CHAPTER - 1

1.1: INTRODUCTION

Orofacial clefts (OFCs) are most challenging congenital malformations with variable phenotype that happen during early pregnancy, affecting about one of every 700 live births worldwide. Prevalence varies by ethnic group, geographic region and sex. Cleft lip and palate (CLP) is more common in boys and cleft palate only (CPO) is common among girls. North American Indians and Asians have the highest prevalence rates (1/500), while Caucasians have intermediate rates (1/1000) and Africans have the lowest prevalence rates (1/2500). Influence of seasonal variations in the incidence of OFCs was also reported[1]. Babies born with oral clefts have a wide range of problems, including but not limited to feeding and speaking difficulties, hearing, dental and impaired cognitive development. These problems can have long lasting consequences on health and social integration.

In general, facial cleft results when medial, lateral, and maxillary nasal processes on either left, right or both sides of the forming craniofacial complex do not fuse completely. Babies may have both cleft lip and palate, or either of them. Broadly speaking, approximately 30% of OFCs are syndromic and are associated with some recognizable anomalies, but 70% are of the non-syndromic and occur as an isolated condition[2,3]. More than 400 syndromes have been already associated with OFC[4]. The study of congenital cleft lip and cleft palate anomalies has been the subject of debate regarding the etiology and mode of transmission[5]. The
etiological factors include heredity, consanguinity, maternal environment, demographic factors and other factors like intra-uterine posture, drugs, vitamins, alcohol consumption, smoking, infections, diet etc[6]. The patients with cleft lip and palate present complex biologic, sociologic and psychological problems and their rehabilitation involves several disciplines. Both genetic and environmental influences are believed to be the cause of cleft lip and palate. Dermatoglyphic features, dental anomalies and handedness are the well-studied sub-phenotypes for the non-syndromic cleft lip and palate. A retrospective study from Canada has revealed that the left-sided clefts are more than the right-sided clefts[7]. The proportion of left-sided clefts among left-handers was higher than the right-handers. Furthermore, non-right-handed patients show a greater tendency of having a cleft than the right-handed patients, but the specific nature of this relationship remains unclear[7]. The correlation between dermatoglyphic patterns and non-syndromic cleft lip and palate has been extensively studied[8-11]. In patients with OFCs several dental traits such as hypodontia, supernumerary teeth, dental asymmetry and delay in tooth development on the side of the cleft were reported[12].

1.1.1: History of cleft lip and palate

The knowledge of cleft lip and the surgical correction was published by the Pierre Franco 19th century “Traité and Traité des Hernies” in which he described the condition as “lievréfendu de nativité” (cleft lip present from birth). In 390 BC the first cleft surgery was documented in China in an 18 years old would be solider, Wey Young-Chi. Germanicus Mirault can be credited to be the originator
of the triangular flap which was later modified by C.W. Tennison in 1952 and Peter Randall in 1959 [13].

1.1.2: Overview of normal embryonic craniofacial development

Craniofacial development is tightly regulated by the complex orchestration of genetic and environmental factors. The factors that influence are the interaction of cell growth, growth factors and receptors, the convergence and fusion of the facial and palatal processes, apoptosis, and adequate nutrient supply. These events are balanced by a complex system that ensures normal development [14, 15].

1.1.3: Formation of primary palate

The primary palate forms the basis of the upper lip and the anterior portion of the palate and is fundamental to mid-face development. On the 14th day of postconception, the formation of the primary palate begins with the orofacial development in the embryo. Due to the cellular interaction of ectodermal-mesodermal cells, ectomesenchymal cells arise and migrate into the subjacent tissue of the designated neural crest cells, which are positioned at the tip of the neural folds prior to the neural tube formation. These ectomesenchymal cells form five facial primordia, which are the frontonasal prominence, the paired maxillary processes, and the paired mandibular processes. At the 24th day post-conception, the frontonasal prominences are bounded by the maxillary prominences and appear distinctive. 28 days after the conception, the surface ectodermal cells thicken, and develop into the olfactory region of the nasal cavity, and form the nasal placodes. These nasal placodes later turn and form inverted horseshoe-shaped ridges on both
Introduction

Role of Transcription Factor Gene Polymorphisms in the Pathogenesis of Non-Syndromic Cleft Lip with or without Cleft Palate - A Case Control Study

sides of the frontonasal prominence. These ridges bulge up, form the medial and lateral nasal prominences, which causes the nasal placodes to sink and form the nasal pits. Between 40th and 48th day post-conception, the maxillary prominences grow medially, push the medial nasal prominences toward the midline, merge together and form the upper lip and the primary palate[16]. The fusion of medial nasal prominences and the maxillary prominences requires critically precise timed correlation in the growth and locations of these two prominences and the degeneration of the frontonasal prominence. Failures in any of the above requirements will result in cleft lip with different levels of severity. In the mildest cases, the clefts are only limited to the lip's vermillion border. In the more severe cases, the clefts develop through the tissue of the lip (unilateral or bilateral cleft lip). In the most severe cases, the clefts also involve the side of the nose (oblique clefts)[16].

1.1.4: Development of palate

The secondary palate includes the anterior and posterior palate and is part of the floor of the nasal cavities and the roof of the mouth. 6 weeks after the conception, the formation of the secondary palate begins and this is made up of three elements: the two secondary palatal shelves and the primary palate. Initially, these three elements are widely separated. During the eighth week of post-conception the two palatal shelves elevate into a horizontal position and grow medially towards each other. Simultaneously, the palatal shelves epithelium is transformed into its different phenotypes which are nasal, medial, and oral, and into mesenchyme. Among these, only the medial epithelium can undergo
cytodifferentiation into peridermal and basal cells. Later these peridermal cells then undergo apoptosis, while basal cells remain healthy. The merging of the shelves begins during the ninth week and is completed by the 12th week post-conception. During this period the medial epithelium undergoes epithelial-mesenchymal transformation, which is essential to mesenchymal merging of the shelves. The edges of the palatal shelves adopt a fibroblastic form when the cells are merged, and this leads to the fusion of the palatal shelves upon contact. The palatal shelves will also fuse anteriorly with the nasal septum in the hard palate region, and subsequently merge with the soft palate region, forming palatal closure [16]. Delay in elevation of the secondary palatal shelves from the vertical to the horizontal position, results in widening the gap between shelves and prevents them from contacting each other and subsequently, prevents the shelves fusing with each other. When the shelves become horizontal, the palatal closure cannot be completed, which results in a cleft palate. Defective shelf fusion, failure of medial edge epithelial apoptosis or epithelium-mesenchymal transformation, and post-fusion ruptures can also lead to a cleft palate[16,17].
Figure 1.1: Schematic diagrams outline the development of the lip and palate in human.

In the fourth week of embryonic development, developing frontonasal prominence, paired maxillary processes, and paired mandibular processes surround the primitive oral cavity (Figure 1.1a). At the fifth week; the nasal pits have formed, which leads to the formation of the paired medial and lateral nasal processes (Figure 1.1b). The medial nasal processes have merged with the maxillary processes to form the upper lip and primary palate by the end of the sixth week (Figure 1.1c). During the seventh week the bilateral outgrowths from the maxillary processes, grow vertically down the side of the tongue (Figure 1.1d). During the eighth week of post-conception, the two palatal shelves elevate into a horizontal position and grow medially towards each other (Figure 1.1e). At the ninth week the palatal shelves fusion begins anteriorly with the nasal septum in the hard palate region, and subsequently merge with the soft palate region, forming palatal closure, and this would get completed by the 12th week of post-conception which ultimately divides the oronasal space into separate oral and nasal cavities (Figure 1.1f)[18].
Introduction

Development of craniofacial region is regulated by many genes, and also by the growth factors (FGFs, TGFs, PDGFs, EGFs, BMPs and respective receptors), signalling molecules (WNT family, SHH and respective receptors) and transcription factors (MSX, DLX, LHX, PRRX and BARX family genes and respective receptors[19]. Interruption of this tightly controlled cascade can result in facial clefts, where the craniofacial complex ultimately fails to meet and fuse or form the appropriate structures.

1.1.5: Classification of cleft lip and palate

There are different ways to classify OFCs. Based on the anatomy, OFCs can be divided three groups. One is cleft lip with or without cleft palate (CLP), second one is cleft palate only (CPO) and third one is cleft lip only (CL). CLP describes OFCs with the primary palate involving of lip, with or without a cleft of the secondary palate and this is either complete or incomplete, and unilateral or bilateral. CPO describes OFCs with the secondary palate only which is either complete or incomplete. CL describes OFCs with the lip only which is either complete or incomplete. In order to standardize reporting of cleft lip and palate cases, in 1971 Kernahan introduced a simplified, symbolic classification system, which is the "striped Y" method. This system has become quite popular and gives information about the degree of cleft figure 1.2[20].
Figure 1.2: Schematic diagrams of Kernahan classification of cleft lip and palate

Here Figure A- is incomplete cleft lip, Figure B-complete cleft lip, Figure C-complete cleft palate, Figure D-complete cleft palate, Figure E-unilateral cleft lip and palate and Figure F-bilateral cleft lip and palate [20].
1.2: REVIEW OF LITERATURE

Gene identification for non-syndromic cleft lip and palate is difficult because of the varied levels of its penetrance, sex differences and other environmental factors[21]. Numerous environmental risk factors are strongly associated with increased prevalence of orofacial clefts, which several epidemiologic studies prove. Though a large number of genes have been speculated to be associated with the OFCs, no susceptible gene has been identified to be the reason consistently. Furthermore, the complex natures of the non-syndromic clefts are attributed to the multiple interacting genes, which confer moderate effects and have been proposed to provide susceptibility to OFCs[22-24]. However, the genetic component of cleft which is responsible for the susceptibility to respond to the environment is still not clearly evident.

With the advent of the genomics era, there have been major advances in identifying the causative genetic mutation and the underlying syndromic forms of OFCs (http://www.ncbi.nlm.nih.gov/omim). On the contrary, owing to its genetic heterogeneity, departure from Mendelian inheritance patterns, the lack of genomic tools and the necessity for very large datasets, there has been less progress in advancing our understanding of the genetic etiology of non-syndromic OFCs. However, the recent development of innovative approaches to phenotyping and powerful genomic tools, together with extrapolation from studies of syndromic forms of OFCs, has increased our understanding of non-syndromic OFCs.
Accurate phenotyping is crucial for understanding the etiology of any congenital malformation, because when the heterogeneous groups are treated as a single entity power to detect the reason, the effects are weakened. Although OFCs show a range of phenotypic expression, they are generally defined as qualitative traits (i.e. Affected or unaffected). Dividing CLP in this simplistic way can lead to loss of important potential information. Furthermore, numerous lines of evidence now suggest that the spectrum is more complex and should include a variety of sub-clinical phenotypic features observed, in either an individual with CLP and/or their “unaffected” relatives.

1.2.1: Insights from syndromic clefts:

Cleft lip or palate can be seen as an associated feature in approximately 300 syndromes. Studies on these syndromes had given sufficient clues to identify the genes that cause the non syndromic clefts[25]. Vander Woude syndrome (VWS) (OMIM 119300) is the most common autosomal dominant disorder which is always associated with CLP[26]. Interferon regulatory factor 6 (IRF6) which is located in the chromosome 1 is the major gene that is associated with VWS,[27,28] and, later its variants were found to be significantly associated with non-syndromic clefts[29,30]. The Wolf-Hirschhorn syndrome (WHS)(OMIM 194190), which is caused by the partial deletion of the short arm of Chromosome 4 (4p), is also characterized by oral clefts[31]. Fluorescence in situ hybridization (FISH) analysis of eight Finnish WHS patients revealed that five patients did not have one copy of MSXI and the other in three where one individual possessing cleft palate has shown the presence of two hybridization signals[32]. Analysis of Dutch family...
with CLP, CPO and selective tooth agenesis revealed a heterozygous nonsense mutation in *MSX1* gene. Furthermore, complete sequencing of the *MSX1* gene in 1000 unrelated CLP individuals showed that mutations in *MSX1* alone could account for 2% of isolated CLP[33]. Kallmann syndrome (KAL2) (OMIM147950) is also inherited in an autosomal dominant fashion with orofacial clefts in 5-10% of the cases. Approximately in 20% of Kallmann syndrome cases, a loss-of-function mutation in *FGFR1* has been reported[34]. Recent analysis has shown 3 to 5% of non-syndromic cleft lip or palate (NSCLP) is also caused by functional impairment in the *FGFR1, FGFR2, FGFR3,* and *FGF8* genes[35]. Cleft lip/palate ectodermal dysplasia syndrome (CLPED) is an autosomal recessive disorder that is also characterised by cleft lip with or without cleft palate. Positional cloning experiments identified mutations of the poliovirus receptor like-1 gene (*PVRL1*) in CLPED families of Margarita Island, Israel, and Brazil[36]. Investigation of the *W185X* nonsense mutation of *PVRL1* revealed its association with cleft in Northern Venezuela[37]. Whereas the same was not found in any of the non-syndromic cleft populations of Italian origin[38] and in the Taiwanese[39]. A common glycine allele of the G361V polymorphism was significantly associated with orofacial cleft phenotypes in all Iowan, Danish, and Filipino families[40]. The ectrodactyly-ectodermal dysplasia-cleft syndrome (EEC, MIM 129900) is a rare form of ectodermal dysplasia, an autosomal dominant disorder, which has given the hints to identify the genes such as *TP63* (OMIM 603273; 3q27)[41]. Analysis of a large family with 10 affected persons revealed maximal linkage between EEC and chromosome 3q27, which contains a known EEC gene and tumor protein 63 (*TP63*)[42]. Analysis of 16 exons of *TP63* gene in 100 Thai non-syndromic CLP
Role of Transcription Factor Gene Polymorphisms in the Pathogenesis of Non-Syndromic Cleft Lip with or without Cleft Palate- A Case Control Study

Introduction

Patients suggesting a causative role of *TP63* gene variants in non-syndromic CLP[43]. Cleft palate with ankyloglossia (CPX; MIM 303400) is inherited as a semi-dominant X-linked disorder previously described in several large families of different ethnic origins and has been the subject of several studies that localized the causative gene to Xq21. Positional cloning identified the CPX locus as the gene encoding T-BOX 22 (*TBX22*)[44]. *TBX22* is a member of the T-box containing transcription factor gene family that is conserved throughout metazoan evolution. Analysis of Thai non-syndromic CP cases for *TBX22* mutations confirms its importance, as a frequent cause of non-syndromic CPO across various populations[45]. Opitz syndrome is characterized by midline abnormalities such as hypertelorism, cleft palate, and hypospadias. Opitz syndrome is heterogeneous with an X-linked recessive form caused by mutations in the *MID1* gene that located at Xp22.3[46]. MID1 encodes a member of the B-box family of proteins, which contain protein-protein interaction domains, including a RING finger, and is implicated in fundamental processes like body axis patterning and regulation of cellular proliferation. A Non-random transmission that observed in *MID1* haplotypes in male NSCLP patients suggests that *MID1* may play an important role in NSCLP[47]. Lymphedema-distichiasis is an autosomal dominant disorder that is present classically with lymphedema at puberty and distichiasis at birth, and some other complications such as cardiac defects, varicose veins, ptosis, cleft palate, spinal extradural cysts, and photophobia[48,49]. This condition has been decisively linked to mutations in the fork head transcription factor *FOXC2* which have been primarily the frame shift mutations truncating the protein[48,49]. Bamforth-Lazarus syndrome (BLS) is a rare syndrome characterized mainly by congenital
Introduction

hypothyroidism, bilateral choanal atresia, cleft palate, bifid epiglottis, and spiky or curly hair. Mutations in the FOXE1 (9q22) gene have been associated with the Bamforth Lazarus syndrome[50]. Analysis of candidate genes within 9q22-q33 revealed that the FOXE1 is a major gene for cleft lip with or without cleft palate[51].

1.2.2: Animal models:

From the beginning of twentieth century, the mouse has been developed into a foremost mammalian model system for genetic research. Scientists from diverse fields have gravitated to the mouse because of its close genetic and physiological similarities to humans, as well as the ease with which its genome can be manipulated and analyzed[52]. Furthermore, the development of the embryonic phase is similar in mice and humans[53]. The development of the upper lip has been well-reviewed and presented for use of the CLP mutants recently[54]. Several genetically altered mouse models such as knockout, knock-in, gene-trapped, spontaneous mutations and chemically induced mutations have provided an excellent opportunity to study CLP. Many CLP mouse models have counterparts in syndrome CLP forms found in humans. Selection of an appropriate genetic model describing the genetic makeup of an underlying phenotype of interest is an intrinsic part of the gene mapping. Animal models with clefts arising spontaneously or as a result of mutagenesis experiments, serve as the best models for gene mapping studies of cleft lip with or without cleft palate[55]. In mice, for CLP, there are 10 spontaneous or induced Mendelian mutants, 2 conditional (tissue specific) knockout mutants, 1 new multifactorial model based on a compound
Introduction

mutant and 1 multifactorial genetic model represented by several related strains. Previous studies on mouse CLP mutants and human population data clearly indicate that the genes and signalling pathways that cause CLP are different from those that cause CPO[56].

In fact, studies in these animal models have helped to identify several genes essential for palatal formation. The Msx1 homeobox gene is expressed at diverse sites including facial primordial, and is required for expression of Bmp2 and Bmp4 in the palatal mesenchyme and Shh in the medial edge epithelia and plays a major role in signalling processes between tissue layers[57]. For Msx1, two independent knockouts that results in 100% cleft palate[58,59]. Mice lacking Msx1 function manifest a cleft secondary palate, a deficiency of alveolar mandible and maxilla, and a failure of tooth development[57,60]. The Tgfβ family represents a class of signalling molecules that play a major role in the development of or facial region, and its isoforms 1, 2 and 3 are expressed during this process and mediated through the Smad signalling system[61]. Mice lacking TGF-beta 3 exhibit an incompletely penetrant failure of the palatal shelves to fuse and leading to orofacial developmental abnormalities, particularly the cleft palate[62,63]. Endothelin-1 (EDN1) is synthesized by the vascular endothelialial cells and is found in plasma. Active endothelin peptides are produced from an inactive transitional form in a step that catalyzed by endothelin-converting enzyme (ECE). Mice lacking EDN1 have shown several craniofacial abnormalities, including cleft palate[64]. The mice deficient in ECA or in endothelin-A receptor genes also have shown almost identical abnormalities to those of EDN1-deficient mice[65]. TBX22
transcription studies in mouse and human tissues revealed posterior-specific expression profile within the developing palate[66]. Tbx22 null mouse exhibited a submucous cleft palate and ankyloglossia, which was similar to the human phenotypes. Further this study demonstrates that the Tbx22 is an important determinant for intramembranous bone formation in the posterior hard palate, which underpins the normal palate development[67].

In zebra fish, Endothelin-1 is expressed ventrally in the primordial of the pharyngeal arches and helps in patterning of pharyngeal cartilage. Kimmel et al[68] has demonstrated the role of EDN1 in the sizing of pharyngeal skeletal elements of the jaw and opercular regions in zebra fish upon injection of a morpholino. But genes involved in the endothelin pathway (EDN1, ECE1, EDNRA, EDNRB) have not shown significant linkage with orofacial clefts[69]. In the mammalian central nervous system the chief inhibitory neurotransmitter is gamma-aminobutyric acid (GABA). Mice homozygous for loss of function mutation in the GABA synthesizing enzyme glutamate decarboxylase 1 (GAD67) causes cleft palate, and indicates role of GABA in the palate development[70,71].

1.2.3: Linkage studies:

Linkage studies screen the whole genome and use parametric or nonparametric methods such as allele sharing methods with no assumptions on the mode of inheritance, penetrance or disease allele frequency. Large-scale family linkage analysis has provided a statistical method of detecting the chromosomal location of possible loci within a population where gene defects might result in a
predisposition to CLP. The first linkage analysis suggests possible linkage between CLP and the blood clotting factor XIII gene (*F13A*) on chromosome 6p[72]. Unfortunately, further evidence in support of the *F13A* gene has been more equivocal[73,74]. Suggestive linkage of chromosome 4q and CLP was shown in one family[75] and is supported by the evidence of association in the same region[76], but a later study showed evidence against linkage in this region[77]. However, successive linkage studies have revealed at least 20 contributing chromosomal regions[78]. A possible interaction between 6p23 and 2p13 chromosomal regions was reported in 38 CLP multiplex families from Italy[79]. Regions on Chromosomes 6p, 2p, 4q and 17q have all shown some evidence of linkage to CLP[80]. However, no evidence of linkage to selected candidate genes on Chromosomes 2, 4, 6 or 19 in Swedish multiplex CLP families was observed[81]. Similarly, 36 CLP multiplex Chinese families did not show significant linkage to any of these regions[82]. Segregation analyses using 11 multigenerational families with CLP found significant linkage with BCL3, a proto-oncogene on Chromosome 19q[83]. Subsequent studies failed to find further evidence of linkage to this marker, but they did find a significant association for an allele at this marker, using the transmission disequilibrium test (TDT)[84,85]. Support for linkage to different chromosome 1 regions (1q21 and 1q32-42.3) and CLP was reported with the positive scores for the 1p36 region near the *MTHFR* locus[86]. Significant linkage disequilibrium between the *GABRB3* gene and CLP was reported[87].
1.2.4: Genomic Rearrangements

Genomic rearrangements are caused by improper recombination of chromosomes and include deletions, duplications, translocations, and inversions that can occur within or between chromosomes. These rearrangements occur in approximately 0.5% of new borns by conventional cytogenetic analysis[88]. The majority of balanced rearrangements do not have phenotypic manifestations, 6% are associated with abnormal phenotypes[89]. In these cases, candidate genes can be identified by determining the genes disrupted at the breakpoints of these rearrangements. This method has been successful in identifying $CLPTM1$[90], $SATB2$[91], $SUMO1$[92] and $FGFR1$[93] as candidate genes for CLP, and implicating 9q and 17q as potential risk loci[94].

1.2.5: Copy number variants (CNVs)

Copy number variants (CNVs) and micro deletions are sub microscopic gains or losses of DNA segments, ranging from kilo bases to mega bases. CNVs can cause a phenotype by several different mechanisms including gene dosage effects[95], position effects[96], deletion of regulatory elements, and by uncovering recessive mutations and functional polymorphisms on the normal allele[97]. Although most CNVs have been found in syndromes forms of clefting, such as DiGeorge syndrome and VWS, recent studies have focused on the role of CNVs and micro deletions in nonsyndromic forms of clefting. Comparative genomic hybridisation analysis of multiple population cleft cases identified Estrogen receptor 1 ($ESR1$) and fibroblast growth factor receptor 2 ($FGFR2$) genes were responsible of CLP[98]. Shi et al. 2009[99] in both Denmark and Norway
population based case control study found deletions of SUMO1, TBX1, and TFAP2A are likely to be etiologic of CLP. Recently in an Iowa case study, one family with CPO found a deletion of MSX2 and additionally, identified deletions of ISM1 in 2 independent families with CLP [100].

1.2.6: Association studies:

A large number of putative genes chosen on the basis of biochemical and pharmacological evidence and candidate sites identified using genome scans and linkage studies have been tested for association with CLP. Genetic association studies intend to test whether single-locus genotype or minor alleles frequencies are different between disease phenotype and healthy controls. A specific allele has to be suspected as susceptible or protective and therefore tested in either a case-control study or a family-based association design. Several case-control association studies that conducted independently have provided evidence that variants in the genes MTHFR, ARNT, TGFA, GAD1, MSX1, RARA, TGFB3, IRF6, MSX1, TFAP2A are associated with the CLP phenotype. Association studies provide a powerful tool for identifying disease susceptibility genes of minor importance. In this regard, using the case-control design, Zucchero et al.[29] first observed that over-transmission of 274V allele in CLP subjects have been identified as a risk factor in Asians and South Americans, but it is not strong in European populations. This positive association between NSCLP and IRF6 was confirmed in American populations[101] and Belgian populations[102]. Several association studies in different populations showed the MSX1 CA 169- bp repeat allele polymorphism was significantly associated with increased risk of with
Introduction

NSCLP[33,103,104]. In a recent study in Korea reports the rs3821949 located in 5’ UTR region of MSX1 gene has been associated with NSCLP[105]. Recent technological advances made it feasible to perform high-throughput sequencing of candidate genes in large patient cohorts to know the genes involved in pathways that regulate transcription factors, growth factors, cell signalling and detoxification metabolisms, but we still have important work to do in the field of OFCs.

1.2.7: Genome-Wide Association Studies

Till date, there have been four GWAS on OFCs[106-109]. The first GWAS for NSCP was conducted in Germany on European origin samples and published by Birnbaum and colleagues. In this study, a 640-kb region at chromosome 8q24.11 was identified for the first time, with SNP rs987525 as a key susceptibility locus for NSCLP and also confirmed the role of IRF6 in the etiology of OFCs, which has been previously found in candidate gene studies, but failed to found any evidence of interaction between IRF6 and 8q24, suggesting that 8q24 and IRF6 confer the risk of OFCs through different pathway[107]. The impact of 8q24.11 on the risk of OFCs was also conformed in the case-control design GWAS conducted on U.S. cohort of European decent, by Grant and colleagues[109]. The third GWAS, published in 2010, found two additional susceptibility loci for NSCLP at chromosome 17q22 and 10q25.3 using an additional set of 177 cases and 940 controls along with the German sample in Birnbaum's study[108]. The fourth GWAS is the only case-parent trio study examining genetic effects in OFCs etiology up until now, using an international consortium that is part of the Gene-
Environment Association Studies Consortium (GENEVA). This study confirmed the $IRF6$ and 8q24 findings and identified risk variants near $MAFB$ and $ABCA4$ that are not previously found to be associated with NSCLP. Interestingly, this study revealed racial differences of genetic impact on NSCP after stratifying trios into European and Asian ancestry. Stronger evidence for association with $MAFB$ and $ABCA4$ were found in Asian families, whereas families of European ancestry have stronger evidence for association with chromosome 8q24[106]. Later Ludwig et al. in 2012[110] performed a meta-analysis by combining the GENEVA Cleft Consortium and the Mangold et al.[108] studies, resulted in six loci reaching genome-wide significance. Three of these (8q24, 10q25, and 17q22) were associated with NSCLP in prior studies. The three new loci on 2p21, 13q31, and 15q22 were implicated. With the addition of the Asian trios from the GENEVA study, five of these loci (except 15q22) had smaller p-values, indicating that these loci contribute to NSCLP in both European and Asian populations. These observations suggest not only that there are multiple genetic variants influencing risk of NSCLP but also that some of these genes may be differentially tagged by polymorphic markers in a population-specific manner.

1.2.8: Environmental factors causing cleft:

Although genes definitely play a role in the development of this defect, environmental factors, including maternal smoking, stress, heavy alcohol intake, infections, folic acid deficiency, and Vitamin A intoxication are suspected[111-113]. Several studies have revealed that high doses of corticosteroids cause cleft palate in rodents and rabbits[113-115]. Stressful life events appear to increase the
Introduction

maternal corticotrophin-releasing hormone and corticosteroid during pregnancy[116,117], and increased risk of clefts among infants born to those mothers[114,118]. Furthermore, usage of corticosteroid medications during the first trimester of pregnancy increased risk of oral clefts[119,120]. Smoking is practised by about a third of the world’s population aged 15 years or older, including some 12% of women[121]. Cigarette smoking during pregnancy is associated with cleft defects[6, 111,122]. In contrast, some studies found that smoking has no impact on any form oral clefts[123,124]. Meta-analysis of 24 case–control and cohort studies revealed significant association with an overall odds ratio of 1.34 (95% CI= 1.25–1.44)[125]. When maternal smoking is considered along with certain genotypes the effect is more significant. The maternal glutathione s-transferase θ-1 (GSTT1) genotype[126] or infant MSX1 genotypes[127] combined with maternal smoking significantly increased the risk of CLP. A recent Chinese case–parent trios study showed evidence of interaction between environmental tobacco smoke and the AC genotype of variant rs1044516 of IRF6[128].

The possible association between maternal alcohol use during pregnancy and risk of oral clefts in the offspring have been contradictory. Earlier studies demonstrated that the alcohol consumption during pregnancy is associated with orofacial clefts[123, 129]. Maternal alcohol consumption during the first trimester of pregnancy increased risk for isolated CP, but not for CLP[130]. Another study, found that maternal alcohol use during pregnancy among oral cleft cases was slightly more frequent than among controls[131]. A recent population-based case-
Introduction

control study demonstrated that even a low level of alcohol consumption during pregnancy can significantly elevate the risk of orofacial clefts in the offspring[132].

Vitamins of B complex including folic acid supplementation during pregnancy have been shown to be effective in preventing cleft lip and palate in humans[133]. Folic acid supplementation (0.4 mg/day) has been recommended by the US Public Health Service for all women of child bearing age to reduce the likelihood of birth defects[134]. Epidemiologic studies have proved that the consumption 4mg of folic acid per day is associated with a reduced risk of many birth defects[135].

1.2.9: Cleft lip and palate research in India

A large number of studies for NSCLP have been conducted mainly in the Caucasoid population. However, it is desirable to perform similar studies in different ethnic groups to reiterate the status of the association or linkage identified. Indian population is very well suited for such studies as it satisfies many criteria required for the investigation of complex traits. Besides, Indian population is an attractive large reservoir of samples due to its population size and the high rate of marital stability, the latter is essential for family based analysis. Work on genetics of NSCLP has been initiated since the last 5–6 years by several groups. Initial studies on genetics of cleft lip and palate was conducted by foreign researchers. Genome-wide scan on 38 CLP multiplex families of West Bengal revealed statistically significant two-point linkage results with markers on
Chromosome 7 (LOD = 1.89), Chromosome 5 (LOD = 1.76), Chromosome 15 (LOD = 1.55), and Chromosome 20 (LOD = 1.46)[136]. Analysis of two Indian pedigrees identified 11 genomic regions that could potentially harbour CLP susceptibility variations with most significant evidence of linkage for Chromosome 13q33.1-34 at marker rs1830756 with heterogeneity LOD score of 5.57[137]. A total of 323 NSCLP patients, 116 of their mothers, 108 of their fathers, and 214 normal controls that examined for IRF6 G820A, MTHFR C677T, and MTHFR A1298C revealed combined genotypes IRF6 GG/MTHFR 677CT together form greater risk for NSCLP[138]. Another south Indian case-control study don’t conform the role of MTHFRC677T and A1298C variants in the pathogenesis of NSCLP[139]. Recent south Indian population case-control study showed significant association between CBS c.844ins68 polymorphism and cleft lip and palate[140]. Another case-control study found that there is protective association between NOS327-bp VNTR polymorphism and NSCLP in the Indian population[141].

NSCLP is a polygenic, multi-factorial disorder, showing complex inheritance and so multiple genes in combination with lifestyle and environmental factors contribute to its aetiology. Considering the complex nature of NSCLP, with no clue to the number and nature of genes involved or their location, it is difficult to search for the genes using the conventional techniques that help unravel the genes for single gene disorders. The challenge is now to fine map the putative regions and identify genes in which variants are more likely to increase the risk for NSCLP. Therefore, it is anticipated that there are additional genes involved in
NSCLP that are yet to be identified, and the functional effects of identified mutations are yet to be discerned. Furthermore, the gene environment interaction will become more evident from the studies that include maternal and foetal genotypes along with gestational environmental exposures. The technological advancements in high-throughput genotyping and powerful statistical methods have accelerated the discovery of loci conferring susceptibility for complex diseases through the use of genome wide scans. Non-parametric method of linkage analysis using large collection of affected sib-pairs along with the genome scan studies using affected sib-pairs would be a more effective and favoured strategy to track genes of minor effect. Haplotype analysis could be informative where genotypic combination might be crucial in providing the susceptibility. More number of samples and families are needed to unravel the genes of minor effect and to establish the interaction of these genes with other genes and with the environment. Considering NSCLP is a polygenic disorder, gene-gene interaction among putative markers is expected to yield important clues to the underlying pathology.

So far, a variety of approaches have been explored to find out how genetic determinants control the risk of NSCLP. These include genome-wide linkage and association analysis as well as study of candidate genes selected on the basis of genes responsible syndromic forms of OFC, gene expression, or animal models with clefts. A major drawback lies in the analysis of patients with heterogeneous aetiology, since this dilutes the chances of finding positive gene phenotype correlations.
1.3: Transcription factors

Analysis of human genome has revealed differences in the binding of master regulators called ‘transcription factors to DNA that affect how genes are expressed in different people. In humans and other organisms, transcription factors account for as much as 10% of the coding genome. When activated, transcription factors turn on or off hundreds or thousands of genes, a cascade that programs the cells to grow or divide. Transcription factors are very diverse protein family and generally function in multi-subunit protein complexes. They may directly bind to the “promoter” regions of DNA or directly to the RNA polymerase molecule and activate or repress the transcription of a gene, a key determinant in whether the gene functions, at a given time. During development transcription factors are responsible for dictating the fate of individual cells. Transcription factors are a common way in which cells respond to extracellular information, such as environmental stimuli and signals from other cells.

"The activity of transcription factors determines what a cell is doing at any given moment". Transcription factors bind to the human genome within regions of the genome still viewed as a black box stretches of DNA sequence between known genes. Progressively, biologists have found that this DNA performs a vital role to turn genes on and off in specific situations. Recent studies show that SNPS and structural variations in transcription factors affect the binding ability to the genes it influence the gene expression[142].
Previous population based case-control studies and animal model studies demonstrated that the functional role of transcription factor genes in craniofacial development and also the variants in the genes played a crucial role in the etiology of NSCLP. Studies on humans showed that common alleles in \textit{IRF6} were associated with NSCLP in different populations[29]. Recent research has shown that \textit{Irf6} mutant mice exhibit a hyper-proliferative epidermis that fails to undergo terminal differentiation, which leads to multiple epithelial adhesions that can occlude the oral cavity and result in cleft palate[143]. Studies showed that complete functional loss of transcription factor \textit{Satb2} leads to increased apoptosis in the developing jaw primordia and subsequent down-regulation of the expression of genes (\textit{Pax9}, \textit{Alx4}, and \textit{Msx1}) involved in craniofacial development in humans and mice[144,145]. \textit{Tbx22} null mouse showed reduced palatal bone formation in the posterior hard palate which prevented normal palate development and causes a submucous cleft palate[67]. Genome-wide association studies also identified three susceptible transcription factor \textit{IRF6}, \textit{MAFB}, \textit{ABCA4} genes for NSCLP[106, 110].

All these findings showed the impartment role of transcription factors in craniofacial development. However, the contribution of transcription factors to NSCLP in India that comprises one of the largest global populations is unclear. Collectively this has led to the following aims and objectives of this research.
1.3.1: Interferon regulatory factor-6 (IRF6):

Interferon regulatory factor-6 (IRF6) is located in chromosome 1q32.2 position. This gene contains 10 exons and in this exons 1, 2, and 10 are noncoding. IRF6 encodes protein contains total 517 amino acids and contains an N-terminal winged-helix DNA-binding domain and a C-terminal SMIR (Smad-interferon regulatory factor-binding) domain. In situ hybridization of mouse embryos demonstrated that Irf6 is highly expressed in the medial edges of the paired palatal shelves immediately before, and during, their fusion. Irf6 expression was also detected in hair follicles, palatal rugae, tooth germ, thyroglossal duct, external genitalia and on the skin throughout the body[28].

**Figure 3: Location of IRF6 gene in Chromosome 1**

Chromosome 1 showing Genomic region of IRF6 gene (Figure from http://ghr.nlm.nih.gov/gene)

1.3.2: Transcription factor AP-2 alpha (TFAP2A):  

The gene coding for AP-2α transcription factor (TFAP2A) is located on 6p24-p22.3 and composed of seven exons with the last six exons encoding the majority of the protein. AP-2 alpha protein acts as a sequence specific DNA-
binding transcription factor, in recognizing and binding to the specific GC-rich regions that are present in the cis-regulatory regions of several cellular genes. Thus it regulates gene expression during the embryogenesis of eye, ear, face, body wall and limbs[146].

**Figure 4: Location of TFAP2A gene in Chromosome 6**

![Chromosome 6 showing Genomic region of TFAP2A gene](http://ghr.nlm.nih.gov/gene)

1.3.3: V-mafmusculoaponeuroticfibrosarcomaoncogene homolog B (*MAFB*):

Transcription factor MafB is encoded by *MAFB* gene located in 20q11.2-q13.1 chromosome position, consists of a single exon and spans around 3 kb. MafB is a basic leucine zipper (bZIP) transcription factor that plays an important role in the regulation of lineage-specific hematopoiesis, and is a putative tumor suppressor in the myeloid lineage. It plays a key role in monopoiesis as well as monocyte-dendritic cell differentiation[147].
Introduction

Figure 5: Location of MFAB gene in Chromosome 20

Chromosome 20 showing Genomic region of MFAB gene (Figure from http://ghr.nlm.nih.gov/gene).

1.3.4: ATP-binding cassette, sub-family A (ABCA4):

ATP-binding cassette, sub-family A (ABC1), member 4, (abca4) encoded by ABCA4 gene belong to a ATP-binding cassette (ABC) superfamily located in 1p22.1, contains 50 exons and spans an estimated 150 kb. ABCA4 is a large retina-specific protein with two transmembrane domains (TMD), two glycosylated extracellular domains (ECD), and two nucleotide-binding domains (NBD) and is almost exclusively expressed in retina localizing in outer segment disk edges of rod photoreceptors.

Figure 6: Location of ABCA4 gene in Chromosome 1

Chromosome 1 showing Genomic region of ABCA4 gene (Figure from http://ghr.nlm.nih.gov/gene).
1.3.5: SATB homeobox 2 (SATB2):

SATB2 gene which is located on chromosome position 2q33.1, encodes AT-rich sequence binding protein and is composed of 733 amino acids with a molecular weight of 82.5 kDa. The protein contains two degenerate homeo domain regions known as CUT domains (amino acid 352–437 and 482–560) and a classical homeo domain (amino acid 614–677). Satb2 is the first cell-type-specific transcription factor that functions as a regulator of the transcription of large chromatin domains[148].

**Figure 7: Location of SATB2 gene in Chromosome 20**

Chromosome 2 showing Genomic region of SATB2 gene (Figure from http://ghr.nlm.nih.gov/gene)

1.3.6: Muscle segment homeobox 1 (MSX1):

Muscle segment homeobox 1 is also known as MSX1 contains 297 amino acids, encoded by the MSX1 gene maps to chromosome 4p16.1[149] and exhibits homology of synteny with the murine MSX1 gene located on chromosome 5[150]. MSX1 gene is a member of MSX family spans 4.05 kb, contains two exons and one intron. MSX1 acts as a transcriptional repressor during embryogenesis through interactions with components of the core transcription complex and other homeoproteins.
1.3.7: T-box transcription factor 22 (TBX22)

The gene encoding T-box transcription factor 22 (TBX22) is composed of seven exons spanning 8.7 kilobases located in Xq21.1 position. The TBX22 encodes a 400-amino-acids protein containing a T-domain in its NH2-terminal region which has the unique feature of missing 20 amino-acids relative to the other known T-domains. Naturally occurring missense mutations in the T-box domain of X-linked cleft patients showed a marked down-regulation or absence of SUMO-1 conjugation[151].

Chromosome 4 showing Genomic region of MSX1 (Figure from http://ghr.nlm.nih.gov/gene).
Given the important role of transcription factors in the mammalian cellular differentiation and embryonic development, the present study is designed to emphasize our current understanding of how common polymorphisms of the transcription factor genes affect orofacial clefts.
1.4: AIM OF THE PRESENT STUDY

The present study is aimed to investigate the role of transcription factors gene polymorphisms in the pathogenesis of NSCLP in south Indian population. Special emphasis was directed to:

1.4.1: OBJECTIVES

   i) Investigate the association of polymorphisms in transcription factors coding genes with the risk for non-syndromic cleft lip and palate.
   
   ii) Study the haplotype and linkage disequilibrium profiles in the transcription factor genes

   iii) Evaluate gene-gene interactions of SNPs of transcription factors genes in relation to non-syndromic cleft lip and palate

   The rest of the thesis is divided into four chapters. Chapter two documents the Methodology used in this thesis. The main findings are put forward and discussed in chapters 3 and 4 respectively. Finally, chapter 5 presents summary and conclusions of the study, based on which directions for future research will be suggested.