A parent who carries the mutation may be given incorrect estimates of recurrence risk in future offspring after the birth of an affected child if they are non-penetrant for craniosynostosis. As described in Section 5.3, a significant non-penetrance rate of 60% was observed in the cohort of 48 families with commonly measured cephalometric indices falling within normal limits. It was therefore hypothesised that the genetic background of non-penetrant *TCF12* mutation carriers may account for the phenotypic variability that was observed and this may also be related to intrinsic differences in *TCF12* expression. The importance of haplotype in disease has already been established from conditions such as Hirschsprung disease (Emison et al., 2005; 2010) and thrombocytopenia with absent radii (Albers et al., 2012).

To obtain normative data for *TCF12* haplotypes in healthy individuals, phased haplotype data for a control population ethnically matched to the families under investigation was obtained from The International HapMap Project. A control population from the UK was also genotyped so the data more accurately reflected the patient population compared to continental Europe from which the Caucasian HapMap data are derived. Use of a

specific set of tagging SNPs facilitated an efficient method to capture the allelic diversity seen in *TCF12*, with a high mean r^2 (0.917) seen across all SNP alleles captured.

Construction of *TCF12* disease haplotypes revealed several novel findings. Firstly, the presence of a non-transmitted bystander allele demonstrated a significant bias toward the ancestral haplotype. This represents a new scientific phenomenon that may require further investigation. Secondly, the mutant *TCF12* allele also showed a bias toward the ancestral haplotype. The presence of a bystander allele was not seen with a mutant *TCF12* allele in any individual, possibly because this would be non-viable. However, no clinical evidence (e.g. recurrent miscarriage history) was found to corroborate this, although this will need to take place on a wider basis across all families surveyed in this thesis Thirdly, the hypothesis that certain alleles present in non-penetrant individuals may be protective from coronal synostosis developing despite the possession of a mutant allele, was not borne out by the data. No significant difference was seen between proportions of *TCF12* alleles in non-penetrant individual and controls and no data points reached statistical significance.

Allele frequencies for partner alleles observed in the presence of craniosynostosis were similar to background. Analysis of data from gene expression studies in lymphoblastoid cells (Cheung et al., 2010) revealed

statistically significant differences for 2 tag-SNPs (rs2585082 and rs12901270) between low and high expressing genotypes in control individuals. This evidence supports the theory that possession of a specific high-expressing normal *TCF12* allele would compensate for the possession of a mutant *TCF12* allele, thus preventing the development of coronal synostosis. This in turn may explain some of the phenotypic variability shown in Figure 6.3. Differential allelic expression is common, occurring in more than 50% of autosomal dominant disorders. To prevent the development of disease, greater expression of wild-type over mutant alleles may be favoured (Cheung & Spielman, 2009). Additionally, a transmission disequilibrium test looking for linkage between the transmission of SNPs from a non-carrier parent in the presence of an association with coronal synostosis, revealed statistically significant results for the same 3 tag-SNPs, with over-transmission of G, C, C alleles.

Haplotyping 159 individuals from 46 families whose members were mutation positive only revealed statistically significant results for the bystander allele for one SNP, rs12901270. Inheritance of opposite *TCF12* alleles does not explain the observed phenotypic variability as noted in a number of cases (Family #34) and the possession of the same haplotype demonstrated a discordance in phenotype. In these three families (Families #11, #35 and #45), children inherited identical mutant *TCF12* alleles and

the same wild-type *TCF12* allele from an unaffected parent, however one sibling developed coronal synostosis and the other did not.

No bias in allele frequencies toward the derived haplotype was observed in non-penetrant individuals and thus may not be biologically relevant. Possession of certain *TCF12* haplotypes could still offer a slight survival advantage, outside the context of craniosynostosis. For example carriers of variants in factor V Leiden demonstrate improved survival in severe sepsis although the primary defect predisposes the individual to a higher risk of venous thrombosis and miscarriage (Kerlin et al., 2003).

Taken together, this implies that *TCF12* haplotype may be less significant than initially thought and other factors in the environment as well as genetic and epigenetic modifiers may exert a greater effect. Overall, the limited correlation of phenotype with *TCF12* haplotype in mutation-positive individuals abrogates its further use in clinical testing.

CHAPTER 7

SEARCH FOR FURTHER CANDIDATE GENES BY EXOME SEQUENCING OF TRIOS AND FURTHER PROBANDS

7.1 Introduction

The discovery of *TCF12* as a bona fide disease gene that was associated with coronal synostosis (Chapter 3) provided an overwhelming proof-of-concept that applying whole-exome sequencing and bioinformatic techniques to previously mutation-negative patients would yield clinically meaningful results. Building on this initial success, further applications of this methodology were used in an attempt to discover further novel disease genes. Additional refinements in reagents and bioinformatic techniques were also implemented over the next phases of this project. As before, genes falling into the categories of transcription factors and growth factor receptors would be prioritised along with those variants exhibiting the most damaging types of mutation (i.e. frameshift, nonsense and splicing) in those groups.

7.2 Patients, materials and methods

Whole-exome sequencing was performed on the genomic DNA of a further 7 patients from the stratified cohort described in Table 2.1, again prepared using the Agilent protocol as before (see Section 2.9). In this second round, the updated SureSelect Human All Exon Kit (v.4; 51 Mb) was used as it reportedly improved coverage of more exons, untranslated regions and read depth uniformity. The bioinformatic pipeline described in Section 3.2 was used and the metrics for this second round of exome sequencing are

summarised in Table 7.1 below. Refinements to the pipeline by Dr S McGowan (Computational Biology Research Group, Weatherall Institute of Molecular Medicine, University of Oxford) included use of Novoalign rather than Bowtie to carry out alignment of the raw data to the human genome due to its superior ability to align reads that contain indels as described in Section 3.5. Additionally, an updated version of dbSNP (dbSNP136 vs dbSNP134) was now used and Steps 3 and 5 were combined with the removal of PCR artefacts (Step 4) performed later in the bioinformatic filtering process.

As described in 1.6.3, another strategy for novel disease gene discovery is to exome sequence patient-parent trios and exclude all inherited variants to ideally yield approximately 0-3 pathogenic variants (Vissers et al. 2010; O'Roak et al. 2011). The parents of 3 patients exome sequenced in Rounds 1 and 2 (OX1662, OX1560, OX3245) along with 2 new patient-parent trios (BCH4972, OX6041) were exome sequenced as described above. Clinical characteristics of each parent is given in Table 2.2. Inherited variants were excluded by removing those present in gene lists that were shared between proband and either parent. Of the remaining apparently *de novo* variants present in the proband only, each was analysed as before (see 3.3) for read depth, quality of sequence surrounding change, and roles in known craniosynostosis or similar disease pathways. These were then validated using PCR and dideoxy sequencing or digestion with an appropriate restriction enzyme.

Finally, 19 mutation-negative patients with coronal synostosis were analysed using the previously successful overlap strategy, looking for variants shared between 2 or 3 patients. This larger cohort included 11 from both rounds of exome sequencing and 8 additional mutation-negative patients from parallel sequencing projects (see Table 2.1 for full clinical details). The results were also intersected with 11 lists of genes thought to be involved or related to various craniosynostosis disease pathways, transcription factors, or those noted not to possess one inactivated allele listed in a large public exome database. This latter resource contained data from 6503 individuals comprised of controls, ischaemic heart disease and lung disease but none with craniosynostosis or any other craniofacial malformations.

Sequence data processing in Round 2 of Exome Sequencing						Number of Reads						
						OX3245	OX3243	OX5098	OX1675	OX4462	OX1799	OX2438
1	Input	Raw data obtained from 100 bp paired-end sequencing				32790039 (x 2)	33684407 (x 2)	30897680 (x 2)	33482643 (x 2)	33153860 (x2)	32690900 (x 2)	33067942 (x2)
2	MAP_SAM_OUTPUT: Novoalign	Map to whole genome – all hits recorded				66224851	68255471	62584554	67926847	67762508	66601852	67165986
3	FIND_BEST_HITS: find_best_hits.pl	Remove pseudogenes (if a read pair maps to >1 location only those which have the lowest number of mismatches over both ends and which map on or near an exome capture bait are retained)				64425036	65129765	59928727	64289224	64219167	60168208	62353867
4	GET_UNIQUE_PAIRS:collapse_pairs.pl	Remove PCR artefacts (only paired-reads with highest total Phred score are retained)				50137750	58178844	55125681	53342792	58551415	49046805	55773202
Variant analysis												
Actual read depth at 10x (% of exons)						80.6	81.6	81.4	81.1	81.4	80.4	81.5
Actual read depth at 30x (% of exons)						59.4	64.2	62.9	61.1	62.4	58.8	63.2
Total number of variants						176828	182916	182388	182176	251884	174293	189852
Number of exonic variants						43434	43999	44485	43411	43850	43356	43908
dbSNP136 filtering						667	590	1227	699	552	600	584
Shared gene variants – Round 2												
Number of patients sharing gene variants						-	7	6	5	4	3	2
Number of genes						-	0	0	0	0	8	58
Combined shared gene variants – Round 2 + *TCF12*-negative Round 1												
Number of patients sharing gene variants		11	10	9		8	7	6	5	4	3	2
Number of genes		0	0	0		0	0	0	0	0	17	110

Table 7.1 Summary of statistics for exome sequencing of 7 further patients with bilateral coronal synostosis (Round 2). Combined results of analysis of 11 patients also shown. See Table 3.1 for statistics from Round 1 and Table 28, Appendix for full combined candidate gene list.

7.3 Gene lists and candidate shortlisting identifies a frameshift mutation in a novel gene accounting for autism but not coronal synostosis observed in one patient

The data from 7 patients in Round 2 of exome sequencing was combined with that of 4 mutation-negative patients in Round 1, producing a list of 138 candidate genes post-bioinformatic filtering for these 11 patients. Again, after excluding variants present in all 11, 10, 9, 8, 7, 6, 5, or 4 patients, the list was narrowed to 127 genes (17 genes with variants in 3 patients and 110 genes with variants in 2 patients – see Table 28, Appendix for full list). However, despite using the same overlap strategy of searching for shared variants, there was not a second gene discovery as significant as *TCF12*.

Analysing gene variants only found in single patients however, revealed a frameshift mutation in *TCF20* (Transcription factor 20) on chromosome 22q13.3, in a proband with right-sided unilateral coronal synostosis, autism, moderate intellectual disability, whose mother also had left-sided unilateral coronal synostosis. This finding was validated by dideoxy sequencing with the mutation confirmed in the child only when both she and her mother were tested. Testing of a paternal sample confirmed this mutation had arisen *de* novo with correct sample relationships confirmed using a panel of microsatellites (analysis by Dr D Lloyd). These findings are summarised in Figure 7.1.

This gene co-regulates transcription and is related very closely to *RAI1* (Retinoic Acid Induced 1), mutations of which cause Smith-Magenis Syndrome (Slager et al., 2003). This is a developmental disorder that includes intellectual disability, speech and language delay and has a distinctive facial appearance, but no craniosynostosis. Other collaborators independently identified a pericentric inversion of chromosome 22 whose breakpoint includes *TCF20* in another patient with multi-suture craniosynostosis. However, as the *TCF20* mutation was *de novo* and therefore did not segregate with the coronal synostosis phenotype in the mother, this gene is not thought to cause coronal synostosis, but is associated with autistic spectrum disorder (Babbs et al., 2014).

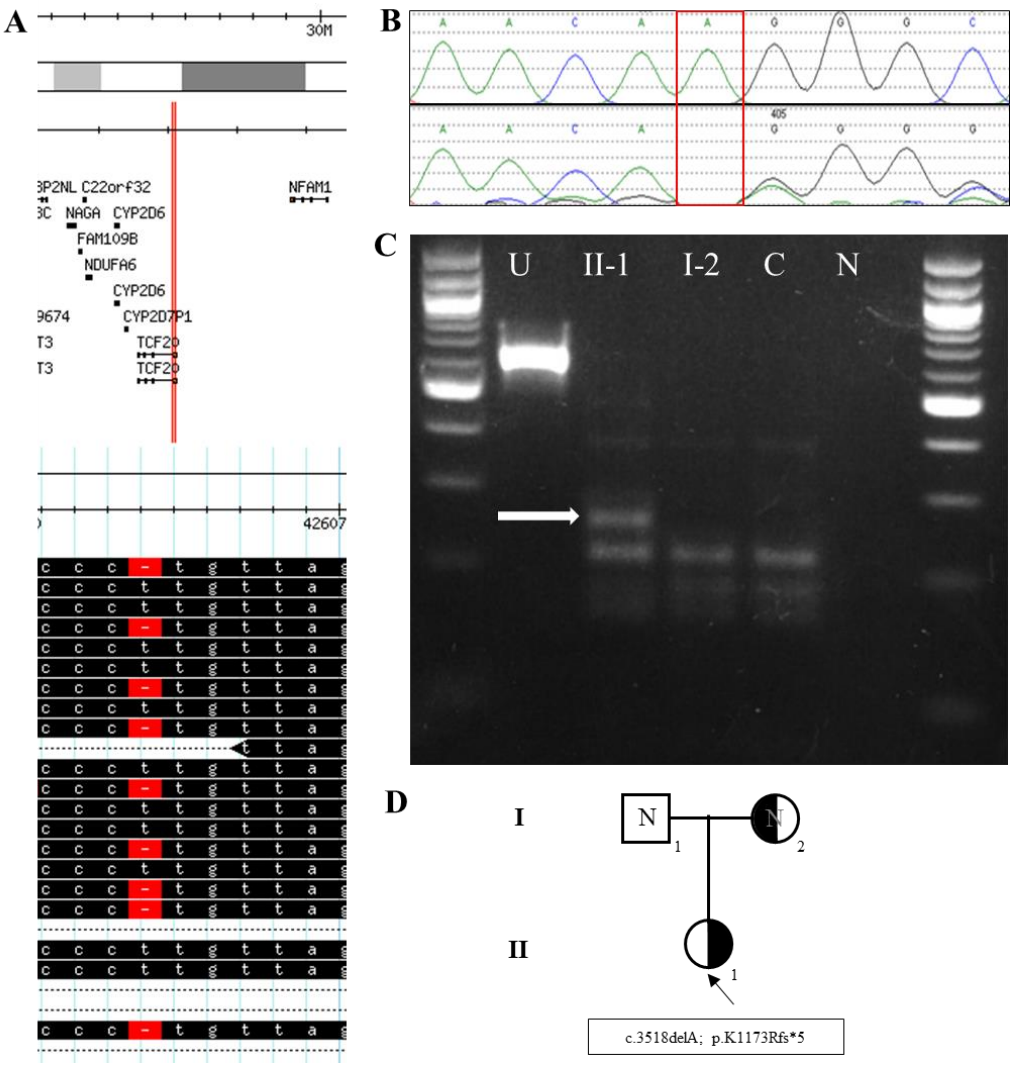


Figure 7.1 Identification of a frameshift mutation in *TCF20* (c.3518delA; p.K1173Rfs*5) by whole-exome sequencing in patient OX2438. A: Visualisation of the mutation using a web-based genome browser. Normally mapped reads to the human genome are in black, with variants indicated in red **B:** Validation of findings using dideoxy sequencing (red box) confirming single base deletion in proband (bottom trace) versus control (top trace). **C:** Restriction digest with BslI demonstrating presence of mutant band in proband (II-1) only (white arrow). Amplicon = 595 bp, wild-type = 215bp, 162bp, 145 bp, 72 bp, 1 bp; mutant = 233 bp, 215 bp, 146 bp, 1 bp. Run with 1 kb DNA ladder on 3%

agarose gel (see Table 38 Appendix for primers and conditions). **D:** Later testing of a paternal sample (by D Lloyd) confirmed mutation had arisen *de novo* in the proband, adding further evidence of pathogenicity of this variant (see Table 29, Appendix for primers and conditions).

7.4 Whole-exome sequencing of 5 proband-parent trios identifies a *de novo* mutation in a gene encoding a muscle protein of unknown significance in craniofacial development

After exclusion of shared variants in the 5 trios that were exome sequenced (OX1662, OX1560, OX3245, BCH4972, and OX6041), the few variants that remained were further analysed as appropriate and are summarised in Table 7.2 below. Following analysis of the top 12 variants seen within the 5 trios, it became apparent that there was a significant issue with false reads as only 4 variants identified by exome sequencing could be validated by dideoxy sequencing. Of those 4 variants, 3 were actually inherited from a parent and may have been missed initially when viewed on the genome browser due to poor read depth in either parents (*TAF1B*, paternal, OX1662; *SPG20*, maternal, OX1560; *LAPTM4B*, maternal, OX6041; see Table 7.4). The only true *de novo* mutation identified by using this trio strategy was in *MYH4* (Myosin Heavy Chain 4; c.4691G>A; p.R1564H) that encodes a myosin motor protein involved in the contraction of skeletal muscle. Mice that are homozygous for the T-allele of an intronic SNP [C/T] in this gene have been proven to have a significant reduction in the muscle

mass of their hind limbs. Possession of the TT genotype demonstrated a 1:1 concordance between genotype and the mini-muscle phenotype in those mice (Kelly et al., 2013). Knockout of *Fgfrl1* in mice with diaphragmatic defects show reduced expression levels of *Myh4* (LopezJimenez et al., 2010), but no association with face or skull development has yet been reported. The implications of this *de novo* mutation are therefore uncertain and a summary of the analysis shown in Figure 7.2.

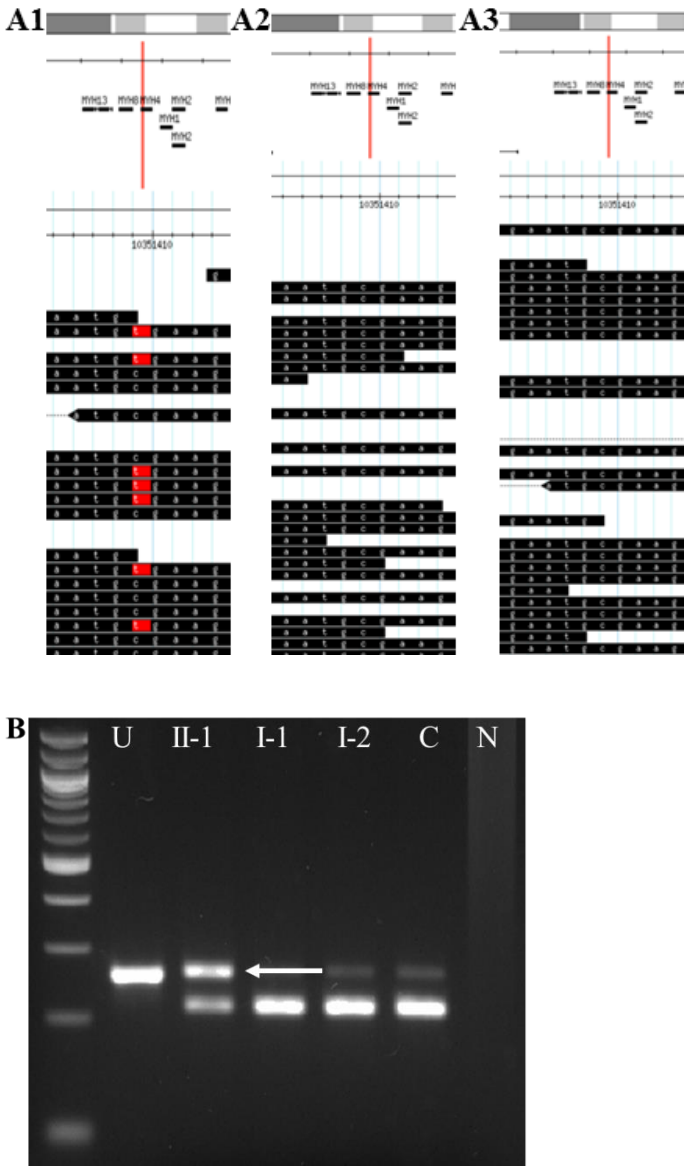


Figure 7.2 Discovery of a *de novo* variant in *MYH4* following exome sequencing of a proband-parent trio. **A1-A3:** The individual sequences of the proband (A1), mother (A2) and father (A3) are shown with the variant in red and normally mapped sequences in black. Absence of the c.4691G>A; p.R1564H variant in the parents indicates this has arisen *de novo* in the child. **B:** Restriction digest with BsmI (amplicon = 259 bp) confirms presence of the mutation in the proband only (white arrow), validating the findings shown in A. Fragment sizes: wild-type = 215 bp, 44 bp; mutant = 259 bp. Samples run on 3% agarose

gel with 1kb DNA ladder (for primers and conditions, see Table 30, Appendix). KEY: U = uncut proband DNA, II-1 = proband, I-1= mother, I-2 = father, C = unaffected control, N = water.

Finally, an in-house craniosynostosis database, 'CranioVAR', was used to conduct a trio analysis *in silico* of variants to identify further *de novo* mutations of the 5 trios described (see Figure 7.3, developed by Dr S McGowan, Computational Biology Research Group, Weatherall Institute of Molecular Medicine). This was performed both including and excluding variants found in dbSNP136 and the data cross-referenced with 11 lists of genes (see Section 7.5 for details). However, no additional *de novo* variants were identified by this method.

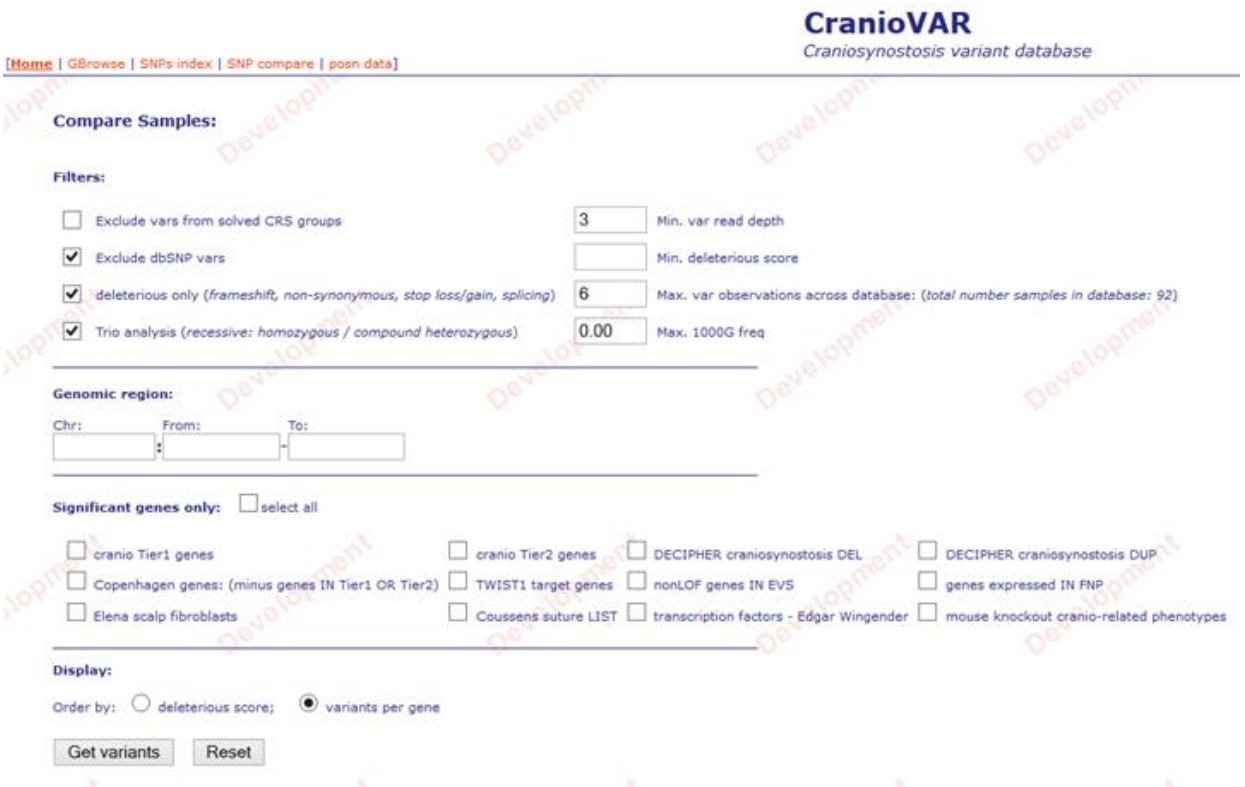


Figure 7.3 Use of an in-house craniosynostosis variant database ('CranioVAR') to identify possible *de novo* mutations in the 5 exome sequenced trios. This analysis was performed including and excluding variants present in dbSNP136 and by varying the read depth for each variant call, however no further *de novo* mutations were identified.

	Variant analysis														
Sequence data processing of 5 patient-proband trios following exome sequencing	***1662*** ***P***	**1664** **M**	**1663** **F**	***1560*** ***P***	**1590** **M**	**1591** **F**	***3245*** ***P***	**3374** **M**	**3375** **F**	***5656*** ***P***	**4972** **M**	**4973** **F**	***6041*** ***P***	**6040** **M**	**6039** **F**
Actual read depth at 10x (% of exons)	*68.7*	88.3	48.3	*66.6*	53.5	81.7	*80.6*	83.9	80.8	88.9	81.1	82.3	84.4	83.7	84
Actual read depth at 30x (% of exons)	*44.1*	66.9	34.6	*35.7*	34.2	65.6	*59.4*	66.8	60.3	69	61.9	66.9	69.8	66.8	67.2
Total number of variants	*408*	656	1125	*475*	548	801	*469*	502	496	495	512	509	856	909	857
Number of exonic variants	*276*	525	955	*344*	420	600	*374*	404	384	392	406	402	667	729	671
dbSNP138 filtering	*41*	213	737	*75*	147	350	*84*	78	74	93	97	134	218	240	212

Table 7.2 Metrics of exome sequencing 10 parents of 5 mutation negative patients (proband data also shown italicised for comparison). Note lower read depth seen in all parents. KEY: P = proband, M = mother, F = father.

	Variant analysis					
Sequence data processing for additional mutation-negative probands with coronal synostosis following exome sequencing	**GOS6568**	**OX4400**	**OX4580**	**OX4809**	**BCH6136**	**OX5569**
Actual read depth at 10x (% of exons)	82.5	98.5	98.6	98.8	82.2	86.0
Actual read depth at 30x (% of exons)	75.9	27.8	27.9	49.2	75.0	57.4
Total number of variants	404	898	1046	670	441	536
Number of exonic variants	325	566	817	519	348	439
dbSNP138 filtering	53	189	222	79	70	110

Table 7.3 Metrics of exome sequencing 6 additional mutation-negative patients with coronal synostosis. These patients then formed part of a larger cohort of 19 patients that were analysed using the overlap strategy and intersected with information from 11 gene lists (see table 7.5). Note the data for patients OX4400, OX4580, OX4809 is derived from a parallel project by other laboratory members (Dr S Twigg, A Fenwick) that used whole-genome sequencing.

Patients affected	Gene	Cytogenetic Locus	SNV/Nucleotide change/Amino acid change	High quality forward and reverse reads (wild-type | mutant)	Consensus Deleterious Score	PhyloP score/Prediction (C= constrained/ A = accelerated)	SIFT score/Prediction (D = damaging/ T=tolerated)	PolyPhen2 score/Prediction (B = benign, P = possibly damaging/D = probably damaging)	LRT score/prediction (U= unknown /D = deleterious /N= neutral)	Mutation Taster score/prediction (A/D= disease causing/ N=polymorphism)	GERP++ score [(-12.3 (not conserved) to 6.17 (most conserved)]	Gene Function (from genecards.com)	Exome variant validated?	True *de novo* variant?
OX1662	*HS6ST1* - Heparan Sulfate 6-O-Sulfo-transferase 1	2q21	C>T; c.553C>T; p.R185X	2, 8 | 0, 6	6	0.982013/C	0.902431 / n/a	0.732602/ n/a	1/D	1/D	1.72	n/a	No	No (artefact)
	TAF1B - TAF1B TATA box binding protein (TBP)-associated factor, RNA polymerase I, B	2p25	c.1411dupA; p.S471fs	1,1 | 1, 0	6							Key role in transcription initiation. SL1/TAF1B form complexes with Runx2, forming functional linkage Runx2 + control of rRNA synthesis,	Yes	No (present in father)
	BHLHE22 (basic helix-loop-helix family, member e22)	8q12	c.703_704insGCA;p.K235delinsSK	1,0 | 2, 0	0							n/a	No	No (artefact)

OX1560	*SPG20* Spastic Paraplegia 20	13q13	c.1793duC; p.G598fs	0,18 | 1, 13	6							Encodes MIT (Microtubule Interacting and Trafficking molecule). Altered BMPR trafficking in Spg20$^{-/-}$ efibroblasts.	Yes	No (variant present in mother)
	TSPAN7 Tetraspanin 7	Xp11.4	C>A; c.705C>A; p.C235X	9, 3 | 2, 1	4	0.998924/C	0.903782 / N/A	0.735383/ N/A	1/D	1/D	5.29	n/a	No	No (artefact)
OX3245	*CDH8* Cadherin-8	16q21	splicing	2, 22 | 0, 4	5							n/a	No	No (artefact)
	KCNQ2 Potassium voltage-gated channel, KQT-like subfamily, member 2	20q13	C>A; c.628C>A: p.R210S	10, 3 | 2, 2	5	0.998041/ C	1/D	1/D	0.999915/D	0.998435/ D	4	n/a	No	No (artefact)
BCH5656	*SP8* Specificity Protein 8	7p21	C>A c.860C>A: p.S287X	1, 1 | 2, 2	6	0.997562/C	0.903789 / N/A	0.733902/ N/A	0.992867/U	1/D	3.84	n/a	No	No (artefact)
	WNT3A Wingless-Type MMTV Integration Site Family, Member 3A	1q42	G>T; c.G568T: p.A190S	0, 8 | 0, 3	4	0.998329/C	1/D	0.977/D	1/D	0.246089/ N	4.14	n/a	No	No (artefact)

	ZIP12 *Zrt- And Irt-Like Protein 12*	10q12	C>A; c.C1614A:p.S538R	16, 5 | 2, 2	3	0.987324/C	0.67/T	0.481/P	0.999193/N	0.990863/D	2.62	n/a	No	No (artefact)
OX6041	*MYH4* Myosin, Heavy Chain 4, Skeletal Muscle	17p13	G>A; c.G4691A:p.R1564H	0, 23 | 3, 28	3	0.987178/C	1/D	0.006/B	0.999121/U	0.220184/N	4.78	Motor protein performing ATP hydrolysis, actin binding and potential for kinetic energy transduction. Down regulated in *Fgfrl1*$^{-/-}$ mice	**Yes**	**Yes**
	LAPTM4B Lysosome Associated Protein Transmembrane 4B	8q22	c.222_223 insGCTCCAGGCGAGGCGG:p.R74fs	7, 0 | 3, 3	6							Protein resides within the endo-lysosomal system. Roles in small molecule transport + membrane protrusions.	Yes	No (variant present in mother)

Table 7.4 Summary of exome patient-parent trio analysis. In addition to the mutation type, *in silico* analysis by various computer programmes (PhyloP, SIFT, Polyphen, LRT, Mutation taster, GERP+++) were used to predict the effects of potentially damaging variants (see 1.6.6 for description of each). Poor read depth and a high number of artefact variant calls were noted during this analysis (for primers and conditions, see Table 30, Appendix).

7.5 Analysis of gene lists of putative disease-causing variants from combined analysis of 19 mutation-negative patients utilising cohort/overlap strategy reveals several potential candidates

Due to the success of the overlap method utilised in Chapter 3, this methodology was repeated and applied to a larger cohort of 19 patients. These comprised 11 mutation-negative patients from both rounds of exome sequencing and 8 additional patients from parallel projects. All patients had coronal synostosis and negative genetic testing. The clinical information for this latter group is also described in Table 2.1 (Patients 15-23).

In addition, the data was cross-referenced or intersected with information from 11 gene lists (compiled by Dr S Twigg, Clinical Genetics Group, Weatherall Institute of Molecular Medicine). These lists contained variants implicated directly or indirectly in craniofacial expression or disease pathways. Additionally, they included genes where one allele was not inactivated based on information from public exome databases (see Table 34.1-34.3). The 'Craniosynostosis Tier 1 and 2' lists contained genes that have known mutations that are associated with craniosynostosis syndromes. Contained within the 'DECIPHER CRS DEL' list were chromosomal regions with copy number variations that overlapped with craniosynostosis phenotypes. This is from the DECIPHER database that contains genotype and phenotype information from 17663 patients (https://decipher.sanger.ac.uk/). The Copenhagen gene list contained genes

found in the DISEASES database excluding Craniosynostosis Tier 1 and 2 genes (http://diseases.jensenlab.org/About). *TWIST1* target genes were acquired from the published literature and non-loss-of-function genes in the Exome Variant Server were those noted not to have any inactivating variants in this database (http://evs.gs.washington.edu/EVS/). For those genes expressed in the frontonasal process (FNP) and other regions of the head in mice as well as mouse knockout craniosynostosis phenotypes, the Eurexpress large scale dataset contains information on the pre-natal mouse (from E14.5) derived from 18,000 coding genes and more than 400 microRNAs (http://discovery.lifemapsc.com/gene-expression-signals/high-throughput/ish-large-scale-dataset-eurexpress). The list of genes derived from the work by Bochukova et al., (2010) was obtained from the expression profile data of genes in scalp fibroblasts that were acquired from syndromic craniosynostosis patients undergoing craniofacial surgery. In a related manner, the prior work by Coussens et al., (2007) provided a gene list with genes that were differentially expressed in fused and unfused sutures in syndromic and non-syndromic children. Finally, a list of all known transcription factors from the Wingender group (http://www.edgar-wingender.de/huTF_classification.html) was also included as mutations in this group have been found in various craniosynostosis syndromes. In particular, those caused by mutations of *TWIST1* and *TCF12* are known to be particularly sensitive to dosages of their respective genes.

Those variants that were present in 1 or more of these lists were prioritised along with their consensus deleterious scores from 6 predictive algorithms, absence from major databases (e.g. dbSNP, 1000G, EVS) in addition to previous shortlisting criteria (read depth, sequence quality, mutation type, etc.) and included only those deemed to be rare variants [i.e. MAF <0.01 (Frazer et al., 2009)]. The summary of this analysis is given in Table 7.5 and Figure 7.4.

In patients OX6136 and OX4481, frameshift and splicing mutations were identified in *PPP2R3A* (c.736delT:p.C246Alafs*5 and c.3223-7_3223-4dupCAAT; (sp)] that encodes Protein Phosphatase 2, Regulatory Subunit B, Alpha or PR72. This splicing mutation is predicted to create a weaker acceptor site in the mutant transcript and may adversely affect splicing (see Figure 7.5). This protein is a negative regulator of canonical Wnt signalling, directly modulating Naked cuticle (*Nkd*), an antagonistic molecule that restricts Wnt signalling in segmentation of *Drosophila* and in *Xenopus*, is thought to affect cell movement and body axis formation (Creyghton et al., 2005).

Variants in *FHOD3* (formin homology 2 domain containing 3) were seen in OX5569 (c.2354C>T; p.P785L) and OX1560 (c.999G>T; Q333H), with the former being validated. These were listed as not having an inactivated allele in the Exome Variant Server. Formin proteins have a role in regulating the actin cytoskeleton and are involved in cell elongation. Although highly

expressed in cardiac muscle, it is also phosphorylated by Casein kinase 2 that affects Wnt signalling by phosphorylating β-catenin (Taniguchi et al., 2009; Kühn & Geyer, 2014).

In patients OX4580 (c.970T>C; p.N324D) and OX794 (c.1277T>C; p.V426A), convincing variants were seen in *ZNF16* (Zinc Finger Protein 16), with the former validated. This gene encodes a transcription factor that has been shown to target and downregulate c-Kit, an important protein tyrosine-kinase that is paralogous to FGFR3, and activates signalling via several pathways including RAS/RAF1/MAPK1/ERK2 and PI3K-AKT1within erythroid and megakaryocyte cells (Chen et al., 2014). Biallelic mutations have been described in this gene in compound heterozygote foetuses that were both terminated due to significant intrauterine growth restriction, complex brain formations, bilateral renal agenesis, flexion arthrogryposis of all joints and a non-viable phenotype. A flattened forehead, microcephaly and low set ears were also noted at autopsy (Filges et al., 2013).

Further variants were identified, but not yet validated (due to lack of experimental time) in *UBQLN4* (Ubiquilin 4), which helps regulate proteosomal protein degradation and *HELQ* (Helicase, POLQ-Like), a DNA helicase (see Table 7.3). These remaining 2 genes in addition to those that failed to be sequenced (OX4580 - *FHOD3* and OX794 – *ZNF16*) will

need to have their apparent variants validated and possibly screened in a larger population to assess their pathogenicity.

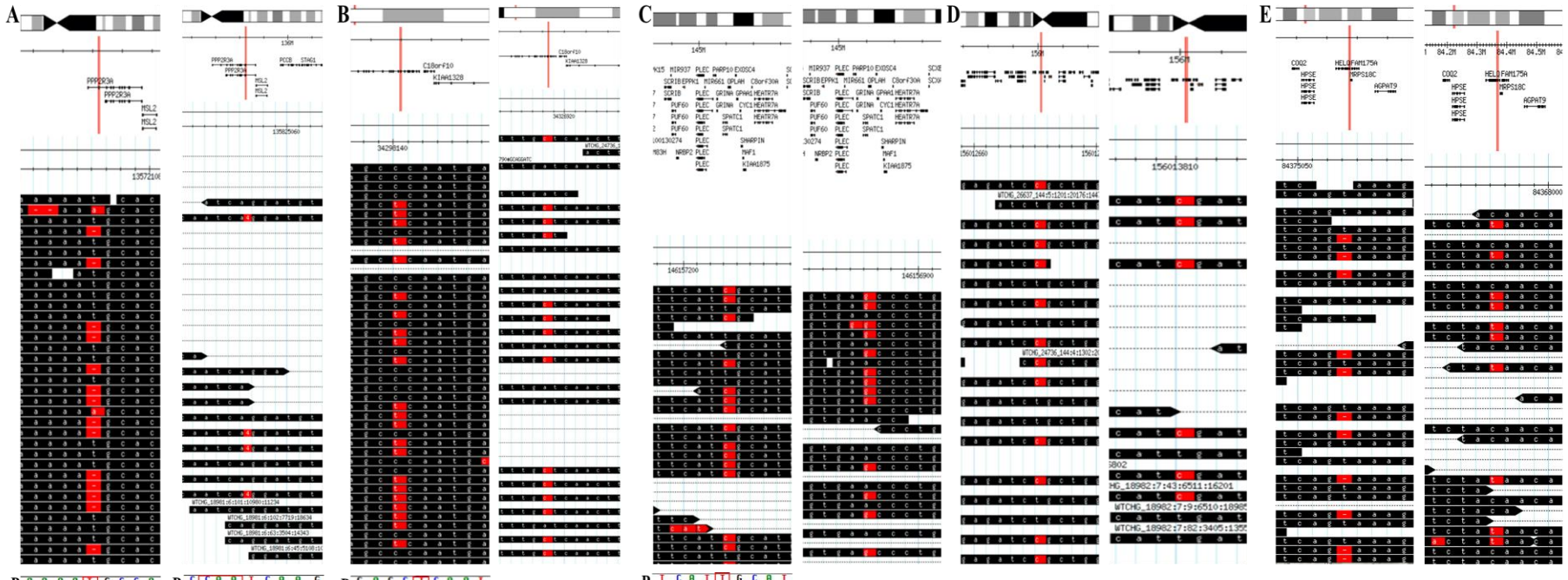
Gene/Cytogenetic location Nucleotide change/Protein change	High quality forward and reverse reads (wild-type | mutant)																			Consensus Deleterious score	Variant in known disease/developmental gene?											Validated?	Inherited? (M= mother /F= father)
	OX6041	OX3245	OX5098	OX1560	BCH6136	OX1662	BCH5656	OX4462	OX4481	OX3245	OX4580	OX797	OX1799	OX5569	OX1584	GOS6568	OX4809	OX2438	OX4400		Craniosynostosis Tier1 genes	DECIPHER CRS DEL	Mouse knockout CRS-related phenotypes	Craniosynostosis Tier2 genes	Copenhagen genes	*TWIST1* target genes	Non LOF genes (EVS)	Genes expressed in FNP	Scalp fibroblasts (Bouchukova et al., 2010)	Coussens suture list (Coussens et al., 2007)	Wingender transcription factors		
PPP2R3A Protein Phosphatase 2, Regulatory Subunit B, Alpha 3p22																															
c.736delT; p.C246Alafs*5 AF = n/a					18, 8, 15, 2															**6**									✓		Y	F_u
c.3223-7_3224-4dupCAAT; (sp)									4, 0, 3, 1											**6**									✓		Y	F_u

AF = 0.01547																							
FHOD3 Formin Homology 2 Domain Containing 3 18q12																							
c.2354C>T; p.P785L AF = n/a c.3527A>C; p.D1176A AF = 0.00001686											2, 6, 6, 8			22, 5, 30, 6						0 3	✓ ✓	Y N	Fu n/a
ZNF16 Zinc Finger Protein 16 8q24																							
c.1277T>C; p.V426A												15, 12								3	✓	n/a	n/a

AF = n/a c.970T>C ; p.N324D AF = n/a											11, 6, 8, 11	, 12, 11								0	✓		Y	?M ?dn
UBQLN4 Ubiquilin 4 1q21																									
c.1166A>G; p.Q389R AF = n/a c.1105A>G; p.N369D AF = n/a											6, 6, 3, 10	3, 0, 3, 3								5 3	✓ ✓		n/a n/a	n/a n/a
HELQ Helicase, POLQ-Like																									
c.343delA ; p.T115Lfs*6 AF = n/a																12, 50, 13				6	✓		n/a	n/a

c.1384G>A; p.V462I AF = n/a															40, 13, 44, 15	, 29				2	✓	n/a	n/a

Table 7.5 Overlap analysis of 19 mutation-negative patients with coronal synostosis with intersection of variants with genes implicated in craniosynostosis disease pathways. The full lists of genes that was cross referenced with variants identified by this process are given in Tables 34.1-34.3, Appendix. See Table 2.1 for full summary of patient phenotypes. Validated n/a = not able to validate due to lack of experimental time. No cases of parental consanguinity reported in any inherited variant identified. M/F_u = maternally/paternally inherited variant, unaffected mother/father (no variants inherited variants from affected parents were observed). dn = *de novo*. AF = allele frequency listed on Exome Aggregation Consortium database (http://exac.broadinstitute.org/). (For primers and conditions, see Table 32, Appendix).



OX6136 *PPP2R3A* c.736delT; p.C246Alafs*5	OX4481 *PPP2R3A* c.3223-7_3223-4 dupCAAT;(sp)	OX5569 *FHOD3* c.2354C>T; p.P785L	OX4580 *FHOD3* c.3527A>C;p.D1176A	OX4580 *ZNF16* c.970T>C;p.N324D	OX794 *ZNF16* c.1277T>C;p.V426A	OX4580 *UBQLN4* c.1166A>G;p.Q389R	OX797 *UBQLN4* c.1105A>G;p.N369D	GOS6568 *HELQ* c.343delA; p.T115Lfs*6	OX1584 *HELQ* c.1384G>A;p.V462I

Figure 7.4 Summary of analysis of 5 top candidate genes identified following overlap analysis of 19 mutation-negative patient exomes who all have coronal synostosis. Each of the shared variants identified were found in at least one list of genes associated with craniofacial development or damaging variants. **A:** Variants in *PPP2R3A* in patients OX6136 (Left) and OX4481 (Right). **B:** Variants in *FHOD3* identified in patients OX5569 (Left) and OX4580 (Right). **C:** Variants in *ZNF16* identified in patients OX4580 (Left) and OX797 (Right). **D:** Variants in *UBQLN4* identified in patients OX4580 (Left) and OX794 (Right) **E:** Variants in *HELQ* identified in patients GOS6568 (Left) and OX1584 (Right). Reads mapped normally to the human genome are in black, with variants indicated in red. Where possible, these variants have been validated by dideoxy sequencing and the chromatogram is given below**.** Conditions and primers for each PCR amplification are given in Table 31, Appendix).

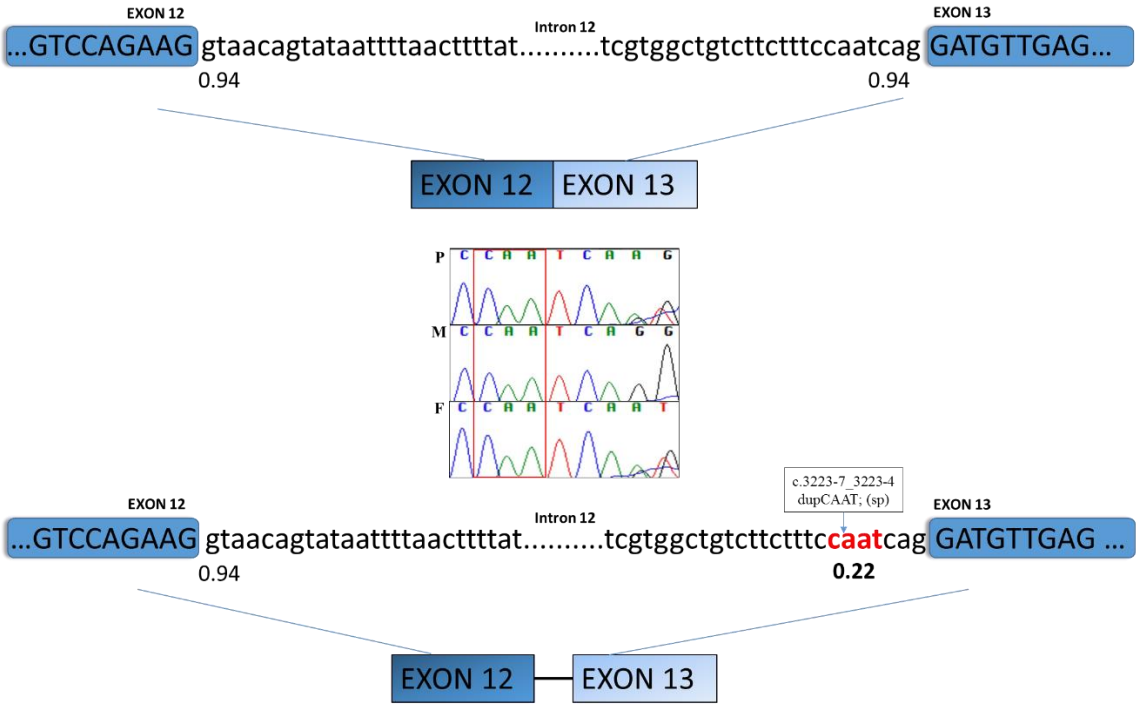


Figure 7.5 The predicted effects of the slicing mutation in *PPP2R3A* (c.3223-7_3223-4dupCAAT) identified in patient OX4481.The intronic region between exons 12 and 13 with normal/wild-type donor and acceptor sites (scoring 0.94 each) is shown (Top) along with dideoxy sequencing traces of proband (P), mother (M) and father (F) (Middle). The mutation (in red box) is located 7 nucleotides before the normal acceptor splice site (Bottom) and is predicted to create a weaker acceptor site (0.22). Further *in vitro* work of complementary DNA of this patient would be required to quantify the actual effects on splicing as per the *TCF12* c.1468-20T>A mutation in patient OX2636 described in Chapter 3. Predicted scores are from the Splice Site Prediction Neural Network program (www.fruitfly.org/seq_tools/splice.html).

7.6 Discussion

The subsequent phases of exome sequencing described in this chapter aimed to identify potential disease genes by applying this technique in 3 ways: a second round of exome sequencing involving 7 patients with coronal synostosis, exome sequencing 5 proband-parent trios to elucidate *de novo* mutations and combining the data from a larger exome cohort of 19 patients with non-syndromic coronal synostosis and applying an overlap strategy for analysis.

Firstly, identification of a *de novo* frameshift mutation in *TCF20* in Round 2 of exome sequencing accounted for the autistic spectrum disorder and moderate intellectual disability in patient OX2438, but not the coronal synostosis. This is because her mother had left-sided unilateral coronal synostosis and did not carry the *TCF20* mutation. Therefore, mutations of this gene are likely to account for the behavioural phenotype only, not the craniofacial abnormalities associated with coronal synostosis. The cause of the coronal synostosis in this family will therefore require further investigation, but an additional genetic locus has be excluded.

Secondly, the analysis of proband and parents aimed to typically identify up to 3 *de novo* mutations per trio analysed after all inherited variants had been excluded (O'Roak et al., 2011; Vissers et al., 2010). It became apparent however, that exome sequencing had been sub-optimal in this instance as a number of cases of poor read depth and exon coverage was

seen (see table 7.2). This meant that potential *de novo* variants were underrepresented and might have been missed during the shortlisting process. Additionally, many variants identified by the bioinformatic pipeline were actually artefacts when it came to the process of validating those variant calls.

Of the 3 candidates shortlisted for the OX1662 trio analysis, 2 were actually artefacts and one was inherited from the father (*TAF1B*). This latter variant was not noted to be inherited initially due to poor read depth in the father when the data was visualised on the genome browser. A variant inherited from the mother was also seen in OX1560 trio analysis in *SPG20* out of the 5 candidates shortlisted. None of the variants identified in the OX3245 or BCH4972 trio analyses could be validated. Therefore, the only true *de novo* variant identified in this analysis of 5 families that was also validated, was *MYH4* in OX6041. However, several lines of evidence appear to go against its role as a disease gene in craniosynostosis. It encodes a contractile component of skeletal muscle with no obvious role in skull and face development described. The variant identified (c.4691G>A; p.R1564H) had an overall moderately deleterious consensus score from the 6 algorithms used for predicting the functional effects of single nucleotide variants (see 1.6.6). Taken together, this probably negates the overall of candidacy of this gene in craniosynostosis.

An analysis of 19 mutation-negative patients with coronal synostosis using the cohort or overlap method (see 1.6.2) that successfully identified *TCF12* in Chapter 3, did not reveal such an outstanding candidate disease gene again, although a few putative candidates have been identified. Of the variants that were discovered by exome sequencing and subsequently validated (see Table 7.3) *PPP2R3A* represents the 'best' putative disease gene due to validated damaging frameshift and splicing mutations in 2 patients and the highest consensus predicted deleterious score. Furthermore, the splice site mutation is predicted to adversely affect normal splicing by creating a weaker acceptor site (see Figure 7.5). In both cases these variants were inherited from unaffected fathers, without a clinically apparent craniofacial phenotype. This in itself does not negate the candidacy of PPP2R3A as significant non-penetrance was observed with carriers of *TCF12* mutations (see Chapter 5). In a study of genes demonstrating a more than 2-fold differential expression between fused and unfused sutures (Coussens et al., 2007), *PPP2R3A* was found to be more highly expressed in unfused sutures. However, this observation was not statistically significant (Log_2 score = -1.0054, *P*-value = 0.4712), and so this gene may only have a minor role in normal suture morphogenesis. It has also been implicated as a negative regulator in body axis formation in model organisms (Creyghton et al., 2005). The impact of this potential disease gene would need to be assessed by screening in a mixed craniosynostosis disease cohort of mutation-negative patients as described

in Chapter 4, although the splicing mutation has been described in the Exome Aggregation Consortium website (http://exac.broadinstitute.org/) and so this particular mutation may be less significant than the other one demonstrating a frameshift mutation.

The remaining candidates identified in Section 7.5 in addition to those that could not be validated initially due to lack of experimental time and will require further analysis in future work to rule them in or out as potential disease genes.

CHAPTER 8

DISCUSSION

8.1 Introduction and general discussion

In this thesis, I have described the successful application of next-generation sequencing and bioinformatic techniques to identify a new disease gene, Transcription Factor 12 (*TCF12)*. This was achieved by interrogating the protein-coding fraction of the genome of patients with coronal synostosis who had otherwise remained without a formal molecular diagnosis. By screening a large, mixed craniosynostosis disease cohort, and surveying the clinical features of mutation-positive individuals, a new clinical entity has been described, and officially recognised on an international database (CRS3; #615314; http://www.omim.org/entry/615314). Mutations of this gene are now as a consequence tested for on a national basis, thereby improving diagnostic outcomes for patients with this disease. This has been approved by the UK Genetic Testing Network (http://ukgtn.nhs.uk/).

This thesis also reports an investigation into the mutational origin of *de novo* cases which assisted further study into the genetic background of *TCF12*-mutation positive carriers who fail to manifest any obvious clinical features. The presence of an association between *TCF12* haplotypes and non-syndromic coronal synostosis was also sought. While a modest correlation between *TCF12* haplotype and phenotypic variability was observed, this was not adequate to be used for clinical testing and no

statistically significant association was seen between *TCF12* haplotype and non-syndromic coronal synostosis.

Finally, this thesis documents attempts to discover further novel disease genes in coronal synostosis by refinement of the previously successful strategy. Additionally, patient-parent trios were analysed for *de novo* mutations and data to form a larger cohort of mutation-negative patients combining for further investigation. These further analyses revealed another disease gene (*TCF20*) that although not implicated in coronal synostosis, did account for a novel association with autistic spectrum disorder. Additionally, *de novo* variant analysis in trios identified *MYH4*, although its role in craniofacial development is currently uncertain. Finally, a number of potential disease genes have been identified as candidates for future study.

8.2 A combined clinical and molecular approach to cases of mutation-negative coronal synostosis identifies a new clinical entity: *TCF12*-related craniosynostosis

As described in Chapter 1, coronal synostosis is a genetically enriched condition with a preponderance of monogenic syndromes presenting with premature fusion of that suture. Cases of apparently non-syndromic coronal synostosis have been shown to have a high diagnostic yield (Wilkie et al., 2010) and embryologically, the coronal suture has a unique

origin, distinct from the other cranial sutures (Jiang et al., 2002). Therefore, as a sub-group amongst all craniosynostosis patients, those possessing this particular pattern of suture fusion represent a good population to target with a novel approach to disease gene identification in those that are mutation-negative.

A combined clinical and novel molecular approach was applied to identify new disease genes in cases refractory to diagnosis by current methods of genetic testing. By carefully stratifying a cohort (see Chapter 2) who with the exception of fused coronal sutures, were mostly non-syndromic in all other regards, the chances of identifying a gene with damaging mutations shared between 2 or more patients was maximised. The power of this approach had been demonstrated previously by other groups that identified the monogenic basis of other syndromes with distinct craniofacial phenotypes (Gilissen et al., 2010; Ng et al., 2010).

Fourteen patients were exome sequenced in two rounds, consisting of 7 patients each. In Round 1, 2 damaging, heterozygous frameshift mutations and a splice site mutation in a third patient were identified in a gene not previously linked with coronal synostosis - *TCF12*. Once patients with mutations in this gene were removed from this initial cohort of 7, no further candidate genes were apparent while using this overlap strategy. The 2 frameshift mutations in patients OX484 (c.1349delC; P450Lfs*69) and OX4581 (c.1642_1645delGAAA; E548fs*14) were missed by the

variant calling software used initially (Bowtie) as it used an algorithm unable to align and read indel mutations. The pipeline was further refined and the data re-processed using Novoalign (Krawitz et al., 2010) that was able to identify indels in short-read data, and used in conjunction with an updated version of ANNOVAR to functionally annotate genetic variants (Wang, et al., 2010).

An important caveat to note when using exome sequencing to identify novel disease-causing variants is that the target regions in commercially available capture kits will miss a significant proportion of non-coding variation located within introns and intergenic regions. This was demonstrated during the course analysis of Round 1 data as a novel cryptic splice site mutation in *TCF12* was missed initially by annotation software that was not able to call this mutation in patient OX2636. The c.1468-20T>A; Val490Gfs*4 variant was only noted following a manual text search of single gene variants amongst the 5 remaining patients after the pipeline had identified the 2 frameshift mutations. The discovery of 3 damaging mutations in my patient cohort encouraged others in our group to re-analyse data in their parallel exome and whole-genome sequencing projects. They were able to identify a *de novo* mutation in patient GOS4437 (c.1283T>G; L428* - see Figure 3.3) that added further evidence of pathogenicity. It also provided further support for the candidacy of *TCF12* as a disease gene as family members of other mutation-positive patients prior to this discovery were non-penetrant for

coronal synostosis. Additionally, a familial case of craniosynostosis with two siblings with bilateral coronal synostosis both possessed maternally-inherited *TCF12* frameshift mutation (c.1376T>G; L459*).

Although not previously linked with craniosynostosis, *TCF12* appeared to fit well within known disease pathways as it encodes a class I bHLH transcription factor that is a dimerisation partner for class II transcription factors whose most notable member in this context was TWIST1 (Connerney et al., 2006). Mutations of *TWIST1* have long been established in their association with Saethre-Chotzen syndrome that presents predominantly with coronal synostosis (el Ghouzzi V et al., 1997; Howard et al., 1997). Further supporting evidence for *TCF12* as an outstanding disease gene came from clinical genetics case reports describing patients with chromosomal deletions at the 15q locus where *TCF12* is located, in patients with severe bilateral coronal synostosis (Fukushima et al., 1990; Shur et al., 2003).

Screening a large, mixed craniosynostosis disease cohort of 344 undiagnosed patients as described in Chapter 4, identified 35 further mutations in *TCF12*, and importantly 13 further *de novo* mutations. Overall, mutations of this gene were associated with 32% of bilateral and 10% of unilateral cases of coronal suture fusion. The commonest pattern was bilateral, followed by right coronal, left coronal then combinations of coronal plus sagittal suture fusion. These findings provided overwhelming

evidence of the pathogenicity of *TCF12* mutations in coronal synostosis. Due to the significance of this discovery, diagnostic testing for *TCF12* mutations was rapidly translated from a research to a clinical setting in national laboratories in Oxford and London.

In an analysis of patients with intellectual disability and microdeletions in the 15q22 region, a 6.5Mb deletion that arose *de novo* was identified in a 3 year old girl who also presented with craniofacial dysmorphism including macrocephaly, prominent forehead and hypertelorism (Yamamoto et al., 2014). While this deletion included *TCF12*, this patient was not reported to have coronal synostosis on cranial CT scan (although no 3D reconstruction of images is described by the authors). This would appear to fit with observations of significant non-penetrance observed in my patient cohort and the smaller sub-group with developmental delay.

Reports by other groups of microdeletions that include the region where *TCF12* is located, describe craniofacial phenotypes that include coronal synostosis. Le Tanno et al. (2014) describe a 2.75 year old boy with a complex unbalanced chromosomal rearrangement that was maternally inherited between chromosomes 2 and 15 [t(2;15)(q21; q21.3)] and a *de novo* 3.64Mb deletion on chromosome 15. This deletion was located in region 15q21.3q22.2, very close to the breakpoints of the translocated chromosome. As a result of *TCF12* haploinsufficiency from the deletion, this patient had a right-sided unilateral coronal synostosis with expected

craniofacial dysmorphism in addition to 2/3 toe syndactyly (simple), a small pituitary gland and developmental delay. These clinical features are again consistent with those described in this thesis (see Table 4.3).

While no gene deletions or duplications were detected by multiplex ligation-dependent probe amplification (MLPA) in 226 patients negative for *TCF12* mutations (work by A Fenwick, see Sharma et al., 2013), a case report by (Piard et al., 2015) describe a patient with a 15q21.3 microdeletion detected by array-CGH. MLPA was used to refine the deleted interval more precisely and demonstrated a heterozygous deletion of exons 19-20 that contains the critical bHLH domain, again resulting in haploinsufficiency of this gene. This patient like those previously described with microdeletions, had intellectual disability with facial dysmorphism and had presented at 72 years old for clinical assessment at the request of her brother. A 3D CT scan of her skull demonstrated thickening of the cranial vault, but unsurprisingly, given her delayed presentation, no obvious craniosynostosis. Therefore, these published reports of microdeletions expand the mutational spectrum of *TCF12*-related craniosynostosis and their association with intellectual disability indicates a possible difference in phenotype compared to *TCF12* point mutations.

Acquisition of further samples from referring diagnostic centres following the initial publication of results (Sharma et al., 2013), allowed clinical

information on a total of 48 families to be collated and the new clinical entity of *TCF12*-related craniosynostosis to be better defined (see Chapter 4). A non-syndromic presentation was commonest in which patients only manifested other craniofacial deformities that were consistent with a fused coronal suture. In other cases, a mild Saethre-Chotzen-like phenotype would be observed with minor ear anomalies, low anterior hairline and blepharoptosis.

Other low frequency features that were seen included minor hand or foot anomalies (single palmar creases or simple soft-tissue syndactyly respectively) and benign radiological findings on imaging of the brain, none of which caused any musculoskeletal or neurological deficits. Malocclusion of the teeth was seen in under 8% of cases with only one patient requiring formal orthognathic surgery under general anaesthetic to correct the dental deformity. Most *TCF12*-muation positive individuals had normal neurocognitive development, but 14% had documented developmental delay, ranging from problems with speech and language to Asperger syndrome and autistic spectrum disorder. This subset of patients will require further investigation to better delineate the nature of the developmental delay and to assess for the presence of any subtle genotype-phenotype correlation. There is also a reported instance of a patient with a maternally inherited frameshift mutation in *TCF12* that was identified in an exome sequencing analysis of 95 trios where the child had autistic spectrum disorder. This patient was not reported to have craniosynostosis

or obviously associated dysmorphology, although a brachycephalic skull shape was reported. Therefore, in the absence of formal skull imaging, an association with coronal synostosis cannot be completely ruled out.

8.3 Investigation into non-penetrance in mutation-positive carriers, parental origins of *de novo* mutations and *TCF12* haplotype

A significant issue that emerged from analysis of 48 families containing *TCF12*-mutation positive individuals was a 60% non-penetrance of coronal synostosis that was observed in carriers. These carriers appeared clinically normal with normal cephalometric measurements, therefore it was hypothesised that their genetic background or haplotype may affect phenotypic variability. Additionally, it may also be related to relative *TCF12* expression levels, whereby possession of a high-expressing wild-type allele may compensate for the effects of a mutant *TCF12* allele, thus preventing the development of coronal synostosis (see Chapter 5). A test of association between non-syndromic coronal synostosis and *TCF12* genotype was also performed.

Previous work into the genetic background of Hirschsprung disease (Emison et al., 2005; Emison et al., 2010) and thrombocytopenia with absent radii (Albers et al., 2012) provided the context and rationale for further investigation into *TCF12* haplotype background (see Chapter 6). Other groups have also reported non-penetrance of mutation-positive

carriers in a smaller cohort of 4 families (di Rocco et al., 2014). Penetrant individuals demonstrated similar clinical features to my cohort (i.e. coronal synostosis with a mild Saethre-Chotzen phenotype including prominent ear crura, blepharoptosis and brachydactyly).

Relating the work on these two diseases to *TCF12* mutations and the phenotypic variability seen in penetrant and non-penetrant individuals, it was felt that a specific SNP or other variant on a certain *TCF12* haplotype or a particular haplotype in non-random association with a specific variant may prove to be a pre-disposing or protective factor in disease development.

An investigation into the parental origins of *TCF12* mutations allowed the correct assignment of phase in some cases of *de novo* mutation identified in the 48 families described in Chapter 4. Use of the HapMap database revealed a limited haplotype structure around *TCF12*, hence an investigation looking into the association of penetrance or expressivity with the parent of origin would be possible. By use of an analogous strategy to one that successfully identified the parent of origin in Apert syndrome (Moloney et al., 1996), I screened 14 trios where the *TCF12* mutation had arisen *de novo* in the child. By sequencing genomic regions flanking the mutation, I identified SNPs that would make analysis informative, so that the parent of origin of each allele could be identified unambiguously. I then designed allele-specific PCRs that would

selectively amplify either parental allele, and performed an appropriate restriction digest or further sequencing to demonstrate on which allele the mutation arose. Of the 14 *de novo* trios analysed, 6 were informative and demonstrated a paternal origin for the *TCF12* mutation.

However, *TCF12*-related craniosynostosis was not found to have an exclusive paternal-origin as is the case for Apert Syndrome since 2 cases of maternal origin were demonstrated, one due to a maternal mosaic mutation and the other via segregation of haplotypes (see Section 5.6). The phenomenon of maternal mosaics in craniosynostosis has been previously described in craniofrontonasal syndrome (Twigg et al., 2006). Additionally, no paternal-age effect was demonstrated as the average age for these 6 fathers was 0.2 years younger than the national UK average.

Investigation of *TCF12* haplotypes utilised 46 of the 48 families identified in Chapter 4 (2 families excluded as haplotype could not be assigned due to lack of DNA samples). A specific set of tagging SNPs were used to capture most of the allelic diversity seen in *TCF12*, thereby obviating the need to genotype every SNP seen across the region occupied by this gene (see Chapter 6). Two control populations (HapMap CEU and UK) were genotyped to obtain allele frequencies of each tag-SNP in healthy, unaffected individuals and to construct validated haplotypes. Samples from *TCF12*-mutation positive individuals and their family members were then constructed and the data plotted according to the ancestral haplotype.

Previous evidence (Emison et al., 2005; Emison et al., 2010) has suggested that possession of a certain haplotype or genotype is associated with disease susceptibility.

It was hypothesised that the ancestral haplotype may pre-dispose to disease, and the derived haplotype may prevent its development. The allele frequencies of the mutant *TCF12* allele showed a bias toward the ancestral haplotype, away from any protective effects of the derived haplotype and thus appeared to fit with the hypothesis. The mutant *TCF12* allele was also observed to be transmitted through a family pedigree with a partner allele either in the presence (i.e. disease penetrant) or absence (i.e. non-penetrant) of craniosynostosis. The allele frequencies of both were similar to background, and none of the data points were statistically significant. A striking bias in the allele frequencies was observed for the non-transmitted bystander allele, being deviated significantly towards the ancestral haplotype, with statistically significant results for 3 tag-SNPs. This in itself poses a new scientific question as it had previously been included in the haplotype analysis as an internal control. No instances of transmission of a mutant plus bystander *TCF12* allele were observed, possibly as this situation would be non-viable. The data from the 6 families with *de novo TCF12* mutations that demonstrated a paternal origin was used to increase the sample size available for haplotype analysis as it allowed additional mutant and partner alleles to be assigned. In uninformative cases, the mutant and partner alleles were apportioned

50:50 rather than presuming paternal origin. However, this scenario also did not give statistically significant results.

An important observation that may negate the importance of the type of *TCF12* allele present in mutation-positive individuals, came from 3 families (Family #11, #35 and #45) which had siblings with the same haplotype, but differing phenotypes - either penetrant or non-penetrant for coronal synostosis. Each had inherited an identical mutant *TCF12* allele from one parent and a normal *TCF12* allele from the other unaffected parent, but an apparent discordant phenotype was observed. Additionally, in Family #34, inheritance of opposite maternal alleles in conjunction with the paternal allele still resulted in coronal synostosis in their offspring. Therefore, other factors in the environment or those with epistatic or epigenetic effects may influence phenotypic variability in mutation-positive carriers to a greater extent. Additionally, no statistically significant association between *TCF12* haplotype and non-syndromic coronal synostosis was observed. Therefore, the modest correlation between phenotype and *TCF12* haplotype illustrated by the analysis described in Chapter 6, is too subtle to be of value in clinical genetic testing.

In the context of gene expression whereby a high expressing normal *TCF12* allele may overcome the effects of a mutant *TCF12* allele, resulting in a normal phenotype, novel results were obtained. Allele-

specific gene expression occurs commonly in many biological pathways such as transcriptional regulation within the human genome (Palacios et al., 2009). Performance of a transmission disequilibrium test to look for linkage between SNP transmission from a non-carrier, unaffected parent (in the presence of an association with coronal synostosis) gave results for 3 tag-SNPs that were statistically significant (rs2585082, rs1628067, rs12901270). Two of the same tag-SNPs also demonstrated statistically significant differences between high and low expressing genotypes in control individuals which correlated with the over-transmitted G and C SNP alleles of rs2585082 [T/**G**] and rs12901270 [A/**C**]. Although this expression data was derived from lymphoblastoid cells not those with osteoblastic lineage, these two lines of evidence may have a biological significance that supports the effects of a high-expressing wild-type *TCF12* allele that prevents coronal synostosis despite the possession of a mutant *TCF12* allele.

8.4 New insights into cranial suture biology from double mutant mice demonstrates importance of dosage sensitivity in coronal suture boundary maintenance

As described in 1.3.1.1, Twist1 can form homo or heterodimers that interact with and inhibit Runx2 to regulate osteoblast differentiation in sutural mesenchyme. For stable boundary formation within the coronal

suture, a full dose of this gene is required as *Twist1*$^{+/-}$ heterozygous mice have coronal synostosis and *Twist1*$^{-/-}$ homozygous mice are non-viable (Chen & Behringer, 1995; Carver et al., 2002). TWIST1 is a Class II bHLH protein that is known to form heterodimers with Class I bHLH proteins including HEB, encoded by *TCF12* (Connerney et al., 2006). A transactivation assay (work by A Fenwick; see Sharma et al., 2013) demonstrated a functional interaction *in vitro* between TWIST1 and TCF12 molecules with specific missense mutations that had been identified in Chapter 4 (c.1912C>G;Q638E in Family#37 and c.1963G>T; E655* in Family #38). Reduced reporter activity of TCF12-TWIST1 heterodimers was seen when specific missense mutations occurring in the functionally important bHLH region were introduced (L624P or Q638E) (Sharma et al., 2013).

Based on these observations, it was hypothesised that TWIST1:TCF12 heterodimer dosage may therefore be important in the development of the normal coronal suture and regulate boundary maintenance. To test this, our collaborators (M Brockop, Dr R Maxson) constructed heterozygous null mutant mice (EIIa-Cre; *Tcf12*$^{flox/+}$ and *Twist1*$^{+/-}$) and compared their phenotype with double mutant mice (EIIa-Cre; *Tcf12*$^{flox/+}$; *Twist1*$^{+/-}$) at 3 weeks of age. Compared with wild-type mice, those with heterozygous mutations in *Tcf12* (EIIa-Cre; *Tcf12*$^{flox/+}$) also had normal coronal sutures demonstrating this species is less sensitive to *Tcf12* dosage than humans. *Twist1*$^{+/-}$ heterozygotes demonstrated partial coronal synostosis as

expected. However, most striking was the severe bilateral coronal synostosis seen in the double mutant mice (EIIa-Cre; *Tcf12*$^{flox/+}$; *Twist1*$^{+/-}$) which also had shorter skulls (Sharma et al., 2013), thereby illustrating the importance of *Tcf12* dosage. In summary, there is compelling *in vitro* and *in vivo* evidence supporting the action of the TCF12-TWIST1 heterodimer in the coronal suture, which is likely to modulate RUNX2 inhibition in the mesenchyme of the coronal suture.

8.5 Investigation into the immune phenotype of patients with *TCF12*-related craniosynostosis implies the immune system is not affected in individuals with heterozygous mutations

TCF12 had previously been shown to have a role in the development of B- and T-lymphocytes. Mice with homozygous *Tcf12* mutations demonstrate a large reduction in the numbers of pro-B cell precursors (Wojciechowski et al., 2007). Alternately spliced forms of the gene forms protein products that are important in uniquely specifying T-cell lineage (HEBAlt; Wang et al., 2006) or act as negative regulators in the thymus by affecting the fate of natural killer cells (HEBCan; Braunstein & Anderson, 2012).

Despite haploinsufficiency of *TCF12*, no instances of recurrent infections or obvious clinical signs of immunosuppression or dysfunction were identified in patients with *TCF12*-related craniosynostosis. However, due to its well documented role in the development of various immune cells,

an investigation into the immune phenotype was performed by collaborators (S Bennet, Prof R Cornall; see Sharma et al., 2013). The numbers of various immune cells and the antigens present on their surface was analysed in 23 individuals from 6 families, of which 11 were mutation-positive. This included B-cells (IgD+, IgM+, CD27+, CD10+, CD14+), T-cells (CD4, CD8), Natural Killer cells (CD161+, CD25+) and monocytes. However no difference was detected in those with *TCF12*-mutations and due to a lack of a definite immune phenotype, this was not investigated any further.

8.6 Evaluation of candidate disease genes and putative disease variants using further exome sequencing with existing and alternate strategies

The discovery of damaging mutations in *TCF12* in 3 patients following Round 1 of exome sequencing represented a major novel disease gene discovery that 'solved' these previously undiagnosed patients. After removing these patients from the initial cohort of 7, no further convincing variants shared between the 4 remaining patients were seen. Round 2 exome sequenced another 7 patients and a combined analysis of variants shared between 2 or more patients was performed on 11 patients in total by including the 4 mutation-negative patients in Round 1. However, no outstanding candidates emerged. Analysis of single gene variants revealed

a frameshift mutation in transcription factor 20 (*TCF20*), in patient OX2438 who had right-sided unilateral coronal synostosis, mild intellectual disability, autistic spectrum disorder and a mother with left coronal synostosis. However, as this mutation was later confirmed as *de novo* (Babbs et al., 2014) and therefore did not segregate with the maternal phenotype, mutations of *TCF20* could only be associated with the behavioural and not the craniofacial phenotype in this patient.

An alternate exome sequencing strategy was then used to identify *de novo* mutations that might be associated with coronal synostosis, utilising 5 patient-parent trios. The basic premise of this strategy was to exclude all inherited variants, leaving approximately 0-3 putative pathogenic variants per trio analysed (Vissers et al., 2010; O'Roak et al., 2011). A significant issue in the quality of the exome data became apparent following this analysis. Poor read depth and exon coverage was encountered in many of the individuals sequenced during this phase of the study. This meant many variant calls were potentially missed, or if they were apparently present, could not be later validated and indicated a high proportion of artefacts. Additionally, in 2 trios, apparently *de novo* variants identified by exome sequencing were actually inherited when validated by dideoxy sequencing.

Only one true *de novo* variant identified by this exome sequencing strategy could be further validated - c.4691G>A; p.R1564H in myosin heavy chain 4 (*MYH4*) in patient OX6041. This finding was facilitated by good read

depth and exon coverage observed in all 3 members of this family trio (see Figure 7.2). However, no association with craniofacial development as yet has been ascribed to the encoded motor protein. The significance of this finding is therefore uncertain.

The cohort or overlap strategy that successfully identified *TCF12* (Chapter 3) was again applied to a larger cohort of 19 mutation-negative patients with coronal synostosis. The shared variants subsequently identified were cross-referenced with 11 lists of genes associated with known craniofacial syndromes, developmental pathways important in the face and skull, no reported inactivated alleles in a large public exome database or are described in a list of all known transcription factors. Of the shared variants identified by exome sequencing that were subsequently validated, the damaging frameshift (c.736delT; p.C246Alafs*5) and splice site [c.3223-7_3224-4dupCAAT; (sp)] mutations observed in *PPP2R3A* (Protein Phosphatase 2, Regulatory Subunit B, Alpha or PR72) represents a candidate gene that could warrant further investigation. The splice site mutation was predicted to create a weaker splice acceptor site *in silico* and will need functional *in vitro* validation as with the splicing mutation identified in Chapter 3. The encoded enzyme has been shown to be a negative regulator in Wnt-signalling and important in body patterning in *Drosophila* and *Xenopus* (Creyghton et al., 2005) as well as roles in PI3K/Akt and c-Myc pathways implicated in cellular transformation that cause aberrant signalling in human disease (Sablina et al., 2010).

Identification of numerous loss of function alleles corresponding to the splice site mutation present in the Exome Aggregation Consortium database may negate its overall significance, although the frameshift mutation has not been reported previously.

Further mutation-positive patients would need to be identified from screening a large craniosynostosis cohort as before, as well as *de novo* mutations to add further lines of evidence that could further prove the pathogenicity of mutations in this gene. In addition, other convincing variants identified by exome sequencing would need further analysis in the following genes: *FHOD3, ZNF16, UBQLN4* and *HELQ*.

8.7 Remaining questions in coronal craniosynostosis

The discovery and delineation of *TCF12*-related craniosynostosis represents a major development in the field, successfully applying exome sequencing to 'solve' a proportion of undiagnosed cases of coronal synostosis. Presently, clinical exome sequencing is now being adding to the routine armamentarium of gene tests available for rare Mendelian diseases (Lee et al., 2014) and the same is likely to occur for craniosynostosis patients. It has identified a new disease locus within craniosynostosis that measurably improves diagnostic outcomes in genetic counselling and craniofacial surgery, further aided by its rapid translation into national testing (Sharma et al., 2013). A Spanish series of 182

craniosynostosis patients and 89 family members that was analysed after the publication of this work puts the frequency of *TCF12* mutations at 4.2% (Paumard-Hernández et al., 2014). A more recent study by our group estimated a prevalence of 1.3% amongst 531 individuals known to the Oxford Craniofacial Unit (Twigg & Wilkie, 2015). Through additional rounds of exome sequencing of mutation-negative patients, further candidate disease genes have been identified, the impact of which will require screening in a larger cohort to identify further mutation-positive patients. Alternatively, proband-parent trios could be analysed to identify pathogenic *de novo* mutations (Twigg et al., 2015).

However, there still remains a large proportion of patients with coronal synostosis within the various cohorts that have been exome sequenced without a formal molecular diagnosis. Technical issues in the exome sequencing process may in part explain this shortcoming with poor coverage and read depth seen in a number of individuals manifesting as numerous artefacts identified in subsequent rounds of exome sequencing. As a result, these may not be called correctly or missed entirely by the bioinformatic algorithms used by our in-house pipeline.

Importantly, the work presented in this thesis results from the interrogation of the 1-3% of the genome that is protein coding. By focussing only on the exome, a very large of fraction of the genome that contains non-coding variants has been left out. Whole-genome sequencing might be the next

logical step in identifying new disease genes, but results in approximately 100 times the amount of data and any such strategy would still begin with the exome as it contains over 85% of the mutational information in a typical genome (Rabbani et al., 2014). These ‘unsolved’ cases of coronal synostosis may not have a simple monogenic answer. Multiple defects in the complex regulatory architecture that controls craniofacial development might synergistically play a greater role. Distal acting enhancers have been shown to influence craniofacial morphology in mice (Attanasio et al., 2014). Although located long genomic distances away from their target genes, various regulatory enhancer sequences are active at many embryological stages of development, modulating the final craniofacial phenotype in the mouse, presently the best model organism for human skull development. Use of three dimensional micro-computed tomography can be utilised as an additional modality to comprehensively quantify changes in dimensions and anatomical features of specific craniofacial bones in mice (Ho et al., 2014).

Three-dimensional (3D) photogrammetry will, in the near future be integrated into the investigation of genotype-phenotype correlations by virtue of principle component analysis. This reduces the number of facial landmarks required to precisely discriminate between face shape differences to a significantly smaller set of data points to represent a facial phenotype. Dense surface modelling allows for delineation of finer facial features by use of a sparse set of facial landmarks corresponding to tens of

thousands of three-dimensional quasi-landmarks to accurately represent a face and has been used to more accurately phenotype in another craniofacial syndrome identified by exome sequencing, *ERF*-related craniosynostosis (Twigg et al., 2013).

Use of facial signatures and signature weight respectively identify statistically significant differences between the face of an individual compared to a matched control and provide an estimate of that individual's dysmorphism. In the absence of a formal molecular diagnosis, a dysmorphology index of an undiagnosed individual may help to more deeply phenotype and stratify them into sub-groups until a specific genetic lesion is identified (Hammond & Suttie, 2012).

There may also be a greater role for extrinsic forces acting *in utero* than previously implicated in craniosynostosis as demonstrated by various epidemiological studies (see 1.2.3). External mechanical forces that act on the developing skull originate not only from the underlying growing brain, but also ectocranial muscle attachments and soft tissue structures such as the eye and oropharynx. Collectively, these forces can apply significant tension at developing sutures, thereby altering their morphology (Percival & Richtsmeier, 2011).

At the cellular level, factors as integrins, extracellular matrix and the cytoskeleton collectively affect mechanostransduction (Ingber, 2010) and micromechanical forces are known to influence processes such as

cutaneuous wound healing by altering inflammatory, angiogeneic and proliferative stages (Lancerotto & Orgill, 2014). Cilia and defects in ciliogenesis may also influence sutures biogenesis due to roles in Hedgehog and WNT signalling as *Fuz* or *Kif3a* mutant mice have coronal and metopic synostosis respectively (Tabler et al., 2013; Liu et al., 2014; Twigg & Wilkie, 2015). In turn, this could influence the three-dimensional development of the sutures as well as other other cranial components. The effect of such forces could be speculated to act more influentially on craniofacial phenotype in the presence of a specific gene mutation, thereby compensating for those effects to a greater extent than haplotype background alone which this thesis has only demonstrated a modest and inconsistent association.

Further refinements in understanding may be sought by use of proteomics, to analyse protein expression profiles of wild-type and mutant *TCF12*-encoded proteins. In cystic fibrosis, a monogenic disorder caused by mutations in *CFTR*, factors other than mutations in the gene itself, contribute to the observed phenotypic variability. Aberrations seen at the protein-level include defective folding, altered cellular trafficking and chloride ion channel functioning (Roxo-Rosa et al., 2004).

Techniques utilising two-dimensional electrophoresis (2-DE) have been further developed and shown by a variety of groups to identify the importance of chaperone proteins and specific protein motifs as potential

drug targets in carriers of the ΔF508del mutation (Kamath et al., 2015). Development of a cellular map in osteoblasts derived from sutures and other calvarial elements in the mouse for example, would allow a comprehensive analysis of protein-protein interactions (an 'interactome'). This could then identify other binding partners of TCF12/HEB that may augment the full effect of heterozygous mutations. An analogous strategy in humans may then identify specific proteins that interact with mutant forms that prevent the development of coronal synostosis, possibly accounting for the significant non-penetrantrance observed in this study.

Therefore, although this thesis has added to the body of knowledge and improved understanding of coronal synostosis by identifying a new mutational 'hotspot', there still remains a significant proportion of individuals with this pattern of suture fusion that remain undiagnosed.

8.8 Conclusion

In conclusion, this thesis has identified a new disease gene and clinical entity (*TCF12*-related craniosynostosis) that accounts for a measurable proportion of diagnoses in patients with coronal synostosis in whom previous gene tests were negative. This was made possible by the successful execution of a combined clinical and molecular strategy that applied next-generation sequencing to identify a new cause of coronal craniosynostosis. Improvements in this technology are occurring at a rapid pace and sequencing exomes or whole genomes is likely to enter standard clinical practice in undiagnosed patients in the near future. Clinical translation of this discovery within the lifetime of this thesis has allowed a measurable improvement in diagnostic outcomes for craniosynostosis patients. It has investigated the genetic background of mutation-positive individuals who fail to manifest the expected coronal synostosis phenotype, identifying limited correlation with *TCF12* haplotype. Finally, it has described the search for further novel disease genes using trio and larger cohort strategies, and while no further major discovery has been delineated, a number of candidate genes and variants of unknown significance have been identified, creating further avenues of enquiry.

8.9 Future Work

Further variants described in the latter stages of this thesis represent candidate genes that will require further analysis. Continuous refinements in exome sequencing and bioinformatic techniques are likely to reveal more potential disease genes. A step-wise progression to whole-genome sequencing may then identify important non-coding variants that may account for unsolved cases of coronal synostosis. The significant issue of non-penetrance still remains and will require further work to look at epistatic and epigenetic effects on gene regulation in addition to environmental factors to better understand why individuals with the same haplotype can manifest distinctly discordant phenotypes. This underscores the continued importance of genetic testing of family members of mutation-positive individuals.

Integration of three-dimensional photogrammetry with advanced imaging software developed with close input from experienced clinicians and dysmorphologists will allow for superior and deeper phenotyping. By association with specific genetic variants, this will allow for more accurate delineation of undiagnosed craniofacial syndromes, to further improve diagnostic outcomes for this group of patients. Progress is already underway with the FaceBase consortium (https://www.facebase.org/), an international collaborative organisation that has created an antatomical, genetic and 3D imaging repository to improve knowledge in craniofacial

development and malformation and this intergrated approach will be pivotal for further progress in the field.

In the short-term, new disease genes identified using the methodologies described can be rapidly translated into clinical testing strategies and improve diagnostic outcomes, thereby increasing the accuracy of genetic counselling offered and better guiding surgical management. As more disease genes are identified, further therapeutic targets are potentially highlighted. This may allow existing pharmaceutical agents to be used in a different way or can focus the development of new agents to prevent the occurrence of coronal synostosis, reduce its progression and severity at an early stage, or augment existing surgical techniques to prevent secondary complications.

BIBLIOGRAPHY

Ababneh FK, AlSwaid A, Youssef T, Al Azzawi M, Crosby A and AlBalwi MA (2013) Hereditary deletion of the entire FAM20C gene in a patient with Raine syndrome. *American Journal of Medical Genetics. Part A*. 161A (12), 3155–60.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS and Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nature Methods*. 7 (4), 248–9.

Agochukwu NB, Solomon BD and Muenke M (2012) Impact of genetics on the diagnosis and clinical management of syndromic craniosynostoses. *Child's Nervous System*. 28 (9), 1447–63.

Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst P a, Jolley JD, Cvejic A, Kostadima M, Bertone P, Breuning MH, Debili N, Deloukas P, Favier R, Fiedler J, Hobbs CM, Huang N, Hurles ME, Kiddle G, Krapels I, Nurden P, Ruivenkamp C a L, Sambrook JG, Smith K, Stemple DL, Strauss G, Thys C, van Geet C, Newbury-Ecob R, Ouwehand WH and Ghevaert C (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nature Genetics*. 44 (4), 435–9, S1–2.

Anastassiadis K, Glaser S, Kranz A, Berhardt K and Stewart AF (2010) A practical summary of site-specific recombination, conditional mutagenesis, and tamoxifen induction of CreERT2. *Methods in Enzymology*. 477 (10), 109–23.

Anderson MW and Schrijver I (2010) Next generation DNA sequencing and the future of genomic medicine. *Genes*. 1 (1), 38–69.

Arman A, Bereket A, Coker A, Kiper PÖS, Güran T, Ozkan B, Atay Z, Akçay T, Haliloglu B, Boduroglu K, Alanay Y and Turan S (2014) Cathepsin K analysis in a pycnodysostosis cohort: demographic, genotypic and phenotypic features. *Orphanet Journal of Rare Diseases*. 9, 60.

Arnaud-López L, Fragoso R, Mantilla-Capacho J and Barros-Núñez P (2007) Crouzon with acanthosis nigricans. Further delineation of the syndrome. *Clinical Genetics*. 72 (5), 405–10.

Atchley WR and Fitch WM (1997) A natural classification of the basic helix – loop – helix class of transcription factors. *Proceedings of the National Academy of Sciences of the USA.* 94 (10), 5172–5176.

Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Denise K, Morrison H, Plajzer-frick I, Holt A, Hosseini R, Akiyama J a, Shoukry M, Afzal V, Edward M, Fitzpatrick DR, Ren B, Hallgrímsson B, Pennacchio LA and Visel A (2013) Fine Tuning of Craniofacial Morphology by Distant-Acting Developmental Enhancers. *Science*. 342 (6157), 1–20.

Babbs C, Lloyd D, Pagnamenta AT, Twigg SRF, Green J, McGowan SJ, Mirza G, Naples R, Sharma VP, Volpi E V, Buckle VJ, Wall S a, Knight SJL, Parr JR and Wilkie AOM (2014) De novo and rare inherited mutations implicate the transcriptional coregulator TCF20/SPBP in autism spectrum disorder. *Journal of Medical Genetics*. 51 (11), 737–47.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA and Shendure J (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews. Genetics*. 12 (11), 745–55.

Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, Bergmann C, Rohrbach M, Koerber F, Zimmermann K, de Vries P, Wirth B, Schoenau E, Wollnik B, Veltman J a, Hoischen A and Netzer C (2011) Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *American Journal of Human Genetics*. 88 (3), 362–71.

Beenken A and Mohammadi M (2009) The FGF family: biology, pathophysiology and therapy. *Nature reviews. Drug discovery*. 8 (3), 235–53.

Bellus GA, Gaudenz K, Zackai EH, Clarke LA, Szabo J, Francomano CA MM (1996) Identical mutations in three different fibroblast growth factor receptor genes in autosomal dominant craniosynostosis syndromes. *Nature Genetics* 14 (2), 174–6.

Bialek P, Kern B, Yang X, Schrock M, Sosic D, Hong N, Wu H, Yu K, Ornitz DM, Olson EN, Justice MJ, Karsenty G, Jolla L and Louis S (2004) A Twist Code Determines the Onset of Osteoblast Differentiation. *Developmental cell.* 6 (3) 423–435.

Bochukova EG, Soneji S, Wall SA and Wilkie AOM (2010) Scalp fibroblasts have a shared expression profile in monogenic craniosynostosis. *Journal of Medical Genetics*. 47 (12), 803–8.

Botstein D and Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*. 33 Suppl (march), 228–37.

Boulet SL, Rasmussen SA and Honein M a (2008) A population-based study of craniosynostosis in metropolitan Atlanta, 1989-2003. *American Journal of Medical Genetics. Part A*. 146A (8), 984–91.

Boyadjiev SA (2007) Genetic analysis of non-syndromic craniosynostosis. *Orthodontics & Craniofacial Research*. 10 (3), 129–37.

Branda CS, Dymecki SM and Cre- S (2004) Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice. *Developmental cell*. 6 (1), 7–28.

Braunstein M and Anderson MK (2012) HEB in the spotlight: Transcriptional regulation of T-cell specification, commitment, and developmental plasticity. *Clinical & Developmental Immunology*. 2012, 678705.

Bruneteau RJ and Mulliken JB (1992) Frontal plagiocephaly: synostotic, compensational, or deformational. *Plastic and Reconstructive Surgery*. 89(1), 21-31.

Bush WS and Moore JH (2012) Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*. 8 (12).

Cai J, Goodman BK, Patel AS, Mulliken JB, Van Maldergem L, Hoganson GE, Paznekas W a, Ben-Neriah Z, Sheffer R, Cunningham ML, Daentl DL and Jabs EW (2003) Increased risk for developmental delay in Saethre-Chotzen syndrome

is associated with TWIST deletions: an improved strategy for TWIST mutation screening. *Human Genetics*. 114 (1), 68–76.

Carver EA, Oram KF and Gridley T (2002) Craniosynostosis in Twist heterozygous mice: a model for Saethre-Chotzen syndrome. *The Anatomical record*. 268 (2), 90–2.

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

Chai Y, Jiang X, Ito Y, Bringas P, Han J, Rowitch DH, Soriano P, McMahon AP and Sucov HM (2000) Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development*. 127 (8), 1671–9.

Chen J, Li X-B, Su R, Song L, Wang F and Zhang J-W (2014) ZNF16 (HZF1) promotes erythropoiesis and megakaryocytopoiesis via regulation of the c- *KIT* gene. *Biochemical Journal*. 458 (1), 171–183.

Chen ZF and Behringer RR (1995) Twist Is Required in Head Mesenchyme for Cranial Neural Tube Morphogenesis. *Genes & Development*. 9 (6), 686–699.

Cheung VG, Nayak RR, Wang IX, Elwyn S, Cousins SM, Morley M and Spielman RS (2010) Polymorphic cis- and trans-regulation of human gene expression. *PLoS biology*. 8 (9).

Cheung VG and Spielman RS (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nature reviews: Genetics*. 10 (9), 595–604.

Chotzen F (1932) Eine eigenartige familiaere Entwicklungsstoerung (Akrocephalosyndaktylie, Dysostosis craniofacialis und Hypertelorismus). *Mschr. Kinderheilk.*. 55, 97–122.

Chun K, Teebi AS, Jung JH, Kennedy S, Laframboise R, Meschino WS, Nakabayashi K, Scherer SW, Ray PN and Teshima I (2002) Genetic analysis of patients with the Saethre-Chotzen phenotype. *American Journal of Medical Genetics*. 110 (2), 136–43.

Chun S and Fay JC (2009) Identification of deleterious mutations within three human genomes. *Genome Research*. 19 (9), 1553–61.

Connerney J, Andreeva V, Leshem Y, Muentener C, Mercado M a and Spicer DB (2006) Twist1 dimer selection regulates cranial suture patterning and fusion. *Developmental Dynamics*. 235 (5), 1345–57.

Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C and Kehrer-Sawatzki H (2013) Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics.* 132(10): 1077-130.

Cooper GM, Stone EA, Asimenos G, Comparative N, Program S, Green ED, Batzoglou S and Sidow A (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research* 15(7), 901–913.

Coussens AK, Wilkinson CR, Hughes IP, Morris CP, van Daal A, Anderson PJ and Powell BC (2007) Unravelling the molecular control of calvarial suture fusion in children with craniosynostosis. *BMC Genomics*. 8, 458.

Creyghton MP, Roël G, Eichhorn PJ, Hijmans EM, Maurer I, Destrée O, Bernards R and R. (2005) PR72, a novel regulator of Wnt signaling required for Naked cuticle function. *Genes & Development*. 19(3): 376–386.

Davydov E V, Goode DL, Sirota M, Cooper GM, Sidow A and Batzoglou S (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS computational biology*. 6 (12), e1001025.

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

Deckelbaum RA, Majithia A, Booker T, Henderson JE and Loomis C a (2006) The homeoprotein engrailed 1 has pleiotropic functions in calvarial intramembranous bone formation and remodelling. *Development*. 133 (1), 63–74.

Doyle AJ, Doyle JJ, Bessling SL, Maragh S, Lindsay ME, Schepers D, Gillis E, Mortier G, Homfray T, Sauls K, Norris R a, Huso ND, Leahy D, Mohr DW, Caulfield MJ, Scott AF, Destrée A, Hennekam RC, Arn PH, Curry CJ, Van Laer L, McCallion AS, Loeys BL and Dietz HC (2012) Mutations in the TGF-β repressor SKI cause Shprintzen-Goldberg syndrome with aortic aneurysm. *Nature Genetics*. 44 (11), 1249–54.

van den Elzen MEP, Twigg SRF, Goos J AC, Hoogeboom AJM, van den Ouweland AMW, Wilkie AOM and Mathijssen IJ (2013) Phenotypes of craniofrontonasal syndrome in patients with a pathogenic mutation in EFNB1. *European Journal of Human Genetics.* 22(8): 995-1001.

Emison ES, Garcia-Barcelo M, Grice E a, Lantieri F, Amiel J, Burzynski G, Fernandez RM, Hao L, Kashuk C, West K, Miao X, Tam PKH, Griseri P, Ceccherini I, Pelet A, Jannot A-S, de Pontual L, Henrion-Caude A, Lyonnet S, Verheij JBGM, Hofstra RMW, Antiñolo G, Borrego S, McCallion AS and Chakravarti A (2010) Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *American Journal of Human Genetics*. 87 (1), 60–74.

Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED and Chakravarti A (2005) A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature*. 434 (7035), 857–63.

Fallin D and Schork NJ (2000) Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *American Journal of Human Genetics*. 67 (4), 947–959.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Davis BT, Baxter RM, Zeng W, Mroske C, Parra MC, Gandomi SK, Lu I, Li X, Lu H, Lu H-M, Salvador D, Ruble D, Lao M, Fischbach S, Wen J, Lee S, Elliott A, Dunlop CLM and Tang S (2014) Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine*. 17(7), 578-86.

Filges I, Nosova E, Bruder E, Tercanli S, Townsend K, Gibson W, Röthlisberger B, Heinimann K, Hall J, Gregory-Evans C, Wasserman W, Miny P and Friedman

J (2013) Exome sequencing identifies mutations in KIF14 as a novel cause of an autosomal recessive lethal fetal ciliopathy phenotype. *Clinical Genetics*. 86(3), 220–228.

Fitzpatrick DR (2013) Filling in the gaps in cranial suture biology. *Nature Genetics*. 45 (3), 231–2.

Florisson JMG, Verkerk AJMH, Huigh D, Hoogeboom AJM, Swagemakers S, Kremer A, Heijsman D, Lequin MH, Mathijssen IMJ and van der Spek PJ (2013) Boston type craniosynostosis: report of a second mutation in MSX2. *American Journal of Medical Genetics. Part A*. 161 (10), 2626–33.

Flück CE, Tajima T, Pandey A V, Arlt W, Okuhara K, Verge CF, Jabs EW, Mendonça BB, Fujieda K and Miller WL (2004) Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nature Genetics*. 36 (3), 228–30.

Forrest CR and Hopper RA (2013) Craniofacial syndromes and surgery. *Plastic and Reconstructive Surgery*. 131 (1), 86e–109e.

Frazer KA, Murray SS, Schork NJ and Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nature reviews. Genetics*. 10 (4), 241–251.

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson D a, Bamshad MJ and Akey JM (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. 493 (7431), 216–20.

Gelb BD, Shi GP, Chapman HA DR (1996) Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. *Science*. 273 (5279), 1236–8.

el Ghouzzi V, Le Merrer M, Perrin-Schmitt F, Lajeunie E, Benit P, Renier D, Bourgeois P, Bolcato-Bellemin AL, Munnich A BJ (1997) Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nature Genetics* 15 (1), 42–6.

Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV a M, Veltman J a and Brunner HG (2010) Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *American Journal of Human Genetics*. 87 (3), 418–23.

Gilissen C, Hoischen A, Brunner HG and Veltman JA (2012) Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics*. 20 (5), 490–7.

Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, Thibodeau P, Bachand I, Bao JYJ, Tong AHY, Lin C-H, Millet B, Jaafari N, Joober R, Dion P a, Lok S, Krebs M-O and Rouleau GA (2011) Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature Genetics*. 43 (9), 860–3.

Goriely A and Wilkie AOM (2012) Paternal age effect mutations and selfish spermatogonial selection: Causes and consequences for human disease. *American Journal of Human Genetics*. 90 (2), 175–200.

Graham JM, deSaxe M and Smith DW (1979) Sagittal craniostenosis: fetal head constraint as one possible cause. *The Journal of Pediatrics*. 95 (5 Pt 1), 747–50.

Graham JM and Smith DW (1980) Metopic Craniostenosis as a Consequence of Fetal Head Constraint: Two Interesting Experiments of Nature. *Pediatrics*. 65 (5), 1000–1002.

Greenwood J, Flodman P, Osann K, Boyadjiev SA and Kimonis V (2013) Familial incidence and associated symptoms in a population of individuals with nonsyndromic craniosynostosis. *Genetics in medicine*. 16(4): 302-10.

Gripp K, Zackai E and Stolle C (2000) Mutations in the human TWIST gene. *Human Mutation*. 15 (5), 479.

Guo Y, Long J, He J, Li C-I, Cai Q, Shu X-O, Zheng W and Li C (2012) Exome sequencing generates high quality data in non-target regions. *BMC Genomics*. 13 (1), 194.

Hajihosseini MK, Duarte R, Pegrum J, Donjacour A, Lana-Elola E, Rice DP, Sharpe J and Dickson C (2009) Evidence that Fgf10 contributes to the skeletal and visceral defects of an Apert syndrome mouse model. *Developmental Dynamics*. 238 (2), 376–85.

Hajihosseini MK, Wilson S, De Moerlooze L and Dickson C (2001) A splicing switch and gain-of-function mutation in FgfR2-IIIc hemizygotes causes Apert/Pfeiffer-syndrome-like phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*. 98 (7), 3855–60.

Hammond P and Suttie M (2012) Large-scale objective phenotyping of 3D facial morphology. *Human Mutation*. 33 (5), 817–825.

Hamosh A, Scott AF, Amberger JS, Bocchini CA and McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*. 33 (Database issue), D514–7.

Harel T, Pehlivan D, Caskey T and Lupski J (2014) Mendelain, Non-mendelian, Multigenic Inheritance, and Epigenetics. In: R. Rosenberg & J. Pascual eds. *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease, 5th Edition*. Elsevier.

Hebrok M, Wertz K and Füchtbauer E-M (1994) M-twist Is an Inhibitor of Muscle Differentiation. *Developmental Biology*. 165 (2), 537–544.

de Heer IM, de Klein A, van den Ouweland AM, Vermeij-Keers C, Wouters CH, Vaandrager JM, Hovius SER and Hoogeboom JM (2005) Clinical and Genetic Analysis of Patients with Saethre-Chotzen Syndrome. *Plastic and Reconstructive Surgery*. 115 (7), 1894–1902.

Ho TV, Iwata J, Ho HA, Grimes WC, Park S, Sanchez-Lara PA and Chai Y (2014) Integration of comprehensive 3D microCT and signalling analysis reveals differential regulatory mechanisms of craniofacial bone development. *Developmental Biology*. 400 (2), 180–190.

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BBA and Veltman JA (2010) De novo

mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature Genetics*. 42 (6), 483–5.

Hoischen A, van Bon BWM, Rodríguez-Santiago B, Gilissen C, Vissers LELM, de Vries P, Janssen I, van Lier B, Hastings R, Smithson SF, Newbury-Ecob R, Kjaergaard S, Goodship J, McGowan R, Bartholdi D, Rauch A, Peippo M, Cobben JM, Wieczorek D, Gillessen-Kaesbach G, Veltman J a, Brunner HG and de Vries BB (2011) De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nature Genetics*. 43 (8), 729–31.

Holder N and Klein R (1999) Eph receptors and ephrins: effectors of morphogenesis. *Development*. 126 (10), 2033–44.

Hollinger, Jeffrey O., Einhorn, Thomas A., Doll, Bruce, Sfeir C (2004) Bone Tissue Engineering. *CRC Press*, 17–20.

Howard, Timothy D, Paznekas WIlliam A, Green Eric D, Chiang Lydia C, Ma Nancy, Ortiz de Luna Rosa I, Delgado, Constanza G, Gonzalez-Ramos M, Kline Antoine D, Jabs EW (1997) Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nature Genetics.* 15 (1), 36–41.

Hu JS, Olson EN and Kingston RE (1992) HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Molecular and Cellular Biology*. 12 (3), 1031–42.

Huang N, Pandey A V, Agrawal V, Reardon W, Lapunzina PD, Mowat D, Jabs EW, Van Vliet G, Sack J, Flück CE and Miller WL (2005) Diversity and function of mutations in p450 oxidoreductase in patients with Antley-Bixler syndrome and

disordered steroidogenesis. *American Journal of Human Genetics*. 76 (5), 729–49.

Hurst JA, Jenkins D, Vasudevan PC, Kirchhoff M, Skovby F, Rieubland C, Gallati S, Rittinger O, Kroisel PM, Johnson D, Biesecker LG and Wilkie AOM (2011) Metopic and sagittal synostosis in Greig cephalopolysyndactyly syndrome: five cases with intragenic mutations or complete deletions of GLI3. *European Journal of Human Genetics*. 19 (7), 757–62.

Ingber DE (2010) From Cellular Mechanotransduction to Biologically Inspired Engineering. *Annals of Biomedical Engineering*. 38 (3), 1148–1161.

International HapMap Consortium H (2005) A haplotype map of the human genome. *Nature*. 437 (7063), 1299–320.

Iseki S, Wilkie AO, Heath JK, Ishimaru T, Eto K and Morriss-Kay GM (1997) Fgfr2 and osteopontin domains in the developing skull vault are mutually exclusive and can be altered by locally applied FGF2. *Development*. 124 (17), 3375–84.

Iseki S, Wilkie AO and Morriss-Kay GM (1999) Fgfr1 and Fgfr2 have distinct differentiation- and proliferation-related roles in the developing mouse skull vault. *Development (Cambridge, England)*. 126 (24), 5611–20.

Jabs EW, Müller U, Li X, Ma L, Luo W, Haworth IS, Klisak I, Sparkes R, Warman ML and Mulliken JB (1993) A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell*. 75 (3), 443–50.

Jabs EW, Li X, Scott AF, Meyers G, Chen W, Eccles M, Mao JI, Charnas LR, Jackson CE JM (1994) Jackson-Weiss and Crouzon syndromes are allelic with mutations in fibroblast growth factor receptor 2. *Nature Genetics*. 8 (3), 275–9.

Janssen A, Hosen MJ, Jeannin P, Coucke PJ, De Paepe A and Vanakker OM (2013) Second family with the Boston-type craniosynostosis syndrome: novel mutation and expansion of the clinical spectrum. *American Journal of Medical Genetics. Part A*. 161 (9), 2352–7.

Jenkins D, Baynam G, Catte L De, Elcioglu N and Gabbett MT (2011). Carpenter Syndrome: Extended RAB23 Mutation Spectrum and Analysis of Nonsense-mediated mRNA Decay. 32(4) E2069–2078.

Jenkins D, Seelow D, Jehee FS, Perlyn C a, Alonso LG, Bueno DF, Donnai D, Josifova D, Josifiova D, Mathijssen IMJ, Morton JE V, Orstavik KH, Sweeney E, Wall S a, Marsh JL, Nurnberg P, Passos-Bueno MR and Wilkie AOM (2007) RAB23 mutations in Carpenter syndrome imply an unexpected role for hedgehog signaling in cranial-suture development and obesity. *American Journal of Human Genetics*. 80 (6), 1162–70.

Jiang X, Iseki S, Maxson RE, Sucov HM and Morriss-Kay GM (2002) Tissue origins and interactions in the mammalian skull vault. *Developmental Biology*. 241 (1), 106–16.

Johnson D (2003) Hunterian Lecture A comprehensive screen of genes implicated in craniosynostosis. *Annals of the Royal College of Surgeons of England.* 85(6), 371–377.

Johnson D, Horsley SW, Moloney DM, Oldridge M, Twigg SR, Walsh S, Barrow M, Njølstad PR, Kunz J, Ashworth GJ, Wall SA, Kearney L and Wilkie AO (1998) A comprehensive screen for TWIST mutations in patients with craniosynostosis identifies a new microdeletion syndrome of chromosome band 7p21.1. *American Journal of Human Genetics*. 63 (5), 1282–93.

Johnson D, Iseki S, Wilkie AO and Morriss-Kay GM (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. *Mechanisms of Development*. 91 (1-2), 341–5.

Johnson D and Wilkie AOM (2011) Craniosynostosis. *European Journal of Human Genetics.* 19 (4), 369–76.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves I a, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG and Todd J a (2001) Haplotype tagging for the identification of common disease genes. *Nature Genetics*. 29 (2), 233–237.

Kamath BM, Stolle C, Bason L, Colliton RP, Piccoli D a, Spinner NB and Krantz ID (2002) Craniosynostosis in Alagille syndrome. *American Journal of Medical Genetics*. 112 (2), 176–80.

Kamath KS, Kumar SS, Kaur J, Venkatakrishnan V, Paulsen IT, Nevalainen H and Molloy MP (2015) Proteomics of hosts and pathogens in cystic fibrosis. *Proteomics - Clinical Applications*. 9 (1-2), 134–146.

Kee BL (2009) E and ID proteins branch out. *Nature reviews. Immunology*. 9 (3), 175–84.

Kelly SA, Bell TA, Selitsky SR, Buus RJ, Hua K, Weinstock GM, Garland T, de Villena FPM and Pomp D (2013) A novel intronic single nucleotide polymorphism in the Myosin heavy polypeptide 4 gene is responsible for the mini-muscle phenotype characterized by major reduction in hind-limb muscle mass in mice. *Genetics*. 195 (4), 1385–1395.

Kerlin BA, Yan SB, Isermann BH, Brandt JT, Sood R, Basson BR, Joyce DE, Weiler H and Dhainaut JF (2003) Survival advantage associated with heterozygous factor V Leiden mutation in patients with severe sepsis and in mouse endotoxemia. *Blood*. 102 (9), 3085–3092.

Kiecker C and Lumsden A (2005) Compartments and their boundaries in vertebrate brain development. *Nature reviews. Neuroscience*. 6 (7), 553–64.

Kim HJ, Rice DP, Kettunen PJ and Thesleff I (1998) FGF-, BMP- and Shh-mediated signalling pathways in the regulation of cranial suture morphogenesis and calvarial bone development. *Development*. 125 (7), 1241–51.

Klopocki E, Lohan S, Brancati F, Koll R, Brehm A, Seemann P, Dathe K, Stricker S, Hecht J, Bosse K, Betz RC, Garaci FG, Dallapiccola B, Jain M, Muenke M, Ng VCW, Chan W, Chan D and Mundlos S (2011) Copy-number variations involving the IHH locus are associated with syndactyly and craniosynostosis. *American Journal of Human Genetics*. 88 (1), 70–5.

Koboldt D (2012) *The Current State of dbSNP*. [Online] Available at: http://massgenomics.org/2012/01/the-current-state-of-dbsnp.html (accessed 02/09/15).

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson S a., Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WSW, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U and Stefansson K (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 488 (7412), 471–475.

Kratz CP, Zampino G, Kriek M, Kant SG, Leoni C, Pantaleoni F, Oudesluys-Murphy AM, Di Rocco C, Kloska SP, Tartaglia M and Zenker M (2009) Craniosynostosis in patients with Noonan syndrome caused by germline KRAS mutations. *American Journal of Medical Genetics. Part A*. 149A (5), 1036–40.

Krawitz P, Rödelsperger C, Jäger M, Jostins L, Bauer S and Robinson PN (2010) Microindel detection in short-read sequence data. *Bioinformatics (Oxford, England)*. 26 (6), 722–9.

Kryukov G V, Pennacchio L a and Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American Journal of Human Genetics*. 80 (4), 727–39.

Kühn S and Geyer M (2014) Formins as effector proteins of Rho GTPases. *Small GTPases*. 5 (June), 1–16.

Kumar P, Henikoff S and Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 4 (7), 1073–81.

Laestander C and Engström W (2014) Role of fibroblast growth factors in elicitation of cell responses. *Cell Proliferation*. 47 (1), 3–11.

Lajeunie E, Heuertz S, El Ghouzzi V, Martinovic J, Renier D, Le Merrer M and Bonaventure J (2006) Mutation screening in patients with syndromic craniosynostoses indicates that a limited number of recurrent FGFR2 mutations accounts for severe forms of Pfeiffer syndrome. *European Journal of Human Genetics*. 14 (3), 289–98.

Lajeunie E, Le Merrer M, Bonaïti-Pellie C, Marchac D and Renier D (1995) Genetic study of nonsyndromic coronal craniosynostosis. *American Journal of Medical Genetics*. 55 (4), 500–4.

Lajeunie E, Cameron R, El Ghouzzi V, de Parseval N, Journeau P, Gonzales M, Delezoide AL, Bonaventure J, Le Merrer M RD (1999) Clinical variability in patients with Apert's syndrome. *Journal of Neurosurgery*. 90 (3), 443–447.

Lancerotto L and Orgill DP (2014) Mechanoregulation of Angiogenesis in Wound Healing. *Advances in wound care*. 3 (10), 626–634.

Langmead B, Trapnell C, Pop M and Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 10 (3), R25.

Laue K, Pogoda H-M, Daniel PB, van Haeringen A, Alanay Y, von Ameln S, Rachwalski M, Morgan T, Gray MJ, Breuning MH, Sawyer GM, Sutherland-Smith AJ, Nikkels PG, Kubisch C, Bloch W, Wollnik B, Hammerschmidt M and Robertson SP (2011) Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. *American Journal of Human Genetics*. 89 (5), 595–606.

Ledent V, Paquet O and Vervoort M (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins. *Genome Biology*. 3 (6), RESEARCH0030.

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E N and S N (2014) Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 312 (18), 1880–7.

Lee HQ, Hutson JM, Wray AC, Lo PA, Chong DK, Holmes AD and Greensmith AL (2012) Changing Epidemiology of Nonsyndromic Craniosynostosis and Revisiting the Risk Factors. *Journal of Craniofacial Surgery*. 23 (5), 1245–1251.

Legué E and Joyner AL (2010) *Genetic fate mapping using site-specific recombinases.* (2nd edition). Elsevier Inc.

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

Liu B, Chen S, Cheng D, Jing W and Helms J a (2014) Primary cilia integrate hedgehog and Wnt signaling during tooth development. *Journal of Dental Research*. 93 (5), 475–82.

Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, Meyers J, Leitch CC, Katsanis N, Sharifi N, Xu FL, Myers L a, Spevak PJ, Cameron DE, De Backer J, Hellemans J, Chen Y, Davis EC, Webb CL, Kress W, Coucke P, Rifkin DB, De Paepe AM and Dietz HC (2005) A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2. *Nature Genetics*. 37 (3), 275–81.

LopezJimenez N, Gerber S, Popovici V, Mirza S, Copren K, Ta L, Shaw GM, Trueb B and Slavotinek AM (2010) Examination of FGFRL1 as a candidate gene for diaphragmatic defects at chromosome 4p16.3 shows that Fgfrl1 null mice have reduced expression of Tpm3, sarcomere genes and Lrtm1 in the diaphragm. *Human Genetics*. 127 (3), 325–36.

Maia AM, da Silva JH, Mencalha AL, Caffarena ER and Abdelhay E (2012) Computational modelling of the bHLH domain of the transcription factor TWIST1 and R118C, S144R and K145E mutants. *BMC Bioinformatics*. 13, 184.

Van Maldergem L, Siitonen H a, Jalkh N, Chouery E, De Roy M, Delague V, Muenke M, Jabs EW, Cai J, Wang LL, Plon SE, Fourneau C, Kestilä M, Gillerot Y, Mégarbané A and Verloes A (2006) Revisiting the craniosynostosis-radial ray hypoplasia association: Baller-Gerold syndrome caused by mutations in the RECQL4 gene. *Journal of Medical Genetics*. 43 (2), 148–52.

Marucci DD, Johnston CP, Anslow P, Jayamohan J, Richards PG, Wilkie AOM and Wall S a (2008) Implications of a vertex bulge following modified strip craniectomy for sagittal synostosis. *Plastic and Reconstructive Surgery*. 122 (1), 217–24.

Massari ME and Murre C (2000) Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology* 20 (2): 429-40.

Mathijssen IM, van Leeuwen H, Vermeij-Keers C and Vaandrager JM (2001) FGF-4 or FGF-2 administration induces apoptosis, collagen type I expression, and mineralization in the developing coronal suture. *The Journal of Craniofacial Surgery*. 12 (4), 399–400.

Merrill AE, Bochukova EG, Brugger SM, Ishii M, Pilz DT, Wall S a, Lyons KM, Wilkie AOM and Maxson RE (2006) Cell mixing at a neural crest-mesoderm boundary and deficient ephrin-Eph signalling in the pathogenesis of craniosynostosis. *Human Molecular Genetics*. 15 (8), 1319–28.

van der Meulen J, van der Hulst R, van Adrichem L, Arnaud E, Chin-Shong D, Duncan C, Habets E, Hinojosa J, Mathijssen I, May P, Morritt D, Nishikawa H, Noons P, Richardson D, Wall S, van der Vlugt J and Renier D (2009) The Increase of Metopic Synostosis. *Journal of Craniofacial Surgery*. 20 (2), 283–286.

Meyers GA, Orlow SJ, Munro IR, Przylepa KA and Jabs EW (1995) Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation in Crouzon syndrome with acanthosis nigricans. *Nature Genetics*. 11 (4), 462–4.

Moloney DM, Slaney SF, Oldridge M, Wall SA, Sahlin P, Stenman G WA (1996) Exclusive paternal origin of new mutations in Apert syndrome. *Nature Genetics*. 13 (1), 48–53.

Momand JR, Xu G and Walter C a. (2013) The Paternal Age Effect: A Multifaceted Phenomenon. *Biology of Reproduction*. 88 (4), 108–108.

Mornet E (2007) Hypophosphatasia. *Orphanet journal of rare diseases*. 2, 40.

Mornet E, Yvard A, Taillandier A, Fauvert D and Simon-Bouy B (2011) A molecular-based estimation of the prevalence of hypophosphatasia in the European population. *Annals of Human Genetics*. 75 (3), 439–45.

Morriss-Kay GM and Wilkie AOM (2005) Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. *Journal of Anatomy*. 207 (5), 637–53.

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR and de Crombrugghe B (2002) The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell*. 108 (1), 17–29.

Nelson MR, Wegmann D, Ehm MG, Kessner D, St. Jean P, Verzilli C, Shen J, Tang Z, Bacanu S -a., Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zollner S, Whittaker JC, Chissoe SL, Novembre J and Mooser V (2012) An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science*. 337 (6090), 100–104.

Ng PC and Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*. 31 (13), 3812–3814.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K-I, Matsumoto N, Ohta T, Niikawa N, Nickerson D a, Bamshad MJ and Shendure J (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics*. 42 (9), 790–3.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson D a, Shendure J, Bamshad MJ and Deborah A (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*. 42 (1), 30–5.

Nie X, Luukko K and Kettunen P (2006) FGF signalling in craniofacial development and developmental disorders. *Oral Diseases*. 12 (2), 102–11.

Nieminen P, Morgan N V., Fenwick AL, Parmanen S, Veistinen L, Mikkola ML, van der Spek PJ, Giraud A, Judd L, Arte S, Brueton LA, Wall SA, Mathijssen IMJ, Maher ER, Wilkie AOM, Kreiborg S and Thesleff I (2011) Inactivation of IL11 Signaling Causes Craniosynostosis, Delayed Tooth Eruption, and Supernumerary Teeth. *American Journal of Human Genetics*. 89 (1), 67–81.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson D a, Bernier R, Fisher SE, Shendure J and Eichler EE (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*. 43 (6), 585–9.

O'Rourke MP and Tam PPL (2002) Twist functions in mouse development. *The International Journal of Developmental Biology*. 46 (4), 401–13.

Opperman LA (2000) Cranial Sutures as Intramembranous Bone Growth Sites. *Developmental Dynamics*. 485 (September), 472–485.

Ornitz DM and Marie PJ (2002) FGF signalling pathways in endochondral and intramembranous bone development and human genetic disease. *Genes & Development*. 16 (12), 1446–65.

Palacios R, Gazave E, Goñi J, Piedrafita G, Fernando O, Navarro A and Villoslada P (2009) Allele-Specific Gene Expression Is Widespread Across the Genome and Biological Processes. *PLoS ONE*. 4 (1), e4150.

Palmer L (2007) *Analysis of Single Nucleotide Polymorphism Haplotypes*. Henry Stewart Talks

Park L (2015) Ancestral Alleles in the Human Genome Based on Population Sequencing Data. *Plos One*. 10 (5), e0128186.

Paten B, Herrero J, Fitzgerald S, Beal K, Flicek P, Holmes I and Birney E (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Research*. 18 (11), 1829–1843.

Paumard-Hernández B, Berges-Soria J, Barroso E, Rivera-Pedroza CI, Pérez-Carrizosa V, Benito-Sanz S, López-Messa E, Santos F, García-Recuero II, Romance A, Ballesta-Martínez JM, López-González V, Campos-Barros A, Cruz J, Guillén-Navarro E, Sánchez Del Pozo J, Lapunzina P, García-Miñaur S and Heath KE (2014) Expanding the mutation spectrum in 182 Spanish probands

with craniosynostosis: identification and characterization of novel *TCF12* variants. *European Journal of Human Genetics*. 23(7), 907-14.

Percival CJ and Richtsmeier JT (2011) The Epigenetics of Dysmorphology: In: *Epigenetics*. Linking Genotype and Phenotype in Development and Evolution. University of California Press. 377–397.

Piard J, Rozé V, Czorny A, Lenoir M, Valduga M, Fenwick AL, Wilkie AOM and Maldergem L Van (2015) *TCF12* microdeletion in a 72-year-old woman with intellectual disability. *American Journal of Medical Genetics Part A*. 167 (8), 1897–1901.

Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE and King M-C (2010) Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *American Journal of Human Genetics*. 87 (2), 282–8.

Pollard KS, Hubisz MJ, Rosenbloom KR and Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*. 20 (1), 110–21.

Przylepa KA, Paznekas W, Zhang M, Golabi M, Bias W, Bamshad MJ, Carey JC, Hall BD, Stevenson R, Orlow S, Cohen MM Jr JE (1996) Fibroblast growth factor receptor 2 mutations in Beare-Stevenson cutis gyrata syndrome. *Nature Genetics*. 13 (4), 492–4.

Pulleyn LJ, Reardon W, Wilkes D, Rutland P, Jones BM, Hayward R, Hall CM, Brueton L, Chun N, Lammer E, Malcolm S WR (1996) Spectrum of craniosynostosis phenotypes associated with novel mutations at the fibroblast

growth factor receptor 2 locus. *European Journal of Human Genetics*. 4 (5), 283–291.

Rabbani B, Tekin M and Mahdieh N (2014) The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*. 59 (1), 5–15.

Reardon W, Winter RM, Rutland P, Pulleyn LJ, Jones BM MS (1994) Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. *Nature Genetics*. 8 (1), 8–103.

Rice DP (2008) Developmental anatomy of craniofacial sutures. *Frontiers of Oral Biology*. 12, 1–21.

Rice DP, Aberg T, Chan Y, Tang Z, Kettunen PJ, Pakarinen L, Maxson RE and Thesleff I (2000) Integration of FGF and TWIST in calvarial bone and suture development. *Development*. 127 (9), 1845–55.

Rice DPC, Rice R and Thesleff I (2003) Molecular mechanisms in calvarial bone and suture development, and their relation to craniosynostosis. *European Journal of Orthodontics*. 25 (2), 139–48.

Rice, D. P. and Rice R (2013) Craniofacial Intramembranous Bone Development and Regeneration. In: G. T.-J. Huang & I. Thesleff eds. *Stem Cells in Craniofacial Development and Regeneration*. Hoboken, NJ, USA: John Wiley & Sons, Inc. 51–69.

Richtsmeier JT and Flaherty K (2013) Hand in glove: brain and skull in development and dysmorphogenesis. *Acta Neuropathologica*. 125 (4), 469–89.

di Rocco F, Baujat G, Arnaud E, Rénier D, Laplanche J-L, Daire VC and Collet C (2014) Clinical spectrum and outcomes in families with coronal synostosis and TCF12 mutations. *European Journal of Human Genetics.* 22 (12), 1413–6.

Roscioli T, Flanagan S, Kumar P, Masel J, Gattas M, Hyland VJ GI (2000) linical findings in a patient with FGFR1 P252R mutation and comparison with the literature. *American Journal of Medical Genetics*. 93 (1), 22–8.

Rosen H, Andrews BT, Meara JG, Stoler JM, Mulliken JB and Rogers GF (2011) Audiologic findings in Saethre-Chotzen syndrome. *Plastic and Reconstructive Surgery*. 127 (5), 2014–20.

Roses AD (2000) Pharmacogenetics and the practice of medicine. *Nature*. 405(6788): 857-65.

Rosset S, Tzur S, Behar DM, Wasser WG and Skorecki K (2011) The population genetics of chronic kidney disease: insights from the MYH9-APOL1 locus. *Nature reviews. Nephrology*. 7 (6), 313–326.

Roxo-Rosa M, Davezac N, Bensalem N, Majumder M, Heda GD, Simas A, Penque D, Amaral MD, Lukacs GL and Edelman A (2004) Proteomics techniques for cystic fibrosis research. *Journal of Cystic Fibrosis*. 3 (SUPPL. 2), 85–89.

Ruffalo M, Laframboise T and Koyut M (2011) Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* 27(20): 2790-6.

Ruiz-Ferrer M, Fernández RM, Antiñolo G, López-Alonso M, Eng C and Borrego S (2006) A complex additive model of inheritance for Hirschsprung disease is supported by both RET mutations and predisposing RET haplotypes. *Genetics in Medicine*. 8 (11), 704–710.

Sood S, Eldadah ZA, Krause WL, McIntosh I DH (1996) Mutation in fibrillin-1 and the Marfanoid-craniosynostosis (Shprintzen-Goldberg) syndrome. *Nature Genetics*. 12 (2), 209–11.

SABiosciences (2010) *PI3K-AKT Signaling Pathway Mutation PCR Array For detecting 87 mutations in the PI3K-AKT pathway*. [Online] Available at: http://www.sabiosciences.com/qbiomarker_product/HTML/SMH-014A.html (accessed 05/09/15).

Sablina AA., Hector M, Colpaert N and Hahn WC (2010) Identification of PP2A complexes and pathways involved in cell transformation. *Cancer Research*. 70 (24), 10474–10484.

Saethre H (1931) Ein Beitrag zum Turmschädel Problem (Pathogenese, Erblichkeit und Symptomatologie). *Deutsche Zeitschrift für Nervenheilkunde*. 117-119 (1), 533–555.

Sanchez-Lara PA, Carmichael SL, Graham JM, Lammer EJ, Shaw GM, Ma C and Rasmussen SA (2010) Fetal constraint as a potential risk factor for craniosynostosis. *American Journal of Medical Genetics. Part A*. 152A (2), 394–400.

Sanger F and Nicklen S (1977) DNA sequencing with chain-terminating inhibitors *Proceedings of the National Academy of Sciences of the United States of America*. 74 (12), 5463–5467.

Schwarz JM, Rödelsperger C, Schuelke M and Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*. 7 (8), 575–6.

Sciences) LG-GSLC (University of UH (2015) *Making SNPs Make Sense*. [Online] Available at: http://learn.genetics.utah.edu/content/pharma/snips/ (accessed 26/08/15).

Sharma VP, Fenwick AL, Brockop MS, McGowan SJ, Goos J a C, Hoogeboom AJM, 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, Johnson D, Wall S a, van der Spek PJ, Mathijssen IMJ, Maxson RE, Twigg SRF and Wilkie AOM (2013) Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nature Genetics*. 45 (3), 304–7.

Sharma VP, Wall S a, Lord H, Lester T and Wilkie AOM (2012) Atypical Crouzon Syndrome With a Novel Cys62Arg Mutation in FGFR2 Presenting With Sagittal Synostosis. *The Cleft Palate-Craniofacial Journal*. 49 (3), 373–7.

Shimoaka T, Ogasawara T, Yonamine A, Chikazu D, Kawano H, Nakamura K, Itoh N and Kawaguchi H (2002) Regulation of osteoblast, chondrocyte, and osteoclast functions by fibroblast growth factor (FGF)-18 in comparison with FGF-2 and FGF-10. *The Journal of Biological Chemistry*. 277 (9), 7493–500.

Shur N, Cowan J and Wheeler PG (2003) Craniosynostosis and congenital heart anomalies associated with a maternal deletion of 15q15-22.1. *American Journal of Medical Genetics. Part A*. 120A (4), 542–6.

Simpson MA, Hsu R, Keir LS, Hao J, Sivapalan G, Ernst LM, Zackai EH, Al-Gazali LI, Hulskamp G, Kingston HM, Prescott TE, Ion A, Patton M a, Murday V, George A and Crosby AH (2007) Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (Raine syndrome), highlighting a crucial molecule in bone development. *American Journal of Human Genetics*. 81 (5), 906–12.

Simpson P (1983) Maternal-Zygotic Gene Interactions during Formation of the Dorsoventral Pattern in Drosophila Embryos. *Genetics*. 105 (3), 615–32.

Singer S, Bower C, Southall P and Goldblatt J (1999) Craniosynostosis in Western Australia, 1980-1994: a population-based study. *American Journal of Medical Genetics*. 83 (5), 382–7.

Slager RE, Newton TL, Vlangos CN, Finucane B and Elsea SH (2003) Mutations in RAI1 associated with Smith-Magenis syndrome. *Nature Genetics*. 33 (4), 466–468.

Slaney SF, Oldridge M, Jane A, Hall CM, Poole MD and Morriss-kay GM (1996) Differential Effects of FGFR2 Mutations Palate in Apert Syndrome Syndactyly and Cleft. *American Journal of Human Genetics*. 58 (5), 923–932.

Slavotinek A, Crawford H, Golabi M, Tao C, Perry H, Oberoi S, Vargervik K and Friez M (2009) Novel FGFR2 deletion in a patient with Beare-Stevenson-

like syndrome. *American Journal of Medical Genetics. Part A.* 149A (8), 1814–7.

Spielman RS, McGinnis RE and Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*. 52 (3), 506–516.

Splendore A, Jabs EW, Félix TM and Passos-Bueno MR (2003) Parental origin of mutations in sporadic cases of Treacher Collins syndrome. *European Journal of Human Genetics.* 11 (9), 718–722.

Standring S (2009) *Gray's Anatomy, 40th Edition. The Anatomical Basis of Clinical Practice* (40th edition). S. Standring ed. Churchill Livingstone

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A and Lewis S (2002) The Generic Genome Browser: A Building Block for a Model Organism System Database. *Genome Research.* 12(10): 1599–1610.

Su N, Jin M and Chen L (2014) Role of FGF / FGFR signaling in skeletal development and homeostasis: learning from mouse models. *Bone Research*. 2 (April), 1–24.

Tabler J, Barrell W, Szabo-Rogers H, Healy C, Yeung Y, Perdiguero E, Schulz C, Yannakoudakis B, Mesbahi A, Wlodarczyk B, Geissmann F, Finnell R, Wallingford J and Liu KJ (2013) Fuz Mutant Mice Reveal Shared Mechanisms between Ciliopathies and FGF-Related Syndromes. *Developmental Cell*. 25 (6), 623–635.

Taniguchi K, Takeya R, Suetsugu S, Kan-o M, Narusawa M, Shiose A, Tominaga R and Sumimoto H (2009) Mammalian formin Fhod3 regulates actin assembly and sarcomere organization in striated muscles. *Journal of Biological Chemistry*. 284 (43), 29873–29881.

Le Tanno P, Poreau B, Devillard F, Vieville G, Amblard F, Jouk P-S, Satre V and Coutton C (2014) Maternal complex chromosomal rearrangement leads to *TCF12* microdeletion in a patient presenting with coronal craniosynostosis and intellectual disability. *American Journal of Medical Genetics. Part A*. 164A (6), 1530–6.

Thomas GPL, Wilkie AOM, Richards PG and Wall S a (2005) FGFR3 P250R mutation increases the risk of reoperation in apparent 'nonsyndromic' coronal craniosynostosis. *The Journal of Craniofacial Surgery*. 16 (3), 347–52; discussion 353–4.

Ting K, Vastardis H, Mulliken JB, Soo C, Tieu A, Do HUY, Kwong E, Bertolami CN, Kawamoto H, Kuroda SI and Longaker MT (1999) Human NELL-1 expressed in unilateral coronal synostosis. *Journal of Bone and Mineral Research*. 14 (1), 80–89.

Ting M-C, Wu NL, Roybal PG, Sun J, Liu L, Yen Y and Maxson RE (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.

Twigg SRF, Babbs C, van den Elzen MEP, Goriely A, Taylor S, McGowan SJ, Giannoulatou E, Lonie L, Ragoussis J, Sadighi Akha E, Knight SJL, Zechi-Ceide

RM, Hoogeboom J a M, Pober BR, Toriello H V, Wall S a, Rita Passos-Bueno M, Brunner HG, Mathijssen IMJ and Wilkie AOM (2013) Cellular interference in craniofrontonasal syndrome: males mosaic for mutations in the X-linked EFNB1 gene are more severely affected than true hemizygotes. *Human Molecular Genetics*. 22 (8), 1654–62.

Twigg SRF, Forecki J, Goos JAC, Richardson ICA, Hoogeboom AJM, van den Ouweland AMW, Swagemakers SMA, Lequin MH, Van Antwerp D, McGowan SJ, Westbury I, Miller KA, Wall SA, van der Spek PJ, Mathijssen IMJ, Pauws E, Merzdorf CS and Wilkie AOM (2015) Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. *The American Journal of Human Genetics*. 97 (3), 378–388.

Twigg SRF, Kan R, Babbs C, Bochukova EG, Robertson SP, Wall S a, Morriss-Kay GM and Wilkie AOM (2004) Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation, cause craniofrontonasal syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 101 (23), 8652–7.

Twigg SRF, Lloyd D, Jenkins D, Elçioglu NE, Cooper CDO, Al-Sannaa N, Annagür A, Gillessen-Kaesbach G, Hüning I, Knight SJL, Goodship J a, Keavney BD, Beales PL, Gileadi O, McGowan SJ and Wilkie AOM (2012) Mutations in multidomain protein MEGF8 identify a Carpenter syndrome subtype associated with defective lateralization. *American Journal of Human Genetics*. 91 (5), 897–905.

Twigg SRF, Matsumoto K, Kidd AMJ, Goriely A, Taylor IB, Fisher RB, Hoogeboom AJM, Mathijssen IMJ, Lourenco MT, Morton JE V, Sweeney E,

Wilson LC, Brunner HG, Mulliken JB, Wall SA and Wilkie AOM (2006) The origin of EFNB1 mutations in craniofrontonasal syndrome: frequent somatic mosaicism and explanation of the paucity of carrier males. *American Journal of Human Genetics*. 78 (6), 999–1010.

Twigg SRF, Vorgia E, McGowan SJ, Peraki I, Fenwick AL, Sharma VP, Allegra M, Zaragkoulias A, Sadighi Akha E, Knight SJL, Lord H, Lester T, Izatt L, Lampe AK, Mohammed SN, Stewart FJ, Verloes A, Wilson LC, Healy C, Sharpe PT, Hammond P, Hughes J, Taylor S, Johnson D, Wall S a, Mavrothalassitis G and Wilkie AOM (2013) Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis. *Nature Genetics*. 45 (3), 308–13.

Twigg SRF and Wilkie AOM (2015) A Genetic-Pathophysiological Framework for Craniosynostosis. *The American Journal of Human Genetics*. 97 (3), 359–377.

Twyman R (2003) *Haplotype mapping*. [Online] Available at: http://genome.wellcome.ac.uk/doc_WTD020781.html (accessed 26/08/15).

Uittenbogaard, M and Chiaramello A (2009) Expression of the bHLH transcription factor Tcf 12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain Research. Gene Expression Patterns*. 1 (November 2001), 115–121.

Vega H, Trainer AH, Gordillo M, Crosier M, Kayserili H, Skovby F, Uzielli MLG, Schnur RE, Manouvrier S, Blair E, Hurst J a, Forzano F, Meins M, Simola KOJ, Raas-Rothschild A, Hennekam RCM and Jabs EW (2010) Phenotypic

variability in 49 cases of ESCO2 mutations, including novel missense and codon deletion in the acetyltransferase domain, correlates with ESCO2 expression and establishes the clinical criteria for Roberts syndrome. *Journal of Medical Genetics*. 47 (1), 30–7.

Vega H, Waisfisz Q, Gordillo M, Sakai N, Yanagihara I, Yamada M, van Gosliga D, Kayserili H, Xu C, Ozono K, Jabs EW, Inui K and Joenje H (2005) Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nature Genetics*. 37 (5), 468–70.

Veistinen L, Aberg T and Rice DPC (2009) Convergent signalling through Fgfr2 regulates divergent craniofacial morphogenesis. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*. 312B (4), 351–60.

Veltman JA and Brunner HG (2012) De novo mutations in human genetic disease. *Nature reviews. Genetics*. 13 (8), 565–75.

Vissers LE, Cox TC, Maga AM, Short KM, Wiradjaja F, Janssen IM, Jehee F, Bertola D, Liu J, Yagnik G, Sekiguchi K, Kiyozumi D, van Bokhoven H, Marcelis C, Cunningham ML, Anderson PJ, Boyadjiev S a, Passos-Bueno MR, Veltman J a, Smyth I, Buckley MF and Roscioli T (2011) Heterozygous mutations of FREM1 are associated with an increased risk of isolated metopic craniosynostosis in humans and mice. *PLoS Genetics*. 7 (9), e1002278.

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BWM, Hoischen A, de Vries

BB a, Brunner HG and Veltman J a (2010) A *de novo* paradigm for mental retardation. *Nature Genetics*. 42 (12), 1109–12.

Vital Statistics Outputs Branch O for NS (2014) *Further Parental Characteristics, England and Wales, 2013*. [Online] Available at: http://www.ons.gov.uk/ons/rel/vsob1/further-parental-characteristics--england-and-wales/2013/index.html (accessed 26/08/15).

Vortkamp A, M Gessler Karl-Heinz G (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature Genetics* 352 (6335), 539–40.

Wang D, Claus CL, Vaccarelli G, Braunstein M, Schmitt TM, Zúñiga-Pflücker JC, Rothenberg E V and Anderson MK (2006) The basic helix-loop-helix transcription factor HEBAlt is expressed in pro-T cells and enhances the generation of T cell precursors. *Journal of Immunology*. 177 (1), 109–19.

Wang K, Li M and Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 38 (16), e164.

Weiss MJ, Cole DE, Ray K, Whyte MP, Lafferty MA, Mulivor R a and Harris H (1988) A missense mutation in the human liver/bone/kidney alkaline phosphatase gene causing a lethal form of hypophosphatasia. *Proceedings of the National Academy of Sciences of the United States of America*. 85 (20), 7666–9.

Wieacker P and Wieland I (2005) Clinical and genetic aspects of craniofrontonasal syndrome: towards resolving a genetic paradox. *Molecular Genetics and Metabolism*. 86 (1-2), 110–6.

Wieland I, Jakubiczka S, Muschke P, Cohen M, Thiele H, Gerlach KL, Adams RH and Wieacker P (2004) Mutations of the ephrin-B1 gene cause craniofrontonasal syndrome. *American Journal of Human Genetics*. 74 (6), 1209–15.

Wilkie AO, Tang Z, Elanko N, Walsh S, Twigg SR, Hurst J a, Wall S a, Chrzanowska KH and Maxson RE (2000) Functional haploinsufficiency of the human homeobox gene MSX2 causes defects in skull ossification. *Nature Genetics*. 24 (4), 387–90.

Wilkie AOM, Byren JC, Hurst JA, Jayamohan J, Johnson D, Knight SJL, Lester T, Richards PG, Twigg SRF and Wall SA (2010) Prevalence and complications of single-gene and chromosomal disorders in craniosynostosis. *Pediatrics*. 126 (2), e391–400.

Wojciechowski J, Lai A, Kondo M and Zhuang Y (2007) E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. *Journal of Immunology*. 178 (9), 5717–26.

Woods RH, Ul-Haq E, Wilkie AOM, Jayamohan J, Richards PG, Johnson D, Lester T and Wall S a (2009) Reoperation for intracranial hypertension in TWIST1-confirmed Saethre-Chotzen syndrome: a 15-year review. *Plastic and Reconstructive Surgery*. 123 (6), 1801–10.

Yamamoto T, Mencarelli MA, Di Marco C, Mucciolo M, Vascotto M, Balestri P, Gérard M, Mathieu-Dramard M, Andrieux J, Breuning M, Hoffer MJV, Ruivenkamp CAL, Shimada S, Sangu N, Shimojima K, Umezu R, Kawame H, Matsuo M, Saito K, Renieri A and Mari F (2014) Overlapping microdeletions

involving 15q22.2 narrow the critical region for intellectual disability to NARG2 and RORA. *European Journal of Medical Genetics*. 57 (4), 163–168.

Yang Z, MacQuarrie KL, Analau E, Tyler AE, Dilworth FJ, Cao Y, Diede SJ and Tapscott SJ (2009) MyoD and E-protein heterodimers switch rhabdomyosarcoma cells from an arrested myoblast phase to a differentiated state. *Genes & Development*. 23 (6), 694–707.

Yen H-Y, Ting M-C and Maxson RE (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. *Developmental Biology*. 347 (2), 258–70.

Yoshida T, Vivatbutsiri P, Morriss-Kay G, Saga Y and Iseki S (2008) Cell lineage in mammalian craniofacial mesenchyme. *Mechanisms of Development*. 125 (9-10), 797–808.

You FM, Huo N, Gu YQ, Luo M-C, Ma Y, Hane D, Lazo GR, Dvorak J and Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*. 9, 253.

Zhang X, Carpenter D, Bokui N, Soo C, Miao S, Truong T, Wu B, Chen I and Vastardis H (2003) Overexpression of Nell-1, a craniosynostosis-associated gene, induces apoptosis in osteoblasts during craniofacial development. *Journal of Bone and Mineral Research*. 18 (12), 2126–2134.

Zhang Y, Flejter WL, Barcroft CL, Rivière M, Szpirer J, Szpirer C BM (1995) Localization of the human HTF4 transcription factors 4 gene (TCF12) to chromosome 15q21. *Cytogenetics and Cell Genetics*. 68 (3), 235–8.

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

PUBLICATIONS ARISING FROM THIS THESIS

1. **Sharma VP**, Fenwick, AL, Goos JA, Twigg SRF, Wall SA, Wilkie, AO. (2014). Diagnostic Outcomes in Craniofacial Surgery are Improved by use of Next-Generation Sequencing. *Journal of Plastic, Reconstructive and Aesthetic Surgery*. 67(10):1462.

2. Babbs C, Lloyd D, Pagnamenta AT, Twigg SRF, Green J, McGowan SJ, Mirza G, Naples R, **Sharma VP**, Volpi EV, Buckle VJ, Wall SA, Knight SJL, Parr JR and Wilkie AOM (2014) De novo and rare inherited mutations implicate the transcriptional coregulator TCF20/SPBP in autism spectrum disorder. *Journal of medical genetics*. 51 (11), 737–47.

3. **Sharma VP**, Fenwick AL, Brockop MS, McGowan SJ, Goos JAC, Hoogeboom AJM, 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, Johnson D, Wall SA, van der Spek PJ, Mathijssen IMJ, Maxson RE, Twigg SRF and Wilkie AOM (2013) Mutations in *TCF12*, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. *Nature genetics*. 45 (3), 304–7.

4. **Sharma VP**, et al. Mutations in *TCF12*, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis (poster abstract). Spring Meeting for Clinician Scientists in Training 2013,

The Academy of Medical Sciences, London, February 2013 – *The Lancet* (Volume 381, Page S114, doi: 10.1016/S0140-6736(13)60554-1)

PRESENTATIONS ARISING FROM THIS THESIS

International

1st author

1. **Sharma VP**, et al. Improved Diagnostic Outcomes in Craniofacial Surgery by use of Next-Generation DNA Sequencing, 5th European Plastic Surgery Research Council meeting, Hamburg, Germany, August 2013
2. **Sharma VP**, et al. Mutations of *TCF12*, encoding a basic-helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. European Human Genetics Conference 2013, Paris, France, June 2013

Contributing author

1. Wilkie AOM, McGowan SJ, Fenwick AL, **Sharma VP,** et al. New insights into craniosynostosis from whole exome and genome sequencing. Society for Craniofacial genetics and developmental biology, Boston, USA, October 2013
2. Goos JAC, **Sharma VP**, et al. Mutations in *TCF12* are a Frequent Cause of Coronal Craniosynostosis. International Society of Craniofacial

Surgery. 15th International Congress, Jackson Hole, Wyoming, September 2013

National

1. **Sharma VP**, et al. Diagnostic Outcomes in Craniofacial Surgery are Improved by use of Next-Generation Sequencing. 12th Congress of the European Society of Plastic, Reconstructive and Aesthetic Surgery, Edinburgh, July 2014
2. **Sharma VP**. *TCF12* Mutational Origin and Haplotyping. Progress in Craniofacial Genetics Research (4 centre national meeting), Oxford, June 2014.
3. **Sharma VP**, et al. Improved Diagnostic Outcomes in Craniofacial Surgery by use of Next-Generation Sequencing. Craniofacial Society of Great Britain and Ireland Annual Scientific Conference, Oxford, April 2014
4. **Sharma VP**. *TCF12* Haplotyping. Progress in Craniofacial Genetics Research (4 centre national meeting), Oxford, June 2013.
5. **Sharma VP**. *TCF12*-related craniosynostosis. 4 Unit National Craniofacial audit meeting, Birmingham, May 2013.
6. **Sharma VP**. Mutations in *TCF12*, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. Spring Meeting for Clinician Scientists in Training 2013, The Academy of Medical Sciences, London, February 2013

7. **Sharma VP.** Coronal Synostosis: A novel disease gene revealed by exome sequencing, 4 Unit National Craniofacial audit meeting, Oxford, May 2012
8. **Sharma VP.** *TWIST*-related craniosynostosis. Craniofacial Research Meeting, Birmingham, January 2012
9. **Sharma VP**, *et al.* A new craniosynostosis syndrome. 4 Unit National Craniofacial audit meeting Liverpool, May 2011
10. **Sharma VP**. Genetics of Craniofacial Malformation. Birmingham and London, February and April, 2011

Contributing author

1. Twigg SRF, Lord H, McGowan SJ, Fenwick AL, **Sharma VP** et al, Bringing gene discovery in craniosynostosis to the clinic, British Society for Genomic Medicine annual conference, Liverpool, September 2013
2. Twigg SRF, Lord H, McGowan SJ, Fenwick AL, **Sharma VP** et al. New insights into craniosynostosis from whole exome and genome sequencing, UK Genome Science Meeting: Biology, Technology & Bioinformatics, Nottingham, September 2013

APPENDIX

(See CD ROM)