KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI, GHANA

Significance of Genetic Mutations in the Molecular Pathogenesis of Human Orofacial Clefts in Ghana

By
Lord Jephthah Joojo Gowans (BSc Biological Sciences)

A Thesis submitted to the Department of Biochemistry and Biotechnology, College of Science, in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

JUNE 2017

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI, GHANA

Significance of Genetic Mutations in the Molecular Pathogenesis of Human Orofacial Clefts in Ghana

By
Lord Jephthah Joojo Gowans (BSc Biological Sciences)

A Thesis submitted to the Department of Biochemistry and Biotechnology, College of Science, in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

JUNE 2017

DECLARATION

I hereby declare that this submission is my own work towards a PhD degree and that, to the best of my knowledge, it contains no material previously published by another person, nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

Lord Jephthah Joojo Gowans	…….	……..
20245827 (PG6120111)	Signature	Date
(Student)		

Certified by

Dr. F. K. N. Arthur	……..	……
(Supervisor)	Signature	Date

Dr. Peter Twumasi	……..	……..
(Supervisor)	Signature	Date

Prof. Peter Donkor	……..	……
(Supervisor)	Signature	Date

Certified by

Dr. Hilary D. Zakpaa	……	……..
(Head of Department)	Signature	Date

ABSTRACT

Human orofacial clefts (OFCs) are congenital craniofacial dysmorphologies, with a global prevalence of 1 per 700 live births. OFCs may be syndromic or nonsyndromic, with the syndromic forms presenting with extra congenital anomalies. Nonsyndromic OFCs (NSOFCs) account for about 70% of all OFCs. The aetiology of the more common NSOFCs is complex, with both genes and environmental factors playing vital roles. The main objective of the present study was to demonstrate association between genetic variants and the pathogenesis of OFCs in Ghanaians. The specific objectives were fivefold: (i) determine pattern of inheritance of OFCs, (ii) establish the role of environmental and other risk factors in OFC aetiology, (iii) ascertain genetic susceptibility loci for OFCs, (iv) detect rare aetiologic mutations in affected individuals and, (v) subsequently validate the pathogenic mechanism of action of some selected rare variants in zebrafish embryos. A questionnaire was administered to participating families mainly at KATH to collect environmental and phenotypic data. Saliva and cheek swab samples were collected from participants using Oragene DNA collection kits. DNA was extracted from samples using Oragene saliva processing protocol and Quibit assay was used to quantify extracted DNA, with subsequent validation of sexes of participants through XY-Genotyping that employed Real Time PCR. A total of 3,585 individuals (872 cases, 1635 relatives and 1078 unrelated controls) were genotyped at 48 SNPs using Fluidigm SNP Genotyping Protocol. Eight genes were also directly sequenced in 184 NSOFC cases and IRF6 gene only on 80 individuals with multiple congenital anomalies (MCAs), using Sanger sequencing technology. Primers used for DNA sequencing were designed with Primer3 software based on human genome assembly

GRCh37/hg19, 2009 (http://genome.ucsc.edu), followed by ascertainment of their optimal

PCR conditions through Gradient PCR. Initial PCR was used to amplify exons and flanking intronic sequences as well as 5' and 3' untranslated regions (UTRs) of the eight genes. PCR amplicons were examined by gel electrophoresis. The Initial PCR product was then sequenced with ABI 3730XL DNA Sequencer. Employing genetic engineering, site-directed mutagenesis and developmental biology techniques, two selected variants of IRF6 were also functionally validated in zebrafish embryos. The study established multifactorial pattern of inheritance for OFCs in Ghana. Low socio-economic status, delayed antenatal attendance and folate deficiency were significantly associated with families with OFCs by increasing OFC susceptibility. In SNP association studies, case-control meta-analyses demonstrated that PAX7 (rs742071, $p=5.10\times10^{-03}$), 8q24 (rs987525, $p=1.22\times10^{-03}$) and VAX1 (rs7078160, p=0.04) were nominally associated with NSCL/P. MSX1 (rs115200552, p=0.01), TULP4 (rs651333, p=0.04), CRISPLD2 (rs4783099, p=0.02) and NOG (rs17760296, p=0.04) were nominally associated with NSCP. Many other loci exhibited threshold association with NSOFCs in TDT and DFAM analyses. Coding, splice site and/or regulatory region variants were observed in all eight sequenced genes. Novel pathogenic mutations were observed in both NSOFCs (p.Glu69Lys, p.Asn185Thr and c.175-2A>C) and MCA (p.Gly65Val, p.Lys320Asn and c.379+1 G>T) cases in IRF6, including probable genetic modifiers. In functional validation of p.Glu69Lys and p.Gly65Val, these variants, through dominantnegative effect, disrupted craniofacial structures, such as pharyngeal arches, in zebrafish. These observations are relevant for prenatal diagnosis of high risk families, understanding the genetic architecture of OFCs and genetic counseling. This is the first ever genetic study on OFCs in Ghana and also the most extensive done in African populations on the African continent to date. The study has for the first time, demonstrated associations between the studied genetic loci and NSOFCs among continental Africans, with striking racial

differences.

ACKNOWLEDGEMENTS

I will like to adore the provisions of the Sovereign God, Jehovah Jireh, for His guidance and love throughout the years of study and research for this submission. Abba Father, the glory is all yours; may Your Name be forever praised. I will also like to express my heartfelt gratitude to my mentors and supervisors for guiding me throughout this work; your encouragements kept me going.

Dr. F. K. N. Arthur, Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology (KNUST)

Dr. Peter Twumasi, Department of Biochemistry and Biotechnology, KNUST

Prof. Peter Donkor, Department of Surgery, SMS, KNUST/Cleft Clinic, Komfo Anokye Teaching Hospital (KATH)/KNUST

I will also like to appreciate deeply my external mentors, in whose laboratories I carried out all laboratory analyses. Together, these two mentors spent over US$200,000 of their research grants to support this research. I adore your impacts on my life.

Prof. Azeez Butali, Department of Oral Pathology, Radiology and Medicine, University of Iowa, Iowa, USA [NIDCR Grant K99/R00 DE022378].

Prof. Jeffrey C. Murray, Department of Pediatrics, University of Iowa, Iowa, USA [NIDCR R37 Grants DE-08559 and DE-016148]

I do appreciate greatly the mentoring of these external supervisors too. The first one sponsored some of my trips for sample collection and provided research support as well. The second person gave me access to his laboratory to carry out the functional analyses in zebrafish. Your contributions are deeply adored.

Prof. Pius Agbenorku, Department Surgery, SMS, KNUST/Cleft Clinic, KATH Prof. Robert A. Cornell, Department of Anatomy and Cell Biology, University of

Iowa, Iowa, USA.

The Ghana Cleft Foundation, through Prof. Peter Donkor, sponsored all trips to all Regional Hospitals where sample collection took place. I appreciate this greatly. Thanks also to the following members from respective laboratories; they provided vital and technical laboratory assistance to this project.

Murray Lab: Nichole Nidey, Jen Standley, Dee Even and Nancy Davin

Butali Lab: Tamara Busch and Erin Brothers-Smith

Cornell Lab: Greg Bonde and Dr. Huan Liu

I also appreciate the contribution of the following to the success of this study:

Doctors and Nurses, Cleft Clinic, KATH

The Directorates of Surgery, Child Health and Oral Health, KATH

Department of Biochemistry and Biotechnology, KNUST

Families who participated in this research; they were cardinal to the success of this study. Your sacrifices have yielded good results.

I adore the sacrifices of my family, who sacrificed their comfort for the success of this study. Special thanks to my twin daughters Lordina Hallel Hillary Gowans and Lordina Berachah Bella Gowans as well as my son, Lord Jephthah Jason Gowans Jnr. Special thanks to my wonderful mother (Emma Hanson), Dr Michael and Mrs Dina Adjaloo, siblings and friends for their consistent prayer support. Finally, I dedicate this research to the memories of my late father (Richard Alfred Ato Gowans) and sister (Mrs Jemima E. Baidoo), who showed so much interest in this research; however, I lost both of them within a year, in the course of this research. I miss you Dad and Sister! May God keep your souls in His bosom.

TABLE OF CONTENTS

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

CONTENTS PAGE

LIST OF TABLES

PAGE

TABLE

LIST OF FIGURES

PAGE

FIGURE

FIGURE PAGE

LIST OF PLATES

ACRONYM FULL NAME

LIST OF ABBREVIATIONS

ACRONYM	FULL NAME
ABCA4	ATP-binding cassette, sub-family A, member 4
ADAMTS20	a disintegrin-like and metalloprotease with thrombospondin type1 motifs
ADCY9	adenylate cyclase 9
AP-2α /TFAP2α	transcription factor AP- 2 alpha (activating enhancer binding protein)
ARGHAP29	Rho GTPase activating protein 29
ATP	Adenosine triphosphate
AXIN2	Axis inhibition protein 2
BMP2	bone morphogenetic protein 2
BMP4	Bone morphogenetic protein 4
cAMP	Cyclic adenosine monophosphate
CD map/number	Coded data map/number
CEPH	Centre d'Etude du Polymorphisme Humain
CL	Cleft lip only
CLP	Cleft lip and Palate
CL/P	Cleft lip and/or palate
CLPTM1	cleft lip and palate associated transmembrane protein 1
CP	Cleft Palate
CDC42	cell division cycle 42
CDH1	cadherin 1, type 1, E-cadherin

CREBBP	CREB (cAMP-response elements binding) binding protein
CRISPLD2	Cysteine-rich secretory protein LCCL domain containing 2
DbGaP	database of Genotypes and Phenotypes
DCAF4L2	DDB1 and CUL4 associated factor 4-like 2
DFAM	Family-Based Association for Disease Traits
DYSF	dysferlin, limb girdle muscular dystrophy 2B
ECM	extracellular matrix
ELAVL2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2
ENCODE	Encyclopedia of DNA elements
EPHA3	Ephrin receptor A3
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1
FBAT	Family-Based Association Test
FGF1	fibroblast growth factor 1
FGF8	fibroblast growth factor 8
FGFR1	fibroblast growth factor receptor 1
FGFR2	fibroblast growth factor receptor 2
FMN1	formin-1
FOXE1	Forkhead box E1
GADD45G	growth arrest and DNA-damage-inducible, gamma
GATA-1	GATA binding protein 1 (globin transcription factor 1)
GAP	GTPase activating protein

ACRONYM	FULL NAME
GREM1	gremlin-1
GRHL3	Grainyhead-like 3
GRID2	glutamate receptor, ionotropic, delta 2
GTP	Guanosine triphosphate
GWAS	Genome-wide association studies
HBAT	haplotype-based association test
HWE	Hardy Weinberg Equilibrium
IGF2	Insulin-like growth factor 2

ACRONYM	FULL NAME
IRF6	Interferon regulatory factor 6
INSIGF	the overlapping region of IGF2 and insulin genes (insulin induced gene)
JHS	Junior High School
KATH	Komfo Anokye Teaching Hospital
KIAA1598	KIAA1598
KDM6A	lysine (K)-specific demethylase 6A
LD	Linkage Disequilibrium
MAF	minor allele frequency
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homologue B
MCA	Multiple congenital anomalies
MEE	medial edge epithelium
MSX1	msh (muscle segment homeobox) homoeobox 1
MYH9	myosin, heavy chain 9, non-muscle
NSCL	Nonsyndromic Cleft Lip
NSCLP	Nonsyndromic Cleft Lip and Palate
NSCL/P	Nonsyndromic cleft lip with or without palate
NSCP	Nonsyndromic Cleft Palate
NSOFC	nonsyndromic orofacial clefts
NTN1	netrin 1
OFC	Orofacial clefts
OMIM	online Mendelian inheritance in man

ACRONYM	FULL NAME
OR	Odds Ratio
Pax6	paired box 6
PAX7	Paired box 7
PCR	Polymerase Chain Reaction
PHF8	PHD (plant homeodomain) finger protein 8
POO	Parent of origin effects
PPS	Popliteal Pterygium Syndrome
PRS	Pierre Robin Sequence
PTHLH/PTHRP	parathyroid hormone-like hormone
PTPN13/PTPL1	protein tyrosine phosphatase, non-receptor type 13
PVRL2	Poliovirus receptor-related 2
p21	p21 protein (Cdc42/Rac)-activated kinase 2
RAC1	RAB10, member RAS oncogene family
RAP2A	RAP2A, member RAS oncogene family
RASIP1	RAS interacting protein 1
RBFOX3	RNA binding protein, fox-1 homolog 3
REDCap	Research Electronic Data Capture
RHOA	ras homolog family member A
ROCK1	Rho-associated, coiled-coil containing protein kinase 1
RR	Relative risk

ACRONYM	FULL NAME
Rx	relaxin/insulin-like family peptide receptor 1
RXR	retinoid X receptor, alpha
SHS	Senior High School
Smad4	SMAD family member 4
SMCP	Submucous cleft palate
SNP	Single nucleotide polymorphism
SOC	Super Optimal broth with Catabolite repression
SOFC	Syndromic orofacial clefts
SPRY2	sprouty homolog 2
Sp1	Sp1 transcription factor
SUMO1	small ubiquitin-like modifier 1
TCF-1α	transcription factor 1
Tcf12	transcription factor 12
TDT	Transmission Disequilibrium Test
TGFα	transforming growth factor alpha
TGFβ	transforming growth factor beta
THADA	thyroid adenoma associated
TP63/ P63	Tumor protein p63
TPM1	tropomysin 1
TULP4	tubby like protein 4
USP9Y	ubiquitin specific peptidase 9, Y-linked
VAX1	Ventral Anterior Homeobox1

ACRONYM	FULL NAME
VWS	Van der Woude Syndrome
WNT	wingless-type MMTV (mouse mammary tumor virus) integration family site
WNT3	wingless-type MMTV (mouse mammary tumor virus) integration site family, member 3
8q24	A gene desert on human chromosome 8

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background

Human orofacial clefts (OFCs) are congenital abnormalities that emanate from the breakdown of normal mechanisms involved in early embryonic development of the face. OFCs are the most frequent congenital craniofacial abnormalities and the global incidence is 1 per 700 live births. However, race, ethnicity, geographical location, environmental factors as well as socio-economic status influence their incidence (Dixon et al., 2011; Gorlin et al., 2001). Asians and Native Americans have the highest incidence of 1.19 to 2.01 and 1.29 to 2.39 per 1000 live births, respectively, followed by Caucasians with an incidence of 1.02 to 2.0 per 1000 live births, with Africans having the lowest incidence of 0.38 to 0.54 per 1000 live births (Mossey and Modell, 2012).

The condition may manifest as cleft lip only (CL), isolated cleft palate (CP) or cleft lip and palate (CLP). Another manifestation of the abnormality is submucous cleft palate (SMCP), which is clinically characterized by a cleft of the soft palate and a bifid or split uvula. OFCs may also appear as facial clefts (Tessier) or macrostomia, an unusually large mouth. Covert forms of cleft include microform cleft which manifests as a slight dent in the red region of the lip or may appear as scar from the lip to the nostrils (Dixon et al., 2011; Gowans et al., 2015).

OFCs may be syndromic or nonsyndromic; syndromic forms are accompanied by other structural aberrations that occur outside the cleft region and thus, present with additional clinical manifestations that can be categorized as structural or cytogenetic chromosomal abnormalities, recognizable Mendelian single gene syndromes, teratogenic effects and other unknown syndromes. However, in nonsyndromic cases, they occur as an

isolated condition that is unlinked to any recognizable abnormality. Nonsyndromic cases account for 70% of OFCs whereas only 30% are syndromic (Dixon et al., 2011).

The aetiology or pathogenesis of OFCs, particularly nonsyndromic forms, is poorly understood. They depict the complexity and diversity of molecular mechanisms in embryogenesis where both genetic and environmental risk factors play influential roles. The role of genetics in the aetiology is buttressed by the fact that sibling risk is thirty times higher than that of normal population prevalence. Also, monozygotic twins have OFC concordance rate of 25 to 45% whereas dizygotic twins have 3 to 6% concordance rate. Nevertheless, the lack of complete concordance in monozygotic twins showcases the importance of environmental risk factors in the aetiology (Gorlin et al., 2001).

OFCs have life-long effects on affected individuals. They may affect speech, hearing, feeding and psychological development of affected individuals. Because of this, patients may undergo many rounds of surgical repair. On many occasions, multidisciplinary care which include hearing and speech therapy, psychotherapy and genetic counseling, nutritional advice, together with extensive dental and orthodontic treatments are required (Stanier and Moore, 2004).

1.2 Problem statement

Though many candidate genetic loci for syndromic OFCs (SOFCs) and nonsyndromic OFCs (NSOFCs) have been studied, the exact roles of these loci in cleft pathogenesis and how these loci work in concert to initiate cleft pathogenesis remain elusive. Moreover, genes implicated in syndromic clefts seem to act as modifiers to influence the occurrence of NSOFCs. For example, IRF6 gene variants that cause Van der Woude Syndrome (VWS) and Popliteal Pterygium Syndrome (PPS) have also been

implicated in the aetiology of NSOFCs (Kondo et al., 2002; Zucchero et al., 2004; Rahimov et al., 2008). Furthermore, some genetic loci associated with the aetiology of clefts in certain populations have failed to show the same association in other populations. For example, the single nucleotide polymorphism (SNP) rs2235371 in IRF6 have shown strong association with NSCL/P in populations of Asian ancestries whereas only a marginal association has been observed for population of European ancestries (Zuchero et al., 2004; Sun et al., 2015). This SNP is not known to exist in the African population and three attempts to replicate the association between IRF6 and NSCL/P in the African population were not successful (Butali et al., 2011; Wheatherley-White et al., 2011; Figueiredo et al., 2014). Thus, some genetic variants implicated in cleft aetiology seem to be population-specific. This is further buttressed by outcomes of various genome-wide association studies (GWAS) of NSCL/P where some loci are known to be population-specific: 8q24 locus show strong signal in populations of European ancestry whereas ABCA4 and MAFB loci are highly associated with NSCL/P in populations of Asian ancestry (Beaty et al., 2010). Even where an association has been established between a gene and NSCL/P, direct DNA sequencing studies have not always been able to detect large numbers of potentially aetiologic rare variants in associated genes (Chawa et al., 2014). Also, current screening programs have not succeeded in revealing major environmental risk factors responsible for clefts (Loenarz et al., 2009). However, race, ethnicity, geographical location, environmental risk factors as well as socio-economic status were reported to influence the occurrence and causes of clefts (Dixon et al., 2011; Gorlin et al., 2001). There is therefore the need to ascertain the patterns of cleft inheritance as well as identify all possible genetic and environmental risk factors that may influence

clefts occurrence and their extent or level of effect or influence in Ghana as well as in every country or kindred.

1.3 Objective

The objective of the study was to establish association between genetic mutations or variants and the pathogenesis of human orofacial clefts in Ghana.

1.3.1 Specific objectives

Five specific objectives were set for the study and they are:

(i) To determine the pattern of inheritance of clefts in the study population through pedigree analysis.

(ii) To establish the role of teratogens and environmental as well as other risk factors in the pathogenesis of clefts through a questionnaire-based study.

(iii) To investigate the role of common variants in cleft pathogenesis by single nucleotide polymorphism (SNP) genotyping of significant genome-wide association studies (GWAS) hit SNPs and other candidate gene SNPs.

(iv) To investigate the role of rare functional variants in cleft pathogenesis by direct DNA sequencing of GWAS hit genes and other candidate genes.

(v) To simulate in zebrafish embryos the mechanism of action of selected rare functional variants.

1.4 Justification

Many genes have been implicated in the aetiology of human OFCs. However, race, ethnicity, geographical locations, environmental factors, socio-economic status and teratogens have been observed to substantially influence the occurrence and causes of clefts. There is no national OFC prevalence or incidence data for Ghana but some studies (Agbenorku et al., 2011a; Agbenorku et al., 2013) have suggested that OFC prevalence is as high as 6.3 per 1000 live births in Ketu South District of Volta Region, which has lower socio-economic status, and an incidence of as low as 1.31 per 1000 live births in Kumasi Metropolis, which is of high socio-economic status. Observations (Donkor et al., 207a; Agbenorku et al., 2011b) from Komfo Anokye Teaching Hospital (KATH) have shown that there has been a progressive increase in the number of reported OFC cases due to rise in the number of free surgeries for patients and community awareness. OFCs have many negative impacts on affected patients: psycho-social challenge, aesthetic defect, feeding difficulty and nasal regurgitation, speech and language pathology, infanticide as well as inner ear infections. High financial burden due to costs of surgeries, orthodontics, feeding appliance and a multidisciplinary cleft team are some of economic burdens OFC place on affected families, community and the country at large. Given the almost racially homogeneous nature of the Ghanaian population, causes and occurrence of clefts in Ghana may thus be different from those occurring elsewhere. Thus, different or unique genetic mutations and risk factors may be responsible for cleft pathogenesis in Ghana and this study seeks to unravel such mutations and risk factors. The study will help to detect high risk families, pattern of inheritance of OFC and mutated genes associated with OFCs; this is crucial for effective genetic counseling. Data from environmental and other risk factors,

geographical distribution as well as socio-economic status of affected families will also impact positively on our public health system through the education of prospective mothers on the probable periconception risk factors that may predispose them to OFCs. The study may also detect possible occurrence of post-operative poor wound healing, scarring and fistula formation in some patients, as mutations in some OFC-predisposing genes (such as IRF6) have been shown to predispose patients to these keratinocyte differentiation-induced processes. This observation may be crucial for post-operative surgical wound management.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Human orofacial clefts

Human orofacial clefts (OFCs) are congenital dysmorphologies of normal facial structures. Cleft lip (CL), cleft palate (CP) and cleft lip and palate (CLP) are the most common forms of OFCs. Thus, OFCs are aetiologically heterogeneous, and this has crucial links in elucidating the biology of facial development and how environmental risk factors interact with genetic factors. Granting that OFCs are not major causes of mortality in developed countries, they do inflict considerable morbidity on affected individuals and impose huge financial burden on affected families, with attendant societal burden (Wehby and Cassell, 2010).

OFCs may be syndromic or nonsyndromic. With the advent of the genomic era, there have been substantial advances in the identification of causative genetic mutations for syndromic OFCs. Howbeit, there has been less advancement in the understanding of the genetic aetiology of NSOFCs due to their genetic heterogeneity, deviation from Mendelian inheritance patterns, scarcity as well as cost of genomic tools and the need for very large data sets. These notwithstanding, recent development of innovative approaches to genotyping and powerful, cost-effective tools, in addition to extrapolation from studies of syndromic OFCs, have enhanced our understanding of NSOFCs (Dixon et al., 2011).

2.1.1 Epidemiology of human orofacial clefts

Collectively, craniofacial anomalies are among the most common manifestations of all birth defects (Dixon et al., 2011). The global incidence of OFCs is 1 per 700 live births.

However, race, ethnicity, geographical locations, environmental factors and socio-economic status influence the incidence of OFCs (Gorlin et al., 2001). The highest incidence occurs in Asians and Native Americans, followed by Caucasians whereas

African-derived populations have the lowest incidence (Mossey and Modell, 2012). Asians and Native Americans have incidence of 1.19 to 2.01 and 1.29 to 2.39 per 1000 live births, respectively, followed by Caucasians with an incidence of 1.02 to 2.0 per 1000 live births, with Africans having the lowest incidence of 0.38 to 0.54 per 1000 live births. Caribbean populations have an incidence of 0.93 per 1000 live births (Mossey and Modell, 2012). These observations presuppose that the relative contributions of individual susceptibility loci may vary across different human populations. The incidence also differs by gender and laterality, i.e. whether it affects the left or right side of the lip or face. A ratio of 2:1 male to female has been observed for CL whereas a ratio of 1:2 male to female has been reported for CP. In terms of laterality, a ratio of 2:1 leftsided to right-sided clefts has been observed among unilateral clefts (Dixon et al., 2011).

Incidence of OFCs in the Ghanaian population may be higher than has been attributed to African-derived populations, due to inefficient ways of reporting clefts and natality in general. A study at KATH from 1991 to 2005 showed that each succeeding year recorded an increase in the number of cleft surgeries. From 2001 to 2005, 58% of OFC cases were males whereas 42% were females; 78% of patients had unilateral CL/P (Donkor et al., 2007a). In a subsequent study from 2006 to 2009 at KATH, higher number (an average of 132) of surgeries was done yearly. Most cases (54.2%, n = 286) were males. Moreover, CL, which was 73.1%, outnumbered CP. Also, unilateral clefts were the majority, being 70.5% (Agbenorku et al., 2011b).

Community-based studies may be powerful tools for accurately estimating the prevalence of OFCs in Africa. In Wudoaba cluster of villages in the Volta Region, CL/P prevalence was as high as 6.3 per 1,000 people: 25 cases per 4,000 live births. Vitamin deficiencies and consanguinity were asserted to be causal factors for the high incidence of CL/P in this cluster of villages (Agbenorku et al., 2011a).

The incidence of the condition may vary in various regions or ethnicities in Ghana. A descriptive prospective survey was done at eleven selected health facilities in Kumasi. An incidence of 1.31 per 1000 live births or 1 in 763 live births was observed. The male: female ratio was 0.5:1 for CL, 1.3:1 for CLP and 4:1 for CP. Majority (69.4%) of clefts were unilateral, with females (=14) outnumbering males (=11). Familial clefts were observed in five babies (13.9%). Seven (19.4%) syndromic clefts were also reported in the study (Agbenorku et al., 2013).

2.1.2 Phenotypic diversity of human orofacial clefts

The common forms of OFCs results from disruption of tissue planes above the lip, extending into the nares and/or the palate. Clefts involving the anterior structures like the lip and primary palate may be differentiated on both embryological and genetic bases from those involving the secondary palate (soft palate and uvula). Though there are many disruptions affecting the orofacial complex, the astounding majority affect only the upper lip and/or palate (Dixon et al., 2011).

The various subdivisions of clefts are even more heterogeneous (plates 2.1, 2.2 and 2.3). For example, CL may be complete or incomplete. Incomplete CL extends from vermillion of upper lip to the base of nares whereas complete CL extends from the upper lip vermillion to the alveolar bone. Also, complete CP affects both hard and soft palates

whereas an incomplete CP involves only the soft palate. A submucous cleft palate (SMCP) is characterized by soft palate cleft and bifid uvula. An oblique facial cleft extends from the base of the eye through the cheek to the upper lip (Gowans et al., 2015; Dixon et al., 2011; Agbenorku et al., 2011b; Donkor et al., 2007a).

Many lines of evidence currently presuppose that the phenotypic spectrum or expressivity of NSOFCs is more complex than earlier thought and should include an array of subclinical phenotypes in either an OFC patient or "unaffected" relatives (Weinberg et al., 2009). Subclinical phenotypes may include minor structural variants, such as lip pits or prints (Neiswanger et al., 2009), dental anomalies (Vieira et al., 2008), abnormalities of orbicularis oris muscles, three-dimensional facial image measurement

(Weinberg et al., 2009) and brain variants as revealed by magnetic resonance imaging (MRI) or surrogate measures (Nopolous et al., 2007). Furthermore, speech or cognitive variations such as velopharyngeal insufficiency (VPI), reading disability and intelligent quotient (IQ) should also be considered when dealing with subclinical phenotypes. Palatal subphenotypes may include bifid uvula, SMCP, congenital absence of uvula and distinction of clefts of the soft and hard palates (Dixon et al., 2011).

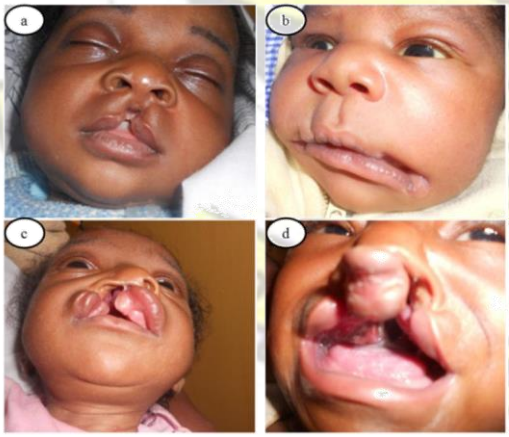


Plate 2.1: Phenotypic diversity of NSOFCs. a: Incomplete Midline CL b: Bilateral Macrostomia c: Right Complete Unilateral CLP d: Complete Bilateral CLP (Gowans et al., 2015).

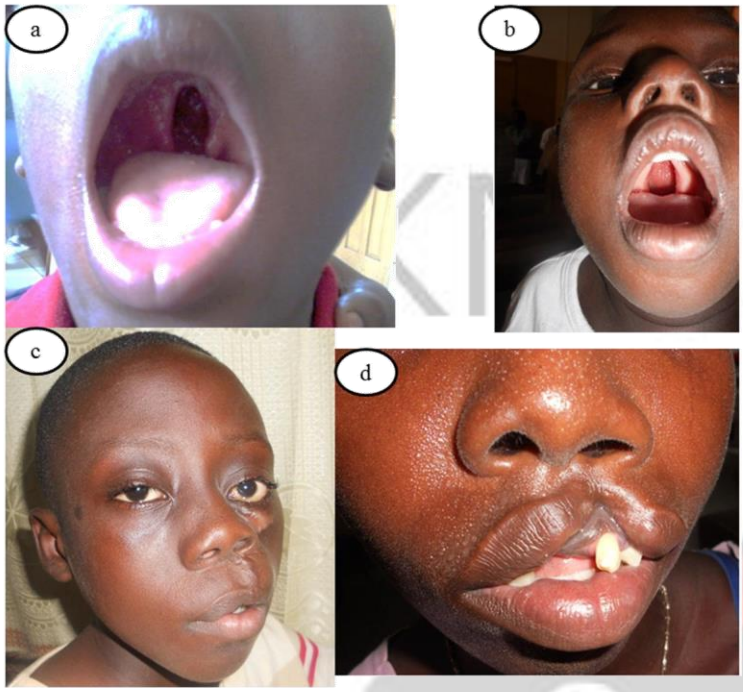


Plate 2.2: Further phenotypic diversity of NSOFCs. a: SMCP b: Wide Complete CP c: Repaired Left Facial Cleft and CLP d: Left Complete CL (Gowans et al., 2015).

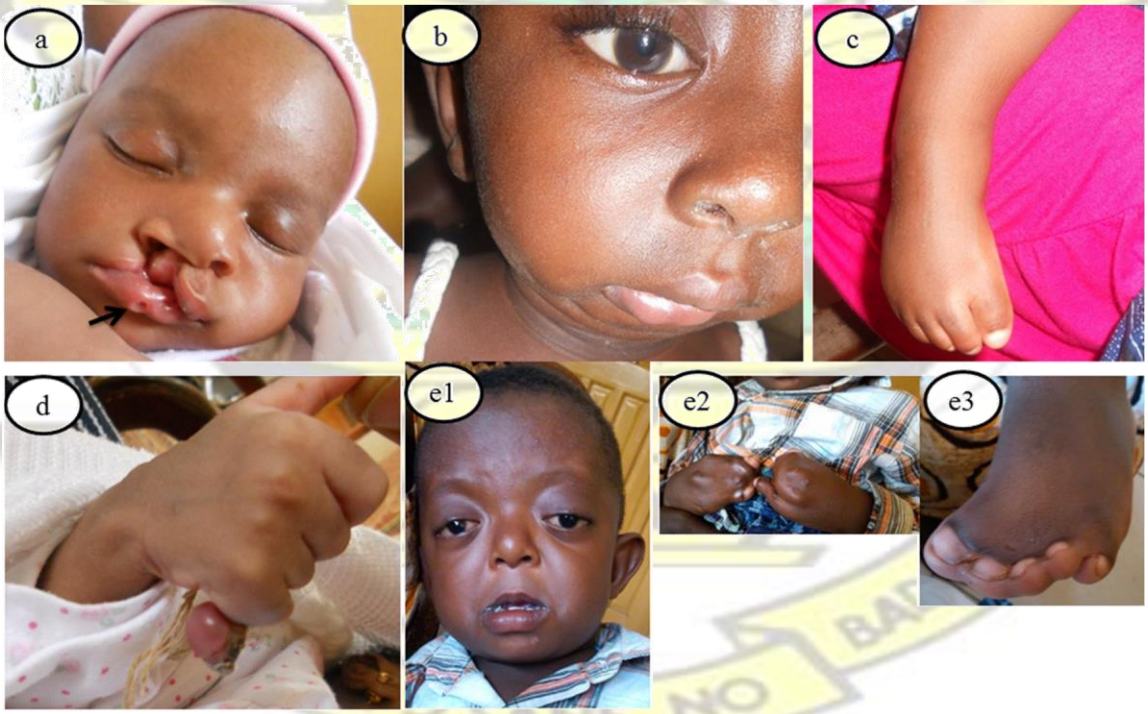


Plate 2.3: Congenital abnormalities which are associated with syndromic clefts. a:Van der Woude Syndrome (CLP and lip pits) b:Pierre Robin Sequence (CP and severe micrognathia) c:Club foot d:hexadactyly e:Apert Syndrome (CP with e1:low-set ears and

characteristic facial appearance, e2:syndactyly of fingers e3:syndactyly of toes) (Gowans et al., 2015).

2.1.3 Pattern of inheritance of human orofacial clefts

Familial clustering studies and twin studies have provided compelling evidence for a genetic component to OFCs. However, few pedigrees showcase well-defined Mendelian inheritance and majority of cases appear to be sporadic (Jugessur et al., 2009). OFCs are known to be influenced by environmental risk factors (Murray, 2002). A multifactorial model of inheritance is most appropriate in which genetic risk factors of small, individual impact may interact with environmental covariates (Rahimov et al., 2008).

The recurrence risk of the abnormality is influenced by an array of factors that are often unique to a particular family. These include number of family members with clefts, level of consanguinity, sex and race of affected individuals, and type of cleft each person has. When parents give birth to a child with a cleft, the risk that each succeeding child will be affected is 2–5%. If more than one individual in the nuclear family has a cleft, the recurrence risk surges to 10–12%. A lone member of a family with cleft has 2– 5% chance of having a child with cleft. Should the individual with cleft also have a close relative with a cleft, the recurrence risk rises to 10–12%. Furthermore, unaffected siblings of an affected patient have a roughly 1% recurrence risk. This might rise to 56% should more than one close family members have cleft. In syndromic clefts, the recurrence risk could be as high as 50% (http://www.cleftline.org/docs/Booklets/GEN-01.pdf , 12/01/2015).

Degree of relationship and anatomical severity of clefts influence recurrence risk of OFCs. In more than 54,000 relatives in Denmark, it was observed that for CL/P probands, the average recurrence risks for first-, second-, and third degree relatives were

3.5%, 0.8 % and 0.6% respectively. Also, most severe form of CL/P (i.e. bilateral CL/P) has the highest recurrence risk of 4.6% while a unilateral CL/P had a recurrence risk of 2.5%. An exception is CP where the least severe form (i.e. incomplete CP) has the highest recurrence risk of 3.9% (Grosen et al., 2010).

Human OFCs thus follow multifactorial pattern of inheritance, which has the following hallmarks: most affected children have normal parents, recurrence risk increases with number of affected individuals in the family, recurrence risk increases with the severity of the abnormality and consanguinity slightly raises the recurrence risk of an affected family. Moreover, the risk of a relative being affected falls off very sharply with a lesser degree of relatedness. Finally, when the two sexes have a varying probability of being affected, the least likely sex, if affected, is the most likely sex to produce an affected offspring. For example, in the Danish study, CL/P which was predominant in males had the highest recurrence risk for offspring when a mother was affected (http://www.uic.edu/classes/bms/bms655/lesson11.html#MODEL, 14/01/2015; Grosen et al., 2010).

2.1.4 The impact of orofacial clefts on affected individuals, families and society

OFCs impose a huge psychosocial and economic burden on affected individuals, families and society (Berk and Marazita, 2002). They come with many health complications early in life, including problems with feeding or ear infections. They could also impact negatively on the quality of life of affected individuals throughout life. This may lead to increased morbidity as well as mortality risks, particularly in less developed societies, where early systematic pediatric care may not be easily accessible (Wehby and Cassell, 2010).

OFC patients have characteristic feeding and hearing problems. Nasal regurgitation is a major feeding challenge, particularly those with CP. Infants may also lack the ability to suck and this complicates feeding. Patients may also develop middle ear infections that may ultimately lead to total hearing loss. CP infants, in particular, may develop compromised hearing, which may drastically reduce the baby's ability to mimic sounds of speech. Therefore, even before expressive language acquisition, infants are at risk of being unable to acquire receptive language. Both lips and palate are crucial in pronunciation. OFC patients cannot pronounce words very well, necessitating the need for a speech therapist (Berk and Marazita, 2002).

Clefts could impose life-long surgical burdens on patients. The diagnosis of OFC is traditionally done at birth by physically examining the baby. Current achievements in prenatal diagnosis enable in utero diagnosis. Though OFCs are very treatable, the kind of treatment hinges on type and severity of the cleft, with treatment procedures differing between craniofacial teams. Timing of treatment also varies for CL and CP. Significantly, craniofacial teams have different protocols for the timing of repair of both CL and CP. Some craniofacial teams wait until the child is between ages 10 to 12 months before CP is repaired whereas CL is repaired at 3 months. Other teams prefer late treatment, arguing that this saves the patient from repeated corrective surgeries, in that growth at this stage is less influential because deciduous teeth are replaced by permanent teeth. Other craniofacial teams prefer early OFC treatment as this reduces the need for speech therapy as speech therapy becomes less effective at advanced ages
(Wehby and Cassell, 2010).

2.2 Embryonic development of craniofacial structures

One of the most complex events during embryonic development is the formation of the face and head. This is coordinated by a network of signaling molecules and transcription factors in addition to proteins that confer cell polarity and cell-cell communications. Perturbation of this tightly regulated cascade of events may result in OFCs, emanating from failure of the facial primordial to meet and fuse (Stanier and Moore, 2004).

2.2.1 Embryology of human face

Human face and palate develop during the fourth to tenth weeks of pregnancy (Figure 2.1). At very early stages, migrating neural crest cells from the dorsal part of the anterior neural tube (i.e., cranial neural crest, CNC) fuse with the mesodermal cells to establish the primordial of the face. Five prominences then grow from the facial primordial. These are the frontonasal prominence (FP), two maxillary prominences (MXPs) and two mandibular prominences (MDPs). The FP is derived from the mesenchyme near the brain. The MXPs are derived from the upper part of pharyngeal arch 1 whereas the lower part of pharyngeal arch 1 gives rise to the MDPs. The MXPs and MDPs are formed around the stomodeum, which is the primitive or primordial mouth (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015; Stanier and Moore, 2004).

During week 5, the FP develops two thickenings, called nasal placodes, near its lateral edges. By week 6, each nasal placode develops a depression at its centre. This depression is called the nasal pit and it develops into the nostril. The remaining nasal placodes then sub-divide into a lateral nasal process or prominence (LNP) and a medial

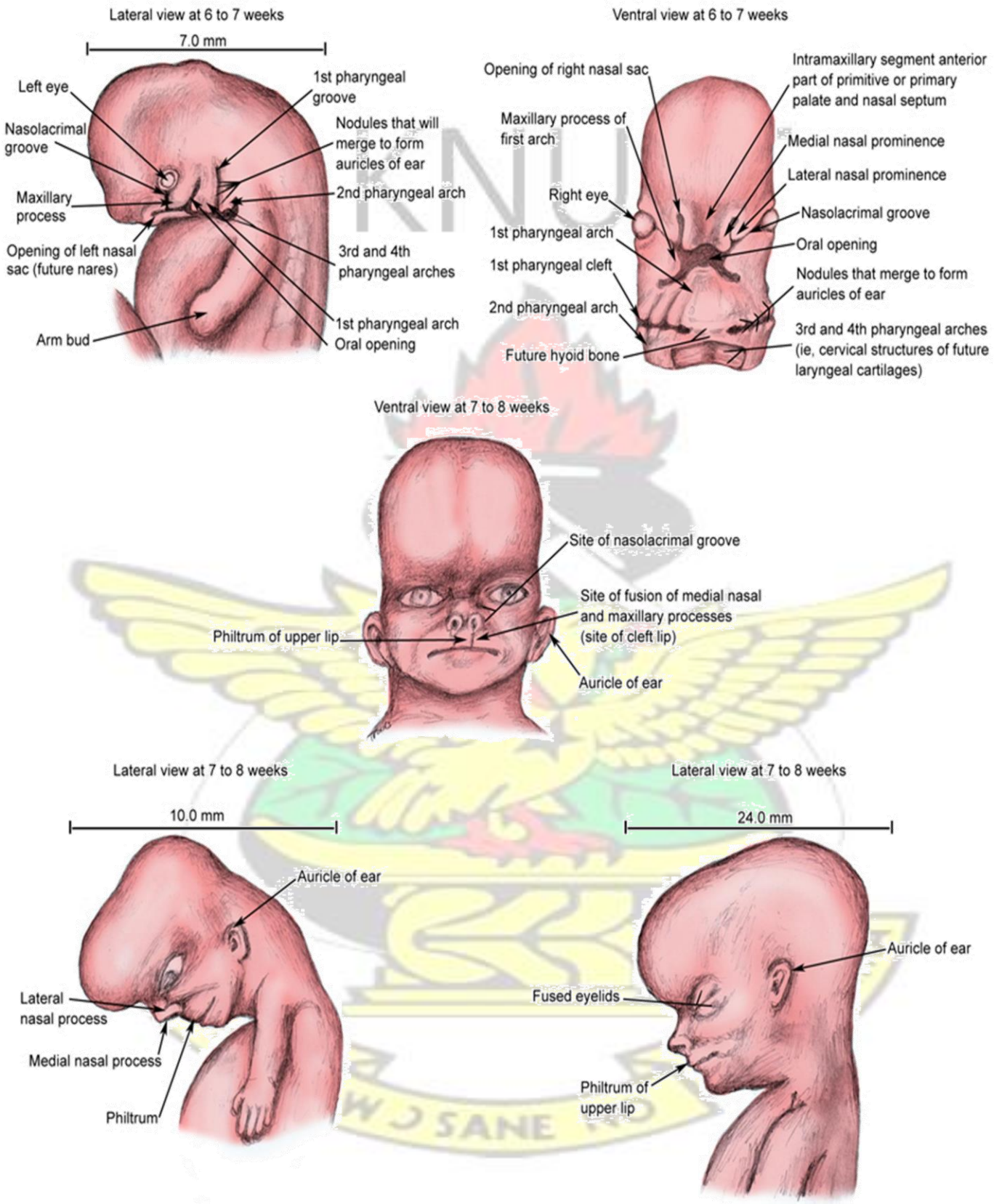


Figure 2.1: Embryonic developmental stages of the human orofacial region (http://emedicine.medscape.com/article/844962-overview, 14/04/2014)

nasal process or prominence (MNP). At this point in ontogeny, a groove, called the nasolacrimal grove, appears between the LNP and MXP on each side of the developing embryo. The LNP then fuses with MXP to form the lateral side of the human face. If this fusion fails to occur, an oblique facial cleft occurs in the neonate. Oblique facial clefts lead to nasolacrimal ducts becoming open grooves (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015; Stanier and Moore, 2004).

During the 7th week, the two MNP fuse to form the intermaxillary segment (IS); this later gives rise to the nasal septum. If this fusion process fails, a median cleft lip occurs in the neonate. Median cleft lip is rare and is generally associated with other problems such as mental retardation or brain abnormalities. By week 8, the IS then fuses with the MXP to form the upper lip and jaw architecture. Perturbation of this fusion results in cleft lip (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015; Stanier and Moore, 2004).

The palate is formed from a primary palate and a secondary palate between weeks 8 to 10 of human embryonic development. The primary or hard, anterior palate is derived from the IS. The secondary palate is derived from two palatine shelves which are tissue extensions that originate from the two MXPs. These right and left palatine shelves fuse to form the secondary or soft, posterior palate. The incisive foramen represents the anatomic boundary between the primary and secondary palates. Complete posterior fusion of the two palatine shelves also lead to uvula formation. If these fusion processes fail to occur, cleft palate and/or cleft uvula results (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015; Stanier and Moore, 2004).

By week 8 of human embryonic development, the nose attains a more humanlike shape. The MXPs grow rapidly, becoming larger than the respective mandibular prominences. By tenth week, the facial features look like that of a neonate. The intermaxillary process, derived from the IS, forms the philtrum of the upper lip whereas the MXPs begin to form well-defined cheeks and parts of the upper jaw. The auricle of the ear is fully formed by week 10 and it originates from the second or hyoid arch by week 5. The eye also emanates from the nasolacrimal groove by week 5 of embryonic development (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015; Stanier and Moore, 2004).

2.2.2 Gene networks in craniofacial development

Murine models give insight into genes involved in craniofacial development. Palatal shelf remodeling, epithelial differentiation and facial primordial identity are modulated by various signaling molecules which act as inductive signals. These molecules include bone morphogenetic proteins (Bmp), sonic hedgehog (Shh), representatives of transforming growth factor β (Tgfβ) subfamily and fibroblast growth factors (Fgf). Shh is involved in the initial induction of facial primordial; it is expressed in the palatal medial edge epithelium (MEE). Howbeit, Bmp2 and Bmp4 are tissuespecifically expressed within the mesenchyme and epithelia of the palatal shelves. The expression of Bmp2 and Bmp4 as well as Shh is mediated by Msx1, a gene which is also expressed in the facial primordia (Stanier and Moore, 2004).

Epidermal growth factor (Egf) promotes glycosaminoglycans production in the palatal shelves whereas Tgfα, which is expressed all over the palatal epithelium and mesenchyme, stimulates the biosynthesis of extracellular matrix (ECM). Collagen III and

fibronectin serve as modulating factors on hyaluronate enlargement during shelf reorientation. Collagen IX, on the other hand, works critically to signal epitheliamesenchyme communications. All these processes occur in the MEE cell surface just before shelf elevation (Stanier and Moore, 2004).

Transcription factors (such as Hox, Gli and Dlx) likewise play critical roles in mandibular and maxillary specification; these are regulated by Shh, Bmp and Fgf signals. These observations showcase the relevance of epithelia-mesenchyme interactions in craniofacial development. Specific regions of gene expression (for example, tooth buds) may act as inductive signaling centres to influence palatogenesis (Stanier and Moore, 2004).

Isoforms (1, 2 and 3) of Tgfβ family play interesting roles in palatogenesis. Current evidence presupposes that Tgfβ function is somehow partly mediated by the Smad signaling system. The first of the Tgfβ isoforms to be expressed is Tgfβ3 and its expression occurs in the epithelial section of the vertical shelves. Tgfβ3 is also lately expressed in the MEE and horizontal shelves; however, once the epithelia seam gets disrupted, Tgfβ3 is not detected. Tgfβ1 is explicitly expressed in the horizontal shelves; howbeit, its expression is halted once epithelial seam disruption occurs, just as is the case for Tgfβ3. Thus, Tgfβ1 and 2 hasten palatal shelf fusion; Tgfβ3 might elicit growth inhibition and is relevant for the initial adhesive interaction (Stanier and Moore, 2004).

Murine Tgfβ3 gene knockout models exhibit cleft palate (CP) because the palatal shelves fail to fuse. In such murine model, palatal shelves develop normally but do exhibit marked diminution in the filopodia present in the MEE surface. The palatal shelves also exhibit down-regulation of condroitin sulphate proteoglycan on the apical surface of the MEE. Both filopodia and condroitin sulphate proteoglycan are needed for efficient MEE

adhesion. In murine antisense forms, Tgfβ isoform-specific antibodies and gene knockouts reveal the failure of palatal fusion when Tgfβ3 but not Tgfβ1 or Tgfβ2 is absent (Stanier and Moore, 2004).

Specific matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs) are overly relevant in tissue remodeling during palatal fusion. This tissue remodeling includes basement membrane degradation and epitheliomesenchymal transformation. The expression of Timp-2 and Mmp-13 are dependent on

Tgfβ3 mediation: murine Tgfβ-/- models show strikingly reduced levels of Timp-2 and Mmp-13. This observation showcases the relevance of proteolytic degradation of ECM prior to palatal fusion. Obviously, cascade of genes are required for absolute palate morphogenesis or development (Stanier and Moore, 2004).

2.3 Pathogenic DNA variants

The human genome is made up of about 3.5 billion base pairs (Strachan and Read, 2011). However, the sequences of these bases are not constant in all humans. In a typical human genome, a single nucleotide polymorphism (SNP) is observed averagely every 1200 bases, leading to the occurrence of approximately 3 million of such SNPs. Other variations include copy number variations (CNVs) in microsatellites and genes, long and short tandem repeats (LTRs and STRs), etc. Most of these variants are not pathogenic but contribute to the genetic variation observed in humans. However, other genetic variations in DNA or chromosomes could be pathogenic because they could, for example, alter the binding motif of cis-regulatory elements, abolish or create splice sites, affect the stability of RNA transcripts, affect the structure and function of translated polypeptides, etc. Figure

2.2 depicts minor gene mutations that can have profound effect on health whereas Figure 2.3 elaborates further on various forms of point mutations. Insertions and deletions (indels) result when one or more nucleotide(s) is/are added or omitted, respectively, from a canonical nucleotide sequence. If the added or omitted

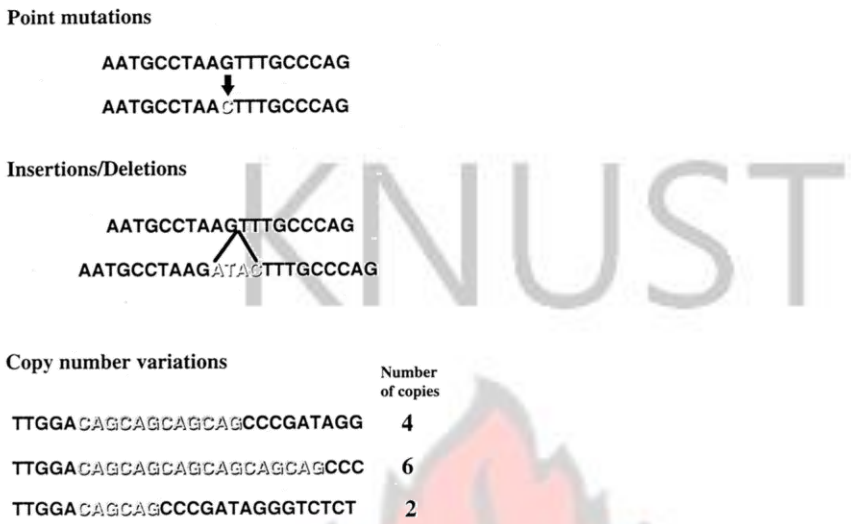


Figure 2.2: Types of gene mutations (www.nwfsc.noaa.gov/publications/scipubs/techmemos/tm17/figures/, 25/01/2015)

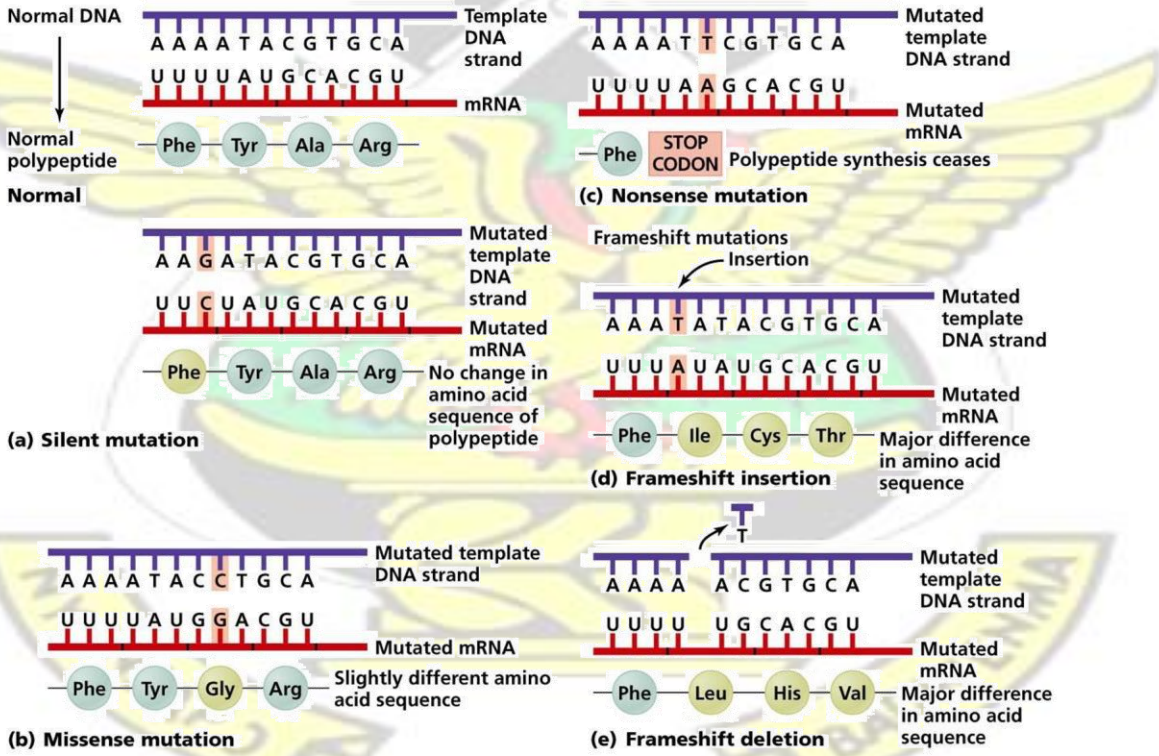


Figure 2.3: Types of point mutations (https://www.pinterest.com/, 25/01/2015) nucleotides are 3 or in multiples of 3, an "in-frame" indel occurs which does not affect the reading frame after the indel. However, if the added or omitted nucleotides are not 3 or not in multiples of 3, the indel

results in a frameshift mutation. Finally, Figure 2.4 delineates various gross, structural chromosomal aberrations that could also have pathogenic effects on human health. These pathogenic genetic variants may cause genetic disorders or diseases such as cystic fibrosis, cancer, OFCs, etc. (Strachan and Read, 2011).

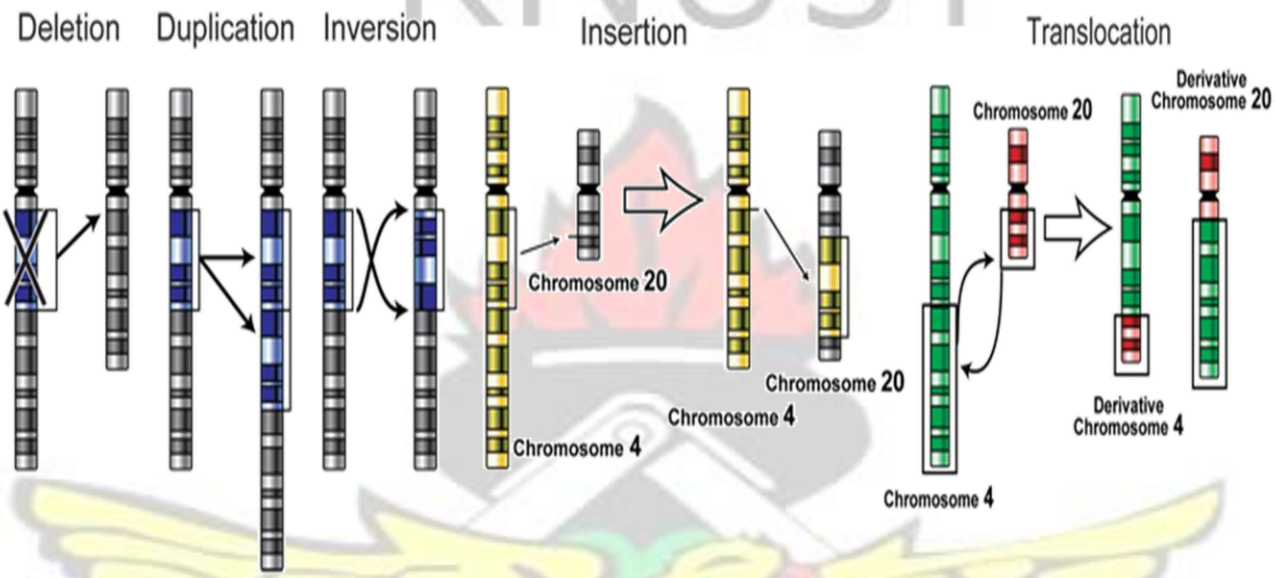


Figure 2.4: Types of gross chromosomal mutations (http://socratic.org/questions/howdo-dna-mutations-occur, 25/01/2015)

2.3.1 Factors influencing allele and genotype frequencies in populations

Population genetics is a subdivision of genetics that deals with genetic variations within and/or between populations and encompass the study of events such as population structure, speciation and adaptation. Four main processes account for the genetic variations within and between populations. These include mutation, genetic drift, gene flow and natural selection; selection is influenced by the phenomena of dominance and epistasis (Strachan and Read, 2011).

Hardy-Weinberg equilibrium (HWE) states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other

evolutionary influences, such as mutation, natural selection, mate choice, meiotic drive, gene flow, founder effect and genetic drift (Strachan and Read, 2011).

Mutation refers to a change in the nucleotide sequence of a gene or DNA and this leads to the formation of new alleles. Mutations may be non-functional, alter the effect of a gene product or may completely abolish the function of a gene product, culminating in varied phenotypes in a population. It may alter allele frequencies of a population over time which may ultimately lead to increased or diminishing of some phenotypes. An allele refers to a variant of a gene or DNA sequence (Strachan and Read, 2011).

Genetic drift refers to alteration of allele frequencies in a population as a result of random sampling from phenomena like random mate choice, founder effects, population bottleneck from catastrophes and meiotic drive. When genetic drift occurs, the alleles in the offspring are random samples of the alleles in the parents and this may cause some genetic variability to vanish from the population. Gene flow occurs as a result of migration from one population to another, leading to alteration in allele frequencies. However, because of physical migration barriers, such as oceans, mountain ranges, deserts and man-made structures, gene flow between populations occur slowly. Moreover, there is a reduced tendency for individual to migrate coupled with the fact that migrants usually may remain or return to their ancestral or natal populations. Therefore, interbreeding between populations that may lead to gene flow is a very slow process. Natural selection alters allele frequencies due to reproductive success which is driven by adaptive and environmental pressures. Dominant alleles (alleles that suppress the expression of other alleles) and epistasis (interaction between different loci in the genome leading to production of modified phenotypes) are key events that drive natural selection (Strachan and Read, 2011).

2.4 Nonsyndromic orofacial clefts

Nonsyndromic orofacial clefts (NSOFCs) are clefts that present with no other congenital anomalies. They are usually grouped as nonsyndromic cleft lip with or without palate (NSCL/P) and nonsyndromic cleft palate (NSCP) based on epidemiological and embryological observations. These two subphenotypes usually do not segregate in the same family, presupposing the existence of distinct genetic aetiologic factors (Dixon et al., 2011).

2.4.1 The complex genetics of nonsyndromic orofacial clefts

Diverse genetic approaches have been used to study NSOFCs. Linkage analysis employs large, multiplex families or smaller but inbred families or the analysis of affected relative pairs. This approach tries to figure out how OFCs segregates in families employing certain genetic markers. Other approaches include association studies that employ case-parent trios or case-control samples to test for association between SNPs in candidate genes and NSOFCs, search for gross chromosomal abnormalities or microdeletions in cases as well as direct sequencing of DNA from both affected and unaffected individuals. These approaches can be applied to candidate genes or employed as genome-wide strategies to detect potentially aetiologic loci (Dixon et al., 2011).

The search for candidate genes for clefts begun when it was shown that variants in transforming growth factor-α (TGFα) are associated with clefts. Gene expression and developmental analyses in model organisms, particularly mouse, gave earlier insight into candidate genes and gave biological evidence of genetic association with clefts.

Moreover, extrapolations from study of syndromic forms of clefts have helped to identify associated loci. These notwithstanding, replication of such studies across diverse

ethnicities have not been consistent, except for IRF6 which have been consistently replicated in populations of Asian and European ancestry (Dixon et al., 2011).

2.4.1.1 Genome-wide association studies

Genome-wide association studies (GWAS), which test for association between selected SNPs in the whole human genome and NSCL/P, have proven to be a powerful tool. To date, six GWAS and a meta-analysis have been published on NSCL/P. These GWAS showed that population heterogeneity may influence a particular genetic loci. Another issue that emanated from GWAS is the need to replicate various GWAS findings in diverse populations (Dixon et al., 2011).

The first GWAS implicated 8q24 (rs987525, $p=3.34\times10^{-24}$) in the aetiology of NSCL/P in Europeans (Birnbaum et al., 2009). The odd ratio (OR) at 95% confidence interval (CI) for the heterozygous genotype (AC) of rs987525 was 2.57 whereas that for the homozygous genotype (AA) was 6.05. The population attributable risk (PAR) was also observed to be 0.41. It was also observed that rs987525 exhibited both multiplicative and additive risk pattern, but not dominant or recessive model of genetic inheritance. An association between rs642961 (1q32.2) and NSCL/P ($p=1.5\times10^{-6}$) was also observed in the replication phase, though no evidence of interaction between IRF6 and 8q24 locus was seen (Birnbaum et al., 2009). The 8q24 locus's association with NSCL/P was replicated (rs987525, $p=9.18\times10^{-8}$, OR=2.09) in a subsequent GWAS involving Europeans from Greater Philadelphia area, US (Grant et al., 2009).

A third GWAS revealed additional NSCL/P susceptibility loci at 17q22 (rs227731, $p=1.07\times10^{-8}$, RR=1.38) and 10q25.3 (rs7078160, $p=1.92\times10^{-8}$, RR=1.36) in Europeans. The PAR for rs227731 (NOG) and rs7078160 (VAX1) were 23.9% and

12.3%, respectively whereas PAR for combined IRF6, 8q24 and these two new loci were 54.6%. NOG codes for noggin protein that is antagonistic to members of TGFβ superfamily, such as BMP4 which regulate mammalian palatogenesis and has also been associated with human OFCs (Mangold et al., 2010). Vax1 null mice exhibit craniofacial deformities that include CL; moreover, two individuals, one with SMCP and the other with CL, with 10q deletion syndrome were shown to have breakpoints in 10q25 (Hallonet et al., 1999, Zhao et al., 2010). Other suggestive loci implicated by this GWAS included rs1258763 at 15q13.3 (GREM1, $p=1.14\times10^{-6}$, RR=1.52), rs9574565 at 13q31.1 (SPRY2, $p=3.44\times10^{-7}$, RR=1.31) and rs7590268 at 2p21 (THADA, $p=8.62\times10^{-8}$, RR=1.42). GREM1 is a known antagonist of BMP4 (Mangold et al., 2010). SPRY2 overexpression in mice leads to craniofacial deformities (Goodnough et al., 2007), as palatogenesis is sensitive to Spry2 dosage (Welsh et al., 2007). THADA has been implicated in cell death receptor pathway and apoptosis as well as two individuals with CP and other malformations (Mangold et al., 2010).

A fourth GWAS identified two novel loci for NSCL/P susceptibility: MAFB at 20q12 (rs13041247, rs17820943 and rs11696257, all with $p=1.44\times10^{-11}$, OR=0.704) and ABCA4 at 1p22.1 (rs560426, $p=5.01\times10^{-12}$, OR=1.432; rs4147811, $p=3.80\times10^{-8}$, OR=0.745; rs481931, $p=8.14\times10^{-8}$, OR=0.750). Both 8q24 (rs987525, $p=1.11\times10^{-16}$, OR=1.781) and IRF6 at 1q32 (rs2073485, $p=1.07\times10^{-10}$, OR=0.691; rs2013162, $p=2.29\times10^{-11}$, OR=0.705; rs861020, $p=1.20\times10^{-9}$, OR=1.437; rs10863790, $p=1.11\times10^{14}$, OR=0.582) also attained genome-wide significance. Stratified analyses based on European and Asian ancestries suggested that some loci were population-specific. Trios

of European ancestry gave the strongest signal for markers at 8q24 (pEuropean=5.0×10^{-14}, OR=2.01 and pAsian=0.0089, OR=1.39), whereas trios of Asian ancestry gave the strongest association for markers in MAFB, ABCA4 as well as IRF6 (Figure 2.5). Three other genes, PAX7 at 1p36, VAX1 at 10q25.3 and NTN1 at 17p13, also showed a trend towards genome-wide significance. Direct DNA sequencing of MAFB revealed a rare, damaging missense variant, p.His131Gln; subsequent genotyping of this variant in Filipinos, showed that this variant was associated (p=0.0003) with NSCL/P. Direct sequence analysis of ABCA4 also yielded 26 missense variants, many of which were predicted to be damaging. Whole-mount in situ hybridization analysis and immuno-detection of expressed mafb mRNA in murine embryos showed that the expression of both mafb mRNA and protein was observed in both craniofacial ectoderm and neural crest-derived mesoderm during embryonic day E13.5 to E14.5. The epithelium that appears near the palatal shelves and MEE during palatal fusion showed the strongest expression levels. Moreover, mafb expression was stronger in the oral epithelium than the mesenchymal tissues after palatal fusion. ABCA4 expression studies exhibited none of the observations for MAFB (Beaty et al., 2010).

A meta-analysis of the two largest GWAS (Mangold et al., 2010; Beaty et al., 2010) revealed additional NSCL/P susceptibility loci. In meta-analysis of all European case-control and TDT data (meta-analysisEuro), six chromosomal loci obtained genomewide significance with NSCL/P (Figure 2.6a): 8q24 (rs987525, $p=3.94\times10^{-34}$, RR=2.074), rs7078160 (VAX1) at 10q25 ($p=2.81\times10^{-8}$, RR=1.459), rs227731 (NTN1) at 17q22 ($p=4.26\times10^{-8}$, RR=1.274), rs7590268 (THADA) at 2p21 ($p=4.05\times10^{-8}$, RR=1.419), rs8001641 (SPRY2) at 13q31 ($p=6.20\times10^{-10}$, RR=1.461) and rs1873147

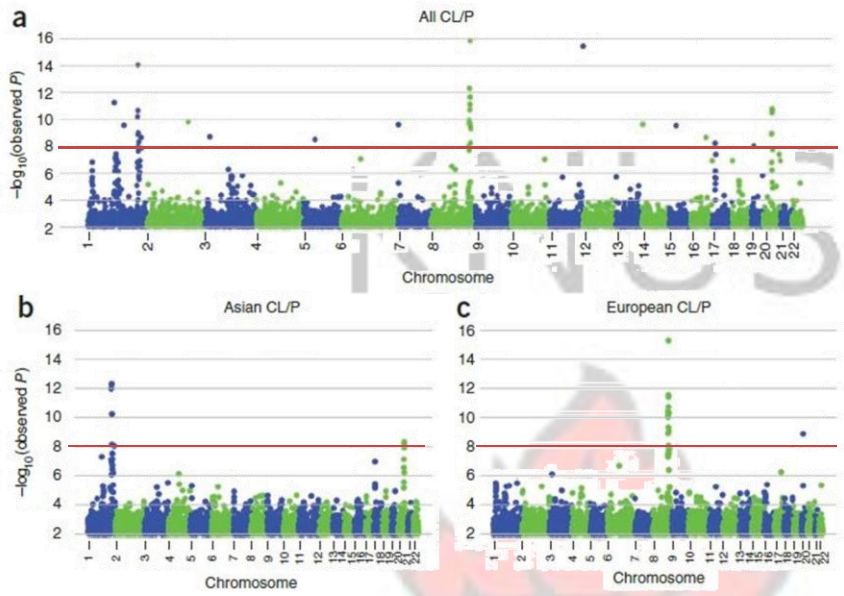


Figure 2.5: Manhattan plots for GWAS. a- result based on all 1,908 NSCL/P trios, b- results based on 1,038 NSCL/P cases of Asian ancestry and c- result based on 825 caseparent trios of European ancestry (Beaty et al., 2010).

(TPM1) at 15q22 ($p=2.81\times10^{-8}$, RR=1.467). In meta-analysis of all European and Asian trios data (meta-analysis$_{All}$), five out of the six regions implicated in meta-analysis$_{Euro}$ showed even higher level of significance (Figure 2.6b), presupposing these loci increased susceptibility to NSCL/P in both ancestries. However, 15q22 in metaanalysis$_{Euro}$ did not demonstrate genome-wide significance in meta-analysis$_{All}$, presupposing that that region may be European ancestry-specific. Moreover, metaanalysis$_{All}$ yielded six additional loci: rs560426 of ABCA4 at 1p22.1 ($p=3.14\times10^{-12}$, RR=1.420), rs861020 of IRF6 at 1q32.2 ($p=3.24\times10^{-12}$, RR=1.443), rs13041247 of MAFB at 20q12 ($p=6.17\times10^{-9}$, RR=0.837), rs742071 of PAX7 at 1p36 ($p=7.02\times10^{-9}$, RR=1.316), rs7632427 of EPHA3 at 3p11.1 ($p=3.90\times10^{-8}$, RR=0.731) and rs12543318 of DCAF4L2 at 8q21.3 ($p=1.90\times10^{-8}$, RR=1.272) (Ludwig et al., 2012).

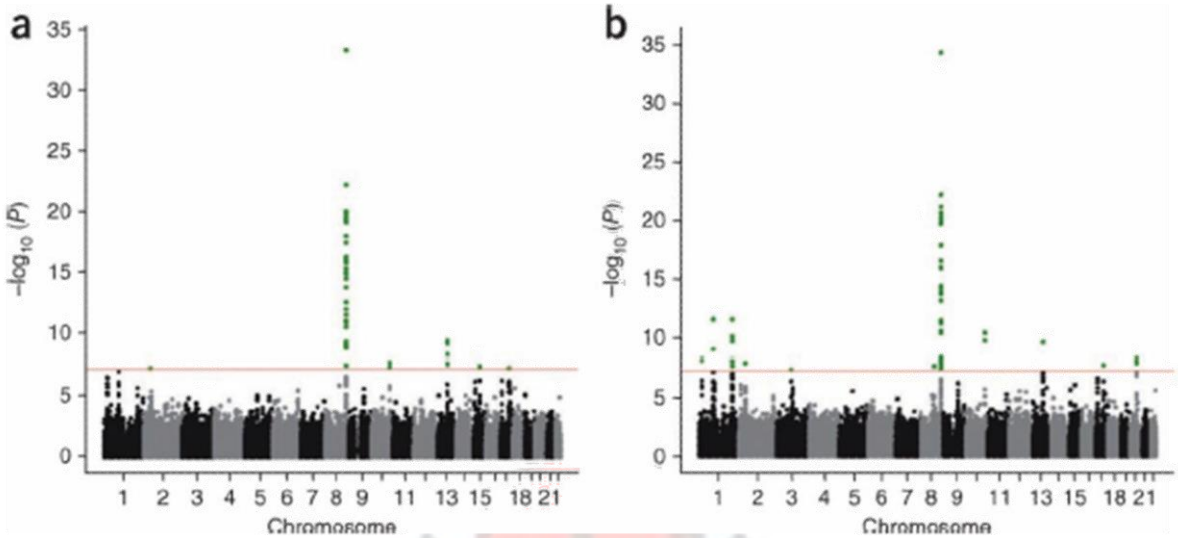


Figure 2.6: Manhattan plot for meta-analyses. a: Genome-wide meta-analyses for NSCL/P pedigrees of European ancestry, b: Meta-analyses for combined European and Asian pedigrees (Ludwig et al., 2012).

Loci implicated in meta-analysis$_{All}$ are thought to play diverse roles in NSCL/P susceptibility. PAX7 plays a role in craniofacial development, and seven PAX7 variants had been observed in NSCL/P cases, two of which had strong parent-of-origin effects. EPHA3 (rs7632427) regulates cell shape and cell-cell contact; also, a highly methylated region was observed approximately 200bp downstream of rs7632427, suggesting the need to study the role of methylation patterns in craniofacial development. Per Encyclopedia of DNA Elements (ENCODE), the associated SNP (rs1873147) of TMP1 has strong enhancer and promoter signatures as well as many transcription factorbinding sites. No functional genetic element, however, was found for the loci at 8q21.3 (Ludwig et al., 2012).

A fifth GWAS implicated ADCY9 in the aetiology of NSCL/P in Chinese. A meta-analysis of both discovery and replication phase cohort revealed a new NSCL/P susceptibility locus at 16p13.3 (rs8049367) that is situated between CREBBP and ADCY9 ($p=8.98\times10^{-12}$; OR=0.74). Deletions that involve CREBBP-ADCY9 region have

been associated with Rubinstein-Taybi Syndrome which is characterized by unique facial features that include CLP. But ADCY9 was favourite gene at this locus because microarray expression analyses yielded corroborative evidence: ADCY9 (but not CREBBP) and NTN1 demonstrated significantly up-regulated levels in dental pulp stem cultures of NSCL/P patients. Further functional analyses of this region showed no functional genetic elements, such as enhancer. However, rs2262251, which is in strong LD (r^2=0.96) with rs8049367, is situated in the exon of the long non-coding RNA (lncRNA) RP11-462G12.2. A potential interaction between lncRNA and CREBBPADCY9 region was observed, suggesting the regulation of 16p13.3 locus by lncRNA may be a possible mechanism by which this locus initiates cleft pathogenesis. Already known loci at 1q32.2, 10q25.3, 17p13.1 and 20q12 were also confirmed to confer susceptibility to NSCL/P in Chinese populations. It was also observed that rs2235371related haplotype at 1q32.2 may be more relevant in conferring susceptibility to NSCL/P in Chinese populations than the enhancer variant rs642961 (Sun et al., 2015).

A sixth GWAS that involved both human and dog subjects suggested a role of ADAMTS20 in cleft aetiology. In dogs, a genome-wide significance was observed between CL/P and a locus on canine chromosome 27. Further whole-genome sequencing of 3 CL/P and Syndactyly (CLPS) and 4 controls at 15X coverage revealed a frameshift variant (c.1360_1361delAA; p.Lys453Ilefs*3) in ADAMTS20 that segregated harmoniously with the CLPS phenotype. In a contemporaneous study involving human subjects, DFAM analysis gave a suggestive association (rs10785430, $p=2.67\times10^{-6}$) between CL/P and ADAMTS20. Subsequent DNA sequencing of 20 Guatemala CL/P cases showed no causative variants in the coding region of ADAMTS20 though four novel risk coding variants (V586A, K601R as well as Q1353H in nonsyndromic cases and

A1108T in a syndromic) were observed in Filipinos. ADAMTS20 is a member of secreted zinc metalloproteases that have been implicated in the cleavage of extracellular matrix (ECM) proteins and processing of procollagen. The protein particularly cleaves proteoglycan and versican of ECM, and acts in diverse biological processes, including melanoblast survival, palatogenesis and inter-digital web regression. Moreover, murine craniofacial expression of Adamts20 occurs in first pharyngeal arch, between nasal processes and a higher expression in palatal mesenchyme where it is involved in sculpting and extension of palate (Wolf et al., 2015).

2.4.1.2 Major replication studies of GWAS signals

A replication study of GWAS signals by Beaty et al (2013) revealed new susceptibility loci for NSCL/P; the cohort included Europeans (Germans, Turkish and Central Europeans) and Asians (Filipinos). GWAS hit loci (IRF6, MAFB, ABCA4 and 8q24) were successfully replicated; however, Europeans gave a stronger signal for 8q24 whiles Filipinos showed marginal significance. Other loci (PAX7, THADA – rs4372955, COL8A1/FILIP1L, DCAF4L2, GADD45G, NTN1, RBFOX3 and FOXE1) also demonstrated evidence of association with NSCL/P. In gene-environment (G×E) interactions tests, two genes (GRID2 and ELAVL2) demonstrated evidence of interaction with maternal smoking in Europeans. A SNP in NTN1 (rs11650357) also demonstrated evidence of possible interactions with three SNPs in IRF6, the most significant SNP being rs6685182 (Beaty et al., 2013).

A possible role of parent of origin effects (POO) in cleft aetiology has been suggested in Europeans and Asians. POOs, such as imprinting, are phenomena whereby the effects of mutations are determined by parental origin. A locus exhibited marginal significance:

maternal-specific transmission bias was observed for rs12543318 at 8q21.3 (Garg et al., 2014).

2.4.1.3 The role of IRF6 gene in nonsyndromic clefts aetiology

Human interferon regulatory factor 6 (IRF6) gene has 9 exons, with only 7 exons (excluding exon 1 and 2) being coding exons. There are multiple transcription variants of IRF6 due to alternate splicing. IRF6 gene encodes a protein of 467 amino acids which is a member of interferon regulatory factor (IRF) family. Members of this family have a highly-conserved tryptophan-pentad, N-terminal helix-turn-helix DNAbinding domain and a less-conserved C-terminal protein-binding domain. IRF6 protein is a transcription activator and has cardinal role in keratinocyte proliferation-differentiation switch which is linked to normal epidermal growth. IRF6 is also involved in regulating mammalian epithelial cell proliferation as well as regulation of the transcription of

WDR65 gene. Variants in IRF6 gene have been associated with syndromic (VWS and PPS) and non-syndromic (OFC6) OFCs (http://genome.ucsc.edu, 22/01/2015).

A ground-breaking study demonstrated IRF6's role in the aetiology of NSOFCs in Asians, Europeans and South Americans. The Val274Ile SNP, rs2235371, showed strong over-transmission of valine (V) allele in the total data set ($p<10^{-9}$); Asians and South Americans exhibited the highest significance. Val274Ile SNP was responsible for 12% of genetic contribution to NSCL/P. Moreover, this locus tripled the recurrence risk of NSCL/P in families with an affected child (Zuchero et al., 2004).

Alleles of IRF6 exhibit variable expressivity and penetrance. Because lip pits do not occur in all cases of VWS, IRF6 mutations can result in phenotypes that are identical to NSOFCs. In sporadic and familial NSCL/P, IRF6 mutations have been observed in two

families with mixed clefting pattern, presupposing that IRF6 mutations do have a role in familial NSOFCs (Rutledge et al., 2010). In syndromic and multiplex NSCL/P families, IRF6 mutations occurred in 62.7% of syndromic and 3.3% of NSCL/P cases. However, in one a priori NSCL/P family with an autosomal dominant inheritance, new inquiries into family history showed the presence, at birth, of lower lip pits in two members and diagnosis was revised as VWS (Desmyter et al., 2010).

Association of IRF6 variants with NSCL/P have been confirmed in the Italian population. Italian triads were genotyped for four markers that span the IRF6 locus, followed by transmission disequilibrium test (TDT). Strong associations were observed between markers and NSCL/P in both single-allele (p=0.002, rs2235375) and haplotype (p=0.0005) analyses (Scapoli et al., 2005). In predominantly White and Hispanic families (72 multiplex and 184 simplex families), family-based association test (FBAT) analysis showed association between entire data set and C allele of rs2013162. Moreover, haplotype-based association test (HBAT) showed a significant overtransmission of haplotypes with C allele of rs2013162 as opposed to those with A allele.

These observations, were however, varied in stratified analyses that were ancestry-based (Blanton et al., 2005).

The IRF6 gene may act as a genetic modifier for NSCL/P in the Belgian population. In genotyping of 195 trios on rs2013162 and rs2235543, IRF6 contributed to the aetiology of sporadic NSCL/P even when no pathogenic coding variants were found in the gene by acting as a genetic modifier. Moreover, pathogenic coding variants were observed in about 80% of European VWS cases when direct DNA sequencing was carried out on the whole-exome of IRF6 (Ghassibe et al., 2005).

IRF6 variants may also be risk alleles for NSCL/P in the Norwegian population. A population-based case-control approach was used to study 377 NSCL/P and 196 nonsyndromic CP cases as well as 763 infant-parent triad controls. Six SNPs in IRF6 were genotyped. Grossly, strong association ($p<0.001$) was observed for NSCL/P but not isolated CP. In single-marker analyses, mothers with a double-dose of 'A' allele at rs4844880 had an increased risk of having a child with CL/P. A relative risk (RR) of 0.38 was observed when the patient carried a single-dose of 'A' allele at rs2235371. In HBAT, many foetal and maternal haplotype relative risks were significant individually but were not strong enough to affect the overall test (Jugessur et al., 2008).

IRF6 has also been implicated in the aetiology of NSCL/P in diverse populations. In a cohort involving European American, Taiwanese, Singaporean and

Korean case-parent trios, 13 SNPs in IRF6 were genotyped, followed by TDT analyses. Evidence of association was observed in all four populations. Two specific haplotypes, GC of rs2235373-rs2235371 and AAG of rs599021-rs2235373-rs595918, exhibited the most significant association in Taiwanese cases. The AGC/CGC haplotype of rs599021rs2235373-rs2013162 exhibited almost a 7-fold increase in risk among the Taiwanese sample, $P<10^{-3}$ (Park et al., 2007).

Pathogenic IRF6 mutations may not always be observed in NSCL/P and VWS cases. In 155 Taiwanese OFC cases (31 syndromic cases that included 19 VWS, 44 NSCL/P multiplex as well as 80 NSCL/P simplex families), a PCR-based mutation analysis for the entire coding regions of IRF6 was done. Eleven mutations were observed in 57.89% of VWS patients but no mutation was observed for the other samples. Five novel VWS mutations were observed: p.Tyr97Cys, p.Gln120HisfsX24,

p.Glu136fsX3, p.Thr291Pro and p.Trp323X (Wu-Chou et al., 2013).

IRF6 variants are also associated with NSCL/P in Malaysia. Allele and haplotype TDT analyses were done on genotyped IRF6 markers. A strong transmission distortion for multiple haplotypes to NSCL/P patients was observed. Haplotypes that carried the 243bp allele of D1S2136 and common alleles at rs861019 and rs2235371 were over-transmitted to patients. However, haplotypes that consisted of the 251bp allele of D1S2136 and rare allele at rs2235371 were more under-transmitted. In addition, many variants and haplotypes exhibited excess maternal transmission, yet none of them was statistically significant in maternal relative risk analyses (Salahshourifar et al., 2012).

IRF6 gene variants may interact with other genes to increase susceptibility to NSCL/P in China. Nine IRF6, 2 MSX1 and 8 PAX9 tag SNPs were genotyped for 204 NSCL/P cases and 226 controls. Significant association was observed for rs2073485, rs2235371, rs2236909 and rs861020 in IRF6 as well as rs17176643 in PAX9. Gene-gene interaction analyses showed that combination of rs2073485, rs2235371 or rs2236909 in IRF6 and rs17176643 in PAX9, increased the risk of NSCL/P (Song et al., 2013).

IRF6 may interact with some environmental factors to produce a cleft phenotype. Five IRF6 SNPs were genotyped in 107 case-parents trios and 100 normal controls of Han Chinese; strong association was observed between these markers and NSCL/P in both single SNP analysis (A allele at rs4844880, G allele at rs2073485, and C allele at rs599021) and haplotype analysis (e.g. A-A for rs861019 and rs4844880). During first trimester of pregnancy, maternal passive smoking and medication increased the risk of NSCL/P, whereas maternal vitamins supplementation (such as folic acid) was a protective factor. Interactions between maternal abortion history and TT genotype of rs2235373 were statistically significant (Jia et al., 2009). A subsequent study that

employed 326 Chinese case–parent trios reinforced the interaction between IRF6 and multivitamin supplementation as well as environmental tobacco smoke. Two SNPs

(rs2076153 and rs17015218) portrayed evidence of interaction with multivitamin. Moreover, rs1044516 exhibited evidence of interaction with maternal environmental tobacco exposure. Haplotype analysis gave a similar hint (Wu et al., 2010).

Mutations in regulatory elements in the IRF6 gene are also associated with OFCs. A total of 41 multispecies conserved sequences (MCSs) were directly sequenced and screened for potential aetiologic variants in 184 NSCL/P cases and same number of controls from Iowa and Philippines. The rs642961 SNP (G>A) was observed in an IRF6 enhancer. The A allele was significantly over-transmitted ($p=10^{-11}$) in families with NSCL/P, particularly CL. Moreover, a dosage effect was observed for A allele, with a relative risk for CL of 1.68 for the AG genotype and 2.40 for the AA genotype. Functional assays showed that the risk allele disrupted the binding site for transcription factor AP-2α in the enhancer. Expression studies in mouse models also localized the enhancer's activity to craniofacial and limb structures. These finding suggest that IRF6 and AP-2α are in the same developmental pathway and that this high-frequency variant in the regulatory element contribute hugely to the aetiology of common but complex OFCs (Rahimov et al., 2008).

Various studies support the role of rs642961 in NSCL/P aetiology. In another study that involved 175 NSCL/P cases and 160 controls from northern China, significant association was observed for rs642961, but not rs2235371 (Shi et al., 2011). Analyses of 106 NSCL/P case-parents trios and 129 control trios from China also demonstrated association between NSCL/P and rs642961 as well as rs2235371 in allele and genotypic frequencies analyses, though only rs642961 still maintained association in FBAT analysis. There was an

apparent dosage effect of allele A at rs642961. The transmission of major G allele of rs2235371 and minor A allele of rs642961 was in linkage disequilibrium (LD) in complete case-parent trios (Zhou et al., 2013). In a Polish FBAT study that included 175 NSCL/P cases, association was observed between NSCL/P and rs642961 as well as rs987525; only rs987525 remained significant after Bonferroni correction (Mostowska et al., 2010). In a Brazilian study, an association was observed between rs642961 and NSCL, with the aetiologic A allele exhibiting dominant model of genetic inheritance (Brito et al., 2012). In direct sequencing of 72 Swedish NSCL/P, 24 Finnish CP and 24 VWS/PPS families and genotyping of rs642961 and rs2235371, no significant association was observed between NSCL/P and these two SNPs. However, aetiologic variants were observed in six out of seven VWS/PPS families sequenced. Interestingly, association was observed between A allele of rs642961 and VWS/PPS as well as CP samples. In CP subset, the G-C haplotype of rs642961-rs2235371 was a risk haplotype (Pegelow et al., 2014). In Hispanic and non-Hispanic white multiplex (122) and simplex (308) NSCL/P families, rs642961 showed modest association whereas rs2235371 exhibited strong association with NSCL/P, particularly in non-Hispanic white simplex families (Blanton et al., 2010). In FBAT analyses, rs642961 did not show association in an Iranian cohort of 352 individuals from 102 Iranian nuclear families (Nouri et al., 2014). In 150 NSCL/P cases and 150 controls, rs642961 was shown, however, to be a marker of severity for NSCL/P in the Iranian population. Though no association was observed for the whole data set, a tagSNP haplotype (AGGT) that carried the risk allele of rs642961 was significantly over-transmitted in individuals with the most severe subphenotype of NSCL/P, complete bilateral CL/P (Kerameddin et al.,

2015).

In linkage analyses involving 820 NSCL/P multiplex families, comprising of 6,565 individuals from Philippines, Colombia, China, India, Turkey and USA, genomewide-significant linkage results were observed for 1q32, 2p13, 3q27-28, 9q21, 12p11, 14q21-24 and 16q24. SNPs in IRF6 (1q32) and in or near FOXE1 (9q21) reached formal genome-wide significance association in a replication phase study. Moreover, results were phenotype-dependent: IRF6 and FOXE1 association were most significant for families with CL and predominant CLP, respectively. These observations pinpoint the relevance of meticulous phenotypic delineation in large samples of families for genetic analyses of complex, heterogeneous traits such as OFCs (Marazita et al., 2009).

A meta-analysis of 20 case-control studies further corroborates the role of IRF6 in cleft aetiology. The A allele of rs2235371 had a significantly decreased risk whereas the A allele of rs642961 had an increased risk for NSCL/P in contrast to the G allele. The A allele of rs987525 was significantly associated with NSCL/P in contrast to the C allele. Moreover, in a stratified ethnic and subphenotypes analyses, significant associations were still achieved (Wang et al., 2012).

IRF6's association with NSCL/P may not be universal. In 128 NSCL/P cases and 105 controls from Kenya, genotyping of GWAS SNPs at 1p22.1, 1q32.2, 8q24, 10q25.3, 17q22 and 20q12, yielded no significant association (Wheatherley-White et al., 2011). Also, in a Nigerian cohort, no association was observed for these loci, though MSX1 showed evidence of association with NSCL/P (Butali et al., 2011). The small sample sizes for these studies as well as population heterogeneity could also account for these deviant observations.

IRF6 may interact with other genes during cleft pathogenesis. Earlier tooth agenesis studies suggested IRF6 and TGFα interact, since tooth agenesis is often observed in NSCL/P cases.

Genotyping of SNPs in and around IRF6 and TGFα in Brazilians, other South Americans, Latvians and individuals from several cohorts, demonstrated that IRF6 and TGFα were associated with NSCL/P in the Brazilian cohort.

Interestingly, IRF6 was also associated with CP with impaction of permanent teeth. Moreover, interaction between IRF6 and TGFα was found in all data sets. IRF6 and TGFα were also associated with subsets of NSCL/P with defined dental anomalies. Also, Tgfα was not expressed in palatal tissues of Irf6 knockout mice (Letra et al., 2012a). In case-control and case-parent analyses in northeast Chinese, TGFα (rs3771494, rs3771523 and rs11466285) and IRF6 (rs2235371 and rs2013162) exhibited significant association with NSCL/P in case-control and FBAT analyses (Lu et al., 2013). In 104 cases and 606 controls from Baltic populations, rs17389541 of IRF6 and rs1793949 of COL2A1 exhibited association with increased risk of NSCL/P. Multiple haplotypes in COL2A1, COL11A2, WNT3, FGFR1 and CLPTM1 also showed association with NSCL/P. The strongest epistatic interactions occurred between MSX1 and BMP2, FGF1 and PVRL2 as well as between COL2A1 and FGF2 genes (Nikopensius et al., 2010).

Animal models have also confirmed the role of IRF6 gene in cleft aetiology. Irf6 mutant mice show a hyper-proliferative epithelium which does not undergo terminal differentiation, resulting in multiple epithelial adhesions and subsequently defect in palatine shelves elevation that can shut the oral cavity, leading to CP. This suggests IRF6 is a cardinal determinant of the keratinocyte proliferation-differentiation switch (Ingraham et al., 2006). IRF6 also plays a crucial role in the formation of the oral periderm, spatiotemporal regulation of which is overly important for ensuring correct palatal adhesion (Richardson et al., 2009). A combination of gene expression analyses, immunoprecipitation, luciferase reporter assays and mouse genetics has established that

IRF6 is a direct transcription target of p63, which has been implicated in many malformation syndromes that exhibit OFCs as hallmark features. p63 activates IRF6 transcription through IRF6 enhancer element whose variants raises susceptibility to CL (Thomason et al., 2010). Murine models suggest that smad4-irf6 interaction and tgfβmediated irf6 signaling cascade are crucial for palatal fusion. Haploinsufficiency of Irf6 in mice with basal epithelial-specific deletion of TGFβ signaling mediator Smad4 leads to compromised p21 gene expression and MEE persistence, resulting in polydactyly that is in consonance with anomalous toe and nail phenotypes in individuals with VWS and PPS. Thus, IRF6 and SMAD4 synergistically regulate the fate of MEE, and TGFβmediated Irf6 activity is responsible for MEE degeneration during palatal fusion in mice

(Iwata et al., 2013).

2.4.1.4 ARHGAP29 gene

Human ARHGAP29 is located on chromosome 1 and has 23 exons, out of which 22 are coding exons. It has many transcript variants due to alternative splicing. It codes for a protein of 1261 amino acids and acts as a GTPase activator for Rho-type GTPases by converting them into an inactive GDP-bound state. The protein has strong activity toward RHOA and a weaker one toward RAC1 and CDC42. It may also act as an effector of RAP2A to regulate Rho. When ARHGAP29 work in concert with RASIP1, Rho signaling is suppressed and they also dampen ROCK1 and MYH9 activities in endothelial cells; they are also involved in blood vessel tubulogenesis. ARHGAP29 protein interacts with PTPN13/PTPL1 as well as with RAP2A and RASIP1 through its coiled domain. ARHGAP29 is widely expressed: it is particularly highly expressed in skeletal muscle and

heart. However, it is also expressed at intermediate levels in placenta, liver and pancreas, and weakly expressed in brain, kidney and lung (http://genome.ucsc.edu, 26/03/2015). RHOA plays diverse roles, including cellular shape, movement, cell-cell interactions and proliferation, which are all crucial for craniofacial development (Birnbaum et al., 2009). RhoA is a downstream effector of both Tgfb and Wnt signaling pathways. These have all been linked to craniofacial development through the activities of Wnt5a, Wnt9b and Tfgb3 in knockout mice with craniofacial defects that encompass CL/P (Leslie et al., 2012).

ARHGAP29 has been shown to be the cleft aetiologic gene at 1p22.1, instead of ABCA4. Since expression and mutation analyses failed to validate ABCA4 as a candidate gene at 1p22.1 (Beaty et al., 2010), the focus shifted to ARHGAP29, an adjacent centromeric gene. Direct sequencing of the coding exons of ARHGAP29 yielded eight potentially deleterious mutations. A frameshift variant (S21Yfs*20) emanating from a 2bp deletion, occurred in a bilateral CLP proband and an affected sibling. A nonsense variant, p.Lys326X, was detected in a bilateral CLP proband and the proband's clinically unaffected mother and grandfather. Two missense variants, p.Thr26Ala and p.Ala832Thr, were observed in individuals with CL and an unaffected parent. Two other missense variants, p.Thr622Met and p.Ile845Val, were also observed in probands with CLP, with the former variant also occurring in the proband's father who had CLP. Two novel, very rare missense variants, p.Lys46Arg and p.Arg1142Gln, were detected in the US population in individuals with CLP. Additionally, two rare missense variants were observed in Filipino cases and controls. These included p.Val1202Leu, which had a frequency of 1.3% in cases and 1.4% in controls, and p.Arg616His, which had a frequency of 4.0% in cases and 3.7% in controls (Leslie et al., 2012).

ARHGAP29 exhibits craniofacial expression, which is reduced in irf6 null mice. During in situ hybridization of murine embryos, Arhgap29 showed strong expression in the medial and lateral nasal processes whereas less intense expression was detected in the mandibular and maxillary processes at E10.5 and the shelves of the secondary palate at E13.5. Arhgap29 transcript and protein were identified in the epithelium and mesenchyme of coronal head sections at E10.5 and E14.5. This spatiotemporal expression pattern of Arghap29 is strongly in consonance with its implicated role in craniofacial development. Moreover, Arhgap29 expression was reduced in all epithelia of irf6 null embryos, presupposing that Arhgap29 may have its activity downstream of irf6. Western blot analysis of E17.5 cutaneous extracts also showed 2.8 fold reductions in Arhgap29 expression in irf6-deficient skin relative to the wildtype, presupposing Arhgap29 is a mediator of irf6 signaling pathway. These observations suggest a novel pathway in which IRF6 gene regulatory network interacts with the Rho pathway through ARHGAP29 (Leslie et al., 2012).

A number of studies suggest that ARHGAP29 variants may exert their effects through a loss-of-function or gain-of-function paradigms. Recently, expression profiling of dental pulp stem cell cultures obtained from individuals with NSCL/P exhibited a nearly 3-fold reduction in ARHGAP29 in contrast to cultures from unaffected controls (Bueno et al., 2011). This presupposes that loss of a GAP function would maintain Rho in an active, GTP-bound form, effectively increasing Rho activity. This could adversely alter cellular migration during embryonic development, and may showcase one possible mechanism by which ARHGAP29 could play a role in the aetiology of NSCL/P. Moreover, a recent study observed that treatment of palate cultures with nocodazole, a drug that destabilizes microtubules, led to heightened RhoA activity, resulting in the formation of a multi-

layered, hypertrophied MEE and lack of palatal fusion. It was further shown that the regulation of microtubule dynamics and actin microfilaments is essential for the remodeling of the palatal MEE during palatogenesis (Kitase and Shuler, 2012). This may also pinpoint another mechanism through which ARHGAP29 may be participating in the aetiology of NSCL/P. Other studies, however, suggest a gain-offunction mechanism which may result in a reduction of Rho activity, the amount of Rho available, or spatial-temporal dysregulation of crucial elements of craniofacial development (Auerbach et al., 2011).

DNA sequence analyses in NSCL/P cases supports ARHGAP29 role in cleft aetiology. In 60 NSCL/P cases from India, a novel nonsense mutation, p.Lys32X, was observed in the heterozygous state in just a case (Chandrasekharan and Ramanathan, 2014). In a cohort of 240 NSCL/P from Nigeria and Ethiopia, two missense variants in ARHGAP29, p.Leu913Ser and p.Arg955His, were found in some individuals (Butali et al., 2014a).

Association between ARHGAP29 and NSCL/P has been replicated in other studies. In 314 Hispanic and 507 non-Hispanic white multiplex and simplex families, strong association was observed between NSCL/P and three SNPs (rs1541098, rs3789688 and rs1576593) in non-Hispanic families (Letra et al., 2014). In a cohort of 1,409 trios from Europe, United States, China and Philippines, targeted sequencing of GWAS loci and TDT analysis replicated the association between NSCL/P and rs560426 in Asians. A subsequent conditioning on rs560426 revealed a second signal, rs77179923. These two peaks are situated within introns of ABCA4 and harbour a craniofacial enhancer. Four nonsense mutations, p.Lys32X, p.Arg647X, p.Trp789X and p.Gly1040X were over-transmitted to the affected children (Leslie et al., 2015a).

2.4.1.5 VAX1 gene

VAX1 codes for a homeo-domain containing transcription factor that is conserved in vertebrates. Genes of the same family as VAX1 are called HOX genes; these are the most conserved genes and are usually found in special gene clusters. HOX genes are associated with the regulation of body development and morphogenesis. VAX1 is a member of VAX subfamily and is closer to the EMX homeobox gene family. VAX1 has two isoforms in humans. The first isoform consists of 3 exons and encodes a protein of 334 amino acids. This isoform's homeodomain is located in 100-159 amino acid positions; it also has Alanine-rich region located in the 216-253 amino acid positions.

The second isoforms has 4 exons, and encodes a protein of 186 amino acids (http://genome.ucsc.edu, 22/01/2015).

VAX1 gene is precisely expressed in the developing forebrain and optic nerve. Vax1 knockout mice exhibit dysgenesis of the optic nerve, have colobomas, malformed basal telencephalon and lobar holoprosencephalon. Using molecular markers, it has been shown that in the developing visual system, Vax1 mutations result in a proximal expansion of the activity of Pax6 and Rx genes, presupposing that Vax1 protein may interfere negatively with the expression of Pax6 and Rx genes. Therefore, Vax1 seems to work in concert with Pax6, Rx, Bmp4 and Shh in the developing embryo to establish the structures of the visual system. Mice having homozygous mutations in Vax1 display craniofacial malformations such as CP (Hallonet et al., 1999, Zhao et al., 2010). VAX1 is involved in human craniofacial development. The coding exons of VAX1 and VAX2 were sequenced in 70 patients with anophthalmia and microphthalmia. In VAX1, two successive homozygous substitutions (c.453G>A and c.454C>A) that resulted in a novel aetiologic missense mutation (p.Arg152Ser) were observed in an Egyptian male patient with diverse

phenotypes: bilateral microphthalmia, small optic nerve, corpus callosus agenesis, malformed hippocampus, lack of pineal gland and bilateral CLP. The pArg152Ser missense mutation was predicted to affect the conserved residue in the third helix of the homeodomain of VAX1 (Slavotinek et al., 2012).

Recent observation suggests that VAX1 protein functions as an activator of a powerful dominant-negative isoform of Tcf12, designated dnTcf12; Tcf12 is a mediator of WNT signaling. VAX1 protein binds to and activates an internal promoter located in the fifth intron of Tcf12. dnTcf12 is deficient in the activating beta-catenin domain of Tcf12 and therefore functions as a strong repressor of WNT target genes. This dominantnegative antagonist is globally expressed in the developing forebrain and its morpholinomediated loss in Xenopus leads to embryos without anterior head region. Thus, wildtype VAX1 protein stimulates the expression of TCF12 mRNA, while the mutant form does not (Vacik and Lemke, 2011). These occurrences presuppose that one way VAX1 p.Arg152Ser mutation may exert its effect is the inability of this mutant to activate dnTCF12 transcription. This results in de-repression of dnTCF12 target genes and hyper-activation of WNT signaling (Slavotinek et al., 2012).

VAX1's association with NSCL/P may not be universal. NSCL/P's association with rs7078160 has been successfully replicated in targeted researches that involved Europeans (Nikopensius et al., 2010b; Mostowska et al., 2012a) and Mesoamericans (Rojas-Martinez et al., 2010). However, two minor studies involving Kenyan (Weatherley-White et al., 2011), Congolese (Figueiredo et al., 2014) and Chinese (Pan et al., 2011) could not replicate the association between NSCL/P and rs7078160. In sequencing of the two isoforms of VAX1 in Central Europeans, 17 rare variants were observed, 7 (p.Asp72Asn, p.Pro92Leu, p.Arg180Pro, p.Leu181Met, p.Glu183Ala,

p.Gly185Asp and p.Ala201Thr) of which were functionally relevant, as they occurred in highly conserved evolutionary nucleotides in the C-terminal domain of the long isoform of VAX1 (Nasser et al., 2012).

Association between VAX1 and NSCL/P has also been established in Asians. In 5,241 individuals, including triads from multiplex families from Mongolia, Japan, Iowan and Philippines, significant associations were found between rs7078160 ($p=2.7\times10^{-6}$) as well as rs4752028 (p=0.0002) and NSCL/P in both Mongolian and Japanese cohorts.

Also, significant association was found between CLP subphenotype and rs7078160 (p=0.001) in multiplex Filipino families. Also, there was strong association between VAX1 (prs7078160=2.3×10^{-8} and prs4752028=0.00004) and NSOFCs in combined Asian cohorts. Moreover, CP also showed significant association with rs7078160 in Asians. Moreover, maternal POO effect was observed for rs7078160 in combined Mongolian and Japanese cohorts. Twelve new functional rare variants were also observed in VAX1. Of these, rs14874160G>T exhibited significant association (p=0.007) with CLP in Filipinos. Two new missense variants, p.Met117Arg and p.Ala233Ser, as well as splice acceptor variant (c.96171G>T) were also observed in Mongolian cases. The p.Met117Arg mutation was predicted to be probably damaging. Also, a proline insertion from c.698_699insCCC (p.Ala233_Pro234insPro) was observed in a single Philippine case and both parents (Butali et al., 2013).

2.4.1.6 PAX7 gene

PAX7 of Homo sapiens has three transcription variants or isoforms, with transcription variant 3 having a total exon count of 9 and encodes a transcription factor consisting of 505 amino acids. PAX7 is a member of paired box (PAX) family of transcription factors.

Members of this family are characteristically made up of a paired box domain (an octapeptide) and a paired-type homeodomain. These genes are thought to play crucial roles in foetal development and cancer growth. PAX7 is thought to be involved in tumor suppression as its fusion with a forkhead domain family member has been linked with alveolar rhabdomyosarcoma (http://genome.ucsc.edu, 22/01/2015).

Pax7 is crucial for neural crest development and expression of the neural crest markers Slug, Sox9, Sox10 and HNK-1 in chick embryos in vivo (Basch et al., 2006). Deficiency in neural crest development has been linked to craniofacial malformations such as CL/P (Wilkie and Moriss-Kay, 2001). Pax7 has also been implicated in neural crest development in mammals: it contributes to many tissues such as cranial, cardiac and trunk neural crests (Murdoch et al., 2012).

Some studies support a role of PAX7 in cleft aetiology. In 297 trios from Maryland, Taiwan, Singapore and Korea, genotyping of 34 SNPs in PAX3, PAX6, PAX7 and PAX9, followed by TDT revealed that rs766325 in PAX7 was associated with NSCL/P. Moreover, rs618941 and rs553934 of PAX7 and four SNPs of PAX3 also exhibited significant excess maternal transmission (Sull et al., 2009). Using Asian and Iowan samples, association was established between PAX7 trinucleotide SNP rs6659735 and NSCL/P in an Iowan cohort for all clefts ($p=0.007$) and when only multiplex families were considered ($p=0.02$). Sequence analysis also revealed missense variants, including p.Gly24Glu, p.Gly411Arg, p.Pro397Leu, p.Gly412Ser, p.Gly466Ser, p.Ala3Thr, p.Ala15Thr, and p.Met387Leu; the first three were predicted to be 'probably damaging' by Polyphen-2 (Butali et al., 2013). Direct DNA sequencing of 220 NSCL/P cases from Nigeria and Ethiopia (Africa) showed novel missense (p.Asp428Asn) and splice-site (c.952+2T>A) variants in PAX7. Three previously reported missense variants were also

observed in this study: p.Gly466Ser in PAX7 and, p.Ser913Leu and p.Arg955His in ARHGAP29. A new rare variant, p.His165Asn, was seen in the MAFB gene (Butali et al., 2014a).

In targeted sequencing of 1p36 GWAS locus in 1,409 trios from Europe, United States, China and Philippines, rs1339062 (but not rs742071) was associated ($p=8.8\times10^{-4}$) with NSCL/P. A de novo, aetiologic missense variant, p.Ala259Val, was also observed. This variant altered DNA-binding ability of PAX7 protein: in electrophoretic mobility shift assays, wildtype PAX7 bound the probe more than p.Ala259Val mutant. Furthermore, in quantitative reporter assays that used HeLa cells transfected with a luciferase reporter vector, co-transfection of a plasmid that encoded wildtype PAX7 resulted in significantly higher levels of expression compared to a plasmid that encoded p.Ala259Val mutant (Leslie et al., 2015a).

2.4.1.7 BMP4 gene

There are three transcript variants of human BMP4 but all encode an identical protein of 408 amino acids. BMP4 is a member of BMP subfamily, a subset of TGFβ superfamily. The superfamily contains substantial families of growth and differentiation factors. Reduced expression of BMP4 leads to a variety of bone diseases, such as Fibrodysplasia Ossificans Progressiva. Thus, BMP4 plays a role in cartilage and bone formation, mesoderm induction, tooth development, limb formation and fracture repair. BMP4 also acts in concert with PTHLH/PTHRP to stimulate dental outgrowth during embryonic mammary development and to inhibit hair follicle induction. Mutations in BMP4 have been associated with some craniofacial anomalies: non-syndromic OFC11

(characterized by a paramedian microform CL) and microphthalmia syndromic type 6 (MCOPS6). MCOPS6 is characterized by microphthalmia or anophthalmia occurring with facial, genital, skeletal, neurologic and endocrine abnormalities (http://genome.ucsc.edu, 22/01/2015). Bmps are expressed at sites of epithelialmesenchymal interaction in several organs and also regulates tooth development at all stages. Bmp4 is expressed in probable epithelium at the commencement of tooth and palate development. Thus, epithelial signaling leads to mesenchyme induction of Bmp4 expression during normal palatal shelves development. Bmp4 also induces the expression of Bmp2, Bmp5, Msx1, Msx2 and Egr-1 (Lu and Hu, 2009).

Variants in BMP4 have been associated with NSCL/P. In 1,085 cases of microform clefts, sub-epithelial anomalies in orbicularis oris muscle (OOM) and overt CL/P from American, European and Asian populations, direct DNA sequencing of BMP4 revealed missense and nonsense mutations in 8 cases in all subphenotypes: p.S91C, p.R162Q, p.R287H, p.A346V, p.R168X, p.V152A, p.G168A and p.T102A (Suzuki et al., 2009). In Han Chinese, rs17563 of BMP4 interacts with environmental factors like maternal passive smoke exposures and multivitamin supplementation as well as paternal high risk drinking and smoking in NSCL/P pedigrees (Jianyan et al., 2010). In a Brazilian cohort of 383 NSCL/P cases and 450 controls, rs17563 was associated with NSCL/P ($p=0.007$), particularly in CL subphenotype, $p=0.00001$ (Antunes et al., 2013). Another Brazilian study of 123 NSCL/P case and 246 controls replicated the association between NSCL/P and rs17563 ($p=0.00018$), with the C allele being protective (Araujo et al., 2012). Direct DNA sequence analysis of BMP4 in 150 Chilean NSCL/P trios revealed no aetiologic coding variants. Subsequent genotyping of rs1957860 and rs762642, which encompass a promoter and an enhancer of BMP4, yielded an association with NSCL/P (Suazo et al., 2010a). Direct sequencing of these regulatory regions in 167 Chilean

NSCL/P cases and 336 controls revealed four novel mutations in BMP4.1 promoter, with three (c.5514G>A, c.5365C>T and c.5049C>T) being risk alleles. Furthermore, BMP4.2 promoter SNP rs2855530C allele carriers portrayed an increased risk for NSCL/P that was confined to males. These promoter variants generated binding motifs for various transcription factors: c.5514A for hematopoietic transcription factor GATA-1, c.5365T for RXR (linked to gene expression regulated by retinoic acid), c.5049T for TCF-1α (a canonical Wnt signaling effector which is expressed during mice mid-face development) and rs2855530G for Sp1 (a ubiquitous transcription factor). This may account for why teratogenic doses of retinoic acid induce CP in Rxr-α knockout mouse at a lower frequency than in wildtypes. Sp1 can modulate gene expression in cellular processes like differentiation, growth and apoptosis (Suazo et al., 2011).

2.4.1.8 FOXE1 gene

FOXE1, also called thyroid transcription factor 2 (TTF-2), is a member of forkhead family of transcription factors which characteristically have a distinct forkhead domain with which DNA-binding occurs. It encodes a polypeptide of 373 amino acids and may probably play a critical role in thyroid morphogenesis. The expression of the gene has been detected in adult brain, placenta, lung, liver, skeletal muscle, kidney, colon, heart, thymus, small intestine and testis. Mutant alleles of this gene have been associated with a number of conditions including congenital hypothyroidism and CP with thyroid dysgenesis. The gene may also play a role in squamous cell epithelioma, hereditary sensory neuropathy type 1 and Bamforth-Lazarus syndrome (BLS), which is characterized by thyroid agenesis or dysgenesis, hypothyroidism, spikey hair, choanal atresia, bifid epiglottis or uvula and CP (http://genome.ucsc.edu, 08/04/2015).

Various studies support a role of FOXE1 in the aetiology of NSOFCs. In a cohort of 388 multiplex families from USA, Philippines, Colombia and Europe, FOXE1 SNPs rs3758249 ($p=5.01\times10^{-13}$) and rs4460498 ($p=6.54\times10^{-12}$) were found to be associated with NSOFCs, including NSCP. The association between FOXE1 and NSOFCs was stronger for Caucasian populations. Maternal POO effects were observed for Colombian and Philippine cohorts for both NSCL/P and NSCP whereas a paternal POO effects were observed for Caucasian data (USA and Scandinavian data sets), possibly indicating different roles of these alleles in different populations. Expression studies also showed that FOXE1 was expressed in the epithelium that is undergoing fusion between the medial nasal and maxillary processes. Mutational screening yielded 32 coding and noncoding variants, including missense variants I59S and P208R (Moreno et al., 2009). Studies in other populations support a role of FOXE1 in the aetiology of NSCL/P and NSCP. In Central Europeans and Mayan Mesoamericans, significant association was observed between NSCL/P and rs4460498 and rs3758249. Further case-parent trios TDT analysis yielded significant association between rs4460498 and NSCP. A meta-analysis of all rs4460498 data revealed a genome-wide significance ($p=1.31\times10^{-08}$) for NSCL/P, which became more significant ($pNSOFC=1.56\times10^{-09}$) in all NSOFCs (Ludwig et al., 2014). A Thai cohort yielded no association with FOXE1 SNP rs111846096 but direct DNA sequencing revealed many missense variants (p.D92Y, p.P190R, p.P190L, p.R222C, p.G364S and p.P191R), mainly in NSCP cases (Srichomthong et al., 2013). In Hispanic Honduran and Colombian cohorts, significant association was observed between NSCL/P and rs3758249's G allele under additive, dominant and recessive models. FOXE1 SNP rs1443434's A allele also exhibited association for only additive and dominant models. ABCA4 and MAFB also demonstrated evidence of association with NSCL/P

(Lennon et al., 2012). In 300 NSCL/P cases and 606 controls from the Baltic region of Northeastern Europe, FOXE1 SNP rs7860144, as well as FGF1, TIMP2 and WNT9B, gained association with NSCL/P (Nikopensius et al., 2011). Employing a combination of zebrafish and mouse transgenesis, three enhancer elements that regulate Foxe1 expression in a tissue-specific pattern (oral epithelium or branchial arches) have been identified (Lidral et al., 2015). However, in 291 multiplex families from European, Asian and Hispanic ancestries, no association was found between NSCL/P and FOXE1 (Letra et al., 2010).

2.4.1.9 MSX1 gene

MSX1, also known as homeobox 7 (HOX7), encodes a protein which is a member of msh gene family. The gene has two exons, all of which encode a protein of 303 amino acids. MSX1 protein is a transcription factor that functions as a repressor during embryogenesis by interacting with members of core transcription complex and homeoproteins. The gene may also play a role in limb-pattern formation, craniofacial development (especially odontogenesis) and tumor growth inhibition. Variants in this gene have been associated with a number of genetic disorders: NSOFC5, Witkop syndrome, Wolf-Hirschom syndrome and autosomal dominant hypodontia. The gene is expressed in the developing nail bed mesenchyme (http://genome.ucsc.edu, 14/04/2015). Various variants in MSX1 have been associated with NSOFCs. In 182 nuclear families that included 83 case-parent triads from Southern Brazil, the 169bp allele of the CA repeat in MSX1 was significantly over-transmitted with NSCL/P, p=0.0005 (Souza et al., 2013). Direct sequence analyses of the gene in a Chilean cohort yielded two missense variants, p.Gly16Asp and p.Ala34Gly, with p.Gly16Asp variant predicted to disrupt a possible splice site

and thus may be involved in the aetiology of OFCs (Vieira et al., 2004). In 150 unrelated case-parent trios with NSCL/P from Chile, four haplotypes in MSX1 exhibited over-transmission from parents to probands, whereas no individual SNP showed association. Furthermore, the A allele of rs12532 presented a 2.08-fold rise in risk to NSCL/P when it was of paternal, but not maternal, origin, suggesting a possible epigenetic control through imprinting (Suazo et al., 2010b). Direct MSX1 sequence analyses of 33 CL/P and 19 OFCs with tooth agenesis cases from Columbus, Ohio, USA, yielded no functionally relevant coding variants, except a known variant, 101C>T (p.Ala34Gly), which was more common (p=0.0008) in individuals with both clefts and tooth agenesis, whereas another variant, named *6C>T, was associated (p=0.001) with CL/P only (Modesto et al., 2006). In 107 case-parent triads and 66 complete control triads of Dutch origin, homozygosity for 169bp allele of CA microsatellite marker interacted significantly with parental smoking (OR=4.9) to increase OFC risk (van den Boogaard et al., 2008). Cases-control analyses of 153 cases and 205 controls from Estonia showed that rs6446693 in MSX1 was significantly associated with NSCL/P, with SNPs in MTHFR and PVRL2 exhibiting nominal association. JAG2 SNP rs11624283 exhibited the strongest association with isolated CP, with epispastic interactions being observed between IRF6 and MSX1 as well as BCL3 and EDN1 (Jagomagi et al., 2010). In 124 Malay families, no significant association was observed between NSCL/P and CA microsatellite, though direct DNA sequencing yielded five variants: p.Ala34Gly, p.Gly110Gly, p.Pro147Gln, p.Met37Leu and p.Gly267Ala (Salahshourifar et al., 2011). For 118 simplex families from Nigeria, p.Ala34Gly variant was significantly associated (p=0.00002) with NSCL/P (Butali et al., 2011). A study of a Korean cohort of 126 NSCL/P cases suggests that some variants in

MSX1 and PAX9 are associated with tooth agenesis within and without the cleft area (Seo et al., 2013). Direct DNA sequencing in a Chinese family with autosomal dominant tooth agenesis and CL yielded a novel, heterozygous nonsense variant (p.Gln189X) in MSX1 (Liang et al., 2012).

Rare functional variants in MSX1 have been observed in individuals with autosomal dominant tooth agenesis with or without NSCL/P. These include p.Arg196Pro in individuals with autosomal dominant oligodontia of second premolars and third molars; p.Ser105X in individuals with autosomal dominant CLP and oligodontia of predominantly the second premolars and third molars; pSer202X in individuals with Witkop tooth-nail syndrome and autosomal dominant oligodontia of predominantly premolars and first and third molars; p.Met61Lys in individuals with autosomal dominant oligodontia of predominantly the second premolars and third molars; p.Gln187X in individuals with hypodontia and oligodontia as well as NSCL/P (De Muynck et al., 2004). Other variants reported for NSCL/P only include p.Glu78Val, p.Gly91Asp, p.Gly98Glu, p.Val114Gly, pGly116Glu, p.Pro147Gln, p.Arg151Ser,

p.Gly267Cys and p.Pro278Ser. Other aetiologic variants in MSX1 reported in connection to isolated tooth agenesis include p.Met61Lys, p.Gly22ArgfsX168, p.Arg151Ser, p.Arg176Trp, p.Ala194Val, p.Gln216GlnfsX125, p.Ala219Thre and p.Ala221Glu. Finally, variants reported in individuals with both oligodontia and NSCL/P includes p.Ser104X and whole gene deletion (Liang et al., 2012).

Studies in other populations support a role of MSX1 in cleft aetiology. In an Iranian cohort of 100 NSOFC cases and 100 controls, significant associations were observed between NSCL/P and CDH1 SNP rs16260 as well as MSX1 SNP rs12532 (Rafighdoost et al., 2013). In 142 Korean families with NSCL/P, significant association (p=0.028) was observed

between NSCL/P and rs3821949 in TDT analysis, the A allele of rs3821949 exhibiting a significant increased risk to NSCL/P under the additive model

(Kim et al., 2013). In a cohort of 602 NSCL/P cases and 605 controls from a Chinese Han population, significant association was observed between rs12532 and the risk of developing NSCL/P. Subsequent luciferase activity and in situ lip tissue assays revealed that rs12532 may interact with miR-3649 miRNA to regulate the expression of MSX1, with the mutant allele of this SNP altering this interaction, probably resulting in NSCL/P (Ma et al., 2014). However, a study in Han Chinese could not replicate the association between rs12532 and NSCL/P (Huang et al., 2011).

2.4.1.10 MAFB gene

Human MAFB encodes a transcription factor that is a basic leucine zipper. This intronless gene encodes a protein of 323 amino acids. It is crucially involved in the regulation of lineage-specific hematopoiesis, by acting as a transcriptional activator or repressor. It acts as a transcription repressor to ETS1-mediated transcription of erythroidspecific genes in myeloid cells. It is also needed for the differentiation of monocytes, macrophages, podocytes and pancreatic islet beta cells. It has also been implicated in renal tube survival and F4/80 macrophage maturation. MAFB also acts as transcriptional activator on the glucagon and insulin promoters. Working in concert with PAX6, MAFB weakly trans-activates the glucagon gene promoter via the G1 element. The ability of this protein to perform these functions is influenced by SUMO modifications. MAFB may also function as an oncogene or a tumor suppressor gene, roles which are influenced by the type of cell (https://genome.ucsc.edu, 12/06/2015).

Various replication studies after the initial GWAS study (Beaty et al., 2010), have reinforced MAFB's role in OFC aetiology. In a Han Chinese cohort of 344 NSOFC cases and 324 controls, AA genotype of rs6065259 (p=0.0027) as well as CC genotype of rs13041247 (p=0.0098; OR=0.50) in MAFB were significantly associated with NSOFCs (Mi et al., 2014). In 396 cases and 384 controls of Han Chinese origin, rs13041247 was significantly associated with NSOFC, with the C allele decreasing susceptibility to NSOFCs, in contrast to the wildtype T allele (Pan et al., 2011). In these two Han Chinese studies, subphenotypes analyses showed that MAFB was associated with both CL/P and CP. Another SNP linked to MAFB, rs17820943, showed evidence of association in a Sothern Han Chinese cohort of 300 NSOFC cases, 354 controls and an additional 168 case-parent trios (Cheng et al., 2012).

Studies in other populations give marginal support to a role of MAFB in cleft pathogenesis. In 895 Asian and 681 European trios, strong evidence of gene-gene (G×G) interaction or epistasis, was observed between markers in WNT5B and MAFB (pAsian=0.0076; pEuropean=0.018). Moreover, epistatic interaction was observed between SNPs in WNT5A, IRF6 and C1orf107 among Asians as well as for markers in 8q24 locus and WNT5B among Europeans (Li et al., 2014). In 182 multiplex and 464 simplex families of Hispanic and non-Hispanic ancestries, only rs481931 of ABCA4 exhibited evidence of association (p=0.00003) with NSCL/P in non-Hispanic dataset whereas both SNPs of ABCA4 showed evidence of association ($p<0.05$) in the Hispanic dataset. Borderline association ($p\leq0.05$) was observed for all MAFB SNPs in the Hispanic dataset but not the non-Hispanic dataset (Yuan et al., 2011). In 400 NSCL/P cases and 412 controls of Caucasian descent from Brazil, neither SNP in MAFB (rs13041247 nor rs11696257) showed evidence of association with NSCL/P. However, rs540426 of ABCA4

demonstrated strong evidence of association (p=0.00007) with NSCL/P, whereas rs481931 exhibited marginal association with NSCL/P (Fontoura et al., 2012).

2.4.1.11 Other genes implicated in the aetiology of nonsyndromic orofacial clefts

AXIN2 encodes a protein of 843 amino acids and down-regulates β-catenin in the Wnt signaling pathway (http://genome.ucsc.edu, 08/08/2015). Eleven SNPs tagging AXIN2 were genotyped in 682 families from Argentina, Guatemala, Madrid, Hungary, Turkey, Shanghai, Beijing and India. FBAT showed that rs7224837 was significantly associated (p=0.001) with NSCL/P in the combined data set. Also, rs3923086 was associated with CLP in Asians (p=0.004). Further genotyping and FBAT analyses of 528 NSCL/P families from USA also confirmed the association between AXIN2 and NSCL/P ($p<0.009$). Murine expression studies showed that AXIN2 mRNA and protein were expressed during murine palatogenesis. Axin2 and Irf6 proteins were also co-localized, particularly in the epithelium (Letra et al., 2012b).

Associations between NTN1 at 17p13 as well as NOG at 17q22 and NSOFCs have been replicated. In targeted sequencing of 1,409 trios from Europe, United States, China and Philippines, a genome-wide significant association was observed between rs9904526 ($pAsian=3.07\times10^{-9}$) and NSCL/P. This SNP tagged NTN1, as anti-NTN1 immunofluorescence studies on murine embryos showed high-level NTN1 activity in mesenchyme, particularly along the basement membrane of palatal shelves, and at highest levels along the presumptive medial edges and oral sides of palatal shelves.

Moreover, a NOG-tagging SNP, rs227727, reached genome-wide significance ($pAsian=7.3\times10^{-8}$); this SNP was in complete LD with rs227731. Gene expression studies

showed that Nog protein was primarily localized to palatal epithelium, including basal and periderm layers as well as mesenchyme. Reporter assays also showed that rs227727 is located in a window with enhancer activity (Leslie et al., 2015a).

In 946 pedigrees of Asian and European ancestries, rs8001641 of SPRY2 was significantly associated with NSCLP. Direct DNA sequence analyses yielded four novel variants in the conserved regions of SPRY2: chr13:80692698A>C, chr13:80692752G>T and chr13:80692951–80692955delAAATT in Filipino cases as well as chr13:80693009T>C in a Mongolian case. These variants altered binding sites for transcription factors POU1F1a and LEF-1 (Jia et al., 2015).

2.5 Syndromic orofacial clefts

About 300 syndromes are known to present with OFCs, according to online Mendelian inheritance in man (OMIM) (http://www.ncbi.nlm.nih.gov/omim, 04/07/2015). Van der Woude (VWS) is the most common syndromic cause of OFCs, increasing the recurrent risk of OFCs threefold (Zuchero et al., 2004). However, OFCs are also occasionally observed in other syndromes: Sticler's syndrome, Siderius Xlinked mental retardation (XLMR) syndrome, Leeys-Deitz syndrome, Hardikar syndrome, Patau syndrome, Malpeuch facial clefting syndrome, hearing loss with craniofacial syndromes, PPS, Treacher Collins syndrome, etc. (Dixon et al., 2011).

2.5.1 Van der Woude Syndrome

VWS (OMIM:119300) is an autosomal dominant disorder that mostly emanates from pathogenic mutations in IRF6. There are two cardinal clinical manifestations of VWS: lower lip pits or lumps and CL/P. VWS may also be expressed as lip pits only due to variance in expressivity. Moreover, it may also manifest as CL/P only and this could

segregate in families as dominant traits (Ghassibe et al., 2004). There are two types of VWS: VWS1 that results from mutations in IRF6 (Kondo et al., 2002) and VWS2 that results from mutation in WDR65 (Rorick et al., 2011).

Variants in IRF6 also cause popliteal pterygium syndrome, PPS (OMIM:119500), making VWS and PPS allelic syndromes. PPS is characterized by CL/P, pterygium, finger and toe syndactyly, oral synechiae and genital abnormalities (Ghassibe et al., 2004). Furthermore, ankyloglossia, hypodontia, hypernasal voice, bifid uvula, syngnathia, narrow high arched palate and other extra-oral phenotypes such as limb anomalies, popliteal webs, accessory nipples, congenital heart defects, Hirschsprung disease and fusion of primary mandibular left lateral incisor and canine, have been observed in PPS cases (Sarode et al., 2011). Thus, IRF6 variants have high penetrance but variable expressivity; stochastic factors or modifier genes may influence IRF6 phenotype (Kondo et al., 2002).

Direct DNA sequence analysis of the then delineated VWS locus in monozygotic twins revealed IRF6 as VWS aetiologic gene: a nonsense mutation in IRF6 was observed in the affected twin but not the unaffected twin or parents. Subsequent sequencing of IRF6 in VWS and PPS cases yielded pathogenic mutations in 45 unrelated VWS families and unique mutations in 13 PPS families. These variants showed non-random distribution among the exons of IRF6: 35 out of 37 missense variants occurred in either the DNA-binding or SMIR/IAD domains, presupposing these domains were critical for gene function. Gene expression studies showed that Irf6 was expressed in murine medial edge of fusing palate, tooth buds, hair follicles, genitalia and skin (Kondo et al., 2002).

VWS and PPS may be familial. Direct sequence analyses of IRF6 in six VWS and PPS families yielded six heterozygous missense mutations: Thre6Ile, Leu22Pro,

Ala16Val, Thre100Ala, Leu251Pro and Pro258Ser. In four generations of a family, some individuals had either VWS or PPS phenotypes, though a common variant was observed in all affected members (Ghassibe et al., 2004). In four generations of an Italian family, with mixed phenotypes of lip pits only, lip pits with bifid uvula and lip pits with CP, a nonsense variant, W217X, was observed in affected members (Gatta et al., 2004). In 17 families from Sweden, Finland, Norway, Thailand and Singapore, aetiologic variants were detected in all cases, except four VWS cases from Finland, probably due to founder effect. Ten variants were observed, six (p.Lys34Glu, p.Gln118fs, p.Ser407fs, p.Ser424X, p.Gln120X and p.Gln359X) of which were novel, with the last two of these six variants being de novo mutations (Peyrard-Janvid et al., 2005).

VWS and PPS phenotypes are variable. A new heterozygous mutation, Arg339Ile, has been observed in a patient with overlapping features of VWS and PPS: CLP, ankyloblepharon, paramedian lip pits, renal aplasia and coronal hypospadias (de Medeiros et al., 2008). In four out of five Honduran VWS families, aetiologic variants were found in both affected and unaffected individuals: p.N88I, p.K101QfsX15, p.Q208X and p.R412X (Birkeland et al., 2011).

IRF6 gene mutations also occur in Asian VWS families. Of four Chinese VWS families screened, four heterozygous mutations were observed: Y111H, S407fsX436 and F165fsX166 which were novel as well as R400W which is a known variant (Ye et al., 2005). Earlier studies in four Chinese families with VWS also showed four aetiologic variants, A2V, R6C, W379X and R400W (Wang et al., 2003). In 11 Chinese families, IRF6 mutations were observed in all affected individuals: p.Ala2Gly, p.Pro76Ser, p.Cys114X, p.Gln118X, p.Gly189fs, p.Arg250X, p.Phe252Leu, p.Ser330X, p.Tyr403X,

p.Arg412X and p.Ala439fs (Du et al., 2006). In 2 unrelated Japanese VWS families, two novel missense variants, R9W and R84C, were observed in both affected and unaffected individuals, presupposing incomplete penetrance and expressivity in these families (Matsuzawa et al., 2004). Using DNA extracted from finger nails, E349V missense variant was observed in all affected members of a Japanese VWS family (Matsuzawa et al., 2006). In two Japanese PPS families, two variants, R84L and S424L, were observed (Matsuzawa et al., 2010). R45W variant has been observed in a Singapore family with VWS (Tan et al., 2008). In simplex and multiplex VWS families from Taiwan, 11 variants were observed in cases only, five of which were novel: p.Tyr97Cys, p.Gln120HisfsX24, p.Glu136fsX3, p.Thr291Pro and p.Trp323X (Wu-Chou et al., 2013). No pathogenic variant was seen in an Iranian individual with lip pits but no cleft, presupposing locus heterogeneity for VWS (Baghestani et al., 2010). A 2.3Mb deletion at 1q32.2 locus was observed in a VWS patient; the deletion was absent in parents and siblings (Tan et al., 2013). In three Thai VWS families with either two lip pits or a heartshaped mass only on lower lip, novel variants, p.Q49X, p.F57L and p.L436V, were observed (Yeetong et al., 2009). In 7 Pakistani VWS families, some with unusual phenotype of lip bumps, six variants, including four novel ones (p.R7fsX52,

p.S212fsX12, p. M1T and p.209fsX15) were observed (Malik et al., 2014).

Aetiologic variants in IRF6 are non-randomly distributed among the exons. In 307 VWS and 37 PPS families, aetiologic variants were found in 68% of VWS and 97% of PPS families. About 106 novel variants were found, 80% of which occurred in exons 3, 4, 7 and 9. Most of the variants observed in PPS cases were missense whereas VWS variants were evenly distributed between missense and protein truncation variants.

Moreover, PPS were significantly localized to exon 4 at residues that are predicted to directly bind DNA (Ferreira de Lima et al., 2009). In 8 VWS and 1 PPS cases from Ethiopia and Nigeria, three novel variants occurred in VWS families: p.Lys66X, p.Pro126Pro (a splice site variant) and p.Phe230Leu. Other known variants were observed in VWS (p.Leu251Pro and an acceptor site variant, c.1061-2A>G) and PPS (p.Arg84His) cases (Butali et al., 2014b).

To date, about 300 different aetiologic variants in IRF6 gene have been found in PPS and VWS cases, including a common, single mutation causing both syndromes, suggesting a role of genetic modifiers. A search for such genetic modifiers yielded marginal associations for FOXE1, TGFβ3 and TFAP2α (Leslie et al., 2013). In direct sequence analyses of 70 VWS cases without IRF6 aetiologic exonic mutations, a rare variant, 350dupA, was observed in a conserved IRF6 enhancer element (MCS9.7). This variant eliminated the cis-binding motifs of p63 and E47 and significantly altered enhancer activity in human cell cultures as well as lacZ expression in mice. This variant also created a binding motif, CAAAGT, for Lef1, creating both a loss- and gain-offunction mechanistic effects that were aetiologic (Fakhouri et al., 2014).

Mutations in certain genes may also cause VWS and PPS. Aetiologic missense mutation (p.Asp523Tyr) was observed in WDR65 (a downstream target of IRF6) in a person with VWS2 in whose IRF6 gene no exonic mutations were observed (Rorick et al., 2011). Variants in IRF6 SMIR domain dysregulate RIPK4 function: both genes are crucial for the regulation of keratinocyte differentiation and variants in RIPK4 occur in individuals with Bartsocas-Papas syndrome (BPS). RIPK4 carry out this function by inducing IRF6 transactivator function through phosphorylation of its C-terminal domains at Ser413 and Ser424. Functional analyses of prevalent p.Arg412X IRF6 variant in VWS cases and

p.Ser376X RIPK4 variant in a BPS cases, showed that truncation of IRF6 at Arg412 led to rapid proteasome-dependent degradation, thereby abolishing the induction of IRF6 transactivator function by RIPK4. Similarly, the p.Ser376X variant also abolished RIPK4's ability to induce IRF6 transactivation and further impaired stabilization of β-catenin by RIPK4, possibly leading to impairment of Wnt signaling (Kwa et al., 2015). In 3 PPS and 3 BPS patients, a novel, segregating, homozygous missense variant, p.Leu439Pro, as well as homozygous, maternal uniparental isodisomyinherited missense variant, p.Arg618His, were found in RIPK4 in individuals with PPS. The last PPS patient also had homozygous RIPK4 missense variant p.Ala448Pro. Two other homozygous missense variants in RIPK4, p.Pro189Leu and p.Arg618His, occurred in BPS cases. A splice acceptor site variant in CHUK was seen in a BPS case (Leslie et al., 2015b). In 8 out of 45 VWS families that lacked pathogenic mutations in IRF6, pathogenic variants were observed in GRHL3: p.Arg298His, p.Arg391Cys, p.Arg520Gln, p.Asn554Ser, p.Tyr90Hisfs*4, p.Phe325Leufs*22, p.Glu522Leufs*10 and p.Val526Cysfs*7. Zebrafish-based assays demonstrated that these variants abrogated oral periderm development in cell-autonomous manner, which is in concordance with dominant negative effect. Grhl3 null mouse also showed abnormal oral periderm and 17% developed CP (Peyrard-Janvid et al., 2014).

2.6 The Role of environmental factors in OFC aetiology

Phenotype is an upshot of interactions between genetics and environment. This is buttressed by lack of 100% concordance of cleft occurrence even among monozygotic twins (Kondo et al., 2002). Therefore, environmental and teratogenic factors that influence gametogenesis and intra-uterine environment may affect the development of human

foetus. A teratogen is any chemical or substance that can cause foetal or congenital malformations. Environmental factors implicated in birth defects aetiology include maternal alcohol intake, hypoxia, hypertensive treatments, pesticide exposure, maternal diet and vitamin intake, retinoids, anticonvulsant drugs, nitrate compounds, organic solvents, parental exposure to lead and illegal drugs (cocaine, crack cocaine, heroin), traditional medicines as well as steroids and mercury in skin lightening creams (Strachan and Read, 2011). For example, mutations in PHF8 interact with hypoxia to cause clefts. The gene codes for a histone lysine demethylase that is crucial for epigenetic regulation. The catalytic activity of this enzyme depends on molecular oxygen and mice exposed to hypoxia have increased risk of developing cleft. F279S mutation in this gene inactivates a highly conserved hydrophobic residue, thereby reducing its oxygen uptake ability under hypoxic conditions (Loenarz et al., 2009).

Folate is crucial for epigenetic regulation and reprogramming during mammalian development. Epigenetic control mechanisms include DNA methylation, posttranslational histones modification, chromatin remodeling, microRNAs (miRNAs), and long noncoding RNAs (lncRNAs). DNA synthesis, repair and methylation as well as histone protein methylation, are some of the biochemical cellular processes that rely on folate as a 1-carbon source (Figure 2.7). Covalent DNA methylation usually occurs on cytosine residues in 5'-CpG-3' dinucleotide islands, which are normally associated with genes, particularly promoters and other regulatory elements. For transcriptionally active genes, promoters are hypo-methylated, gene body and repetitive elements are hypermethylated while epialleles are hypo-methylated. Excessive maternal folate intake and cancer can hyper-methylate epialleles and promoters, respectively, leading to repression of genes. DNA methylation stabilizes the genome by recognizing and inactivating parasitic viral

DNA sequences that infiltrate the genome. It is also crucial for transcriptional regulation, X-chromosome inactivation, parent of origin genomic imprinting and tissue-specific gene expression (Crider et al., 2012).

Methylation patterns are passed on during cell division; this is influenced by availability of folate. During cell division, DNA methyltransferase 1 (DNMT1) methylates all hemimethylated daughter cells, whereas de novo methylation is carried out by DNMT3a and DNMT3b that function primarily during mammalian development to establish DNA methylation patterns as epigenetic remodeling and reprogramming proceed during differentiation. In gametogenesis, methylation patterns are erased and reprogrammed to conform to that of an egg or sperm. Except imprinted genes, there occurs genome-wide erasure of DNA methylation during early embryogenesis (from 4cell stage up to blastocyst formation), followed by re-methylation in late morula; this open windows for environmental influences, such a folate deficiency, which may lead to birth defects (Cider et al. 2012).

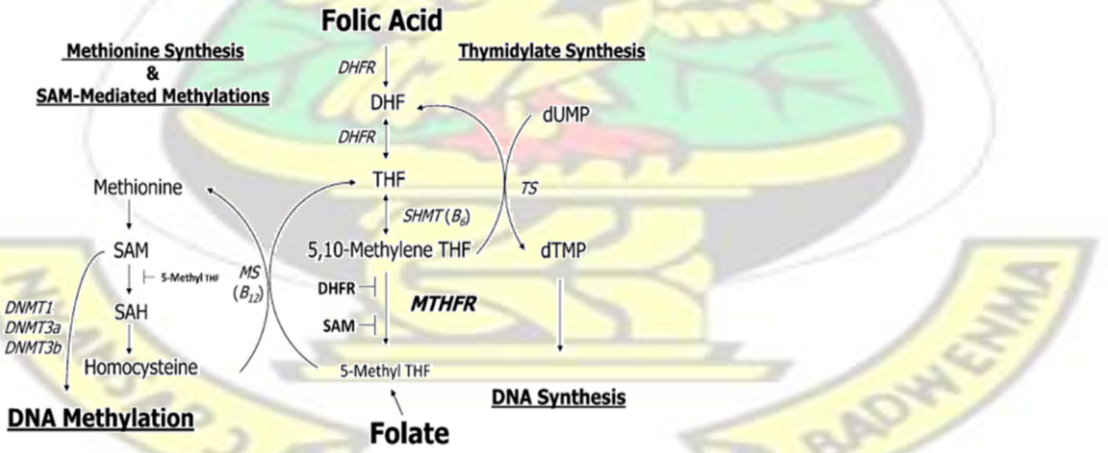


Figure 2.7: Role of folate in DNA-related metabolism. DHF, dihydrofolate; DHFR, dihydrofolate reductase; DNMT, DNA methyltransferase; dTMP, thymidylate; dUMP, deoxyuridine monophosphate; MS, methionine synthase; MTHFR, methylenetetrahydrofolate reductase; SAH, S-adenosylhomocysteine; SAM, Sadenosylmethionine; SHMT, serine hydroxymethyltransferase; THF, tetrahydrofolate; TS, thymidylate synthase (Crider et al., 2012).

CHAPTER THREE

3.0 SUBJECTS, MATERIALS AND METHODS

3.1 Subjects, ethical clearance and study design

Sampling of participants was carried out mainly at the Cleft Clinic of Komfo Anokye Teaching Hospital (KATH) in Kumasi. Through the outreach programmes of the Ghana Cleft Foundation, other participants were also recruited from Ho Regional Hospital, Sogakope District Hospital, Koforidua Regional Hospital and Bolgatanga Regional Hospital. Unrelated control subjects were recruited from the Child Health Directorate of KATH as well as Trede and Kentenkyire Basic Schools, all in Ashanti Region. Recruitment of subjects spanned the period from September 2012 to May 2015.

3.1.1 Ethical approval and informed consent

Since this is a study involving human subjects, ethical clearance was obtained from the Committee on Human Research, Publication and Ethics (CHRPE), School of Medical Sciences, Kwame Nkrumah University of Science and Technology, KNUST (Appendix A). The ethical approval number was CHRPE/AP/217/13 dated 3rd October 2013 and was renewed annually; prior to this, a conditional approval was obtained.

Written, informed consent (Appendix B) was obtained from each participating family before administration of questionnaire and sample collection. Subjects who were at least 18 years signed their own informed consent forms whereas for subjects below 18 years, the consent form was signed by a parent or guardian.

To preserve confidentiality, each participant was assigned an individual identification (ID) number. For example, GH20130836 represented an ID for the entire family, with each participant in the family being given a unique personal ID:

GH20130836_1 for proband, GH20130836_2 for father, GH20130836_3 for mother, etc. These ID numbers were used to identify all research information and biospecimens. Moreover, for easy and effective management of samples and research data, the Research Electronic Data Capture (REDCap) (https://www.icts.uiowa.edu/apps/redcap) was used to store de-identified data. The data was subsequently transferred to the PROGENY intra-net based software at the Jeffrey C. Murray Laboratory. PROGENY allows researchers to easily link sequence and genotype data to participant information.

3.1.2 Study design

The study design and methodology (Figure 3.1) included both case-parents trios (family-based studies) and case-control approaches for the SNP genotyping and direct DNA sequencing of selected genes, with subsequent functional validation of selected rare variants. In case-parent triads or trios, samples were collected from the proband as well as the father and mother. However, in instances where a parent was not available, samples were also collected from siblings and other second degree relatives such as grandparents. Case-parent trios design enhances the detection of transmission of alleles or traits from relatives to the patient or proband. In the case-control approach, samples were collected from probands with no family history of cleft. Comparison of data from these unrelated controls and cleft cases allows detection of alleles that are present at higher frequency in families with family history of clefts but are not present or are present at lower frequency in the background population. Case families were selected consecutively (based on their willingness to participate in the research), from cleft patients who were encountered at the KATH Cleft Clinic. Hospital records of patients gave additional information on phenotypes and environmental exposures for participating families. For some case

families, a visit was paid to communities in which they lived in order to collect additional samples, genetic and environmental information about them.

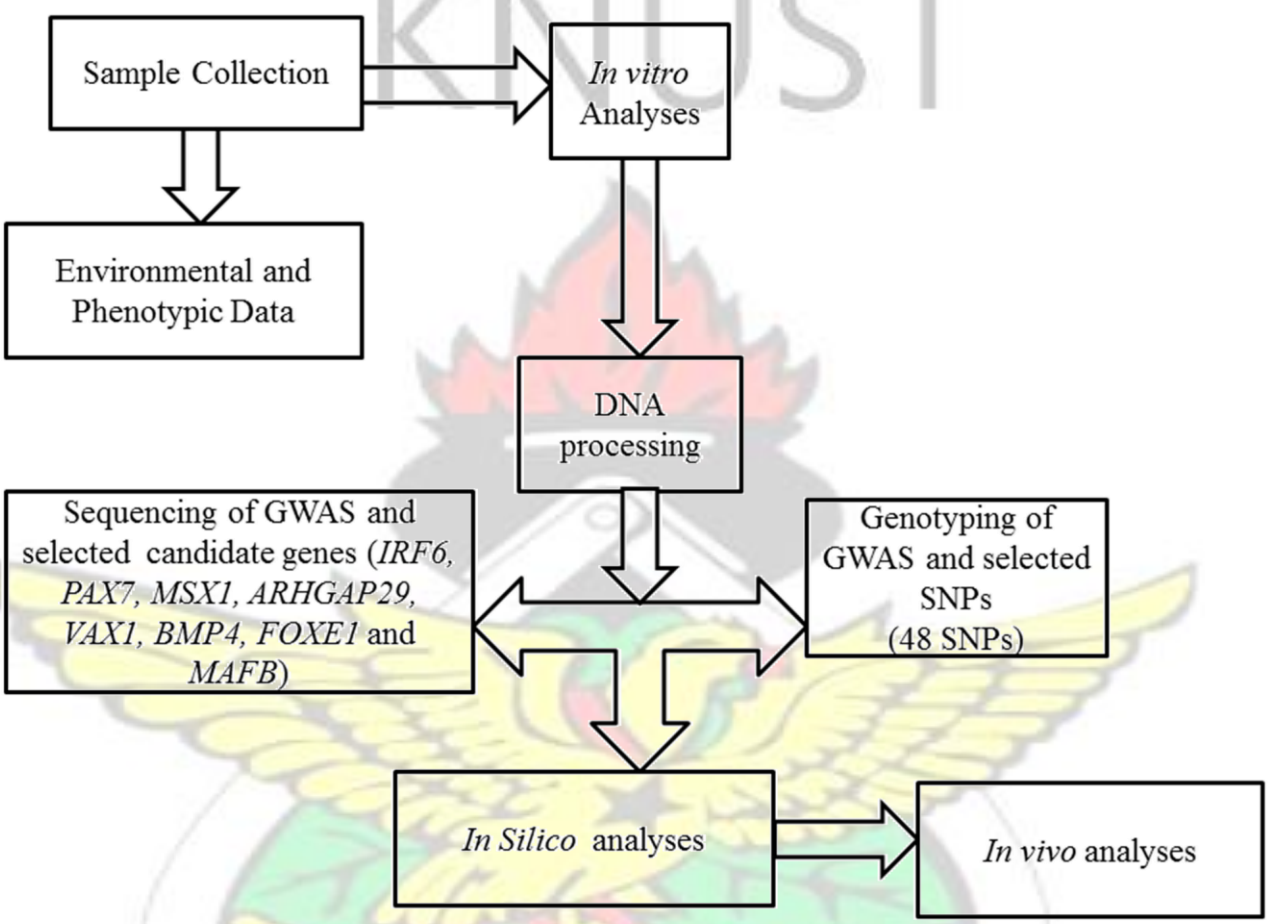


Figure 3.1: Study design and methodologies that were employed for the study

3.1.2.1 Inclusion criteria

For a subject to be classified as a case proband or patient, the subject must have any of or a combination of the following congenital cleft phenotypes:

(i) Cleft palate alone (CP): cleft of the hard and/or soft palate.

(ii) Cleft lip alone (CL): cleft of the lip alone, whether unilateral or bilateral.

(iii) Combined cleft lip and palate (CLP).
(iv) Facial or oblique clefts, also called Tessier clefts.

(v) Microform cleft: characterized by congenital scar from the lip to the nostrils.

(vi) Submucous cleft palate (SMCP): characterized by bifid uvula as well as translucent soft palate or soft palate with midline furrow.

(vii) Subclinical cleft phenotypes, such as ankyloglossia or tongue-tie.

Moreover, case families with just a member having cleft were classified as simplex families whereas those with more than one member with cleft were classified as multiplex families.

3.1.2.2 Exclusion criteria

A case proband was classified as having syndromic cleft, and was excluded from NSOFCs, whenever any other clinical, congenital, phenotypic malformations, such as club foot, presented with cleft. Also, proband was classified as having nonsyndromic cleft if no clinical, congenital, phenotypic abnormalities occurred together with cleft.

For families with no history of any of the cleft phenotypes stipulated in the inclusion criteria, this study considered them as unrelated control families. These control families also had no family history of other congenital anomalies, such as neural tube defects. Thus, families with clinically palpable congenital defects were excluded from the control families. The unrelated controls were ethnically matched to the cases.

3.1.3 Sample size determination and sample selection

Donkor et al (2007a) reported that, on average, 69 patients undergo cleft surgery each year at KATH. However, a review of the records of the Cleft Clinic at KATH by the researcher revealed that, on average, four new cleft cases are reported each week.

Thus, on average, 208 cleft patients report to the cleft clinic yearly (Medical Records of Cleft Clinic, KATH, 14/11/2012). Corollary to this, about 135 case families were recruited annually for a period of three years. In all, a minimum of 405 families with OFCs were targeted. The sample size was calculated using an online sample size calculator at http://www.surveysystem.com/sscalc.htm. The sample size was calculated at a confidence level of 95% and a confidence interval of 5. Over 1,000 participants with family history of clefts participated. Moreover, about 400 participants with no family history of clefts were also recruited as unrelated controls.

The sample size was justified for the case-control analyses using Epitools (http://epitools.ausvet.com.au/content.php?). The highest prevalence of OFCs reported in Africans is 0.54 per 1,000 live births (Mossey and Modell, 2012). Moreover, the average odds ratio (OR) for the genotyped SNPs was 3. Considering these observations as well as a confidence level of 0.95 and a power of 0.99, Epitools indicated that the average sample size per year for the OFC cases only should be 136 whereas the sample size for both control and cases should be 272. So for the three year period that sample collection was carried out, Epitools suggested a total sample size of 408 each for cases and controls, giving a total sample size of 816 participants. So the sample size used for this study has at least 99% power to detect associations at an assumed OR of 3 in the Ghanaian population.

For SNP genotyping, 48 SNPs (Table 3.1) were genotyped (Figure 3.5). These SNPs were selected based on GWAS and candidate gene studies. A total of 1,414 Ghanaian samples were genotyped at these markers. The Ghanaian samples came from 408 unrelated control probands, 413 NSOFC case probands and 593 relatives of case probands, primarily parents. Furthermore, 2,171 samples from Ethiopia and Nigeria that included NSOFC cases were also added to the genotyping cohort to increase the sample

size. In all, 3,585 African samples were genotyped: 872 NSOFC cases, 1635 unaffected relatives and 1078 unrelated controls. The 872 NSOFC cases comprised 163 NSCP cases, 340 NSCL cases, 361 NSCLP cases and 8 "un-typed" cases. "Un-typed" refers to samples from case probands that failed quality control checks which were carried out after genotyping and were therefore not included in the final statistical analyses.

3.2 Questionnaire administration

A questionnaire (Appendix C) was administered to case families. After identification of case families, any congenital abnormalities, apart from clefts, if they occurred, were noted. The generations of each case pedigree was carefully scrutinized for any congenital birth defects. The generations included first degree relatives (siblings and parents), second degree relatives (grandparents) as well as third degree relatives (such uncles, aunties and cousins).

Moreover, since phenotype is an upshot of the interplay between genetics and environment, the questionnaire also probed into environmental exposures of these families. Twenty-two environmental exposures were ascertained: consanguinity, enemas, herbal remedies, family planning, trauma, attempted abortion, skin lightening creams, alcohol, cigarette, anti-hypertensive drugs, anti-convulsant drugs, pesticide, nitrate compounds (fertilizers), mining communities, lead/lip stick, heroine/cocaine, hypoxia (smoky environment), dietary folate deficiency, lack of antenatal care, miscarriage, still birth and prematurity. Since the uterine environment of mothers greatly influences the development of foetus (Dixon et al., 2011), these exposures were ascertained primarily from the mother. For almost all these environmental factors, the mothers were asked a "yes or no" question as to whether they got exposed to any of them, particularly during

periconception – three months before and three months after conception. However, for dietary folate deficiency, mothers were asked to describe their daily dietary intake in detail.

Daily and weekly dietary folate intake was assessed based on the consumption of fruits (such as citrus fruits and fruit juices), slightly cooked or uncooked dark green leafy vegetables (such as lettuce), dried beans and legumes (peas), and other folate-fortified foods (such as cereals, flours, etc.). Folate is a water-soluble type of vitamin B that is not stored in fat tissues of the body and excess folate is secreted through urine. Because of this, poor dietary folate, for even few weeks, could drastically reduce blood folate levels (http://www.nlm.nih.gov/medlineplus/ency/article/000354.htm, 09/07/2015). Corollary to these observations, a mother that did not consume folate-rich foods daily was classified as having "dietary folate deficiency". Folate supplementation was also ascertained: daily consumption of folate-fortified foods and peri-conception intake of folic acid supplements were inquired from mothers.

Other environmental data collected from case families included socio-economic status, level of education and when mothers accessed antenatal care. Socio-economic status was assessed based on the occupation of both mother and father. No questionnaire was administered to control families on environmental factors. However, they were interviewed to ensure that no clinical birth defects occurred in the family. Samples were then taken from only the probands. Time of antenatal care attendance was also ascertained based on the time lapse after conception before mother visited hospital; this is mainly the time that folate supplementation was done.

3.3 Sample collection

Oragene DNA Saliva Collection Kits (Plate 3.1) (DNA Genotek, Ottawa, Canada, http://www.dnagenotek.com) were used to collect samples. Before sample collection, participants were asked to refrain from introducing any food item (such as gums, water or food) into the oral cavity for a period of at least thirty minutes. This practice resulted in highly concentrated DNA samples. For participants who could spit, saliva was collected into the vial of the DNA collection kit, followed by closing of the lid of the vial so that the preservative on the lid would be spilled into the saliva. This preservative can keep the sample for up to five years at room temperature. For children and other older people who could not spit, a cheek swab was used to collect samples. The swab was used to wipe the cheeks and also placed in the space between the cheeks and gum so as to collect enough cells and saliva. Caution was taken to avoid continuous

Plate 3.1: Oragene saliva and cheek swab collection kits. A: Saliva kit, with the actual saliva kit at the left and the screw cap at the right, B: Swab kits, with the cheek swabs or sponges at the left and the scissors for severing the sponges at the right.

contact between the swab and the teeth, as this could lead to uptake of a large number of oral microbiota. Each cheek swab kit for every participant had at least five swabs and all were used on a participant. A pair of scissors, provided with each cheek swab kit, was then used to excise the swabs from the handle into an Oragene DNA Collection Kit. After sample collection, the original lid of the DNA Collection kit was removed and the vial was covered tightly with a different screw cap provided along with the DNA kit.

Data from sample collection and questionnaire were then captured into REDCap.

Samples were then packaged and shipped through FedEx to the Jeff Murray Laboratory, University of Iowa, Iowa, where actual genetic analyses were carried out.

3.4 DNA extraction and quality controls

DNA was processed from saliva and cheek swab samples (Figure 3.2). DNA was extracted from saliva and cheek swabs employing the Oragene Saliva Processing Protocol and was subsequently quantified using Qubit Assay (http://www.invitrogen.com/site/us/en/home/brands/Product-Brand/Qubit.html). Finally, sex of each participant was verified using XY Genotyping (http://genetics.uiowa.edu/protocols.php) that employed Real Time Polymerase Chain Reaction (PCR); this was carried out to validate the sanctity of the DNA samples. For XY Genotyping, the Y-specific primer amplified the USP9Y gene on chromosome Y whereas the X-specific primer amplified the KDM6A gene on X chromosome.

3.4.1 DNA extraction

The Oragene Saliva Processing Protocol (Plate 3.2) was employed to extract DNA from saliva and cheek swabs. Saliva Processing Sheet (Appendix D) was created and filled by scanning the sample IDs and typing in the Site information into PROGENY. Cheek swabs

or sponge samples were accordingly noted in the comments column of the processing sheet. Twenty samples were processed at a time. The lids and sides of 15 mL conical tubes were labeled 1 to 20 and lined up accordingly in a rack. Saliva or sponge swab vials were then lined up in numerical order in front of the conical tubes, per the order on the Processing Sheet.

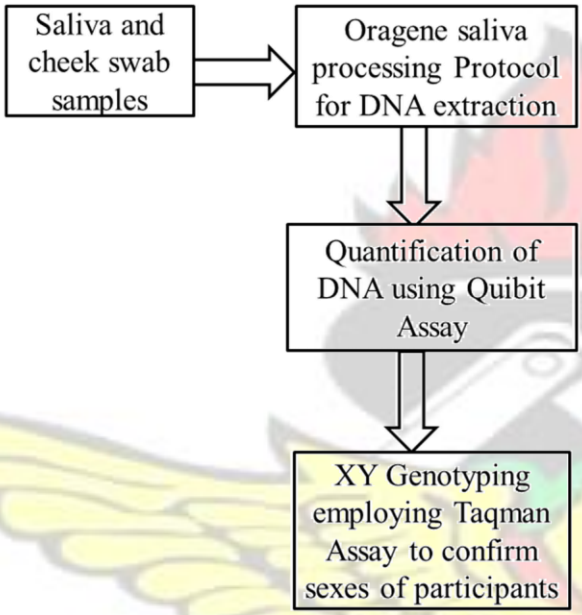


Figure 3.2: DNA extraction and quality controls flow chat

For cheek swab or sponge samples, an additional step was carried out. One mL of phosphate-buffered saline, PBS (QuickGene, Holliston, USA) was added to each vial containing sponges 1 to 7 days prior to processing. This was meant to rinse the vials and also dislodge any cell that had tightly adhered to the sponges. On the day of processing, all fluid in the saliva kit or vial was transferred into a labeled 15 mL conical tube, per the order on the processing sheet. The barrel of a 5 mL syringe was then placed into the same 15 mL conical tube and forceps were used to transfer the sponges into the syringe barrel. The IDs and number on the conical tubes were always verified at every step to make sure they corresponded. The 15 mL conical tube was then centrifuged at 1000 rpm for 10

minutes using AllegraTM 25R Centrifuge (Beckman Coulter Life Sciences, Indianapolis, USA). Fluids that were collected in the conical tubes were then processed along with the regular saliva samples.

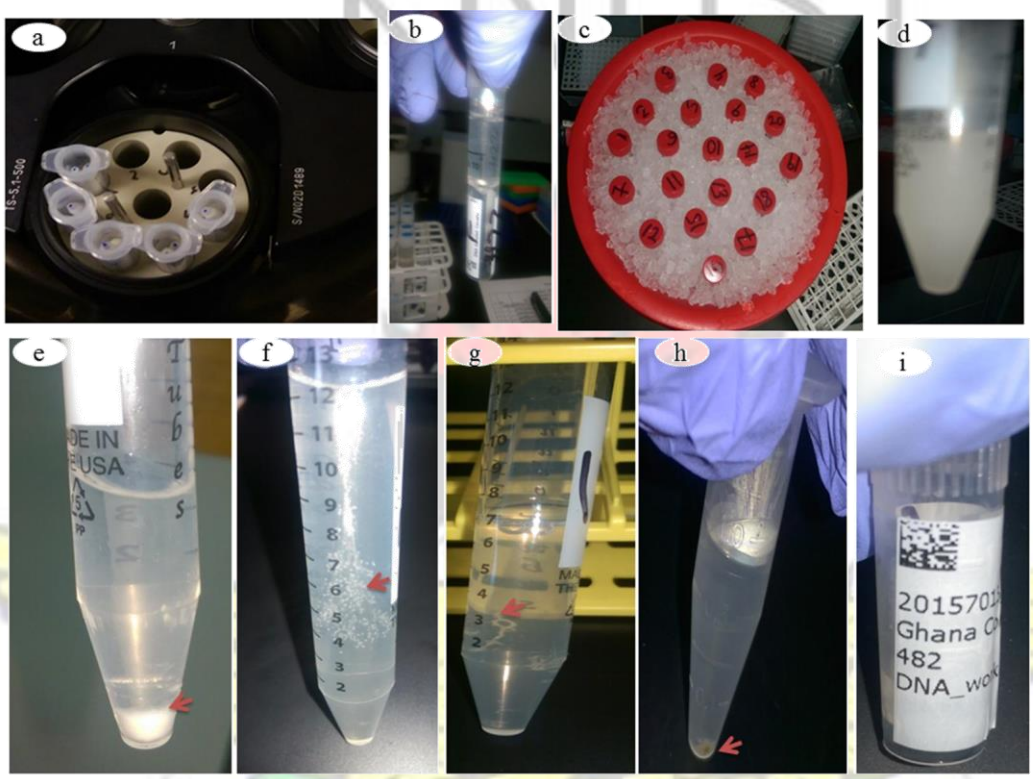


Plate 3.2: Steps in DNA extraction using the Oragene Saliva Processing Protocol. a: cheek swabs were transferred into a barrel of a syringe that had been placed in a 15 mL conical tube. b: saliva sample in Oragene DNA Collection Kit. c: incubation on ice after addition of Oragene purifier. d: sample becomes colourless after incubation on ice. e: Cellular debris (indicated by red arrowhead) forms pellet while the DNA remains in the supernatant after centrifugation. f: after transferring the supernatant to a different 15 mL conical tube, addition of 100% ethyl alcohol precipitates DNA (red arrowhead). g: DNA forms bands after inverting the conical tubes up and down. h: final spin in microcentrifuge to remove any cellular debris (indicated by arrowhead) from the DNA. i: samples were then transferred into screw caps, were quantified with Qubit Assay and labels were printed and pasted on tubes.

Samples were transferred into 15 mL conical tubes that were already labeled 120, verifying that the sample name from the vial corresponded to the number on the processing sheet. The volume of each sample was then noted on the processing sheet.

Samples were then incubated overnight in a 50 °C water bath (ThermoFisher Scientific, Hampton, USA). Oragene Purifier (prepIT®-L2P by DNA Genotek) was then pipetted into each 15 mL conical tube in a ratio of 40 uL of purifier to 1 mL of sample. The content of the conical tubes was then mixed by vortexing for a few seconds using Fisher Vortex, Genie 2 (ThermoFisher Scientific, Hampton, USA). The conical tubes were then incubated on ice for 10 minutes. This was followed by centrifugation of tubes at 3500 rpm for 10 minutes to separate cellular debris from DNA in the supernatant. Another set of conical tubes were then labeled 1 to 20. The clear supernatant was carefully transferred into a new 15 mL tube with the same number on it, for example, 1 to 1, 2 to 2 etc. The pellet was discarded.

Equal volume of 100% EtOH (QuickGene, Holliston, USA) as the supernatant was then added to each tube. Content was mixed by inverting tube 10 times; no vortexing or pipetting was carried out at this stage as that could potentially shear the DNA. Strands of DNA were observable at this stage. Samples were allowed to stand at room temperature for a minimum of 10 minutes. This was followed by centrifugation at 3500 rpm for 10 minutes. The supernatant was carefully transferred into a glass beaker, taking care not to disturb the DNA pellet at the bottom of the conical tube. The conical tubes were subsequently turned upside-down in a rack on a Teri wipe or paper towel for about 5 minutes so as to drain all ethanol from the DNA pellet. The DNA pellet was subsequently rehydrated with Elution Buffer (QuickGene, Holliston, USA): 1 mL for saliva samples and 750 uL for cheek swab samples. Vortexing or pipetting up and down was avoided

here because that could shear the DNA. Instead, the bottom of the tube was flicked gently about 10 times to loosen the DNA pellet. Samples were then allowed to stand at room temperature on bench overnight. If the pellet of DNA was not completely re-suspended after sitting overnight, tubes were flicked again, followed by placement of samples in a 35°C water bath for an hour.

The DNA was transferred into a 1.5 mL Eppendorf tubes (QuickGene, Holliston, USA) that had been labeled with the same number as the corresponding conical tube. Subsequently, the DNA was centrifuged at 14000 rpm for 15 minutes using Eppendorf Centrifuge 5417C (GMI Inc., Greater Minneapolis, USA). This final spinning removed any remaining turbid material or debris. DNA from each sample was carefully pipetted into two 2.0 mL screw cap freezer tubes (QuickGene, Holliston, USA) that had the same ID as the corresponding Eppendorf tubes: one as a working stock and the other for storage. Care was taken not to disturb the debris that had formed a pellet at the bottom of the Eppendorf tube. After DNA quantification by Qubit Assay, labels with the sample ID and concentration of DNA were printed and pasted on each screw cap freezer tube.

3.4.2 Qubit Assay

Amount of DNA obtained for each sample was quantified by employing the Qubit Assay (ThermoFisher Scientific, Hampton, USA). The Qubit Assay measured DNA concentration by measuring the level of fluorescence given off in the presence of a DNA-binding dye. Two sets of dyes were used: broad range (BR) molecular probes for all samples assayed, and high sensitive (HS) molecular probes, if the BR dyes could not detect any DNA. The BR molecular probes could detect DNA concentrations in the range of 100 pg/uL to 1000 ng/uL whereas the HS molecular probes detect DNA concentrations

between 10 pg/uL to 100 ng/uL. However, the same protocol works for both BR and HS molecular probes. The working solution was only stable for two hours, so care was taken not quantify too many samples at a go; large sample sets were divided into smaller groups.

Part 1- Preparing Samples: Two Qubit® Assay Tubes were labeled S1 and S2 for the standards provided with the kit; a tube each was also labeled for each sample to be assayed. Ten uL of standard #1 and standard #2 were pipetted into assay tubes labeled S1 and S2, respectively. Care was taken to use the correct BR standards. Two uL of each sample was then pipetted into the corresponding sample assay tube. A

Working Solution was subsequently prepared by diluting the Qubit™ reagent 1:200 in Qubit™ buffer (n uL reagent and 199 x n uL buffer, Figure 3.3). The solution was either prepared in a 15 mL or 50 mL conical tube, contingent upon the number of samples to be assayed. A volume of 190 uL of working solution was then added to each standards tube (S1, S2) whereas 198 uL of working solution was added to each sample tube. All tubes were vortexed for 2-3 seconds and were subsequently incubated at room temperature for two minutes.

Part 2-Reading Samples using the Qubit® Fluorometer: Upon turning on the Qubit Fluorometer, "DNA" was selected, followed by selection of "dsDNA broad range" (double stranded DNA) when the "Choose Your Assay" screen appeared. The two standards (S1 and S2) were read first; every time a new working solution was made, new standards were made and read. This was followed by reading of each sample at a time on the machine at unit of concentration of ng/uL. The concentration of each sample was recorded on the sample processing sheet (Appendix D). This information was scanned into PROGENY. Labels that bear the sample name and concentration were printed, and

were pasted on each sample accordingly. Care was also taken not to generate bubbles in the assay tubes during reading as this could give incorrect values;

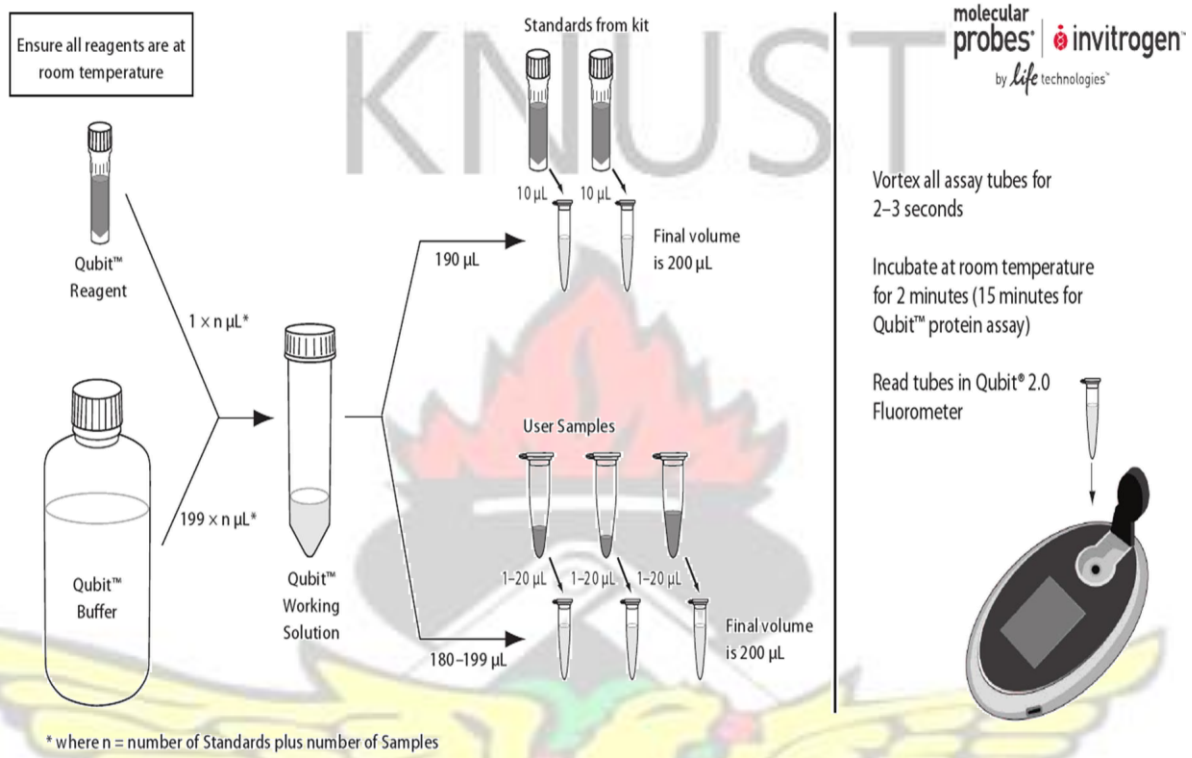


Figure 3.3: Quick reference guide that illustrates the making of working solution and its subsequent distribution into assay tubes. Note: n = 2 (standards) + number of samples + 2 (error). For example, for quantification of 32 samples, n was calculated as n = 32 + 2 standards + 2 =36). The extra 2 that was added was to compensate for pipetting errors.

tubes were tapped lightly to remove the pockets of air if they occurred. If during sample reading a message was received about the concentration of DNA being too low, the sample was re-vortexed and re-read. If the BR kit was in use and a message about concentration being too low was persistent, the HS kit was then used with that particular sample.

3.4.3: XY genotyping using real time PCR

This was a quality control measure that was carried out to ensure the sanctity of the samples by verifying the sex of each sample. A 384-well PCR plate (QuickGene,

Holliston, USA) was used, but wells were divided into quadrants so that only 96 samples could be genotyped at a time. One uL of each DNA sample at a concentration of 2 ng/uL was added to all 4 wells of a quadrant based on plate map; 2 wells for X-specific reaction and 2 wells for Y-specific reaction. Each sample was set up in duplicates for both primer sets because XX genotype is determined based on lack of curve for Yspecific primers (a failed assay could cause wrong genotyping XY > XX). The wells were set up as follows: A1, A2 = X primer reaction and B1, B2 = Y primer reaction. After the addition of the DNA, the PCR plate was spun down in centrifuge at 1500 rpm for 1 minute. The plate was then dried in a drying oven (GMI Inc., Greater Minneapolis, USA) for 10-15 minutes at a temperature of 75^{o}C.

Two separate Master Mix (MM), each for X or Y primer sets were prepared as follows: Syber Green = 300 uL, UTX – Forward primer (10 uM) = 7 uL, UTX – Reverse primer (10 uM) = 7 uL and dH_20 = 291 uL. For Y-specific MM, DFFRY forward and DFFRY reverse primers were used instead. All reagents (QuickGene Invitrogn, Holliston, USA) were kept on ice at all times. Below are the primer sequences that were designed with Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi).

DFFRY-F GAGCCCATCTTTGTCAGTTTAC
DFFRY-R CTGCCAATTTTCCACATCAACC } Y specific primers for USP9Y gene

UTX-F TAACAGGAAGTGGAAGTAATGG
UTX-R GGAGTTAGATAGTTGGTTTTTCC } X specific primers for KDM6A gene

Three uL of each MM was added to each well: X MM into top two wells of a quadrant and Y MM to the bottom two wells of a quadrant. After the MM was added to the dried plate, the plate was spun down for 1 minute at 1500 rpm before it was put on the 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, USA) to run. The 7900HT

system was run for a maximum of 2½ hours based on the Sequence Detection System (SDS) software on the attached computer under the following conditions for 40 cycles: hot-start – 95^{o}C for 10 minutes, denaturing – 95^{o}C for 15 seconds, annealing and extension – 60^{o}C for 1 minute and hold – 50^{o}C for 2 minutes. All wells that did not have samples were omitted in instances where the PCR plates were not full; the changes were then saved. It was ensured that "Method Running" together with the time remaining was listed before the 7900HT was left to run.

After plate run, file generated was saved into a lab folder for onward analyses at a different computer. The sex of each individual sample was checked by highlighting one quadrant at a time. For males, both the X and Y curves appear together on the plot whereas for females, the Y primer does not work at all or may come up very late on the curve (Figure 3.4).

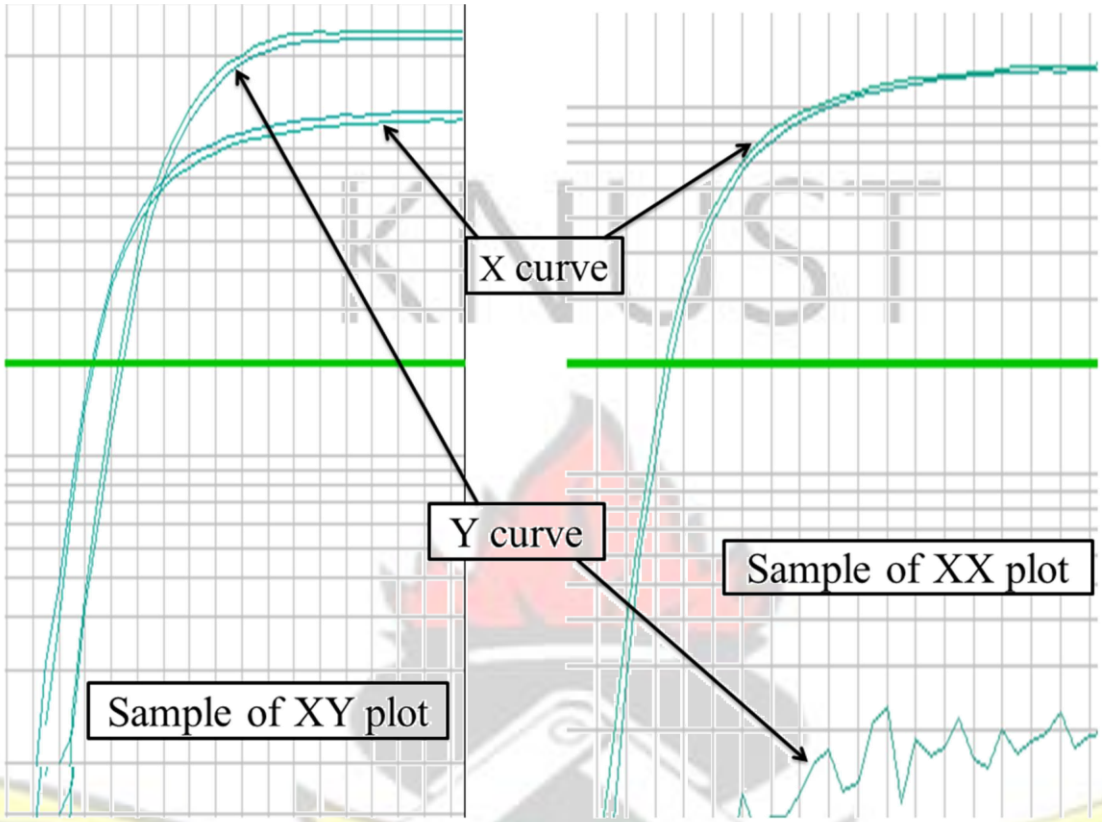


Figure 3.4: An XY Genotyping plots for males and females

3.5 SNP genotyping

Forty-eight SNPs were genotyped (Figure 3.5; Table 3.1) on 3,585 African samples. Each of the 3,585 DNA samples was randomly assigned to a well in a labeled 96-well microplate to form a plate map (Appendix E). Samples on microplates were diluted to a concentration of 2 ng/uL. Each of these also contained two template controls, NA18856 (male) and NA18855 (female), which are HapMap samples (http://browser.1000genomes.org/index.html). They therefore served as a guide in calling the genotype of individuals in this study. Each microplate also had provision for at least two No Template Controls (NTCs), which was deionized water (dH_2O);

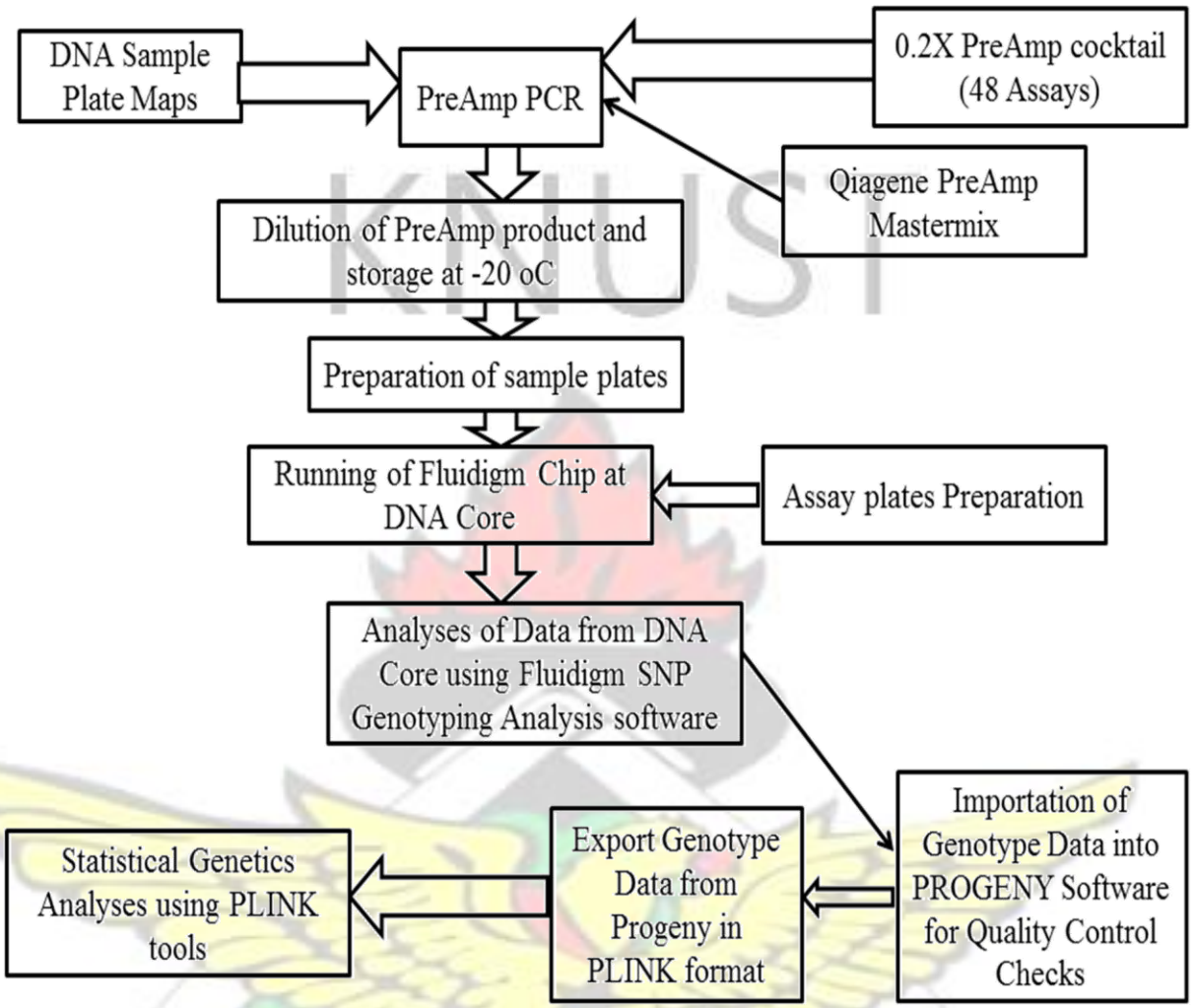


Figure 3.5: Flow chat for SNP genotyping, quality control checks and data analyses

Table 3.1: List of 48 SNPs that were genotyped

Chr.	coordinate	SNP	Probable gene/loci	Alleles/ Variation	NA18856 Genotype in 1000 Genomes	NA18855 Genotype in 1000 Genomes	Averge Call Rate (%)	Reference study	Study population
1	11854476	rs1801131	MTHFR	T>G	T/T	T/T	99.4	Boyles et al., 2008	Europeans
1	11856378	rs1801133	MTHFR	G>A	G/A	G/G	99.7		
1	18956458	rs766325	PAX7	A>G	A/A	A/A	99.6	Beaty et al., 2010	Europeans and Asians
1	18979874	rs742071	PAX7	G>T	T/G	G/G	99.4		
1	94553438	rs560426	ABCA4	T>C	C/T	T/C	99.5		
1	94570016	rs481931	ABCA4	G>T	G/G	No data	99.5		
1	94575056	rs4147811	ABCA4	C>T	C/C	No data	99.3		
1	94650805	rs138751793	ARHGAP29	T>C	T/T	No data	99.3	Present Study	Africans
1	108699730	rs6677101	SLC25A24	T>G	G/T	G/G	99.1	Butler et al., 2015	Europeans
1	209977111	rs861020	IRF6	G>A	G/G	G/G	99.5	Rojas-Martinez et al., 2010	Europeans
1	209979529	rs34743335	IRF6	A>T	A/A	No data	97.9	Present Study	Africans
1	209989270	rs642961	IRF6	G>A	G/G	G/G	99.5	Rahimov et al., 2008	Europeans and Asians
1	210174417	rs11119388	SYT14	A>G	A/A	A/A	99.6	Leslie et al., 2014	Europeans and Asians
2	43540125	rs7590268	THADA	T>G	T/T	T/T	99.5	Mangold et al., 2010	Europeans
2	71674476	rs4332945	DYSF	T>G	T/T	T/T	99.3	Brayton et al., 2009	Mouse screen
2	71780215	rs2303596	DYSF	C>T	C/T	C/C	99.2		
2	71866842	rs227782	DYSF	A>G	A/G	A/G	99.3		
4	4865146	rs12532	MSX1	A>G	G/A	A/G	99.4	Suzuki et al. 2004	Asians
4	4864991	rs115200552	MSX1	G>C	G/G	No data	99.5	Present study	Africans
6	93506409	rs2674394	Gene Desert	A>C	C/C	C/C	99.4	Ludwig et al. 2012	Europeans and Asians
6	158885758	rs651333	TULP4	C>T	C/T	T/T	99.3	Ludwig et al., 2012	Europeans and Asians
8	27389542	rs6558002	EPHX2	C>T	C/C	C/C	99.1		
8	129946154	rs987525	8q24	A>C	A/A	C/C	99.4	Birnbaum et al., 2009	Europeans

Table 3.1 continued…

Chr.	coordinate	SNP	Probable gene/loci	Alleles/ Variation	NA18856 Genotype in 1000 Genomes	NA18855 Genotype in 1000 Genomes	Averge Call Rate (%)	Reference study	Study population
9	100612270	rs894673	FOXE1	A>T	T/T	T/T	99.3	Moreno et al., 2009	Europeans, Asians and Hispanics
9	100614140	rs3758249	FOXE1	T>C	C/C	C/C	99.4		
10	118827560	rs7078160	VAX1	G>A	A/G	A/G	99.2	Beaty et al., 2010	Europeans and Asians
10	118834991	rs4752028	VAX1	T>C	C/C	C/T	99.6		
12	43819298	rs10785430	ADAMTS20	A>G	A/A	A/A	99.5	Wolf et al., 2015	Hispanics
13	80668874	rs9574565	SPRY2	T>C	C/T	T/T	99.4	Ludwig et al., 2012	Europeans and Asians
13	80692811	rs8001641	SPRY2	G>A	A/G	G/G	99.5		
13	92003297	rs375489721	MIR17HG	T>C	T/T	No data	99.1	Amendt et al., unplished	Mouse screen
13	92003356	rs185831554	MIR17HG	T>G	T/T	No data	99.3		
14	54417522	rs17563	BMP4	A>G	A/A	G/A	99.5	Chen et al., 2008	Asians
15	33050423	rs1258763	GREM1	C>T	C/C	C/C	99.3	Ludwig et al., 2012	Europeans and Asians
16	3980445	rs8049367	ADCY9	C>T	C/T	C/T	99.4	Sun et al., 2015	Asians
16	68771034	rs16260	CDH1	C>A	C/A	A/C	99.2	Song et al., 2011	Asians
16	68790394	rs11642413	CDH1	G>A	G/G	A/G	99.5		
16	84872051	rs1546124	CRISPLD2	C>G	C/C	C/G	99.5	Chiquet et al., 2007	Hispanics
16	84941329	rs4783099	CRISPLD2	C>T	C/T	C/C	99.3		
17	8956285	rs8069536	NTN1	G>T	G/G	G/T	99.4	Beaty et al., 2010	Europeans and Asians
17	8965551	rs8081823	NTN1	G>A	A/G	G/G	99.4		
17	54615617	rs17760296	NOG	T>G	T/T	T/T	99.5	Mangold et al., 2010	Europeans
17	54773238	rs227731	NOG	G>T	T/G	T/G	99.3		
17	63528123	rs7224837	AXIN2	A>G	A/G	A/A	99.3	Letra et al., 2012	Europeans and Asians
17	63549488	rs3923086	AXIN2	A>C	A/A	A/A	99.7		
20	39268516	rs17820943	MAFB	C>T	T/C	C/C	99.6	Beaty et al., 2010	

20	39269074	rs13041247	MAFB	T>C	C/T	T/T	99.5		Europeans
20	39270816	rs11696257	MAFB	C>T	T/C	C/C	99.3		and Asians

however, NTCs were not added unto microplates until the preparation of the sample plates. In all, sixteen and twenty-two 96-well microplates were generated for Ghanaian and Ethiopian-Nigerian cohorts, respectively, and were genotyped by the Fluidigm SNP Genotyping Protocol (http://genetics.uiowa.edu/protocols.php). The 192.24 Fluidigm Chip was used: this chip genotyped 24 SNPs at a time on 192 samples. Therefore, two microplates (for example, Plates 1 and 2) were combined to form a chip map (Appendix F). Each chip map was genotyped twice, a run for each of the two sets of 24 SNPs. Thus, the 48 SNPs genotyped were divided into two sets: Assay Set 1 and Assay Set 2, each comprising 24 SNPs and were called assay plates (Appendix G). Each assay plate indicated the position of each SNP assay, the allele (X) attached to VIC dye as well as the allele (Y) attached to FAM dye. SNPs were designed based on human genome assembly GRCh37/hg19, 2009 (http://genome.ucsc.edu). A total of 38 Fluidigm chips were run.

3.5.1 Fluidigm SNP genotyping protocol

Many reagents were used for the SNP genotyping (Table 3.2).

Table 3.2: Reagents for fluidigm protocol

Reagents	Supplier	Product Number
40X Assays	ABi/Life Technologies	rs number of SNP
Low TE buffer	Quickgene	CBD-02
2X Assay Loading Reagent	Fluidigm Kit/DNA Core	100-3459
Fast GT Loading Reagent *192.24 only*	Fluidigm Kit/DNA Core	100-3459
Pressure Fluid	Fluidigm Kit/DNA Core	100-3459
Control Line Fluid Syringes	Fluidigm Kit/DNA Core	100-3459
ROX	Enzyme Core/Invitrogen	12223-012
GTXpress MasterMix *192.24 only*	ABi/Life Technologies	4401892
PreAmp MasterMix	Qiagen	206145
192.24 Chip	Fluidigm	BMK-M-192.24GT

Pre-Amplification of DNA: The Pre-Amp was carried out to amplify, simultaneously, all the 48 regions of the human genome that harboured the selected SNPs (Table 3.1). The 0.2X PreAmp Cocktail was created by combining 40X assay with low-TE buffer (Table 3.3); these numbers are given per chip and were scaled up appropriately when the need arose. The PreAmp cocktail was combined with the PreAmp master mix to make the sample pre-mix (Table 3.4). The sample pre-mix was added to the DNA in each well of 96-well PCR plates; 1.3 uL of DNA was initially transferred from each 96-well microplate (plate map) to a 96-well PCR plate. This was followed by the running of the PCR amplification program using the following parameters: Hold at 95^{o}C for 10 minutes and 14 cycles - 95^{o}C at 15 seconds for denaturing and 60^{o}C at 4 minutes for annealing.

Table 3.3: 0.2X PreAmp cocktail preparation

Chip Type 192.24	Reagent	Volume (uL)
	40X assay	1.5 of each (48 total)
	Low-TE Buffer	228
	Total	300

Table 3.4: Sample pre-mix preparation

Chip Type 192.24	Reagent	Volume (uL)	Volume/well
	PreAmp MasterMix	580	
	0.2X PreAmp Cocktail	290	
	Total	870	4
	DNA (@ 2 ng/uL)		1.3

The Fluidigm SNP Genotyping Software relied on negative controls (NTCs) to make all other calls; moreover, correct positive control (NA18856 and NA18855) calls were also needed to trust the rest of the calls. To be able to test the quality of the results, each chip

map (that consisted of two 96-well PCR plates) contained at least three positive and three negative controls. The positive controls were amplified in the PreAmp PCR, but the negative controls were not amplified but were added later during the preparation of the sample plates; amplification of NTCs in PreAmp PCR affects the quality of the genotype calls made. After the PreAmp PCR, each well was diluted to a 1:5 ratio by the addition of 20 uL of low-TE buffer to a final volume of 25 uL. Plates were either used right away or stored at -20 °C for later use.

Preparation of assays: The two assay plates were prepared and used the same day or were stored for 3-4 days in advance of running chip(s) and frozen at -20°C. Various reagents were combined to form the assay master mix (Table 3.5) for one chip and were scaled up when the need arose. The first 24 SNP assays were transferred into the first three columns of a 96-well PCR plate, an assay per well (Assay Set 1); the other 24 SNP assays were also transferred to the last three columns of the same 96-well PCR plate, one assay per well (Assay Set 2) – Appendix G. The assay mastermix was then added to each well.

Table 3.5: Assay setup

Chip Type 192.24	Reagent	Volume (uL)	Volume/well
	2X Assay Loading Reagent	64	
	ROX	6.4	
	dH_2O	25.6	
	Total Assay Mastermix	96	3
	40X assay		1

Preparation of samples: The sample plates were created the morning of the chip run. Various reagents were combined to form the sample master mix (Table 3.6) for one chip and were scaled up when the need arose. Samples were transferred from two 96well PreAmp PCR plates (this forms one 192.24 chip) into new, correspondingly labeled PCR

plates - one sample per well. Care was taken to ensure that each sample was always placed in the appropriately, assigned well and plate. The sample mastermix was then added to each well. At this point, the negative controls (NTCs), which were dH_2O and of the same volume as that used for DNA, were added to the sample plates.

Table 3.6: Sample setup

Chip Type 192.24	Reagent	Volume (uL)	Volume/well
	GTXpress MasterMix (2X)	480	
	Fast GT Loading Reagent (20X)	48	
	dH_2O	48	
	Total Sample Mastermix	576	2.5
	PreAmped DNA		1.6

Fluidigm run: the assay and sample plates were vortexed and centrifuged at 1500 rpm for 1 minute at the Murray Lab. The chip run was carried out at the DNA Core, Institute of Human Genetics, University of Iowa. At the DNA Core, the sample and assay plates were re-vortexed and re-centrifuged at 3000 rpm for 1 minute by selecting bucket "6449" of the centrifuge. Next, the chip was unwrapped and oriented so that A1 was in the upper left-hand corner (Figure 3.6). The chip number, the plate numbers of the two sample plates to be used on the chip as well as the assay set to be used for the chip were all recorded in a book.

Subsequently, the control line fluid was loaded into the chip. The 192.24 chip has only one syringe with 160 uL of control fluid which was dispensed in the accumulator at the top of the chip (Acc2, Figure 3.6). The successful loading of this fluid was critical to the success of the chip run: any spillage of fluid on any part of the chip or inlets rendered the chip unusable. The syringe was removed from its packaging. While the syringe cap was

on, the syringe was used to push down the check valve gently to ensure that the valve moved freely up and down. While the syringe was facing up and away from the chip, the cap was removed. The chip was held at a 45° angle, the syringe tip was then gently inserted into the check valve opening. Care was taken to ensure that the valve was open and that the black O-ring was pushed down and moved to the side. Subsequently, the plunger of the syringe was pushed slowly to release the fluid into the accumulator. It was ensured that the O-ring returned to its normal position after the syringe was removed. The blue backing on the chip was subsequently removed.

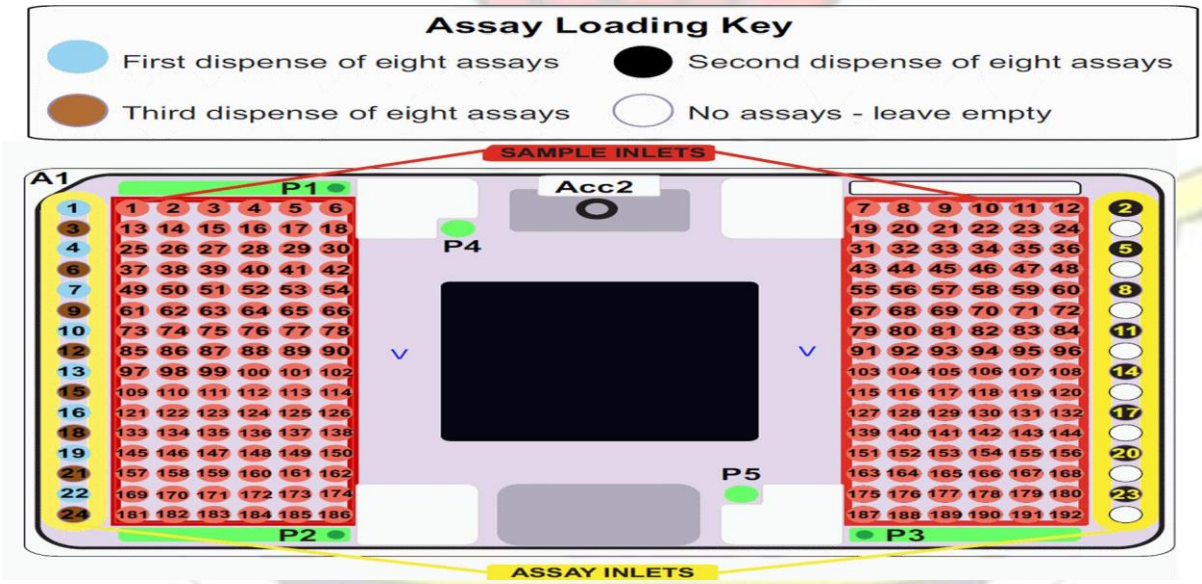


Figure 3.6: Fluidigm chip as well as sample and assay loading guide

Loading samples and assays: Using a multichannel pipette, 3 uL of sample was transferred from each sample plate into the sample inlets on the chip. Plate 1, for example, was pipetted into the top row of the sample inlets (# 1-12) and every other row down to the second to bottom (# 169-180, Figure 3.6). Plate 2 was pipetted into the second top row of sample inlets (# 13-24) and every other row down to the bottom row (# 181-192,

Figure 3.6). While pipetting, care was taken not to go past the first stop on the pipette, otherwise bubbles could be introduced into the inlets, which could prevent the samples from reaching the reaction chambers. Basically, in most cases, an inlet with a bubble does not yield any genotype for that sample or assay. After this, 3 uL of each assay was transferred from the assay plate to the assay inlets on the chip (Figure 3.6). A final check was then carried out on the chip once all assays and samples were loaded to ensure that no bubbles were visible in any of the inlets. Whenever bubbles were observed, needles and propanol (in a tube that creates surface tension) were used to dislodge and/or pop any bubbles.

A total of 150 uL of pressure fluid was pipetted into each of P1, P2 and P3 reservoirs. This was followed by pipetting of 20 uL of pressure fluid into each of P4 and P5 reservoirs (Figure 3.6). The chip was subsequently placed into the IFC Controller RX, making sure the notched corner aligned with A1. This was followed by running the 'LoadMix (166x)' script on the machine for 30 minutes. When the LoadMix script was complete, the chip was ejected from the controller and was transfered to the FC1 Cycler, making sure it was in proper alignment. FC1 Cycler was run for 30 minutes by selecting specific programmes: genotyping tab ('GT') and 'GT 192X24 Fast v1.pcl' programme. The chip was then transferred to the Chip Reader, which generated files that were read as VIC/FAM, but not FAM/VIC.

Preparation of files for reading results: this involved using Fluidigm SNP Genotyping Software to call out the actual genotype of each sample from the files that were generated by the Chip Reader. The chip map (Appendix F) was foremostly uploaded onto the SNP Genotyping software and NTCs (water samples) were set up.

Next, the assay and allele maps were set up by copying them from the assay plate map (Appendix G) and pasting them on the SNP Genotyping software. The VIC dye was attached to the X alleles whereas the FAM dye was attached to the Y alleles. Next, 'Detail Views', 'Analyze', 'Results Table', 'Image View', 'ROX' and '1' commands on the SNP Genotyping software were followed to verify the success of the chip run. The "image viewer" gave a good picture of how successful the chip run was; a good run looked pretty even in spacing and brightness (Figure 3.7a). If any bubbles got into any of the inlets, they usually affected genotype calls of a sample and the whole column showed up odd and glassy (Figure 3.7b); 8 of such inlets are indicated by arrows. If something went very wrong with the run, the image may not appear at all or will be extremely varied in brightness and evenness (Figure 3.7c) and such genotype data must be discarded and the chip must be re-run; however, nothing like this occurred for all the chip ran.

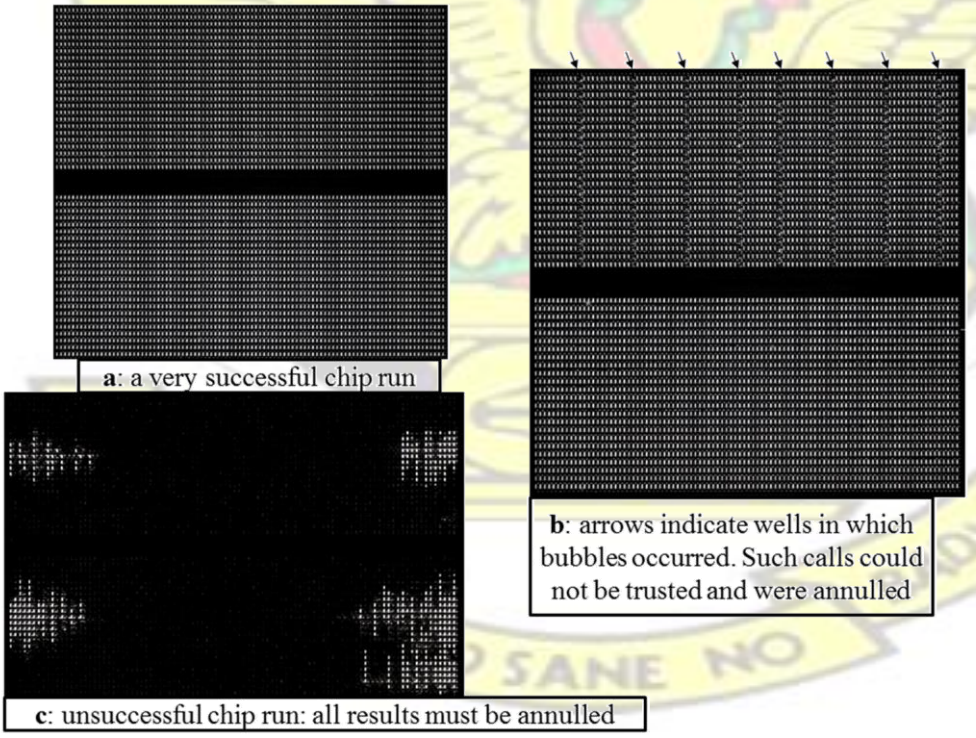


Figure 3.7: Image view for various chip runs

When a chip's image view looked like Figure 3.7a, it meant everything went on well; colours for alleles were then set, followed by the checking of positive controls and making of manual calls. If any of the wells looked like those shown in Figure 3.7b, those calls were invalidated as follows (Figure 3.8). First, in the image view, each red box that corresponded to a bubble (Figure 3.7b) was selected by holding down Ctrl and clicking each box individually; this turned each red box to green. With the boxes selected, the Call Map Viewer was chosen; all boxes that were selected appeared highlighted as calls for a sample and were invalidated by clicking the "Invalid" button.

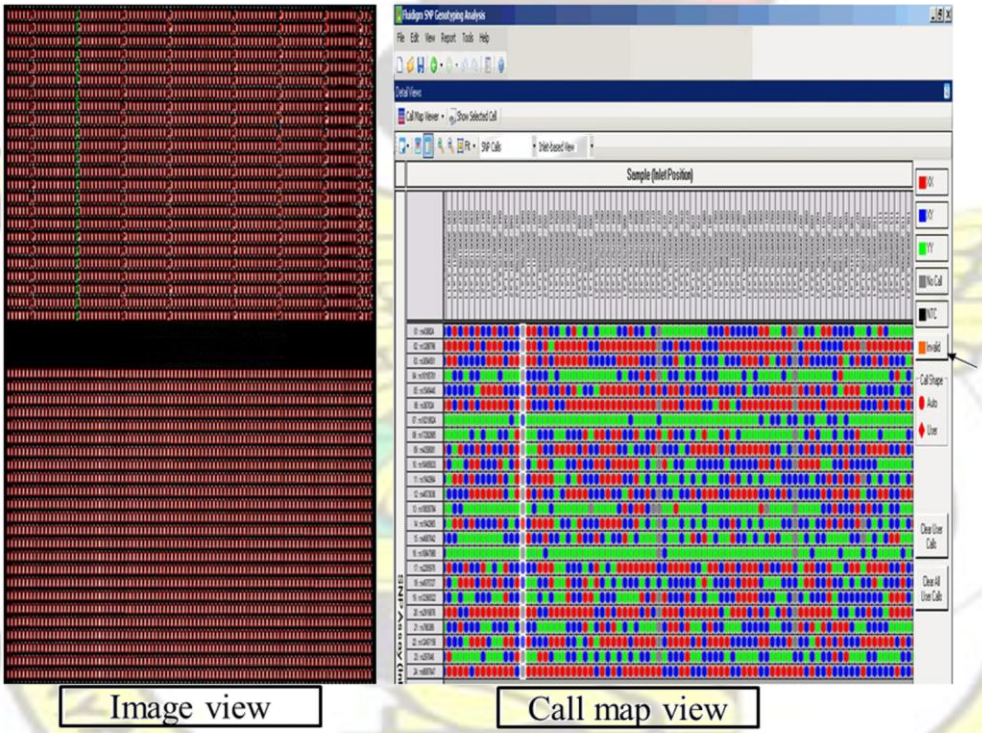


Figure 3.8: Invalidation of calls from samples that had bubbles

Colours for VIC (X) and FAM (Y) dyes were subsequently set. To set the colours of the calls, 'Assay Setup' and 'Change' buttons under 'Allele Settings' were selected. If they were entered correctly, VIC corresponded with Allele Symbol 'X' and scatter plot axis

'X-axis' and FAM matched up with 'Y' and 'Y-axis'. To match to TaqMan genotype calls, XX was assigned to red, XY to green and YY to blue.

Reading results and exporting the data: By clicking "Detail Views" tab, the scatter plot for each marker was displayed. These scatter plots were opened one after the other and manual calls were made. During genotype calling, it was checked to ensure that positive controls, NA18856 and NA18855, were reading as the right genotypes as reported in 1000 Genomes database (Table 3.1). Also, negative controls were monitored. NTCs should normally orient close to the origin of the scatter plot, usually around (0.1, 0.1). If any negative control amplified and appeared elsewhere apart from quadrat (1, 1), it was invalidated, as failure to do so affected other calls on the plot. Once genotype calling was done, the raw data was exported in excel format as well as "analysis report" file. The Analysis Report, a PDF file, was sent to senior investigators and other laboratory technicians to also critique the calls made.

Importing the raw data into PROGENY for quality control checks: Once genotyping was completed (all sample plates and all markers done, all files read and exported), the raw data, in excel format, was imported into PROGENY. It was ensured that all markers (SNPs) had been created in the database before attempt was made to import the data.

3.6 Statistical analyses

3.6.1 Quality control checks after SNP genotyping

Upon importing genotype data into PROGENY (Figure 3.5), two reports were generated through FBAT (Family-based association tests) package that is linked to

PROGENY: Mendelian Errors and Discrepancies reports. Mendelian errors occurred when a variant in a child was not found in either parent. These may result from sample switches, sample

contamination, unsuccessful genotyping of a sample from a member of a triad and non-paternity. However, since no sample switch and contamination were observed in the Ghanaian samples through XY Genotyping, Mendelian errors were mainly due to non-paternity and less frequently, ungenotyped samples. This premise is based on the fact that de novo mutations were highly unlikely to occur for common variants or SNPs that are usually used for association studies. All samples whose Mendelian errors could not be resolved were excluded from statistical analyses.

Discrepancies report had two phases: sample and marker issues. Some samples were not successfully genotyped on all 48 SNPs and were thus flagged as sample discrepancies. Therefore, if a sample was not successfully genotyped on at least 95% of markers, it was excluded from further genetic analyses because the sample could be of low quality and its results could not be trusted. Marker discrepancies occurred when a marker or SNP was not successfully genotyped on 100% of all the samples used for genotyping. However, no marker was excluded from further statistical analyses at this stage because all 48 SNPs were successfully genotyped on at least 95% of the samples, which was the benchmark. All NTCs and positive control data were also excluded from any statistical analysis.

Final detection and resolution of marker typing incompatibility or suspected errors in pedigrees were carried out using PedCheck (https://watson.hgen.pitt.edu/register/docs/pedcheck.html). PedCheck is a programme that is used to ascertain errors in family structure, as well as to detect Mendelian errors and individuals responsible for the error. This is an advantage over FBAT programme in that PedCheck specifically states the individual responsible for the family structure error or inconsistency. Two types of errors were generated by PedCheck: Level 0 and Level 1 errors. Level 0 errors were associated with pedigree structure errors, such as missing

genotype data for father, mother or affected child. Where no parental genotype data was available because of lack of sample or ungenotyped sample, a "fake" individual was imputed for that parent. For example, for a missing mother, the gender is given as female, affected status was designated as unknown, while genotype data was given as 0, implying there is no actual genotype data for the mother. It must be noted that such "fake" maternal information were not incorporated into genetic analyses, but were just meant to account for the missing individual in a family triad. Level 1 errors were solely Mendelian inconsistencies and stated the pedigree as well as the individual(s) genotype(s) responsible for the error. This type of error was resolved by correcting sample switches or contaminations, if any, or by ignoring the error if it was certain that all individuals stated on the pedigree were indeed the affected child and parents. All samples and pedigrees with unresolved PedCheck errors were not included in the final statistical analyses. All samples that passed all quality control checks were included in the final statistical analyses (Table 3.7).

3.6.2 Association studies: statistical genetics

Statistical analyses were carried out using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) and Family-Based Association Test, FBAT (http://www.hsph.harvard.edu/fbat/fbat.htm) (Figure 3.5). All parameters used in the genotype data were coded after correcting errors and exporting data from PROGENY. This was done before they were used for PLINK and FBAT analyses.

Males were coded as "1" while females had "2". Unaffected individuals were coded as

Table 3.7: Subphenotypes, gender and sample types of study cohort that passed quality control checks and were included in statistical analyses

	Number of samples per population	Total

Cleft Subphenotype of probands	Ghana	Ethiopia	Nigeria	
	Case control cohort			
NSCL	162	101	77	340
NSCLP	144	143	74	361
NSCP	102	21	40	163
Unrelated Controls	408	357	313	1078
	Case-parents trios			
NSCL	52	2	20	74
NSCLP	48	3	26	77
NSCP	34	1	7	42
	Case-parent dyads			
NSCL	77	84	51	212
NSCLP	76	134	47	257
NSCP	53	20	32	105
	other trios			
NSCL	18	0	0	18
NSCLP	14	0	0	14
NSCP	11	0	0	11
	other dyads			
NSCL	8	0	0	8
NSCLP	3	0	0	3
NSCP	3	0	0	3
	Singletons			
NSCL	5	13	6	24
NSCLP	1	8	1	10
NSCP	2	0	1	3
	Tetrads			
NSCLP	2	0	0	2
	Pentads			
NSCLP	1	0	0	1

Note: Case probands consisted of 423 males and 441 females whereas unrelated controls were made up of 441 males and 637 females. The probands in the case-control arm of the study were the same probands in the family-based studies. In some of the designated singletons, parental samples failed data cleaning and were dropped from statistical analyses, hence the designation of such families as singletons. Singletons were informative in the case-control arm of the study but not the family-based studies. Tetrads and pentads were collected from families where two

individuals were affected with clefts. Case-parent trios, tetrads and pentads were employed in Transmission Disequilibrium Test (TDT) whereas all sample types, except singletons and unrelated controls, were used for Family-Based Association for Disease Traits (DFAM) analyses. Only case probands and unrelated controls were included in the case-control analyses. "1" while affected individuals were coded as "2"; individuals with unknown affected status were coded as "0". Alleles that were tagged by A or X dye (VIC) were coded as "1" whiles those that were tagged by B or Y dye (FAM) were coded "2".

Hardy Weinberg Equilibrium (HWE) for each SNP was ascertained in controls using the "--hardy" command of PLINK. All SNPs whose allele and genotype frequencies were usually biased towards controls were designated as deviating from HWE. Only one SNP (rs34743335, p=0.0025) out of the 48 genotyped SNPs deviated from HWE in the Ghanaian cohort so it was excluded from further statistical analyses, since it was not informative. In the African cohort, EPHX2 (rs6558002, p=0.03) failed HWE test. So in both the Ghanaian and combined African cohorts, only one SNP each deviated from HWE. This observation largely suggested the absence of selection bias in the study cohorts. Corollary to this, the normal p-value for all statistical analyses was pegged at $p \leq 0.05$ while the Bonferroni or Corrected p-value for the most inclusive case control association study was pegged at $p < 3.54 \times 10^{-4}$. Bonferroni Correction was used to counteract the effect of multiple testing. The normal p-value for each of the SNP tested was $p \leq 0.05$. However, because 47 SNPs did not deviate from HWE, three main subphenotypes (NSCL, NSCLP and NSCP) formed the study cohort and the cohort involved only one race, the effect of multiple testing was counteracted by dividing 0.05 by 141 (47×3×1) to get the corrected or Bonferroni p-value of 3.54×10^{-4} (Jia et al, 2015; Beaty et al., 2010).

FBAT analysis was used to test for transmission disequilibrium: this refers to departure from the expected 50:50 transmission of alleles from heterozygote parents to an offspring due to chance. The parent may or may not be affected. This test considered only pedigrees where genotype data for both parents were available; at least, a parent should also be

heterozygote for the particular SNP being tested and also have an affected child. Thus, all pedigrees with homozygous parents and no affected offspring were excluded from FBAT analysis. So for FBAT analyses, only pedigrees with sample triads and had a heterozygous parent for a particular SNP were informative (http://www.hsph.harvard.edu/fbat/fbat.htm).

Family-based association designs, in contrast to case-control designs, are particularly attractive, in that they test for both linkage and association; spurious associations that emanate from population admixtures are also avoided. Linkage refers to simultaneous transmission of more than one SNP or trait to an offspring, usually due to closeness of the loci responsible for such SNPs or traits. Family-based analysis also prevents false association or otherwise that may result from population stratification. Population stratification refers to differences in allele frequencies in sub-populations of a larger population due to factors that topple HWE, such as ethnicities, founder effect and non-random mate choice (http://www.hsph.harvard.edu/fbat/fbat.htm).

The Transmission Disequilibrium Test (TDT) tool of PLINK ("--tdt") was used to verify results from FBAT analyses, since it does a similar type of analyses as the FBAT software. TDT, however, test for over-transmission of an allele (not necessarily from a heterozygote parent, as is the case in FBAT) from parents to affected offspring. Thus, the basic TDT Chi Square statistics is calculated simply as $(b-c)^2/(b+c)$, where b and c are the number of transmitted (T) and un-transmitted (U) alleles, respectively. Parent of origin effect (POO) was also tested using the PLINK tool '--tdt --poo". POO tested for a possible paternal or maternal imprinting of alleles that are over-transmitted. Thus, POO tries to trace the source of the allele that is associated with the disease, whether it always come from the mother or father (http://pngu.mgh.harvard.edu/~purcell/plink/).

To include all relatives of cases into the statistical analyses, the family-based association for disease traits (DFAM) function of PLINK ("--dfam") was employed. This analysis combined data from siblings, parent-offspring trios, parent-offspring dyads as well as data from other relatives in a single clustered-analysis that uses the Cochran-Mantel-Haesnzel test. This analysis thus combined all familial genotype data, in a single clustered-analysis and corrected for effects of covariates, such as population stratification, through the Cochran-Mantel-Haesnzel test (http://pngu.mgh.harvard.edu/~purcell/plink/).

Case-control association studies were also performed using the PLINK tool "-assoc" for the Ghanaian cohort only. For the combined African samples (Ghana, Ethiopia and Nigeria), the effect of ethnicity or population heterogeneity was counteracted in case-control analyses by employing the "meta-analyses" function of PLINK. Case-control analysis compared the minor allele frequencies of each SNP in cases to those observed in controls. This approach increased the number of informative families, as it did not rely solely on over-transmission of an allele from a heterozygous or homozygous parent to a patient. However, this analysis is limited by population stratification: unmatched selection of controls may give false results, as allele frequencies may vary between sub-populations (http://pngu.mgh.harvard.edu/~purcell/plink/).

3.6.3 Other statistical analyses

Statistical analyses on environmental factors, phenotypic diversity and zebrafishbased assays were calculated using In Silico Support for Life Sciences by employing the Chi Square function

(http://in-silico.net/tools/statistics/chi2test). All charts were drawn with Excel and error bars indicated are standard errors of mean.

3.7 Sample preparation, primer design and direct DNA sequencing

Direct DNA sequencing was carried out using Sanger Sequencing Technology (Figure 3.9). IRF6, PAX7, MSX1, ARHGAP29, VAX1, BMP4, FOXE1 and MAFB were directly sequenced on 184 case probands with NSOFCs (60 NSCL, 71 NSCLP and 53 NSCP); these probands were randomly selected from the cohort that was used for SNP genotyping. Each individual sample was assigned to a well in a 96-well microplate to form a plate map. Two of such plate maps (Ghana Plate 1 and Ghana Plate 2) were generated for the NSOFC probands sequenced (Appendix I). Also, IRF6 was sequenced on 54 Ghanaian OFC cases that were designated as syndromic cases, including those with VWS. A total of 26 syndromic cases from Nigeria and Ethiopia were also included in the syndromic cohort. In all, IRF6 was also sequenced on 80 syndromic OFC cases. These syndromic cases were individuals with multiple congenital anomalies (MCAs), such as an OFC that presented with club foot.

Each plate map used for sequencing contained case samples, positive controls (NA18856 and NA18855) and negative controls (dH_2O). Sample concentration of 4 ng/uL was used for sequencing. Primer design and sequence results were based on human genome assembly GRCh37/hg19, 2009 (http://genome.ucsc.edu). Primers were designed with Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). Optimal conditions for each primer set were ascertained through Gradient PCR. Exons and flanking intronic sequences, as well as 5' and 3' UTRs of each gene were amplified by Initial PCR. Gel electrophoreses was then carried out on all Initial PCR products.

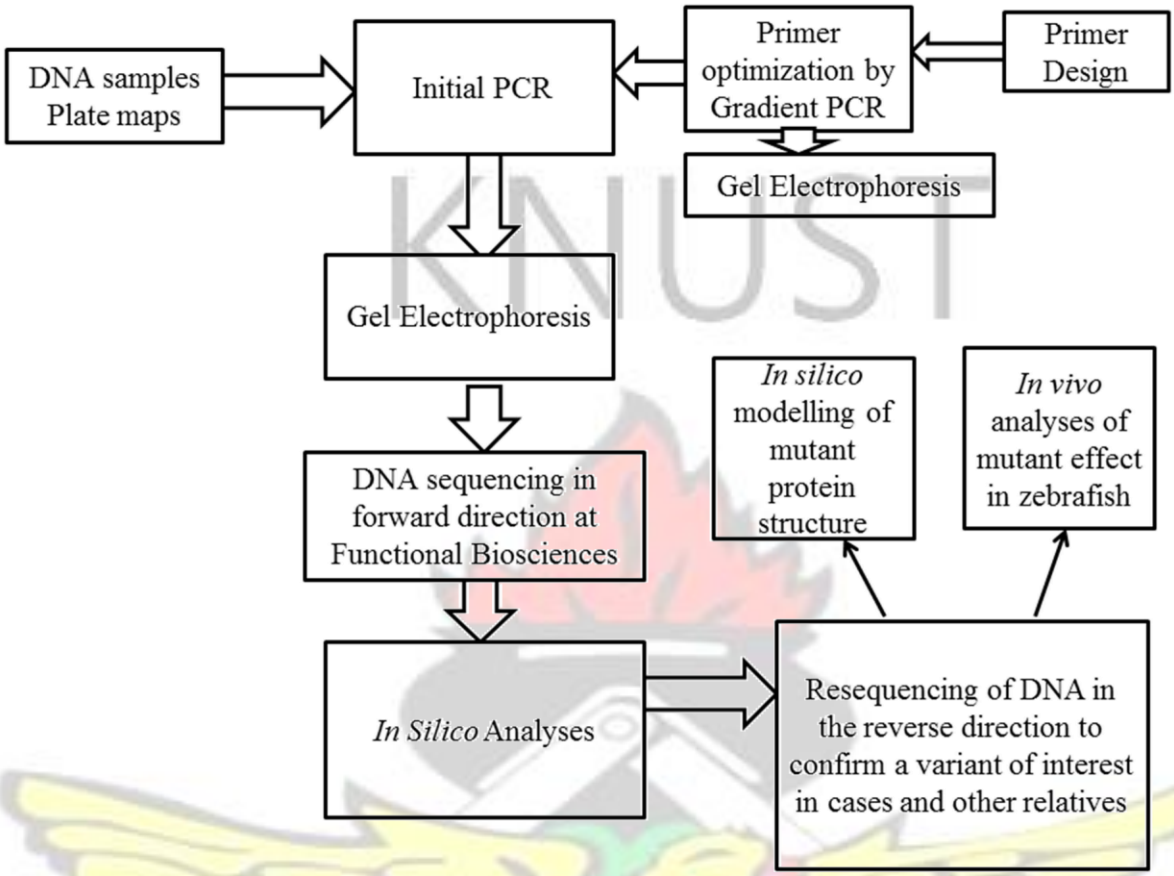


Figure 3.9: Various methodologies employed for primer design and direct DNA sequencing as well as the analyses pipelines

Subsequently, a coded data (CD) number and a corresponding CD map (which is a plate map that shows the primer set that was used for a particular PCR reaction, Appendix I) were generated for each PCR product at http://genetics.uiowa.edu/. Each CD number was then attached to the appropriate PCR plate while each CD map was uploaded to the website of Functional Biosciences (http://order.functionalbio.com/seq/index). These allowed easy matching of sequence results with the corresponding individual sample. The PCR plates were then packaged in a box with CO_2 dry ice and were shipped to Functional Biosciences, Madison, Wisconsin, USA, for DNA sequencing. Usually, the forward primer for each primer set was sent with the PCR plates; however, where sequencing with forward primer failed, the reverse

primer was sent. The concentration of each primer for sequencing was 10 uM/uL and was prepared by diluting 12 uL of stock primer solution with 108 uL of dH_2O to a total volume of 120 uL. This was sufficient for sequencing one 96-well PCR plate and was thus scaled up appropriately when the need arose.

In all, 130 96-well PCR plates were sent to Functional Bioscience for sequencing. These included 31 plates for ARHGAP29, 8 for BMP4, 14 for FOXE1, 18 for IRF6, 6 for MAFB, 6 for MSX1, 18 for PAX7 and 14 for VAX1. Nine plates were also sequenced for IRF6 in the syndromic cohort. Finally, 6 PCR plates were sent for resequencing: this was meant to confirm an observed variant of interest in probands and to ascertain if the variant also occurred in relatives or not.

3.7.1 Primer design

Primers for each gene was designed based on human genome assembly GRCh37/hg19, 2009 (http://genome.ucsc.edu). The desired gene name was imputed into UCSC Genome Browser, followed by zooming into the section of interest for the gene, normally an exon and some few intronic sequences. When two exons of a gene were smaller and close, a single primer set was designed for them. However, when an exon was very large, such as over 1000 base pairs (bp), two primer sets were designed for such an exon. Primer sets were normally designed not to be more than 1000 bp because the optimal read length for Sanger Sequencing Technology is around 800 bp. DNA sequences for all genes were obtained from UCSC Genome Browser (http://genome.ucsc.edu).

Primer was then designed using "Primer3" online software

(http://biotools.umassmed.edu/bioapps/primer3_www.cgi). DNA sequence was copied from the word document and was pasted into the big box, “paste source sequence below”. Brackets, [ ], were then placed around specific sequence of interest. Each line of DNA sequence had 50 bp, so counting 4 lines from the top and adding a bracket, [, left out the extra 200 bp that were added in UCSC; this enabled the primer to be designed in the specific region of interest. Moreover, counting from the bottom of the DNA sequence, a bracket,], was added after 200 bp (each line of DNA sequence had 50 bp). The software then designed the primers using the 200 bp that were out of the brackets on each side of the sequence of interest. The following parameters were selected during the primer design: ‘Pick Left Primer’ and ‘Pick Right Primer’, Number to Return – 20, Product Size - adjusted if necessary to a maximum of usually 1000 bp, Primer Size – 20 to 27 with a minimum of 18, Primer melting temperature (Tm) - 57 to 65, Max Self Complement – 3 and Max 3’ Complement – 2. This was followed by clicking of ‘Pick Primers’.

The best primer set was then selected from the primer sets generated by the system using the following guidelines. Primers that had low values (<4) for the ‘any’ and ‘3’’ columns were picked; these values represent the chance that the primer will bind somewhere else or won’t anneal. The forward and reverse primers were always checked in UCSC BLAT Search to make sure they were specific for the gene of interest; only one result was always expected in BLAT search and had to match 100% and show the correct chromosome and region. The primers were also checked always with “In Silico PCR” on UCSC, making sure that only one product (the correct one) was amplified. Forward and Reverse Primer Tm’s were also ensured that there were within 2^{o}C of each other. Primers were then named based on the gene ID. For example, a primer set for exon 1 of IRF6 was named i6e1m13f and i6e1m13r, representing forward and reverse primers,

respectively (Table 3.8). The primer sequences were then sent to the DNA Core of Institute of Human Genetics, University of Iowa, where the oligoes were synthesized.

3.7.2 Primer optimization through Gradient PCR

After receiving the oligoes or primers from the DNA Core, Gradient PCR was carried out to ascertain the optimum annealing temperatures and chemical conditions that worked best for each primer set. Optimization was carried with DNA from CEPH (Centre d'etude du polymorphisme humain) samples. This was followed by running of gel electrophoreses to pick the best parameters for running Initial PCR, whose product was sent for direct DNA sequencing. The following PCR reagents were used for both Gradient PCR and Initial PCR: 10X NH_4 buffer, $MgCl_2$ (50 mM), deoxynucleotide triphosphates - dNTPs (100 mM stock), Taq Polymerase (5U/uL) and dimethylsulfoxide (DMSO). In few instances where the normal Taq polymerase could not carry out the PCR successfully, a special type of polymerase called Bio-XACT and its additive, High Spec Additive, were used to replace Taq Polymerase and DMSO, respectively. All these reagents were supplied by Bioline, USA (http://www.bioline.com/).

A 2X Biolase was prepared in bulk, usually 10 mL, and was distributed in volumes of 500 uL into 1.5 mL Eppendorf tubes for later use. Prior to use, all reagents were thawed and vortexed well (except Taq) and spun down in a microfuge; this brought any reagent down from the cap of the tube. Taq was kept on ice until it thawed by itself. Biolase was prepared by combining the reagents in the volumes indicated (Table 3.9). The dNTPs were four: dATP, dCTP, dGTP and dTTP; the volume indicated for the dNTPs was picked for each of these four. All these reagents were mixed in a conical tube, which was subsequently vortexed very well before

Table 3.8: List of primers for DNA sequencing

Gene or region	Primer name	Primer Sequence	Product size	Annealing Temperature/ DMSO Treatment
IRF6	i6e1m13f	TGTAAAACGACGGCCAGTATCTGGAAAAGGGCGACAGG	573	62.5(1X DMSO)
IRF6	i6e1m13r	CAGGAAACAGCTATGACCAGAAGCGGAGGAGTAGGGTG		
IRF6	i6e2m13f	TGTAAAACGACGGCCAGTAAAGTTATGGAAACAGCAAC	417	57(1X DMSO)
IRF6	i6e2m13r	CAGGAAACAGCTATGACCTTATTCTAGGGCTTCTGAGC		
IRF6	i6e3m13f	TGTAAAACGACGGCCAGTCATGCCCCCAAAAGAGGAAT	593	57(1X DMSO)
IRF6	i6e3m13r	CAGGAAACAGCTATGACCGGCTAGAGCATGAAGTGTAA		
IRF6	i6e4m13f	TGTAAAACGACGGCCAGTAGGCTTTCTTGCTTTATCCA	607	57(1X DMSO)
IRF6	i6e4m13r	CAGGAAACAGCTATGACCTCCTGCTGCAGTCTCTGTAA		
IRF6	i6e5m13f	TGTAAAACGACGGCCAGTTGCTTTCAGGGCAGTGGTGG	461	62.5(1X DMSO)
IRF6	i6e5m13r	CAGGAAACAGCTATGACCCAGTGAATCTAGGGAGGTCC		
IRF6	i6e6m13f	TGTAAAACGACGGCCAGTTTTACTTCTTCCCTGGTGAC	468	57(1X DMSO)
IRF6	i6e6m13r	CAGGAAACAGCTATGACCCAGTGTTTGGTTCTTGTCTA		
IRF6	i6e7m13f	TGTAAAACGACGGCCAGTCTTGACCTCCTCCAGACTAA	683	57(1X DMSO)
IRF6	i6e7m13r	CAGGAAACAGCTATGACCAGTGGCCTTCCTGAATGCTG		
IRF6	i6e8m13f	TGTAAAACGACGGCCAGTGTTTCAGCAAGACTCTAAGG	471	57(1X DMSO)
IRF6	i6e8m13r	CAGGAAACAGCTATGACCAAAGATGGTATTTGTTGAGT		
IRF6	i6e9m13f	TGTAAAACGACGGCCAGTGTCTTCCTCAGGGCCTCTTT	502	57(1X DMSO)
IRF6	i6e9m13r	CAGGAAACAGCTATGACCGGCATATTTGGAGAATCACAAAC		
PAX7	bv5UTRfw	AGGCTCCTTCTTCCGTCTGT	848	61(1X DMSO)
PAX7	bv5UTRrv	AACGGGATCCCTCTGCTATC		
PAX7	bvex1fwd	TGACTCCTCTATCCATCTCTGC	596	61(1X DMSO)
PAX7	bvex1rvs	TGGGAGGACATCTGGGAGTA		
PAX7	bvex2fwd	AGACCCCCAGCTGCCTTC	482	61(1X DMSO)
PAX7	bvex2rvs	TGCCTTAACTCTGCCTTTCTG		
PAX7	bvex3fwd	TCAATGCCTAAATGCCTGTG	448	61(1X DMSO)

PAX7	bvex3rvs	CTGTCGGCAAAGCCTGAG		

Table 3.8 continued….

Gene or region	primer name	Primer Sequence	Product size	Annealing Temperature/ DMSO Treatment
PAX7	bvex4fwd	GAAGTGCAGTCATGGGTCCT	498	61(1X DMSO)
PAX7	bvex4rvs	CCAACATTTGCTGCCAACTA		
PAX7	bvex5fwd	AGGAGGATGCTGGCTGGAT	398	61(1X DMSO)
PAX7	bvex5rvs	TGCTTACATGAAAGTAAGAAATGGTC		
PAX7	bvex6fwd	GCCCAGGAGCCAACTTGT	500	61(1X DMSO)
PAX7	bvex6rvs	TTCTAGAACAAAAACCCCAACC		
PAX7	bvex9fwd	CAATGTGGGTAATCTTCTGTGC	838	61(1X DMSO)
PAX7	bvex9rvs	AGACTCCAAGGGCATCACTCT		
PAX7	bv3UTRfw	GCTGAGTTCTCAGTGGAAGGA	595	61(1X DMSO)
PAX7	bv3UTRrv	CACCCTGACACCACCTTGTA		
MSX1	msx1e1fd	GCCTCGCCTTATTAGCAAGT	916	56.3 (2x DMSO)
MSX1	msx1e1rv	CTTCTTCCTGGGTGGGAGT		
MSX1	mse2p1fd	AGCCTTCAACGTGGGTATTTT	815	56.3 (1X DMSO)
MSX1	mse2p1rv	CAGTGTGAGGGTTAAAGGGAAG		
MSX1	mse2p2fd	CTCGCTCTACGGTGCCTCTG	983	58.9 (1X DMSO)
MSX1	mse2p2rv	CCGTCAAACTGAAAAGCAACTG		
ARHGAP29	daex2fwd	GTGGGCTGTTTGTGAAAATCTA	447	62.1 (1XDMSO)
ARHGAP29	daex2rev	GCAGAATACCCTGTTCCTACCA		
ARHGAP29	daex3fwd	ATGTGCTTCAGATGACGTTTTG	326	60 (2XDMSO)
ARHGAP29	daex3rev	ACATCACACAAATGCAATGAGA		
ARHGAP29	daex4fwd	ATGCCAGAAAGCAGTAAGGAAG	612	60 (1XDMSO)
ARHGAP29	daex4rev	GCCATACAGAAGGTAGTTTTTCC		
ARHGAP29	daex5fwd	TGAGAGTATGTTGCATTATCCAGT	884	61.3 (NO DMSO)
ARHGAP29	daex5rev	CATAAGTCCCTGATGCAGTTTTT		

Table 3.8 continued…

Gene or region	primer name	Primer Sequence	Product size	Annealing Temperature/ DMSO Treatment
ARHGAP29	dae6e7fw	GATCTTGGGAACTCCTGTTGTAAA	971	61.3 (NO DMSO)
ARHGAP29	dae6e7rv	GAGTCTTTTCTGGATAATGGTTTT		
ARHGAP29	daex8fwd	CCTTCTGATACCCTCTTTGAACTT	829	61.3 (NO DMSO)
ARHGAP29	daex8rev	GCAGTGGCATGAACTCCTAAA		
ARHGAP29	dae1011F	ATTGCAGCTCTCCAGGCTAAC	843	57.5 (1XDMSO)
ARHGAP29	dae1011R	TTCTTTCACAAACTCCAAAATTCA		
ARHGAP29	daex12fw	CTGATAGCAAGGATAGGCAAAAG	644	60 (1x DMSO)
ARHGAP29	daex12rv	AGAGTGGGTAATCATAATCGAAGC		
ARHGAP29	dae1415F	TTCACTTCCCATGTTTTCACTG	770	61.4 (NO DMSO)
ARHGAP29	dae1415R	ATGAGTTTTCATTGCCAGACTT		
ARHGAP29	daex16fw	CTTACCATCAGAAGTTGGCAGT	609	59.6 (NO DMSO)
ARHGAP29	daex16rv	ACATTTTACATCCAGTAGTTTTGCT		
ARHGAP29	dae1718F	TGACTTGGTGACGATTTTTACTC	937	61 (NO DMSO)
ARHGAP29	dae1718R	CCCCTGAGCATTATTTTCCTTT		
ARHGAP29	daex19fw	ATCTGCATTTAGCCATTTGTACT	310	55.5 (NO DMSO)
ARHGAP29	daex19rv	CCGTCTATGCCAATGAAAAA		
ARHGAP29	daex20fw	CCACATCTGTAGCAAACATTTACA	907	59.4 (NO DMSO)
ARHGAP29	daex20rv	TTCAATGTCTCCCTTTCCCTTA		
ARHGAP29	dae2122F	TGAAACTCTATTTGGCATGTGG	841	62.2 (NO DMSO)
ARHGAP29	dae2122R	TTCTCTGGCTTCCCTCATTG		
ARHGAP29	daex23AF	AACTCCTCCACATAGGTAACTGAG	968	57.5 (1XDMSO)
ARHGAP29	daex23AR	TGGTCTCTGACACATTGGATTC		
ARHGAP29	daex23BF	GGTCTCGTGGTGAAGTCAATG	934	57.3 (NO DMSO)
ARHGAP29	daex23BR	AAGGGGTGAAAAACTATCAAATGT		

Table 3.8 continued…

Gene or region	primer name	Primer Sequence	Product size	Annealing Temperature/ DMSO Treatment
VAX1	bw5UTRfw	GCTGTTCGCTACCTGATCGCC	710	57.4 (2X DMSO)
VAX1	bw5UTRrv	GCCAACAACTTTCTCCCAAG		
VAX1	bwex1fwd	GCCCTCCACACAGTGTCTTT	500	63.3 (NO DMSO)
VAX1	bwex1rvs	AGGCCTGTCTTACCCATCCT		
VAX1	bwisoa3f	GAGCTCGCTTCCTTTGACG	728	60 (2X DMSO)
VAX1	bwisoa3r	AGCTCTGGGCACCTAATGC		
VAX1	bwex23UF	TCTGAAGCAAGCGAAAAACA	978	62.1 (NO DMSO)
VAX1	bwex23UR	GTGCAGGCGCTACAAGAAG		
BMP4	b4ex1fwd	GTGAGTGGATGGGAACGTGT	600	62.1 (NO DMSO)
BMP4	b4ex1rvs	GACTCACCTCCATCAGACTCG		
BMP4	b4ex2fwd	CGCTGTGAGTGATGCTTAGG	827	63.3 (BioX/No DMSO)
BMP4	b4ex2rvs	TGATTGTGGTGCCATAGAGG		
BMP4	b4ex3fwd	GAACACCTCCCCCTCTGTCT	900	60 (1X DMSO)
BMP4	b4ex3rvs	AGAAGCCACGCTGAGATCAT		
BMP4	b4ex4fwd	CAGGGAGCAGAGTGTGGATA	801	57.5 (1X DMSO)
BMP4	b4ex4rvs	GCTCGTGTTTGCTTGTGTGT		
FOXE1	e1ex1p1f	CCCTTTAAGGAGGGGAAGCCG	628	61.4 (1X DMSO)
FOXE1	e1ex1p1r	TTCTGCGGAGAGCTCAGGGGA		
FOXE1	e1ex1-2f	GTCACTCCCGAGCCTCTGT	700	62.5 (1X DMSO)
FOXE1	e1ex1-2r	TAAGCCGGGTAGGTGGAGAG		
FOXE1	e1ex1-3f	CAAGCGCTCGGACCTCTC	700	55.5 (1X DMSO)
FOXE1	e1ex1-3r	ATGAGTTTTCGTCCCTACGC		
FOXE1	e1ex1-4f	CTGGGAGCCTGCTACAACC	658	61.4 (1X DMSO)
FOXE1	e1ex1-4r	CAGGTCAACCAGAGCAGAGA		

Table 3.8 continued…

Gene or region	primer name	Primer Sequence	Product size	Annealing Temperature/ DMSO Treatment
FOXE1	e1ex1-5f	TACGCGGTTCGTCCTCTAGT	699	57.3 (2X DMSO)
FOXE1	e1ex1-5r	CCCTCCCCAGCTACTGAAAG		
FOXE1	e1ex1-6f	GCAGGACGTGCTGGTAATT	696	57.3 (2X DMSO)
FOXE1	e1ex1-6r	CCCATTTGGACTGAACCAAG		
FOXE1	e1fxgapf	CTACAGCTACATCGCGCTCA	520	61.4 (1X DMSO)
FOXE1	e1fxgapr	AGGAACCAGGCCGAAGAC		
MAFB	axex1fw1	GAGCAAGAGAGCTAGAGAGCGAGC	847	62.5 (1X DMSO)
MAFB	axex1rv1	ACGCTTGGTGATGATGGTGATG		
MAFB	axex2fw2	CGAGCAACTACCAGCAGATGAAC	116	61.4 (1X DMSO)
MAFB	axex2rv2	AAAGCTGTCGAAGCTTTGCAG		
MAFB	axex2fw3	ACACCGCTCAGCACTCCGTG	1048	62.5 (1X DMSO)
MAFB	axex2rv3	TCTCCGGCTCTGCTCGAGTCTAG		

onward distribution into labeled Eppendorf tubes

Table 3.9: Preparation of biolase

Component	For 10mL	For 5 mL	Final Concentration
dH_2O	7.14 mL	3.57 mL	
10X NH_4Buffer	2.00 mL	1.00 mL	2X
$MgCl_2$	600 uL	300 uL	3 mM
dNTP (100 mM stock)	40 uL each	20 uL	400 uM each
Taq @ 5U/uL	100 uL	50 uL	

Rehydrating primers and making dilutions: When the primers were received from DNA Core, they were dry at the bottom of a blue-cap tube. The tubes were spun to make sure the primer was down at the bottom. The primers were rehydrated with deionized water (dH_2O). The primers usually need to be 100 uM stocks, so the volume of dH_2O that was added was ten times the amount of primer listed in nmol on the original tube from DNA Core. For example, if the primer tube said 32.0 nmol, 320 uL of water was added. The tube was then vortexed really well to properly re-suspend the primers; else dilutions will not be accurate. Subsequently, 20 uM dilution was made as the working solution to be used for PCR. This was done by labeling new 2.0 mL screw cap tubes with the labels that came with the primer, followed by the pipetting of 20 uL of 100 uM stock and 80 uL dH_2O into the screw cap tubes. The screw cap tubes were then vortexed very well to properly re-suspend the primers before using them for PCR.

Optimizing primers: Primers were optimized before they were used with the Ghanaian samples in order to determine the appropriate conditions for the PCR reaction. To achieve this, the C1000 Touch^{TM} thermal cycler (BIO-RAD, USA) was used to run a "gradient PCR" which varies the annealing temperature across each row of the plate.

This allowed the testing of the quality of the PCR reaction for a range of temperatures at the same time. The primers were tested on CEPH DNA, which were diluted to an aliquot of 20 ng/uL. A gradient PCR form (Appendix J) was filled based on the name of each primer set (forward and reverse) as well as the chemical gradients. A gradient PCR reaction usually lasted for about 3 hours.

Eppendorf tubes were labeled for the preparation of a mastermix for each primer set to be tested. Reagents were removed from freezer boxes and allowed to thaw on ice, followed by vortexing. Biolase was always kept on ice. The PCR plate was also labeled with appropriate information, such as primer name, date, DNA plate, etc. One uL of diluted CEPH DNA was pipetted into each well of the PCR plate, using a 25 uL repeat pipette. The mastermix for each primer set was prepared by combining the appropriate amounts of each reagent (Appendix J) into the mastermix labeled tubes; usually the 10X volume on the gradient PCR form was used. The appropriate chemical gradients were created by varying the volumes of DMSO and dH_2O. For example, three tubes were labeled for each primer set: one had no DMSO (dH_2O = 37.0 uL, DMSO = 0), the other tube had 1X DMSO (dH_2O = 32 uL, DMSO = 5.0 uL) whiles the last tube had 2X DMSO (dH_2O = 27.0 uL, DMSO = 10.0 uL). The content of each tube was then vortexed thoroughly.

A total of 9 uL of each mastermix was pipetted into a well on the PCR plate using a repeat pipette. Each mastermix for a primer set was distributed into wells in a column on the PCR plate (for example, A1 to H1). The PCR plate was tightly sealed with plastic PCR covers; the plate was sealed very tightly to avoid evaporation. The PCR plate was centrifuged at 1500 rpm for 1 minute to thoroughly mix its content. The PCR plate was placed on the gradient PCR machine (C1000 Touch^{TM} thermal cycler), making sure that A1 on the PCR plate corresponded with A1 on the machine. The Gradient PCR machine was run by

selecting a desired range of annealing temperatures (usually 52 to 65) at 10 uL reaction volume. The annealing temperature for each row on the PCR plate was recorded on the corresponding row on the gradient PCR form

(Appendix J). Usually, 8 different temperatures appeared for each of the 8 rows (A1 to H1) on the PCR plate. After the gradient PCR, gel electrophoreses was run to visualize which temperature and chemical treatment worked best for a particular primer set. The best DNA band on the electrophoretic gel for a particular primer set was chosen as the optimal annealing tempertature (Appendix K).

3.7.3 Gel electrophoresis

Electrophoretic procedure was carried after running every gradient PCR or Initial PCR. A total of 4 g of 2% agarose (Research Products International, USA) was weighed into a glass flask. Then 200 mL of 0.5X TBE buffer (Tris/Borate/EDTA - Ethylenediaminetetraacetic acid) was measured by a beaker and added to the same glass flask. The content of the glass flask was microwaved for 3 minutes. The flask was picked up thereafter with a rubber potholder and was observed to ensure that the agarose had dissolved and that the solution was clear and not grainy. If this was not observed, the flask was microwaved again for one minute.

While allowing the agarose to cool, the gel rig was set up. The gel bridge was placed in the plastic tray, followed by placement of the desired combs (usually six 17well combs) on the gel bridge. The combs were positioned so that they were 1 mm above the gel tray so as to avoid the tearing of the gel during the removal of the combs. After complete cooling of the gel by running the flask under a stream of water, 10 uL of ethidium bromide was added to the flask using the labeled ethidium bromide pipetteman.

This was done with gloved hands and the pipette tip was discarded into the biohazard container, since ethidium bromide is carcinogenic and toxic. The flask was then swirled until the ethidium bromide was well mixed into the agarose. If the gel does not cool before the addition of ethidium bromide (EtBr) to the gel, the EtBr may aerosolize and this is detrimental to one's health! The agarose gel was poured into the gel rig gently. Air bubbles, when they occurred, were poked with a pipette tip and the tip discarded into a biohazard container. The gel was allowed to solidify for 30 minutes at room temperature, followed by removal of the combs. The gel bridge was then placed into an electrophoretic tank that was filled with 0.5X TBE that was sufficient to cover the gel. The buffer in the electrophoretic tank was changed after every five run of electrophoreses.

While the gel was solidifying, the PCR product was prepared for loading by adding 50% loading dye: 5 uL of loading dye to the 10 uL gradient PCR product in the same PCR plate and 2 uL of loading dye to 4 uL of Initial PCR product into a different PCR plate. The dye and the sample were mixed very well by vortexing and centrifuging. The sample was loaded into the gel using the "loading" pipetteman by gently inserting the pipette tip into the top of the well, then slowly depressing the plunger to load the sample. This was followed by loading of 3 uL of 100 bp ladder into the middle-most wells. This enabled the determination of the success of the PCR run: it helped to know if the PCR reaction worked well or not, and also helped to determine if the product observed was of the expected size. It was ensured that the gel was oriented such that the products would travel through the gel towards the positive lead (red lead). The cover of the electrophoretic tank or gel box was then placed on it, followed by connection of the leads to power supply (red- positive, black-negative) and the power supply turned on.

Electrophoreses was run for 20 minutes at 200V.

When the run was finished, the migration of the product was checked using an ultraviolet (UV) light source by turning off the power supply to the gel box, removing the leads and transporting the gel to the UV light box. The ethidium bromide intercalates with the DNA and fluoresces in UV light. All light in the gel room was turned off and the UV source was put on. Bands of DNA appeared as glowing orange in the gel if the PCR reaction was successful. A picture of the gel was taken using the "Alpha Imager 2200" programme on the attached computer (Appendices K and M).

3.7.4 Initial PCR and direct DNA sequencing

GeneAmp^{R} PCR System 9700 (Applied Biosystems, Foster City, USA) was used for all initial PCR under the conditions indicated earlier on (Table 3.8). All Initial PCRs went through 35 cycles of amplification. Initial PCR form (Appendix L) was filled for each primer set. This information was critical for tracking possible machine breakdowns, contamination of DNA plates or faulty reagents such as operator error in making biolase. Reagents were removed from freezer boxes and allowed to thaw on ice. Primers were thoroughly vortexed once they had fully thawed and biolase was also kept on ice always. This was followed by the preparation of a mastermix for each primer set.

The volume of each reagent for the mastermix was usually the volume indicated for 100X of Appendix L, because almost all initial PCR reactions were carried out on full 96-well PCR plates. When a lower number of samples had to be sequenced, initial PCR was carried out by appropriately scaling up the volumes indicated under 1X of Appendix L. Once all reagents for the mastermix were in a 1.5 Eppendorf tube, the tube was thoroughly vortexed to mix its content. Mastermix preparation made pipetting easier, increased accuracy of the process and also helped to account for pipetting errors. A 96-

well PCR plate was labeled with the appropriate information, such as primer name, date, DNA plate, and the researcher's name. A total of 1 uL of DNA was pipetted from the 96-well microplate working stock into the 96-well PCR plates. This was carried out using a multichannel pipette and it was ensured each tip had pulled up DNA and was emptied into the well. The mastermix was re-vortexed and distributed into each well of the PCR plate at a volume of 9 uL using a motorized repeater pipetteman.

The PCR plate was tightly sealed with plastic PCR covers so as to avoid evaporation. This was followed by centrifuging the PCR plate at 1500 rpm for 1 minute to mix its content thoroughly.

The PCR plate was placed on the initial PCR machine, followed by running of the appropriate programme for the primers used (Table 3.8; Appendix L). The PCR programme was run on "hot-start", by heating the machine for 5 minutes at 94□C before putting the PCR plate on it. This prevented the reaction from occurring at temperatures other than the specific ones wanted. After an Initial PCR, gel electrophoreses (Section 3.7.3) was carried out on 4 uL out of the 10 uL reaction volume of each well to ascertain the success of the of the PCR run.

3.7.5 Analyses of DNA sequence results

Amplified DNA products were sequenced at Functional Biosciences in Madison, Wisconsin (http://order.functionalbio.com/seq/index) using an ABI 3730XL DNA Sequencer (http://www.appliedbiosystems.com/absite/us/en/home.html; Figure 3.9) that employed Sanger Sequencing Technology. Chromatograms were transferred to a Unix workstation in the University of Iowa, base called with PHRED (http://www.phrap.org/phredphrapconsed.html, v.0.961028), assembled with PHRAP

(http://www.phrap.org/, v.0.960731), scanned by POLYPHRED (http://droog.gs.washington.edu/polyphred/, v. 0.970312) and viewed with the CONSED programme (http://www.phrap.org/consed/consed.html, v. 4.0).

The genomic location of each variant revealed by CONSED was ascertained using the "Blat" function which is under the "Tools" tab of UCSC Genome Browser (https://genome.ucsc.edu/). Variants and their effects were recorded on the appropriate CD maps. The effect of a coding or exonic variant on protein structure was predicted using POLYPHEN-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/) and Ensembl (http://www.ensembl.org/Homo_sapiens/Tools/VEP). Simulation of mutant protein structure was carried out using HOPE (http://www.cmbi.ru.nl/hope/method). Effect of a variant on mRNA splicing was ascertained using Human Splicing Finder 3.0 (http://www.umd.be/HSF3/). Furthermore, effect of a mutation on a regulatory region was predicted using RegulomeDB (http://regulomedb.org/) and Haploreg (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php).

The Minor Allele Frequencies (MAF) or novelty of a mutation was ascertained by comparing it to variants found in 1000 Genomes (http://browser.1000genomes.org/index.html), Exome Variant Server (http://evs.gs.washington.edu/EVS/), dbSNP (www.ncbi.nlm.nih.gov/SNP/) and ExAC Browser (http://exac.broadinstitute.org/). Mutations were classified as "novel" if they have never been reported in any of these databases or literature. Mutations were classified as "rare" if they have been reported in these databases or literature but have a minor allele frequency (MAF) of 1% or less. Finally, variants were classified as "common" if they have been previously reported in any of these databases and/or literature and had a MAF

of more than 5%. Examples of such “common variants” are the SNPs that were used for SNP genotyping.

Upon detection of a variant of interest in proband, samples from other relatives of the proband, usually parents, were also pulled. These samples were then sequenced; the sample for the proband was also re-sequenced with those of the relatives. This was achieved by re-carrying out initial PCR and gel electrophoreses for the samples, followed by submission to Functional Biosciences for re-sequencing. However, this time, the reverse primer of a primer set was sent with the initial PCR product. This resequencing was intended to confirm the variant by re-sequencing the gene in the reverse direction. It also helped to detect whether a variant was de novo (originated solely from the affected child) or segregated (was transmitted from an affected or unaffected parent to the affected child).

3.8 In Vivo analyses of rare functional variants

Two novel variants that were observed in IRF6 through direct DNA sequencing were selected for functional analyses (Figure 3.10). These variants were p.Glu69Lys that was observed in two NSOFC cases and p.Gly65Val that was observed in a VWS case. The functional analyses were intended to model the mechanisms of action of the mutant proteins: whether dominant-negative, haploinsufficiency or null. In dominant-negative, a gene gets mutated on one chromosome, but not both, of a diploid organism. The protein from the mutant allele gains competitive advantage which then selectively binds to the target site and prevents the functional protein from the wildtype allele from carrying out its function. The end result is that none of the protein from either allele functions normally, leading to a disease phenotype. A gene also gets mutated on one chromosome, but not both, in haploinsufficiency; however, the disease phenotype results because not enough

protein is made. Finally in null mechanism, no protein is made at all or no functional protein is made because a gene gets deleted or mutated, respectively, on both chromosomes of a diploid organism. The disease phenotype then results from complete lack of the protein (Peyrard-Janvid et al., 2014, Kondo et al., 2002). The dominantnegative mechanism was tested in variants that were observed in IRF6.

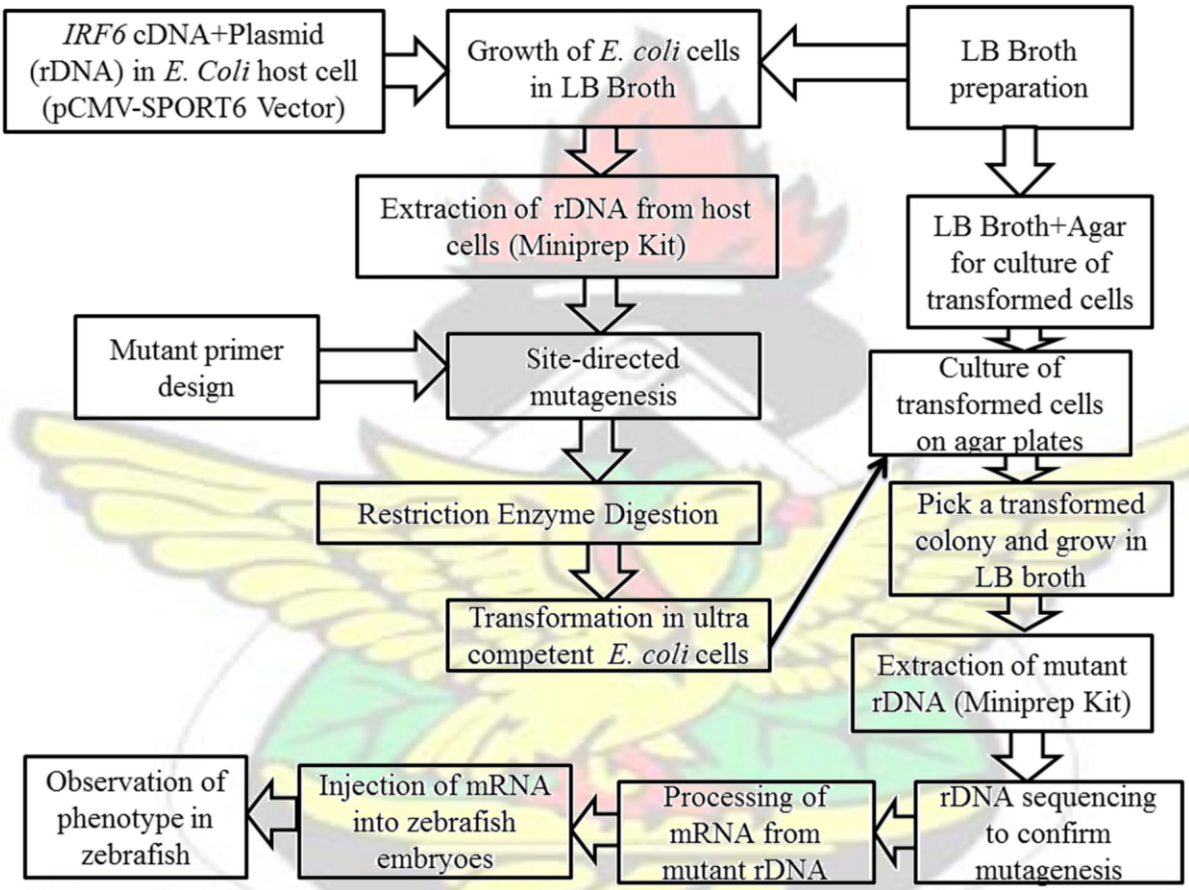


Figure 3.10: Flow chart for functional analyses or in vivo assays

Genetic engineering and developmental biology techniques were used to model the effects of these variants in zebrafish in vivo. The pCMV-SPORT6 expression vector (Figure 3.11), with IRF6 cDNA cloned into it, was obtained from the Robert Cornell Laboratory, University of Iowa, USA. The size of the recombinant DNA (rDNA) was about 9kb: about 4.2kb vector backbone and about 4.6kb IRF6 cDNA. This rDNA was

housed in dam^{+} (DNA adenine methylation) E. coli host cells as a starter culture. The starter culture was multiplied or cloned by growing a colony in Lysogeny broth (LB broth). Upon successful growth of the starter culture colony, some recombinant E. coli cells were stored as a backup while majority of the cells were processed to get the rDNA. Site-directed mutagenesis was then carried out in order to introduce the mutations into the desired position in the cDNA that was situated in the rDNA that was extracted. Dpn I restriction enzyme was used to digest all parental rDNA, leaving only plasmids with solely mutant rDNA intact. Mutant rDNA was used to transform ultracompetent E. coli cells, followed by plating of transformed cells on agar plates. A colony of transformed cells was cultured in an LB broth. Mutant rDNA was subsequently extracted from the transformed cells, using the QIAprep® Spin Miniprep

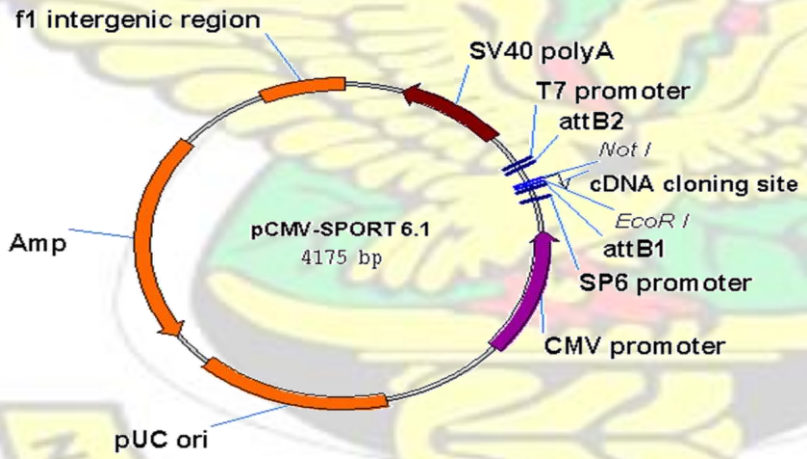


Figure 3.11: Vector map of pCMV-SPORT6 construct. attB1 and attB2: Bacterial attachment sites sequence, which enhances site-specific recombination and insertion of PCR product into vector, Amp: ampicillin resistant gene, which acts as a selectable marker.

Kit. The rDNA concentration was ascertained using the Qubit Assay. To confirm the presence of the desired mutation in the rDNA, sample of the rDNA from each reaction was sent to Functional Biosciences for sequencing. The mMESSAGE mMACHINE® SP6 Transcription Kit was used to transcribe mRNA from the rDNA, followed by injection of

the mRNA into wildtype zebrafish embryos. The zebrafish embryos were observed at various time points, followed by knocking-down and staining of the embryos with Alcian blue (a dye that stains extracellular matrix associated with chondrocytes) at day 7 for observation under fluorescence microscope.

3.8.1 Growth of recombinant plasmid in E. coli host cells

Preparation of LB broth: The following reagents were measured into a labeled glass jar: 500 mL of MilliQ dH_2O, 5 g of tryptone, 2.5 g of yeast extract and 5 g of NaCl. The content was subsequently mixed by inverting it up and down several times. This was followed by autoclaving for 20 minute, with subsequent cooling of the content. The LB was stored at room temperature for later use.

Preparation of agar plates: agar plates were prepared following the same procedure for the preparation of LB broth. However, in addition to the reagents stated above, 7.5 g of agar was added to the glass jar that was meant for agar plates. Upon cooling of the jar to a temperature of 60^oC after autoclaving, 5 mL of ampicillin at a concentration of 10 mg/mL was added. The content of the jar was dispensed into labeled petri dishes, allowed to solidify and were stored for later use. The ampicillin solution was prepared by weighing 0.1 g of ampicillin powder into 10 mL of dH_2O, followed by thorough vortexing.

Growth of starter culture: 10 mL of LB broth was pipetted into a labeled 50 mL conical tube. A total of 100 uL of ampicillin at a concentration of 10 mg/mL was pipetted into the conical tube. The ampicillin acted as selectable marker that allowed only cells with the pCMV-SPORT6 vector to grow. A single colony of the starter culture was picked into the conical tube using a loop, followed by vortexing for 10 seconds. The colony was subsequently grown by placing the conical tube on an orbital shaker for at least 16 hours

at 225 rpm and a temperature of 37ºC. The caps of the tubes were loosened a bit to allow air into the culture. After 16 hours, the colour of the culture turned from the deep yellow colour of the LB broth into a cloudy, colourless culture (Plate 3.3a and b).

3.8.2 Extraction of rDNA from E. coli host cells

Freezing down glycerol stock of E. coli: 1.5 mL of the culture was pipetted into a labeled 1.5 mL Eppendorf tubes. The tube was centrifuged at 8100 rpm for 3 minute. The supernatant was discarded, leaving the pelleted cells. One mL of culture was added to the tube again, re-centrifuged and the supernatant was discarded again. A total of 350 uL of already autoclaved 50% glycerol was pipetted into the tube with the pellet. The content was mixed thoroughly by vortexing, followed by transfer of the content into a labeled 2.0 mL screw cap cryo vial. The screw cap vial was stored at -80ºC freezer as a backup.

Extraction of rDNA from E. coli using QIAprep® Spin Miniprep Kit: All reagents, except Buffer P1, were stored at room temperature (15-25ºC), per the manufacturer's instruction (https://www.qiagen.com/). Buffer P1 was prepared for use by adding LyseBlue® to it at a ratio of 1 to 1000, followed by the addition of the provided RNase A solution, with subsequent vortexing and storage at 2-8ºC. Buffer PE was prepared for use by adding 24 mL of 100% ethanol to it.

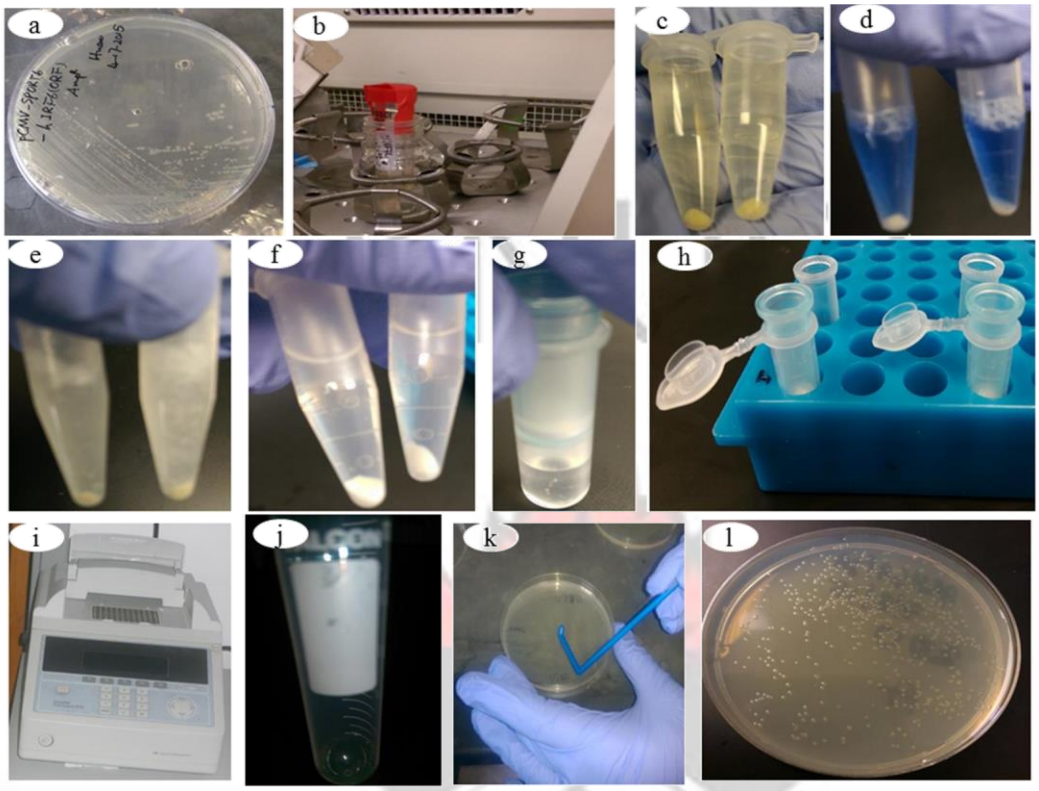


Plate 3.3: Plasmid culture, site-directed mutagenesis and transformation. a: E. coli starter culture that harbours pCMV-SPORT6 plasmid. b: Culture of a colony of the starter culture in LB broth on orbital shaker. c: After culturing, LB broth was centrifuged to get E. coli host cells as a pellet for onward extraction of rDNA. d: Addition of lysing buffer changes the colour of reaction to blue. e: addition of neutralizing buffer turns reaction colourless. f: After centrifugation, cell debris form pellet whereas rDNA remains in the supernatant. g: the spin column that was used for the various rDNA washing steps. h: elution of rDNA from spin column into Eppendorf tubes. i: PCR machine for sitedirected mutagenesis. j: transformation in falcon tubes. k: plating of transformed cells on agar plates. l: growth of transformed cells after incubating them overnight.

The remaining 7.5 mL culture was processed for rDNA. A total of 1.5 mL of the culture was distributed into three labeled 1.5 mL Eppendorf tubes. The tubes were spun at 8100 rpm for 3 minutes, followed by discarding of the supernatant. One mL of the culture was added to each of the three tubes again and the tubes were centrifuged again as indicated. The supernatant was discarded, leaving the pellet of E. coli cells (Plate 3.3c). A total of 250 uL of Buffer P1 was added to each tube and vortexed thoroughly to re-suspend the pelleted bacterial cells. Furthermore, 250 uL of Buffer P2 was added to

each tube and the content mixed thoroughly by inverting the tube 4 to 6 times. This lysing reaction was allowed to proceed for 5 minutes and the solution turned blue because of the usage of LyseBlue® (Plate 3.3d). Subsequently, 350 uL of Buffer N3 was added to each tube and content mixed immediately and thoroughly by inverting the tube 4 to 6 times. The solution turned colourless (Plate 3.3e). The tubes were centrifuged at 13,000 rpm for 10 minutes in a microcentrifuge to pellet all cellular debris (Plate 3.3f).

Subsequently, the supernatant from each tube was applied to the corresponding labeled QIAprep spin column (Plate 3.3g) by decanting and pipetting. The spin columns were centrifuged at 13,000 rpm for 1 minute, followed by the discarding of the flowthrough in the collection tube. Each of the QIAprep spin column was washed by adding 500 uL of Buffer PB, followed by centrifugation at 13,000 rpm for 1 minute and the discarding of the flow-through. This step was required to reduce the nuclease activity of the solution under processing. Each column was washed further by adding 750 uL of Buffer PE, followed by centrifugation at 13,000 rpm for 1 minute, with subsequent discarding of the flow-through. With the QIAprep spin column in the collection tube, the column was centrifuged again at 13,000 rpm for 1 minute to remove any residual buffer.

Each of the spin columns was then placed in a corresponding labeled 1.5 mL Eppendorf tubes (Plate 3.3h). The rDNA was eluted by adding 50 uL of Buffer EB (10 mM Tris.Cl, pH 8.5) to each column. The columns were allowed to stand for 1 minute, followed by 1 minute of centrifugation at 13,000 rpm to collect the rDNA. The concentration of rDNA in each tube was obtained through Qubit Assay.

3.8.3 Site-directed mutagenesis

The primer for each mutation was designed using QuikChange Primer Design (http://www.genomics.agilent.com/primerDesignProgram). The generated forward

(i6E69Kfd) and reverse (i6E69Krv) primer sequences for the p.Glu69Lys variant were 5'-CAGGGAAGTACCAGAAAGGGGTGGATGACC-3' and 5'-GGGTCATCCACCCCTTTCTGGTACTTCCCTG-3', respectively. Furthermore, the generated forward (i6G65Vfd) and reverse (i6G65Vrv) primer sequences for p.Gly65Val variant were 5'-GGGCTGTAGAGACAGTGAAGTACCAGGAAGG-3' and 5'-CCTTCCTGGTACTTCACTGTCTCTACAGCCC-3', respectively. The mutant allele in each primer is shown in red. The designed primers were synthesized by Agilent Technologies, USA. The primers were rehydrated to a stock concentration of 100 nM by adding dH_2O, followed by thorough vortexing. This was achieved by multiplying the quantity of primer stated on the tubes by 10 to get the volume of dH_2O to be added. The working primer solution was obtained by combining 2 uL of stock solution and 18 uL of dH_2O, followed by thorough vortexing. The rDNA was also diluted to a concentration of 10 ng/uL before use.

The Agilent QuikChange II XL Site-Directed Mutagenesis Kit was employed for the site-directed mutagenesis. The following reagents were combined in a well of a PCR plate for each of the mutants: 5 uL of 10X buffer, 1.25 uL of rDNA at a concentration of 10 ng/uL, 1.25 uL each of forward and reverse primers, 1 uL of dNTP mix, 3 uL of QuikSolution, 37.25 uL of dH2O and 1 uL of PfuUltra HF DNA polymerase at a concentration of 2.5 U/uL. The PCR plate was vortexed thoroughly followed by centrifugation at 1500 rpm for 1 minute and was run on the PCR machine under the indicated conditions (Table 3.10; Plate 3.3i). Upon completion of PCR, the reaction was transferred into labeled, pre-chilled Eppendorf tubes and was placed on ice for 2 minutes to cool the reactions to less than 37^{o}C.

Table 3.10: PCR parameters for site-directed mutagenesis

Segment	Cycles	Temperature	Time
1	1	95°C	1 minute
2	18	95°C	50 Seconds
		60°C	50 Seconds
		68°C	9 minutes
3	1	68°C	7 minutes

3.8.4 Restriction enzyme digestion and transformation

Dpn I digestion of Amplification product: Upon cooling of the PCR reaction below 37°C, 1 uL of Dpn I restriction enzyme (10 U/uL) was directly added to each amplification reaction. Each reaction was mixed thoroughly and gently by pipetting the solution up and down several times, followed by centrifugation in a microcentrifuge for 1 minute. The reaction was incubated at 37°C in water bath for 1 hour to digest the parental (non-mutated) supercoiled dsDNA. The Dpn I endonuclease cleaves DNA at the target sequence 5'Gm^6ATC-3' and is specific for methylated and hemimethylated DNA. It is thus used to digest the parental DNA template and to select for mutation-containing PCR amplified DNA. DNA isolated from almost all E. coli strains is dam methylated and therefore susceptible to Dpn I digestion (http://www.genomics.agilent.com).

Transformation of XL10-Gold Ultracompetent Cells: The XL10-Gold ultracompetent cells were gently thawed on ice, followed by pipetting of 45 uL of cells into two pre-chilled 14 mL Falcon polypropylene round-bottom tubes, one tube for each variant (Plate 3.3j). A total of 2 uL of β-ME mix was added to each reaction tube. The tubes were swirled gently, followed by incubation on ice for 10 minutes, with intermittent swirling every two minutes. Sequel to this, 5 uL of the Dpn I-treated rDNA was transferred into each tube with the ultracompetent cells. The reaction was swirled gently to mix and incubated on ice for 30 minutes to condition cells for the subsequent heat shock.

Subsequently, the Falcon tubes were heat-pulsed at 42ºC in water bath for 30 seconds. The duration and temperature for this step were critical for transformation efficiency and were not exceeded. The Falcon tubes were incubated on ice for 2 minutes. A total of 500 uL of pre-heated SOC medium (QuickGene Invitrogn, Holliston, USA) was pipetted into each reaction tube, followed by incubation at 37ºC and 225-250 rpm on orbital shaker for 1 hour. The SOC medium was preheated to 42ºC in a water bath. Subsequently, each reaction was plated on agar plates that contained ampicillin (Plate 3.3k); this allowed only cells harbouring the mutant plasmid to grow. Each Falcon tube had 500 uL volume of reaction which was plated on 3 agar plates; one plate was plated with 100 uL while the other two agar plates were plated with 200 uL each of reaction. Spatula was used to evenly spread the transformed cells on the agar plates. The agar plates with transformed cells were incubated in an incubator at 37ºC for over 16 hours. Growth of transformed cells and extraction of mutant rDNA: After 16 hours of incubation (Plate 3.3l), a colony of transformed cells was picked by a spatula from each agar plate and was grown in 10 mL of LB broth in a conical tube. A total of 100 uL of ampicillin at a concentration of 10 mg/mL was pipetted into the conical tube to act as a selectable marker. This reaction was carried out at 37ºC and 225-250 rpm in an orbital shaker and was left overnight. Mutant rDNA was subsequently extracted from the transformed cells, using the QIAprep® Spin Miniprep Kit. The concentration of the mutant rDNA was detected by Qubit Assay. The rDNA was sent to Functional Biosciences for direct DNA sequencing. This was carried out to ensure that the mutations had been introduced into the correct location or coordinate on the IRF6 cDNA.

3.8.5 In vitro transcription of mRNA from rDNA

The mMESSAGE mMACHINE® SP6 Transcription Kit (Life Technologies, USA) was used to transcribe mutant mRNA from the rDNA. The mRNA was then injected into zebrafish embryos.

The mutant rDNA was cut open to linearize it by employing the following reaction in an Eppendorf tube – 10 µL of plasmid (mutant rDNA), 2.5 µL of enzyme buffer, 2.5 µL of bovine serum albumin (BSA), 9 µL of dH_2O and 1 µL of EcoR I enzyme, all adding up to a total volume of 25 uL. The reaction was incubated in water bath at 37^oC overnight. Subsequently, gel electrophoresis was run, as described earlier, to confirm the successful cutting of the rDNA, using uncut rDNA as control.

Histones were removed from the linearized rDNA by setting up the following reaction in an Eppendorf tube: 1 µL of 1 mg/mL Prot K, 2.5 µL of 10% SDS (Sodium dodecyl sulfate) and 22.5 µL of dH_2O, resulting in a total reaction volume of 50 µL. The reaction was incubated in water at 50^oC for 30 minutes.

Mutant, linearized rDNA was purified further. A total of 50 uL of dH_2O, 300 uL of Buffer QG and 100 uL of isopropanol were added to the Eppendorf tube that contained the 50 uL reaction volume. The reaction was transferred into a spin column. The Eppendoff tube was rinsed with 500 uL of Buffer QG to pick up any remaining DNA, which was also added to the spin column. The spin column was spun for 1 minute and the filtrate was pipetted out. Sequel to this, 500 uL of Buffer QG was added again and the spin column was spun for 1 minute, followed by pipetting out of the filtrate. Then 750 µL of Buffer PE was slowly added to the spin column and was spun for 1 minute. A pippet was then used to remove the filtrate and the spin column was spun again for 1 minute. The spin column was subsequently placed in 1.5 mL Eppendorf tube. A total of 20 µL of dH_2O was added to the center of the spin column and was allowed to sit for 1 minute. The reaction was then

spun for 1 minute with the lid of the Eppendorf tube dragging behind the direction of spin in the centrifuge. The purified, cut or linearized rDNA was collected into the Eppendorf tube.

mRNA was transcribed from the cut rDNA by setting the following reaction in a 1.5 mL Eppendorf tube using the "mMessage mMachine Kit" for SP6 promoter: 4 µL of dH_2O, 10 uL of 2X NTP/CAP, 2 uL of 10X buffer, 2 uL of cut rDNA template and 2 uL of Enzyme mix, all adding up to a total volume of 20 uL. The reaction was set in water bath at 37^oC overnight. Subsequently, 1 µL of reaction was pulled into a capped PCR tube with load dye, followed by running of gel electrophoreses, as described above, to assess the success of the transcription. One µL of Turbo DNAse from the mMessage mMachine Kit was added to the main reaction in the 1.5 mL Eppendorf tube and was left to sit for 15 minutes. This was meant to degrade any DNA in the reaction tube.

The mRNA was purified by the RNeasy kit (Life Technologies, USA). The volume of the reaction in the Eppendorf tube was brought to 100 uL by adding 80 uL of Gibco water to the 20 uL reaction. A total of 350 uL of RLT buffer was added to the 100 uL of RNA and was mixed by inverting tube up and down. Then 250 uL of ethanol was added, followed by inverting of tube to mix it content. The 700 uL reaction was transferred into a spin column and was spun at 10,000 rpm for 15 seconds. The flowthrough was discarded. Subsequently, 500 uL of RPE buffer was added, followed by spinning for 15 seconds at 10,000 rpm and discarding of the flow-through. Again, 500 uL of RPE buffer was added to the spin column, spun for 2 minutes and the flowthrough discarded. The spin column was then placed in an empty Eppendorf tube and was spun at 10,000 rpm for 1 minute to remove excess ethanol. The flow-through was discarded.

Corollary to these, the spin column was placed in a new Eppendorf tube, followed by addition of 30 uL of RNase free water to the column and was spun at 10,000 rpm for one minute to collect the purified mRNA. The concentration of the mRNA was quantified using the Qubit Assay. However, the RNA probes and buffer were used, followed by selection of single stranded RNA (ssRNA) on the Quite Fluorometer.

3.8.6 Breeding of zebrafish and injection of mRNA into embryos

Collection of wildtype zebrafish embryos: Wildtype zebrafish, Danio rerio, were obtained from the Rob Cornell laboratory and were stored in darkness at room temperature, with a shield separating the sexes. Usually 2 males to 2 females of zebrafish were hosted in every tank. To get wildtype embryos, the fish tank was set in light on the day of embryo collection. The shield separating the sexes was removed. The fish were allowed to breed for over 1 hour (Plate 3.4a). A sieve was placed near the bottom of the fish tank to prevent the fish from getting access to the embryos that fell to the bottom of the tank. This prevented the zebrafish from eating its own embryos.

Subsequently, the fish were transferred into different tank, followed by sieving of water that was in the original tank to get the wildtype embryos. The wildtype embryos were further washed clean with the provided fish water and were transferred into a Petri dish with fish water (Plate 3.4b). The wildtype embryos were loaded into troughs that had been created on agar in Petri dishes (Plate 3.4c). This was done with the aid of light microscope, a spatula and suction pipette. Each trough was loaded with about 40 embryos. This agar plate also contained fish water.

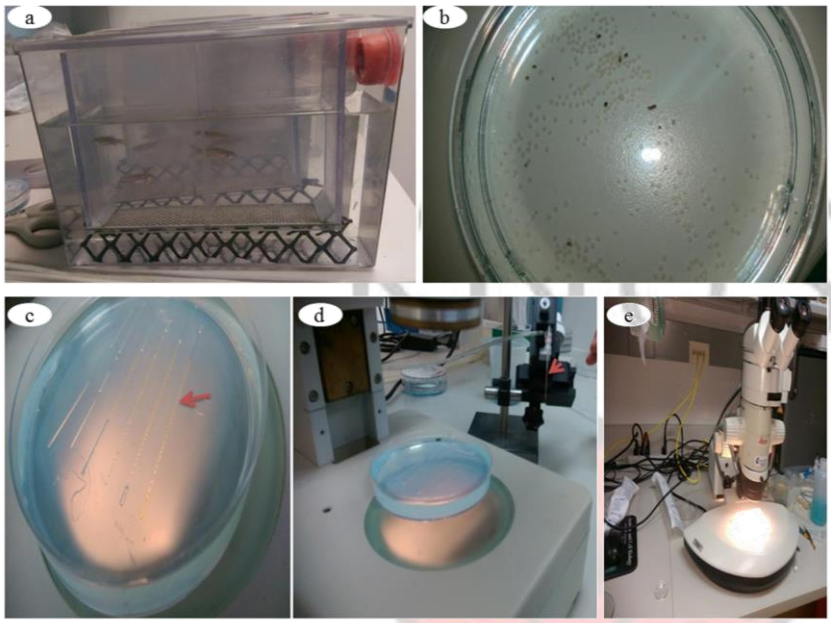


Plate 3.4: Zebrafish breeding and mRNA injection. a: breeding of zebrafish to get embryos. b: collected zebrafish embryos. c: Embryos loaded into troughs, indicated by red arrowhead. d: mutant mRNA was loaded into microinjection needle (pointed by red arrowhead), which was subsequently used to inject the embryos. e: embryos were knocked-down after seven days and were viewed under a fluorescence microscope.

Injection of mutant mRNA and observation of phenotype: The mRNAs were diluted to three different concentrations with low-TE buffer: 60 ng, 100 ng and 200 ng. This was done for wildtype IRF6 mRNA which served as the control as well as the mRNA for the two mutants, p.Glu69Lys and p.Gly65Val. A total of 40 uL dilution volume of the mRNAs was used. Sequel to this, 10 uL of Phenol Red dye with 0.2X KCl_2 (QuickGene Invitrogn, Holliston, USA) was added to each volume. The dye imparted colour and helped to visualize the successful injection of mRNA into the embryos. The samples were loaded into microinjection needles and the needles were inserted into the MPPI-2 pressure injector (Plate 3.4d). The tip of the microinjection needles was then broken off. The needle was slowly lowered to each embryo at a time, with the aid of a light microscope. When the needle was between the one cell of the embryo and the yolk, the foot peddle of the MPPI-12 pressure injector was pressed to release the mRNA into the wildtype embryo.

Subsequently, spatula was used to dislodge the wildtype embryos from the troughs, followed by transfer of the embryos into correspondingly labeled Petri dishes, taking note of the type and concentration of mRNA that was injected into every set of embryos. These new Petri dishes contained fish water as well. Each set of embryo was incubated in an oven at 32.6^{o}C. The wildtype embryos were observed at various time intervals, starting from 6 hours after incubation, to check for the phenotype. The embryos were observed at 6 hours, 24 hours, 48 hours and 7 days. Embryos were classified as abnormal if they exhibited any of the following phenotypes: head pointing upward instead of downward, missing eyes, ruptured embryos, defective arches, yolk bolus, heart oedema and other major gastrulation problems such as missing or shortened tails and malformed bodies. Embryos were classified as dead if they ruptured through the animal pole, which was indicative of disruption of periderm development (PeyrardJanvid et al., 2014).

The wildtype embryos, now fingerlings by 7 days, were knocked-down in Eppendorf tubes by adding Alcian Blue fluorescent dye that stained cartilages in the embryos. The Eppendorf tubes were placed on an orbital shaker for 30 minutes to uniformly mix and stain the cartilages. The fingerlings were then observed under a fluorescent microscope and pictures of them were taken (Plate 3.4e).

CHAPTER FOUR

4.0: RESULTS

4.1: Subjects for the study

A total of 1,554 Ghanaian participants were recruited for the research. These participants came from 875 families or pedigrees: 467 case pedigrees and 408 control pedigrees. The case pedigrees consisted of 413 NSOFC case pedigrees and 54 syndromic case pedigrees. About 5% (n=467) of the case pedigrees were multiplex families while the remaining

pedigrees were simplex families. A multiplex family refers to a family where there is more than one individual with cleft whereas a simplex family designates a family with only one individual with cleft. In 10 out of the 24 multiplex families, clefts occurred only in the nuclear family whereas in the remaining multiplex families, clefts occurred in other extended family members. Figure 4.1 shows three case pedigrees, panels A and B being multiplex pedigrees and panel C being a simplex pedigree. In pedigree A (GH20134972), all females in the family are affected with NSCL/P whereas neither of the males nor their parents have cleft. For pedigree B (GH20130786), the mother has NSCLP which was transmitted to a male child while the other two male siblings show no clinical cleft. For the simplex family C (GH20140578), the syndromic proband of monozygotic twins, has cleft and club foot whereas neither of the parents nor siblings show any of these congenital malformations. In total, 12 twins that were born alive were observed in this study; apart from the twin probands that had clefts, none of the other twins had cleft or any other congenital anomaly. The age range of probands was 1 week to 75 years; about 95% of probands were below 7 years.

Varied number of family members participated in the study by contributing samples. The 413 NSOFC case pedigrees had 1,006 participants: 413 case probands and 593 case relatives. These comprised 134 case-parent triads, 191 maternal dyads, 17 paternal dyads, 14 case-mother-sibling triads, 2 case-father-sibling triads, 23 casemother-maternal grandmother triads and 7 case-father-paternal grandmother triads. Other combination of participants included 1 case- father-maternal grandmother triad, 1 case-parents-maternal grandmother tetrad, 1 case-mother-sibling-maternal grandmother tetrad, 3 case-siblings triads, 3 case-maternal grandmother dyads, 9 case-siblings dyads, 1 case-parents-sibling tetrad, 1 case-parents-siblings pentad, 1 case-mother-maternal grandfather triad, 1 case-

mother-paternal grandmother triads, 2 case-paternal grandfather dyads and 3 singleton probands. The 54 syndromic pedigrees had 140 participants: 23 case-parents triads, 22 maternal dyads, 3 case-mother-maternal grandmother triads, 1 case-father-paternal grandmother triad, 3 case-mother-sibling triads, 1 case-mother-

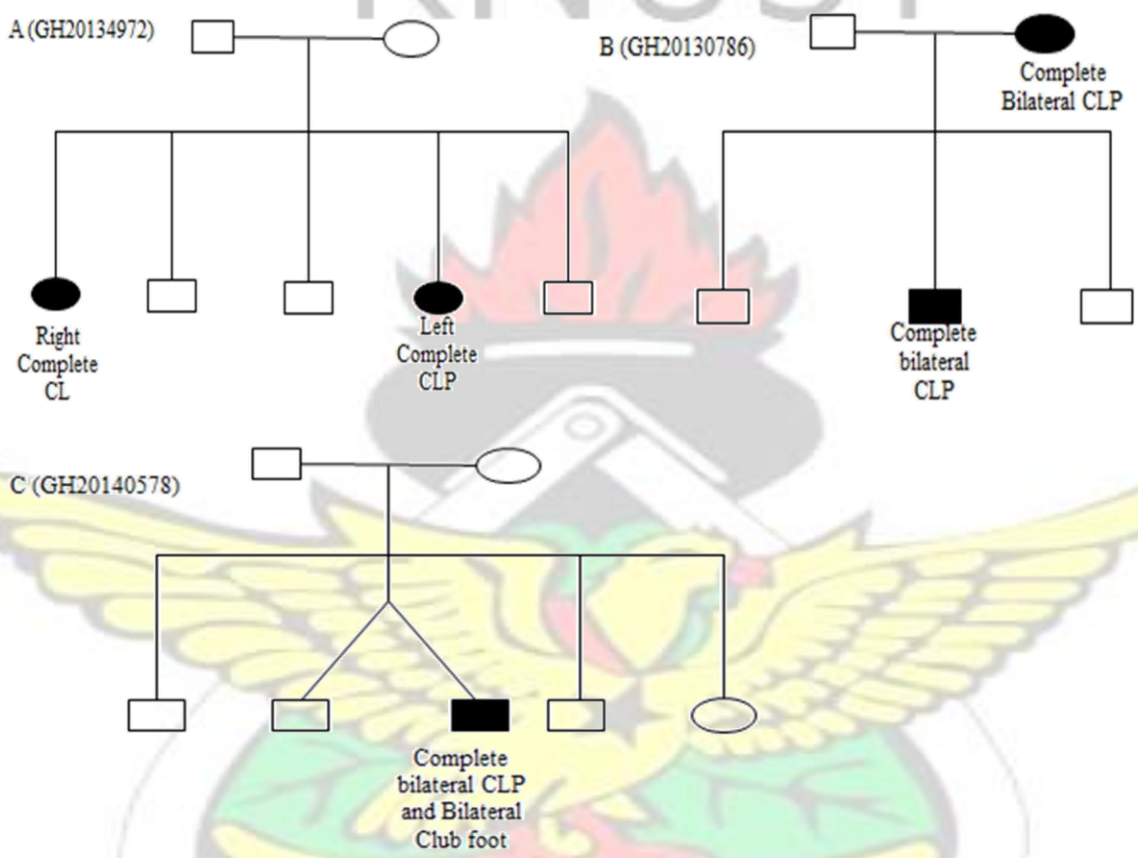


Figure 4.1: Multiplex (A and B) and simplex (C) case pedigrees

paternal grandmother triad and 1case-siblings triad. The control pedigrees were solely made up 408 singletons.

4.2 Phenomics of orofacial clefts in Ghana

Diverse cleft phenotypes were observed in the study population (Table 4.1). These classifications were based on affected part of the face, the severity of the cleft as well as the laterality. Detailed classifications that grouped clefts into complete and incomplete phenotypes are shown in Appendix N. All p-values reported under this section were

generated by chi-square test. There was significant difference (p=0.0073) between the number of males (n=213) and females (n=254) affected with clefts in the study population, with higher number of females being affected. This effect emanated from the 1:2 ratio of males (n=43) to females (n=83) affected with CP (p<0.0001), since there was no significant difference between number of males and females affected with CL (p=0.5865), CLP (p=0.4996) and special clefts (p=0.9999). The CL subphenotype consisted of 87 males and 82 females whereas the CLP subphenotype consisted of 76 males and 82 females. Equal number, 7, of males and females were observed for the special clefts subphenotype. There was also significant difference in laterality for CL (p<0.0001) as well as CLP (p=0.0119), with left unilateral clefts being more common. Based on clinical evidence, about 88% of all clefts were nonsyndromic whereas only about 12% were syndromic.

The syndromic forms of clefts (Table 4.2) observed in the study population were based on clinical evidence as stated in Online Mendelian Inheritance in Man (OMIM:www.ncbi.nlm.nih.gov/omim) but not molecular evidence. Syndromes were assigned to some probands based on clinical evidence whereas for some probands, only

Table 4.1: Diversity of orofacial clefts classified based on severity and laterality

Type of Cleft	Nonsyndromic Clefts			Syndromic Clefts			Total A+B	Subphenotype (Total A+B) as a percentage of total clefts
	Male	Female	Total A	Male	Female	Total B		
Cleft Lip only (CL)								
Right Unilateral CL	20	23	43	5	0	5	48	10.28%
Left Unilateral CL	46	44	90	2	1	3	93	19.91%
Bilateral CL	7	8	15	1	0	1	16	3.43%
Midline CL	1	1	2	0	1	1	3	0.64%
Macrostomia	5	4	9	0	0	0	9	1.92%
Totals of CL	79	80	159	8	2	10	169	36.19%

Cleft Palate (CP)								
CP only	21	60	81	8	13	21	102	21.84%
Submucous CP (CP + bifid uvula)	13	8	21	1	2	3	24	5.14%
Totals of CP	34	68	102	9	15	24	126	26.98%
Cleft Lip and Palate (CLP)								
Right CLP	22	17	39	2	0	2	41	8.78%
Left CLP	31	30	61	0	1	1	62	13.28%
Bilateral CLP	14	28	42	6	6	12	54	11.56%
Midline CLP	1	0	1	0	0	0	1	0.21%
Totals of CLP	68	75	143	8	7	15	158	33.83%
Special Clefts								
Oblique facial clefts or Tessier	3	1	4	1	1	2	6	1.28%
Ankyloglossia (tongue-tie)	1	0	1	0	1	1	2	0.43%
Midline Lower Lip CL	0	1	1	0	0	0	1	0.21%
Microform Clefts	1	2	3	1	1	2	5	1.07%
Totals	5	4	9	2	3	5	14	3.00%

phenotypes or symptoms were described since it was difficult assigning a particular syndrome to such probands based on clinical records and were just classified as having multiple congenital anomalies (MCAs). Nine cases of Pierre Robin Sequence (PRS) were observed, making this the most common syndromic cleft. However, it was difficult to classify these cases as a single syndrome, since there could be more than one underlying syndrome, such as Stickler and Treacher Collins Syndromes, causing PRS Table 4.2: Syndromes associated with OFCs in the study population

Name of Syndrome	Males	Females	Total	A Syndromic form as a percentage of total number of syndromic forms
Pierre Robin Sequence (PRS)	2	7	9	16.67%
Van der Woude Syndrome (VWS)	4	2	6	11.11%

Cleft-club foot only	5	1	6	11.11%
Goldenhar Syndrome	2	3	5	9.26%
Mobius Syndrome	1	1	2	3.70%
DiGeorge Syndrome	1	0	1	1.85%
Fragile-X Syndrome	0	1	1	1.85%
Edward Syndrome	1	0	1	1.85%
Opitz Syndrome	1	0	1	1.85%
Apert Syndrome	1	0	1	1.85%
Stickler Syndrome	0	1	1	1.85%
Cri du Chat Syndrome	0	1	1	1.85%
Amniotic Band Syndrome	0	1	1	1.85%
Kabuki Syndrome	1	0	1	1.85%
Holoprosencephalon	1	0	1	1.85%
Other multiple congenital anomalies (MCAs)	7	9	16	29.62%
Total	27	27	54	

(OMIM). PRS was characterized by micrognathia (very small mandibles), glossoptosis (distally placed tongue leading to blockage of pharyngeal air passage) and bifid uvula with soft palate cleft (SMCP). The most well-characterized syndromic form of clefts observed was Van der Woude Syndrome (VWS) (Table 4.2). Diagnosis of VWS was based on the presence of OFCs together with lower lip pits, being bilateral or unilateral lip pits.

Many patients with cleft who had other multiple congenital anomalies (MCAs) were also observed. Of the other probands classified as having just MCAs, clefts that presented with club foot only was the most common subphenotype; 6 of such MCAs were observed. Eight other MCAs presented with a spectrum of abnormalities that included overt OFCs, hexadactyly, microphthalmia, anophthalmia, microform clefts, ankyloglossia, talipes equinovarus (club foot), penoscrotal webbing, undescended testes, hypospadias, ocular hypertelorism, wide low-set ears, malformed ear lobe and coloboma. This spectrum of

phenotypes is characteristic of pterygial syndromes, such as popliteal pterygium syndrome, PPS (OMIM). Three other probands had clefts with other anomalies such as frontonasal dysplasia, microcephaly, developmental delays, toe aplasia and hypoplasia as well as hyperdontia. Moreover, cleft with hole-in-heart, stunted growth, motor dysfunction, deafness, dumbness, bent or defective vertebral column as well as seizures was observed in a female proband. Two other individuals also had clefts that presented with a number of developmental delays: severe mental retardation, learning disability, language disability, other global developmental delays, microcephaly, ocular hypertelorism, low-set ears as well as hearing and visual impairments. A female case proband also had cleft together with imperforate anus, which led to faecal discharge through the vagina. Finally, an individual also had cleft that presented with ptosis and proptosis of eyes, club foot as well as choanal atresia.

4.3 Demographics and socio-economic status of Ghanaian case families

Case pedigrees or families had similar demographics. About 66% of families with orofacial clefts lived in the rural areas whereas the remaining 34% lived in the urban areas. Even for urban dwellers, most families lived in "zongos" (enclaves predominantly inhabited by Moslems), slums and other low-standard areas. About 66.27% of the families recruited in this study resided in the Ashanti Region probably due to the geographical location of KATH Cleft Clinic, 7.95% in the Western Region, 6.75% in the Volta Region, 6.02% in the Brong Ahafo Region, 5.78% in the Eastern Region, with the remaining families residing in the Central, Greater, Northern, Upper East and Upper West Regions. However, clefts seemed to be endemic at specific locations in these regions: Manso and Asante Akyem Agogo areas of the Ashanti Region, Begoro area of the Eastern region,

Sefwi area of the Western Region and Dunkwa-On-Offin area of the Central Region. Clefts also occured in almost all ethnic groups: Ewes = 8.19%,

Brongs and Ahafo = 4.58%, Asantes = 45.54%, Fante and other tribes of Central Region = 7.71%, Sefwi and other tribes of Western Region = 6.99% and tribes of the Northern, Upper East and Upper West Regions = 19.27%. Therefore, human OFCs may occur in all tribes of Ghana, though they were endemic at certain rural settings of various regions where poverty and low level of education were high.

Case families were also characterized by low socio-economic status and youthful parents. Majority of mothers (90%) and fathers (75%) never went to school or ended their education at the elementary level (Figure 4.2). Most of these parents could neither read nor write. Since illiteracy level in Ghana is 26% (2010 Census), illiteracy was significantly ($p<0.0001$) associated with families with clefts. Furthermore, about 87% of mothers and 70% of fathers gave birth to the probands when they were within the age range of 21 to 40 years (Figure 4.3). Thus, clefts patients were mostly born to active, youthful, reproducing parents. About 95% of both mothers and fathers were low income

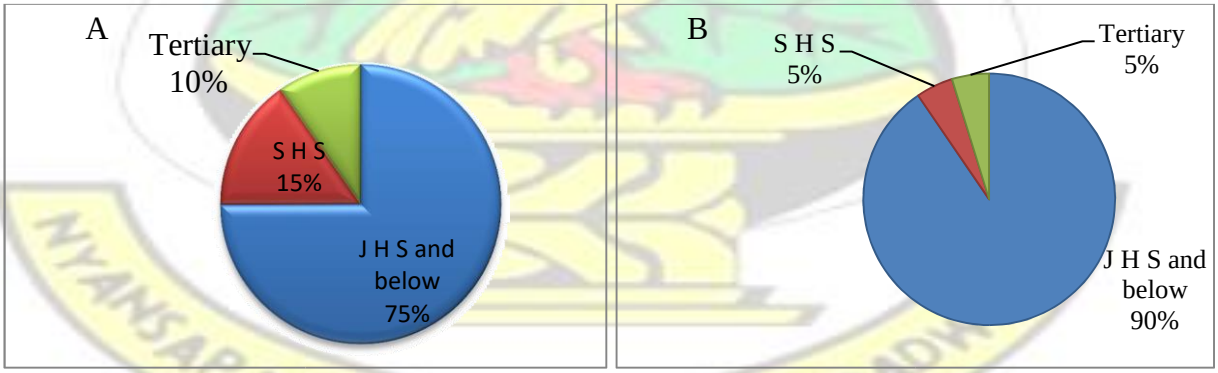


Figure 4.2: Level of education of parents. A: fathers, B: mothers.

earners (Figure 4.4). Thus, only about 5% of families with clefts had reliable source of income of at least GH₵500.00 per month. High income earners included a doctor,

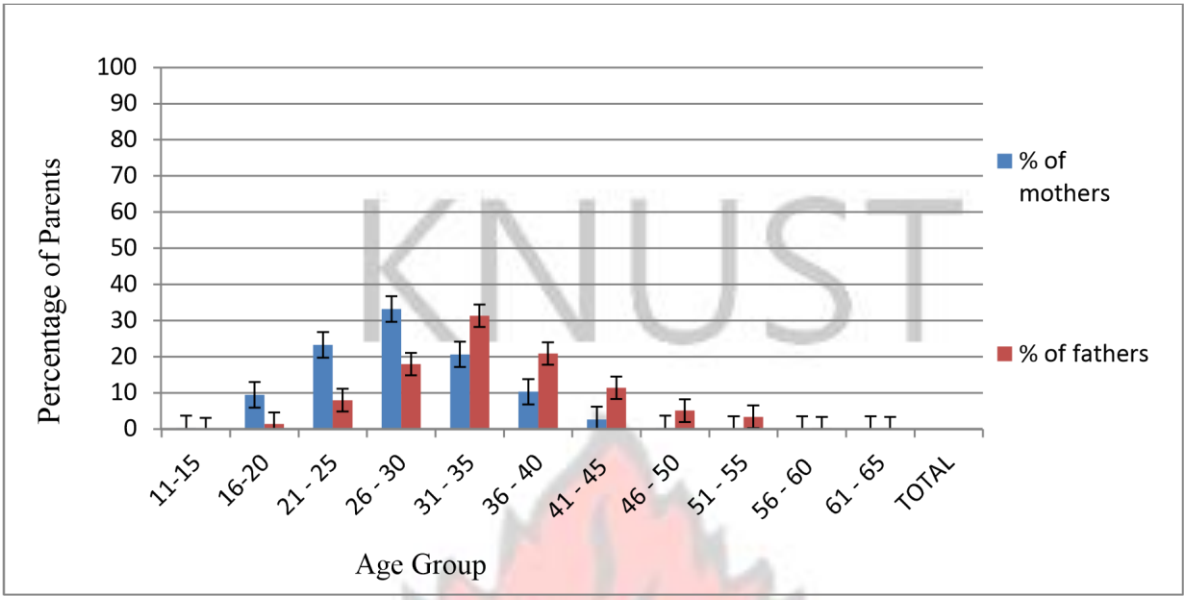


Figure 4.3: Age distribution of parents at the time of proband's birt h

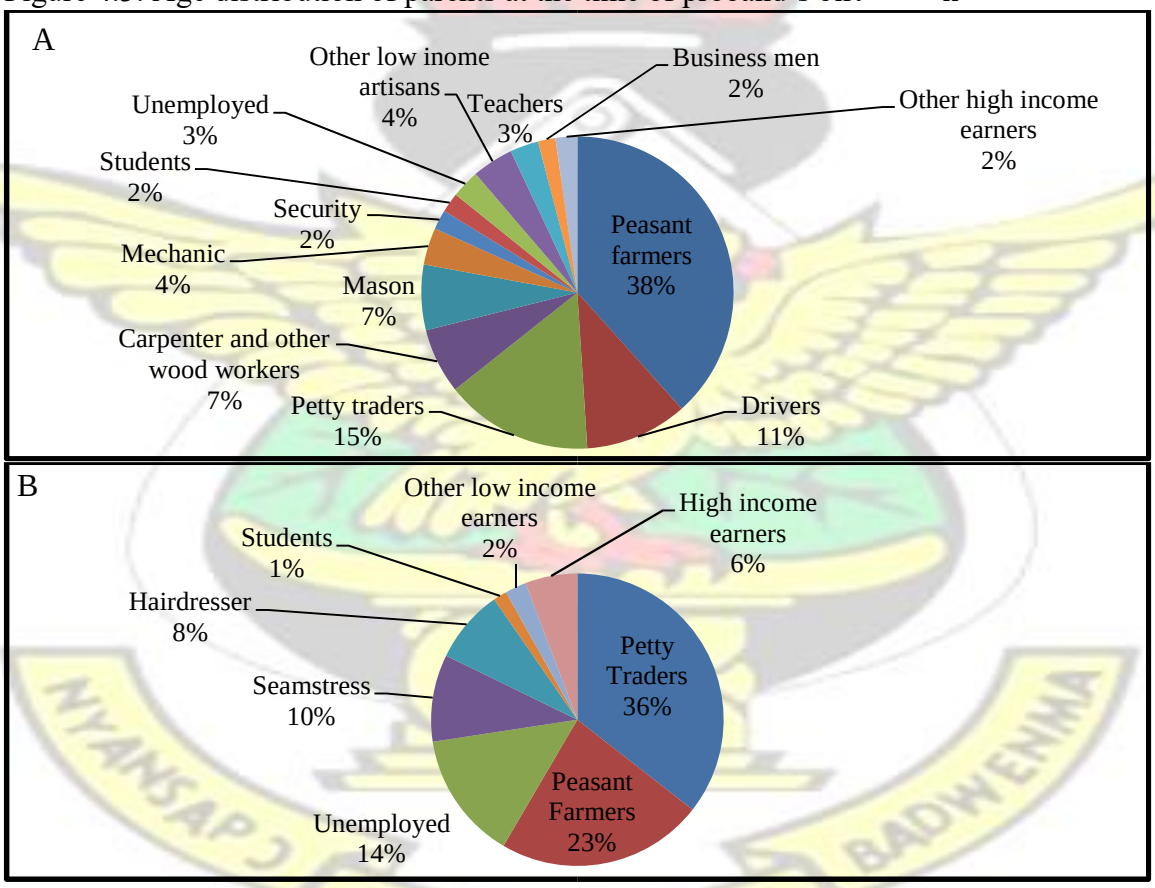


Figure 4.4: Occupation or economic status of case families. A: fathers, B: mothers.

teachers, civil servants, public workers as well as traders who owned permanent stores and were making reliable sales every month.

4.4: Exposomics of OFCs: environmental and other risk factors

Some mothers in case families were exposed to diverse environmental and other risk factors (Figure 4.5). A very significant ($p<0.0001$) percentage (96%) of mothers had poor dietary folate intake. This assessment was based on lack of consumption of slightly cooked or uncooked green leafy vegetables, folate-fortified cereals, fruits, etc. Most

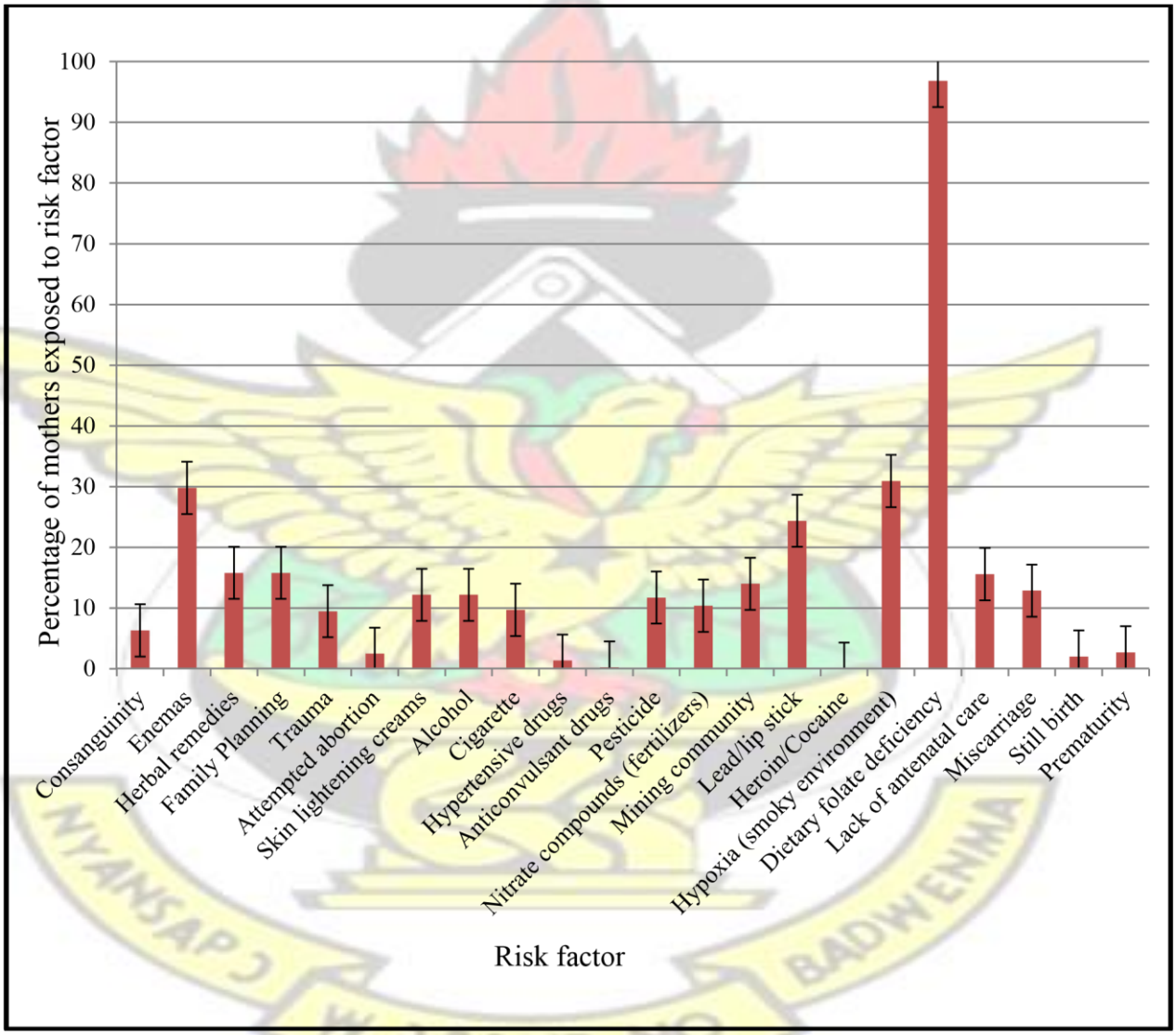


Figure 4.5: Maternal exposure to risk factors

mothers considered eating of folate-rich foods as the privilege of the rich. Moreover, about 90% of mothers hardly ate any folate-fortified food, such as fortified cereals.

Furthermore, only about 14% of mothers had folate supplementation within the first trimester whereas about 70% of mothers took folic acid supplements after the first trimester. About 16% of mothers never took any folate supplements during gestation. In all families, folate supplementation was initiated during the first visit to antenatal clinic. Over 85% of mothers either accessed antenatal care after 3 months of gestation or did not access antenatal care at all (Figure 4.6).

Enema usage was also common among some mothers, with about 30% of mothers using enemas during conception. Also worthy of note is the high number of mothers (31%) that were consistently exposed to smoke from sources such as firewood, refuse dumps and factories.

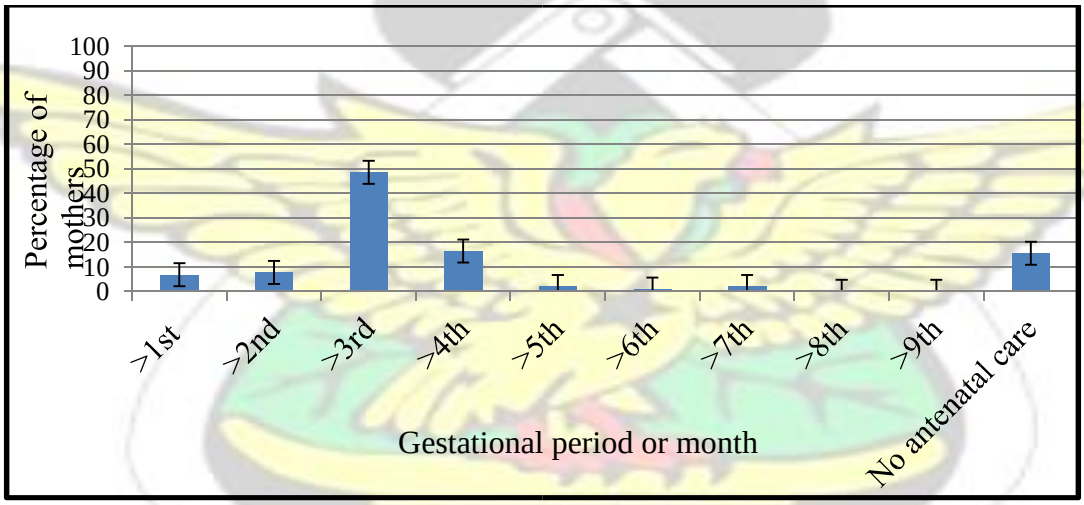


Figure 4.6: Number of months after conception after which mother got antenatal care. > implies after the indicated month.

4.5: DNA yield, SNP call rates and association studies

DNA yield ranged from 0.01 ng/uL to 890 ng/uL. Only 10 saliva samples had DNA yield of 0.01 ng/uL but were re-concentrated to 50 ng/uL using SpeedVac machine. The average saliva DNA yield was about 100 ng/uL, with some saliva samples yielding as high as 890 ng/uL. Furthermore, the average DNA yield for cheek swabs was about 50

ng/uL, though some cheek swabs gave a yield as high as 372 ng/uL. No sample failed the quality control checks that were done through XY Genotyping. Individual genotypes were represented as scatter plots (Figure 4.7). The average call rate for all 48 markers was 99.4% (Table 3.1). Call rate refers to the percentage of total samples that was accurately genotyped for a marker.

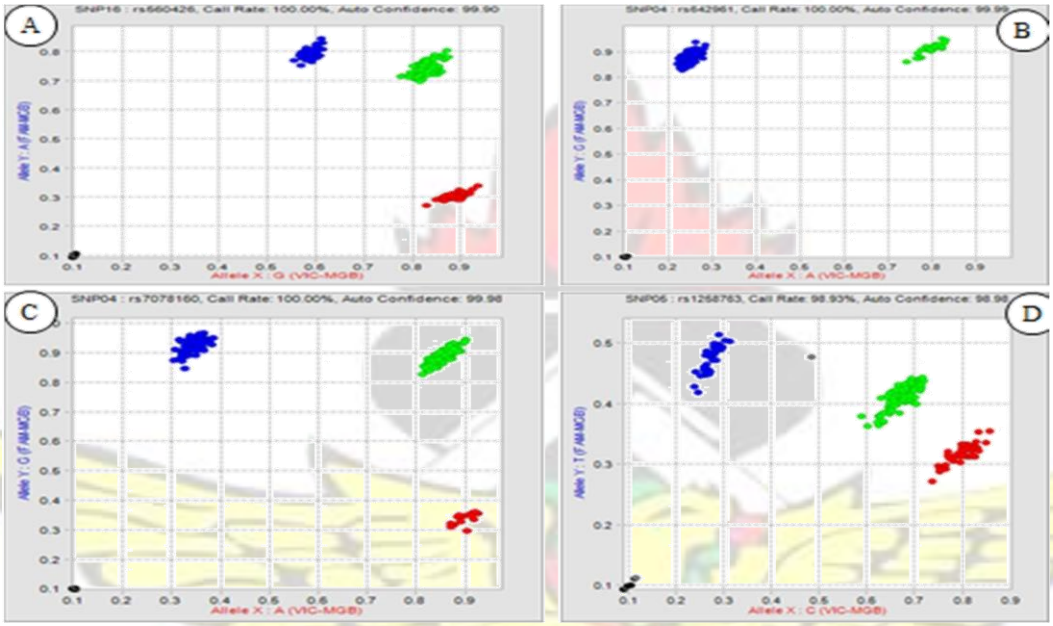


Figure 4.7: Fluidigm Scatter Plot for four SNPs that constantly showed association with NSOFCs in family-based association analyses. A: rs560426 of ABCA4, B: rs642961 of IRF6, C: rs7078160 of VAX1 and D: rs1258763 of GREM1. Red indicates the homozygous allele that was attached to the VIC dye whereas blue connotes the homozygous allele that was attached to the FAM dye. Green represents the heterozygous conditions for the alleles for a particular SNP.

4.5.1 Genomics of OFCs: association analyses

All results presented showed SNPs that had a $p<0.10$ in at least one of the analyses done, except the case-control meta-analyses for Africa (Tables 4.4 and 4.5) where all SNPs that passed the analysis pipeline are shown. All SNPs are shown in

Tables 4.4 and 4.5 because additional information on minor as well as risk alleles and

test of heterogeneity for the genotyped SNPs are shown. All p-values indicated are for the minor alleles in Africans. Bonferroni Correction of $p<3.54\times10^{-4}$ for formal significance was applied to the most inclusive case-control association studies only. For TDT and DFAM, many loci demonstrated formal significance with NSOFCs at $p\leq0.05$. Formal significance for TDT, FBAT and DFAM was evaluated at $p\leq0.05$ because these were secondary analyses compared with case-control analyses, and are not true independent tests.

Very interesting results emanated from the case-control association studies (Tables 4.3, 4.4 and 4.5). ABCA4 locus (rs560426, p=0.03, OR=1.23) and VAX1 (rs7078160, p=0.03, OR=1.25) demonstrated evidence of nominal association with NSCL/P in the Ghanaian cohort, with the ABCA4 signal being stronger for NSCLP (p=0.01, OR=1.39). Moreover, VAX1 (rs7078160, p=0.04, OR=1.16) and ABCA4 (rs560426, p=0.03, OR=1.20) also showed evidence of nominal association with NSCL/P and NSCLP, respectively, in the African cohort. PAX7 (rs742071, $p=5.10\times10^{-3}$, OR=1.19) also gained nominal association with NSCL/P, with subphenotype analyses suggesting the NSCL subphenotype exhibited some level of heterogeneity. The 8q24 locus was also nominally associated (rs987525, $p=1.22\times10^{-3}$, OR=0.81) with NSCL/P in Africans, with the minor C allele being protective, suggesting the major A allele is also the risk allele for the combined NSCL/P phenotype in Africans. This notwithstanding, subphenotype analyses of the NSCL/P cohort (Table 4.5) showed that either of C or A allele may be a risk allele in Africans, depending on the subphenotype: NSCL (rs987525, $p=5.38\times10^{-3}$, OR=1.28) and NSCLP (rs987525, p=0.01, OR=0.80). Subphenotype analyses also showed that MSX1 (rs115200552, p=0.01, OR=1.81),

TULP4 (rs651333, p=0.04, OR=1.29), CRISPLD2 (rs4783099, p=0.02, OR=0.74) and

Table 4.3: Case-control association studies for Ghana

SNP	Tagging Gene or loci	NSCL/P		NSCL		NSCLP		NSCP		NSOFC	
		pvalue	OR	pvalue	OR	pvalue	OR	pvalue	OR	pvalue	OR
rs1801131	MTHFR	0.93	1.01	0.51	0.88	0.41	1.17	0.40	0.81	0.96	0.99
rs766325	PAX7	0.82	0.97	0.61	1.09	0.30	0.82	0.65	0.91	0.65	0.95
rs742071	PAX7	0.81	1.02	0.82	0.97	0.48	1.10	0.88	0.98	0.84	1.02
rs560426	ABCA4	0.03	1.23	0.35	1.13	0.01	1.39	0.16	1.24	0.02	1.22
rs6677101	SLC25A24	0.33	1.10	0.98	1.00	0.09	1.26	0.27	1.19	0.20	1.12
rs861020	IRF6	0.89	0.98	0.71	1.08	0.62	0.90	0.27	0.73	0.66	0.94
rs34743335	IRF6	0.93	1.04	0.89	1.08	0.87	0.90	0.52	0.52	0.93	1.04
rs642961	IRF6	0.48	0.89	0.89	0.97	0.39	0.81	0.13	0.60	0.28	0.84
rs7590268	THADA	0.80	1.03	0.86	0.97	0.25	1.20	0.26	0.79	0.68	0.96
rs4332945	DYSF	0.89	1.02	0.86	1.03	0.82	1.04	0.70	0.92	0.87	0.98
rs2303596	DYSF	1.00	1.00	0.51	1.11	0.69	0.94	0.05	1.40	0.48	1.07
rs115200552	MSX1	0.24	1.28	0.29	1.34	0.39	1.27	0.06	1.72	0.11	1.33
rs2674394	Gene Desert	0.49	1.09	0.07	1.34	0.46	0.87	0.75	1.07	0.64	1.05
rs6558002	EPHX2	0.54	1.07	0.71	0.94	0.14	1.25	0.67	0.92	0.56	1.06
rs987525	8q24	0.11	0.85	0.06	0.78	0.56	0.92	0.80	0.96	0.12	0.87
rs7078160	VAX1	0.03	1.25	0.04	1.32	0.17	1.22	0.59	1.10	0.04	1.22
rs4752028	VAX1	0.12	0.86	0.11	0.82	0.27	0.87	0.54	0.91	0.20	0.90
rs8001641	SPRY2	0.78	1.04	0.64	1.10	0.75	0.93	0.55	0.86	0.97	1.01
rs1258763	GREM1	0.56	1.06	0.39	1.11	0.65	1.06	0.65	0.93	0.79	1.02
rs8049367	ADCY9	0.31	1.11	0.53	1.09	0.49	1.10	0.48	0.89	0.55	1.05
rs11642413	CDH1	0.86	0.98	0.88	0.98	1.00	1.00	0.24	0.80	0.61	0.95
rs1546124	CRISPLD2	0.87	1.02	0.53	0.90	0.30	1.18	0.44	1.15	0.84	1.02
rs4783099	CRISPLD2	0.43	1.08	0.80	1.03	0.17	1.20	0.28	0.84	0.90	1.01
rs8069536	NTN1	0.69	1.04	0.71	0.95	0.31	1.15	0.88	0.98	0.77	1.03
rs8081823	NTN1	0.45	0.92	0.77	0.96	0.33	0.86	0.73	1.06	0.68	0.96
rs17760296	NOG	0.90	1.05	0.71	1.21	0.66	1.29	0.07	2.66	0.56	1.21
rs227731	NOG	0.45	1.09	0.65	1.07	0.53	1.11	0.25	1.23	0.25	1.12
rs3923086	AXIN2	0.36	2.00	0.15	2.87	0.40	0.00	0.33	0.00	0.83	1.16
rs17820943	MAFB	0.85	1.02	0.94	1.01	0.72	0.94	0.82	1.04	0.62	1.05
rs13041247	MAFB	0.67	1.05	0.82	1.04	0.99	1.00	0.58	1.11	0.45	1.08
rs11696257	MAFB	0.80	1.03	0.89	1.02	0.76	0.95	0.88	1.03	0.61	1.05

Note - NSOFC: combination of NSCL/P and NSCP, NSCL/P: nonsyndromic cleft lip with or without cleft palate, NSCL: nonsyndromic cleft lip, NSCLP: nonsyndromic cleft lip with palate, NSCP:

nonsyndromic cleft palate. The p-values that reached nominal significance are shown in bold characters whereas those that exhibited a trend towards nominal association are shown in italics.

Table 4.4: Case-control meta-analyses for combined African NSCL/P and NSCP cohorts

SNP	Probable gene/loci	Minor alleles[a]	African MAF	NSCL/P			NSCP		
				p	OR	I	p	OR	I
rs1801131	MTHFR	C/A[r]	0.15	0.32	1.08	0.00	0.19	0.79	0.00
rs1801133	MTHFR	A/G[p]	0.09	0.49	1.08	18.19	0.44	0.83	0.00
rs766325	PAX7	G/A[c,d,r]	0.18	0.29	0.92	0.00	0.23	0.82	0.00
rs742071	PAX7	T/G[r]	0.39	5.10E-03[f]	1.19	54.68	0.76	0.96	0.00
rs560426	ABCA4	C/T[r]	0.49	0.10	0.90	6.15	0.16	1.18	0.00
rs481931	ABCA4	T/G[p]	0.10	0.40	1.09	11.13	0.49	0.85	0.00
rs4147811	ABCA4	T/C[p]	0.11	0.23	1.13	67.35	0.93	1.02	0.00
rs138751793	ARHGAP29	C/T[e]	0.02	0.24	1.32	0.00	0.47	1.34	27.90
rs6677101	SLC25A24	G/T[b,d,r]	0.33	0.80	0.98	12.11	0.87	1.02	53.89
rs861020	IRF6	A/G[r]	0.11	0.23	1.11	0.00	0.83	0.96	24.15
rs34743335	IRF6	T/A	0.02	0.59	0.90	0.00	0.84	0.89	38.34
rs642961	IRF6	A/G[r]	0.09	0.32	1.11	68.47	0.57	0.88	44.17
rs7590268	THADA	G/T[r]	0.20	0.74	0.98	0.00	0.38	0.87	0.00
rs4332945	DYSF	T/G[b,d,r]	0.16	0.94	0.99	0.00	0.97	1.01	0.00
rs2303596	DYSF	T/C[c,d,p]	0.22	0.20	0.91	75.32	0.57	1.09	73.54
rs227782	DYSF	A/G[b,r]	0.42	0.33	1.06	0.00	0.35	1.12	61.90
rs115200552	MSX1	C/G[e]	0.02	0.38	1.16	28.63	0.01[f]	1.81	0.00
rs12532	MSX1	G/A[d,p]	0.44	0.49	0.96	0.00	0.37	0.90	0.43
rs2674394	Gene Desert	A/C[r]	0.17	0.62	1.04	0.00	0.68	1.07	0.00
rs651333	TULP4	C/T[b,c,r]	0.34	0.97	1.00	0.00	0.04[f]	1.29	0.00
rs6558002	EPHX2	C/T[b,r]	0.24	0.39	1.06	0.00	0.87	1.02	0.00
rs987525	8q24	A/C[b,r]	0.38	1.22E-03[f]	0.81	40.55	0.22	0.86	0.00
rs894673	FOXE1	A/T[p]	0.33	0.42	0.95	0.00	0.93	1.01	0.00
rs3758249	FOXE1	T/C[p]	0.33	0.56	0.96	0.00	0.90	1.02	0.00
rs7078160	VAX1	A/G[r]	0.25	0.04[f]	1.16	0.00	0.88	1.02	0.00
rs4752028	VAX1	C/T[b,r]	0.45	0.51	0.96	0.00	0.80	0.97	0.00
rs10785430	ADAMTS20	G/A[r]	0.32	0.90	0.99	0.00	0.49	1.09	0.00
rs9574565	SPRY2	T/C[b,p]	0.35	0.75	1.02	0.00	0.45	1.10	0.00
rs8001641	SPRY2	G/A[b,c,d,p]	0.10	0.35	1.08	0.00	0.37	0.85	0.00
rs17563	BMP4	T/C[b,c,d,r]	0.18	0.95	0.99	0.00	0.77	1.04	0.00
rs1258763	GREM1	C/T[b,c,d,p]	0.49	0.11	1.11	0.00	0.50	0.92	0.00
rs8049367	ADCY9	C/T[c,d,p]	0.30	0.20	1.09	0.00	0.10	0.81	0.00

rs16260	CDH1	A/C[r]	0.13	0.59	1.05	0.00	0.39	0.85	0.00
rs11642413	CDH1	G/A[b,d,r]	0.28	0.83	1.02	0.00	0.21	0.83	0.00
rs1546124	CRISPLD2	G/C[d,r]	0.25	0.60	0.96	0.00	0.89	0.98	0.00
rs4783099	CRISPLD2	T/C[r]	0.33	0.59	1.04	0.00	0.02[f]	0.74	0.00
rs8069536	NTN1	T/G[r]	0.32	0.13	1.11	0.97	0.88	0.98	0.00
rs8081823	NTN1	A/G[p]	0.24	0.08	0.88	0.00	0.63	0.94	32.54

Table 4.4 continued…

SNP	Probable gene/loci	Minor alleles[a]	African MAF	NSCL/P			NSCP		
				p	OR	I	p	OR	I
rs17760296	NOG	G/T[r]	0.02	0.92	0.99	0.00	0.04[f]	1.74	0.00
rs227731	NOG	G/T[b,r]	0.22	0.86	0.99	0.00	0.26	1.17	0.00
rs7224837	AXIN2	G/A[r]	0.11	0.75	1.04	0.00	0.81	0.95	0.00
rs3923086	AXIN2	A/C[b,c,d,r]	0.02	0.25	1.15	0.00	NA	NA	NA
rs17820943	MAFB	T/C[p]	0.25	0.33	0.93	15.15	0.68	1.06	22.99
rs13041247	MAFB	C/T[p]	0.25	0.37	0.94	34.01	0.42	1.12	0.00
rs11696257	MAFB	T/C[p]	0.25	0.30	0.93	32.24	0.61	1.07	0.00

[a]The first allele is the minor allele in Europeans and unless otherwise indicated, the first allele is also the minor allele in Europeans, East Asians, South Asians and Africans, [b]the first allele is the major allele while the second allele is the minor allele in Africans, [c]the first allele is the major allele while the second allele is the minor allele in South Asians, [d]the first allele is the major allele while the second allele is the minor allele in East Asians, [e]first allele is the minor allele and the variation exists only in Africans, [f]loci that reached nominal significance in meta-analyses, [r]minor allele was the risk allele in initial study, [p]minor allele was protective in initial study, MAF: minor allele frequency, p: pvalues, OR: odds ratio, I: test of heterogeneity of which 0 to 40 represents no heterogeneity; NA: not applicable. All p-values reported are for the minor alleles. All initial studies were either carried out in Asians and/or Caucasians, but not Africans. Source of minor alleles and MAF is http://browser.1000genomes.org.

Table 4.5 Case-control meta-analyses of subphenotypes of NSCL/P for combined African cohort

SNP	Probable gene/loci	Minor alleles[a]	African MAF	NSCL			NSCLP		
				p	OR	I	p	OR	I
rs1801131	MTHFR	C/A[r]	0.15	0.78	1.03	0.00	0.22	1.13	0.00
rs1801133	MTHFR	A/G[p]	0.09	0.71	1.06	8.24	0.30	0.30	0.00
rs766325	PAX7	G/A[c,d,r]	0.18	0.91	0.99	0.00	0.17	0.86	0.00
rs742071	PAX7	T/G[r]	0.39	0.02[f]	1.23	68.74	0.03[f]	1.19	0.00

rs560426	ABCA4	C/T[r]	0.49	0.73	1.03	0.00	0.03[f]	1.20	10.33
rs481931	ABCA4	T/G[p]	0.10	0.81	0.97	0.00	0.08	1.27	63.75
rs4147811	ABCA4	T/C[p]	0.11	0.50	1.10	65.82	0.15	1.21	15.35
rs138751793	ARHGAP29	C/T[e]	0.02	0.19	1.53	66.38	0.41	1.29	0.00
rs6677101	SLC25A24	G/T[b,d,r]	0.33	0.92	0.99	0.00	0.98	1.00	58.97
rs861020	IRF6	A/G[r]	0.11	0.18	1.17	17.72	0.57	1.07	0.00
rs34743335	IRF6	T/A	0.02	0.87	0.96	0.00	0.50	0.85	23.72
rs642961	IRF6	A/G[r]	0.09	0.96	0.99	15.60	0.15	1.21	62.97
rs7590268	THADA	G/T[r]	0.20	0.45	0.92	0.00	0.50	1.07	0.00
rs4332945	DYSF	T/G[b,d,r]	0.16	0.54	0.94	10.40	0.71	1.04	0.00
rs2303596	DYSF	T/C[c,d,p]	0.22	0.29	0.89	63.58	0.44	0.93	75.54
rs227782	DYSF	A/G[b, r]	0.42	0.85	0.98	0.00	0.13	1.14	0.00
rs115200552	MSX1	C/G[e]	0.02	0.18	1.37	61.30	0.68	1.10	0.00

Table 4.5 continued…

SNP	Probable gene/loci	Minor alleles[a]	African MAF	NSCL			NSCLP		
				p	OR	I	p	OR	I
rs12532	MSX1	G/A[d,p]	0.44	0.55	0.95	0.00	0.51	0.95	0.00
rs2674394	Gene Desert	A/C[r]	0.17	0.06	1.22	0.00	0.42	0.91	0.00
rs651333	TULP4	C/T[b,c,r]	0.34	0.63	0.96	0.00	0.74	0.97	0.00
rs6558002	EPHX2	C/T[b,r]	0.24	0.82	1.02	0.00	0.11	0.11	0.00
rs987525	8q24	A/C[b,r]	0.38	5.38E-03[f]	1.28	0.00	0.01[f]	0.80	54.21
rs894673	FOXE1	A/T[p]	0.33	0.54	0.95	42.39	0.45	0.94	0.00
rs3758249	FOXE1	T/C[p]	0.33	0.53	0.94	46.73	0.68	0.96	0.00
rs7078160	VAX1	A/G[r]	0.25	0.03[f]	1.23	0.00	0.20	1.13	24.04
rs4752028	VAX1	C/T[b,r]	0.45	0.55	1.05	16.64	0.50	0.95	0.00
rs10785430	ADAMTS20	G/A[r]	0.32	0.88	1.01	41.30	0.86	0.98	3.00
rs9574565	SPRY2	T/C[b,p]	0.35	0.53	1.06	72.62	0.43	1.07	65.44
rs8001641	SPRY2	G/A[b,c,d,p]	0.10	0.99	1.00	0.00	0.26	1.13	0.00
rs17563	BMP4	A/G[b,c,d,r]	0.18	0.89	0.99	25.84	0.98	1.00	0.00
rs1258763	GREM1	C/T[b,c,d,p]	0.49	0.22	0.90	0.00	0.10	1.15	0.00

rs8049367	ADCY9	C/T[c,d,p]	0.30	0.36	1.09	10.19	0.35	1.08	0.00
rs16260	CDH1	A/C[r]	0.13	0.46	0.91	10.51	0.20	1.16	0.00
rs11642413	CDH1	G/A[b,d,r]	0.28	0.98	1.00	0.00	0.55	1.05	0.00
rs1546124	CRISPLD2	G/C[d,r]	0.25	0.26	0.90	0.00	0.88	1.01	0.00
rs4783099	CRISPLD2	T/C[r]	0.33	0.85	1.02	0.00	0.32	1.09	0.00
rs8069536	NTN1	T/G[r]	0.32	0.72	1.03	3.47	0.04[f]	1.20	0.00
rs8081823	NTN1	A/G[p]	0.24	0.55	0.95	0.00	0.05	0.83	0.00
rs17760296	NOG	G/T[r]	0.02	0.83	1.04	5.85	0.85	0.97	0.00
rs227731	NOG	G/T[b,r]	0.22	0.38	0.92	0.00	0.59	1.05	0.00
rs7224837	AXIN2	G/A[r]	0.11	0.61	1.08	0.00	0.81	1.04	0.00
rs3923086	AXIN2	A/C[b,c,d,r]	0.02	0.62	1.10	40.28	NA	NA	0.00
rs17820943	MAFB	T/C[p]	0.25	0.25	0.89	15.55	0.43	0.93	0.00
rs13041247	MAFB	C/T[p]	0.25	0.25	0.89	31.03	0.54	0.94	0.00
rs11696257	MAFB	T/C[p]	0.25	0.24	0.89	27.17	0.40	0.92	0.00

[a]The first allele is the minor allele in Europeans and unless otherwise indicated, the first allele is also the minor allele in Europeans, East Asians, South Asians and Africans, [b]the first allele is the major allele while the second allele is the minor allele in Africans, [c]the first allele is the major allele while the second allele is the minor allele in South Asians, [d]the first allele is the major allele while the second allele is the minor allele in East Asians, [e]first allele is the minor allele and the variation exists only in Africans, [f]loci that reached nominal significance in meta-analyses, [r]minor allele was the risk allele in initial study, [p]minor allele was protective in initial study, MAF: minor allele frequency, p: p-values, OR: odds ratio, I: test of heterogeneity of which 0 to 40 represents no heterogeneity; NA: not applicable. All p-values reported are for the minor alleles. All initial studies were either carried out in Asians and/or Caucasians, but not Africans. Source of minor alleles and MAF is http://browser.1000genomes.org.

NOG (rs17760296, p=0.04, OR=1.74) were nominally associated with NSCP in Africans. However, none of these case-control associations passed Bonferroni correction.

In TDT analyses (Table 4.6, 4.7, 4.8 and 4.9), 16 SNPs either demonstrated formal significance with NSOFCs or showed a trend towards association. IRF6 (rs642961) showed a trend towards association in the overall Ghanaian samples, but the same SNP showed association with NSCP (p=0.02, O/R=0.00). This unconventional or dissenting observation suggests the minor A allele of rs642961 is protective in Ghanaians whereas the G allele is the risk allele for NSCP. VAX1 (rs7078160) also interestingly demonstrated evidence of association with NSOFCs in both the Ghanaian and African cohorts, with the

conventional observation that the minor A allele is associated with NSOFCs. Another SNP of VAX1, rs4752028, was unconventionally associated with NSCP, with the T allele being the minor and risk allele in Africans (Table 4.4). ABCA4 (rs560426) showed a trend towards association in the Ghanaian cohort, with subphenotypes analyses suggesting this SNP was associated with NSCLP (Tables 4.7 and 4.9). Interestingly, MTHFR (rs1801131) showed association with NSCL in both Ghanaian and combined African cohorts (Tables 4.7 and 4.9), with the minor C allele being protective whereas the A allele was the risk allele for NSCL. Two SNPs of MAFB also showed a trend towards association, with the O/R=1.54 suggesting that the minor T allele of rs17820943 and C allele of rs13041247 are the risk alleles for NSOFCs in Africans. AXIN2 also showed a trend towards association with NSCP and NSCLP (Tables 4.8 and 4.9) whereas PAX7 (rs742071, p=0.03, OR=1.67) demonstrated evidence of threshold association with NSCL in the combined African cohort.

TDT results were corroborated using FBAT. Genes that showed either

Table 4.6: Transmission disequilibrium test (TDT) for Ghanaian NSCL/P and NSCP cohorts

SNP	Probable gene/loci	NSCL/P			NSCP			NSOFC		
		T/NT	p	OR (95% CI)	T/NT	p	OR (95% CI)	T/NT	p	OR (95% CI)
rs1801131	MTHFR	16/25	0.16	0.64 (0.34 - 1.20)	7/8	0.80	0.88 (0.32 - 2.41)	23/33	0.18	0.70 (0.41 - 1.19)
rs742071	PAX7	51/53	0.84	0.96 (0.66 - 1.41)	13/9	0.39	1.44 (0.62 - 3.38)	64/62	0.86	1.03 (0.73 - 1.46)
rs560426	ABCA4	49/35	0.13	1.40 (0.91 - 2.16)	12/14	0.69	0.86 (0.40 - 1.85)	61/49	0.25	1.25 (0.85 - 1.81)
rs481931	ABCA4	15/20	0.40	0.75 (0.38 - 1.47)	2/8	0.06	0.25 (0.05 - 1.18)	17/28	0.10	0.61 (0.33 - 1.11)
rs4147811	ABCA4	12/18	0.27	0.67 (0.32 - 1.38)	3/9	0.08	0.33 (0.09 - 1.23)	15/27	0.06	0.56 (0.30 - 1.04)
rs6677101	SLC25A24	43/49	0.53	0.88 (0.58 - 1.32)	19/10	0.09	1.90 (0.88 - 4.09)	62/59	0.79	1.05 (0.74 - 1.50)
rs642961	IRF6	14/20	0.30	0.70 (0.35 - 1.39)	0/5	0.03	0.00 (0.00 - NA)	14/25	0.08	0.56 (0.29 - 1.08)
rs7590268	THADA	34/29	0.53	1.17 (0.71 - 1.92)	7/7	1.00	1.00 (0.35 - 2.85)	41/36	0.57	1.14 (0.73 - 1.78)
rs115200552	MSX1	9/10	0.82	0.90 (0.37 - 2.22)	4/2	0.41	2.00 (0.37 - 10.92)	13/12	0.84	1.08 (0.49 - 2.37)
rs7078160	VAX1	43/28	0.08	1.54 (0.95 - 2.47)	14/8	0.20	1.75 (0.73 - 4.17)	57/36	0.03	1.58 (1.04 - 2.40)
rs4752028	VAX1	46/46	1.00	1.00 (0.66 - 1.51)	18/8	0.05	2.25 (0.98 - 5.18)	64/54	0.36	1.19 (0.83 - 1.70)
rs9574565	SPRY2	47/32	0.09	1.47 (0.94 - 2.30)	12/15	0.56	0.80 (0.37 - 1.71)	59/47	0.24	1.26 (0.86 - 1.84)
rs4783099	CRISPLD2	52/35	0.07	1.49 (0.98 - 2.28)	11/17	0.26	0.65 (0.30 - 1.38)	63/52	0.31	1.21 (0.84 - 1.75)
rs7224837	AXIN2	11/14	0.55	0.79 (0.36 - 1.73)	1/3	0.32	0.33 (0.03 - 3.21)	12/17	0.35	0.71 (0.34 - 1.48)
rs17820943	MAFB	30/20	0.16	1.50 (0.85 - 2.64)	13/8	0.28	1.63 (0.67 - 3.92)	43/28	0.08	1.54 (0.95 - 2.47)
rs13041247	MAFB	30/20	0.16	1.50 (0.85 - 2.64)	13/8	0.28	1.63 (0.67 - 3.92)	43/28	0.08	1.54 (0.95 - 2.47)

T: transmitted alleles, NT: not transmitted alleles, OR: odds ratio, CI: confidence interval, p: p values. The p-values that reached threshold significance are shown in bold characters whereas those that exhibited a trend towards association are shown in italics.

Table 4.7: TDT subphenotype analyses for the Ghanaian NSCL/P cohort

SNP	Probable gene/loci	NSCL			NSCLP		
		T/NT	p	OR (95% CI)	T/NT	p	OR (95% CI)
rs1801131	MTHFR	5/15	0.03	0.33 (0.12 - 0.92)	10/10	1.00	1.00 (0.42 - 2.40)
rs742071	PAX7	30/24	0.41	1.25 (0.73 - 2.14)	21/28	0.32	0.75 (0.43 - 1.32)
rs560426	ABCA4	23/25	0.77	0.92 (0.52 - 1.62)	26/10	7.66E-03	2.60 (1.25 - 5.39)
rs481931	ABCA4	7/10	0.47	0.70 (0.27 - 1.84)	8/10	0.64	0.80 (0.32 - 2.03)
rs4147811	ABCA4	2/8	0.06	0.25 (0.05 - 1.18)	10/10	1.00	1.00 (0.42 - 2.40)
rs6677101	SLC25A24	18/28	0.14	0.64 (0.36 - 1.16)	25/21	0.56	1.19 (0.67 - 2.13)
rs642961	IRF6	9/11	0.65	0.82 (0.34 - 1.97)	5/9	0.29	0.56 (0.19 - 1.66)
rs7590268	THADA	15/20	0.40	0.75 (0.38 - 1.47)	19/9	0.06	2.11 (0.96 - 4.67)
rs115200552	MSX1	5/2	0.26	2.50 (0.49 - 12.89)	3/8	0.13	0.38 (0.10 - 1.41)
rs7078160	VAX1	24/17	0.27	1.41 (0.76 - 2.63)	18/11	0.19	1.64 (0.77 - 3.47)
rs4752028	VAX1	21/27	0.39	0.78 (0.44 - 1.38)	25/18	0.29	1.39 (0.76 - 2.55)
rs9574565	SPRY2	25/18	0.29	1.39 (0.76 - 2.55)	21/13	0.17	1.62 (0.81 - 3.23)
rs4783099	CRISPLD2	26/23	0.67	1.13 (0.65 - 1.98)	25/12	0.03	2.08 (1.05 - 4.15)
rs7224837	AXIN2	5/6	0.76	0.83 (0.25 - 2.73)	6/8	0.59	0.75 (0.26 - 2.16)
rs17820943	MAFB	13/10	0.53	1.30 (0.57 - 2.97)	17/9	0.12	1.89 (0.84 - 4.24)
rs13041247	MAFB	13/10	0.53	1.30 (0.57 - 2.97)	17/9	0.12	1.89 (0.84 - 4.24)

T: transmitted alleles, NT: not transmitted alleles, OR: odds ratio, CI: confidence interval, p: p values. The p-values that reached threshold significance are shown in bold characters whereas those that exhibited a trend towards association are shown in italics.

Table 4.8: TDT analyses for African (Ghanaian, Ethiopian and Nigerian) NSOFCs

SNP	Probable gene/loci	NSCL/P			NSCP		
		T/NT	p	OR (95% CI)	T/NT	p	OR (95% CI)
rs1801131	MTHFR	27/34	0.37	0.79 (0.48 - 1.32)	10/9	0.82	1.11 (0.45 - 2.73)
rs742071	PAX7	82/75	0.58	1.09 (0.80 - 1.50)	16/11	0.34	1.46 (0.68 - 3.13)
rs560426	ABCA4	78/59	0.10	1.32 (0.94 - 1.85)	18/18	1.00	1.00 (0.52 - 1.92)
rs481931	ABCA4	28/25	0.68	1.12 (0.65 - 1.92)	3/8	0.13	0.38 (0.10 - 1.41)
rs4147811	ABCA4	26/25	0.89	1.04 (0.60 - 1.80)	5/10	0.20	0.50 (0.17 - 1.46)
rs6677101	SLC25A24	65/75	0.40	0.87 (0.62 - 1.21)	21/14	0.24	1.50 (0.76 - 2.95)
rs642961	IRF6	29/29	1.00	1.00 (0.60 - 1.67)	2/7	0.10	0.29 (0.06 - 1.38)
rs7590268	THADA	49/48	0.92	1.02 (0.69 - 1.52)	8/8	1.00	1.00 (0.38 - 2.66)
rs115200552	MSX1	10/13	0.53	0.77 (0.34 - 1.75)	7/2	0.10	3.50 (0.72 - 16.85)
rs7078160	VAX1	60/44	0.12	1.36 (0.92 - 2.01)	18/10	0.13	1.80 (0.83 - 3.90)
rs4752028	VAX1	73/76	0.81	0.96 (0.70 - 1.32)	27/13	0.03	2.08 (1.07 - 4.03)
rs9574565	SPRY2	69/55	0.21	1.26 (0.88 - 1.79)	18/17	0.87	1.06 (0.55 - 2.05)

rs4783099	CRISPLD2	75/64	0.35	1.17 (0.84 - 1.64)	15/21	0.32	0.71 (0.37 - 1.39)
rs7224837	AXIN2	19/27	0.24	0.70 (0.39 - 1.27)	1/6	0.06	0.17 (0.02 - 1.38)
rs17820943	MAFB	49/42	0.46	1.17 (0.77 - 1.76)	15/12	0.56	1.25 (0.59 - 2.67)
rs13041247	MAFB	49/43	0.53	1.14 (0.76 - 1.72)	15/12	0.56	1.25 (0.59 - 2.67)

T: transmitted alleles, NT: not transmitted alleles, OR: odds ratio, CI: confidence interval, p: p values. The p-values that reached threshold significance are shown in bold characters whereas those that exhibited a trend towards association are shown in italics.

Table 4.9: TDT subphenotype analyses for African (Ghanaian, Ethiopian and Nigerian) NSCL/P

SNP	Probable gene/loci	NSCL			NSCLP		
		T/NT	p	OR (95% CI)	T/NT	p	OR (95% CI)
rs1801131	MTHFR	9/20	0.04	0.45 (0.20 - 0.99)	18/14	0.48	1.29 (0.64 - 2.59)
rs742071	PAX7	50/30	0.03	1.67 (1.06 - 2.62)	32/45	0.14	0.71 (0.45 - 1.12)
rs560426	ABCA4	32/35	0.71	0.91 (0.57 - 1.48)	46/24	8.55E-03	1.92 (1.17 - 3.14)
rs481931	ABCA4	10/13	0.53	0.77 (0.34 - 1.75)	18/12	0.27	1.50 (0.72 - 3.14)
rs4147811	ABCA4	8/10	0.64	0.80 (0.32 - 2.03)	18/15	0.60	1.20 (0.60 - 2.38)
rs6677101	SLC25A24	26/41	0.07	0.63 (0.39 - 1.04)	39/34	0.56	1.15 (0.72 - 1.82)
rs642961	IRF6	16/15	0.86	1.07 (0.53 - 2.16)	13/14	0.85	0.93 (0.44 - 1.98)
rs7590268	THADA	21/32	0.13	0.66 (0.38 - 1.14)	28/16	0.07	1.75 (0.95 - 3.23)
rs115200552	MSX1	6/3	0.32	2.00 (0.50 - 8.00)	4/10	0.05	0.30 (0.13 - 1.28)
rs7078160	VAX1	37/23	0.07	1.61 (0.96 - 2.71)	23/21	0.76	1.10 (0.61 - 1.98)
rs4752028	VAX1	32/38	0.47	0.84 (0.53 - 1.35)	41/38	0.74	1.08 (0.69 - 1.68)
rs9574565	SPRY2	35/29	0.45	1.21 (0.74 - 1.97)	34/26	0.30	1.31 (0.78 - 2.18)
rs4783099	CRISPLD2	39/35	0.64	1.11 (0.71 - 1.76)	36/29	0.39	1.24 (0.76 - 2.02)
rs7224837	AXIN2	10/9	0.82	1.11 (0.45 - 2.73)	9/18	0.08	0.50 (0.22 - 1.11)
rs17820943	MAFB	18/22	0.53	0.82 (0.44 - 1.53)	31/20	0.12	1.55 (0.88 - 2.72)
rs13041247	MAFB	18/22	0.53	0.82 (0.44 - 1.53)	31/21	0.17	1.48 (0.85 - 2.57)

T: transmitted alleles, NT: not transmitted alleles, OR: odds ratio, CI: confidence interval, p: p values. The p-values that reached threshold significance are shown in bold characters whereas those that exhibited a trend towards association are shown in italics.

association or a trend towards association with NSOFCs in FBAT as far as the Ghanaian cohort is concerned included IRF6 (rs642961, p=0.06), VAX1 (rs7078160, p=0.02), BMP4 (rs17563, p=0.10) and MAFB (rs17820943, p=0.05 and rs13041247, p=0.07).

Parent of origin (poo) effects were not observed for almost all SNPs, except rs16260 of CDH1 and rs8001641 of SPRY2. For rs16260, a trend towards association was observed for both the Ghanaian (p=0.06) and African (p=0.08) cohorts for NSOFCs.

The rs16260 SNP exhibited a maternal imprinting or maternal over-transmission effect. On the other hand, rs8001641 showed a paternal over-transmission effect with NSCL (p=0.06) in the Ghanaian cohort. However, rs16260 did not attain statistical significance with NSOFCs in any of the association studies, though rs8001641 demonstrated evidence of association with NSCL/P in DFAM analyses in the Ghanaian cohort only.

Interestingly, in the most inclusive family-based test of DFAM (Table 4.10), 22 SNPs either showed formal significance or a trend towards significance with NSOFCs. Here again, IRF6, VAX1, ABCA4, MAFB and BMP4 demonstrated evidence of association with NSOFCs in the Ghanaian but not in the combined African cohorts. Subphenotype analyses revealed that some genes may be responsible for specific subphenotypes. For example, rs642961 of IRF6 still dissentingly maintained its association with NSCP whereas CRISPLD2 and THADA were associated with NSCL/P

Table 4.10: Family-Based Association for Disease Traits (DFAM)

SNP	Tagging Gene	p-values for Ghana					p-values for Africa			
		NS CL/P	NS CL	NS CLP	NS CP	NS OFC	NSCL/P	NS CL	NSCLP	NS CP
rs742071	PAX7	0.75	0.41	0.64	0.63	0.68	0.32	0.02	0.29	0.96
rs560426	ABCA4	0.07	0.98	3.17E-03	0.94	0.13	2.59E-02	0.72	4.75E-03	0.80
rs138751793	ARHGAP29	0.19	0.12	0.63	0.15	0.07	0.38	0.66	0.43	0.40
rs6677101	SLC25A24	0.73	0.35	0.59	0.03	0.46	1.00	0.80	0.64	0.24
rs861020	IRF6	0.92	0.86	0.70	0.08	0.39	0.43	0.23	0.98	0.35
rs642961	IRF6	0.45	0.77	0.40	0.01	0.09	0.83	0.99	0.98	0.15
rs7590268	THADA	0.26	0.82	0.04	0.26	0.62	0.85	0.30	0.18	0.77

rs4332945	DYSF	0.16	0.07	0.93	0.66	0.13	0.04	0.02	0.60	0.62
rs115200552	MSX1	0.95	0.09	0.10	0.16	0.36	0.89	0.13	0.04	0.08
rs987525	8q24	0.07	0.08	0.44	0.98	0.12	0.80	0.50	0.52	0.99
rs7078160	VAX1	0.04	0.11	0.17	0.47	0.02	0.21	0.18	0.77	0.28
rs4752028	VAX1	0.73	0.17	0.25	0.10	0.67	0.88	0.43	0.30	0.06
rs9574565	SPRY2	0.05	0.14	0.20	0.87	0.11	0.07	0.16	0.28	0.22
rs17563	BMP4	0.14	0.41	0.17	0.24	0.05	0.66	0.15	0.80	0.70
rs1258763	GREM1	0.20	0.79	0.09	0.40	0.11	0.14	1.00	0.06	0.98
rs11642413	CDH1	0.90	0.91	0.97	0.56	0.77	0.33	0.81	0.08	0.88
rs4783099	CRISPLD2	0.03	0.20	0.05	0.59	0.09	0.17	0.14	0.89	0.37
rs227731	NOG	0.86	0.57	0.35	0.21	0.44	0.24	0.41	0.43	0.09
rs3923086	AXIN2	0.55	0.51	0.86	0.85	0.60	0.89	0.70	2.88E-03	0.85
rs17820943	MAFB	0.03	0.10	0.15	0.32	0.02	0.31	0.88	0.16	0.65
rs13041247	MAFB	0.03	0.07	0.18	0.29	0.02	0.37	0.83	0.21	0.63
rs11696257	MAFB	0.04	0.09	0.25	0.40	0.04	0.46	0.89	0.26	0.77

Note - NSOFC: combination of NSCL/P and NSCP, NSCL/P: nonsyndromic cleft lip with or without cleft palate, NSCL: nonsyndromic cleft lip, NSCLP: nonsyndromic cleft lip with palate, NSCP: nonsyndromic cleft palate. The p-values that reached threshold significance are shown in bold characters whereas those that exhibited a trend towards association are shown in italics.

and NSCLP, respectively, in the Ghanaian cohort. Furthermore, in the Ghanaian cohort, SLC25A24 and SPRY2 also gained threshold association with NSCP and NSCL/P, respectively, with BMP4 showing threshold association with all NSOFCs. The 8q24 gene desert also exhibited a trend towards association with NSOFCs in the Ghanaian cohort. In addition, AXIN2, PAX7, ABCA4 and MSX1 showed evidence of association with NSOFCs in the combined African cohort but not the isolated Ghanaian cohort. Other genes also exhibited a trend towards association with NSOFCs in both Ghanaian and African cohorts. These observations suggest that family-based study designs may be more informative and successful in the Ghanaian population.

4.5.2 Gene-Gene and Gene-Environment interactions

In Gene-Gene (G×G) or epistatic interactions, three SNPs exhibited evidence of epistasis with other SNPs in the combined African cohort. Each of these epistatic interactions

yielded p=0.02. A SNP for ABCA4, rs560426, interacted with rs2674394 (gene desert). Moreover, rs2303596 of DYSF interacted with rs3923086 of AXIN2. Finally, rs8069536 of NTN1 interacted with rs17820943, rs13041247 and rs11696257, all of MAFB. Furthermore, no significant Gene-Environment (G×E) interactions were observed between any SNP and environmental factors.

4.6: Genomics of OFCs: detection of pathogenic variants by direct DNA sequencing

Chromatograms generated (Figure 4.8a; Appendix O) for each PCR plate were aligned to the consensus human genome and viewed (Figure 4.8b) with Consed programme. Figure 4.8a shows chromatograms for 3 individuals; the orange arrow indicates an individual that has a heterozygote mutation at that locus of the genome. I Figure 4.8b, an aligned reads is displayed. Here, the bases were called out for each

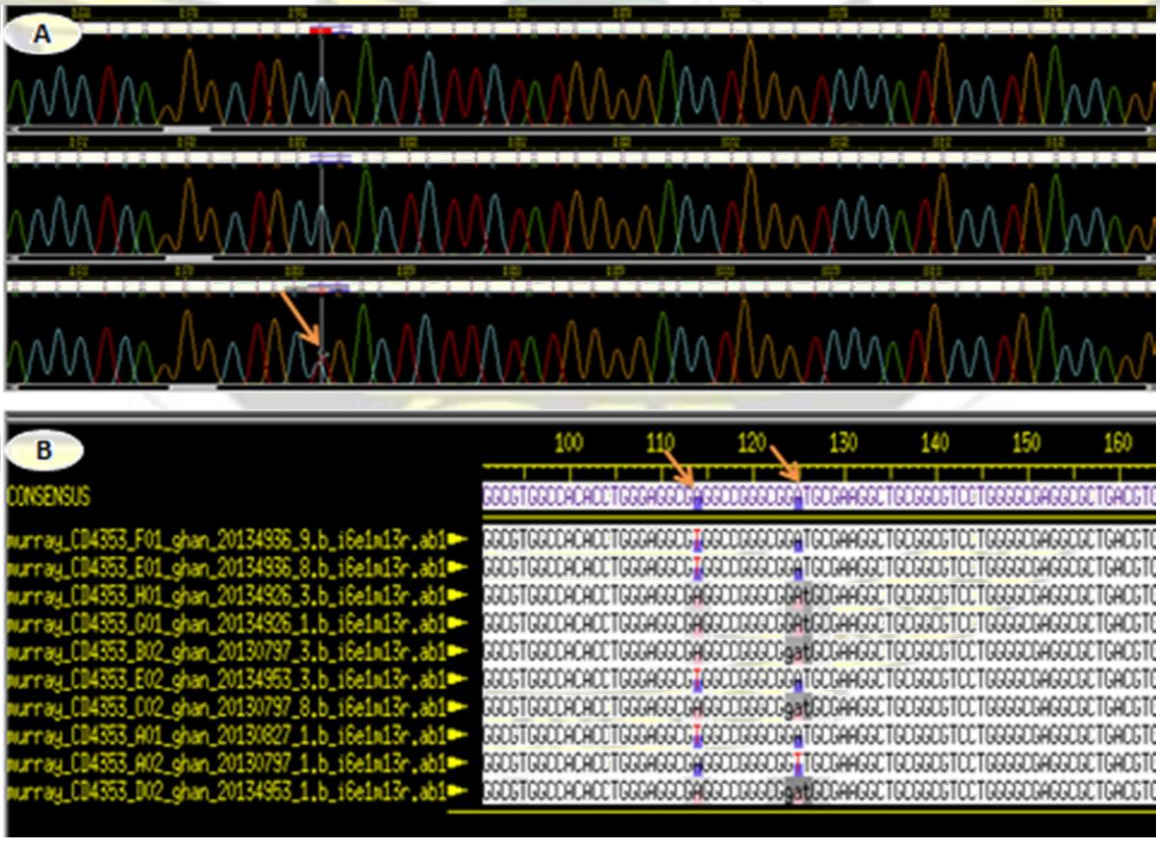


Figure 4.8: Direct DNA sequence results. A: sequence chromatograms for 3 individuals. B: aligned reads created by Consed programme. Orange arrows indicate sites of SNPs.

individual and were aligned to the reference genome. Blue columns and orange arrows indicate sites of genetic variations that had been called out. However, not all variants were successfully called out by Consed. So apart from calls that were made by Consed, each sequence was also perused meticulously to look for variants. For example, insertions and deletions were hardly called by Consed; these could only be detected through careful perusal of the DNA sequence.

4.6.1 Rare functional variants in IRF6

After the various analyses pipelines stipulated in the methodology, 14 rare functional variants were observed in IRF6 in both syndromic and nonsyndromic cleft cases (Table 4.11, Appendix O). Seven of these mutations were novel: they have never

Table 4.11: Rare functional variants observed in IRF6

Variant	cDNA position	Amino acid change	a	b	Polyphen-2	Sift	Human Splice Finder	RDB	Reference
chr1:209,979,529 A>T (rs34743335)	N/A	N/A	5	1	N/A	N/A	N/A	i	dbSNP
chr1:209,969,899 T>G	c.175-2A>C	N/A	1	0	N/A	N/A	c	N/A	Novel
chr1:209,969,878 C>A	c.194G>T	p.Gly65Val	0	1	Probably Damaging	Deleterious	N/A	N/A	Novel
chr1:209,969,867 C>T	c.205G>A	p.Glu69Lys	2	0	Probably Damaging	Deleterious	N/A	N/A	Novel
chr1:209,969,809 T>C	c.263A>G	p.Asn88Ser	1	0	Probably Damaging	Deleterious	N/A	N/A	de Lima et al 2009
chr1:209,969,738 G>C (rs116264750)	c.334C>G	p.Gln112Glu	2	0	Benign	Tolerated	N/A	N/A	dbSNP
chr1:209,969,692 C>A	c.379+1 G>T	N/A	0	1	N/A	N/A	d	N/A	Novel
chr1:209,968,879 A>T (rs148969434)	c.380-16T>A	N/A	5	2	N/A	N/A	e, f, g	N/A	dbSNP
chr1:209,965,727 T>G	c.554A>C	p.Asn185Thr	1	0	Benign	Tolerated	h	N/A	Novel
chr1:209,964,151 C>T	c.749G>A	p.Arg250Gln	0	1	Probably Damaging	Deleterious	N/A	N/A	de Lima et al 2009
chr1:209,964,152 G>A	c.748C>T	p.Arg250X	1	2	N/A	N/A	N/A	N/A	de Lima et al 2009
chr1:209,963,955 C>T	c.945G>A	p.Arg315Arg	1	0	Benign	Tolerated	h	N/A	ExAC
chr1:209,963,940 C>G	c.960G>C	p.Lys320Asn	0	1	Probably Damaging	Deleterious	N/A	N/A	Novel
chr1:209,963,814 G>A	c.1060+26C>T	N/A	1	0	N/A	N/A	N/A	i	Novel

a: Total number observed in NSOFC cases, b: Total number observed in syndromic clefts, c: Alteration of the wildtype acceptor site-most probably affecting splicing, d: Alteration of the wildtype donor site-most probably affecting splicing, e: Activation of an intronic

cryptic acceptor site, f: Activation of an intronic cryptic donor site, g: creation of an intronic ESE site, h: Alteration of an exonic ESE site; potential alteration of splicing, i: 2b-Likely to affect binding of POLR2A. RDB implies RegulomeDB software.

been reported in literature or in any of the reference databases stated in the methodology. In silico analyses were based on Ensemble transcript number ENST00000367021 for IRF6. It must be noted that p.Asn185Thr was observed in an individual with the p.Glu69Lys variant. Apart from this, all variants stated in the table were observed in different individuals.

Figure 4.9 shows chromatograms and pedigrees in which a rare, novel aetiologic missense variant, p.Glu69Lys occurred. The top chromatogram (Figure 4.9a) is from an unaffected individual whereas the middle and bottom ones are from affected individuals. The white vertical line delineates the nucleotide position where the substitution mutation occurred. The mutation changes cytosine to thymine (C>T) on chromosome 1 coordinate chr1:209,969,867 and cDNA position c.205G>A, which result in wildtype glutamate at position 69 of IRF6 protein being replaced with Lysine (Table 4.11). This variant occurred in two NSOFC patients: one with incomplete NSCP (soft palate cleft only) and the other with right complete unilateral NSCLP with bifid uvula. This variant segregated in pedigree GH20135021, where the variant occurred in both the proband and father, though the father had no overt cleft. In pedigree GH20135024, the variant occurred in the proband, but not the mother. However, there was no sample from the father so as to determine whether it was segregating or de novo.

Computational simulation of mutant protein structure by Project Hope predicted that the p.Glu69Lys mutation altered IRF6 protein structure and function (Figure 4.10). The mutation occurred in the highly conserved, N-terminal helix-turn-helix DNAbinding domain of IRF6 transcription factor. The p.Gly65Val missense variant was also functional (Figure 4.11a, b) in that the wildtype residue also occurred in the helix-turnhelix DNA-binding domain. Valine is bigger and more hydrophobic than glycine.

Finally, the p.Lys320Asn variant was also detrimental to protein structure and function (Figure 4.11b, d), as it is located in the protein-binding, transactivation domain of IRF6 transcription factor. Figure 4.12 shows a RegulomeDB prediction for an enhancer element.

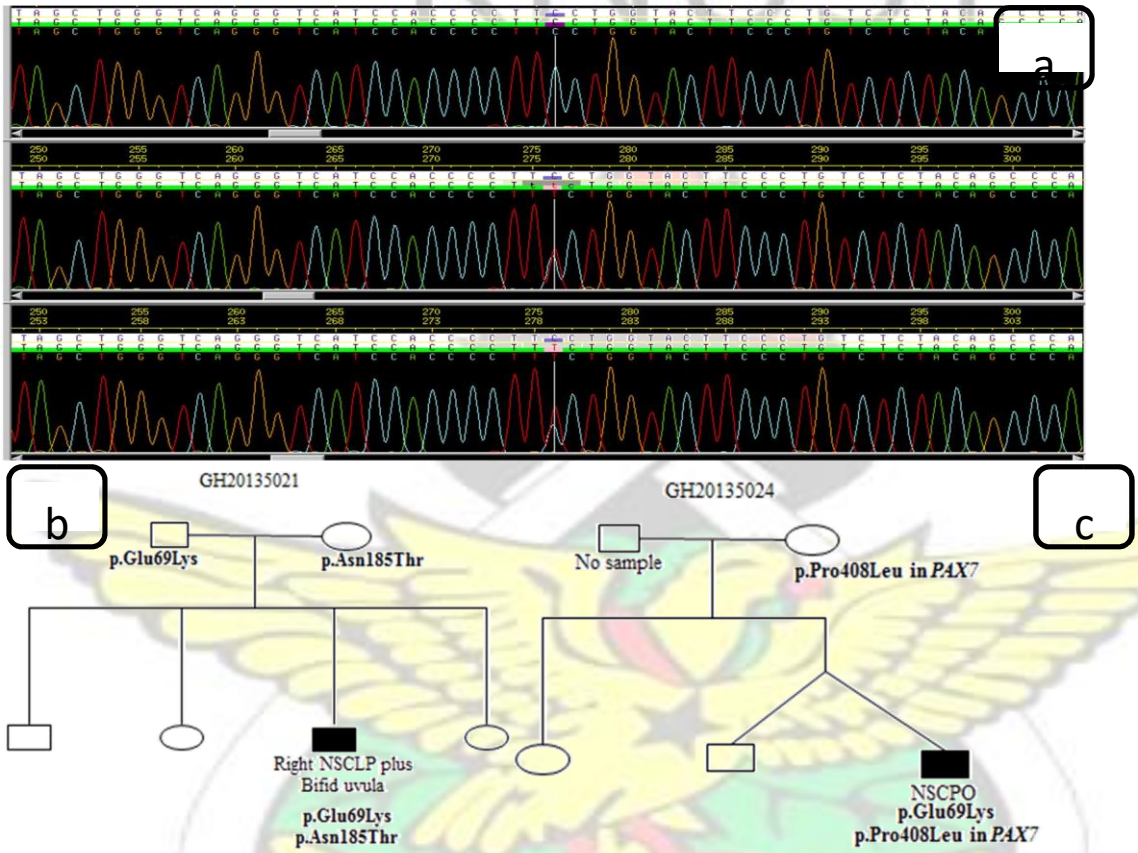


Figure 4.9: DNA Sequence chromatograms and pedigrees for p.Glu69Lys variant

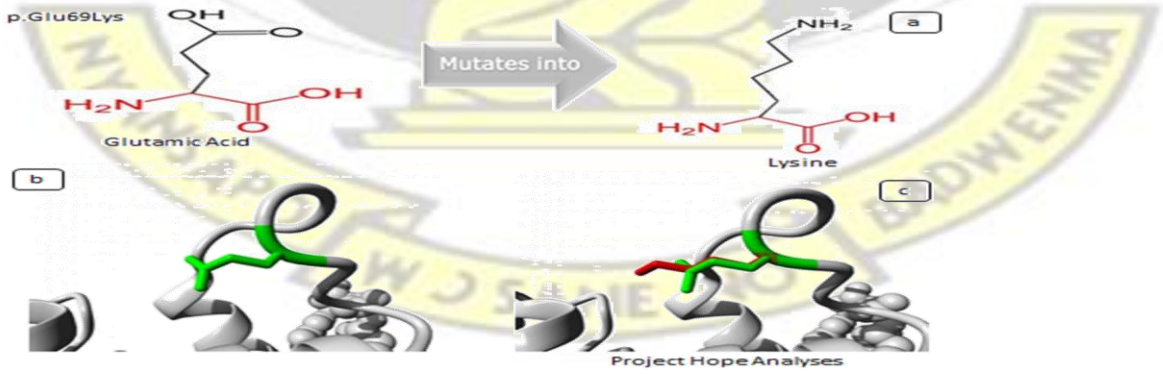


Figure 4.10: The p.Glu69Lys mutant protein structure as predicted by Project HOPE. a: substitution of Glutamate for Lysine, b: Green indicates wildtype residue, c: Red indicates the effect of mutant residue on protein folding.

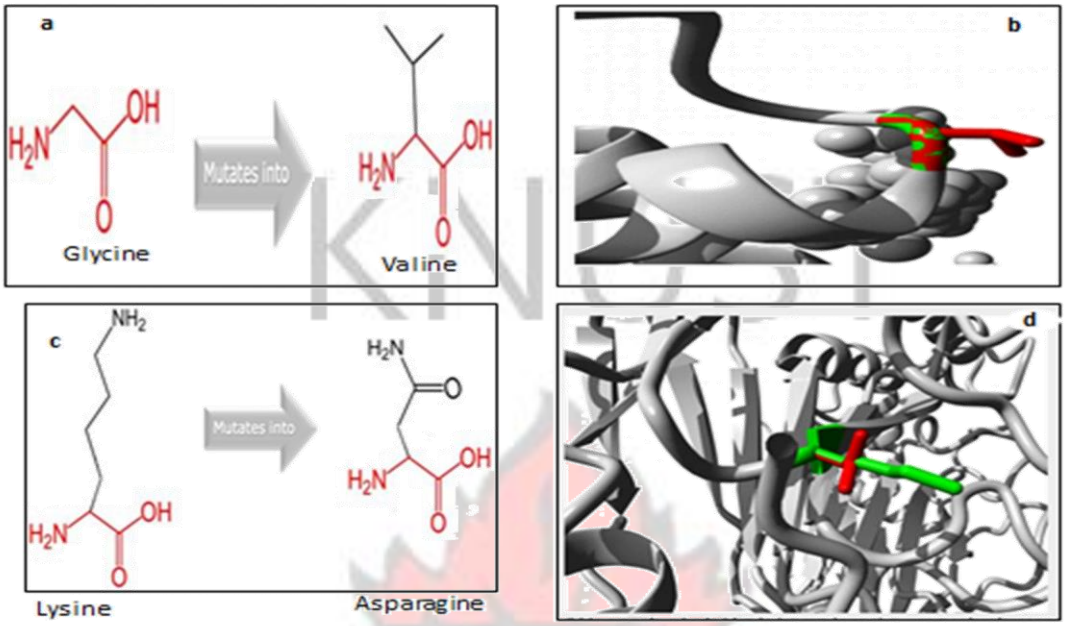


Figure 4.11: Project Hope analyses for two other novel variants. a: wildtype and mutant amino acid residues for p.Gly65V al , b: p.Gly65Val mutant protein structure, c: p.Lys320Asn wildtype and mutant amino acid residues and d: mutant protein structure of p.Lys320Asn. In b and d, green and red indicate wildtype and mutant amino acid residues, respectively.

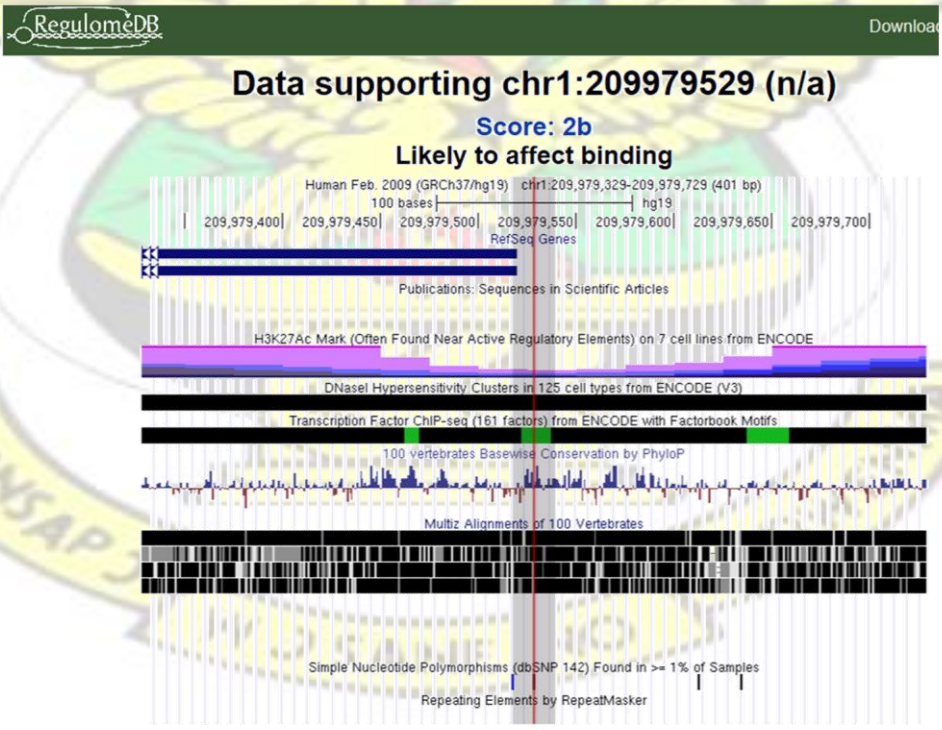


Figure 4.12: The rs34743335 enhancer element. Red vertical line indicates the mutated nucleotide position in the gene.

4.6.2: Rare functional variants in other seven genes

Many rare functional variants were observed in ARHGAP29 (Table 4.12; Appendix O). Some of these variants affect mRNA splicing whereas others affect protein structure. A novel protein-truncating frameshift mutation, p.Lys426IlefsTer6, resulting from a homozygous insertion, leads to the termination of ARHGAP29 protein after 6 amino acids. This variant was found in a female case with right complete CLP but the variant was absent in the mother, though there was no sample from the father. This variant could therefore have emerged by de novo mutation (if the variant is not in the father) or paternal uniparental isodisomy (if the variant is from the father), since this is a homozygous insertion. The p.Ser913Leu variant was found in four unrelated cases with diverse phenotypes and segregated from mothers to cases in all four families: a female with right complete CLP, a male with complete CP, a male with left complete CL and a female with CLP born to consanguineous parents. The p.Asn323Asp mutation was observed in a female case with left incomplete CL and was not observed in the available paternal sample. The splice donor site variant, c.1281+4A>G, was observed in both the mother and a male proband with right complete CLP.

A male and a female probands, all with left unilateral CLP plus bifid uvula had the p.Asp988Tyr variant; the variant segregated from mother to the male proband while no parental samples were available for the female proband. Interestingly, these two probands are not related: the male proband comes from Agona Swedru in the Central Region whereas the female proband hails from Sogakope in the Volta. This suggests this mutation arose spontaneously and independently in these families. Globally, this variant has been reported in only 3 African Americans, suggesting this variant is extremely rare and may therefore be pathogenic (http://browser.1000genomes.org/index.html). Project

Table 4.12: Rare functional variants observed in ARHGAP29

Variant	cDNA Position	Amino Acid Change	Total Number of Cases with Variant	Polyphen-2	Sift	Human Splice Finder	Reference
chr1:94,674,936 T>A	c.341-30T>A	N/A	1	N/A	N/A	ESS Site broken	Novel
chr1:94,671,346 T>C	c.511-107T>C	N/A	2	N/A	N/A	ESS Site broken and New ESE Site	Novel
chr1:94,670,953 T>C (rs191674548)	c.560-199T>C	N/A	1	N/A	N/A	ESS Site broken	dbSNP
Chr1:94668276 A>G	c.967A>G	p.Asn323Asp	1	Benign	Deleterious	N/A	Novel
chr1:94667280 T>TA	c.1277delAinsTA	p.Lys426IlefsTer6	1	N/A	N/A	N/A	Novel
chr1:94,667,272 A>G	c.1281+4A>G	N/A	1	N/A	N/A	Alteration of the WT donor site	Novel
chr1:94667431 A>G (rs200743606)	c.1144-18T>C	N/A	2	N/A	N/A	ESS Site broken and New ESE Site	dbSNP
chr1:94,643,466 C>T (rs139252732)	c.2738C>T	p.Ser913Leu	4	Benign	Deleterious	N/A	dbSNP
chr1:94,640,254 T>C (rs142506942)	c.2957T>C	p.Ile986Thr	1	Benign	Tolerated	N/A	dbSNP
chr1:94,640,249 G>T (rs139412088)	c.2962G>T	p.Asp988Tyr	2	Probably Damaging	Deleterious	N/A	dbSNP

chr1:94,640,188 G>A (rs140638899)	c.3023G>A	p.Arg1008Lys	2	Benign	Tolerated	N/A	dbSNP

HOPE predicts that this variant alters ARHGAP29 protein structure and function (Figure 4.13) for a number of biochemical reasons.

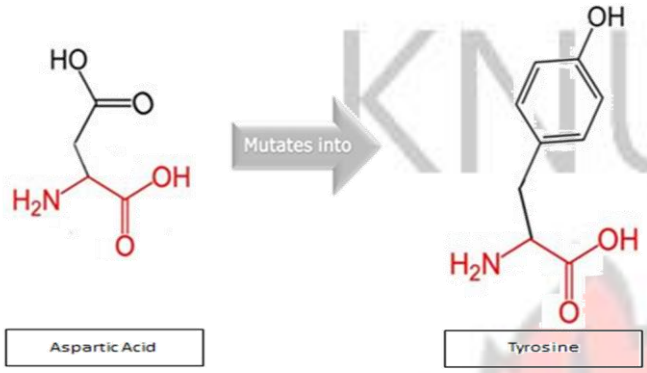


Figure 4.13: Project Hope analysis for p.Asp988Tyr mutation in ARHGAP29

Regulatory, splicing and missense variants were observed in VAX1 gene (Table 4.13; Appendix O). Regulatory region variants were predicted by RegulomeDB to affect the binding of EZH2 (enhancer of zeste 2, a polycomb repressive complex 2 [PRC2] subunit), a histone-lysine N-methyltransferase enzyme that participates in DNA methylation and inadvertently, transcriptional repression. The enzyme is responsible for the methylation activity of PRC2. Splicing variant may affect the stability of VAX1 mRNA transcript. A novel silence or synonymous mutation, p.Ala231Ala, is predicted by Human Splicing Finder to affect VAX1 mRNA stability and inadvertently, translation, for a number of reasons. The variant introduces a new acceptor site with branch points and also alters an ESE site. The p.Ala231Ala mutation was observed in four cases: a male with incomplete CP as well as congenital absence of uvula, a female with SMCP, a male with left complete CLP and a female with complete CP.

A novel missense variant, p.Gly252Cys in VAX1, was observed in a male case with left complete CL; this variant was absent in paternal sample, though no maternal sample was available. Project Hope predicts (Figure 4.14) that the mutant residue

Table 4.13: Rare functional variants observed in VAX1, PAX7, MSX1, BMP4, FOXE1 and MAFB

Variant	cDNA Position	Amino Acid Change	§	Polyphen-2	Sift	¥	λ	Reference
VAX1								
chr10:118,897,605 to chr10:118,897,608 delA-G-C haplotype (rs376169796rs572236555rs369158710-rs573343467)	5'UTR	N/A	8	N/A	N/A	N/A	η	dbSNP/ 1000 Genome
chr10:118,896,022 G>A	c.390G>A	p.Arg130Arg	1	Benign	Tolerated	γ	N/A	Novel
chr10:118,895,946 G>C (rs372955877)	c.429+37G>C	N/A	1	N/A	N/A	ε	N/A	dbSNP/ 1000 Genome
chr10:118,895,933 C>A (rs200558256)	c.429+50C>A	N/A	4	N/A	N/A	δ	η	dbSNP/ 1000 Genome
chr10:118,893,836 G>A (rs571879048)	c.688G>A	p.Ala230Thr	12	Benign	Tolerated	μ, γ	η	dbSNP only
chr10:118,893,831 C>A	c.693C>A	p.Ala231Ala	4	Benign	Tolerated	β, γ	η	Novel
chr10:118,893,770 G>T	c.754G>T	p.Gly252Cys	1	Probably Damaging	Deleterious	N/A	N/A	Novel
chr10:118,891,384 T>C	3'UTR	N/A	1	N/A	N/A	N/A	η	Novel
PAX7								
chr1:19,018,364 G>A (rs147079707)	c.703G>A	p.Ala235Thr	2	Probably Damaging	Deleterious	N/A	N/A	dbSNP and ExAC
chr1:19,062,193 C>T (rs564200011)	c.1223C>T	p.Pro408Leu	1	Probably Damaging	Deleterious	N/A	N/A	dbSNP and ExAC
chr1:19,062,197 G>A	c.1227G>A	p.Leu409Leu	1	Benign	Tolerated	γ	N/A	Novel
MSX1								
chr4:4861721 C>T	c.95C>T	p.Ala32Val	4	Benign	Tolerated	α, γ	N/A	ExAc

chr4:4861745 C>G (rs36059701)	c.119C>G	p.Ala40Gly	28	Benign	Tolerated	μ, γ	N/A	dbSNP
chr4:4,861,844 C>T (rs201156596)	c.218C>T	p.Pro73Leu	3	Possibly Damaging	Deleterious	N/A	N/A	dbSNP
chr4:4,861,923 G>C	c.297G>C	p.Pro99Pro	1	Benign	Tolerated	N/A	N/A	ExAc

Table 4.13 continued …

Variant	cDNA Position	Amino Acid Change	§	Polyphen-2	Sift	¥	λ	Reference
chr4:4861974 C>T (rs34165410)	c.348C>T	p.Gly116Gly	24	Benign	Tolerated	α, γ	N/A	dbSNP
chr4:4,864,480 G>A	c.522G>A	p.Lys174Lys	1	Benign	Tolerated	N/A	N/A	Novel
chr4:4,864,991 G>C (rs115200552)	c.*121G>C	N/A	15	N/A	N/A	N/A	η	dbSNP
BMP4								
chr14:54,417,117 C>T (rs121912768)	c.860G>A	p.Arg287His	1	Benign	Tolerated	N/A	N/A	dbSNP
chr14:54,417,362 G>T	c.615C>A	p.Val205Val	2	Benign	Tolerated	N/A	N/A	ExAc
chr14:54,417,770 C>T	c.371-164G>A	N/A	2	N/A	N/A	ε, δ	N/A	Novel
chr14:54,418,661 C>T	c.280G>A	p.Glu94Lys	1	Benign	Tolerated	N/A	N/A	Novel
chr14:54,418,713 A>T (rs114957446)	c.228T>A	p.Ser76Arg	3	Possibly Damaging	Damaging to two isoforms	N/A	N/A	dbSNP
FOXE1								
chr9:100,616,303 C>T (rs554583478)	c.107C>T	p.Thr36Met	1	Possibly Damaging	Deleterious	N/A	N/A	ExAc
chr9:100,616,712 C>A	c.516C>A	p.Ala172Ala	50	Benign	Tolerated	N/A	N/A	ExAc

chr9:100,616,765 C>G (rs182535331)	c.569C>G	p.Pro190Arg	6	Possibly Damaging	Deleterious	N/A	N/A	dbSNP
MAFB								
chr20:39,317,491 G>T (rs369936070)	c.-1G>A	N/A	2	N/A	N/A	θ	N/A	dbSNP
chr20:39,317,565 C>T (rs574807151)	c.-75C>A	N/A	1	N/A	N/A	N/A	η	dbSNP
chr20:39,316,888 G>C	c.603G>C	p.Ala201Ala	1	Benign	Tolerated	N/A	N/A	Novel

§: Total number of cases with variant, ¥: Human Splice Finder, λ: RegulomeDB, γ: Alteration of wildtype ESE site, ε: Alteration of wildtype ESS site, δ: Creation of new ESE site, μ: Creation of new ESS site, β: New Acceptor site with branch points created, α: New Donor Site created, θ: Alteration of wildtype Donor Site, η: 2b - Likely to affect binding of EZH2.

(cysteine) of p.Gly252Cys is bigger and more hydrophobic than the wildtype residue (glycine) and this biochemical alteration may affect VAX1 protein function.

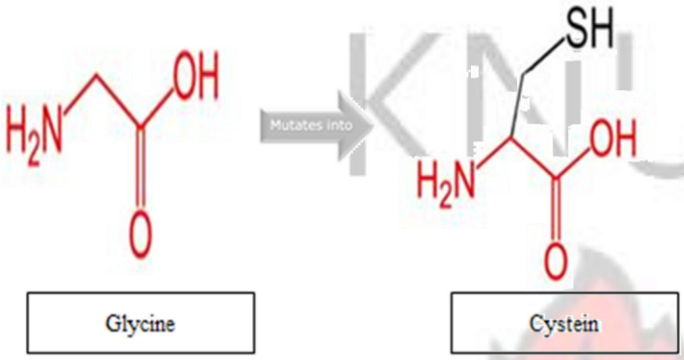


Figure 4.14: Project Hope analysis for p.Gly252Cys in VAX1

Three predicted aetiologic variants, p.Leu409Leu, were observed in PAX7 (Table 4.13). The p.Leu409Leu mutation affects incomplete CP who also had aetiologic p.Glu69Lys variant in IRF6. The variant was inherited from clinically unaffected mother. Project Hope predicts that the mutant residue (leucine) is bigger than the wildty pe residue (proline) and this may leads to bumps (Figure 4.15).

p.Ala235Thr, p.Pro408Leu and

PAX7 mRNA stability by altering an ESE site and was observed in a female case with right complete CL. The p.Pro408Leu mutation was also observed in a male case with

Figure 4.15: Project Hope analysis for p.Pro408Leu in PAX7

Two case probands had the p.Ala235Thr mutation in PAX7 DNA-binding domain: a male with complete bilateral CLP and a female with left complete CLP. In the male case, the variant segregated from the mother to the case whereas in the female case the variant was inherited from the father; however, both parents were clinically unaffected. According to Project Hope (Figure 4.16), the mutant residue (threonine) is bigger and less hydrophobic than the wildtype residue (alanine).

Direct DNA sequence analysis of PAX7 gene also revealed a SNP, rs4920523, in PAX7 being over-transmitted in OFC cases. A total of 46 C/C homozygous and 19 C/G heterozygous genotypes were observed, giving a total minor allele count of 111. Therefore, the MAF for this SNP is about 30% in the study cohort. In 1000 genomes database, rs4920523 has an African MAF of 12%, and a global MAF of 22%.

Comparing the MAF observed in the present study (30%) with the African MAF in 1000

Genomes (12%), suggests that this SNP was significantly associated ($p<0.0001$) with OFCs in the present cohort. Future association studies involving African cohorts should aim at genotyping rs4920523 in PAX7.

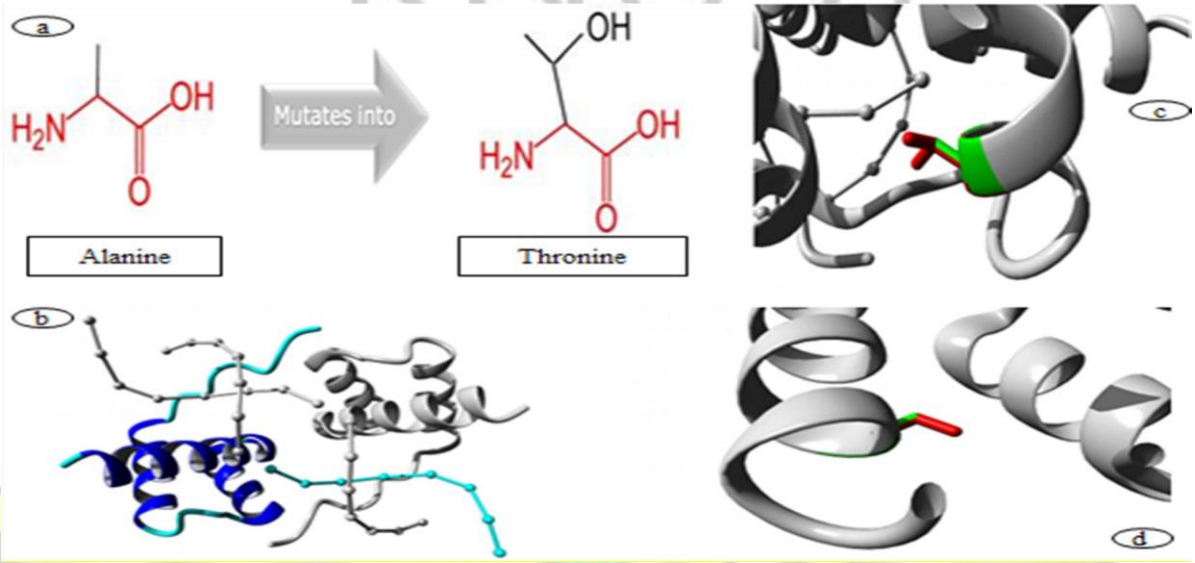


Figure 4.16: Project Hope analysis for p.Ala235Thr variant in PAX7. a: wildtype and mutant amino acid residues, b: wildtype PAX7 protein structure, c and d: effect of the mutation on PAX7 protein structure.

Many predicted pathogenic variants were observed in MSX1 (Table 4.13). A missense variant, p.Pro73Leu, was observed in three cases. In two females, one with complete bilateral CL and the other with incomplete bilateral CL, the variant segregated from the clinically unaffected mothers to the cases. In the third male case with left incomplete CL, the variant was absent in the maternal sample, but no paternal sample was available. Though p.Pro73Leu mutation has been reported in dbSNP, its global minor allele frequency (MAF) is extremely low (less than 1%), suggesting this is an extremely rare mutation and may be aetiologic. However, in dbSNP, no phenotypes are displayed for individuals that had this mutation. This notwithstanding, the present study is the first to report this variant in OFC patients globally. According to Project Hope, this mutation

alters protein structure and function because of the same biochemical alterations stipulated for p.Pro408Leu mutation in PAX7.

Rare variants were also observed in BMP4 (Table 4.13). A missense variant, p.Ser76Arg, was observed in three female cases with diverse phenotypes: left complete CLP, right complete CLP and right complete CL. In the case with right complete CLP, the variant segregated from the mother to the case whereas some parental samples were not available for the other two cases to determine whether the variant was segregating or de novo. Again though this variant has been reported in dbSNP, its global MAF is extremely low (less than 1%), suggesting this is an extremely rare mutation and may be aetiologic. Moreover, in dbSNP, no phenotypes are displayed for individuals that had this mutation. This notwithstanding, the present study is the first to report this variant in OFC patients globally. Project Hope predicts that p.Ser76Arg is pathogenic for a number of biochemical reasons (Figure 4.17).

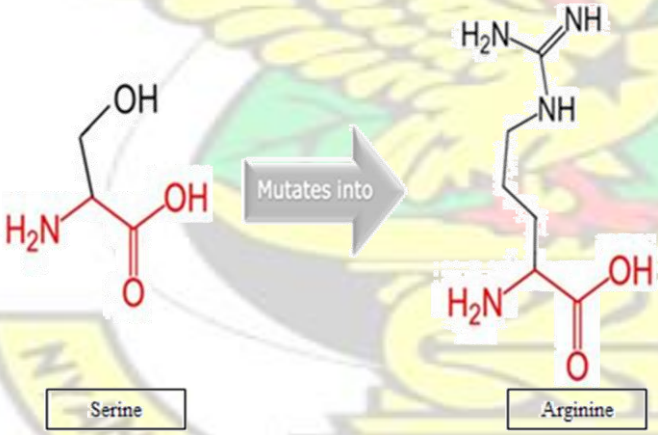


Figure 4.17: Project Hope analysis of p.Ser76Arg mutation in BMP4

In direct DNA sequence analysis of FOXE1 (Table 4.13), a pathogenic mutation, p.Thr36Met, was observed in a male case with left complete CLP. The case was heterozygous for the mutation (A/G) whereas the clinically unaffected mother was

homozygous for the mutation (A/A) but no paternal sample was available (Figure 4.18a). Interestingly, the maternal grandmother was homozygous for wildtype allele (G/G) but no sample from maternal grandfather was available. Either of two genetic mechanisms may account for the homozygous mutation in the mother: de novo mutation (if the variant is not in the maternal grandfather) or paternal uniparental isodisomy (if the variant occurs in the maternal grandfather). According to Project Hope (Figure 4.18b), the bigger mutant residue (methionine) of p.Thr36Met mutation may cause bumps in the protein structure. The mutation also introduces a more hydrophobic residue at position 36, and this can result in loss of hydrogen bonds and/or disturb correct folding. This variant has a global MAF of less than 1% (http://browser.1000genomes.org/index.html), suggesting it is rare and pathogenic. The present study is the first to report this mutation in an OFC patient, since 1000 genomes browser does not give the phenotypes of individuals in which this mutation was observed.

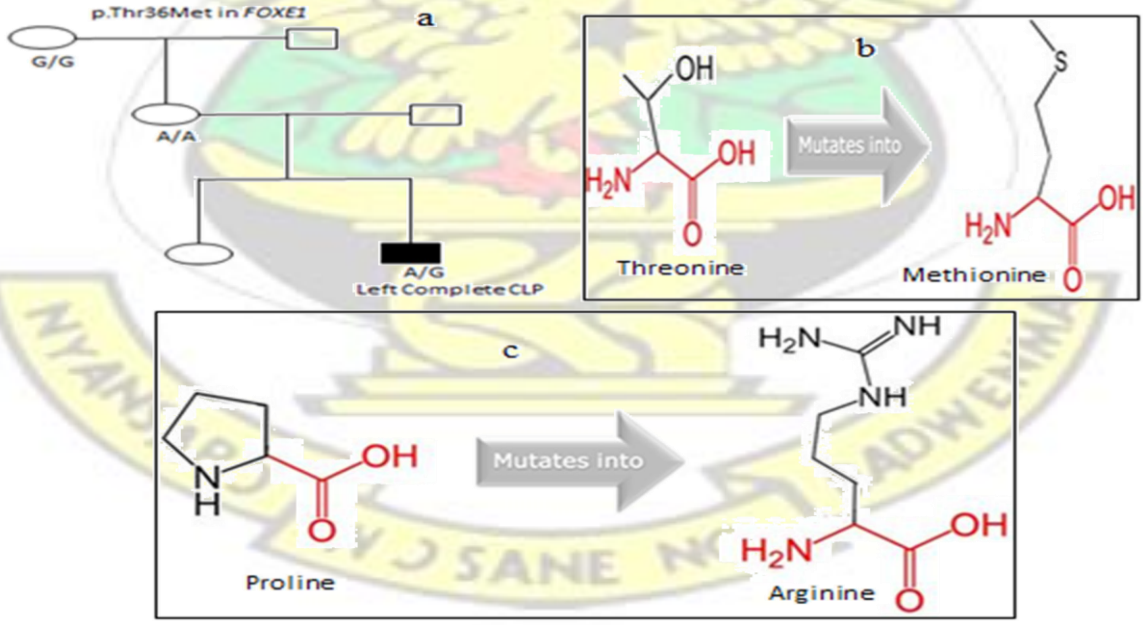


Figure 4.18: Pathogenic missense variants in FOXE1. a: pedigree for p.Thr36Met mutation, b: Project Hope analysis for p.Thr36Met mutation and c: Project Hope analysis for p.Pro190Arg mutation.

Six case probands had p.Pro190Arg mutation in FOXE1 (Table 4.13). A male and a female case with right complete CLP had this variant but some parental samples were not available. In a male twin case with left complete CLP, the variant was not observed in the mother but no paternal sample was available. The mutation was transmitted from heterozygous, clinically unaffected mothers to a female case with right complete CL and a male case with left incomplete CL. Finally, in a male case with ankyloglossia, the variant segregated from the father to the case. For p.Pro190Arg mutation, the mutant residue is bigger than the wildtype residue and this might lead to bumps (Figure 4.18c). The p.Pro190Arg mutation also has a global MAF of less than 1% and this study is the first to report it in OFC patients.

A synonymous variant, p.Ala172Ala (rs775448711), though predicted to be nonaetiologic by various algorithmic tools, is significantly associated with NSOFCs in Ghanaians (Table 4.13). This variant has never been reported in dbSNP, 1000 Genomes or Exome Variant Server. However, it has been reported in ExAC browser with a global MAF of less than 1% and an African and African-American MAF of 10.5% (http://www.ensembl.org/Homo_sapiens/Tools/VEP). The present study observed a minor allele count of 68 (18 homozygous and 32 heterozygous) in 184 cases, resulting in a MAF of about 19%. Therefore, this variant was significantly associated (p=0.0181) with NSCL/P in the Ghanaian population and future association studies must consider genotyping of this SNP.

4.7: Zebrafish-based functional assays

Injection of IRF6 mRNA into wildtype zebrafish embryos yielded diverse phenotypes (Table 4.14). At 60 ng/uL, no phenotype was observed for both wildtype and mutant mRNAs (Plate 4.1, Panel A). Interestingly, none of the embryos died even at 7

Table 4.14: Impact of mutant mRNA on zebrafish embryos

Type of mRNA injected	Number of embryos injected with mRNA	Number of embryos dead at 6 hours	Number of embryos dead at 24 hours	Viable embryos at 48hrs	Number of abnormal embryos at 48 hours	% of abnormal embryos at 48 hours
	100 ng/uL					
Wildtype	26	10	0	16	6	37.50
p.Gly65Val	27	10	1	16	11	68.75
p.Glu69Lys	20	3	2	15	5	33.33
	200 ng/uL					
Wildtype	23	8	0	15	9	60.00
p.Gly65Val	36	9	6	21	14	66.67
p.Glu69Lys	39	11	2	26	23	88.46
	60 ng/uL					
Wildtype	45	0	0	45	0	0
p.Gly65Val	56	0	0	56	0	0
p.Glu69Lys	37	0	0	37	0	0

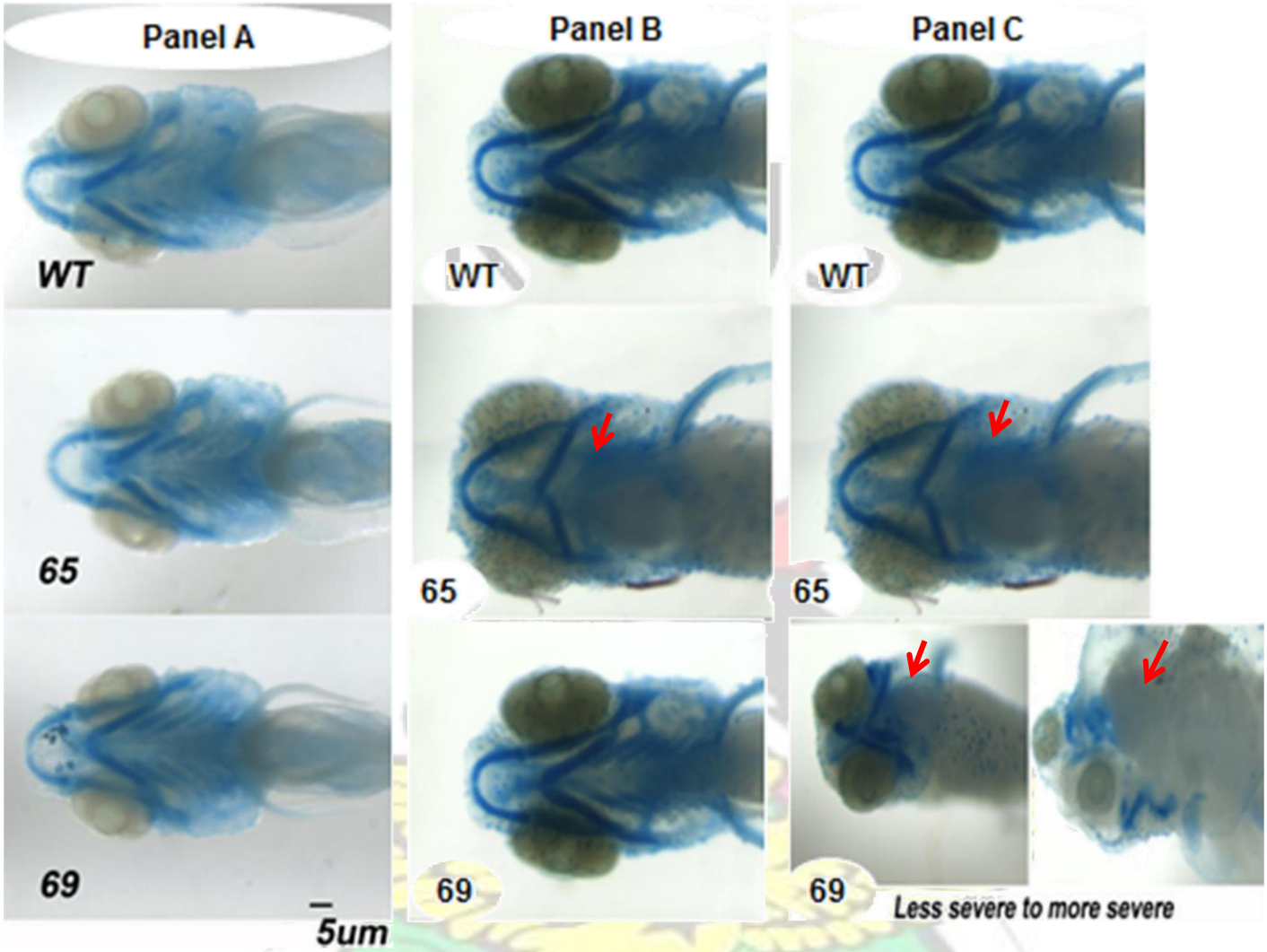


Plate 4.1: Phenotypes observed after treating wildtype zebrafish embryos with both wildtype and mutant IRF6 mRNAs. Panel A: embryos treated with mRNA concentration of 60 ng/uL where no craniofacial phenotype was observed, Panel B: embryos treated with mRNA concentration of 100 ng/uL, Panel C: embryos treated with mRNA concentration of 200 ng/uL, WT: embryos treated with wildtype or canonical IRF6 mRNA served as controls, 65: embryos treated with mutant p.Gly65Val IRF6 mRNA, 69: embryos treated with mutant p.Glu69Lys IRF6 mRNA. Red arrows indicate disrupted arches. Top of Panels A, B and C: phenotype observed in embryos treated with wildtype mRNA at 60, 100 and 200 ng/uL, demonstrating no disruption of pharyngeal arches. Middle of both Panels B and C: most ceratobranchial arches were absent, together with malformed hyoid and mandibular arches in embryos treated with p.Gly65Val at both 100 and 200 ng/uL. Bottom of Panel C: almost all arches were either severely malformed or completely absent in embryos treated with p.Glu69Lys mRNA at 200 ng/uL.

days after fertilization and injection of mRNA. This observation suggested that all embryos collected for the 60 ng/uL treatment were all fertilized eggs. An egg that is not fertilized by a sperm does not produce viable embryos and such embryos die by 6 hours

after embryo collection. Moreover, it also suggested that the 60 ng/uL dosage used had no effect on the development of zebrafish.

Phenotypes for the various treatments were palpable by 48 hours postfertilization. At 100 ng/uL, wildtype mRNA yielded no major defects, apart from few embryos that had very minor head deformities. However, at 200 ng/uL, about 50% of embryos treated with wildtype mRNA had minor abnormalities: the heads of these embryos were slightly pointing upward instead of the expected downward position. This effect probably resulted from the high dosage of mRNA. This notwithstanding, after staining embryos with Alcian blue by day 7, no major deformities of the arches were observed for embryos treated with wildtype mRNA at both 100 ng/uL and 200 ng/uL (Plate 4.1, top of Panels B and C). Moreover, about 90% of wildtype-treated embryos that were viable at 48 hours post-fertilization were still alive by day 7, when they were knocked-down and stained with Alcian blue. This presupposes the wildtype treatment had no major effect on the embryos.

Embryos that were treated with mutant p.Gly65Val mRNA showed similar phenotypes and pattern of death at both 100 ng/uL and 200 ng/uL treatments (Plate 4.1, middle of Panels B and C). Abnormal phenotypes observed for both dosages of p.Gly65Val-treated embryos at 48 hours post-fertilization included missing eyes, head severely pointing upward instead of downward, moderate heart oedema and malformed pharyngeal arches. After staining with Alcian blue, severe abnormalities of the pharyngeal arches (Plate 4.2B) were observed. The Meckel's cartilage and palatoquadrate (mandibular arches) were very broad, instead of the expected almost pointed arch; this suggests lack of complete fusion of these arches. The two hyosymplectic arches, were almost absent, affecting the joint formed between these arches with palatoquadrate and ceratohyal arches. Moreover, the two ceratohyal arches were extremely broad, with each ceratohyal arch

being far apart, almost 180^{o} apart. This also suggests disruption of a major hyoid arch, which emanated from lack of proper formation and fusion of the two ceratohyal arches. Interestingly, the basihyal arch was severely malformed, with most of the ceratobranchial arches being either completely missing or severely malformed. Of further interest is the observation that over 70% of mutant p.Gly65Val-treated embryos that were viable at 48 hours post-fertilization died by day 7.

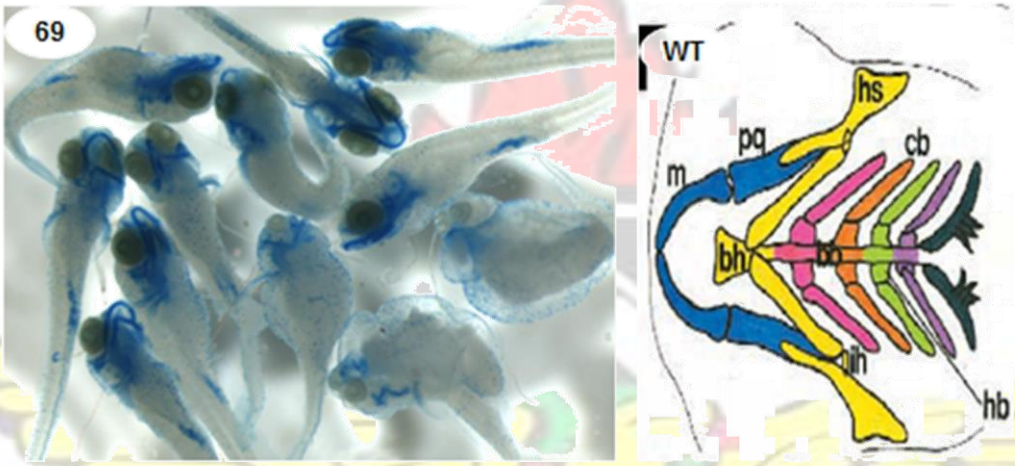


Plate 4.2: Plethora of phenotypes observed in embryos treated with 200 ng/uL p.Glu69Lys mutant mRNA and pharyngeal arches in wildtype zebrafish embryos. 69: array of phenotypes observed in p.Glu69Lys-treated embryos, WT: wildtype pharyngeal arches in zebrafish, bh: basihyal, cb: ceratobranchial, ch: ceratohyal, hs: hyosymplectic, m: meckel's cartilage, pq: palatoquadrate, blue: mandibular cartilages, yellow: hyoid cartilages, pink/orange/green/purple/black: ceratobranchial arches (1st to 5th branchial arches, respectively).

The two dosage treatments of p.Glu69Lys mRNA produced varied phenotypes at 48 hours post-fertilization. At 100 ng/uL, no major abnormalities were observed in embryos that were treated with mutant p.Glu69Lys mRNA at 48 hours (bottom of Panel B). The phenotypes observed were almost the same as was observed for the wildtypetreated embryos. Interestingly, at 200 ng/uL, embryos that were treated with mutant p.Glu69Lys mRNA showed a plethora of phenotypes at 48 hours: severe yolk bolus, missing eyes, head severely pointing upward instead of downward as well as severe heart

oedema. Other major gastrulation problems observed included missing, bent or short tails as well as short or compact body (Plate 4.2A). Of further interest is the observation that almost all pharyngeal arches (Meckel's cartilage, palatoquadrate, hyosymplectic, ceratohyal, basihyal and ceratobranchial) were malformed for the 200 ng/uL mutant p.Glu69Lys-treated embryos, with the phenotype varying from less severe to complete loss of pharyngeal arches (Plate 4.1, bottom of Panel C). Finally, over 50% of 200 ng/uL mutant p.Glu69Lys-treated embryos which were viable at 48 hours postfertilization died by day 7 of embryonic development.

Percentage of abnormal embryos observed for each treatment at 100 ng/uL and 200 ng/uL were statistically significant (Table 4.14 and Figure 4.19). The level of significance was higher at 100 ng/uL ($p<0.0001$) than at 200 ng/uL ($p=0.0011$). These observations presuppose that the mutant mRNAs were responsible for the observed abnormalities in wildtype zebrafish embryos. These observations also buttress the pathogenic nature of these variants in humans.

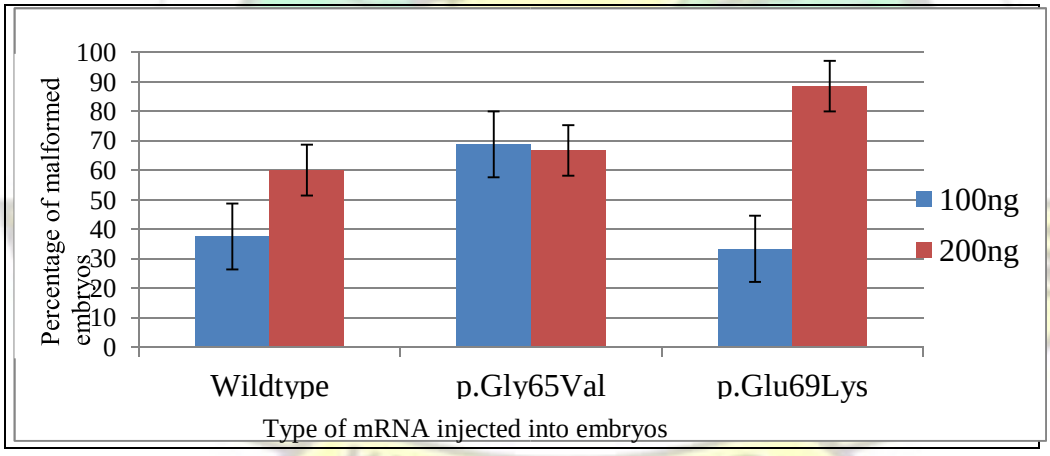


Figure 4.19: Percentage of abnormal embryos observed for each mRNA treatment

CHAPTER FIVE

5.0 DISCUSSION

5.1 Distribution and pattern of inheritance of orofacial clefts in Ghana

Recent advances in cleft care in Ghana have resulted in higher number of reported cases (Donkor et al., 2007a). Though no reliable national incidence rate is known, population-based and hospital-based epidemiological studies suggest OFCs may be among the most prevalent congenital abnormalities in Ghana. In a population-based study at Wudoaba cluster of villages in Ketu South District of the Volta Region, a prevalence of as high as 6.3 per 1000 live births was recorded (Agbenorku et al., 2011a). However, in a hospital-based study in Kumasi, a low incidence rate of 1.31 per 1000 live births was observed (Agbenorku et al., 2013), probably due to the relatively higher socio-economic status of the Kumasi population. Though the present study did not attempt to determine incidence of OFCs, it was interestingly observed that clefts were endemic in certain parts of Ghana. For example, many cases were reported from Manso and Asante Akyem Agogo areas of the Ashanti Region, Begoro area of the Eastern Region as well as Sefwi area of the Western Region. This observation suggests nonrandom distribution of clefts in Ghana, buttressing the assertion that ethnicity, geographical locations and environmental factors could influence the incidence of clefts (Gorlin et al., 2001).

Diverse OFC phenotypes were observed in the Ghanaian population (Table 4.1). No statistically significant differences were observed in the subphenotypes of CL, CP and CLP, though earlier studies suggested CP is a less common subphenotype (Agbenorku et al., 2011b; Donkor et al., 2007a). Furthermore, there was significant difference between the number of males and females affected with clefts, with higher number of females being affected. This observation is corroborated by a recent study in Kumasi that encompassed 11 hospitals (Agbenorku et al., 2013). The higher incidence of clefts in females emanated mainly from the about 1:2 males to females affected with CP in the study cohort; this CP

observation is also consistent with most literature (Dixon et al, 2011), except in some few studies (Agbenorku et al., 2013). Earlier studies in

KATH suggested the contrary: males were more affected with clefts than females (Agbenorku et al., 2011b; Donkor et al., 2007a), probably because CP is less visible at birth and may be recognized later in life. Also, no significant differences were observed between males and females affected with CL and CLP. This observation is contrary to literature, which suggests a 2:1 male to female ratio for these subphenotypes (Dixon et al., 2011). In harmony with reports elsewhere (Dixon et al., 2011), left unilateral clefts were as twice more common than right unilateral clefts in the present study.

Most OFC cases in Ghana were non-familial and had complex pattern of genetic inheritance. Only about 5% of case pedigrees were multiplex families. Interestingly, even within these multiplex families, no clear pattern of genetic inheritance was observed. For example, in pedigree GH20134972 of Figure 4.1, all females in that nuclear family are affected whereas no parent or male sibling is affected. Even within this pedigree, clefts differed in severity and laterality. Also, in pedigree GH20130786 of Figure 4.1, a mother with complete bilateral CLP transmits the exact phenotype to a male child while all other male siblings remain unaffected. For this particular pedigree, future generations must be studied before a conclusive pattern of inheritance can be obtained. Observations from these two pedigrees suggest to a large extent a nonMendelian pattern of inheritance of OFCs even in some familial cases and the urgent need for molecular data in pedigree analyses. It must be noted that these two pedigrees are nonsyndromic ones. About 95% of cleft cases were non-familial in that the proband is the first person to be born with cleft in these families. In the simplex pedigree GH20140578 of Figure 4.1, an identical twin presented with complete bilateral CLP as well as bilateral club foot, though the other twin

had no congenital anomalies. This discordance in occurrence of clefts even among identical or monozygotic twins, suggests an interaction between genetic and environmental factors in the pathogenesis of OFCs (Dixon et al., 2011). Moreover, of the 12 twins that were observed in this study, only the twin probands had OFCs whereas none of the other twin pairs had OFCs or any other congenital malformations. This evidence from twins also suggests that sporadic, somatic mutations that emanate after fertilization and embryo cleavage may significantly account for cleft aetiology (Jugessur et al., 2009). Thus, genetic mosaicism may be common among cleft patients. Corollary to these observations, a multifactorial pattern of inheritance (instead of a simple Mendelian inheritance) may account for cleft occurrence in the Ghanaian population, which is similar to observations made by Dixon et al. (2011).

The multifactorial pattern of inheritance proposed here is based on a number of observations made by this study that have also been reported elsewhere (http://www.uic.edu/classes/bms/bms655/lesson11.html#MODEL, 14/01/2015; Grosen et al., 2010). Firstly, most affected children (about 98%) had normal or unaffected parents. Secondly, the recurrence risk increased with the number of affected individuals in the family. For example, in pedigree GH20140648, a mother with CL gave birth to eleven children: both the proband and six other siblings were also born with CL. Moreover, in 10 multiplex families where clefts occurred within the nuclear family, 70% of these clefts were transferred from parents to probands whereas only 30% were clefts that occurred among siblings. In 5 out of 7 cases of parent to proband transmission, cleft was transferred from affected mothers to affected daughters. In one case, the exact phenotype of complete bilateral CLP was transferred from a mother to a male child whereas in the last pedigree, CL was transferred from a father to a female child.

Interestingly, the exact cleft subphenotypes in parents were transferred to the probands: 2 complete CLP, 4 complete CL as well as 1 complete CP were transmitted from parents to offspring. Thus, CP and CL/P, which are embryologically as well as aetiologically distinct (Dixon et al., 2011), did not segregate together in a family. Thirdly, recurrence risk increased with the severity of the abnormality: all 10 nuclear multiplex families observed in the present study had severe subphenotypes of clefts (complete CL, CLP and CP). Lastly, consanguinity slightly raised the recurrence risk of an affected family: familial OFCs were more common in parts of the Northern, Upper East, Upper West and Volta (Agbenoku et al., 2011a) Regions of Ghana where consanguinity was high.

Clinically, the vast majority of clefts in Ghana were nonsyndromic clefts; 88% of case pedigrees were nonsyndromic whereas only 12% were syndromic. This classification was based on clinical data only as there were no genetic data from genetic testing. This observation is statistically in line with global cleft distribution of 70% nonsyndromic cases and 30% syndromic cases (Dixon et al., 2011), though NSOFC was extremely high in the present cohort. Interestingly, molecular evidence or data (Table 4.11) gathered from the present study suggests that the percentage of syndromic clefts may actually be higher than was ascertained through just clinical data. These observations presuppose the urgent need to combine both clinical and molecular data before making any absolute cleft diagnosis. VWS was the most well-characterized and common syndromic form of cleft in the study cohort and this also is consistent with literature (Dixon et al., 2011). Though a higher number of Pierre Robin Sequence (PRS) cases were observed, it was difficult to cluster them as one syndrome since PRS has a heterogeneous genetic architecture (OMIM). Also worthy of note was the high number of cleft-club foot only phenotype. Though no syndrome was assigned to these multiple congenital

anomalies, it was observed that 80% occurred in males. This male dominance suggests an X-linked trait. Moreover, CP was observed more in syndromic pedigrees than other cleft subphenotypes, an observation that buttress earlier studies (Dixon et al., 2011).

Most case families were characterized by low socio-economic status and youthful parents. About 75% of fathers and 90% of mothers never went to school or dropped out of school at elementary level and therefore could neither read nor write. With regards to the 2010 Population and Housing Census of Ghana, general illiteracy level in Ghana is 26%. Therefore, illiteracy was found to be significantly ($p<0.0001$) associated with families with clefts. Moreover, over 90% of parents were low income earners, as a result of lack of employment or were in occupations with low income.

These observations are in harmony with those made elsewhere (Gorlin et al., 2001; Donkor et al., 2007b), that suggest that clefts distribution is influenced by low socioeconomic status. Lastly, clefts were mostly observed in families with parent within the age range of 21 to 45 years. Gross chromosomal abnormalities-related genetic defects are very predominant in children born to old-aged parents (Strachan and Read, 2011). Since most parents observed in the present study were within the age range of 21 to 45 years which is in harmony with earlier observations (Donkor et al., 2007b), old ageassociated chromosomal defects may not significantly account for cleft occurrence in the Ghanaian population.

5.2 Maternal risk factors contributing to OFC aetiology

Phenotypes usually emanate from interactions between genes and the environment (Strachan and Read, 2011). Assessment of maternal environmental exposure yielded many factors (Figure 4.5). However, not all these factors were statistically significantly associated with OFCs. This notwithstanding, each individual factor may play a particular

role in a family. For example, even though attempted abortion was not common among the cohort, a child with Moebius Syndrome was born to a mother who attempted abortion; the syndrome usually results from attempted abortion (OMIM), and this factor is significant in this particular family. Some of these environmental factors (such as pesticide, anticonvulsant drugs and cocaine usage) are also thought to alter DNA methylation patterns during foetal development (Crider et al., 2012).

The most striking environmental factor in cleft families was folate deficiency (Figure 4.5). Over 95% of mothers had poor dietary folate intake whereas about 86% of mothers either took folic acid supplements after the first trimester or had no folate supplements at all; this is complicated by delayed antenatal care (Figure 4.6). Folate is crucial for DNA methylation, repair and synthesis as well as histone modifications (Crider et al., 2012). These processes are overly crucial during embryogenesis and gametogenesis: except for few imprinted genes, DNA methylation patterns are globally erased very early in these processes and are reprogrammed later in development. This window of ontogeny gives room for environmental influences: lack of folate, which provides a one-carbon source for DNA methylation, may alter methylation patterns. This may lead to "epimutations", with attendant dysregulation of genes. Since DNA methylation also stabilizes the genome, alterations in methylation patterns could also lead to high level of mutations in genes, which are usually marked by methylated CpG islands (Crider et al., 2012). All these observations suggest that folate deficiency could contribute significantly to the occurrence of genetic mutations.

Maternal peri-conception folate levels affect the development of the foetus. During the Dutch hunger that followed World War II, it was observed that babies born to mothers, who were exposed to hunger before or during very early conception, were more prone to

congenital deformities. Methylation pattern of certain genes, such as IGF2 and INSIGF, were significantly altered in such babies. Such altered methylation patterns were even transferred to subsequent generations. Ironically, babies that were born to mothers that were exposed to hunger late in conception had very low probability of developing congenital defects. The gist of the study was that nutritional status of mothers during peri-conception was crucial (Cider et al., 2012). Human face and palate develop as early as by 10 weeks of gestation (www.indiana.edu/~anat550/hnanim/face/face.html, 17/01/2015). Interestingly, over 80% of case mothers in the present study never went for antenatal care or went for it after three months (Figure 4.6); folate supplementation is usually started by access to antenatal care. Folate supplementation after three months of gestation is less helpful for craniofacial development, since the face will be fully formed by then. Therefore, dietary folate deficiency, as well as late antenatal care attendance, may alter epigenetic regulation and stability of the genome, with attendant genetic mutations which may be detrimental to the developing foetus.

Hypoxia and use of traditional medicine (enemas and other herbal remedies) were also common among case mothers (Figure 4.6). No active cigarette smoker was observed among the case mothers. However, passive smoking and exposure to environmental smoke from refuse dumps, factories and use of firewood was common in about 30% of mothers. Low molecular oxygen levels, resulting from hypoxia, have been shown to alter the function of PHF8 gene, which is a histone lysine demethylase (Loenarz et al., 2009). Moreover, over 30% of mothers used traditional medicines, which were usually crude plant concoctions. Some of these crude plants preparation have been shown to cause

congenital deformities by disrupting some morphogenetic processes (Crider et al., 2012; Strachan and Read, 2011).

There are many lines of evidence supporting gene-environment interactions in cleft aetiology (Murray, 2002). In Han Chinese, rs17563 of BMP4 was found to interact with maternal passive smoke exposures and multivitamin supplementation as well as paternal high risk drinking and smoking in NSCL/P pedigrees (Jianyan et al., 2010).

Glutamate receptor, ionotropic, delta 2 (GRID2) and embryonic lethal, abnormal vision, Drosophila)-like 2 (ELAVL2) genes demonstrated evidence of interaction with maternal smoking in Europeans (Beaty et al., 2013). Interactions between maternal abortion history and TT genotype of rs2235373 of IRF6 have been observed in Han Chinese (Jia et al., 2009). Another study in 326 case-parent Chinese trios also showed interaction between maternal multivitamin intake and exposure to tobacco smoke (Wu et al., 2010). Moreover, there is lack of 100% concordance in occurrence of clefts for monozygotic twins (Kondo et al., 2002; Dixon et al., 2011; the present study), presupposing the intrauterine environment plays a huge role in cleft aetiology. These observations suggest both genes and environmental factors may be involved in the pathogenesis of clefts in the Ghanaian population.

5.3 Association studies: role of common variants in cleft pathogenesis

5.3.1 IRF6 is associated with NSCP in Ghanaians

IRF6 (rs642961) consistently and interestingly showed evidence of association with NSCP in all family-based association tests in the Ghanaian cohort (Tables 4.6 and 4.10). However, the IRF6 signal did not show up in the case-control association studies,

presupposing that family-based study designs are more powerful strategies for teasingout the IRF6 signal in the Ghanaian cohort. In European-derived populations, the minor A allele of rs642961 is a risk allele for NSCL/P, with the signal being strongest for NSCL only (Rahimov et al., 2008). The same allele has been shown to be associated with the most extreme subphenotype of NSOFCs, bilateral NSCLP, in some Asians (Kerameddin et al., 2015). The present study has for the first time shown evidence of association of rs642961 (IRF6) with NSOFCs in an African population, but interestingly with NSCP. Another significant observation is the fact that the minor A allele is protective in Africans (rs642961, p=0.03, OR=0.00, Table 4.6) whereas the major G allele is the risk allele for NSCP. In all races, A is the minor allele whereas G is the major allele (Table 4.4), with A being the risk allele for NSCL/P in both Asians and Caucasians (Rahimov et al., 2008). These dissenting observations may be due to population heterogeneity among the major populations of the world. These suggest the need to replicate GWAS signals in populations of diverse ancestries as well as the need to conduct GWAS for each of the major ancestries of the world, particularly for complex traits.

IRF6 is a gene that has consistently been associated with clefts in diverse populations, particularly those of Asian and European ancestries. Pathogenic variants in this gene were first observed in individuals with syndromic clefts, such as VWS and PPS, from populations of European ancestries (Kondo et al., 2002). Since then, aetiologic variants in this gene have been found in VWS and PPS patients from diverse ethnicities, including Africa (de Lima et al., 2009; Butali et al., 2014; Wang et al., 2003; Wu-Chou et al., 2013). However, aetiologic exonic variants are not always observed in this gene in VWS and PPS patients; only about 70% of VWS and PPS cases have

aetiologic mutations in this gene (Fakhouri et al., 2014).

Subsequent studies on IRF6 established that this gene is even associated with NSCL/P in populations of European and Asian ancestries, with the classic SNP being rs2235371 (Zuchero et al., 2004). This notwithstanding, this SNP has consistently shown evidence of association with NSCL/P, mostly in Asian populations, and to a lesser extent, European-derived populations, but not Africans (Sun et al., 2015; Park et al., 2007; Jugessur et al., 2008). The rs2235371 SNP is not known to exist in the African population (http://browser.1000genomes.org/index.html, 06/08/2015). A functionally validated SNP, rs642961, has shown that it is a universal marker that can be used to test for IRF6's association with NSOFCs, howbeit with population-specific subphenotypes and risk alleles. This variant disrupts an enhancer element that is a binding motif for AP2α transcription factor (Rahimov et al., 2008). Asian and European-derived populations have consistently shown evidence of association of this SNP with NSCL/P (Shi et al., 2011; Zhou et al., 2013; Mostowska et al., 2010; Pegelow et al., 2014). However, some studies could not replicate this association (Nouri et al., 2014), probably due to population heterogeneity and disregard for cleft subphenotypes, as it has been shown that this SNP is associated with the most severe form of cleft, complete bilateral NSCLP, in some Asian populations (Kerameddin et al., 2015). Earlier attempts to replicate association between IRF6 and NSCL/P in Africans from Kenya, Congo and Nigeria were not successful (Wheatherley-White et al., 2011; Figueiredo et al., 2014; Butali et al., 2011).

Interestingly, the present study has successfully replicated this association in the African population, with Ghana driving the IRF6 signal. Here, it has been shown that IRF6's association with NSOFCs among Africans resides in the functional enhancer variant, rs642961, but not rs2235371 which has proven to be a strong marker in Asians,

or rs861020 which was revealed by a GWAS (Beaty et al., 2010). This is of molecular pathogenesis interest because rs642961 variant alters the level of expression of IRF6 (Rahimov et al., 2008) and may contribute hugely to the aetiology of NSOFCs. Though this SNP is aetiologic, it may also be in LD with other aetiologic variants in the gene. For example, direct IRF6 sequence analysis of 184 NSOFCs case probands by this study revealed aetiologic variants in some of these patients (Table 4.11).

IRF6 play crucial role in embryology and alterations in its function have been proven to result in clefts (Kondo et al., 2002; Ingraham et al., 2006; Iwata et al., 2013; Plate 4.1). It has a cardinal role in keratinocyte proliferation-differentiation switch which is linked to normal epidermal growth. IRF6 is also involved in regulating mammalian epithelial cell proliferation as well as regulation of the transcription of WDR65 gene (http://genome.ucsc.edu, 22/01/2015). Significantly, IRF6 that is associated with NSCP in the Ghanaian population has also been shown to interact with environmental factors like folate deficiency and maternal smoking to initiate cleft pathogenesis (Wu et al., 2010; Jia et al., 2009). The present study also witnessed high folate deficiency among mothers of case probands (Figure 4.5), presupposing that folate deficiency could contribute to the occurrence of the aetiologic variants in the enhancer variants in IRF6, since folate is needed to methylate CpG islands that are associated with genes in the genome (Crider et al., 2012).

5.3.2 ARHGAP29 is mostly associated with NSCLP in Africans

Another SNP that consistently demonstrated evidence of association with NSCL/P is rs560426, situated in an intron of ABCA4 (Tables 4.3, 4.5, 4.7 and 4.9). Subphenotype analyses also presupposed that this SNP was associated with the most

severe form of NSOFCs, NSCLP, in Africans. Earlier studies suggested that ARHGAP29, but not ABCA4, may be the aetiologic gene at 1p22.1 locus, though rs560426 resides in an intron of ABCA4. The rs560426 SNP was first demonstrated to be associated with NSCL/P in a GWAS that comprised cohorts of Asian and European ancestries, though Asians showed the strongest signal (Beaty et al., 2010). In the study by Beaty et al (2010), C was the minor and risk allele for rs560426 among Europeans and Asians (Table 4.4). In harmony with studies in Europeans and Asians, the present study has also demonstrated that C is both the minor and risk allele for NSCLP in Africans in TDT analyses (Ghana – $p=7.66\times10^{-3}$, OR=2.60, 95% CI=1.25-5.39 and combined African cohorts - $p=8.55\times10^{-3}$, OR=1.92, 95% CI=1.17-3.14). Subsequent expression study in the same GWAS showed that ABCA4 did not exhibit craniofacial expression pattern (Beaty et al., 2010). This suggested that though rs560426 was situated in ABCA4, it may be in LD with aetiologic variants in nearby genes. Expression studies of ARHGAP29, a nearby gene, showed that it was the probable cleft candidate gene at this locus. Aetiologic rare variants were also observed in this gene by the same study

(Leslie et al., 2012).

Various replication studies have also confirmed ARHGAP29 as the aetiologic gene linked to rs560426 (Fontoura et al., 2012; Yuan et al., 2011). In a study involving Hispanics and Non-Hispanics, association was established between SNPs in and around ARHGAP29 and NSCL/P (Letra et al., 2014). Using targeted sequencing and TDT in a large number of trios from diverse population, rs560426 was again shown to be strongly associated with NSCL/P in Asian cohorts. Another SNP, rs77179923, also demonstrated strong association. Other aetiologic rare variants were also observed. These two SNPs were observed to be located in regions that harbour regulatory elements, of particular

interest is a craniofacial enhancer (Leslie et al., 2015a). This suggests that rs560426associated regulatory element probably regulates the expression of ARHGAP29, by acting as a craniofacial enhancer. Variants in rs560426, therefore, disrupt this enhancer activity. Other direct DNA sequence analyses also support ARHGAP29's role in craniofacial development (Chandrasekharan and Ramanathan, 2014; Butali et al., 2014a). Interestingly, a SNP in ARHGAP29, rs138751793, which was identified by the present study through direct DNA sequencing of this gene, showed a trend towards association in the Ghanaian cohort in DFAM analysis (Table 4.10).

5.3.3 VAX1 may be involved in the pathogenesis of all NSOFCs

VAX1 demonstrated evidence of association with NSOFCs in all association studies (Tables 4.3, 4.4, 4.5, 4.6, 4.8 and 4.10). The minor A allele for rs7078160, as has been reported elsewhere (Mangold et al., 2010), was also the risk allele in Africans, with subphenotype analyses suggesting this allele was largely associated with NSCL/P (NSOFC, p=0.03, OR=1.58, 95%CI=1.04-2.40 in Table 4.6). Another VAX1 SNP, rs4752028, was unconventionally and particularly associated with NSCP among Africans (Ghana, Table 4.6 – p=0.05, OR=2.25, 95%CI=0.98-5.18 and combined

African cohorts, Table 4.8 – p=0.03, OR=2.08, 95%CI=1.07-4.03). In the initial study by Beaty et al (2010), C was the minor and risk allele for rs4752028 in NSCL/P cases among Europeans and Asians (Table 4.4). However, in the present study, T was the minor and risk allele for rs4752028 in NSCP cases among Africans.

VAX1 is crucial for regulation of body development and morphogenesis (http://genome.ucsc.edu, 22/01/2015) and alterations in its structure could lead to malformations, including clefts. It is a member of the most conserved genes in vertebrates,

the HOX homeobox genes. Vax1 knockout mice develop many craniofacial deformities, including CP (Hallonet et al., 1999, Zhao et al., 2010). VAX1 is also an upstream effector of WNT signaling through dnTCF12, by activating the transcription of dnTCF12 which is a dominant negative isoform of TCF12 which up-regulates WNT signaling. The dnTCF12 represses WNT signaling and mutant VAX1 may abolish this function, leading to hyper-activation of WNT signaling with attendant craniofacial malformations (Vacik et al., 2011).

Many lines of evidence support a role of VAX1 in cleft aetiology in humans. Three successive GWAS and a meta-analysis all implicated VAX1 at 10q25 as a cleft candidate gene (Mangold et al., 2010; Beaty et al., 2010; Ludwig et al., 2012; Sun et al., 2015). In 70 patients with various congenital ocular defects, aetiologic rare variant was found in an individual with many craniofacial anomalies, including CLP (Slavotinek et al., 2012). Many aetiologic rare variants have also been reported in Asians (Butali et al., 2013) as well as European-derived populations (Nasser et al., 2012). NSCL/P's association with rs7078160 has also been successfully replicated in targeted researches that involved Europeans (Nikopensius et al., 2010; Mostowska et al., 2012a) and Mesoamericans (Rojas-Martinez et al., 2010). Earlier attempts to replicate VAX1 signal in Kenyan, Congolese and Nigerian cohorts were not successful (Weatherley-White et al., 2011; Butali et al., 2011; Figueiredo et al., 2014), probably due to population heterogeneity or small sample size. Interestingly, the present study has successfully replicated the VAX1 signal in Africans, with Ghana mainly driving the signal. Moreover, direct DNA sequence analyses of 184 NSOFC patients by the present study revealed aetiologic variants in VAX1 (Table 4.13).

5.3.4 The 8q24 locus is associated with NSCL/P in Africans

The 8q24 (rs987525) locus showed the strongest association ($p=1.22\times10^{-03}$, OR=0.81) with NSCL/P in case-control analysis of the African cohort (Table 4.4). In the combined NSCL/P cohort, the C allele was the associated allele and it was protective in Africans, presupposing the A allele was the risk allele. However, subphenotype analyses of the NSCL/P cohort demonstrated that both C and A alleles could be risk alleles depending on the subphenotype (Table 4.5). The C allele was observed to be the risk allele for NSCL ($p=5.38\times10^{-03}$, OR=1.28) whereas the A allele was the risk allele for NSCLP (p=0.01, OR=0.80). Subphenotype analyses of the NSCL/P cohort further suggested it may be largely important in NSCL pathogenesis, buttressing earlier observations (Beaty et al., 2010). Among Africans, the C allele of rs987525 is the minor allele whereas A is the major allele; in Asians and Europeans, A is the minor allele whereas C is the major allele (www.1000genomes.org; Table 4.4). Irrespective of these differences in minor alleles among Europeans and Africans, the observation that A is the risk allele for combined NSCL/P phenotype at this locus in Africans is in harmony with earlier studies (Birnbaum et al., 2009; Grant et al., 2009; Mangold et al., 2010; Beaty et al., 2010; Ludwig et al., 2012). This notwithstanding, observations made by the present study suggest the actual risk locus may be in LD with either the C or A allele of rs987525, depending on the NSCL/P subphenotype (NSCL or NSCLP). Fine mapping of the African Haplotype (which is smaller in the 8q24 region, Appendix P) will help identify the risk loci.

Observations from this work further buttress the notion that the varied ethnic impact of the rs987525 locus seems to be directly related to its minor allele frequency

(MAF) in various populations. This signal is very strong in populations of European ancestry whereas Asian-derived populations show marginal association for this locus (Beaty et al., 2010). Various replication studies have confirmed that 8q24 signal is European ancestry-specific (Beaty et al., 2013). Intriguingly, the present study has shown that this locus is strongly associated with NSCL/P in the African population, though earlier attempts to replicate this signal were not successful (Wheatherley-White et al., 2011; Butali et al., 2011), probably due to small sample sizes. Though this locus is a gene desert, current epistatic studies suggest that the 8q24 (rs987525) locus interact with WNT5B among Europeans (Li et al., 2014). This suggests there could be regulatory RNAs (such as miRNA, lncRNA, etc.) as well as enhancer and promoter elements at this locus that alter the level of expression of various distant genes. The interaction between this locus and WNT5B is very crucial, as WNT genes are very crucial for craniofacial development (Vacik et al., 2011). Current evidence also suggests that the 8q24 (rs987525) window harbors very remote cis-acting craniofacial enhancer elements that regulate the expression of oncogenic MYC in the developing face; perturbation of this regulatory network leads to craniofacial dysmorphologies, including sporadic CL/P, in mice (Uslu et al., 2014).

5.3.5 NSCP and NSCLP pathogenesis may be influenced greatly by MSX1 in Africans

MSX1 (rs115200552) showed evidence of association (p=0.01, OR=1.81) with NSCP in case-control association studies in the African cohort and a trend towards association (p=0.06, OR=1.72) in the isolated Ghanaian cohort (Tables 4.3 and 4.4). Furthermore, the trend shown suggests that with larger sample size, evidence of association may be achieved for a Ghanaian cohort. DFAM analyses (Table 4.10) also suggested that rs115200552 may be associated with NSCLP among Africans (p=0.04).

The CA microsatellite marker and rs12532 of MSX1 have mostly been associated with NSCL/P in all major populations of the world (Souza et al., 2013; Suazo et al., 2010b; van den Boogaard et al., 2008 Rafighdoost et al., 2013; Kim et al., 2013; Seo et al., 2013), but not Africans. Earlier study involving Africans from Nigeria established association between NSCL/P and p.Ala34Gly, a missense variant in exon 1 of MSX1.

Direct DNA sequence analysis of this gene in 184 NSOFC cases from Ghana showed that p.Ala34Gly (rs36059701) had a MAF of about 9%. This variant is now coated as p. Ala40Gly because of the fine-tuning of the human genome after the release of the initial draft in 2005. In 1000 Genomes, this variant also has an African MAF of 9%, suggesting this variant is not over-transmitted in NSOFC cases from Ghana. Though the CA microsatellite was not genotyped in the present study, genotyping of rs12532 did not show any evidence of association. Interestingly, rs115200552, which was identified through direct sequencing of MSX1 in 184 NSOFC cases from Ghana, showed evidence of association. This variant, located at the 3'UTR of MSX1 is predicted by RegulomeDB to alter binding site for EZH2, which is a histone-lysine N-methyltransferase enzyme. Haploreg also predicts (Appendix Q) that rs115200552 is situated in an enhancer element where POL2 binds, suggesting this SNP is functional and may regulate the level of expression of MSX1. Thus, this SNP may be responsible for the cleft phenotype in individuals harbouring the aetiologic and risk C allele.

Role of MSX1 in cleft aetiology, as well as odontogenesis, is undisputable. Many gene expression and knockout studies in mouse have shown that MSX1 mutant mouse show an array of craniofacial anomalies, including CL/P and tooth agenesis (Lu and Hu, 2009). Direct DNA sequence analyses of this gene in NSOFC with tooth agenesis and autosomal dominant tooth agenesis cases have revealed many aetiologic variants in these

individuals (De Muynck et al., 2004; Salahshourifar et al., 2011; Liang et al., 2012). Interestingly, the present study also witnessed a missense variant, p.Pro73Leu, in three NSCL/P cases from Ghana (Table 4.13). Though the p.Pro73Leu variant is in 1000 Genomes and other public databases, the minor allele frequency (MAF) is 0.0046, which means this variant is extremely rare.

5.3.6 AXIN2 is highly associated with NSCLP in Africans

In DFAM analyses (Table 4.10), rs3923086 of AXIN2 demonstrated evidence of threshold association with NSCLP among Africans ($p=2.88\times10^{-03}$). Another SNP tagging AXIN2, rs7224837, showed a trend towards association with NSCP (Table 4.8) and NSCLP (Table 4.9) in TDT analyses. For rs3923086, A was the minor and risk allele for Europeans (Table 4.4) whereas for Asians and Africans, C is the minor allele. TDT analyses (Table 4.9) suggested the associated minor C allele of rs3923086 was protective, demonstrating that the A allele may be the risk allele in all races.

AXIN2 down-regulates β-catenin in the Wnt signaling pathway, a pathway that is crucial for craniofacial development (http://genome.ucsc.edu, 08/08/2015). AXIN2 has been implicated in the aetiology of NSOFCs in multiple populations, except Africans, with rs3923086 demonstrating association with NSCLP among Asians (Letra et al., 2012b). Other studies (Mostowska et al., 2012b; Araujo et al., 2015) have corroborated the association between AXIN2 and NSCL/P. Here, it has been demonstrated that rs3923086 is also conventionally associated NSCLP among Africans. AXIN2 also exhibited craniofacial expression pattern and co-localization with IRF6, presupposing they exhibit their effects contemporaneously (Letra et al., 2012b). These observations suggest AXIN2

may be relevant in cleft aetiology. Intriguingly, the present study, has for the first time, replicated the association of this gene with NSCLP in African populations.

5.3.7 MAFB is mostly associated with NSCL/P in Ghanaians

MAFB demonstrated association with NSOFCs in the Ghanaian cohort in DFAM analyses (Table 4.10; rs17820943, p=0.02; rs13041247, p=0.02; rs11696257, p=0.04) and showed a trend towards association in TDT analyses (Table 4.6; rs17820943 and rs13041247, all with p=0.08, OR=1.54, 95%CI=0.95-2.47). Subphenotype analyses also suggested this gene may be largely associated with NSCL/P but not NSCP, as has been reported elsewhere (Beaty et al., 2010). Interestingly, the minor alleles of rs17820943 (T) and rs13041247 (C) are protective in Asians and Europeans (Table 4.4) whereas the same alleles are risk alleles in Africans. MAFB may act as a transcriptional activator or repressor, depending on the tissue and is extremely crucial for many cellular functions (https://genome.ucsc.edu, 12/06/2015).

The MAFB gene has demonstrated strong evidence of association with NSCL/P in populations of Asian ancestry (Beaty et al., 2010; Mi et al., 2014; Pan et al., 2011; Cheng et al., 2012), with lesser association being achieved for populations of European ancestries (Yuan et al., 2011; Fontoura et al., 2012). However, this study presents the first ever association of this gene with any African population. Larger numbers of caseparent trios will be needed to increase the power of association of this particular signal in the Ghanaian population, as larger data sets are required for genes of small effects (Dixon et al., 2011). This notwithstanding, this gene is crucial for craniofacial development, as it interacts with WNT5B, a member of the WNT signaling which are overly crucial for craniofacial development (Li et al., 2014).

5.3.8 PAX7 and MTHFR are largely associated with NSCL among Africans

A SNP in PAX7 (rs742071) demonstrated evidence of association with NSCL/P, largely NSCL (Tables 4.4, 4.5, 4.9 and 4.10) in the combined African cohort only. This notwithstanding, direct DNA sequence analysis of PAX7 revealed that the minor C allele of rs4920523G>C, located in an intron of this gene, was over-transmitted ($p<0.0001$) in NSOFC patients in the Ghanaian cohort. Haploreg predicts (Appendix Q) that rs4920523 is situated in an exhancer element where SUZ12 binds and the C allele disrupts binding motif for Pax-2, STAT and p300, presupposing this SNP may alter the level of expression of PAX7 gene. This observation suggests that different SNPs or markers may be needed to tease-out certain associations between common variants and NSOFCs in different populations or ancestries.

PAX7 has been shown to be associated with NSCL/P in both GWAS (Beaty et al., 2010; Ludwig et al., 2012) and replication studies (Butali et al., 2013; Leslie et al., 2015a; Sull et al., 2009). Moreover, aetiologic functional rare variants in PAX7 were observed in some of these studies. PAX7 plays a vital role in neural crest development as well as the expression of vital neural crest markers in chicken and mammals.
Perturbations of these functions lead to craniofacial anomalies, including CL/P (Murdoh et al., 2012; Wilkie and Moriss-Kay, 2001; Basch et al., 2006). Interestingly, the present study also detected aetiologic variants in the DNA-binding domain of PAX7 in NSCL/P cases from Ghana (Table 4.13) through direct DNA sequencing.

MTHFR (rs1801131) exhibited association with NSCL (Ghana, $p=0.03$, OR=0.33, 95%CI=0.12-0.92; combined African cohorts, $p=0.04$, OR=0.45, 95%CI=0.20-0.99) in TDT analyses. For rs1801131, though C is the minor allele in all races, C has been

suggested to be the risk allele in Asians and Europeans (Table 4.4). The present study, howbeit, suggests the C allele is protective in Africans and it is rather the A allele that is the risk allele for NSCL in Africans. MTHFR is crucial for folate metabolism (Crider et al., 2012) and variants in this gene have been associated with NSCL/P (Jagomagi et al., 2010). The C677T (rs1801133) SNP of MTHFR, but not A1298C (rs1801131), has largely been associated with reduced risk for NSCL/P in Asians (Martinelli et al., 2016; Pan et al., 2015; Zhao et al., 2014) and to some extent, in European-derived populations (Estandia-Ortega et al., 2014; de Aguiar et al., 2015), though not all studies (Sozen et al., 2009) replicated the association. Interestingly, the present study has demonstrated that MTHFR is, unconventionally, largely associated with NSCL among Africans and that it is the A1298C (rs1801131) SNP that confers a reduced risk to NSCL. The present study has also implicated folate deficiency as a major environmental risk factor for OFCs in Ghana (Figures 4.5 and 4.6). These two folaterelated observations suggest that folate deficiency and dysfunctional folate metabolism may play crucial roles in OFC aetiology.

5.3.9 Other genes associated with NSOFCs in Africans

In both TDT (Table 4.7; rs4783099, p=0.03, OR=2.08, 95%CI=1.05-4.15) and DFAM (Table 4.10; rs4783099, p=0.03), CRISPLD2 showed threshold association with NSCLP and NSCL/P, respectively, for the Ghanaian cohort. The present study has also buttressed earlier observations (Table 4.4) that the T minor allele of rs4783099 is both the minor and risk allele for NSCL/P cases. Case control analyses also demonstrated that this gene was nominally associated (p=0.02, OR=0.74) with NSCP in the African cohort (Table 4.4), with the minor T allele being dissentingly protective in NSCP cases.

Genotyping of SNPs of CRISPLD2 in Caucasian and Hispanic multiplex and simplex families showed that rs1546124 was significantly associated with NSCL/P in all these ancestries. Expression studies demonstrated that CRISPLD2 was expressed in the mandible, palate and nasopharynx during murine craniofacial development (Chiquet et al., 2007). Both CRISPLD2 and CRISPLD1 interact with genes in the folate pathway to initiate cleft pathogenesis (Chiquet et al., 2011). Also, in a study involving 200 NSCL/P cases and 180 controls from China, significant association was observed between NSCL/P and CRISPLD2 SNP rs1546124 (Mijiti et al., 2015). CRISPLD2 protein promotes matrix assembly and binds to herparin, dermatan sulphate and chondroitin sulphate (https://genome.ucsc.edu, 08/08/2015). These observations suggest that this gene may play vital role in "zipping" together the various parts of the developing craniofacial region.

DYSF is a promising candidate gene for NSCL/P among Africans. In DFAM analyses, DYSF (rs4332945) gained threshold association (p=0.04) with NSCL/P, with subphenotype analyses suggesting this loci may be largely important in NSCL aetiology among Africans (Table 4.10; rs4332945, p=0.02). DYSF encodes a skeletal muscle protein that is associated with sarcolemma. It plays a role in muscle contraction and harbours C2 domains that are involved in calcium-mediated membrane fusion events, presupposing it may be crucial for membrane repair and regeneration

(https://genome.ucsc.edu, 08/08/2015). These observations suggest this protein may be crucial for orbicularis oris muscle development, alterations of which could result in OFCs.

TULP4 (rs651333) demonstrated nominal association (p=0.04, OR=1.29) with NSCP in case-control meta-analyses of the combined African cohorts. The gene product may act as a substrate-recogition component of a SCF-like ECS (Elogin-Cullin-SOCSbox protein) E3 ubiquitin ligase complex which modulates the ubiquitination and subsequent

proteosomal degredation of target proteins (http://genome.ucsc.edu, 20/08/2016). Earlier community-based study involving Phillipines showed that the minor allele of this SNP was over-transmitted in NSCL/P but not NSCP cases (Vieira et al., 2015). The present study has shown that among Africans, the minor T allele was unconventionally associated with NSCP with the T allele being the risk allele whereas earlier studies suggested the C allele as the risk allele (Table 4.4). A study which anyalysed copy number variants (CNVs) due to deletions and duplications in the genome also implicated TULP4 in the aetiology of OFCs (Conte et al., 2016).

Some loci may be relevant to NSOFCs aetiology largely in the Ghanaian population. THADA (rs7590268) exhibited threshold association (p=0.04) with NSCLP in DFAM analyses; the same locus also demonstrated a trend towards association with NSCLP (p=0.06, OR=2.11, 95%CI=0.96-4.67) in TDT analyses. THADA has been implicated in cell death receptor pathway and apoptosis as well as the aetiology of NSOFCs and other malformations (Mangold et al., 2010; Ludwig et al., 2012; Beaty et al., 2013). In DFAM analyses, SPRY2 (rs9574565) showed evidence of threshold association (p=0.05) with NSCL/P among Ghanaians; the same loci also demonstrated a trend towards association (p=0.09, OR=1.47, 95%CI=0.94-2.30) with NSCL/P in Ghanaians in TDT analyses. SPRY2 showed evidence of association with NSCL/P in a GWAS (Mangold et al., 2010), a meta-analysis (Ludwig et al., 2012) as well as a replication studies (Jia et al., 2015). In DFAM analyses, SLC25A24 (rs6677101) showed threshold association (p=0.03) with NSCP among Ghanaians; the same loci also showed a trend towards association (p=0.09, OR=1.90, 95%CI=0.88-4.09) with Ghanaian NSCP cases in TDT analyses. SLC25A24 encodes a solute carrier protein that transports ATPMg, exchanging it for phosphate in the mitochondria (https://genome.ucsc.edu,

08/08/2015). The risk alleles observed by the present study for both THADA (rs7590268) and SPRY2 (rs9574565) are the same as has been reported in other races (Table 4.4), although the minor allele for SPRY2 (rs9574565) vary between Africans and other races. For SLC25A24 (rs6677101), G is the minor and risk alleles for NSCL/P in Europeans and South Asians whereas T was observed to be the minor and risk alleles for NSCP among Africans.

FOXE1 never showed any evidence of association with NSOFCs among Africans in various SNP genotyping studies. However, direct DNA sequence analysis (Table 4.13) of this gene in 184 NSOFC cases from Ghana showed that a SNP, rs775448711 (chr9:100,616,712, C>A leading to p.Ala172Ala) was over-transmitted in NSOFC cases (p=0.0181). This association trend suggests other SNP(s), other than the genotyped SNPs, may play a significant role in the aetiology of clefts in the Ghanaian population, probably due to population heterogeneity. Many lines of evidence support a role of FOXE1 in cleft aetiology. Various linkage and association studies in major populations of the world, except Africans, have implicated this gene in the aetiology of both NSCL/P and NSCP, which are thought to be embryologically and epidemiologically distinct subphenotypes of clefts (Marazita et al., 2009; Moreno et al., 2009; Ludwig et al., 2014; Lennon et al., 2012; Nikopensius et al., 2011), though a study by Letra et al (2010) could not replicate the association. Direct DNA sequence analyses have also revealed some potentially aetiologic variants in this gene in NSCL/P and NSCP cases (Srichomthong et al., 2013). Furthermore, this gene is mutated in individuals with Bamforth-Lazarus syndrome (BLS), a condition that present with many craniofacial abnormalities, including CP (http://genome.ucsc.edu, 08/04/2015). Finally, mouse and zebrafish transgenesis have shown that FOXE1 is preferentially expressed in the oral

epithelium and branchial arches, presupposing FOXE1 is relevant for normal craniofacial development (Lidral et al., 2015). Ironically, direct DNA sequence analysis of 184 NSOFCs cases from Ghana (Table 4.13) further revealed potentially aetiologic missense mutations in FOXE1.

BMP4 (rs17563) showed threshold association (p=0.05) with NSOFCs in DFAM for the Ghanaian cohort (Table 4.10). For rs17563, T is the minor and risk allele in Europeans (Table 4.4) whereas C is the minor allele for Africans and Asians. BMP4 is highly expressed at the site of epithelial-mesenchyme interaction and is relevant for tooth and palate development (Lu and Hu, 2009). An array of ocular and bone defects, as well as microform clefts have been associated with variants in BMP4 (http://genome.ucsc.edu, 22/01/2015). Direct BMP4 sequence analyses in Asian and European-derived populations have shown that variants in this gene are associated with microform clefts, orbicularis oris muscle defects as well as overt CL/P (Suzuki et al., 2009). In Han Chinese, rs17563 of BMP4, which showed threshold association with NSOFCs in the present study, have been shown to interact with maternal and paternal environmental risk factors, such as vitamin deficiency and smoking, to initiate cleft pathogenesis (Jianyan et al., 2010). Two successive studies in the Brazilian population have also shown that this SNP is associated with NSCL/P (Antunes et al., 2013; Araujo et al., 2012). Promoter variants in this gene, that creates new binding sites for various transcription factors, have also been associated with NSCL/P in Chilean cohorts (Suazo et al., 2011; Suazo et al., 2010a). All these lines of evidence support a role of BMP4 in the aetiology of NSCL/P in the Ghanaian cohort.

NOG (rs17760296, p=0.04, OR=1.74) showed nominal association with NSCP (Tables 4.4) in case-control analyses whereas another SNP, rs227731, tagging the same

gene exhibited a trend towards association with NSCP in DFAM analyses (Table 4.10). A GWAS (Mangold et al., 2010) and targeted sequencing of NSCL/P families from major populations of the world, except Africans, as well as gene expression studies in mice (Leslie et al., 2015a), have associated this gene with NSCL/P. NOG is also a component of BMP4 signaling pathway. However, direct sequencing of this gene in European NSCL/P cases yielded no aetiologic variants (Chawa et al., 2014). This notwithstanding, this gene may be crucial for craniofacial development due to its expression pattern and interactions with BMP4 pathway.

In DFAM analyses, GREM1 showed a trend towards association with NSCLP in both the Ghanaian (p=0.09) and combined African (p=0.06) cohorts (Table 4.10). This gene is a known BMP antagonist and plays crucial roles in regulating body patterning, tissue differentiation, limb bud outgrowth and organogenesis (http://genome.ucsc.edu, 08/08/2015). This gene yielded a suggestive association in a GWAS (Mangold et al., 2010). However, two replication studies have shown that this gene is associated with NSCL/P among Europeans (Chawa et al., 2014; Mostowska et al., 2015), though direct DNA sequence analyses is yet to reveal aetiologic variants in this gene. Considering the crucial role BMP4 plays in craniofacial development, GREM1 as an antagonist of this gene may be likely involved in craniofacial anomalies, such as NSOFCs.

NTN1 demonstrated a trend towards association (rs8081823, p=0.08, OR=0.88) with NSCL/P in the African cohort (Table 4.4). Subphenotype analyses also demonstrated that another SNP, rs8069536, tagging NTN1 was nominally associated (p=0.04, OR=1.20) with NSCLP (Table 4.5). The other NTN1 SNP, rs8081823, also exhibited nominal association (p=0.05, OR=0.83) with NSCLP. Two GWAS (Beaty et al., 2010; Sun et al.,

2015), as well as various replication studies (Beaty et al., 2013; Leslie et al., 2015a) found association between NTN1 and NSCL/P.

All genes discussed under this paragraph have been associated with NSCL/P in all major populations of the world, except Africans. The present study suggests that with greater number of NSCL/P case-parent trios, these signals could be "teased-out" of the African population.

5.3.10 Parent of origin effects may not be significant among Africans

No significant parent of origin (POO) effect was observed for any of the SNPs that showed association with NSOFCs. However, rs16260 of CDH1 showed evidence of maternal imprinting whereas rs8001641 of SPRY2 demonstrated paternal overtransmission. POO effects are hard to determine though it has long been suspected that some paternally or maternally imprinted alleles may increase the risk of NSCL/P.

A recent analysis of large dataset by Garg et al (2014) did not yield any genomewide significance for POO, though a suggestive association was obtained for some genes. This notwithstanding, a maternally imprinted POO effects was observed for rs7078160 of VAX1 in Mongolian and Japanese NSCL/P cases (Butali et al., 2013). In a FOXE1 study, maternal POO effects were observed for Colombian and Philippine cohorts for both CL/P and CP whereas paternal POO effects were observed for Caucasian data (USA and Scandinavian data sets), possibly indicating different roles of these alleles in different populations (Moreno et al., 2009).

5.3.11 Epistatic interactions may play crucial roles in cleft aetiology

The present study observed three epistatic or gene-gene interactions, all with p=0.02. Firstly, The ABCA4 locus (rs560426) on human chromosome 1 interacted with a

gene desert (rs2674394) on human chromosome 6. Though the the rs2674394 locus harbour no known protein-coding gene, there could be regulatory RNAs (such as miRNA, lncRNA, etc.) that may regulate the expression of ARHGAP29 through the enhancer element for ARHGAP29 which is located in an intron of ABCA4 that harbours the rs560426 SNP (Lesli et al., 2015a). Earlier studies (Lesie et al., 2012) had shown that ARHGAP29, but not ABCA4, was the probable NSOFC aetiologic gene at the rs560426 locus.

Secondly, DYSF (rs2303596) also showed evidence of epistasis with AXIN2 (rs3923086). DYSF encodes a skeletal muscle protein that has been associated with sarcolemma. It plays a role in muscle contraction and harbours C2 domains that are involved in calcium-mediated membrane fusion events, presupposing it may be crucial for membrane repair and regeneration (https://genome.ucsc.edu, 08/08/2015). AXIN2 down-regulates β-catenin in the Wnt signaling pathway, a pathway that is crucial for craniofacial development (http://genome.ucsc.edu, 08/08/2015). Since DYSF protein is a structural protein whereas AXIN2 protein is a signalling or polarizing protein, it remains elusive how these two genes interact. However, it is possible that AXIN2 protein may signal the recruitment of DYSF protein onto the sarcolemma of the orbicularis oris muscle, which is crucial for the formation of the human lip.

Finally, a SNP tagging NTN1 (rs8069536) interacted with all the geneotyped SNPs of MAFB (rs17820943, rs13041247 and rs11696257), suggesting this epistatic interaction may be strong. NTN1 has been implicated in axon guidance and cell migration during human development. NTN1 may act as a survival factor through its association with some receptors to hinder the initiation of apoptosis, suggesting this gene plays a crucial role in tumorigenesis by regulating apoptosis (https://genome.ucsc.edu, 14/08/2016). MAFB, a

transcription factor, acts as a transcriptional activator or repressor to many genes, including ETS-1 and glucagon promoter. MAFB may also function as an oncogene or a tumor suppressor gene, roles which are influenced by the type of cell (https://genome.ucsc.edu, 12/06/2015). As a transcription factor, MAFB may interact with NTN1 by activating or repressing its expression. Moreover, since both MAFB and NTN1 are involved in tumorigenesis, it is possible both genes may regulate cellular migration, survival and death during fusion events in craniofacial development.

Some earlier studies support a role of epistasis or gene-gene interactions in cleft aetiology. A SNP in NTN1 (rs11650357) has demonstrated evidence of possible interactions with three SNPs in IRF6, the most significant SNP being rs6685182 (Beaty et al., 2013). Tooth agenesis studies have shown that IRF6 interacts with TGFα, since tooth agenesis is often observed in individuals with NSCL/P (Letra et al., 2012a). In a NSCL/P study, epistatic interactions were observed between MSX1 and BMP2, FGF1 and PVRL2 as well as COL2A11 and FGF2 (Nikopensius et al., 2010).

5.4 Direct DNA sequencing: fine-tuning association studies signals to elucidate molecular pathogenesis

Many rare functional variants were observed in direct DNA sequencing of IRF6, ARHGAP29, PAX7, MSX1, BMP4, VAX1, FOXE1 and MAFB genes in 184 NSOFCs as well as IRF6 only in 80 MCA cases. These variants included silent or synonymous, missense, protein-truncation or nonsense, splice site and regulatory region variants. Those that were predicted to alter protein structure and mRNA splicing are discussed here. Effect of a variant on protein structure is dependent on the structure, size, charge, hydrophobicity and conservation of both mutant and wildtype amino acid residues.

5.4.1 Deciphering the functional effects of rare pathogenic variants in IRF6 gene

It was observed that aetiologic variants in IRF6 were non-randomly distributed among the 7 coding exons of the gene; moreover, aetiologic IRF6 variants were not detected in all VWS patients. All aetiologic coding and splice site variants (12 out of 13 variants, about 92%), except p.Asn185Thr in exon 6, were distributed between exons 4 and 7 of the gene, confirming earlier observations (Kondo et al., 2002; Ferreira de Lima et al., 2009; Butali et al., 2014a). This observation is relevant for genetic testing in Africa: with limited resources, one can focus on exons 4 and 7 when the need arises to sequence the DNA of a subject. Moreover, potentially aetiologic IRF6 variants were detected in 9 out of the 13 (about 69%) patients with VWS. The lack of aetiologic IRF6 variants in all VWS patients buttresses the point that other upstream or downstream IRF6 target genes or genes in the same regulatory network as IRF6 may also contribute to clefting in VWS and PPS patients, as aetiologic IRF6 variants are observed in only about 70% of VWS patients (Peyrard-Janvid et al., 2014; Fakhouri et al., 2014; WuChou et al., 2013; Rorick et al., 2011; Kwa et al., 2015). It was also observed that 18 out of the 184 patients with NSOFCs had potentially aetiologic IRF6 variants, suggesting aetiologic IRF6 variants occurred in about 9% of our study cohort. These 18 variants included both coding and splice site variants (a total of 12), as well as potentially aetiologic regulatory region (enhancer) variants (a total of 6). Furthermore, no aetiologic IRF6 variants were observed in other MCAs, save VWS patients. Interestingly, some of these MCA patients had phenotypes that have been reported in PPS patients (Ghassibe et al., 2004; Sarode et al., 2011]. These included hexadactyly, syndactyly, undescended testes, hypospadias, nail and toe aplasia, club foot, microphthalmia, etc. This observation suggests other known or unknown syndromes may account for these extra-congenital anomalies, since over 300 syndromes present with cleft and other birth defects (OMIM). The observation of aetiologic variant in overt cleft

patients with variable phenotypes as well as clinically normal parent portrays the variable penetrance of IRF6 variants (Figure 4.9 and 4.10). In a case out of the two cases in which the novel p.Glu69Lys variant was observed, the variant was also observed in clinically unaffected father. Based on molecular evidence, the two patients with p.Glu69Lys variant were actually VWS patients, instead of NSOFC cases. However, during clinical assessment, the classical VWS phenotype that occurs with CL/P, lip pits, were absent. This suggests that though these a priori NSOFC cases are VWS cases, the lip pit phenotype was not penetrant. Variable VWS phenotypes have been observed in VWS families which are usually caused by a single, aetiologic mutation (Gatta et al., 2004; Ghassibe et al., 2004;

Birkeland et al., 2011; de Medeiros et al., 2008). Moreover, sequencing of a priori NSCL/P cases, especially from multiplex families, have shown that VWS can occur as CL/P without lower lip pit or may present with atypical lip pits, such as lip bumps, sinuses, etc. Some of these lip pits are also very shallow and may heal very early in life; clinical assessment will usually miss such lip pit phenotypes (Rutledge et al., 2010;

Desmyter et al., 2010; Malik et al., 2014). Finally, there is the need to do deep (subclinical) phenotyping in families with clefts (Marazita et al., 2009).

Variable expressivity was also observed for some IRF6 aetiologic variants. For example, the protein-truncating variant p.Arg250X was observed in both nonsyndromic and VWS cases: the NSCL case had no lip pits and the variant was not observed in the available maternal sample whereas the VWS cases had lip pits. Even the two VWS cases that had the p.Arg250X mutation had variable phenotypes: a case had complete bilateral CLP, ankyloglossia and bilateral lip pits whereas the other case had right complete

unilateral CL and bilateral lip pits. This variant was not observed in fathers, though there were no samples from mothers to determine whether the variant segregated or was de novo in these syndromic families. The p.Arg250X nonsense variant is a proteintruncation variant and may therefore lead to haploinsufficiency. Though OFCs are quantitative traits with a spectrum of phenotypes, subjects are often classified as "affected" and "unaffected". These classifications are simplistic, as subclinical features, such as orbicularis oris muscle defects, ankyloglossia, atypical lip pits or bumps, alterations in brain structure, facial asymmetry, velopharyngeal insufficiency (hypernasal voice), dental anomalies, congenital absence of uvula, etc., are actually covert or subclinical cleft subphenotypes. IRF6 aetiologic variants have been found in some of these subclinical phenotypes (Weinberg et al., 2009; Neiswanger et al., 2009; Vieira et al., 2008; Nopolous et al., 2007; Dixon et al., 2011; Yeetong et al., 2009).

The variable penetrance and expressivity of IRF6 variants may be due to modifying genetic loci (Kondo et al., 2002). Interestingly, the present study detected some of these long-suspected modifying loci, though none have been found as yet. One of such modifying factors is compound heterozygote: in pedigree GH20135021, the proband inherited two functional variants, p.Glu69Lys from father and p.Asn185Thr from mother. All these variants are novel. The p.Glu69Lys variant alters protein structure (Figure 4.10), whereas p.Asn185Thr, though a missense variant, does not alter protein structure. The p.Asn185Thr variant is predicted to alter an Exonic Splicing Enhancer (ESE) site in the DNA, thereby affecting mRNA splicing (Table 4.11), suggesting this variant is a functional allele. Therefore either of the variant is not sufficient to cause the phenotype meaning both p.Glu69Lys and p.Asn185Thr compound heterozygote are necessary for the severe form of cleft observed in the proband. However, it is also possible the parents of

this case probands could have subclinical phenotypes that could not be clinically ascertained, since even lip pit phenotypes are variable. Moreover, the other case (GH20135024) with p.Glu69Lys variant in IRF6 also had aetiologic p.Pro408Leu variant in PAX7; the proband shared the later variant with clinically unaffected mother. This IRF6-PAX7 compound heterozygote could also explain the presence of the cleft phenotype in the case proband but not the mother. Interactions between IRF6 and other genes have been suggested by a study (Leslie et al., 2013). Therefore, compound heterozygosity and epistasis (gene-gene interactions) may explain why phenotypes of IRF6 aetiologic variants are variable.

Other missense mutations were observed in IRF6: p.Gly65Val, p.Asn88Ser, p.Arg250Gln and p.Lys320Asn (Table 4.11, Figure 4.11). The p.Gly65Val and p.Lys320Asn variants are also novel missense variants that were discovered in VWS patients in the present study. The p.Arg250Gln mutation was observed in a VWS patient, p.Asn88Ser was observed in a NSCLP case whereas the nonsense variant, p.Arg250X, was observed in two VWS cases and one NSCL case; these three variants have been reported elsewhere in VWS patients (Ferreira de Lima et al., 2009). Both clinically unaffected father and a priori NSCLP patient had the p.Asn88Ser variant. The p.Lys320Asn variant was observed in a VWS case with complete bilateral CLP, bilateral lip pits and hypodontia; no parental samples were available. A VWS case with bilateral CLP and lip pits had the p.Gly65Val variant; the variant was not observed in the mother but there was no sample from the father. The p.Arg250Gln mutation was observed in a VWS patient with alveolar cleft only. All these variants cause biochemical alterations to protein folding, with attendant loss of wildtype protein function. These biochemical alterations therefore may account for clefts in cases, since IRF6 is crucial for keratinocyte

proliferation-differentiation switch which is linked to normal epidermal growth (http://genome.ucsc.edu, 22/01/2015).

Splice site and regulatory region variants were also observed in IRF6 (Table 4.11). The splice acceptor site variant, c.175-2A>C, was observed in a priori NSCLP patient with left complete unilateral CLP whereas the splice donor site variant, c.379+1 G>T, was found in a VWS patient with complete bilateral CLP and bilateral lip pits. These two variants are novel; however, some parental samples were not available to determine whether they were segregating or de novo. Other variants that alter mRNA splicing include c.380-116T>A, c.554A>C and c.945G>A. These variants either altered splicing enhancer sites or they activated cryptic intronic donor and acceptor sites. These splicing defects could lead to mRNA with more or less number of nucleotides, thereby shifting the reading frame of the codons. Two other variants, c.1060+26C>T and rs34743335, altered transcriptional enhancer motifs, thereby affecting the binding of POLR2A, the RNA polymerase enzyme that transcribes IRF6 (RegulomeDB). The rs34743335 enhancer element is located about 11 nucleotides upstream of IRF6 (Figure 4.12), a locus with highly conserved nucleotide and H3K27Ac mark, which marks active transcriptional enhancers. Haploreg predicts (Appendix Q) that rs34743335 (A>T) is located in a window that has promoter activity in 20 tissues, enhancer activity in 3 tissues, DNAse activity in 43 tissues and is a binding motif for as many as 18 proteins. Haploreg further asserts that the minor T allele disrupts binding motif for BCL, Egr-1, Ets, Irf, Klf7, Myc, Pou2f2, SP1 and TATA proteins. All these observations buttress the aetiologic nature of T minor allele of rs34743335 and may account for the cleft phenotype in individuals harbouring this variant.

5.4.2 Bioinformatics-predicted functional effects of pathogenic variants on IRF6 protein structure

Project HOPE (http://www.cmbi.ru.nl/hope/method) predicted that the p.Glu69Lys altered IRF6 protein structure and function for a number of reasons. Glutamate is relatively smaller and negatively-charged whereas Lysine is relatively larger and positively-charged. Though the mutated residue does not interact directly with DNA, one of its neighboring residues interacts directly with DNA. This interaction might be affected by the mutation. Moreover, the wildtype residue forms a hydrogen bond with lysine at position 58 of wildtype IRF6 protein. The size difference between wildtype and mutant residue hinders the new residue from making the same hydrogen bond as the original wildtype residue did. The residue is also located on the surface of the protein; mutation of this residue can disturb interactions with other molecules or other parts of the protein. Lastly, the wildtype residue forms a salt bridge with lysine at position 58 of wildtype IRF6 protein. The difference in charge will disturb the ionic interaction made by the original, wildtype residue.

The p.Gly65Val variant also altered IRF6 protein structure and function. According to Project Hope, the mutated residue is not in direct contact with a ligand; however, the mutation could affect the local stability which in turn may affect the ligand-contacts made by one of the neighboring residues. The wildtype residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function. The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for binding. The mutation might disturb the interaction between these two domains and as such affect the function of the protein. The residue is

located on the surface of the protein; mutation of this residue can disturb interactions with other molecules or other parts of the protein. The torsion angles for this wildtype residue are unusual. Only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

For the novel p.Lys320Asn variant, asparagine is relatively smaller and neutral in charge whereas lysine is relatively larger and positively-charged. According to Project Hope, the wildtype residue forms a salt bridge with glutamic acid at position 378. The difference in charge will disturb the ionic interaction made by the original, wildtype residue. The mutated residue is also involved in a multimer contact; the mutation introduces a smaller residue at this position and this new residue might be too small to make multimer contacts. The mutated residue is also located in a domain that is important for binding of other molecules and in contact with residues in a domain that is important for the activity of the protein. The mutation might affect this interaction and thereby disturb signal transfer from binding domain to the activity domain. These biochemical alterations to IRF6 protein structure and function may account for the VWS phenotype in patient.

5.4.3 Pathogenic mutations in other seven genes

Mutations that alter one or more process(es) in the Central Dogma of molecular genetics were also observed in the 7 other genes (Tables 4.12 and 4.13). These included mRNA splicing, missense and frameshift mutations. Donor and acceptor sites variants alter the canonical highly conserved dinucleotides GT and AG, respectively, resulting in loss of sequences that are recognized by the spliceosome (Strachan and Read, 2011). Such

variants produce mRNAs with more than the usual number of nucleotides, leading to frameshift mutations with inadvertent production of non-functional proteins. Such mRNAs may also be degraded largely through the nonsense-mediated decay mechanism and not translated into protein at all, since they are highly unstable. Moreover, mutations that alter splicing enhancer and silencer sites may also affect the amount of functional mRNAs produced. Therefore, splicing variants observed in sequenced genes may also be pathogenic and thus account for the cleft phenotype in affected individuals.

Some variants observed also exhibited variable penetrance and expressivity. For example, in the two cases with CLP that had the p.Ala235Thr mutation in PAX7, a parent of each case also had the heterozygous mutation, though these parents were clinically normal. This particular mutation is predicted to alter the DNA-binding ability of PAX7 transcription factor and may be aetiologic. However, the existence of the mutation in unaffected parents suggests the mutation is not 100% penetrant, presupposing there could modifying genetic loci. However, it is also possible these clinically unaffected parents may have subclinical phenotypes, such as orbicularis oris muscle defects, dental anomalies, alterations in brain structure, VPI, facial asymmetry, etc. (Marazita et al., 2009; Weinberg et al., 2009; Neiswanger et al., 2009; Vieira et al., 2008; Nopolous et al., 2007; Dixon et al., 2011), that could not be ascertained by the present study. Some of the observed variants also exhibited variable expressivity: for example, of the three cases that had p.Ser76Arg mutation in BMP4, a case had left complete CLP, another with right complete CLP and the last case had right complete CL. Variations in laterality and severity as far as the phenotypes emanating from this variant is concerned suggest the mutation has variable expressivity and genetic modifying loci may account for this (Kondo et al., 2002).

Two or more genes may be mutated in some OFC patients and this may account for the variable penetrance and expressivity of some of these mutations. For example, in pedigree GH20134932, the case proband has two mutated genes: p.Pro190Arg in FOXE1 and c.1284+4A>G (splice site variant) in ARHGAP29. The ARHGAP29 mutation, but not the FOXE1, was also observed in the mother but no paternal samples were available. Thus, the combined effects of these mutated genes could explain why the case has CLP while the mother is clinically unaffected. Also, in pedigree GH20130876, a proband with CLP had two mutated genes: p.Pro190Arg in FOXE1 and p.Lys426Ilefs*Ter6 in ARHGAP29. None of these two variants were observed in the available maternal sample though no paternal sample was available. Finally, in pedigree GH20134962, a proband with CL had p.Pro190Arg in FOXE1 and p.Arg250X in IRF6. Here, the FOXE1 mutation, but not the IRF6 mutation, was observed in the available maternal sample. These observations presuppose that multiple developmental or homeobox genes may get mutated during embryogenesis, resulting in OFCs.

Biochemical alterations in mutant protein structures may be directly responsible for the pathogenic effects of missense mutations. Project Hope analyses for various missense variants suggested that changes in amino acid size, charge and hydrophobicity, as far as mutant and wildtype residues are concerned, may be responsible for the cleft phenotypes in cases. The conservation and location of each wildtype residue are also crucial for protein function. Therefore, mutations that alter biochemical properties (and phylogenetically conserved DNA sequences) may produce dysfunctional proteins, inadvertently leading to diseased phenotypes.

5.4.4 Deciphering the effects of mutations on protein structure and function

HOPE (http://www.cmbi.ru.nl/hope/method) predicts that the p.Asp988Tyr variant alters ARHGAP29 protein structure and function (Figure 4.13) for a number of biochemical reasons. The mutant residue, tyrosine, is bigger, neutral and more hydrophobic whereas the wildtype residue, aspartic acid, is smaller, negatively charged and less hydrophobic. Therefore, the negative charge of the wildtype residue will be lost and this can cause loss of interactions with other molecules or residues. The bigger mutant residue might also lead to bumps. Also, the higher hydrophobicity of tyrosine can result in loss of hydrogen bonds and/or disturb correct folding. Finally, the wildtype residue is predicted to be situated in its preferred secondary structure, a turn, while the mutant residue prefers to be in another secondary structure; therefore the local conformation will be slightly destabilized.

For the p.Gly252Cys mutation in VAX1, HOPE predicts (Figure 4.14) that the mutant residue (cysteine) is bigger and more hydrophobic than the wildtype residue (glycine). The bigger mutant residue might lead to bumps. Also, the wildtype residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function. Finally, the torsion angles for the residue at position 252 of VAX1 protein are unusual. Only glycine is flexible enough to make these torsion angles and mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

According to HOPE, the mutant residue (leucine) of p.Pro408Leu observed in PAX7 is bigger than the wildtype residue (proline) and this may leads to bumps (Figure 4.15). Moreover, proline is known to have a very rigid structure, sometimes forcing the backbone in a specific conformation. Possibly, the p.Pro408Leu mutation changes a proline with such a function into another residue, thereby disturbing the local structure.

For the p.Ala235Thr mutation in PAX7, simulation by HOPE (Figure 4.16) suggests the mutant residue (threonine) is bigger and less hydrophobic than the wildtype residue (alanine). The wildtype residue is buried in the core of the protein, presupposing that the bigger mutant residue will probably not fit at this position. The less hydrophobic nature of the mutant residue might also cause loss of hydrophobic interactions in the core of the protein. Also, the 3D-structure of PAX7 protein shows that the wildtype residue is located in an α-helix. However, p.Ala235Thr mutation converts the wildtype residue into another residue that does not prefer α-helices as secondary structure. The wildtype residue is also highly conserved and is part of a domain named homeobox domain or DNA-binding domain. Thus, the mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for DNA binding. The mutation might therefore disturb the interaction between these two domains and as such affect the function of the protein.

Computational simulation by HOPE demonstrates that the p.Ser76Arg mutation in BMP4 is pathogenic for a number of biochemical reasons (Figure 4.17). The mutant residue (arginine) is bigger than the wildtype residue (serine) and this may lead to bumps. The wildtype amino acid is neutral whereas the mutant residue is positively charged; the mutation therefore introduces a charge, which can cause repulsion of ligands or other residues with the same charge. The wildtype residue is also more hydrophobic than the mutant residue; hydrophobic interactions, either in the core of the protein or on the surface, will be lost. Moreover, the mutation is predicted to be located within the signal peptide. This sequence of this peptide is crucial because it is recognized by other proteins and often cleaved off to generate the mature protein. The new residue that is introduced in the signal peptide differs in its properties from the original one. It is possible that this mutation

disturbs recognition of the signal peptide. Finally, the mutant residue is located in the transforming growth factor-beta (TGFβ) N-terminal domain which is critical for BMP4 protein's growth factor and protein-binding activities. Thus, the mutated residue is located in a domain that is important for binding of other molecules and such mutational event might disturb this function.

For p.Pro190Arg mutation in FOXE1, the mutant residue is bigger than the wildtype residue and this might lead to bumps (Figure 4.18c). Also, the wildtype residue was neutral whereas the mutant residue is positively charged. The introduction of a positive charge can cause repulsion of ligands or other residues with the same charge. Moreover, the wildtype residue is more hydrophobic than the mutant residue, presupposing hydrophobic interactions, either in the core of the protein or on the surface, will be lost. Finally, proline is known to have a very rigid structure, sometimes forcing the backbone in a specific conformation. It is possible the substitution of a proline with such a function for another residue, might disturb the local structure of the protein.

5.5 The corroborative evidence from zebrafish embryology

Two aetiologic variants from IRF6, p.Glu69Lys and p.Gly65Val, were selected for functional or in vivo validation in zebrafish. These were hypothesized to exert their aetiologic effect through dominant-negative mechanism: mutant protein from mutant allele prevents wildtype protein from wildtype allele from exerting its functional effects (Peyrard-Janvid et al., 2014; Hor et al., 2015; Wu et al., 2015). Mutant mRNAs were injected into wildtype zebrafish embryos. Wildtype IRF6 protein was produced by wildtype zebrafish IRF6 gene whereas mutant protein was also expressed in the same embryos by mutant exogenous mRNA. The mutant protein, though dysfunctional, gains

competitive advantage and selectively binds to IRF6 downstream targets, preventing wildtype proteins from functioning, thereby causing a dominant-negative effect. This phenomenon creates an embryo with nonfunctional IRF6 gene. Since IRF6 is crucial for keratinocyte proliferation-differentiation switch which is linked to normal epidermal growth (http://genome.ucsc.edu, 22/01/2015), this dominant-negative effect disrupt normal embryo development, leading to ruptured, malformed or dead embryos.

Embryos injected with mutant p.Glu69Lys mRNA showed an array of phenotypes (Plate 4.1, bottom of Panel C and Plate 4.2A) that exactly recapitulated the phenotypes observed in human subjects. At 60 ng/uL and 100 ng/uL, no major abnormalities were observed, probably because not enough mutant proteins were produced to exert the dominant-negative effect. However, at 200 ng/uL a plethora of phenotypes resulted, such as yolk bolus and heart oedema. Interestingly, the craniofacial anomalies observed in zebrafish were similar to CL/P. The craniofacial arches (Meckel's cartilage, palatoquadrate, hyosymplectic, ceratohyal, basihyal and ceratobranchial) in zebrafish, that are analogous to human alveolar, maxillary and mandibular bones which emanate from first pharyngeal arch of developing human foetus, were either severely malformed or completely absent. This recapitulates the effect of this variant in humans: it was found in two NSOFC patients, one with incomplete NSCP (soft palate cleft) and the other with Complete NSCLP with bifid uvula. These observations buttress the in silico predictions that this variant is aetiologic, as well as its variable expressivity observed in humans. Ironically, about two embryos treated with the same mRNA were still normal at the time of knock-down, presupposing the variant is not completely penetrant and may account for why this variant is also found in a clinically unaffected father.

The p.Gly65Val mutant gave the same phenotype at both 100 ng/uL and 200 ng/uL (Plate 4.1, middle of both Panels B and C). The ventral components of mandibular and hyoid arches (meckel's cartilage and ceratohyal cartilages) as well as dorsal structures (palatoquadrate and hyosymplectic cartilages) were malformed whereas some of the supporting pharyngeal arches (basihyal and ceratobranchial) were completely absent. This variant was observed in a VWS patient with complete bilateral CLP and bilateral lower lip pits, hence the similar phenotype observed in treated embryos. Similar malformations of pharyngeal arches observed for both p.Glu69Lys- and p.Gly65Val-treated embryos have been made elsewhere in zebrafish which were mutants for various craniofacial developmental genes (Schilling et al., 1996; Piotrowski et al., 1996). Moreover, the occurrence of malformed heads, eyes and tails, as well as compact bodies, heart oedema and yolk bolus in embryos treated with mutant mRNAs suggest periderm development had been disrupted (Peyrard-Janvid et al., 2014). Such dominant-negative effects that disrupt specific zebrafish morphological structures and function, but not necessarily leading to embryo rupture or death, have been reported in recent studies (Hor et al., 2015; Wu et al., 2015).

The differences in proportion of malformed embryos for each mRNA treatment and dosage were statistically significant. The level of significance was higher at 100 ng/uL ($p<0.0001$) than at 200 ng/uL ($p=0.0011$). This observation stems from the fact that a larger proportion of p.Gly65Val-treated embryos were malformed at 100 ng/uL whereas increasing dosage to 200 ng/uL did not significantly alter the percentage of abnormal embryos for this variant. This presupposes that the variant exerts its effect even at lower dosage (Table 4.14; Plate 4.1, middle of both Panels B and C). The p.Glu69Lys variant was more deleterious at 200 ng/uL, presupposing the mutant allele may be up-regulated

than the wildtype allele in the patient. This also suggests the effect of p.Glu69Lys mutant allele may be modified by p.Asn185Thr in IRF6 and p.Pro408Leu variant in PAX7 in these compound heterozygote patients. These statistical significances, coupled with the observed phenotypes, buttress the fact that the two IRF6 variants tested in zebrafish embryos are responsible for the CL/P phenotypes in humans.

5.6 The gist of the present study

The present study has shown evidence of association of certain loci with NSOFCs among Africans. Candidate genes that have never reached genome-wide significance in various GWASs published on NSOFCs to date may play significant roles in the aetiology of NSOFCs among Africans. Subphenotype as well as sub-population analyses and genotyping of other SNPs, other than those already reported for some loci, may be crucial in identifying NSOFC loci in various ethnicities and populations. This study further suggests that earlier studies (Weatherley-White et al., 2011; Figueiredo et al., 2014; Butali et al., 2011) that could not replicate various association signals among Africans could have resulted from disregard for subphenotype analyses, small sample size and/or genetic heterogeneity among populations. The present study is the first to demonstrate any of these associations between these loci and NSOFC among Africans.

These observations are crucial for understanding the genetic architecture of NSOFCs in Africans and further suggest the need to carry out GWAS for every ethnicity as far as complex traits are concerned.

The present study could not detect formal association between some GWAS hit loci as well as candidate gene loci and NSOFCs, presupposing either these loci may not

play a role in the aetiology of NSOFCs in Africans or the genotyped SNPs may not be the tag SNPs for Africans. Lack of statistical power due to small sample size and low MAF of the genotyped SNPs in Africans could also be possible reasons. For example, a SNP, rs2235371 of IRF6 which is in high LD and same locus as rs642961, that has been associated with NSCL/P mostly among Asians (Sun et al., 2015) and in some Europeans (Zucchero et al., 2004), does not exist in the African population (http://browser.1000genomes.org/index.html). It is also possible that even when no associations are detected between reported loci and NSOFCs, potentially pathogenic variants may be observed in NSOFC cases through direct DNA sequencing of such loci. Therefore, GWAS and whole genome sequencing (WGS) of NSOFC cases from Africa is required to detect more risk loci.

Direct DNA sequencing and SNP association studies are complementary techniques in the search for aetiologic loci for OFCs. Analyses of both association studies and direct DNA sequence results suggest that different SNPs or markers may be responsible for the association between certain loci and NSOFCs in Ghana or Africa. Though none of the reported SNPS that were genotyped for PAX7, FOXE1 and MSX1 exhibited evidence of association with NSOFCs in Ghana, direct DNA sequence analysis showed that other SNPs may be responsible for the association between these genes and NSOFCs in Ghana or Africa. These include p.Ala172Ala (rs775448711) in FOXE1, rs4920523 in PAX7 and rs115200552 in MSX1.

Observations made by the present study recapitulate out of Africa (OOA) human migration history: all humans have a common African origin and migrated out of East Africa less than 200,000 years ago to their present locations. OAA has shown that humans migrated from Africa and settled first in Europe (forming the Caucasian population) before

some migrated further to Asia (forming the Asian population). The Native Americans or Amerindians migrated further from Asia to the Americas (McEvoy et al., 2011; Mathias et al., 2016; Henn et al., 2016). The incidence of cleft also increases with this pattern of migration and distance from the African population:

Africans have the lowest incidence of OFC, followed by Caucasians whereas Asians and Native Americans (Amerindians) have the highest incidence of OFC (Mossey and Modell, 2012). The ABCA4/ARHGAP29, MAFB and IRF6 signals have strongly been associated with NSCL/P in populations of Asian ancestry whereas the 8q24 locus is much stronger in populations of European ancestry (Beaty et al., 2010; Sun et al., 2015).

For cohorts from Africa to show that all these loci contribute to the aetiology of NSOFCs, buttresses the common origin of humans from Africa. This notwithstanding, the African genome may have developed microevolutionary mechanisms by maintaining the highest number of heterozygosity for rare deleterious genetic variants as well as lower number of deleterious common variants (MAF>10%) of large effect size whereas European and Asian populations have the highest number of homozygosity for the same rare deleterious genetic variants as well as higher number of deleterious common variants of large effect size (Henn et al., 2016). These phenomena may have combatted the effect of some of these aetiologic loci, culminating in reduced clefting in Africans (Modell and Mossey, 2012; Butali et al., 2014c). So the older a population, the more time that population has to diversify genetically and microevolutionally (Mathias et al., 2016), phenomena which subsequently act through purifying selection to combat the effects of aetiologic variants (Henn et al., 2016). The global distribution and genetic causes of clefts thus recapitulate the OOA model of human migration history.

The incidence of OFC in Africans is much lower than in Europeans, Asians and Native Americans or Amerindians (Mossey and Modell, 2012; Butali et al., 2014c), even though these populations may share the same or similar genetic susceptibility loci for OFCs, as observed in the present study. Though there is no national prevalence data for Ghana, a prevalence estimate of 0.5 per 1000 has been observed for Nigeria (Butali et al., 2014c). Among Ethiopians, estimated incidence of 0.44 per 1000 live births and a prevalence of 0.20 per 1000 individuals were observed in a recent study (Eshete et al., 2016). Inasmuch as under-ascertainment due to lack of birth defect registries in most African countries could be a contributing factor (Butali et al., 2014c), the low incidence of OFCs among Africans may be real, as African-derived populations in the Caribbean as well as African Americans in the United States of America have lower OFC incidence that is similar to their ancestral African population (Mossey and Modell, 2012). Therefore, to account for the disparities in the incidence or prevalence of OFCs among Caucasians, Asians and Africans, though these racial populations share similar genetic susceptibility loci, I hypothesize the possible existence of genetic protective variants in the African genome, whose "rescue mission" reduces clefting. These protect genetic variants may be copy number variations (CNVs) of genes at OFC risk loci, the ability of other genes to perform similar functions as their orthologous genes as well as mutations that can activate both nonprocessed and processed (also called retrotransposed) pseudogenes. A pseudogene is a defective copy of a functional gene to which there is significant sequence homology. A nonprocessed pseudogene usually encompasses the 5'UTR, promoter, introns, exons and 3'UTR of a whole gene whereas a retrotransposed pseudogene is made from matured mRNA and is usally a complementary DNA (cDNA) sequence from the exons of a gene and therefore lacks regulatory regions like 5'UTR,

promoter and 3'UTR as well as introns. Insertion of retrotransposed pseudogene near a functional promoter can lead to its expression (Strachan and Read, 2011). Thus, the identification and elucidation of such protective variants in the African genome can be translated to European and Asian populations to bring about reduced OFC incidence, and eventually prevention.

5.7 Significance of the present study to public health

The present study is relevant for genetic counseling. Couples that give birth to children with cleft usually inquire about the possibility of giving birth to another child with cleft. It was observed by this study that only 5% of affected families were multiplex, i.e. had more than one individual affected with cleft. This presupposes that the recurrence risk is generally very low, though individuals with syndromic clefts may have as high as 50% recurrence risk (http://www.cleftline.org/docs/Booklets/GEN-01.pdf, 12/01/2015). Through direct DNA sequencing, this study also detected high risk families in that the affected individuals harboured aetiologic mutations in some genes that may be transferred to subsequent generations. Individuals with IRF6 mutations, for example, have three-fold increase in OFC recurrence risk (Ghassibe et al., 2004). OFCs were also more prevalent (over 90% of cases) in families with low socio-economic status.

Measures geared toward OFC prevention may also benefit from the present study. It was observed that affected families had poor diet, largely did not consume folate-fortified foods and majority of mothers (about 86%) either had antenatal care after 3 months of gestation or did not access antenatal care at all. Folic acid intake is crucial during periconception (three months before and after fertilization), since the human face forms by 10 weeks of gestation. Observations made by the present study are in harmony with

those reported elsewhere (Darmstadt et al., 2016) that have shown that folate fortification and periconception health care are crucial for prevention of OFCs, and birth defects in general. This strategy has led to reduced incidence of OFCs in the United States of America (Mossey and Modell, 2012).

Identification and elucidation of pathogenic genetic mechanisms of OFC may also be relevant for prenatal foetal screening as well as embryo screening and selection. Foetus may be screened for mutations in any of the OFC susceptibility genes and this may predict the chance of having a child with OFCs. Parents may be be given the option to either keep the pregnancy or abort it. In vitro fertilization may also be carried out for high risk couples. The fertilized embryos may then be screened for mutations in OFC susceptibility genes. Embryos without any OFC susceptibility mutations may subsequently be implanted into the uterus of the mother.

The ultimate goal of modern genetics and genomics studies is personalized medicine, i.e. giving specific treatment to patients based on their genetic profile. Individuals could be screened for their OFC carrier status based on the occurrence of certain genetic variants in their genomes and this could be used to predict the probability of giving birth to children with OFCs. Patients with inborn errors of metabolism may benefit immensely from molecular protein therapeutics that may ameliorate the debilitating effects of mutaed genes. However, individuals with structural birth defects, such as OFC, may not benefit immensely from protein therapeutics because re-initiation of morphological growth to close the cleft may be difficult. This notwithstanding, modern biotechnology may grow histologically compatible and "personalized" tissues that may aid OFC surgical repair. For structural birth defects, surgical interventions may always be necessary irrespective of the advances in genetics and genomics.

Individuals with IRF6 pathogenic mutations (VWS patients) have poor wound healing capability and there is high possibility of the occurrence of scar, fistula and dehiscence after surgery (Jones et al., 2010). The observation of pathogenic IRF6 mutation in some patients by the present study may give a hint to surgeons about the possible occurrence of these observations.

CHAPTER SIX

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Corollary to the observations made by the present study, the following conclusions were made:

- Orofacial clefts showed multifactorial pattern of inheritance in the Ghanaian population, because both genes and environmental factors interacted in cleft pathogenesis.

- Families with cleft were mostly characterized by low socio-economic status. Significant percentage (96%) of case mothers had poor dietary folate deficiency, coupled with no or delayed antenatal care and folate supplementation.
- Association studies that employed common variants showed that many loci (e.g. IRF6, ARHGAP29, VAX1, PAX7, MAFB, etc.) were involved in cleft aetiology among Ghanaians and Africans, with striking racial differences.
- Many novel or known rare functional variants were observed through direct DNA sequencing and various in silico predictions and simulations showed that these variants were probably pathogenic.
- In vivo functional assays in zebrafish showed that the selected variants disrupted craniofacial structures through dominant negative effect.

6.2 Limitations

- Lack of relevant expertise in clinical genetics and dysmorphology might have affected the classification of clefts into syndromic and nonsyndromic forms.
- Small number of case-parents trios, multiplex families and control probands might have affected the power to detect certain associations.
- Sub-clinical phenotypes were not effectively ascertained due to lack of equipment.
- Lack of Birth Defects Registry (and effective birth surveillance system in general) might have affected sample size calculation and other observations.

6.3 Recommendations

Based on the outcomes of the present study, the following are recommended. ✓ Larger sample size should be used in future studies and this can be achieved by continuous sample collection by a Principal Investigator.

- ✓ Future studies must employ equipment that can detect sub-clinical cleft phenotypes, such as orbicularis muscle defects.
- ✓ Lip tissue samples could also be collected from patients for in situ expression studies of some of the genes implicated by this study.
- ✓ Birth defect registry must be set-up at various health centres. This will help health practitioners to track patients with congenital deformities and will also enable large family-based studies which are essential for "teasing out" associated genetic factors.
- ✓ Newborn Screening and thorough examination of babies at birth for birth defects will enhance detection of clefts at birth.
- ✓ A nutritionist may design a data collection instrument that may further probe into the folate deficiency observation made by this study.
- ✓ Peri-conception healthcare: maintaining a balanced diet, especially folate-rich foods, must be encouraged among mothers, particularly during three months prior to conception and first trimester of conception.

6.4 Publications

- Gowans LJJ, Adeyemo WL, Eshete M, Mosey PA, Busch T, Aregbesola B, Donkor P, Arthur FKN, Bello SA, Martinez A, Li M, Augustine-Akpan E-A, Deressa W, Twumasi P, Olutayo J, Deribew M, Agbenorku P, Oti AA, Braimah R, Plange-Rhule G, Gesses M, Obiri-Yeboah S, Oseni GO, Olaitan PB, Abdur-Rahman L, Abate F, Hailu T, Gravem P, Ogunlewe MO, Buxó CJ, Marazita ML, Adeyemo AA, Murray JC, Butali A (2016). Association Studies and Direct DNA Sequencing Implicate Genetic Susceptibility Loci in the Etiology of Nonsyndromic Orofacial Clefts in Sub-Saharan African Populations. J Dent Res.; 95(11):1245-1256. DOI: 10.1177/0022034516657003.
- Gowans LJJ, Busch TD, Mossey PA, Eshete MA, Adeyemo WL, Aregbesola B, Donkor P, Arthur FKN, Agbenorku P, Olutayo J, Twumasi P, Braimah R, Oti AA,

Plange-Rhule G, Obiri-Yeboah S, Abate F, Hoyte-Williams PE, Hailu T, Murray JC, Butali JC (2017). The Prevalence, Penetrance and Expressivity of Aetiologic IRF6 variants in Orofacial Clefts Patients from sub-Saharan Africa. Mol Genet Genomic Med.; 5(2):164-171. DOI:10.1002/mgg3.273.

REFERENCES

Agbenorku, P., Agbenorku, M., Iddi, A., Abude., F., Sefenu, R., Matondo, P. and Schneider, W. (2011a). A study of cleft lip/palate in a community in the South East of Ghana. Eur J Plast Surg.; 34(4):267-272.

Agbenorku, P., Ansah, S., Acheampong, A., Sabbah, D., Bankas, D., Adu, E., Plange-Rhule, G., Adiyiah, Y. and Donkor, P., (2011b). Komfo Anokye Teaching Hospital Multidisciplinary Cleft Clinic. J Craniofac Surg.; 22(2):532-536.

Agbenorku, P., Yorke, M., Danso, K. A. and Turpin, C., (2013). Incidence of Orofacial Clefts in Kumasi, Ghana. ISRN Plast Surg.; 2013:28903.

Antunes, L. S., Kuchler, E. C., Tannure, P. N., Costa, M. C., Gouvea, C. V., Olej, B. and Granjeiro, J. M., (2013). BMP4 polymorphism is Associated With Nonsyndromic Oral Cleft in a Brazillian Population. Cleft Palate Craniofac J.; 50(6):633-638.

Araujo, T. K., Simioni, M., Felix, T. M., de Souza, L. T., Fontes, M. I., Monlleo, I. L., Souza, J., Fett-Conte, A. C., Secolin, R., Lopes-Cendes, I., Maurer-Morelli, C. V. and Gil-daSilva-Lopes, V. L., (2012). Preliminary Analysis of the Nonsynonymous Polymorphism rs17563 in BMP4 Gene in a Brazillian Population Suggests Protection for Nonsyndromic Cleft Lip and Palate. Plast Surg Int.; 2012:247104.

Araujo, T. K., Secolin, R., Felix, T. M., Souza, L. T., Fontes, M. I., Monlleo, I. L., Souza, J., Fett-Conte, A. C., Ribeiro, E. M., Xavier, A. C., Rezende, A. A., Simioni, M., RibeiroDos-Santos, A. K., Santos, S. E. and Gil-da-Silva-Lopes V. L., (2015). A multicentric association study between 39 genes and nonsyndromic cleft lip and palate in a Brazilian population. J Craniofac Surg.; S1010-5182(15)00252-00258.

Auerbach, B. D., Osterweil, E. K. and Bear, M. F., (2011). Mutations causing syndromic autism define an axis of synaptic pathophysiology. Nature; 480(7375):63–68.

Baghestani, S., Sadeghi, N., Yavarian, M. and Alghasi, H., (2010). Lower Lip Pits in a Patient With van der Woude Syndrome. J Craniofac Surg; 21(5):1380-1381.

Basch, M. L., Bronner-Fraser, M. and Garcia-Castro, M. I., (2006). Specification of the neural crest occurs during gastrulation and requires Pax7. Nature; 441(7090):218–222.

Beaty, T. H., Murray, J. C., Marazita, M.L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., Liang, K. Y., Wu, T., Murray, T., Fallin, M. D., Redett, R. A., Raymond, G., Schwender, H., Jin, S.C., Cooper, M. E., Dunnwald, M., Mansilla, M. A., Leslie, E., Bullard, S., Lidral, A. C., Moreno, L. M., Menezes, R., Vieira, A. R., Petrin, A., Wilcox, A. J., Lie, R. T., Jabs, E. W., Wu-Chou, Y. H., Chen, P. K., Wang, H., Ye, X., Huang, S., Yeow, V., Chong, S. S., Jee, S. H., Shi, B., Christensen, K,, Melbye, M., Doheny, K. F., Pugh, E. W., Ling, H., Castilla, E. E., Czeizel, A. E., Ma, L., Field, L. L., Brody, L., Pangilinan, F., Mills, J. L., Molloy, A. M., Kirke, P. N., Scott, J. M., Arcos-Burgos, M. and Scott, A. F., (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet.; 42(6):525–529.

Beaty, T. H., Taub, M. A., Scott, A. F., Murray, J. C., Marazita, M. L., Schwender, H., Parker, M. M., Hetmanski, J. B., Balakrishnan, P., Mansilla, M. A., Mangold, E., Ludwig, K. U., Noethen, M. M., Rubini, M., Elcioglu, N. and Ruczinski, I., (2013). Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. Hum Genet.; 132(7):771–781.

Berk, N. W. and Marazita, M. L., (2002). The Costs of Cleft Lip and Palate: Personal and Societal Implications. In: Wyszynski DF, editor. Cleft Lip and Palate: From Origin to Treatment. Oxford: Oxford University Press.

Birkeland, A. C., Larrabee, Y., Kent, D. T., Flores, C., Su, G. H., Lee, J. H. and Haddad Jr, J., (2011). Novel IRF6 mutations in Honduran Van Der Woude syndrome patients. Mol Med Rep.; 4(2):237-241.

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeidade Assis, N., Alblas, M. A., Barth, S., Freudenberg, J., Lauster, C., Schmidt, G., Scheer, M., Braumann, B., Berge, S. J., Reich, R. H., Schiefke, F., Hemprich, A., Potzsch, S., Steegers-Theunissen, R. P., Potzsch, B., Moebus, S.,

Horsthemke, B., Kramer, F. J., Wienker, T. F., Mossey, P. A., Propping, P., Cichon, S., Hoffmann, P., Knapp, M., Nothen, M. M. and Mangold. E., (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet.; 41(4):473–477.

Blanton, S. H., Cortez, A., Stal, S., Mulliken, J. B., Finnell, R. H. and Hecht, J.T., (2005). Variation in IRF6 contributes to nonsyndromic cleft lip and palate. Am J Med Genet A.; 137A(2):259–262.

Blanton, S. H., Burt, A., Garcia, E., Mulliken, J. B., Stal, S. and Hecht, J. T., (2010). Ethnic heterogeneity of IRF6 AP-2α binding site promoter SNP association with nonsyndromic cleft lip and palate. Cleft Palate Craniofac J.; 47(6):574–577.

Boyles, A. L., Wilcox, A. J., Taylor, J. A., Meyer, K., Fredriksen, A., Ueland, P. M., Drevon, C. A., Vollset, S. E. and Lie, R. T., (2008). Folate and one-carbon metabolism gene polymorphisms and their associations with oral facial clefts. Am J Med Genet. A.; 146A(4): 440-449.

Brito, L. A., Bassi, C. F. S., Masotti, C., Malcher, C., Rocha, K. M., Schlesinger, D., Bueno, D. F., Cruz, L. A., Barbara, L. K., Bertola, D. R., Meyer, D., Franco, D., Alonso, N. and Passos-Bueno, M. R., (2012). IRF6 is a risk factor for nonsyndromic cleft lip in the Brazilian population. Am J Med Genet A.; 158A(9):2170–2175.

Bueno, D. F., Sunaga, D. Y., Kobayashi, G. S., Aguena, M., Raposo-Amaral, C. E., Masotti, C., Cruz, L. A., Pearson, P. L., Passos-Bueno, M. R. (2011). Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. Stem Cell Rev.; 7(2):446–457.

Butali, A., Mossey, P. A., Adeyemo, W. L., Jezewski, P. A., Onwuamah, C. K., Ogunlewe, M. O., Ugboko, V. I., Adejuyigbe, O., Adigun, A. I., Abdur-Rahman, L. O., Onah, Audu, R. A., Idigbe, E. O., Mansilla, M. A., Dragan, E. A., Petrin, A. L., Bullard, S. A., Uduezue, A. O., Akpata, O., Osaguona, A. O., Olasoji, H., Ligali, T. O., Kejeh, B. M., Iseh, K. R., Olaitan, P. B., Adebola, A. R., Efunkoya, E., Adesina, O. A., Oluwatosin, O. M., Murray, J. C. and NigeriaCRAN Collaboration (2011). Genetic studies in the Nigerian population implicate an MSX1 mutation in complex oral facial clefting disorders. Cleft Palate Craniofac J.; 48(6):646–653.

Butali, A., Suzuki, S., Cooper, M. E., Mansilla, A. M., Cuenco, K., Leslie, E. J., Suzuki, Y., Niimi, T., Yamamoto, M., Ayanga, G., Erkhembaatar, T., Furukawa, H., Fujiwawa, K., Imura, H., Petrin, A. L., Natsume, N., Beaty, T. H., Marazita, M. L. and Murray, J. C., (2013). Replication of genome wide association identified candidate genes confirm the role of common and rare variants in PAX7 and VAX1 in the etiology of nonsyndromic CL[P]. Am J Med Genet A.; 161A(5):965–972.

Butali, A., Mossey, P., Adeyemo, W. L., Eshete, M., Gaines, L. A. L., Braimah, R. O., Aregbesola, B. S., Rigdon, J., Emeka, C., Olutayo, J., Ogunlewe, O., Ladeinde, A., Abate, F., Hailu, T., Mohammed, I., Gravem, P., Deribew, M., Gesses, M., Adeyemo, A., Marazita, M. and Murray, J., (2014a). Rare Functional Variants in Genome–wide Association Identified Candidate Genes for Non-syndromic Clefts in the African Population. Am J Med Genet A.; 164A(10):2567–2571.

Butali, A., Mossey, P. A., Adeyemo, W. L., Eshete, M. A., Gaines, L. A., Even, D., Braimah, R. O., Aregbesola, B. S., Rigdon, J. V., Emeka, C. I., Olutayo, J., Ogunlewe, M. O., Ladeinde, A. L., Abate, Fikre, T., Hailu, Mohammed, I., Gravem, P. E., Deribew, M.,

Gesses, M., Adeyemo, A. A. and Murray J. C., (2014b). Novel IRF6 mutations in families with Van Der Woude syndrome and popliteal pterygium syndrome from sub-Saharan Africa. Mol Genet Genomic Med.; 2(3):254–260.

Butali, A., Adeyemo, W. L., Mossey,P. A., Olasoji, H. O., Onah, I. I., Adebola, A., Efunkoya, Akintububo, A., James, O., Adeosun, O. O., Ogunlewe, M. O., Ladeinde, A. L., Mofikoya, B. O., Adeyemi, M. O., Ekhaguere, O. A., Emeka, C., Awoyale, T. A., and The NigeriaCRAN collaboration, (2014c). Prevalence of orofacial clefts in Nigeria. Cleft Palate Craniofac J.; 51(3):320-325.

Chandrasekharan, D. and Ramanathan, A., (2014). Identification of a novel heterozygous truncation mutation in exon 1 of ARHGAP29 in an Indian subject with nonsyndromic cleft lip with cleft palate. Eur J Dent.; 8(4):528-532.

Chawa, T. A., Ludwig, K. U., Fier, H., Potzsch, B., Reich, R. H., Schmidt, G., Braumann, B., Daratsianos, N., Bohmer, A. C., Schuencke, H., Alblas, M., Fricker, N., Hoffmann, P., Knapp, M., Lange, C., Nothen, M. M. and Mangold, E., (2014). Nonsyndromic Cleft Lip with or without Cleft Palate: Increased Burden of Rare Variants within Gremlin-1, a Component of the Bone Morphogenetic Protein 4 Pathway. Birth Defects Res A Clin Mol Teratol.; 100(6):493–498.

Cheng, H., Huang, E., Tan, S., Xu, M. and Shu, S., (2012). Association between single nucleotide polymorphisms of v-maf musculoaponeurotic fibrosarcoma oncogene homologue B gene and non-syndromic cleft lip with or without cleft palate. Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi; 26(6):845-848.

Chiquet, B. T., Lidral, A. C., Stal, S., Mulliken, J. B., Moreno, L. M., Arcos-Burgos, M., Valencia-Ramirez, C., Blanton, S. H. and Hetch J. T., (2007). CRISPLD2: a novel NSCLP candidate gene. Hum Mol Genet.; 16(18):2241-2248.

Chiquet, B. T., Henry, R., Burt, A., Mulliken, J. B., Stal, S., Blanton, S. H. and Hetch, J. T., (2011). Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection. Birth Defects Res A Clin Mol Teratol.; 91(1):44-49.

Conte, F., Oti, M., Dixon, J., Carels, C. E., Rubini, M. and Zhou, H., (2016). Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. Hum Genet.; 135(1):41-59.

Crider, K. S., Yang, T. P., Berry, R. J. and Bailey, L. B., (2012). Folate and DNA Methylation: A Review of Molecular Mechanisms and the Evidence for Folate's Role. Adv Nutr.; 3(1):21–38.

Darmstadt, G. L., Howson, C. P., Walraven, G., Armstrong, R. W., Blencowe, H. K., Christianson, A L., Kent, A., Malherbe, H., Murray, J. C., Padilla, C. D., Walani, S. R.; Participant Working Group of the Dar es Salaam Seventh International Conference on Birth Defects and Disabilities in the Developing World, (2016). Prevention of Congenital Disorders and Care of Affected Children: A Consensus Statement. JAMA Pediatr.; 170(8):790-793.

de Aguiar, P. K., Coletta, R. D., de Oliveira, A. M., Machado, R. A., Furtado, P. G., de Oliveira, L. A., de Aquino, S. N., Martelli-Junior, H., de Almeida Reis, S. R., Moreira, H. S. and Persuhn, D. C., (2015). Rs1801133C>T polymorphism is a risk factor for nonsyndromic cleft lip with or without cleft palate in the Brazilian population. Birth Defects Res A Clin Mol Teratol.; 103(4):292-298.

de Medeiros, F., Hansen, L., Mawlad, E., Eiberg, H., Asklund, C., Tommerup, N. and Jakobsen, L. P., (2008). A Novel Mutation in IRF6 Resulting in VWS–PPS Spectrum Disorder With Renal Aplasia. Am J Med Genet A.; 146A(12):1605–1608.

De Muynck, S., Schollen, E., Matthijs, G., Verdonck, A., Devriendt, K. and Carels, C., (2004). A Novel MSX1 Mutation in Hypodontia. Am J Med Genet A.; 128A(4):401–403.

Desmyter, L., Ghassibe, M., Revencu, N., Boute, O., Lees, M., François, G., VerellenDumoulin, C., Sznajer Y., Moncla, A., Benateau, H., Claes, K., Devriendt, K., Mathieu, M., Van Maldergem, L., Addor, M. C., Drouin-Garraud, V., Mortier, G., Bouma, M., Dieux-Coeslier, A., Genevieve, D., Goldenberg, A., Gozu, A., Makrythanasis, P., McEntagart, M., Sanchez, A., Vilain, C., Vermeer, S., Connell, F., Verheij, J., Manouvrier, S., Pierquin, G., Odent, S., Holder-Espinasse, M., Vincent-Delorme, C., Gillerot, Y., Vanwijck, R., Bayet, B. and Vikkula, M., (2010). IRF6 Screening of Syndromic and a priori Non-Syndromic Cleft Lip and Palate Patients: Identification of a New Type of Minor VWS Sign. Mol Syndromol.; 1(2):67–74.

Dixon, M. J., Marazita, M. L., Beaty, T. H. and Murray, J. C., (2011). Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet.; 12(3):167–178.

Donkor, P., Bankas, D. O., Agbenorku, P., Plange-Rhule, G. and Ansah S. K., (2007a). Cleft lip and palate surgery in Kumasi, Ghana: 2001-2005. J Craniofac Surg.; 18(6):13761379.

Donkor, P., Plange-Rhule, G. and Amponsah, E. K., (2007b). A prospective survey of patients with cleft lip and palate in Kumasi. West Afr J Med.; 26(1):14-16.

Du, X., Tang, W., Tian, W., Li, S., Li, X., Liu, L., Zheng, X., Chen, X., Lin, Y. and Tang, Y., (2006). Novel IRF6 Mutations in Chinese Patients with Van der Woude Syndrome. J Dent Res.; 85(10):937-940.

Eshete, M., Butali, A., Deressa, W., Pagan-Rivera, K., Hailu, T., Abate, F., Mohammed, I., Demissie, Y., Hailu, A., Dawson, D. V., Deribew, M., Gessese, M., Gravem, P. E., Mossey, P. (2016). Descriptive Epidemiology of Orofacial Clefts in Ethiopia. J Craniofac Surg.; Epub ahead of print. DOI: 10.1097/SCS.0000000000003234.

Estandia-Ortega, B., Velazquez-Aragon, J. A., Alcanta-Ortigoza, M. A., Reyna-Fabian, M. E., Villagomez-Martinez, S. and Gonzalez-Del Angel A., (2014). 5,10Methylenetetrahydrofolate reductase single nucleotide polymorphisms and geneenevironment interaction analysis in non-syndromic cleft lip/palate. Eur J Oral Sci.; 122(2):109-113.

Fakhouri, W. D., Rahimov, F., Attanasio, C., Kouwenhoven, E. N., Ferreira De Lima, R. L., Felix, T. M., Nitschke, L., Huver, D., Barrons, J., Kousa, Y. A., Leslie, E., Pennacchio, L. A., Bokhoven, H. V., Visel, A., Zhou, H., Murray, J. C. and Schutte, B. C., (2014). An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. Hum Mol Genet.; 23(10):2711–2720.

Ferreira de Lima, R. L. L., Hoper, S. A., Ghassibe, M., Cooper, M. E., Rorick, N. K., Kondo, S., Katz, L., Marazita, M. L., Compton, J., Bale, S., Hehr, U., Dixon, M. J., DaackHirsch, S., Boute, O., Bayet, B., Revencu, N., Verellen-Dumoulin, C., Vikkula, M., Richieri-Costa, A., Moretti-Ferreira, D., Murray, J. C. and Schutte, B. C., (2009). Prevalence and non-random distribution of exonic mutations in Interferon Regulatory Factor 6 (IRF6) in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. Genet Med.; 11(4):241–247.

Figueiredo J. C., Ly S., Raimondi H., Magee K., Baurley J. W., Sanchez-Lara P. A., Ihenacho U., Yao C., Edlund C. K. and van den Berg D. (2014). Genetic risk factors for orofacial clefts in Central Africans and Southeast Asians. Am J Med Genet A.; 164(10):2572– 2580.

Fontoura, C., Silva, R. M., Granjeiro, J. M. and Letra, A., (2012). Further evidence of association of the ABCA4 gene with cleft lip/palate. Eur J Oral Sci.; 120(6):553–557.

Garg, P., Ludwig, K. U., Bohmer, A. C., Rubini, M., Steegers-Theunissen, R., Mossey, P. A., Mangold, E. and Sharp, A. J., (2014). Genome-wide analysis of parent-of-origin effects in non-syndromic orofacial clefts. Eur J Hum Genet.; 22(6):822–830.

Gatta, V., Scarciolla, O., Cupaioli, M., Palka, C., Chiesa, P. L. and Stuppia, L., (2004). A novel mutation of the IRF6 gene in an Italian family with Van der Woude syndrome. Mutat Res.; 547(1-2):49–53.

Ghassibe´, M., Bayet, B., Revencu1, N., Verellen-Dumoulin, C., Gillerot, Y., Vanwijck, R. and Vikkula, M., (2004). Six families with van der Woude and/or popliteal pterygium syndrome: all with a mutation in the IRF6 gene. J Med Genet.; 41(2):e15.

Ghassibe, M., Bayet, B., Revencu, N., Verellen-Dumoulin, C., Gillerot, Y., Vanwijck, R. and Vikkula, M., (2005). Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. Eur J Hum Genet.; 13(11):1239–1242.

Goodnough, L. H., Brugmann, S. A., Hu, D. and Helms, J. A., (2007). Stage-dependent craniofacial defects resulting from sprouty2 overexpression. Dev. Dyn.; 236(7):19181928.

Gorlin, R. J., Cohen, M. M. and Hennekam, R. C. M. (2001). Syndromes of the neck and head. Oxford University Press, New York.

Gowans, L. J. J., Donkor, P., Agbenorku, P., Plange-Rhule, G., Obiri-Yeboah , S., Arthur, F. K. N., Twumasi, P., Murray J.C. and Butali, A., (2015). Environmental Factors Contributing to Genetic Mutations in the Pathogenesis of Human Orofacial Clefts in Ghana. 72nd Annual Meeting of American Cleft Palate-Craniofacial Association; 244:Poster Presentation.

Grant, S. F. A., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C. E., Bradfield, P. J., Glessner, J. T., Thomas, K. A., Garris, M., Frackelton, E. C., Otieno, F. G., Chiavacci, R. M., Nah, H. D., Kirschner, R. E. and Hakonarson, H., (2009). A genomewide association study identifies a locus for non-syndromic cleft lip with or without cleft palate on 8q.24. J Pediatr.; 155(6):909–913.

Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Molsted, K., Sivertsen, A., Murray, J. C. and Christensen, K., (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: Support for the multifactorial threshold model of inheritance. J Med Genet; 47(3):162–168.

Hallonet, M., Holleman, T., Pieler, T. and Grruss, P., (1999). Vax1, a novel homeoboxcontaining gene, directs development of the basal forebrain and visual system. Genes Dev.; 13(23):3106-3114.

Henn, B. M., Botigué, L. R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B. K., Martin, A. R., Musharoff, S., Cann, H., Snyder, M. P., Excoffier, L., Kidd, J. M., Bustamante, C. D., (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci USA; 113(4):E440-9. doi: 10.1073/pnas.1510805112.

Hor, H., Francescatto, L., Bartesaghi, L., Ortega-Cubero, S., Kousi, M., Lorenzo-Betancor, O., Jimenez-Jimenez FJ, Gironell A, Clarimon J, Drechsel O, Agundez JA, Kenzelmann

Broz, D., Chiquet-Ephrismann, R., Lleo, A., Coria, F., Garcia-Martin, E., AlonsoNavarro, H., Marti, M. J., Kulisevsky, J., Hor, C. N., Ossowski, S., Chrast, R., Katsanis, N., Pastor, P. and Estivil, X., (2015). Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. Hum Mol Genet.; 24(20):56775686.

Huang, Y. Q., Ma, J., Ma, M., Deng, Y., Li, Y. D., Ren, H. W., Zhao, G. Z., Guo, S. S., Wang, Y. Y., Zhang G. X. and Shi, B., (2011). Association between MSX1 variants and oral clefts in Han Chinese in western China. DNA Cell Biol.; 30(12):1057-1061.

Hurst, J. A., Houlston, R.S., Roberts, A., Gould, S.J. and Tingey, W.G., (1995). Transverse limb deficiency, facial clefting and hypoxic renal damage: an association with treatment of maternal hypertension. Clin Dysmorphol.; 4(4):359–363.

Ingraham, C. R., Kinoshita, A., Kondo, S., Yand, B., Sajan, S., Trout, K. J., Malik, M. I., Dunnwald, M., Guddy, S. L., Lovett, M., Murray, J. C. and Schutte, B. C., (2006). Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (irf6). Nat Genet.; 38(11):1335-1340.

Iwata, J., Suzuki, A., Pelikan, R. C., Ho, T., Sanchez-Lara, P. A., Urata, M., Dixon, M. J. and Chai, Y., (2013). Smad4-Irf6 genetic interaction and TGFβ-mediated IRF6 signaling cascade are crucial for palatal fusion in mice. Development; 140(6):1220-1230.

Jagomagi, T., Nikopensius, T., Krjutskov, K., Tammekivi, V., Viltrop, T., Saag, M. and Metspalu A., (2010). MTHFR and MSX1 contribute to the risk of nonsyndromic cleft lip/palate. Eur J Oral Sci.; 118(3):213–220.

Jia, .Z, Li, Y., Li, L., Wu, J., Zhu, L., Yang, C., Chen, C. and Shi, B., (2009). Association Among IRF6 Polymorphism, Environmental Factors, and Nonsyndromic Orofacial Clefts in Western China. DNA and Cell Biol.; 28(5):249–257.

Jia, Z., Leslie, E. J., Cooper, M. E., Butali, A., Standley, J., Rigdon, J., Suzuki, S., Gongorjav, A., Shonkhuuz, T. E., Natsume, N., Shi, B., Marazita, M. L. and Murray, J. C., (2015). Replication of 13q31.1 association in nonsyndromic cleft lip with cleft palate in Europeans. Am J Med Genet A.; 167A(5):1054–1060.

Jianyan, L., Zeqiang, G., Yongjuan, C., Kaihong, D., Bing, D. and Rongsheng, L., (2010). Analysis of interactions between genetic varainats of BMP4 and environmental factors with nonsyndromic cleft lip with or without cleft palate susceptibility. Int J Oral Maxillofac Surg.; 39(1):50-56.

Jones, J., Canady, J. W., Brookes, J. T., Wehby, G. L., Schutte, B. C., Murray, J. C., and Dunnwald, M., (2010). Wound Complications Following Cleft Repair in Children with Van der Woude Syndrome. J Craniofac Surg.; 21(5):1350–1353.

Jugessur, A., Rahimov, F., Lie, R. T., Wilcox, A. J., Gjessing, H. K., Nilsen, R. M., Nguyen, T. T. and Murray J. C., (2008). Genetic Variants in IRF6 and the Risk of Facial Clefts: Single-Marker and Haplotype-Based Analyses in a Population-Based Case-Control Study of Facial Clefts in Norway. Genet Epidemiol.; 32(5):413–424.

Jugessur, A., Shi, M., Gjessing, H. K., Lie, R. T., Wilcox, A. J., Weinberg, C. R., Christensen, K., Boyles, A. L., Daack-Hirsch, S., Trung, T. N., Bille, C., Lidral, A. C. and Murray, J. C., (2009). Genetic Determinants of Facial Clefting: Analysis of 357 Candidate Genes Using Two National Cleft Studies from Scandinavia. PLoS ONE; 4(4):e5385.

Kerameddin, S., Namipashaki, A., Ebrahimi, S. and Ansari-Pou, N., (2015). IRF6 Is a Marker of Severity in Nonsyndromic Cleft Lip/Palate. J Dent Res. 94(9 Suppl):226S232S.

Kim, N. Y., Kim, Y. H., Park J. W. and Baek, S., (2013). Association between MSX1 SNPs and Nonsyndromic Cleft Lip with or without Cleft Palate in the Korean Population. J Korean Med Sci.; 28(4):522-526.

Kitase, Y. and Shuler, C. F., (2012). Multi-layered hypertrophied MEE formation by microtubule disruption via GEFH1/RhoA/ROCK signaling pathway. Dev Dyn.; 241(7):1169-1182.

Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A. S., Watanabe, Y., Howard, E., Ferreira de Lima, R. L. L., Daack-Hirsch, S., Sander, A., McDonaldMcGinn, D. M., Zackai, E. H., Lammer, E. J., Aylsworth, A. S., Ardinger, H. H., Lidral, A. C., Pober B. C., Moreno, L., Arcos-Burgos, M., Valencia, C., Houdayer, C., Bahuau, M., Moretti-Ferreira, D., Richieri-Costa, A., Dixon, M. J., and Murray, J. C., (2002). Mutations in IRF6 cause van der Woude and popliteal pterygium syndromes. Nat Genet.; 32(2):285–289.

Kwa, M. Q., Huynh, J., Reynolds, E. C., Hamilton, J. A. and Scholz, G. M., (2015). Diseaseassociated mutations in IRF6 and RIPK4 dysregulate their signalling functions. Cell Signal.; 27(7):1509–1516.

Lennon, C. J., Birkeland, A. C., Nunez, J. A. P., Su, G. H., Lanzano, P., Guzman, E., Celis, K., Eisig, S. B., Hoffman, D., Rendon, M. T. G., Ostos, H., Chung, W. K. and Haddad Jr., J., (2012). Association of Candidate Genes With Nonsyndromic Clefts in Honduran and Colombian Populations. Laryngoscope; 122(9):2082–2087.

Leslie, E. J., Mansilla, M. A., Biggs, L. C., Schuette, K., Bullard, S., Cooper, M., Dunnwald, M., Lidral, A. C., Marazita, M. L., Beaty, T. H. and Murray, J. C., (2012). Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome wide association on chromosome 1p22. Birth Defects Res A Clin Mol Teratol; 94(11):934–942.

Leslie, E. J., Mancuso, J. L., Schutte, B. C., Cooper, M. E., Durda, K. M., L'Heureux, J., Zucchero, T. M., Marazita, M. L. and Murray, J. C., (2013). Search for Genetic Modifiers of IRF6 and Genotype-Phenotype Correlations in Van der Woude and Popliteal Pterygium Syndromes. Am J Med Genet A.; 161A(10):2535–2544.

Leslie, E. J., Taub, M. A., Liu, H., Steinberg, K. M., Koboldt, D. C., Zhang, Q., Carlson, J. C., Hetmanski, J. B., Wang, H., Larson, D. E., Fulton, R. S., Kousa, Y. A., Fakhouri, W. D., Naji, A., Ruczinski, I., Begum, F., Parker, M. M., Busch, T., Standley, J., Rigdon, J., Hecht, J. T., Scott, A. F., Wehby, G. L., Christensen, K., Czeizel, A. E., Deleyiannis, F. W.B., Schutte, B. C., Wilson, R. K., Cornell, R. A., Lidral, A. C., Weinstock, G. M., Beaty, T. H., Marazita, M. L. and Murray, J. C., (2015a). Identification of Functional Variants for Cleft Lip with or without Cleft Palate in or near PAX7, FGFR2, and NOG by Targeted Sequencing of GWAS Loci. Am J Hum Genet.; 96(3):397–411.

Leslie, E. J., O'Sullivan, J., Cunningham, M. L., Singh, A., Goudy, S. L., Ababneh, F., Alsubaie, L., Ch'ng, G., van der Laar, I. M. B. H., Hoogeboom, A. J. M., Dunnwald, M., Kapoor, S., Jiramongkolchai, P., Standley, J., Manak, R. J., Murray, J. C. and Dixon, M. J., (2015b). Expanding the Genetic and Phenotypic Spectrum of Popliteal Pterygium Disorders. Am J Med Genet A.; 167A(3):545–552.

Letra, A., Menezes, R., Govil, M., Fonseca, R. F., McHenry, T., Granjeiro, J. M., Castilla, E. E., Orioli, I. M., Marazita, M. L. and Vieira, A. R., (2010). Follow-Up Association Studies

of Chromosome Region 9q and Nonsyndromic Cleft Lip/Palate. Am J Med Genet A.; 152A(7):1701–1710.

Letra, A., Fakhouri, W., Fonseca, R. F., Menezes, R., Kempa, I., Prasad, J. L., McHenry, T. G., Lidral, A. C., Moreno, L., Murray, J. C., Daack-Hirsch, S., Marazita, M. L., Castilla, E. E., Lace, B., Orioli, I. M., Granjeiro, J. M., Schutte, B. C. and Vieira, A. R., (2012a). Interaction between IRF6 and TGFA genes contribute to the risk of nonsyndromic cleft lip/palate. PloS ONE; 7(9):e45441.

Letra, A., Bjork, B., Cooper, M. E., Szabo-Rogers, H., Deleyiannis, F. W., Field, L. L., Czeizel, A. E., Ma, L., Garlet, G. P., Poletta, F. A., Mereb, J. C., Lopez-Camelo, J. S., Castilla, E. E., Orioli, I. M., Wendell, S., Blanton, S. H., Liu, K., Hetch, J. T., Marazita, M. L., Vieira, A. R. and Silva, R. M., (2012b). Association of AXIN2 with nonsyndromic oral clefts in multiple populations. J Dent Res.; 91(5):473-478.

Letra, A., Maili, L., Mulliken, J. B., Buchanan, E., Blanton, S. H. and Hecht, J. T., (2014). Further Evidence Suggesting a Role for Variation in ARHGAP29 Variants in Nonsyndromic Cleft Lip/Palate. Birth Defects Res A Clin Mol Teratol.; 100(9):679–685.

Li, Q., Kim, Y., Suktitipat, B., Hetmanski, J. B., Marazita, M. L., Duggal, P., Beaty, T. H. and Bailey-Wilson, J. E., (2014). Gene-Gene Interaction Among WNT Genes for Oral Cleft in Trios. Genet Epidemiol.; 39(5):385-94.

Liang, J., Zhu, L., Meng, L., Chen, D. and Bian, Z., (2012). Novel nonsense mutation in MSX1 causes tooth agenesis with cleft lip in a Chinese family. Eur J Oral Sci; 120(4):278–282.

Lidral, A. C., Liu, H., Bullard, S. A., Bonde, G., Machida, J., Visel, A., Moreno Uribe, L. M., Li, X., Amendt, B. and Cornell, R. A., (2015). A Single Nucleotide Polymorphism Associated with Isolated Cleft Lip and Palate, Thyroid Cancer and Hypothyroidism Alters the Activity of an Oral Epithelium and Thyroid Enhancer Near FOXE1. Hum Mol Genet.; 24(14):3895-3907.

Loenarz, C., Ge, W., Coleman, M. L., Rose, N. R., Cooper, C. D. O., Klose, R. J., Ratcliffe, P. J. and Schofield, C. J., (2009). PHF8, a gene associated with cleft lip/palate and mental retardation, encodes an N{varepsilon}-dimethyl lysine demethylase. Hum Mol Genet.; 19(2):217-222.

Lu, H Z. and Hu, J. G., (2009). Expression of bone morphogenetic proteins-2/4 in neural stem cells and their lineages. Acta Neurobiol Exp. (Wars); 69(4):441-447.

Lu Y., Liu, Q., Xu, W., Li, Z., Jiang, M., Li, X., Zhao, N., Liu, W., Sui, Y., Ma, C., Feng, W., Han, W. and Li, Jianxin., (2013). TGFA and IRF6 Contribute to the Risk of Nonsyndromic Cleft Lip with or without Cleft Palate in Northeast China. PLoS ONE; 8(8):e70754.

Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., Alchawa, T., Nasser, E., Bohmer, A. C., Mattheisen, M., Alblas, M. A., Barth, S., Kluck, N., Lauster, C., Braumann, B., Reich, R. H., Hemprich, A., Potzsch, S., Blaumeiser, B., Daratsianos, N., Kreusch, T., Murray, J. C., Marazita, M. L., Ruczinski, I., Scott, A. F., Beaty, T. H., Kramer, F. J., Wienker, T. F., Steegers-Theunissen, R. P., Rubini, M., Mossey, P. A., Hoffmann, P., Lange, C., Cichon, S., Propping, P., Knapp, M. and Nothen M. M., (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet.; 44(9):968–971.

Ludwig, K.U., Böhmer, A.C., Rubini, M., Mossey, P.A., Herms, S., Nowak, S., Reutter, H., Alblas, M.A., Lippke, B., Barth, S., Paredes-Zenteno, M., Muñoz-Jimenez, S.G., OrtizLopez, R., Kreusch, T., Hemprich, A., Martini, M., Braumann, B., Jäger, A., Pötzsch,

B., Molloy, A., Peterlin, B., Hoffmann, P., Nöthen, M.M., Rojas-Martinez, A., Knapp, M., Steegers-Theunissen, R.P. and Mangold E., (2014). Strong Association of Variants around FOXE1 and Orofacial Clefting. J Dent Res.; 93(4):376-381.

Ma, L., Xu, M., Li, D., Han, Y., Wang, Z., Yuan, H., Ma, J., Zhang, W., Jiang1, H., Pan, Y. and Wang, L., (2014). A miRNA-binding-site SNP of MSX1 is Associated with NSOC Susceptibility. J Dent Res.; 93(6):559-564.

Malik, S., Wilcox, E. R. and Naz, S., (2014). Novel lip pit phenotypes and mutations of IRF6 in Van der Woude syndrome patients from Pakistan. Clin Genet.; 85(5):487–491.

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N. A., Chawa, T. A., Mattheisen, M., Steffens, M., Barth, S., Kluck, N., Paul, A., Becker, J., Lauster, C., Schmidt, G., Braumann, B., Scheer, M., Reich, R. H., Hemprich, A., Potzsch, S., Blaumeiser, B., Moebus, S., Krawczak, M., Schreiber, S., Meitinger, T., Wichmann, H. E., Steegers-Theunissen, R. P., Kramer, F. J., Cichon, S., Propping, P., Wienker, T. F., Knapp, M., Rubini, M., Mossey, P. A., Hoffmann, P. and Nothen, M. M., (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet.; 42(1):24–26.

Marazita, M.L., Lidral, A. C., Murray, J. C., Field, L. L., Maher, B. S., McHenry, T. G., Cooper, M. E., Govil, M., Daack-Hirsch, S., Riley, B., Jugessur, A., Felix, T., Moreno, L., Mansilla, M. A., Vieira, A. R,, Doheny, K., Pugh, E., Valencia-Ramirez, C. and Arcos-Burgos, M., (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype specific differences in linkage and association results. Hum Hered.; 68(3):151–170.

Martinelli, M., Girardi, A., Cura, F., Nouri, N., Pinto, V., Carinci, F., Morselli, P. G., Salehi, M. and Scapoli, L., (2016). Non-syndromic cleft lip with or without cleft palate in Asian populations: Association analysis on three gene polymorphisms of the folate pathway. Arch Oral Biol.; 61:79-82.

Mathias, R. A., Taub, M. A., Gignoux, C. R., Fu, W., Musharoff, S., O'Connor, T. D., Vergara, C., Torgerson, D. G., Pino-Yanes, M., Shringarpure, S. S., Huang, L., Rafaels, N., Boorgula, M. P., Johnston, H. R., Ortega, V. E., Levin, A. M., Song, W., Torres, R., Padhukasahasram, B., Eng, C., Mejia-Mejia, D. A., Ferguson, T., Qin, Z. S., Scott, A. F., Yazdanbakhsh, M., Wilson, J. G., Marrugo, J., Lange, L. A., Kumar, R., Avila, P. C., Williams, L. K., Watson, H., Ware, L. B., Olopade, C., Olopade, O., Oliveira, R., Ober, C., Nicolae, D. L., Meyers, D., Mayorga, A., Knight-Madden, J., Hartert, T., Hansel, N. N., Foreman, M. G., Ford, J. G., Faruque, M. U., Dunston, G. M., Caraballo, L., Burchard, E. G., Bleecker, E., Araujo, M. I., Herrera-Paz, E. F., Gietzen, K., Grus, W. E., Bamshad, M., Bustamante, C. D., Kenny, E. E., Hernandez, R. D., Beaty, T. H., Ruczinski, I., Akey, J.; CAAPA, Barnes, K. C., (2016). A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun.; 7:12522.

Matsuzawa, N., Yoshiura, K., Machida, J., Nakamura, T., Niimi, T., Furukawa, H., Toyoda, T., Natsume, N., Shimozato, K. and Niikawa, N., (2004). Two missense mutations in the IRF6 gene in two Japanese families with Van der Woude syndrome. Oral Surg Oral Med Oral Pathol Oral Radiol Endod.; 98(4):414-417.

Matsuzawa, N., Shimozato, K., Natsume, N., Niikawa, N. and Yoshiura, K., (2006). A Novel Missense Mutation in Van der Woude Syndrome: Usefulness of Fingernail DNA for Genetic Analysis. J Dent Res.; 85(12):1143-1146.

Matsuzawa, N., Kondo, S., Shimozato, K., Nagao, T., Nakano, M., Tsuda, M., Hirano, A., Niikawa, N. and Yoshiura, K-i., (2010). Two missense mutations of the IRF6 gene in two Japanese families with popliteal pterygium syndrome. Am J Med Genet A.; 152A(9):2262–2267.

McEvoy, B. P., Powell, J. E., Goddard, M. E. and Visscher, P. M., (2011). Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res.; 21(6): 821–829.

Mi, N., Hao, Y., Jiao, X., Zheng, X., Song, T., Shi, J. and Donga, C., (2014). Association study of single nucleotide polymorphisms of MAFB with non-syndromic cleft lip with or without cleft palate in a population in Heilongjiang Province, northern China. Br J Oral Maxillofac Surg.; 52(8):746–750.

Mijiti, A., Ling, W., Maimaiti, A., Tuerdi, M., Tuerxun, J. and Moming, A., (2015). Preliminary evidence of an interaction between the CRISPLD2 gene and non-syndromic cleft lip with or without cleft palate (nsCL/P) in Xinjiang Uyghur population, China. Int J Pediatr Otorhinolaryngol.; 79(2):94-100.

Modesto, A., Moreno, L.M., Krahn1, K., King, S. and Lidral1, A.C., (2006). MSX1 and Orofacial Clefting with and without Tooth Agenesis. J Dent Res.; 85(6):542–546.

Moreno, L. M., Mansilla, M. A., Bullard, S. A., Cooper, M. E., Busch, T. D., Machida, J., Johnson, M. K., Brauer, D., Krahn, K., Daack-Hirsch, S., L'Heureux, J., ValenciaRamirez, C., Rivera, D., Lo´pez, A. M., Moreno, M. A., Hing, A., Lammer, E. J., Jones, M., Christensen K., Lie, R. T., Jugessur, A., Wilcox, A. J., Chines, P., Pugh, E., Doheny, K., Arcos-Burgos, M., Marazita, M. L., Murray, J. C. and Lidral, A. C., (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum Mol Genet.; 18(24):4879–4896.

Mossey, P. A. and Modell, B., (2012). Epidemiology of oral clefts 2012: an international perspective. Front Oral Biol.; 16:1–18.

Mostowska, A., Hozyasz, K. K., Wojcicki, P., Biedziak, B., Paradowska, P. and Jagodzinski, P. P., (2010). Association between genetic variants of reported candidate genes or regions and risk of cleft lip with or without cleft palate in the Polish population. Birth Defects Res A Clin Mol Teratol.; 88(7):538–545.

Mostowska, A., Hozyasz, K. K., Wojcicka, K., Biedziak, B. and Jagodzinski, P. P., (2012a). Polymorphic variants at 10q25.3 and 17q22 loci and the risk of non-syndromic cleft lip and palate in the Polish population. Birth Defects Res A Clin Mol Teratol.; 98(1):42-46.

Mostowska, A., Hozyasz, K. K., Wojcicki, P., Lasota, A., Dunin-Wilczynskal, I. and Jagodzinski, P. P., (2012b). Association of DVL2 and AXIN2 gene polymorphisms with cleft lip with or without cleft palate in a Polish population. Birth Defects Res A Clin Mol Teratol.; 94(11):943-9450.

Mostowska, A., Hozyasz, K. K., Wojcicki, P., Zukowski, K., Dabrowska, A., Lasota, A., Zadurska, M., Radomska, A., Dunin-Wilczynska, I. and Jagodzinski, P. P., (2015). Association between polymorphisms at the GREM1 locus and the risk of nonsyndromic cleft lip with or without cleft palate in the Polish population. Birth Defects Res A Clin Mol Teratol.; 103(10):847-56.

Murdoch, B., DelConte, C. and Garcia-Castro, M. I., (2012). Pax7 Lineage Contributions to the Mammalian Neural Crest. PLoS ONE; 7(7):e41089.

Murray, J. C., (2002). Gene/environment causes of cleft lip and/or palate. Clin Genet.; 61(4): 248–256.

Nasser, E,, Mangold, E., Tradowsky, D. C., Fier, H., Becker, J., Boehmer, A. C., Herberz, R., Fricker, N., Barth, S., Wahle, P., Nowak, S., Reutter, H., Reich, R. H., Lauster, C., Braumann, B., Kreusch, T., Hemprich, A., Potzsch, B., Hoffmann, P., Kramer, F. J., Knapp, M., Lange, C., Nothen, M. M. and Ludwig, K. U., (2012). Resequencing of VAX1 in patients with nonsyndromic cleft lip with or without cleft palate. Birth Defects Res A Clin Mol Teratol.; 94(11):925–933.

Neiswanger, K., Walker, K., Klotz, C. M., Cooper, M. E., Bardi,K. M., Brandon, C. A., Weinberg, S. M., Vieira, A. R., Martin, R. A., Czeizel, A. E., Castilla, E. E., Poletta, F. A. and Marazita, M. L., (2009). Whorl patterns on the lower lip are associated with nonsyndromic cleft lip with or without cleft palate. Am J Med Genet A.; 149A(12):26732679.

Nikopensius, T., Birnbaum, S., Ludwig, K. U., Jagomagi, T., Saag, M., Herms, S., Knapp, M., Hoffmann, P., Nothen, M. M., Metspalu, A. and Mangold, E., (2010b). Susceptibility locus for non-syndromic cleft lip with or without cleft palate on chromosome 10q25 confers risk in Estonian patients. Eur J Oral Sci.; 118(3):317–319.

Nikopensius, T., Kempa, I., Ambrozaityte, L., Jagomagi, T., Saag, M., Matuleviciene, A., Utkus, A., Krjutskov, K., Tammekivi, V., Piekuse, L., Akota, I., Barkane, B., Krumina, A., Klovins, J., Lace, B., Kucinskas, V. and Metspalu, A., (2011). Variation in FGF1, FOXE1, and TIMP2 genes is associated with nonsyndromic cleft lip with or without cleft palate. Birth Defects Res A Clin Mol Teratol.; 91(4):218-225.

Nopoulos, P., Richman, L., Andreasen, N. C., Murray, J. C. and Schutte, B., (2007). Abnormal brain structure in adults with Van der Woude syndrome. Clin Genet.; 71(6):511–517.

Nouri, N., Memarzadeh, M., Carinci, F., Cura, F., Scapoli, L., Nouri, N., Jafary, F., Sedghi, M., Sadri, L. and Salehi, M., (2014). Family-based association analysis between nonsyndromic cleft lip with or without cleft palate and IRF6 polymorphism in an Iranian population. Clin Oral Invest.; 19(4):891–894.

Pan, Y., Zhang, W., Du, Y., Tong, N., Han, Y., Zhang, H., Wang, M., Ma, J., Wan, L. and Wang, L., (2011). Different roles of two novel susceptibility loci for nonsyndromic orofacial clefts in a Chinese Han population. Am J Med Genet A.; 155A(9):2180–2185.

Pan, X., Wang, P., Yin, X., Liu, X., Li, D., Li, X., Wang, Y., Li, H. and Yu, Z., (2015). Association between Maternal MTHFR Polymorphisms and Nonsyndromic Cleft Lip with or without Cleft Palate in Offspring, A Meta-Analysis Based on 15 Case-Control Studies. 2015. Int J Fertil Steril.; 8(4):463-480.

Park, J. W., McIntosh, I., Hetmanski, J. B., Jabs, E. W., Vander Kolk, C. A., Wu-Chou, Y. H., Chen, P. K., Chong, S. S., Yeow, V., Jee, S. H., Park, B. Y., Fallin, M. D., Ingersoll, R., Scott, A. F. adn Beaty, T. H., (2007). Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. Genet Med.; 9(4):219–227.

Pegelow, M., Koillinen, H., Magnusson, M., Fransson, I., Unneberg, P., Kere, J., Karsten, A. and Peyrard-Janvid, M., (2014). Association and Mutation Analyses of the IRF6 Gene in Families With Nonsyndromic and Syndromic Cleft Lip and/or Cleft Palate. Cleft Palate Craniofac J.; 51(1):49–55.

Peyrard-Janvid, M., Pegelow, M., Koillinen, H., Larsson, C., Fransson, I., Rautio, J., Hukki, J., Larson, O., Karsten, A. L. A. and Kere, J., (2005). Novel and de novo mutations of the

IRF6 gene detected in patients with Van der Woude or popliteal pterygium syndrome. Eur J Hum Genet.; 13(12):1261–1267.

Peyrard-Janvid, M., Leslie, E. J., Kousa, Y. A., Smith, T. L., Dunnwald, M., Magnusson, M., Lentz, B. A., Unneberg, P., Fransson, I., Koillinen, H. K., Rautio, J., Pegelow, M., Karsten, A., Basel-Vanagaite, L., Gordon, W., Andersen, B., Svensson, T., Murray, J. C., Cornell, R. A., Kere, J., and Schutte, B. C., (2014). Dominant Mutations in GRHL3 Cause Van der Woude Syndrome and Disrupt Oral Periderm Development. Am J Hum Genet A.; 94(1):23–32.

Piotrowski, T., Schilling, T. F., Brand, M., Jiang, Y-J., Heisenberg, C-P., Beuchle, D., Grandel, H., van Eeden, F. J. M., Furutani-Seiki, M., Granato, M., Haffter, P., Hammerschmidt, M., Kane, D. A., Kelsh, R. N., Mullins, M. C., Odenthal, J., Warga, R. M., and Nüsslein-Volhard, C., (1996). Jaw and branchial arch mutants in zebrafish II: anterior arches and cartilage differentiation. Development; 123:345-356.

Rafighdoost, H., Hashemi, M., Narouei, A., Eskanadri-Nasab, E., Dashti-Khadivaki, G. and Taheri, M., (2013). Association Between CDH1 and MSX1 Gene Polymorphisms and the Risk of Nonsyndromic Cleft Lip and/or Cleft Palate in a Southeast Iranian Population. Cleft Palate Craniofac J.; 50(5):e98–e104.

Rahimov, F., Marazita, M. L., Visel, A., Cooper, M. E., Hitchler, M. J., Rubini, M., Domann, F. E., Govil, M., Christensen, K., Bille, C., Melbye, M., Jugessur, A., Lie, R. T., Wilcox, A. J., Fitzpatrick, D. R., Green, E.D., Mossey, P. A., Little, J., Steegers-Theunissen, R. P., Pennacchio, L. A., Schutte, B. C. and Murray, J. C., (2008). Disruption of an AP-2α binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet.; 40(11):1341–1347.

Richardson, R. J., Dixon, J., Jiang, R. and Dixon, M. J., (2009). Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. Hum Mol Genet.; 18(14):2632-2642.

Riley, B. M., Schultz, R. E., Cooper, M. E., Goldstein-McHenry, T., Daack-Hirsch, S., Lee, K. T., Dragan, E., Vieira, A. R., Lidral, A. C., Marazita, M. L. and Murray, J. C., (2007). A genome-wide linkage scan for cleft lip and cleft palate identifies a novel locus on 8p11-23. Am J Med Genet A.; 143A(8):846-852.

Rojas-Martinez, A., Reutter, H., Chacon-Camacho, O., Leon-Cachon, R. B. R., MunozJimenez, S. G., Nowak, S., Becker, J., Herberz, R., Ludwig, K. U., Paredes-Zenteno, M., Arizpe-Cantu, A., Raeder, S., Herms, S., Ortiz-Lopez, R., Knapp, M., Hoffmann, P., Nothen, M. M. and Mangold, E., (2010). Genetic Risk Factors for Nonsyndromic Cleft Lip with or without Cleft Palate in a Mesoamerican Population: Evidence for IRF6 and Variants at 8q24 and 10q25. Birth Defects Res A Clin Mol Teratol.; 88(7):535–537.

Rorick, N. K., Kinoshita, A., Weirather, J., Peyrard-Janvid, M., Ferreira de Lima, R.L. L., Dunnwald, M., Shanske, A. L., Moretti-Ferreira, D., Koillinen, H., Kere, J., Mansilla, M. A., Murray, J. C., Goudy, S. L. and Schutte B. C., (2011). Genomic Strategy Identifies a Missense Mutation in WD-Repeat Domain 65 (WDR65) in an Individual with Van der Woude Syndrome. Am J Med Genet A.; 155A(6):1314–1321.

Rutledge, K. D., Barger, C., Grant, J. H. and Robin, N. H., (2010). IRF6 mutations in mixed isolated familial clefting. Am J Med Genet A.; 152A(12):2107–2019.

Salahshourifar, I., Halim, A. S., Sulaiman, W. A. W. and Zilfalil, B. A., (2011). Contribution of MSX1 variants to the risk of non-syndromic cleft lip and palate in a Malay population. J Hum Genet.; 56(11):755–758.

Salahshourifar, I., Sulaiman, W. A. W., Halim, A. S. and Zilfalil, B. A., (2012). Mutation screening of IRF6 among families with non-syndromic oral clefts and identification of two novel variants: Review of the literature. Eur J Med Genet.; 55(6-7):389-393.

Sarode, G. S., Desai, R. S., Sarode, S. C. and Kulkarni, M.A., (2011). Van der Woude syndrome with an unusual intraoral finding. Indian J Dent Res.; 22(1):164-165.

Scapoli, L., Palmieri, A., Martinelli, M., Pezzetti, F., Carinci, P., Tognon, M. and Carinci, F., (2005). Strong Evidence of Linkage Disequilibrium between Polymorphisms at the IRF6 Locus and Nonsyndromic Cleft Lip With or Without Cleft Palate, in an Italian Population. Am J Hum Genet.; 76(1):180–183.

Schilling, T. F., Piotrowski, T., Grandel, H., Brand, M., Heisenberg, C-P., Jiang, Y-J., Beuchle, D., Hammerschmidt, M., Kane, D. A., Mullins, M. C., van Eeden, F. J. M., Kelsh, R. N., Furutani-Seiki, M., Granato, M., Haffter, P., Odenthal, J., Warga, R. M., Trowe, T. and Nüsslein-Volhard, C., (1996). Jaw and branchial arch mutants in zebrafish I: branchial arches. Development; 123:329-344.

Slavotinek, A. M., Chao, R., Vacik, T., Yahyavi, M., Abouzeid, H., Bardakjian, T., Schneider, A., Shaw, G., Sherr, E. H., Lemke, G., Youssef, M. and Schordere, D. F., (2012). Brief Report: VAX1 mutation associated with microphthalmia, corpus callosum agenesis and orofacial clefting – the first description of a VAX1 phenotype in humans. Hum Mutat.; 33(2):364–368.

Seo, Y., Park, J. W., Kim, Y. H. and Baekd, S., (2013). Associations between the risk of tooth agenesis and single-nucleotide polymorphisms of MSX1 and PAX9 genes in nonsyndromic cleft patients. Angle Orthodontist; 83(6):1036-1042.

Shi, J., Song, T., Jiao, X., Qin, C. and Zhou, J., (2011). Single-nucleotide polymorphisms (SNPs) of the IRF6 and TFAP2A in non-syndromic cleft lip with or without cleft palate (NSCLP) in a northern Chinese population. Biochem Biophys Res Commun.; 410(4):732–736.

Song, T., Wu, D., Wang, Y., Li, H., Yin, N. and Zhao, Z., (2013). SNPs and interaction analyses of IRF6, MSX1 and PAX9 genes in patients with non-syndromic cleft lip with or without palate. Mol Med Rep.; 8(4):1228-1234.

Srichomthong, C., Ittiwut, R., Siriwan, P., Suphapeetiporn, K. and Vorasuk, S., (2013). FOXE1 mutations in Thai patients with oral clefts. Genet Res. (Camb).; 95(5):133–137.

Stanier, P. and Moore, G. E., (2004). Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. Hum Mol Genet.; 13 Spec No 1:R73–R81.

Strachan, T. and Read, A., (2011). Human Molecular Genetics. 4th Edition, Gerland Science, Taylor and Francis Group, LLC.

Souza, L.T., Kowalski, T.W., Collares, M.V.M. and Felix, T.M., (2013). MSX1 gene and nonsyndromic oral clefts in a Southern Brazilian population. Braz J Med Biol Res.; 46(7):555-558.

Sozen, M. A., Tolarova, M. M. and Spritz, R. A., (2009). The common MTHFR C677T and A1298C variants are not associated with the risk of non-syndromic cleft lip/palate in northern Venezuela. J Genet Genomics; 36(5):283-288.

Suazo, J., Santos, J. L. and Blanco, R., (2010). Association between bone morphogenetic protein 4 gene polymorphisms with nonsyndromic cleft lip with or without cleft palate in a Chilean population. DNA Cell Biol.; 29(2):59-64.

Suazo J., Santos, J. L., Jara, L. and Blanco, R., (2010b). Parent-of-origin effects for MSX1 in a Chilean population with nonsyndromic cleft lip/palate. Am J Med Genet A.; 152A(8):2011–2016.

Suazo, J., Tapia, J. C., Santos, J. L., Castro, V. G., Colombo, A. and Blanco, R., (2011). Risk variants in BMP4 promoters for nonsyndromic cleft lip/palate in a Chilean population. BMC Med Genet.; 12:163.

Sull, J. W., Liang, K. Y., Hetmanski, J. B., Fallin, M. D., Ingersoll, R. G., Park, J., Wu-Chou, Y. H., Chen, P. K., Chong, S. S., Cheah, F., Yeow, V., Park, B. Y., Jee, S. H., Jabs, E. W., Redett, R., Scott, A. F. and Beaty, T. H., (2009). Maternal transmission effects of the PAX genes among cleft case-parent trios from four populations. Eur J Hum Genet.; 17(6):831–839.

Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., Du, Y., Wang, M., Lan, F., Hu, Z., Wang, G., Jiang, J M., Ma, J., Zhang, X., Ma, H., Ma, J., Zhang, W., Huang, Q., Zhou, Z., Ma, L., Li, Y., Jiang, H., Xie, L., Jiang, Y., Shi, B., Cheng, J., Shen, H., Wang, L. and Yang, Y., (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun.; 6:6414.

Suzuki, S., Marazita, M. L., Cooper, M. E., Miwa, N., Hing, A., Jugessur, A., Natsume, N., Shimozato, K., Ohbayashi, N., Suzuki, Y., Niimi, T., Minami, K., Yamamoto, M., Altannamar, T. J., Erkhembaatar, T., Furukawa, H., Daack-Hirschm S., L'Heureux, J., Brandonm, C. A., Weinberg, S.M., Neiswanger, K., Deleyiannis, F. W., de Salamanca, J. E., Vieira, A. R., Lidral, A. C., Martin, J. F. and Murray, J, C., (2009). Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. Am J Hum Genet.; 84(3):406–411.

Tan, E., Lim, E. C., Yap, S., Lee, S., Cheng, J., Por, Y. and Yeow, V., (2008). Identification of IRF6 gene variants in three families with Van der Woude syndrome. Int J Mol Med.; 21(6):747-751.

Tan, E., Lim, E. C. P. and Lee, S., (2013). De novo 2.3 Mb microdeletion of 1q32.2 involving the Van der Woude Syndrome locus. Mol Cytogenet.; 6:31.

Thomason, H. A., Zhou, H., Kouwenhoven, E.N., Dotto, G.P., Restivo, G., Nguyen, B. C., Little, H., Dixon, M. J., van Bokhoven, H. and Dixon. J., (2010). Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. J Clin Invest.; 120(5):1561–1569.

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C. and Spitz, F., (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat Genet. 46(7):753-758.

Vacik, T. and Lemke, G., (2011). Dominant-negative isoforms of Tcf/Lef proteins in development and disease. Cell Cycle; 10(24):4199-4200.

van den Boogaard, M. H., de Costa, D., Krapels, I. P. C., Liu, F., van Duijn, C., Sinke, R. J., Lindhout, D., Steegers-Theunissen, R. P. M., (2008). The MSX1 allele 4 homozygous child exposed to smoking at periconception is most sensitive in developing nonsyndromic orofacial clefts. Hum Genet.; 124(5):525–534.

Vieira, A. R., Castillo Taucher, S., Aravena, T., Astete, C., Sanz, P., Tastets, T. E., Monasterio, L. and Murray, J. C., (2004). Mutational analysis of the muscle segment homeobox gene 1 (MSX1) in Chilean patients with cleft lip/palte. Rev Med Chil.; 132(7): 816-822.

Vieira, A. R., McHenry, T. G., Daack-Hirsch, S., Murray, J. C. and Marazita, M. L., (2008). A genome wide linkage scan for cleft lip and palate and dental anomalies. Am J Med Genet A.; 146A(11):1406-1413.

Vieira, A. R., de Carvalho, F. M., Johnson, L., DeVos, L., Swailes, A. L., Weber, M. L. and Deeley, L., (2015). Fine Mapping of 6q23.1 Identifies TULP4 as Contributing to Clefts. Cleft Palate Craniofac J.; 52(2):128-134.

Wang, X., Liu, J., Zhang, H., Xiao, M., Li, J., Yang, C., Lin, X., Wu, Z., Hu, L. and Kong, X., (2003). Novel mutations in the IRF6 gene for Van der Woude syndrome. Hum Genet.; 113(5):382–386.

Wang, M., Pan, Y., Zhang, Z. and Wang, L., (2012). Three polymorphisms in IRF6 and 8q24 are associated with nonsyndromic cleft lip with or without cleft palate: Evidence from 20 studies. Am J Med Genet A.; 158A(12):3080–3086.

Weatherley-White, R. C., Ben, S., Jin, Y., Riccardi, S., Arnold, T. D. and Spritz, R. A., (2011). Analysis of genome-wide association signals for nonsyndromic cleft lip/palate in a Kenya African Cohort. Am J Med Genet A.; 155A(10):2422–2425.

Wehby G. and Cassell, C., (2010). The impact of orofacial clefts on quality of life and healthcare use and costs. Oral Dis.; 16(1):3–10.

Weinberg, S. M., Naidoo, S. D., Bardi, K. M., Brandon, C. A., Neiswanger, K., Resick, J. M., Martin, R. A. and Marazita, M. L., (2009), Face shape of unaffected parents with cleft affected offspring: combining three-dimensional surface imaging and geometric morphometrics. Orthod Craniofac Res.; 12(4):271-281.

Welsh, I. C., Hagge-Greenberg A. and O'Brien, T. P., (2007). A dosage-dependent role for spry2 in growth and patterning during palate development. Mech Dev.; 124(9-10):746761.

Wilkie, A. O. and Morriss-Kay, G. M., (2001). Genetics of craniofacial development and malformation. Nat Rev Genet.; 2(6):458-68.

Wolf, Z.T., Brand, H. A., Shaffer, J. R., Leslie, E. J., Arzi, B., Willet, C. E., Cox, T. C., McHenry, T., Narayan, N., Feingold, E., Wang, X., Sliskovic, S., Karmi, N., Safra, N., Sanchez, C., Deleyiannis, F. W. B., Murray, J. C., Wade, C. M., Marazita, M. L. and Bannasch, D. L., (2015). Genome-Wide Association Studies in Dogs and Humans Identify ADAMTS20 as a Risk Variant for Cleft Lip and Palate. PLoS Genet; 11(3): e1005059.

Wu, T., Liang, K. Y., Hetmanski, J. B., Ruczinski, I., Fallin, M. D., Ingersoll, R. G., Wang, H., Huang, S., Ye, X., Wu-Chou, Y., Chen, P. K., Jabs, E. W., Shi, B., Redett, R., Scott, A. F. and Beaty, T. H., (2010). Evidence of gene–environment interaction for the IRF6 gene and maternal multivitamin supplementation in controlling the risk of cleft lip with/without cleft palate. Hum Genet.; 128(4):401–410.

Wu, T. F., Yao, Y. L., Lai, I. L., Lai, C. C., Lin, P. L. and Yang, M. W., (2015). Loading of PAX3 to Mitotic Chromosomes Is Mediated by Arginine Methylation and Associated with Waardenburg Syndrome. J Biol Chem.; 290(33):20556-20564.

Wu-Chou, Y., Lo, L., Chen, K. P., Chang, C. F. and Chen, Y., (2013). A combined targeted mutation analysis of IRF6 gene would be useful in the first screening of oral facial clefts. BMC Med Genet.; 14:37.

Ye, X. Q., Jin, H. X., Shi, L. S., Fan, M. W., Song, G. T., Fan, H. L. and Bian, Z., (2005). Identification of novel mutations of IRF6 gene in Chinese families with Van der Woude syndrome. Int J Mol Med.; 16(5):851-856.

Yeetong P., Mahatumarat, C., Siriwan, P., Rojvachiranonda, N., Suphapeetiporn, K. and Shotelersuk, V., (2009). Three novel mutations of the IRF6 gene with one associated with an unusual feature in Van der Woude syndrome. Am J Med Genet A.; 149A(11):2489–2492.

Yuan, Q., Blanton, S. H. and Hecht, J. T., (2011). Association of ABCA4 and MAFB with nonsyndromic cleft lip with or without cleft palate. Am J Med Genet A.; 155A(6):1469–1471.

Zhao, L., Saitsu, H., Sun, X., Shiota, K. and Ishbashi, M., (2010). Sonic hedgehog is involved in formation of the ventral optic cup by limiting Bmp4 expression to the dorsal domain. Mech Dev.; 127(1-2):62-72.

Zhao, M., Ren, Y., Shen, L., Zhang, Y. and Zhou, B., (2014). Association between MTHFR C677T and A1298C polymorphisms and NSCL/P risk in Asians: a meta-analysis. PLoS ONE; 9(3):e88242.

Zhou, Q., Li, M., Zhu, W., Guo, J., Wang, Y., Li, Y. and Li, S., (2013). Association Between Interferon Regulatory Factor 6 Gene Polymorphisms and Nonsyndromic Cleft Lip With or Without Cleft Palate in a Chinese Population. Cleft Palate Craniofac J.; 50(5):570– 576.

Zucchero, T. M., Cooper, M. E., Maher, B. S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., Caprau, D., Christensen, K, Suzuki, Y., Machida, J., Natsume, N., Yoshiura, K-I., Vieira, A. R., Orioli, I. M., Castilla, E. E., Moreno, L., Arcos-Burgos, M, Lidral, A. C., Field, L. L., Liu, Y., Ray, A., Goldstein, T. H., Schultz, R. E., Shi, M., Johnson, M. K., Kondo, S., Schutte, B. C., Marazita, M. L. and Murray, J. C. (2004). Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med.; 351(8):769-780.

Web Resources (Bioinformatics Tools) OMIM: http://ghr.nlm.nih.gov/condition
Human Splicing Finder: http://www.umd.be/HSF3/
REDCaP: https://www.icts.uiowa.edu/apps/redcap
Sample Size Calculator: http://www.surveysystem.com/sscalc.htm
1000 Genomes: http://browser.1000genomes.org/index.html
UCSC Genome Browser: http://genome.ucsc.edu
PedCheck: https://watson.hgen.pitt.edu/register/docs/pedcheck.html
CD Map Generator: http://genetics.uiowa.edu/
PLINK: http://pngu.mgh.harvard.edu/~purcell/plink/
FBAT: http://www.hsph.harvard.edu/fbat/fbat.htm
Primer3: http://biotools.umassmed.edu/bioapps/primer3_www.cgi
Phred: http://www.phrap.org/phredphrapconsed.html, v.0.961028
Phrap: http://www.phrap.org
Polyphred: http://droog.gs.washington.edu/polyphred
Consed: http://www.phrap.org/consed/consed.html

Polyphen-2: http://genetics.bwh.harvard.edu/pph2/
SIFT: http://sift.jcvi.org/
Ensemble: http://www.ensembl.org/Homo_sapiens/Tools/VEP
Project Hope: http://www.cmbi.ru.nl/hope/method
RegulomeDB: http://regulomedb.org/
Exome Variant Server: http://evs.gs.washington.edu/EVS/
ExAC Browser: http://exac.broadinstitute.org/ dbSNP: www.ncbi.nlm.nih.gov/SNP/
Mutant Primer Design: http://www.genomics.agilent.com/primerDesignProgram
Chi Square tests: http://in-silico.net/tools/statistics/chi2test

Other URLs

http://elementsofmorp hology.nih.gov/anatomy-oral.shtml [Accessed 2015, January 11] http://www.cleftline.org/docs/Booklets/GEN-01.pdf [Accessed 2015, January 12] http://www.uic.edu/classes/bms/bms655/lesson11.html#MODEL [Accessed 2015, January 14] www.indiana.edu/~anat550/hnanim/face/face.html [Accessed 2015, January 17] http://emedicine.medscape.com/article/844962-overview [Accessed 2014, April 14] http://www.fastbleep.com/biology [Accessed 2015, January 25] www.mywage.org/ghana/home [Accessed 2015, May 30]
http://socratic.org/questions/how-do-dna-mutations-occur [Accessed 2015, January 25]
http://www.nlm.nih.gov/medlineplus/ency/article/000354.htm [Accessed 2015, July 9]

APPENDICES

Appendix A: Ethical approval letter

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF HEALTH SCIENCES
SCHOOL OF MEDICAL SCIENCES / KOMFO ANOKYE TEACHING HOSPITAL
COMMITTEE ON HUMAN RESEARCH, PUBLICATION AND ETHICS

Our Ref: CHRPE/AP/217/13

3rd October, 2013.

Mr. Lord Jephthah Joojo Gowans
Department of Biochemistry
and Biotechnology
KNUST.

Dear Sir,

LETTER OF APPROVAL

Protocol Title ***"Significance of Genetic Mutations in Molecular Pathogenesis of Cleft Lip, Cleft Palate and Cleft Lip and Palate among Selected Cases in Ghana."***

Proposed Site: ***Cleft Clinic, KATH, Kumasi and Murray laboratory, University of Iowa, Iowa City, Iowa, USA.***

Sponsor: ***Murray laboratory, University of Iowa, Iowa, Iowa, USA.***

Your submission to the Committee on Human Research, Publications and Ethics on the above named protocol refers.

The Committee reviewed the following documents:

- A notification letter of 21st March, 2013 from the Komfo Anokye Teaching Hospital (study site) indicating approval for the conduct of the study in the Hospital.
- A completed CHRPE Application Form.
- Participant Information Leaflet and Consent Form.
- Research Proposal.
- Questionnaire.
- Material Transfer Agreement.

The Committee has considered the ethical merit of your submission and approved the protocol. The approval is for a fixed period of one year, renewable annually thereafter. The Committee may however, suspend or withdraw ethical approval at anytime if your study is found to contravene the approved protocol.

Data gathered for the study should be used for the approved purposes only. Permission should be sought from the Committee if any amendment to the protocol or use, other than submitted, is made of your research data.

The Committee should be notified of the actual start date of the project and would expect a report on your study, annually or at close of the project, whichever one comes first. It should also be informed of any publication arising from the study.

Thank you Sir, for your application.

Yours faithfully,

Rev. Prof. John Appiah-Poku
Honorary Secretary
FOR: CHAIRMAN

Appendix B: Informed Consent Form

Participant Information Leaflet and Consent Form

This leaflet must be given to all prospective participants to enable them know enough about the research before deciding to or not to participate

Title of Research: Significance of genetic mutations in molecular pathogenesis of cleft lip, cleft palate and cleft lip and palate among selected cases in Ghana.

Name(s) and affiliation(s) of researcher(s): This study is being conducted by Lord Jephthah Joojo Gowans, a Biotechnology postgraduate student of the Department of Biochemistry and Biotechnology, KNUST and a collaborator of Jeff Murray Laboratory, University of Iowa, Iowa City, Iowa, USA. This research is being supervised by Dr Peter Twumasi and Dr F. K. N. Arthur, all of Department of Biochemistry and Biotechnology, KNUST and Professor Peter Donkor of School of Medical Sciences, KNUST.

Background: Cleft lip and/or palate have various patterns of inheritance. Genetic and/or environmental factors are taught to cause clefts. For example, a mutation or change in the IRF6 gene causes a genetic disorder called Van der Woude Syndrome, which increases the occurrence of clefts threefold. This same mutation also causes Popliteal Pterygium Syndrome (PPS), which also displays clefts as associated symptoms. This study thus mainly seeks to establish the role of genetic mutations in the occurrence of clefts in the study population.

Purpose(s) of research: The study mainly seeks to establish the pattern of inheritance of clefts in the study population and the significance of genetic mutations in the aetiology of clefts.

Procedure of the research, what shall be required of each participant and approximate total number of participants that would be involved in the research: About 135 cleft patients will be randomly selected from the Cleft Clinic of KATH every year. So for the about three year period that the study would be carried out, about 1215 participants (consisting of 405 families) are expected to take part in the study. These patients and their parents (and sometimes other family members) will be visited at their places of abode. The patients and their families will be physically examined for any genetic or phenotypic abnormality. Saliva or buccal/cheek swab samples will be taken from cleft patients and their parents, and sometimes other family members of the patients if the cleft is familial or syndromic. DNA will be extracted from the samples for genetic analyses. This will help establish the role of genetic mutations in cleft pathogenesis. The environment of participants will also be probed in order to establish the role of the environment in the pathogenesis of clefts. Participants will also be required to answer questions in a questionnaire. The principal investigator also requests your permission to have access to your medical record.

Individual samples, information and data may be stripped of identifiers (such as name, date of birth, address, etc) and sent to the USA (Murray Laboratory) and will be placed in a central repository. Data may be placed in the database of genotypes and phenotypes (dbGaP) and samples sent to other national repository at National Institute for Health (NIH). Other qualified researchers who obtain proper permission may gain access to the

sample and/or data for use in approved research studies that may or may not be related to craniofacial birth defects. Sample and data used through such repositories will be monitored and usage approved by the repositories' administrators. Samples may be used for sequencing (reading out all the genetic information) and / or genome wide studies (which are studies of genetic variation across the entire human genome). The tests the Principal Investigator (PI) and the Murray Laboratory might want to use to study the biological samples may not even exist at this time. Therefore, the Murray Laboratory requests the investigator's and subjects' permission to store biological sample so that the Laboratory can study it in the future. The Research Electronic Data Capture (REDCap) software of the University of Iowa, USA, will be used to manage samples and data in this research.

Risk(s): The study may cause discomfort due to collection of saliva or buccal swab samples. It may also cause inconvenience, in that the families would have to sacrifice their time and energy to help in the study. Finally, there is possible loss of confidentiality as the families will have to tell the investigator some of the medical conditions they may deem to be extremely dear or secrete to the family.

Benefit(s): Participants may know the possible cause of the cleft in the family. The participants would also be offered genetic counseling; they will be taught how to manage the clefts in the family. The family will also benefit psychologically, as they would be made to know that clefts are also "normal" medical conditions.

Confidentiality: All information collected in this study will be given code numbers, such as GH20130001_1_a for a proband, GH20130001_2_a for a father, GH20130001_3_a for a mother, etc. Names may be recorded. Data collected, however, cannot be linked to anybody in the published research. Thus, no name or identifier will be used in any publication or reports from this study. However, as part of research team's responsibility to conduct this research properly, the team may allow officials from ethics committees to have access to your records.

Voluntariness: Taking part in this study should be out of your own free will. Nobody is under any obligation to be part.

Alternatives to participation: Your participation or otherwise in this study will not affect your treatment in this hospital in any way.

Withdrawal from the research: You may choose to withdraw from the research at anytime without having to explain yourself. You may also choose not to answer any question you find uncomfortable or private.

Consequence of Withdrawal: There will be no consequence, loss of benefit or care to you if you choose to withdraw from the study. Please note however, that some of the information that may have been obtained from you without identifiers (name etc), before you chose to withdraw, may have been modified or used in analysis reports and

publications. These cannot be removed anymore. We do promise to make good effort to comply with your wishes as much as practicable.

Costs/Compensation: For your time and any inconvenience to you due to this study, the investigator may, in consultation with KATH, relay to you the genetic cause of the cleft and give you the necessary counsel.

Contacts: If you have any question concerning this study, please do not hesitate to contact Lord Jephthah Joojo Gowans on 0244866389/0261802515.

Further, if you have any concern about the conduct of this study, your welfare or your rights as a research participant, you may contact:

The Office of the Chairman
Committee on Human Research and Publication Ethics
Kumasi
Tel: 03220 63248 or 020 5453785

CONSENT FORM

Statement of person obtaining informed consent:

I have fully explained this research to _____ and have given sufficient information about the study, including that on procedures, risks and benefits, to enable the prospective participant make an informed decision to or not to participate.

DATE: _____

NAME: Lord Jephthah Joojo Gowans

Statement of person giving consent:

I have read the information on this study/research or have had it translated into a language I understand. I have also talked it over with the interviewer to my satisfaction.

I understand that my participation is voluntary (not compulsory).

I know enough about the purpose, methods, risks and benefits of the research study to decide that I want to take part in it.

I understand that I may freely stop being part of this study at any time without having to explain myself.

I have received a copy of this information leaflet and consent form to keep for myself.

NAME:_____

DATE: _____

SIGNATURE/THUMB PRINT: _____

Statement of person witnessing consent (Process for Non-Literate Participants):

I _____ (Name of Witness) certify that information given to

_____ (Name of Participant), in the local language, is a true reflection of what l have read from the study Participant Information Leaflet, attached.

WITNESS' SIGNATURE (maintain if participant is non-literate): _____

MOTHER'S SIGNATURE (maintain if participant is under 18 years): _____

MOTHER'S NAME: _____

FATHER'S SIGNATURE (maintain if participant is under 18 years): _____

FATHER'S NAME: _____

Appendix C: Questionnaire for genetic analyses

The response to this questionnaire is for academic purposes only. All information provided would be kept confidential.

SECTION A: Proband

Name:…… Sex:male/female Age:……

Type of cleft: □Cleft lip and palate (CLP) □Cleft lip (CL) □Cleft palate (CP) □Microform cleft □Submucous cleft palate (SMCP) □Facial cleft

Form of cleft: □Incomplete CP □Complete CP □Complete unilateral CLP □Complete bilateral CLP □Complete Unilateral CL □Incomplete Unilateral CL□ Complete Bilateral CL □ Incomplete Bilateral CL □Unilateral Facial Cleft□ Bilateral Facial Cleft □ Incomplete unilateral CLP □Incomplete bilateral CLP

Presence of other congenital phenotypic symptoms/abnormalities? (Yes/No)

VWS/PPS symptoms: □Lower lip pits □Lip malformation □Microforms □Syngnathia □Ankyloglossia □Pterygium □Finger and/or toe syndactyly □Oral synechiae □Genital abnormalities □Hypodontia □Alteration in brain structure/function □Hypernasal voice

Other phenotypic malformations observed: ……

SECTION B: Same-sex sibling of proband

Sex: male/female Age…… Presence of cleft? (Yes/No)
Type of cleft:……
Other phenotypic malformations: ……

SECTION C: Opposite-sex sibling of proband

Sex: male/female Age…… Presence of cleft? (Yes/No) Type of cleft:……

Other phenotypic malformations: ……

SECTION D: Father of proband

Age of father…… Age at which he gave birth to proband:……

Hometown/Ethnicity:……. Residential town: ……..

Occupation: …… Level of education:……..

Presence of cleft? (Yes/No) Type of cleft:……

Other phenotypic malformations: ……
SECTION E: Mother of proband
Age of mother…… Age at which she gave birth to proband:……
Hometown/Ethnicity:……. Residential town: ……..
Occupation: …… Level of education:……..
Presence of cleft? (Yes/No) Type of cleft:…… Other phenotypic malformations: ……

Exposure to possible teratogens: □Alcohol □Cigarette □Hypertensive drugs □Preterm □Pesticide □Anticonvulsant drugs □Nitrate compounds □Mining community □Lead/Lip sticks □Heroin/Cocaine □Hypoxia (smoky environment)□Attempted abortion □Consanguinity □Skin lightening cream □Enemas □Herbal remedies □Trauma □Family Planning □Miscarriage □Stillbirth

Maternal diet deficient in folate? (Yes/No) Access to antenatal care (Yes/No):……
Time of first ANC attendance:……

SECTION F: Paternal grandfather of proband

Age:…… Presence of cleft? (Yes/No)
Type of cleft:……..
Other phenotypic malformations: ……

SECTION G: Paternal grandmother of proband

Age:…… Presence of cleft? (Yes/No)
Type of cleft:……..
Other phenotypic malformations: ……
Possible teratogens encountered:……

SECTION H: Maternal grandfather of proband

Age:…… Presence of cleft? (Yes/No)

Type of cleft:……..

Other phenotypic malformations: ……

SECTION I: Maternal grandmother of proband

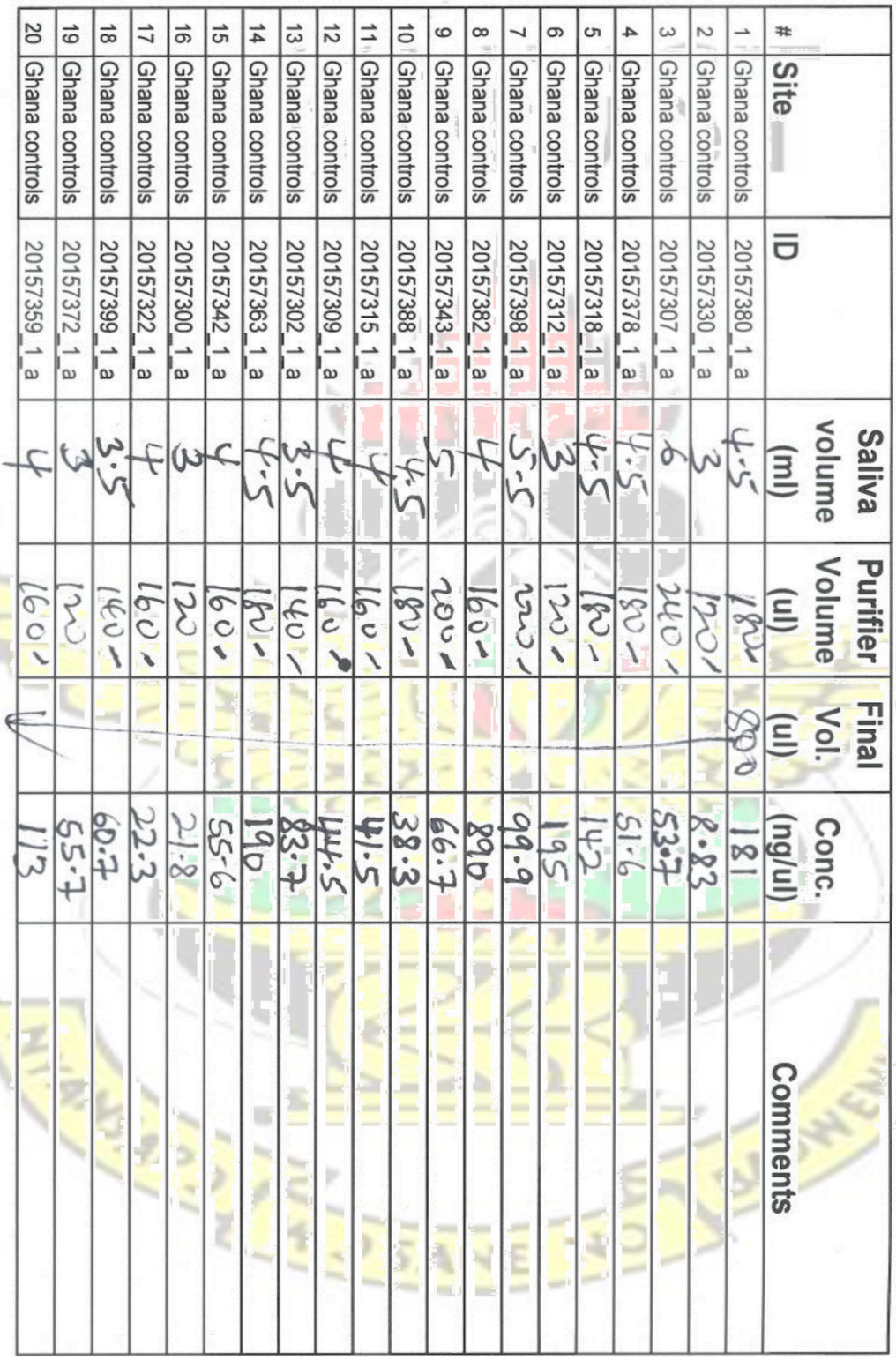
Age:…… Presence of cleft? (Yes/No)

Type of cleft:……..

Other phenotypic malformations: ……

Possible teratogens encountered:……

Appendix D: Saliva Processing Sheet



Ghana Controls Saliva Processing 6-7-2015 set 4

#	Site	ID	Saliva volume (ml)	Purifier Volume (ul)	Final Vol. (ul)	Conc. (ng/ul)	Comments
1	Ghana controls	20157380_1_a	4.5	180	800	181	
2	Ghana controls	20157330_1_a	3	120		8.83	
3	Ghana controls	20157307_1_a	6	240		53.7	
4	Ghana controls	20157378_1_a	4.5	180		51.6	
5	Ghana controls	20157318_1_a	4.5	180		142	
6	Ghana controls	20157312_1_a	3	120		195	
7	Ghana controls	20157398_1_a	5.5	220		99.9	
8	Ghana controls	20157382_1_a	4	160		89.0	
9	Ghana controls	20157343_1_a	5	200		66.7	
10	Ghana controls	20157388_1_a	4.5	180		38.3	
11	Ghana controls	20157315_1_a	4	160		41.5	
12	Ghana controls	20157309_1_a	4	160		144.5	
13	Ghana controls	20157302_1_a	3.5	140		83.7	
14	Ghana controls	20157363_1_a	4.5	180		190	
15	Ghana controls	20157342_1_a	4	160		55.6	
16	Ghana controls	20157300_1_a	3	120		21.8	
17	Ghana controls	20157322_1_a	4	160		22.3	
18	Ghana controls	20157399_1_a	3.5	140		60.7	
19	Ghana controls	20157372_1_a	3	120		55.7	
20	Ghana controls	20157359_1_a	4	160		113	

Appendix

E: Sample of genotype plate map

Sample Name	2015 dilution	Well
NTC	plate 01	A01
20130846_3_a.2	plate 01	B01
NTC	plate 01	C01
20135066_1_a.2	plate 01	D01
20134945_3_a.2	plate 01	E01
20130812_1_a.2	plate 01	F01
20134951_2_a.2	plate 01	G01
20134945_1_a.2	plate 01	H01
20130809_1_a.2	plate 01	A02
20140633_1_a.2	plate 01	B02
20130833_1_a.2	plate 01	C02
20140562_1_a.2	plate 01	D02
20134914_2_a.2	plate 01	E02
20134889_3_a.2	plate 01	F02
20134939_1_a.2	plate 01	G02
20130804_1_a.2	plate 01	H02
20130834_1_a.2	plate 01	A03
20130859_2_a.2	plate 01	B03
20130813_1_a.2	plate 01	C03
20135056_1_a.2	plate 01	D03
20134899_1_a.2	plate 01	E03
20134945_2_a.2	plate 01	F03
20130862_3_a.2	plate 01	G03
20134922_2_a.2	plate 01	H03
20130822_3_a.2	plate 01	A04
20130793_1_a.2	plate 01	B04
20130826_1_a.2	plate 01	C04
20134903_3_a.2	plate 01	D04
20134921_1_a.2	plate 01	E04
20134889_1_a.2	plate 01	F04
20130755_3_a.2	plate 01	G04
20134913_1_a.2	plate 01	H04
20130850_3_a.2	plate 01	A05
20130832_1_a.2	plate 01	B05
20130816_1_a.2	plate 01	C05
20134895_1_a.2	plate 01	D05
20134946_3_a.2	plate 01	E05
20134905_1_a.2	plate 01	F05
20134931_1_a.2	plate 01	G05
20134954_3_a.2	plate 01	H05
20130821_3_a.2	plate 01	A06
20130819_2_a.2	plate 01	B06
20130828_1_a.2	plate 01	C06
20134900_1_a.2	plate 01	D06
20134899_3_a.2	plate 01	E06
20134962_1_a.2	plate 01	F06

Sample Name	2015 dilution	Well
20130831_2_a.2	plate 01	A07
20130802_2_a.2	plate 01	B07
20130829_1_a.2	plate 01	C07
20134936_1_a.2	plate 01	D07
20130854_1_a.2	plate 01	E07
20134890_1_a.2	plate 01	F07
20134908_1_a.2	plate 01	G07
20130787_2_a.2	plate 01	H07
20130814_3_a.2	plate 01	A08
20130831_1_a.2	plate 01	B08
20130802_3_a.2	plate 01	C08
20130761_1_a.2	plate 01	D08
20130872_3_a.2	plate 01	E08
20134901_1_a.2	plate 01	F08
20134927_3_a.2	plate 01	G08
20130808_1_a.2	plate 01	H08
20130835_1_a.2	plate 01	A09
20130856_2_a.2	plate 01	B09
20140591_1_a.2	plate 01	C09
20130878_3_a.2	plate 01	D09
20134933_1_a.2	plate 01	E09
20134950_1_a.2	plate 01	F09
20134928_1_a.2	plate 01	G09
20134916_1_a.2	plate 01	H09
20130811_1_a.2	plate 01	A10
20130811_3_a.2	plate 01	B10
20135064_2_a.2	plate 01	C10
20130850_1_a.2	plate 01	D10
20130846_2_a.2	plate 01	E10
20130864_2_a.2	plate 01	F10
20130808_3_a.2	plate 01	G10
20134929_3_a.2	plate 01	H10
20130819_3_a.2	plate 01	A11
20130759_2_a.2	plate 01	B11
20135068_1_a.2	plate 01	C11
20130848_1_a.2	plate 01	D11
20134914_1_a.2	plate 01	E11
20134927_1_a.2	plate 01	F11
20130803_1_a.2	plate 01	G11
20134954_1_a.2	plate 01	H11
20130820_1_a.2	plate 01	A12
20134891_3_a.2	plate 01	B12
20140577_1_a.2	plate 01	C12
20134884_2_a.2	plate 01	D12
20130841_1_a.2	plate 01	E12
20134919_1_a.2	plate 01	F12

20134924_1_a.2	plate 01	G06
20130869_3_a.2	plate 01	H06

CEPH	plate 01	G12
CEPH	plate 01	H12

Appendix F: Sample of chip map

Plates 1 & 2 Ghana	1	2	3	4	5	6
A	NTC	20130834_1_a.2	20130809_1_a.2	20130822_3_a.2	20130850_3_a.2	20130821_3_a.2
A2	20134956_1_a.2	NTC	20134935_1_a.2	20130865_1_a.2	20130794_1_a.2	20130787_1_a.2
B	20130846_3_a.2	20140633_1_a.2	20130859_2_a.2	20130793_1_a.2	20130832_1_a.2	20130819_2_a.2
B2	NTC	20130795_1_a.2	20130786_3_a.2	20130772_1_a.2	20130753_1_a.2	20130778_2_a.2
C	NTC	20130833_1_a.2	20130813_1_a.2	20130826_1_a.2	20130816_1_a.2	20130828_1_a.2
C2	20130791_1_a.2	NTC	20130755_1_a.2	20130743_1_a.2	20130740_3_a.2	20130745_1_a.2
D	20135066_1_a.2	20140562_1_a.2	20135056_1_a.2	20134903_3_a.2	20134895_1_a.2	20134900_1_a.2
D2	20130782_1_a.2	20134956_2_a.2	20130877_2_a.2	20134937_3_a.2	20134947_3_a.2	20134951_1_a.2
E	20134945_3_a.2	20134914_2_a.2	20134899_1_a.2	20134921_1_a.2	20134946_3_a.2	20134899_3_a.2
E2	20130864_1_a.2	20134950_2_a.2	20130818_3_a.2	20134903_1_a.2	20130745_3_a.2	20130874_1_a.2
F	20130812_1_a.2	20134889_3_a.2	20134945_2_a.2	20134889_1_a.2	20134905_1_a.2	20134962_1_a.2
F2	20134909_3_a.2	20134920_3_a.2	20130759_3_a.2	20130764_1_a.2	20130869_1_a.2	20134911_2_a.2
G	20134951_2_a.2	20134939_1_a.2	20130862_3_a.2	20130755_3_a.2	20134931_1_a.2	20134924_1_a.2
G2	20134908_3_a.2	20134944_1_a.2	20134953_1_a.2	20134947_2_a.2	20134911_3_a.2	20134938_1_a.2
H	20134945_1_a.2	20130804_1_a.2	20134922_2_a.2	20134913_1_a.2	20134954_3_a.2	20130869_3_a.2
H2	20134947_1_a.2	20135011_1_a.2	20134997_1_a.2	20130769_2_a.2	20130794_3_a.2	20135033_1_a.2

7	8	9	10	11	12
20130831_2_a.2	20130814_3_a.2	20130835_1_a.2	20130811_1_a.2	20130819_3_a.2	20130820_1_a.2
20130756_1_a.2	20130772_3_a.2	20130779_3_a.2	20130785_1_a.2	20130769_1_a.2	20130784_1_a.2
20130802_2_a.2	20130831_1_a.2	20130856_2_a.2	20130811_3_a.2	20130759_2_a.2	20134891_3_a.2
20130778_1_a.2	20130750_3_a.2	20130789_3_a.2	20130780_1_a.2	20130782_3_a.2	20130739_1_a.2
20130829_1_a.2	20130802_3_a.2	20140591_1_a.2	20135064_2_a.2	20135068_1_a.2	20140577_1_a.2
20130766_1_a.2	20130741_1_a.2	20130788_1_a.2	20130797_1_a.2	20130768_1_a.2	20130790_3_a.2
20134936_1_a.2	20130761_1_a.2	20130878_3_a.2	20130850_1_a.2	20130848_1_a.2	20134884_2_a.2
20130822_1_a.2	20134888_1_a.2	20134885_3_a.2	20130872_1_a.2	20130871_3_a.2	20134914_3_a.2
20130854_1_a.2	20130872_3_a.2	20134933_1_a.2	20130846_2_a.2	20134914_1_a.2	20130841_1_a.2
20130870_3_a.2	20134908_2_a.2	20130871_2_a.2	20134922_1_a.2	20130749_3_a.2	20130870_2_a.2

Appendix

20134890_1_a.2	20134901_1_a.2	20134950_1_a.2	20130864_2_a.2	20134927_1_a.2	20134919_1_a.2
20134912_1_a.2	20130870_1_a.2	20134912_2_a.2	20130835_3_a.2	20134884_3_a.2	20134938_2_a.2
20134908_1_a.2	20134927_3_a.2	20134928_1_a.2	20130808_3_a.2	20130803_1_a.2	NA18856
20130810_2_a.2	20134931_3_a.2	20130795_2_a.2	20135008_1_a.2	20134884_1_a.2	NA18856
20130787_2_a.2	20130808_1_a.2	20134916_1_a.2	20134929_3_a.2	20134954_1_a.2	NA18855
20135002_1_a.2	20130807_3_a.2	20130768_2_a.2	20135032_1_a.2	20130752_1_a.2	NA18855

G: Assay plate and allele maps

Assays Set 1	1	2	3
A	rs34743335	rs138751793	rs7590268
B	rs642961	rs1801133	rs12532
C	rs861020	rs1801131	rs115200552
D	rs742071	rs11119388	rs651333
E	rs766325	rs6677101	rs2674394
F	rs560426	rs227782	rs987525
G	rs4147811	rs2303596	rs6558002
H	rs481931	rs4332945	rs894673

Assays Set 2	10	11	12
A	rs3758249	rs17563	rs8069536
B	rs7078160	rs1258763	rs227731
C	rs4752028	rs8049367	rs17760296
D	rs10785430	rs1546124	rs3923086
E	rs8001641	rs4783099	rs7224837
F	rs9574565	rs16260	rs13041247
G	rs185831554	rs11642413	rs11696257
H	rs375489721	rs8081823	rs17820943

VIC (X) VIC allele

	1	2	3
A	A	T	G
B	A	C	A
C	A	C	C
D	G	A	G
E	A	G	A
F	G	T	A
G	G	C	C
H	C	G	A

allele (X)

	10	11	12
A	G	T	G
B	A	C	A
C	C	C	G
D	A	C	T
E	A	C	A
F	C	A	C
G	T	A	C
H	T	A	C

FAM (Y) FAM allele

	1	2	3

allele (Y)

	10	11	12

A	T	C	T
B	G	T	G
C	G	A	G
D	T	G	A
E	G	T	C
F	A	C	C
G	A	T	T
H	A	T	T

A	A	C	T
B	G	T	C
C	T	T	T
D	G	G	G
E	G	T	G
F	T	C	T
G	G	G	T
H	C	G	T

Appendix

H: Selected steps in Fluidigm SNP Genotyping Protocol

a: Set of 48 assays (SNPs) that were combined to form the 0.2X PreAmp cocktail; they were also used to prepare the assay plates. b: PreAmp was carried out on this PCR machine. c: 192.24 Fluidigm Chip that was used to genotype the samples. d: After loading samples and assays, the chip was first run on the IFC controller (left) and was subsequently transferred to the FC1 Cycler (right). The chip was then read on a Chip Reader to generate the genotype files.

I: Sample of DNA sequencing plate map

Appendix

Well	Sample Name	Primer Name
A01	murray_CD4353_A01_ghan_20130827_1.b	i6e1m13r
B01	murray_CD4353_B01_ghan_20130827_2.b	i6e1m13r
C01	murray_CD4353_C01_ghan_20130827_3.b	i6e1m13r
D01	murray_CD4353_D01_ghan_20134936_1.b	i6e1m13r
E01	murray_CD4353_E01_ghan_20134936_8.b	i6e1m13r
F01	murray_CD4353_F01_ghan_20134936_9.b	i6e1m13r
G01	murray_CD4353_G01_ghan_20134926_1.b	i6e1m13r
H01	murray_CD4353_H01_ghan_20134926_3.b	i6e1m13r
A02	murray_CD4353_A02_ghan_20130797_1.b	i6e1m13r
B02	murray_CD4353_B02_ghan_20130797_3.b	i6e1m13r
C02	murray_CD4353_C02_ghan_20130797_8.b	i6e1m13r
D02	murray_CD4353_D02_ghan_20134953_1.b	i6e1m13r
E02	murray_CD4353_E02_ghan_20134953_3.b	i6e1m13r
F02	murray_CD4353_F02_ghan_20134904_1.b	i6e4m13r
G02	murray_CD4353_G02_ghan_20134904_2.b	i6e4m13r
H02	murray_CD4353_H02_ghan_20130877_1.b	i6e4m13r
A03	murray_CD4353_A03_ghan_20130877_2.b	i6e4m13r
B03	murray_CD4353_B03_ghan_20130877_3.b	i6e4m13r
C03	murray_CD4353_C03_ghan_20130792_1.b	i6e4m13r
D03	murray_CD4353_D03_ghan_20130792_2.b	i6e4m13r
E03	murray_CD4353_E03_ghan_20130792_3.b	i6e4m13r
F03	murray_CD4353_F03_ghan_20135024_1.b	i6e4m13r
G03	murray_CD4353_G03_ghan_20135024_3.b	i6e4m13r
H03	murray_CD4353_H03_ghan_20135021_1.b	i6e4m13r
A04	murray_CD4353_A04_ghan_20135021_2.b	i6e4m13r
B04	murray_CD4353_B04_ghan_20135021_3.b	i6e4m13r
C04	murray_CD4353_C04_ghan_20135021_1.b	i6e6m13r
D04	murray_CD4353_D04_ghan_20135021_2.b	i6e6m13r
E04	murray_CD4353_E04_ghan_20135021_3.b	i6e6m13r
F04	murray_CD4353_F04_ghan_20134923_1.b	i6e7m13r
G04	murray_CD4353_G04_ghan_20134923_2.b	i6e7m13r
H04	murray_CD4353_H04_ghan_20134962_1.b	i6e7m13r

Appendix

A05	murray_CD4353_A05_ghan_20134962_3.b	i6e7m13r
B05	murray_CD4353_B05_ghan_20130759_1.b	bvex4rvs
C05	murray_CD4353_C05_ghan_20130759_2.b	bvex4rvs
D05	murray_CD4353_D05_ghan_20130759_3.b	bvex4rvs

J: Gradient PCR form

Appendix

GRADIENT PCR

Date: 3/20/15 Sample: CEPH1 Gene: MAPB

Master Mix	1x	(10x)	15x
2x Biolase Mix	5.0µl	50µl	75µl
F Primer (20uM)	0.15µl	1.5µl	2.25µl
R Primer (20 uM)	0.15µl	1.5µl	2.25µl
DMSO (5%)	0.5µl	5.0µl	7.5µl
ddH_20	3.2µl	32µl	48µl
DNA	1.0µl		

C1000 Touch™ Thermal Cycler

PCR Conditions		PCR Machine used: MJ
94°C	5 min.	
94°C	45 sec.	
58.2 to 63 °C	45 sec.	Number of cycles 35
72°C	60 sec.	
72°C	7 min.	
4°C	hold	

Dispense 9µl of master mix into each well (A1-H1 or A2-H2) containing 1µl of DNA.

Primers (Columns):

1- 9xex1fw1 & rv1 } No DMSO
2- 9xex2fw2 & rv2 }
3- 9xex2fw3 & rv3 }
4- 9xex1fw1 & rv1 } 1x DMSO
5- 9xex2fw2 & rv2 }
6- 9xex2fw3 & rv3 }
7- 9xex1fw1 & rv1 } 2x DMSO
8- 9xex2fw2 & rv2 }
9- 9xex2fw3 & rv3 }
10- _____
11- _____
12- _____

Gradient Temperatures (rows):

A 63
B 62.5
C 61.4
D 59.4
E 57.1
F 55.2
G 53.9
H 53.2

After the PCR is done, check the product on a gel.

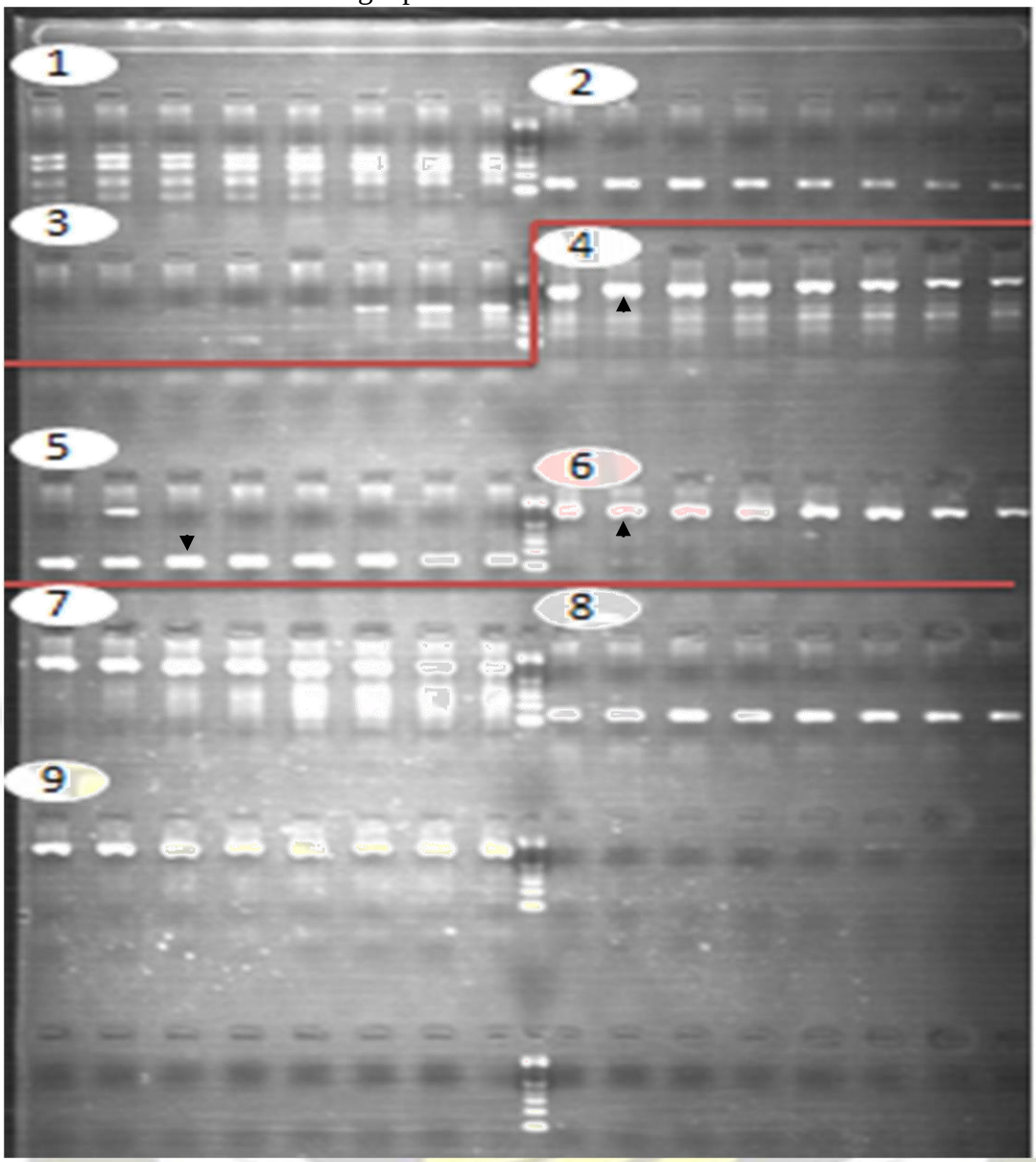
Add 5 ul loading dye, mix, and load 10 ul into the wells of the gel.

Run on a 2% agarose gel for 15 min. @ 200V. If the bands haven't migrated far enough you can let it run longer.

Save your picture and print it.

Appendix

K: Gradient PCR gel picture



This is a gradient PCR gel picture for MAFB. Numbers on picture corresponds to the numbers on the section captioned “Primers (columns)” in Appendix J. 1, 2 and 3 indicate when no DMSO was added to the gradient PCR reaction for the three primer sets, leading to poor DNA bands for all the three primer sets. In 4, 5 and 6, 5% DMSO was added to the reaction and the DNA bands were very clear. In 4, the primer worked for almost all the temperatures indicated under “temperature (rows)” in Appendix J. However, 62.5 ^{o}C, indicated by the black arrowhead, had the best band and was chosen as the optimal annealing temperature for that primer set. Therefore, the Initial PCR for that primer set was run with 5% DMSO at 62.5^{o}C. In 5 and 6 too, 61.4 and 62.5, respectively, were the temperatures that produced the best DNA bands and were chosen for the subsequent Initial PCR with 5% DMSO; these temperatures are indicated by black arrowheads. At 10%

DMSO (7, 8 and 9), the DNA band quality begun to deteriorate and therefore this treatment was not optimal for these primer sets.

L: Initial PCR form

Appendix

INITIAL PCR

Date: 4/18/15 Plate (samples): Syndromics – Africa

Gene: IRF6 Exon: 2 primers 16e2rn13f2r

1x Dmso

Master Mix:	1x	100x
Biolase mix	5.0µl	500µl
F primer _____	0.15µl	15µl
R primer _____	0.15µl	15µl
DMSO5%	0.5µl	50µl
ddH_2O	3.2µl	320µl
DNA	1.0µl	

Dispense 9µl of master mix into each well containing 1µl of DNA.

PCR Conditions:

94°C - 5 min.
94°C - 45 sec.
57°C - 45 sec.
72°C - 45 sec.
72°C - 7 min.
4°C - hold

PCR Machine used: GeneAmp® PCR System 9700

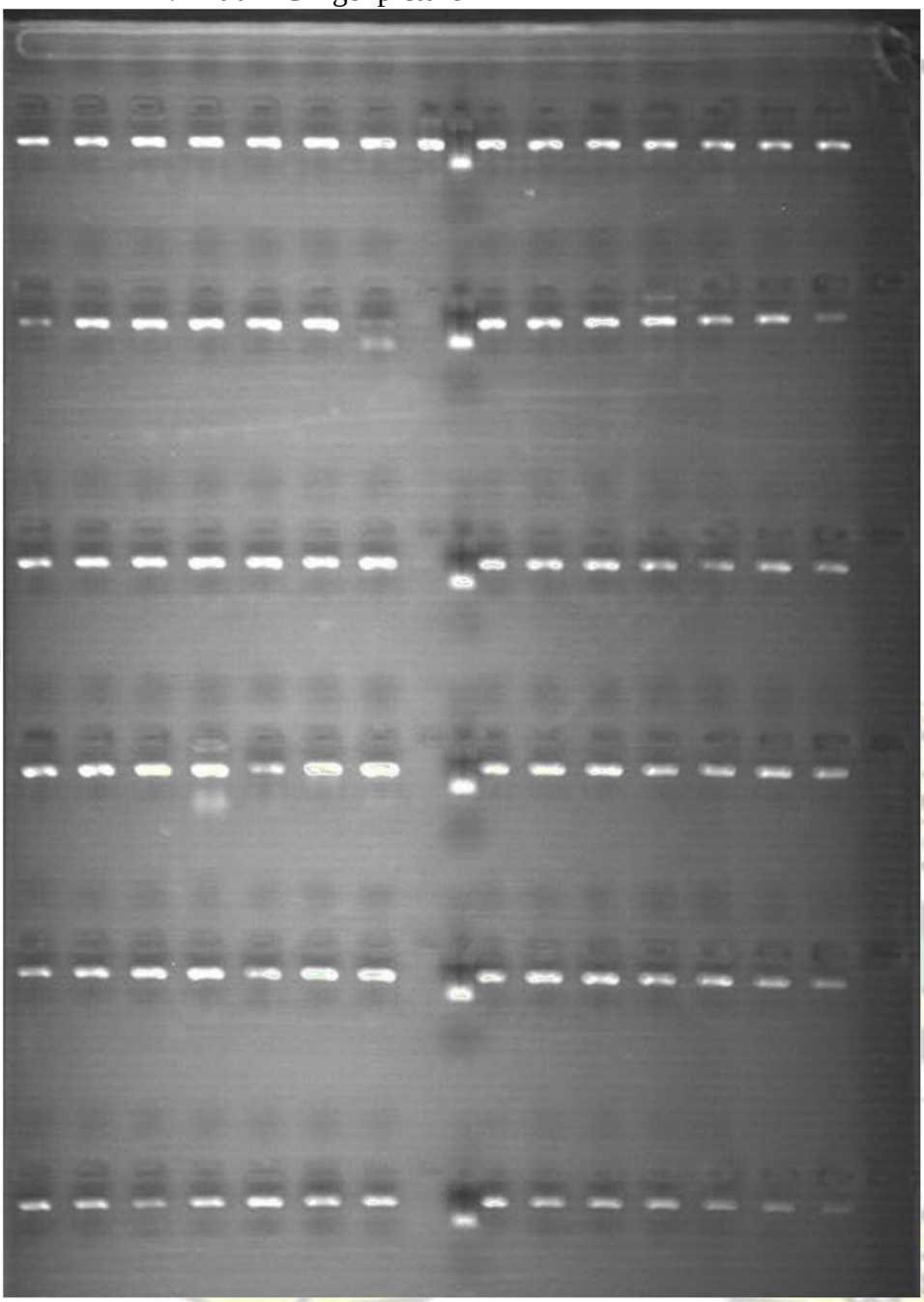
Number of cycles 35

Agarose gel: 4/18/15

Dispense 2µl of loading dye dilution into each well and add 4µl of DNA.
Run on a 2% agarose gel for 15 min. @ 200V. If the bands haven't migrated far enough you can let it run longer.
Save the picture and print it.

Appendix

M: Initial PCR gel picture



This is an Initial PCR gel picture for IRF6 exon 2 primer sets. The optimal temperature and DMSO treatment for this primer set are shown in Appendix L. Here, each band represents an individual's DNA. After each Initial PCR, a gel of this type was run on 4 uL out of 10 uL reaction volume. If the DNA bands looked very good like this, the remaining Initial PCR product was sent to Functional Biosciences, Wisconsin, for direct DNA sequencing. However, if the Initial PCR reaction did not successfully amplify the exon of

interest, no or very poor DNA bands were observed on the gel picture. In such instances, the Initial PCR product was discarded and a new Initial PCR was re-run, followed by re-running of gel electrophoreses, before the PCR product was sent for sequencing.

Appendix N: Deep phenotyping of clefts in the study population

Classification of Clefts Based on laterality and severity

	Nonsyndromic Clefts			Syndromic Clefts			
	Male	Female	Total A	Male	Female	Total B	Total A+B
Cleft Lip only (CL)							
Right Complete Unilateral CL	15	19	34	3	0	3	37
Left Complete Unilateral CL	35	33	68	2	0	2	70
Complete Bilateral CL	6	7	13	1	0	1	14
Right Incomplete Unilateral CL	5	4	9	2	0	2	11
Left Incomplete Unilateral CL	11	11	22	0	1	1	23
Incomplete Bilateral CL	1	1	2	0	0	0	2
Midline CL	1	1	2	0	1	1	3
Macrostomia	5	4	9	0	0	0	9
Totals	79	80	159	8	2	10	169
Cleft Palate (CP)							
Complete CP	9	23	32	4	7	11	43
Incomplete CP	13	37	50	3	6	9	59
Submucous CP (CP + Cleft Uvula)	12	8	20	1	2	3	23
Complete CP + Bifid Uvula	1	0	1	0	0	0	1
Totals	35	68	103	8	15	23	126
Cleft Lip and Palate (CLP)							
Right Complete Unilateral CLP	19	12	31	2	0	2	33
Left Complete Unilateral CLP	28	26	54	0	1	1	55
Left Complete Unilateral CLP + Right Complete CP	2	0	2	0	0	0	2
Complete Bilateral CLP	9	20	29	6	6	12	41
Right Incomplete Unilateral CLP	0	3	3	0	0	0	3
Left Incomplete Unilateral CLP	2	2	4	0	0	0	4
Incomplete Bbilateral CLP	0	5	5	0	0	0	5
Right Complete Unilateral CLP + Cleft Uvula	3	2	5	0	0	0	5
Left Complete Unilateral CLP + Cleft Uvula	1	2	3	0	0	0	3

Complete BilateralL CLP + Cleft Uvula	3	2	5	0	0	0	5
Incomplete Bilateral CLP + Cleft Uvula	0	1	1	0	0	0	1

Appendix N continued…

	Nonsyndromic Clefts			Syndromic Clefts			
	Male	Female	Total A	Male	Female	Total B	Total A+B
Midline CLP	1	0	1	0	0	0	1
Totals	68	75	143	8	7	15	158
Special Clefts							
Right Facial Cleft + CL	0	1	1	0	1	1	2
Left Facial Cleft + CL	1	0	1	0	0	0	1
Left Facial Cleft + CLP	1	0	1	1	0	1	2
Bilateral Facial Cleft	1	0	1	0	0	0	1
Akyloglossia (Tongue-tie)	1	0	1	0	1	1	2
Midline Lower Lip CL	0	1	1	0	0	0	1
Midline Microform CL	0	1	1	0	0	0	1
BilateralL Microform CL	0	0	0	0	1	1	1
Left Microform CL + Right Complete CL	1	0	1	0	0	0	1
Left Microform CL + Alveolar Cleft+ Ccomplete CP	0	0	0	1	0	1	1
Left Incomplete CL + Left Distal Microform CL	0	1	1	0	0	0	1
Totals	5	4	9	2	3	5	14

NB: Complete CL: Clefts of the Lip and Alveolus (tooth socket), Incomplete CL: Cleft of the Lip only, Macrostomia: unusually large mouth resulting from incomplete fusion of the maxilla and mandible at the point of the mouth, Complete CP: Clefts of the Soft and Hard Palates, Incomplete CP: Cleft of the Soft Palate only, Complete CLP: Clefts of the Lip, Alveolus, Hard Palate and Soft Palate, Incomplete CLP: Clefts of the Lip, Alveolus and Soft Palate only.

Appendix O: Chromatograms for observed aetiologic variants

NB: In all these chromatograms, green, red, blue and orange represent A, T, C and G nucleotides, respectively. The mutated residue is indicated by a white vertical line.

Plate

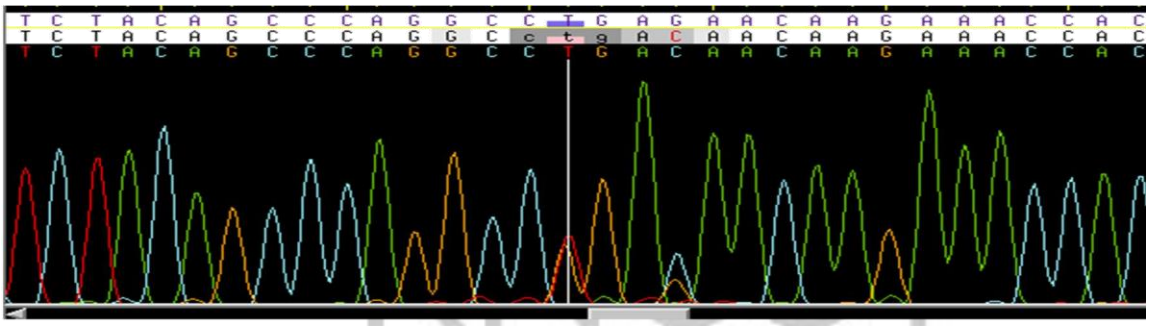


Plate Q1: Splice acceptor site variant (c.175-2A>C) in IRF6 (T>G)

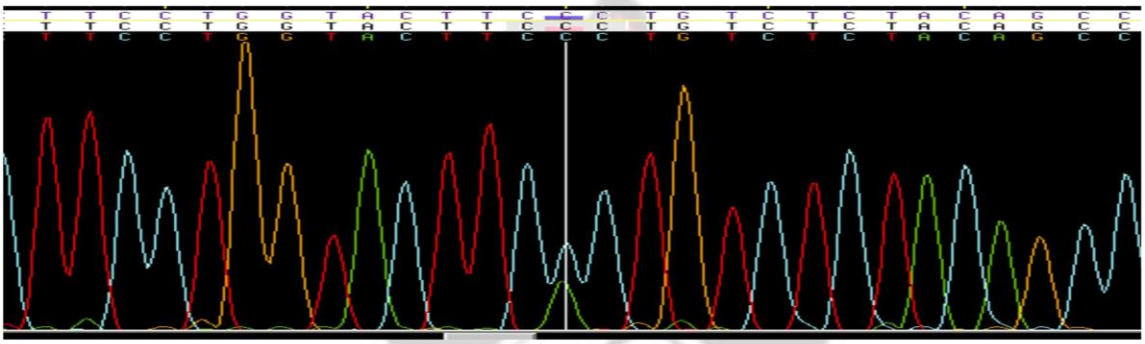


Plate Q2: p.Gly65Val in IRF6 (C>A)

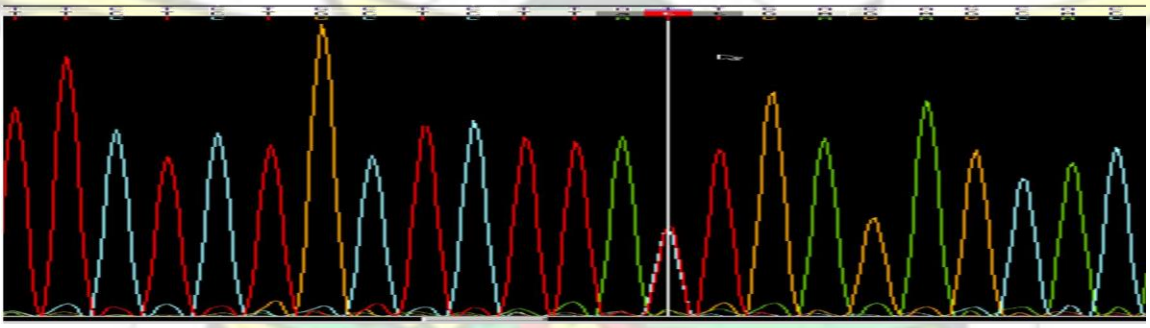
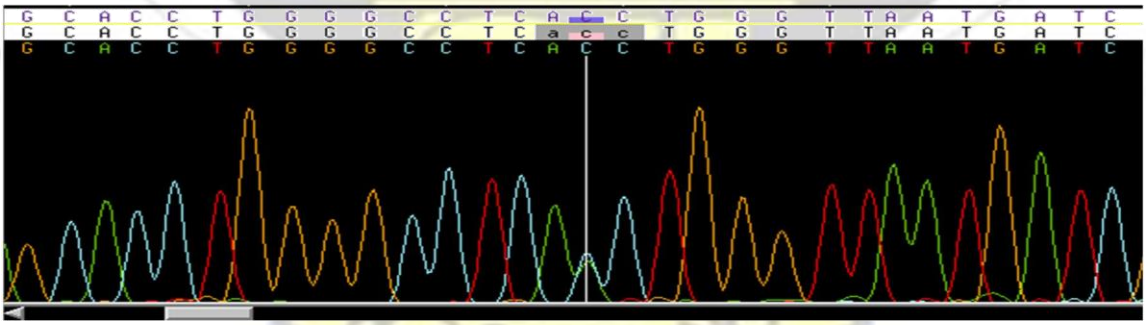


Plate Q3: p.Asn88Ser in IRF6 (T>C)



Q4: Splice donor site variant (c.379+1 G>T) in IRF6 (C>A)

Plate

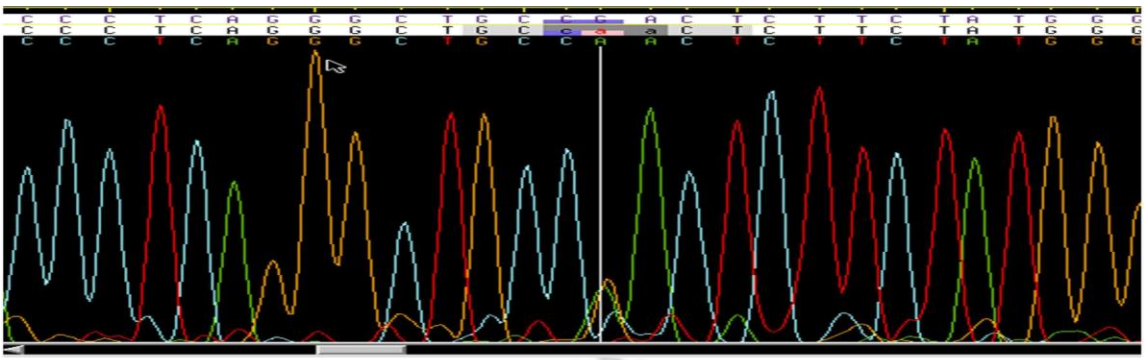


Plate Q5: p.Agr250Gln in IRF6 (G>A)

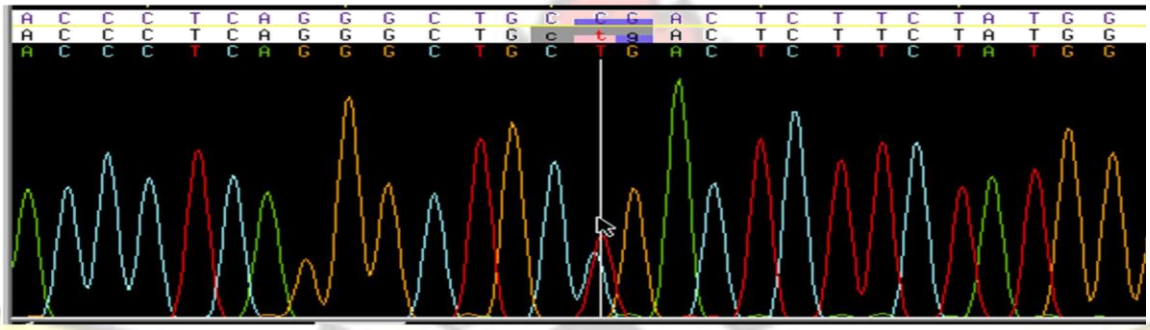


Plate Q6: p.Arg250X in IRF6 (C>T).

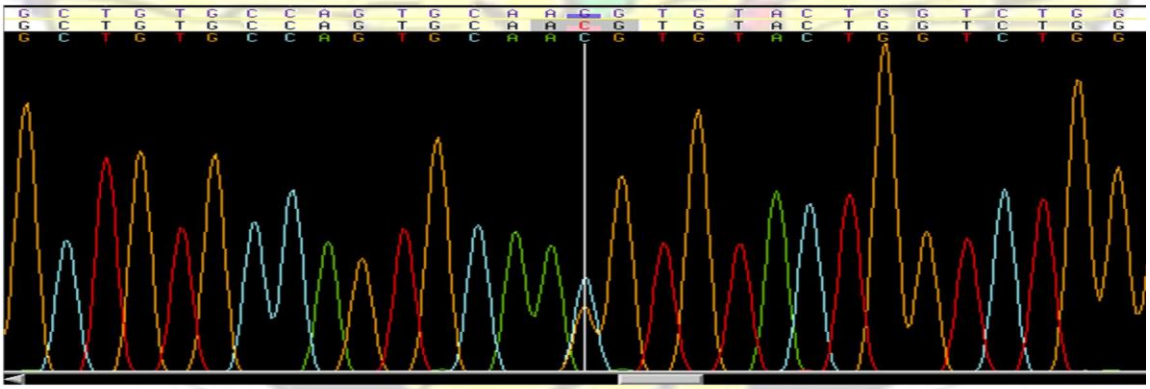
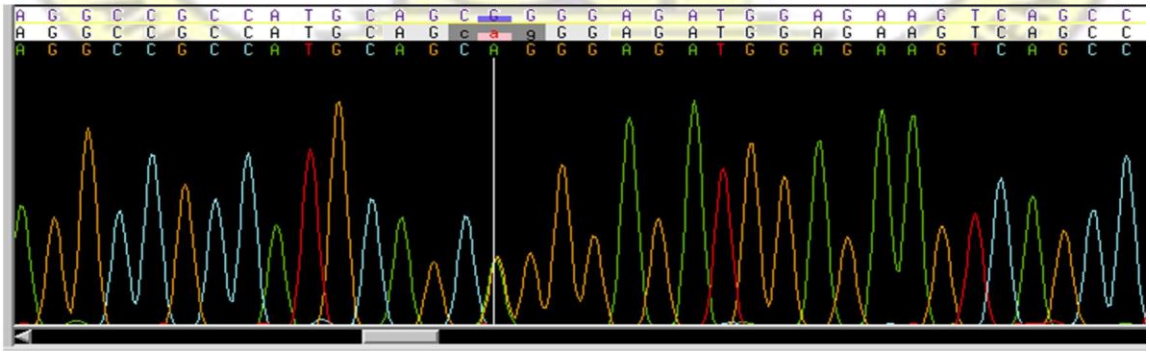


Plate Q7: p.Lys320Asn in IRF6 (G>C)



Plate

Q8: p.Pro408Leu in PAX7 (G>A)

Plate

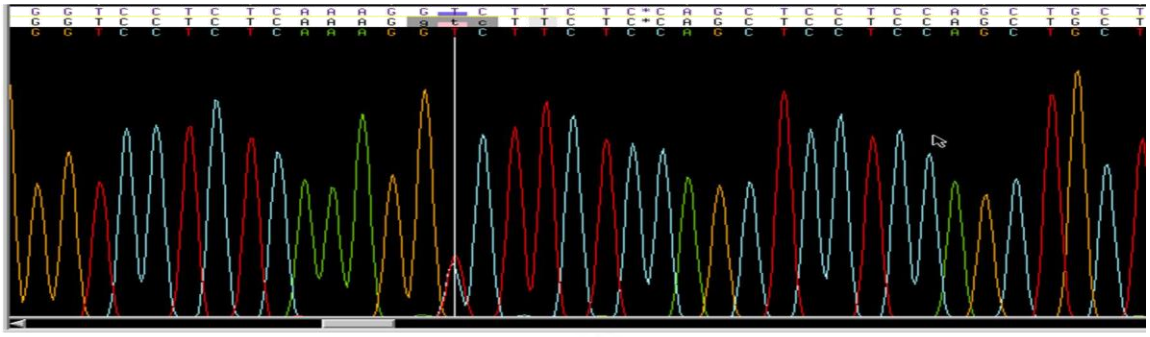


Plate Q9: p.Ala235Thr in PAX7(C>T)

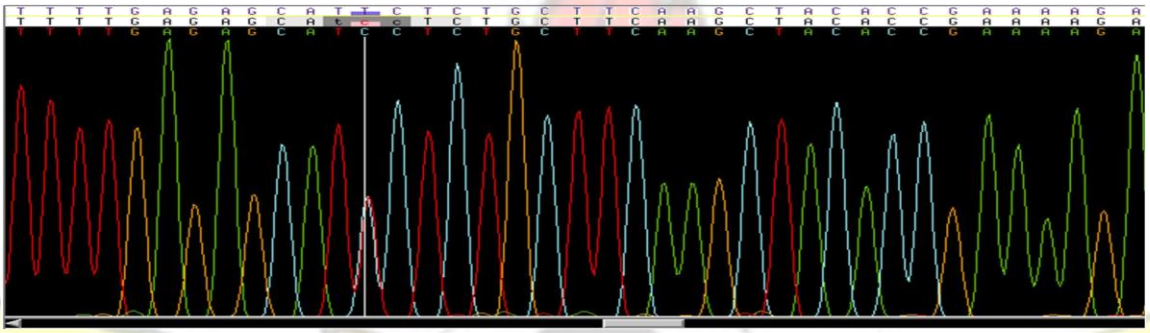


Plate Q10: p.Asn323Asp in ARHGAP29(T>C)

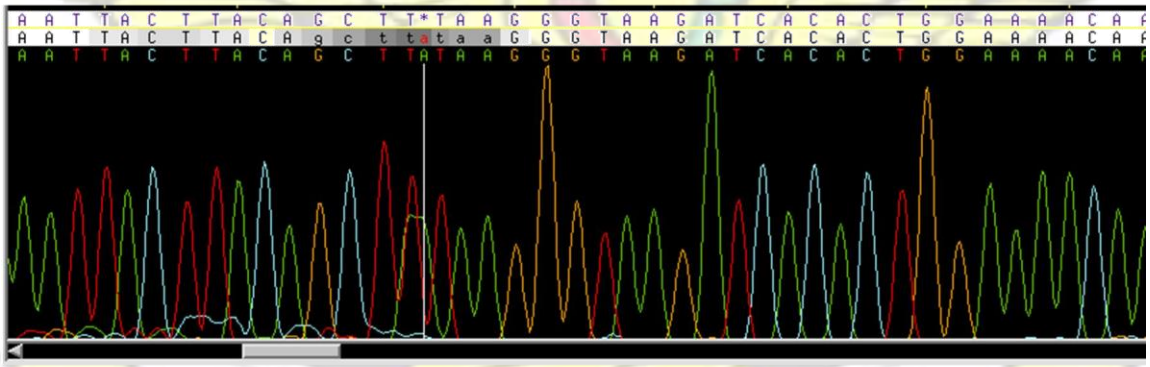
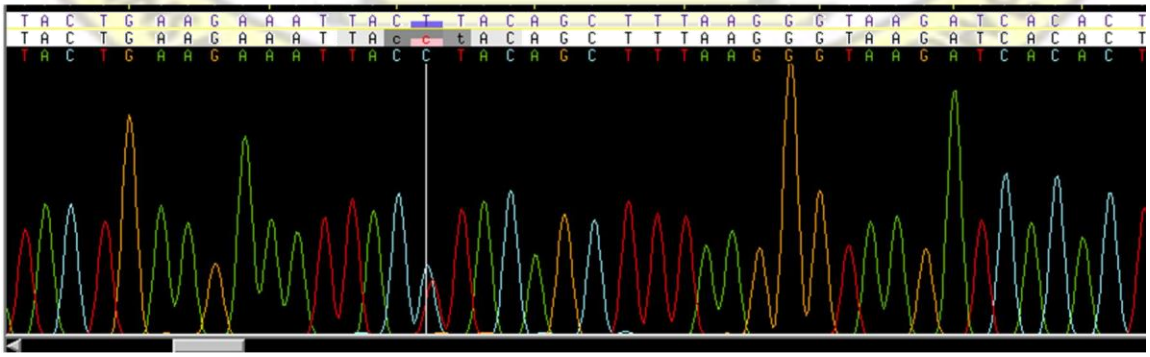


Plate Q11: p.Lys426IlefsTer6 in ARHGAP29(T>TA)



Plate

Q12: Splice donor site variant (c.1281+4A>G) in ARHGAP29 (T>C)

Plate

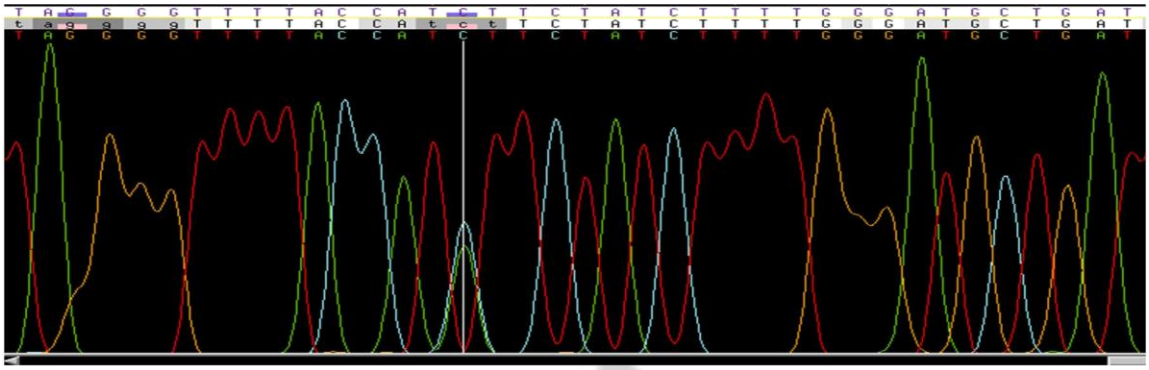


Plate Q13: p.Asp988Tyr in ARHGAP29(C>A)

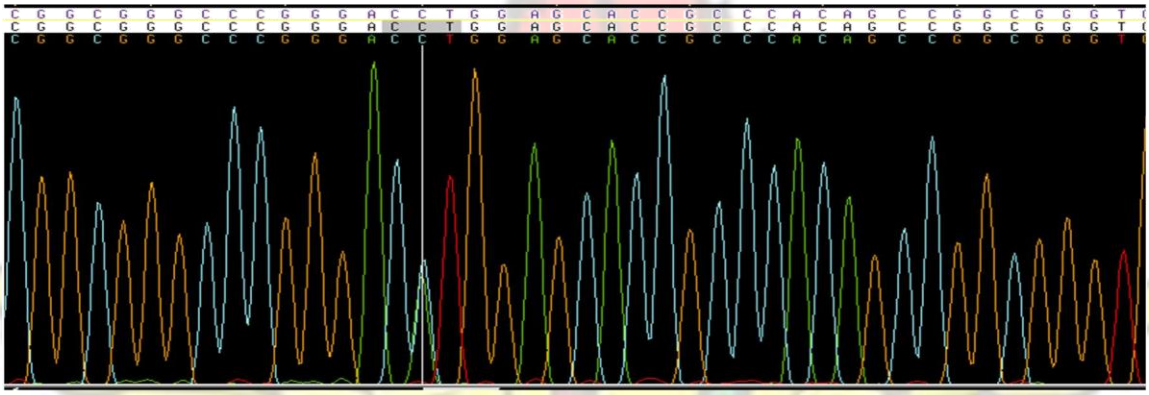


Plate Q14: p.Gly252Cys in VAX1(C>A)

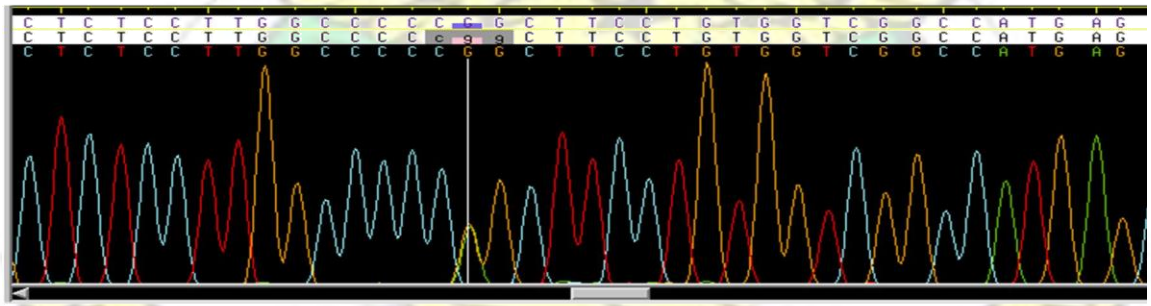
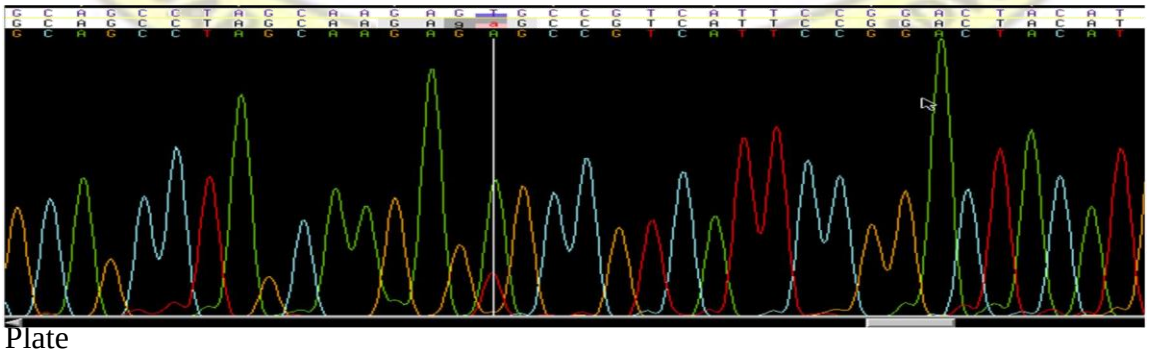


Plate Q15: p.Pro73Leu in MSX1(G>A)



Plate

Q16: p.Ser76Arg in BMP4 (T>A)

Plate

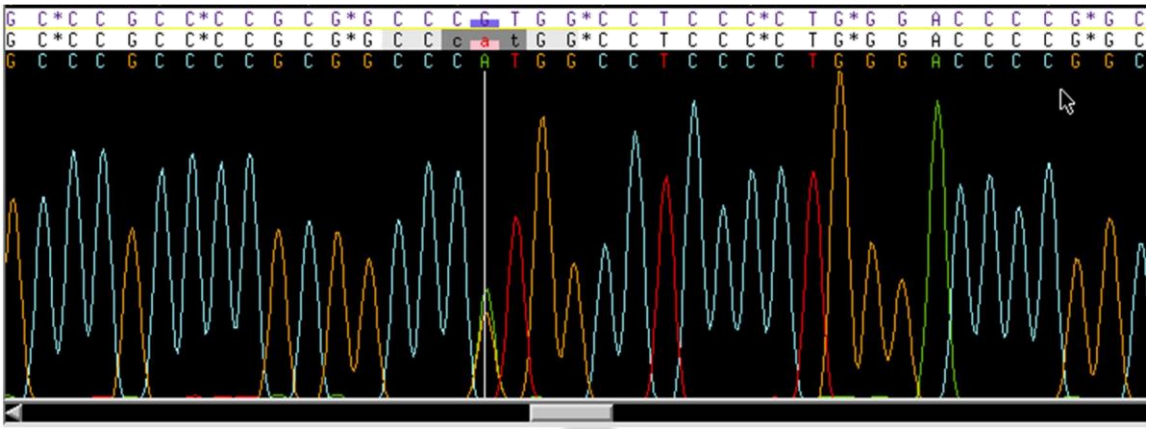


Plate Q17: p.Thr36Met in FOXE1 (G>A)

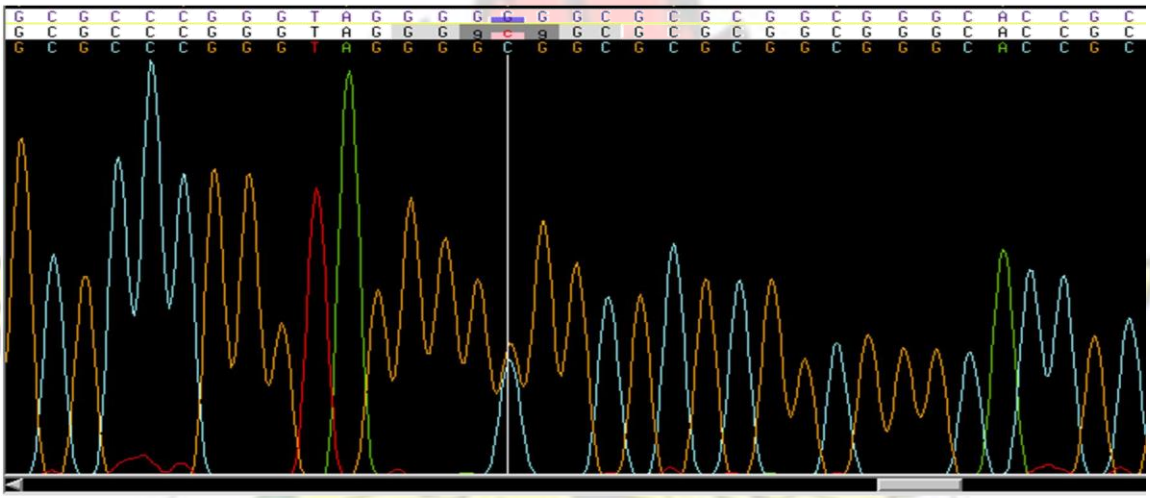


Plate Q18: p.Pro190Arg in FOXE1 (G>C)

Appendix P: Linkage disequilibrium (LD) block for 8q24 (www.1000genomes.org)

Linkage disequilibrium (LD) blocks for rs987525 of 8q24 in Africans and Europeans. The black arrowhead indicates the position of rs987525 in these LD blocks. a: LD block for rs987525 of 8q24 for Yoruba (Africans) – this shows that there are as many as 111 SNPs within this 20kb region, suggesting the haplotype block for Africans in this region is smaller, i.e. an average 180bp between SNPs. Moreover, within this window, rs987525 is in perfect LD with rs1119880 (purple arrowhead) and rs1579748 (green arrowhead) - which are dinucleotide SNPs- in Africans b: LD block for rs987525 of 8q24 for British in England and Scotland (Europeans) – this demonstrate that there are 64 SNPs within this 20kb window, presupposing the haplotype block for Europeans in this region is larger, i.e. an average of 315bp between SNPs. Among Europeans, rs987525 is in perfect LD with rs17241908 (green arrowhead), rs55865336 (blue arrowhead) and rs72728736 (purple arrowhead).

Appendix Q: Haploreg predictions for genotpted SNPs that have functional effects

SNP	Gene	Promoter histone marks	Enhancer histone marks	DNAse activity	Proteins bound	Motifs changed
rs34743335	IRF6	Observed in 20 tissues	Observed in 3 tissues	Observed in 43 tissues	POL2, ZEB1, JUND, TBP, TCF12, POL24H8, CCNT2, CMYC, ELF1, ETS1, GABP, HMGN3, IRF1, MAX, NFYB, SP1, TAF1, ZNF263	BCL, Egr-1, Ets, Irf, Klf7, Myc, Pou2f2, SP1, TATA
rs115200552	MSX1	Observed in 3 tissues	Observed in 4 tissues	Observed in 2 tissues	POL2	Irf, PRDM1, RXRA
rs4920523	PAX7	Observed in 1 tissues	Observed in 3 tissues	Observed in 2 tissues	SUZ12	Pax-2, STAT, p300
rs560426	ARHGAP29	None	Observed in foetal thymus tissue	None	None	Nkx2

Source: http://archive.broadinstitute.org/mammals/haploreg/haploreg.php