DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
199

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
199

TIIT NIKOPENSIUS

Genetic predisposition to nonsyndromic orofacial clefts

TARTU UNIVERSITY PRESS

Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in molecular diagnostics) on 12.07.2011 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Prof. Andres Metspalu, MD, PhD
Department of Biotechnology, Institute of Molecular and Cell Biology, and Estonian Genome Center, University of Tartu, Estonia

Opponent: Assist. Prof. Alexandre R. Vieira, DDS, PhD
Department of Oral Biology and Department of Pediatric Dentistry, School of Dental Medicine, University of Pittsburgh, Pittsburgh, USA

Commencement: Room No 217, 23 Riia Str., Tartu, on August 22nd 2011, at 14.00

The publication of this dissertation is granted by the University of Tartu.

The publication of the current thesis was supported by Targeted Financing from the Estonian Ministry of Education and Research (SF0180142s08), the Estonian Science Foundation grant ETF7076, the European Science Foundation Frontiers of Functional Genomics Exchange grant 2182, the EU FP7 grant ECOGENE (#205419, EBC), the EU FP7 grant OPENGENE (#245536, EGC/University of Tartu), and by the EU via the European Regional Development Fund grant to the Centre of Excellence in Genomics, the Estonian Biocentre, and the University of Tartu.

ISSN 1024–6479
ISBN 978–9949–19–810–8 (trükis)
ISBN 978–9949–19–811–5 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 485

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

Ref. II **Nikopensius T***, Jagomägi T*, Krjutškov K, Tammekivi V, Saag M, Prane I, Piekuse L, Akota I, Barkane B, Krumina A, Ambrozaitytė L, Matulevičienė A, Kučinskienė ZA, Lace B, Kučinskas V, Metspalu A. Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. *Birth Defects Res A Clin Mol Teratol* 2010; 88(9): 748–756.

Ref. III **Nikopensius T***, Kempa I*, Ambrozaitytė L*, Jagomägi T, Saag M, Matulevičienė A, Utkus A, Krjutškov K, Tammekivi V, Piekuse L, Akota I, Barkane B, Krumina A, Klovins J, Lace B, Kučinskas V, Metspalu A. 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* 2011; 91(4): 218–225.

Ref. IV **Nikopensius T***, Ambrozaitytė L*, Ludwig KU, Birnbaum S, Jagomägi T, Saag M, Matulevičienė A, Linkevičienė L, Herms S, Knapp M, Hoffmann P, Nöthen MM, Kučinskas V, Metspalu A, Mangold E. Replication of novel susceptibility locus for non-syndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients. *Am J Med Genet A* 2009; 149(11): 2551–2553.

Ref. V **Nikopensius T**, Birnbaum S, Ludwig KU, Jagomägi T, Saag M, Herms S, Knapp M, Hoffmann P, Nöthen MM, Metspalu A, Mangold E. Susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 10q25 confers risk in Estonian patients. *Eur J Oral Sci* 2010; 118(3): 317–319.

* These authors contributed equally to this work

Author's contributions:

Ref. I	statistical analysis, participated in study design and manuscript writing
Ref. II, III	study design, statistical analysis, manuscript writing
Ref. IV, V	performed the experiments in Estonian sample, participated in study design, statistical analysis, and manuscript writing

LIST OF ABBREVIATIONS

BMP	bone morphogenetic protein
CEU	U.S. residents (Utah) with northern and western European ancestry
CL	cleft lip
CLP	cleft lip and palate
CL/P	cleft lip with or without cleft palate
CLPTM	cleft lip and palate transmembrane
CNC	cranial neural crest
COL	collagen
CNV	copy number variation
CP	cleft palate
CYP	cytochrome P450
DNA	deoxyribonucleic acid
ECM	extracellular matrix
EEC	ectrodactyly, ectodermal dysplasia, and clefting (syndrome)
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
FOXE1	forkhead box E1
GST	glutathione S-transferase
GWAS	genome-wide association study
HWE	Hardy-Weinberg equilibrium
IRF	interferon regulatory factor
kb	kilobase, 1.000 base pairs
LD	linkage disequilibrium
LHX	LIM homeobox
MAF	minor allele frequency
MMP	matrix metalloproteinase
MSX	muscle segment homeobox
MTHFR	methylenetetrahydrofolate reductase
OR	odds ratio
PVRL	poliovirus receptor-like
SNP	single nucleotide polymorphism
tagSNP	haplotype-tagging SNP
TBX	T-box transcription factor
TGF	transforming growth factor
TIMP	tissue inhibitor of matrix metalloproteinases
UTR	untranslated region
WNT	wingless-type MMTV integration site family

INTRODUCTION

Orofacial clefts are among the most common congenital malformations worldwide, requiring complex multidisciplinary treatment throughout childhood and having lifelong medical and psychosocial implications for affected individuals. The worldwide prevalence of these defects is estimated as approximately 1 per 600 liveborn babies, and their incidence varies according to geographic origin, ethnicity and socioeconomic status. Orofacial clefts comprise a heterogeneous group of disorders affecting the lips and oral cavity. They can be classified into several subgroups on the basis of anatomical, genetic and embryological findings. The two most common types of orofacial clefts are cleft lip with or without cleft palate (CL/P) and isolated cleft palate (CP), representing a complex phenotype and reflecting a breakdown of the molecular mechanisms involved during craniofacial morphogenesis. Collectively, orofacial clefts have significant clinical impact requiring extensive dental and orthodontic treatment together with speech and hearing therapy. Likewise, referring to psychotherapy and genetic counselling may be recommended as well. Understanding the etiology of orofacial clefts is an important step towards improved treatment and proposing preventive strategies.

Nonsyndromic cases of both CL/P and CP are considered to have multifactorial etiology, determined by the interaction of multiple genetic, environmental and lifestyle risk factors. Several genes known to underlie Mendelian syndromic forms of clefts seem also to play a role in the etiology of nonsyndromic clefts. The large extent of familial clustering, high recurrence risks for siblings and greatly increased concordance rates in monozygotic versus dizygotic twins indicates the importance of genetic factors in the etiology of CL/P. However, only a small proportion of the overall contribution of genetic variants causing nonsyndromic clefts is estimated to have been identified to date. Despite the most recent developments, genetics of orofacial clefts remains controversial due to uncertain mode of inheritance, incomplete penetrance, and heterogeneity both within and among populations.

Human studies have used both association and linkage analyses to identify genetic determinants of orofacial clefting. The results of the earlier candidate gene-based association studies, performed in different populations, have been mostly inconclusive or conflicting, with only a few candidate loci being implicated in cleft phenotypes. This inconsistency could have been caused mostly by inadequate power and genetic heterogeneity. However, recent years have witnessed significant advances in the understanding of genetic determinants of nonsyndromic clefting and how these genes influence the embryonic development of the face. Moreover, advances in microarray-based genotyping technology have made large-scale association studies using thousands of cases and control individuals feasible. In particular, genome-wide association studies (GWAS) are emerging as a powerful technique in studies of complex traits, and according to the expectations, the first achievements in the efforts to discover additional major susceptibility loci for CL/P have already been made.

Large sample sizes are required in genetic epidemiological studies of complex traits, such as orofacial clefts, to identify etiologically relevant risk factors with adequate statistical power, and studies relying on linkage disequilibrium (LD) can be more effective in detecting more subtle gene effects derived from multiple common and low-penetrance alleles. Nonetheless, phenotypic and genetic heterogeneity may remarkably hamper the discovery of the genuine disease-causing variants. In addition, confounding factors like allelic heterogeneity and population diversity may often lead to spurious association findings. Recent advances in bioinformatics have made possible to combine case-control samples from different populations in the scenario when study sample(s) derived from one population suffer from limited sample size. Well-characterized study populations are a prerequisite to study the genetic component of the complex traits, and therefore, prior assessment of the genetic structure appears to be most benefitial to combine large genotype datasets to achieve a significant gain in study power.

The intent of the present study was to follow-up published reports of genes that have been implicated in previous studies for evidence of association with nonsyndromic oral clefts and to confirm the relevance of previous findings in an independent study population. Therefore, we have assembled a new clefting sample of 1010 subjects from a previously uncharacterized ethnic background, comprised of three genetically close populations from the Baltic region – Estonia, Latvia and Lithuania.

The main goals of the present study were: 1) to investigate the role of genetic variants in 40 candidate genes for clefting in susceptibility to nonsyndromic cleft palate; 2) to investigate the role of genetic variants in 40 candidate genes for clefting in susceptibility to nonsyndromic cleft lip with or without cleft palate; and 3) to carry out a replication study in Estonian and Lithuanian CL/P case-control samples to test the best findings from the first genome-wide association study, conducted in German population and extended further to other Central European populations.

1. REVIEW OF LITERATURE

1.1. Orofacial clefts

Craniofacial development comprises one of the most complex events during early stages of embryonic development, coordinated by a network of transcription factors and signaling molecules together with proteins conferring cell polarity and cell-cell interactions. Disturbance of this tightly controlled cascade may result in a facial cleft where the facial primordia ultimately fail to join and form the appropriate structures (Stanier & Moore, 2004). The most common features of craniofacial abnormalities are the orofacial clefts, which include cleft lip, cleft lip and palate, and cleft palate alone. Clefts of the lip or clefts of the lip with the palate arise in the primary palate, whereas clefts of the palate alone occur in the secondary palate (Murray & Schutte, 2004). The clinical manifestations of these defects are diverse, ranging from unilateral clefts of the lip and isolated clefts of the secondary palate to complete bilateral clefts of the lip and palate (Figure 1).

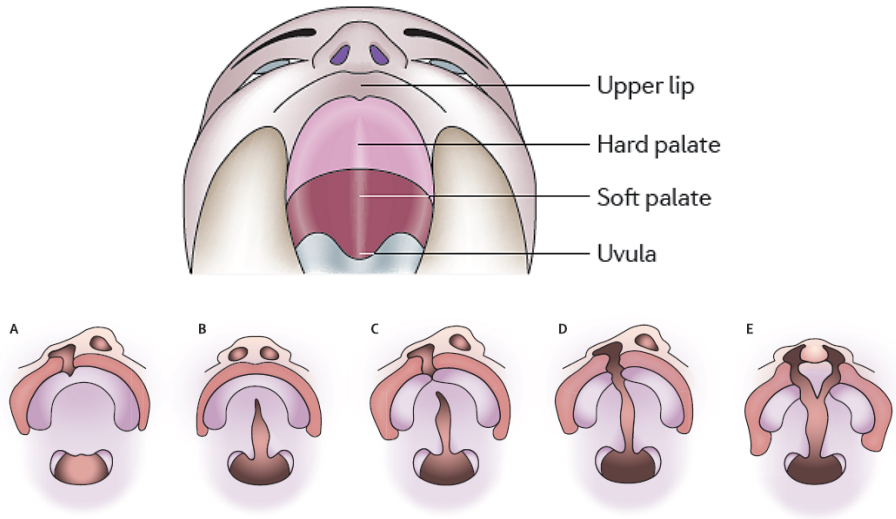

Figure 1. Nonsyndromic orofacial clefts (Muenke, 2002; Mossey et al., 2009).
(A) Cleft lip and alveolus. (B) Cleft palate. (C) Incomplete unilateral cleft lip and palate. (D) Complete unilateral cleft lip and palate. (E) Complete bilateral cleft lip and palate.

Orofacial clefts result in complications affecting feeding, speech, hearing, appearance, and cognition. These effects can lead to long-term adverse outcomes for health and social integration, and affected children need complex multidisciplinary care from birth until adulthood (Mossey et al., 2009). Significantly increased lifetime mortality ratios of 1.4 for males and 1.8 for females are associated with CL/P, with an increased risk of all major causes of death

(Christensen et al., 2004). Long-term outcomes in affected individuals include higher overall morbidity rates contributed mostly by higher incidence of psychiatric disorders in adults (Christensen & Mortensen, 2002) and cancer: an increased occurrence of breast and brain cancer among adult females born with clefts, and an increased occurrence of primary lung cancer among adult males born with clefts have been reported (Bille et al., 2005). Currently, in developed countries, the mortality for isolated clefts is equivalent to background neonatal mortality (Druschel et al., 1996), but may be strikingly elevated for infants born with clefts and additional anomalies suggestive of an underlying syndrome, especially in settings with inadequate health care (Murray et al., 1997).

Orofacial clefts are classified as nonsyndromic (isolated) or syndromic based upon the presence of other congenital anomalies or developmental delay. The majority of CL/P cases (~70%) and approximately 50% of CP cases are regarded as nonsyndromic, occurring as an isolated condition without other recognizable anomalies, while the remaining syndromic cases have additional characteristic features that can be subdivided into categories of chromosomal abnormalities, recognizable Mendelian single gene syndromes, teratogenic effects and various uncategorized syndromes (Stanier & Moore, 2004). Our understanding of the etiology and pathogenesis of orofacial clefts, particularly the nonsyndromic forms, is still in relatively early stages, reflecting the genetic complexity of clefting and diversity of the mechanisms involved at the molecular level during craniofacial morphogenesis in early stages of embryonic development, with both genetic and environmental factors having an important role (Murray, 2002; Cobourne, 2004).

It is generally accepted that CL/P and CP are genetically distinct phenotypes in terms of their inheritance patterns. Significant advances have been made in the identification of numerous genes and pathways critical for craniofacial development using direct sequencing of primary candidate genes, mutagenesis experiments in animal models, gene expression patterns in facial morphogenesis, breakpoint mapping, association studies with candidate genes or loci and genome-wide scans in large multiplex families. However, the overall contribution of genetic variants to clefting phenotypes has comprised to date only a modest fraction of the recognized etiologies (Vieira, 2008b). Implementation of increasingly powerful combination of careful collection of large samples from diverse ethnic background integrated with the most up-to-date microarray-based genotyping technologies and sophisticated statistical evaluations will largely expand our knowledge of genetic mechanisms underlying orofacial clefting.

1.1.1. Developmental pathogenesis

Integration of findings of human genetic studies (including positional cloning strategies, parametric linkage analysis, nonparametric affected sib-pair approaches, analysis of chromosomal rearrangements, and candidate gene-based association studies) with data of experimental embryological techniques in

model organisms has increased our knowledge of fundamental mechanisms driving normal craniofacial morphogenesis and how these mechanisms are disturbed in CL/P and isolated CP (Mossey et al., 2009).

Embryonic development (Bender et al., 2000; Stanier & Moore, 2004; Mossey et al., 2009).

Development of the lip and palate (Figure 2) comprises a precisely co-ordinated cascade of developmental processes involving cell migration, growth, differentiation and apoptosis results in the development of craniofacial structures from the originating oropharyngeal membrane (Sperber, 2002).

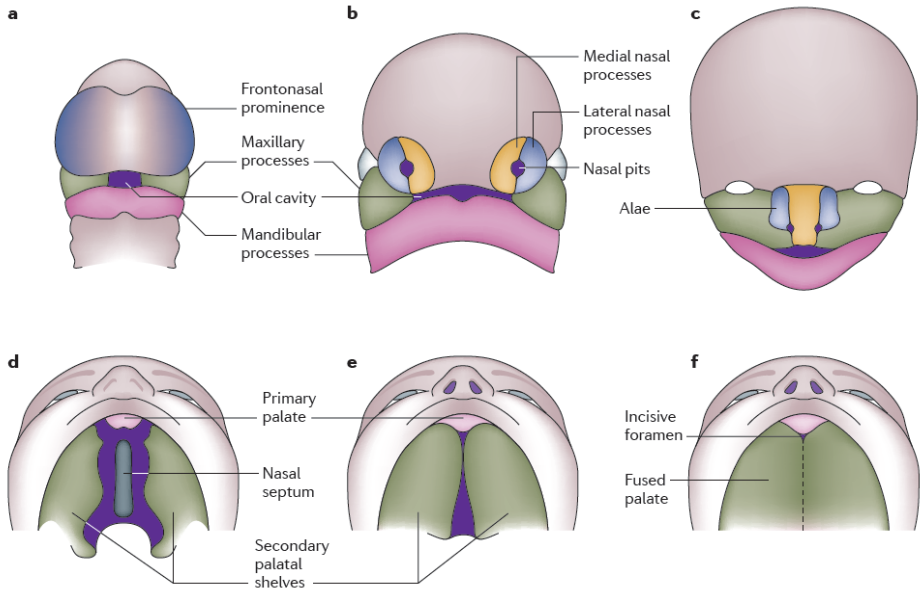

Figure 2. Development of the lip and palate in humans (Dixon et al., 2011).
a) The developing frontonasal prominence, paired maxillary prominences and paired mandibular prominences surround the primitive oral cavity by the fourth week of embryonic development. **b)** By the fifth week, the nasal pits have formed, which leads to the formation of the paired medial and lateral nasal prominences. **c)** The medial nasal prominences have merged with the maxillary prominences to form the upper lip and primary palate by the end of the sixth week. The lateral nasal prominences form the nasal alae. Similarly, the mandibular prominences fuse to form the lower jaw. **d**) During the sixth week of embryogenesis, the secondary palate develops as bilateral outgrowths from the maxillary prominences, which grow vertically down the side of the tongue. **e)** Subsequently, the palatal shelves elevate to a horizontal position above the tongue, contact one another and commence fusion. **f)** Fusion of the palatal shelves ultimately divides the oronasal space into separate oral and nasal cavities.

Development of the human face begins in the 4th week of human embryonic development, when neural crest cells from the dorsal region of the anterior neural tube (cranial neural crest, CNC) proliferate and migrate into the frontonasal and visceral arch region and combine with mesodermal cells to form the five facial primordia consisting of the frontonasal prominence, two maxillary prominences, and two mandibular prominences. The maxillary prominences enlarge and grow towards each other and the nasal prominences. The frontonasal prominence forms the forehead and the nose. The maxillary prominences are bilateral and form the lateral stomodeum (primitive mouth), while the mandibular prominences are also bilateral and responsible for the caudal growth of the stomodeum. The neural crest cells within these prominences differentiate into skeletal and connective tissue of the face, bone, cartilage, fibrous connective tissue, and all dental tissues except enamel (Lettieri, 1993). During the 4th week, the medial ends of the mandibular prominences merge to form the mandible, lower lip, and lower cheek region. Formation of the nasal placodes by end of the 4th week divides the lower portion of the frontonasal prominence into paired medial and lateral nasal processes. The nasal or olfactory pits form and extend into the primitive mouth and become the nostrils (Gorlin et al., 1990). By the end of the 6th week, rapid proliferation of the maxillary prominences results in the medial nasal prominences merging with each other and with the lateral maxillary prominences leading to formation of the upper lip and the primary palate. The lower lip and jaw are produced by the mandibular prominences, which merge across the midline (Jugessur & Murray, 2005).

Palate development begins during the 5th week of embryogenesis, after fusion of the upper lip. Fusion of the hard palate is completed by the 10th week and development of the soft palate and uvula is completed in the 12th week with successful merging of the secondary growth centers (Gorlin et al., 1990).

Development of the secondary palate begins early in the 6th week with outgrowth of two palatal shelves from the maxillary prominences, initially by growing vertically down the sides of the developing tongue. During the 7th–8th week of development, apoptosis and epithelial-mesenchymal transformation (EMT) at the medial edges enable the palatal shelves, consisting of rapidly proliferating mesenchymal cells, to fuse to form a midline epithelial seam after the shelves have ascended to an appropriate position above the flattening tongue. Proteins such as integrins, matrix metalloproteinases, microtubules and actin cytoskeletons are involved in the EMT process (Cox, 2004). The palatal mesenchyme then differentiates into bony and muscular elements that are correlated with the position of the hard and soft palate, respectively. In addition to fusing in the midline, the secondary palate fuses with the primary palate and the nasal septum. These fusion processes are complete by the 10th week of embryogenesis; development of the mammalian secondary palate thereby divides the oronasal space into separate oral and nasal cavities, allowing respiration and mastication to take place simultaneously (Sperber et al., 2002).

Developmental gene networks

The molecular events that underlie the formation of orofacial structures are under the strict control of a cascade of genes encoding for a variety of molecules implicated in signaling facial primordia identity, epithelial differentiation and palatal shelf remodeling. These include ECM molecules and growth factors including bone morphogenetic proteins (Bmp), sonic hedgehog (Shh), fibroblast growth factors (FGF), members of the transforming growth factor b (Tgfb) superfamily, and various transcription factors (Stanier & Moore, 2004).

A combination of apoptosis and epithelial-mesenchymal transformation interactions are crucial in fusion events that contribute to formation of the lip and primary palate. These events are suggested to include SHH, MSX1 and MSX2; and control of signaling by bone morphogenetic proteins and fibroblast growth factors in part by TP63 (Sun et al., 2000; Thomason et al., 2008). Molecular control of palatal shelf initiation and vertical growth involves complex signaling cascades with transcription factors and various growth factors and their receptors, including *Lhx8*, *Msx1*, *Osr2*, *Fgf10*, *Fgfr2b*, *Tgfb2*, and *Tgfbr2* (Gritli-Linde et al., 2007). Signaling between the palatal epithelium and mesenchyme has an important role in palatal growth regulation – for example, fibroblast growth factor 10 (FGF10) signals from the palatal mesenchyme to its receptor FGFR2b, which is expressed in the palatal epithelium. Loss of function of either FGF10 or FGFR2b causes a reduction in mesenchymal proliferation and an increase in apoptosis, leading to truncation of the palatal shelves. In addition, activation of FGFR2b by FGF10 is crucial for maintenance of SHH expression in the palatal epithelium: loss of SHH function in this tissue leads to cleft palate (Rice et al., 2004). Signaling between the epithelium and mesenchyme during palatal growth occurs also between *Msx1*, *Bmp4*, *Shh*, and *Bmp2*. Molecular studies have shown that *Bmp2* and *Bmp4* are expressed within the epithelia and mesenchyme of the palatal shelves, while *Shh* plays an important role in the early induction of facial primordia. The *Msx1* homeobox gene, which is also expressed in the facial primordia, regulates expression of *Bmp2* and *Bmp4* in the palatal mesenchyme and *Shh* and *Bmp4* in medial edge epithelium. In turn, *Shh* stimulates *Bmp2* expression in the mesenchyme, which regulates growth of the palatal shelves (Zhang et al., 2002). Transcription factors such as the distal-less (Dlx), Hox, Gli and T-box families have also important roles in maxillary and mandibular specification and are regulated by Shh, Bmp and Fgf signals (Richman & Lee, 2003).

Palatal fusion appears to be driven by several cell adhesion molecules (including PVRL1), desmosomal components, and growth factors including transforming growth factor α (TGFA) and epidermal growth factor receptor (EGFR) (Suzuki et al., 2000; Mogass et al., 2000; Miettinen et al., 1999). The transforming growth factor β superfamily is particularly interesting in palate development and isoforms 1, 2 and 3 are all expressed during this process. TGFB3 is expressed earliest and is found in the epithelial component of the vertical shelves. It is also expressed later in the horizontal shelves and medial edge epithelia, but expression is undetected once the epithelial seam disrupts

(Stanier & Moore, 2004). While TGFB1 and 2 accelerate palatal shelf fusion, TGFB3 may play a role in growth inhibition and is crucial for the first adhesive interaction (Fitzpatrick et al., 1990). Recent evidence suggests that their function in the embryonic palate is at least in part mediated through the Smad signaling system (Greene et al., 2003). The Tgfb3 knock-out mice exhibit an isolated CP through failure of palatal shelf fusion (Kaartinen et al., 1995; Proetzel et al., 1995). Additional genetic factors involved in palate development have been described using mouse transgenic models; in particular, both Msx1 and Lhx8 have been implicated in palatal mesenchymal proliferation during palatogenesis, and the respective null mice have CP because of the palatal shelves failure to meet and fuse (Satokata & Maas, 1994; Zhao et al., 1999). As a general model, insufficient mesenchyme is believed to be the most common reason for CP in mice (Wilkie & Morriss-Kay, 2001).

Subsequent developmental studies have suggested that TGFB3 may promote palatal fusion via synergistic effects – by stimulating initial adhesion of the palatal shelves, increasing the surface area of the medial edge epithelium and by promoting degeneration of medial edge epithelium (Kaartinen et al., 1997; Taya et al., 1999; Tudela et al., 2002). Tissue remodeling during palatal fusion involves a combination of basement membrane degradation and epithelial-mesenchymal transformation, which are both under the control of specific matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs), which have been implicated in proteolytic degradation of the ECM. In the Tgfb3 knock-out mice, the palatal expression levels of TIMP2 and MMP13 are markedly reduced and their expression is dependent on TGFB3 (Blavier et al., 2001). Importantly, IRF6 is down-regulated in the medial edge epithelium of mice with mutations in *Tgfb3* and *Tgfbr2*, which suggests strongly that IRF6 lies downstream of TGFB3 signaling for the fate of medial edge epithelium (Knight et al., 2006; Xu et al., 2006).

1.1.2. Descriptive epidemiology

A remarkable interpopulation variation in birth prevalence of orofacial clefts has been described, with Asian and Amerindian populations tending to have the highest frequencies, often at ≥1:500, with Caucasian populations intermediate, and African populations the lowest at 1:2500 (Murray, 2002). Nonsyndromic CL/P is reported to occur in approximately 1 in 700 newborns worldwide, with the incidence varying according to parental race/ethnicity and geographic origin, and the socioeconomic status of the affected families (Bender et al., 2000). Native American Indians have the highest incidence of CL/P, 3.6 in 1000 newborns, followed by Japanese 2.1 in 1000 newborns, Chinese 1.7 in 1000 newborns, and Caucasian populations, with 1 in 800 to 1 in 1000 livebirths, while African-Americans have the lowest incidence of CL/P at 0.3–0.4 in 1000 newborns (Croen et al., 1998; Tolarova & Cervenka, 1998). These observations suggest that the relative contribution of individual susceptibility genes may vary

across different populations (Mossey et al., 2009; Beaty et al., 2010). In contrast, CP is less common, with a prevalence of approximately 1 in 1500 births in most racial backgrounds (Lidral & Murray, 2004). Several studies indicate that the incidence of nonnative Philippine and Chinese infants born in the United States have a lower incidence of CL/P than those born in their native country (Murray et al., 1997; Croen et al., 1998). The incidence of orofacial clefts is relatively uniform in European populations; however, there are exceptions with some particular geographic areas (e.g., Finland) which have higher frequencies, possibly related to founder effects or environmental triggers (Figure 3).

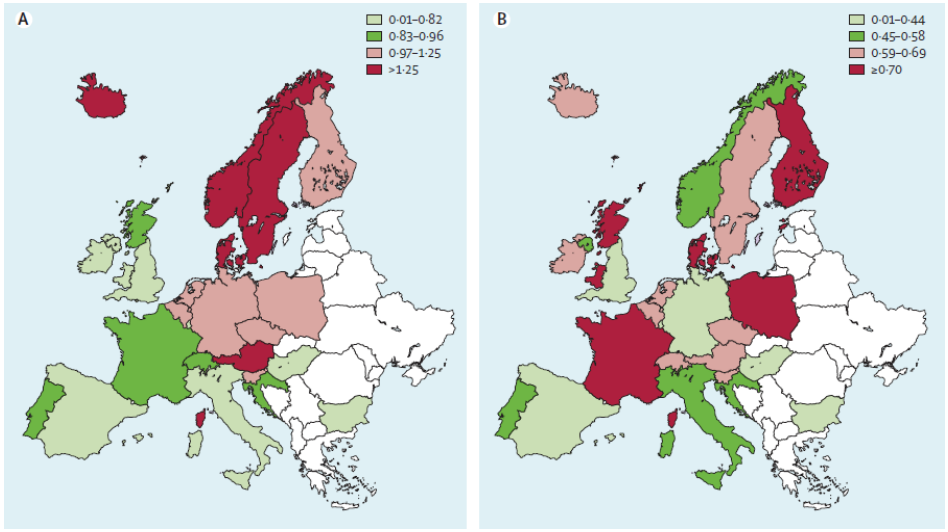

Figure 3. European birth prevalence per 1000 livebirths of nonsyndromic cleft lip and palate (Mossey et al., 2009; http://www.eurocran.org).
(A) Cleft lip with or without cleft palate. (B) Isolated cleft palate.

Numerous epidemiological studies across various ethnic groups have reported an unequal sex distribution for both nonsyndromic CL/P and CP. CL/P occurs more frequently in males, while affected females outnumber affected males in CP. This may suggest the existence of genes with a differential impact on cleft formation depending on the sex of the affected individual (Mangold et al., 2009). In populations of European descent, the sex ratio for CL/P is approximately 2:1 (male to female) and the male excess becomes more apparent with increasing severity of cleft (Mossey & Little, 2002). The sex ratio is reversed for CP, with male to female ratio estimates ranging from 1:1.25 to 1:2 (Grosen et al., 2010; Dixon et al., 2011). The sex ratio varies with the presence of additional malformations, number of affected siblings in a family, and ethnic

origin. The male predominance in CL/P is smaller when more than one sibling is affected in the family, and likewise, the male excess is less apparent when the infant has malformations of other systems (Mossey & Castilla, 2003). By contrast, findings from a large study suggest predominance in females when the father's age exceeds 40 years (Rittler et al., 2004).

The incidences of clinical subtypes for CL/P are as follows: left unilateral clefts are the most frequent, followed by right unilateral and bilateral clefts with a ratio of 6:3:1. Approximately 70% of the unilateral clefts and 85% of the bilateral clefts involve the palate (Lettieri, 1993).

Clefts of the lip and/or palate are more severe when both cleft types are present. In a large population-based study, 18% of CL cases were severe (i.e., complete cleft of the primary palate) in the absence of CP, compared with 81% when CP was also present. Similarly, among babies with CP, 40% were severe (complete cleft of the secondary palate) in the absence of CL, compared with 93% when CL was also present. Girls were more likely to have severe clefts, as were patients who had other types of congenital anomalies. Although cleft lip was more frequent on the left side, clefts were not more severe on the left side. In bilateral CL, the severity was similar on both sides (Sivertsen et al., 2008a).

In combined data from European registries for 1995–1999, 3.5% of babies with CL/P were stillborn and 9.4% were from terminated pregnancies; respective proportions for isolated CP were 2.4% and 8.1% (Mossey et al., 2009). No consistent seasonal patterns in birth prevalence of either defect have been recorded (Fraser & Gwyn, 1998).

Oral clefts have one of the highest rates of familial recurrence among all birth defects (Lie et al., 2001). The relative risk for siblings (λ_s), defined as the prevalence in siblings of an affected individual divided by the normal population prevalence, is 30 to 40 times higher; and there is a 2–5% increased risk for first degree relatives of affected individuals (Lidral & Moreno, 2005; Grosen et al., 2010). In a large cohort-based study, the relative recurrence risk was 32 for CL/P and 56 for CP among first degree relatives, suggesting a larger genetic component for CP. The risk of clefts was similar among the children of affected fathers, the children of affected mothers, and the full siblings of affected cases. It was suggested that autosomal fetal genes make the major contribution to risk of recurrence, with little additional contribution from heritable aspects of the maternal phenotype (Sivertsen et al., 2008b). More important, anatomical severity does have an effect on recurrence in first degree relatives and the type of cleft is predictive of the recurrence type. Individuals affected by the most severe form – bilateral CLP – have a significantly higher recurrence risk among both offspring and siblings compared to unilateral CLP. Recurrence risks for siblings in families with one proband have shown a consistent pattern of recurrence specificity – CLP families have a recurrence of CLP in ≥95% of cleft recurrences and CP families have CP in ~95% of cleft recurrences (Grosen et al., 2010).

1.1.3. Gene discovery in nonsyndromic clefts

The Online Mendelian Inheritance in Man catalog (OMIM) lists more than 400 syndromes known to have clefts of the lip and/or palate as an associated feature. The proportion of orofacial clefts associated with specific syndromes is between 5% and 7% (Tolarova & Cervenka, 1998). Genetic causes of clefting also include chromosomal rearrangements, genetic susceptibility to environmental and teratogenic exposures, and complex genetic contributions of multiple genes (Lidral & Murray, 2004). It has been estimated that around 30% of all CL/P cases and around 50% of CP cases occur in the context of an underlying malformation syndrome (Murray, 2002). In a report of almost 4000 patients with cleft palate from European populations, 55% of cases were classified as isolated, 18% were recorded in association with other congenital anomalies, and 27% were described as syndromic cases (Calzolari et al., 2004). In a report of more than 5000 patients with cleft lip with or without cleft palate, 71% of cases were isolated and 29% were seen in association with other anomalies (Calzolari et al., 2007).

Nonsyndromic cases of both CL/P and CP are considered to have a multifactorial etiology, which involves both genetic and environmental factors. Evidence for a genetic component in CL/P has been obtained from studies of familial recurrence, which indicate that the relative risk for siblings is 30–40 times higher than the average population risk (Mangold et al., 2009). A segregation analysis of familial recurrence patterns has estimated that 3–14 multiplicatively interacting genes may be involved the etiology of non-syndromic CL/P (Schliekelmann & Slatkin, 2002). The high familial aggregation rates and greatly increased concordance rates in monozygotic twin pairs than in dizygotic pairs provide the evidence for a strong genetic component in CL/P (Stanier & Moore, 2004). Predominance of left-sided clefting and the male excess among CL/P cases also suggest the importance of genetic susceptibility (Mossey & Little, 2002). The lack of complete concordance in monozygotic twins illustrates the importance of environmental factors in the etiology of CL/P (Cobourne, 2004).

It is generally accepted that cleft lip with or without cleft palate (CL/P) and cleft palate (CP) have distinct etiology on the basis of different embryonic timing and epidemiology. However, recent epidemiologic data suggest that cleft lip only may have unique etiologic features, including strong genetic associations, whereas some individuals with cleft palate only show evidence of subclinical cleft lip (Harville et al., 2005; Rahimov et al., 2008; Weinberg et al., 2008). Nevertheless, this broad division of anatomical defects is consistent with the distinct developmental origins of the lip/primary palate versus the secondary palate. Furthermore, separate cellular and genetic etiologies for CL/P and CP are consistent with the general observation that these two conditions do not segregate in the same pedigree, although exceptions – referred as “mixed clefting” – have been reported for families with etiologic mutations in specific genes (Dixon et al., 2011). Mixed clefting disorders suggest that identical

mechanisms cause these two forms, which previously had been separated based on embryologic and genetic evidence (Murray, 2002). Examples of mixed clefting disorders include *IRF6* (van der Woude syndrome) (Kondo et al., 2002); *FGFR1* (Kallmann syndrome) (Dodé et al., 2003), *TP63* (EEC syndrome) (Barrow et al., 2002) and *MSX1* (orofacial clefts and hereditary tooth agenesis) (van den Boogaard et al., 2000). The presence of dental anomalies in some individuals who have mutations in each of these genes suggests that the same pathways are common to tooth development (Murray & Schutte, 2004). In addition, CP may also be caused by mechanisms that interfere with palatal shelf elevation, such as micrognathia, which contribute to the genetic heterogeneity for CP and are unlikely to be causal for CL/P (Lidral & Murray, 2004). In an ENU-induced mouse model of nonsyndromic cleft palate that is caused by a mutation in the zinc finger transcription factor Prdm16, encoding a transcriptional cofactor that regulates TGFβ signaling, the cleft palate appears to be the result of micrognathia and failed palate shelf elevation due to physical obstruction by the tongue, resembling human Pierre Robin sequence (PRS)-like cleft of the secondary palate (Bjork et al., 2010).

Variants in genes linked to syndromic forms of orofacial clefts that have a Mendelian mode of inheritance can also produce phenocopies of isolated clefts, and the latest data from mouse and human studies have identified several genes known to underlie Mendelian syndromic forms of clefts as also playing a role in the etiology of nonsyndromic clefts. These include *IRF6* (Zucchero et al., 2004), *MSX1* (Jezewski et al., 2003), *FGFR1* (Dodé et al., 2003), *PVRL1* (CLPED1 – Margarita Island ectodermal dysplasia) (Sözen et al., 2001), *PTCH1* (Gorlin syndrome) (Kimonis et al., 1997), and *TBX22* (X-linked cleft palate and ankyloglossia) (Marçano et al., 2004). *SATB2* has been identified as an important gene in development of the human secondary palate in a study of rare balanced chromosomal rearrangements associated with isolated CP (Fitzpatrick et al., 2003).

To date, genetic approaches to nonsyndromic clefts have included: linkage analysis using large, multiplex families or smaller but inbred families, or analysis of affected sibpairs; association studies using case-control samples or case-parent trios; direct sequencing of DNA samples from affected individuals; and identification of chromosomal anomalies or microdeletions in cases. These methods can be applied to candidate genes or genome-wide strategies can be used, both approaches having advantages and disadvantages, depending mostly on the underlying genetic mechanisms of the disease, as well as the cost-effectiveness and limitations in technology. Most studies of nonsyndromic clefts to date have focused on CL/P rather than isolated cleft palate, possibly relying on the larger numbers of available cases, easier ascertainment and less pronounced impact from confounding syndromes (Dixon et al., 2011).

Genome-wide linkage studies have generated several chromosomal candidate regions for CL/P. Significant linkage has been reported for 16 loci in a metaanalysis of 13 genome-wide linkage scans in CL/P families: 1p12-q13, 1q32, 2q32-q35, 3p25, 6p23, 6q23-q25, 7p12, 8p21, 8q23, 9q21, 12p11, 14q21-

q24, 15q15, 17q21, 18q21, and 20q13 (Marazita et al., 2004). The loci on 1p-q, 7p, 8p, 9q, 14q, 20q corresponded with the loci of nominal significance identified in CL/P families of Central European descent, where suggestive evidence of linkage was also obtained for the loci 4q21-q26 and 1p31-p21, with the chromosome 1 locus showing a male-specific genetic effect (Mangold et al., 2009). Other loci for which significant linkage have also been reported, including 13q33.1-q34 (Radhakrishna et al., 2006), 18q21.1 (Beiraghi et al., 2007), and 19p13.12-q12 (Vieira et al., 2008a), but these three were not replicated (Mangold et al., 2009). Moreover, the 9q21 region surrounding the *FOXE1* gene reached genome-wide levels of significance with subsequent fine-mapping and replication (Marazita et al., 2009; Moreno et al., 2009). A single genome-wide linkage study of CP in Finnish multiplex families produced suggestive evidence of linkage for three chromosomal loci and no obvious overlap was observed with regions implicated in CL/P (Koillinen et al., 2005).

Various genetic polymorphisms have been investigated in population-based association studies using either family-based or case-control study design. The choice of candidate genes has traditionally relied on analyses of gene expression patterns during facial development, cleft phenotype in transgenic or knockout mouse models, association with syndromic forms of clefting, previous positive findings in humans, cytogenetic location adjacent to chromosomal anomalies associated with orofacial cleft phenotypes, and role in xenobiotic or nutritional pathways to provide biological plausibility for the association. Some gene products studied are growth factors (e.g., TGFA, TGFβ3), transcription factors (e.g., IRF6, MSX1, TBX22), or factors that play a part in xenobiotic metabolism (e.g., CYP1A1, GSTM1), nutrient metabolism (e.g., MTHFR, RARA – retinoic acid receptor α), or immune response (e.g., IRF6, PVRL1) (Mossey et al., 2009). As with candidate gene studies of many complex diseases, replication is not common, with only *IRF6* yielding consistent evidence of association with nonsyndromic CL/P and this finding has been replicated in many different populations and ethnic groups (Zucchero et al., 2004; Park et al., 2007; Birnbaum et al., 2009b; Jugessur et al., 2009). This inconsistency indicates the challenges in searching associations with a relatively rare phenotype such as nonsyndromic clefting.

Resequencing studies have identified several private missense mutations (i.e. those seen only in one family) as rare causes of nonsyndromic orofacial clefts, including *TGFB3* (Lidral et al., 1998); *MSX1* (Jezewski et al., 2003; Vieira et al., 2005); *FOXE1*, *GLI2, JAG2, LHX8, MSX2, SKI, SPRY2*, and *TBX10* (Vieira et al., 2005); *PTCH1* (Mansilla et al., 2006); *PVR* and *PVRL2* (Warrington et al., 2006); *RYK* (Watanabe et al., 2006); *FGF*s and *FGFRs* (Riley et al., 2007a; Riley et al., 2007b), and *TBX22* (Marçano et al., 2004). The missense mutations reported in the candidate genes listed above do not clearly segregate in the families and variable impact on gene expression levels during embryonic development and incomplete penetrance have been suggested (Vieira et al., 2008b). However, several of the reported mutations are conserved in other mammals, may disrupt exonic splicing enhancer sequences, and were not found in between 400 to 2000 control

chromosomes (Vieira et al., 2005). Missense mutations could affect expression of the genes during embryonic development, and consequently, cause inability to inhibit or activate transcription of target genes, which leads to lack of growth due to insufficient cell proliferation and/or differentiation, and finally, lack of cell adhesion and/or excess apoptosis leading to failure of fusion. Nonetheless, definitive evidence regarding the biological consequences of the described mutations is yet to be unraveled (Vieira et al., 2008b).

Several microdeletions have been recently identified in CL/P families. Deletions involving *CYP1B1*, *FGF10*, *SP8*, *SUMO1*, *TBX1*, *TFAP2A*, and *UGT7A1* have been confirmed, including both *de novo* and familial cases, and deletions of *SUMO1*, *TBX1*, and *TFAP2A* were proposed to be etiologic. These deletions suggest the potential roles of genes or regulatory elements within deleted regions in the etiology of clefting (Shi et al., 2009). In addition, analysis of chromosomal anomalies in patients has proven to be a productive route for the identification or confirmation of CL/P loci, with recent successes for fibroblast growth factor receptor 2 (*FGFR2*) and *SUMO1* (Osoegawa et al., 2008; Alkuraya et al., 2006).

Whole-exome sequencing has been successful in identifying causative genetic variants for Mendelian traits, including Miller syndrome (Ng et al., 2010a) and Kabuki syndrome (Ng et al., 2010b) that can both include cleft palate, but is as yet to be efficacious for complex and heterogeneous traits such as nonsyndromic clefts.

Data from direct sequencing, linkage and association studies have affirmed *BMP4* (Suzuki et al., 2009), *FGFR2* (Riley et al., 2007a; Osoegawa et al., 2008), *FOXE1* (Moreno et al., 2009; Marazita et al., 2009), *MSX1* (Vieira et al., 2003; Suzuki et al., 2004), and *MYH9* (Martinelli et al., 2007; Birnbaum et al., 2009c) as genes that seem likely (at least one study with compelling data and other supportive studies) to be involved in clefts of the lip and/or palate, whereas *CRISPLD2* (Chiquet et al., 2007), *FGF8* (Riley et al., 2007a), *GSTT1* (Shi et al., 2007), *MTHFR* (Mills et al., 2008), *PDGFC* (Choi et al., 2009), *PVRL1* (Avila et al., 2006; Sözen et al., 2009), *SUMO1* (Shi et al., 2009; Carter et al., 2010), *TGFA* (Vieira et al., 2006; Carter et al., 2010), and *TGFB3* (Vieira et al., 2003; Suazo et al., 2010) have been characterized as intensively studied (multiple studies, no consensus or convincing meta-analysis) clefting genes (Dixon et al., 2011). However, before GWA studies were introduced in craniofacial genetics, approximately 25% of the overall contribution of genetic variants causing CL/P was estimated to be identified, including private mutations and microdeletions identified in genes potentially involved in clefting (Vieira, 2008b; Shi et al., 2009).

The first attempts to identify genetic determinants of nonsyndromic cleft palate have been inconclusive or conflicting, particularly in the earlier association studies performed in different populations, with only a few candidate loci being implicated in CP, including *TGFA*, *MSX1*, and *TGFB3* (Shiang et al., 1993; Shaw et al., 1996, Maestri et al., 1997; Lidral et al., 1998; Mitchell et al., 2001; Beaty et al., 2002; Hecht et al., 2002). This inconsistency

could have been caused mostly by inadequate power and genetic heterogeneity. However, recent years have witnessed considerable success in the mapping of loci for nonsyndromic CP, with additional genes to include *TBX22*, *TCOF1*, *FOXE1*, and *SUMO1* (Marçano et al., 2004; Pauws et al., 2009; Sull et al., 2008; Moreno et al., 2009; Carter et al., 2010) among other genes contributing increased risk for CP.

Genome-wide association studies (GWAS) have successfully identified hundreds of genetic risk factors predisposing individuals to many common complex diseases. Likewise, findings from recent GWA studies have significantly improved our understanding of genes and pathways that contribute to the etiology of nonsyndromic CL/P. The first GWAS conducted in German case-control sample reported a major susceptibility locus (rs987525) on chromosome 8q24.21 and confirmed the impact of *IRF6*, which had previously been identified in candidate gene studies (Birnbaum et al., 2009a). This association was independently confirmed in a sample of European-American cases and controls (Grant et al., 2009). Two additional susceptibility loci that achieved genome-wide significance were subsequently identified at chromosomes 10q25.3 (closest gene *VAX1* – ventral anterior homeobox 1) and 17q22 (closest gene *NOG* – noggin) using an extended German GWAS sample and a replication sample of mixed European origin (Mangold et al., 2010). The most recent GWAS which included case-parent trios from multiple populations confirmed the *IRF6* findings and replicated the chromosome 8q24 and 10q25 associations (Beaty et al., 2010). Interestingly, in this GENEVA Cleft Consortium study, the level of statistical significance for rs987525 was much higher among case-parent trios of European ancestry than among those of Asian ancestry, whereas the evidence for linkage and association for markers in *IRF6* was much stronger in trios of Asian ancestry. The GENEVA study identified two additional loci (near *MAFB* on 1q22.1 and *ABCA4* on 20q12) not previously associated with CL/P that attained genome-wide significance with stronger signals in Asian compared to European populations (Beaty et al., 2010). These findings suggest that there are multiple genetic variants influencing risk of CL/P and also that some of these genes may be differentially tagged by polymorphic markers in a population-specific manner, where some putative causal genes or loci have been identified and confirmed in most populations (e.g., *IRF6*), whereas others (e.g., 8q24, *MAFB* and *ABCA4*) seem to be more population-specific, which could reflect variable coverage by available marker panels or true allelic heterogeneity (Dixon et al., 2011).

1.1.4. Environmental and lifestyle risk factors

Epidemiological and experimental data suggest that environmental risk factors might be important in the development of cleft lip and palate, and maternal exposure to tobacco smoke, alcohol, poor nutrition, occupational hazards, infection, medicinal drugs, and teratogens in early pregnancy have all been investigated (Mossey et al., 2009). The relatively high familial recurrence rates among orofacial clefts are likely not only because of shared genetic factors but also as a result of shared environmental factors (Hayes et al., 2002). Large prospective cohort studies are required to be effective in the identification of environmental triggers contributing to clefting and studies of gene-environment interaction. Moreover, the identification of environmental risks, particularly if they can be personalized with genetic covariates, provides the best short-term opportunities for prevention (Dixon et al., 2011).

Maternal smoking during pregnancy has been linked consistently with increased risk of having a child with orofacial cleft (Little et al., 2002; Shi et al., 2007; Shi et al., 2008). This association might be underestimated because passive exposure to smoke has not been assessed in most studies (Mossey et al., 2009). Population attributable risk calculations suggest that maternal smoking contributes to 4% of the total cleft lip and palate cases and 12% of bilateral cleft lip and palate cases (Honein et al., 2007). Maternal smoking has been associated with increased risk of both CL/P and isolated CP, and meta-analysis on the effects of smoking indicates a moderately increased relative risk for having CL/P of 1.34 and for having CP of 1.22 among offspring of mothers who smoke (Little et al., 2002).

Maternal alcohol consumption has also been suggested as a risk factor for orofacial clefts, but the evidence has been more inconsistent, with positive associations reported in some studies (Chevrier et al., 2005; Bille et al., 2007) but not others (Meyer et al., 2003; Romitti et al., 2007). A recent study suggests that 'binge' drinking patterns (high doses of alcohol in short periods of time) increase risk of infant oral clefts (DeRoo et al., 2008). Social and dietary contexts of alcohol consumption are varied and complex and can include modifying or confounding effects of nutrition, smoking, or drug use (Mossey et al., 2009).

Maternal nutrition has also been implicated as one of the environmental factors that may have a role in clefting. In most studies, maternal use of multivitamin supplements in early pregnancy has been linked to decreased risk of orofacial clefts; in a meta-analysis, multivitamin use was associated with a 25% reduction in birth prevalence of orofacial clefts (Johnson & Little, 2008). Folate deficiency has also been suggested to influence risk of CL/P, based on both observational studies and interventional trials using folate supplementation to prevent recurrences of CL/P in families (Wehby & Murray, 2010). Folate antagonists are associated with increased risk of orofacial clefts in people (Hernandez-Diaz et al., 2000). Folic acid supplements appear to reduce the risk of a baby having cleft lip and palate by about a third, but with no apparent effects on CP (Wilcox et al., 2007). Findings of case-control studies focusing

on multivitamin supplements containing folic acid, maternal dietary folate intake, and plasma folate have been inconsistent (van Rooij et al., 2004; Badovinac et al., 2007; Little et al., 2008a,b). Furthermore, food fortification programs using folic acid have shown evidence suggesting a decline in the birth prevalence rates of clefting in some (Canfield et al., 2005; Yazdy et al., 2007), but not all studies (López-Camelo et al., 2010).

Raised serum concentrations of homocysteine in mothers of infants with both CL/P and CP have been reported (Wong et al., 1999). Moreover, vitamin B6 (pyridoxine) acts as a cofactor in homocysteine metabolism and biomarkers of vitamin B6 deficiency were associated with increased risk of orofacial clefts (Wong et al., 1999; Munger et al., 2004).

Zinc is important in fetal development, and lower concentrations of zinc in erythrocytes were observed in mothers of children with both CL/P and CP compared to mothers of children without clefts (Krapels et al., 2004). Likewise, zinc deficiency may increase risk of oral clefts in populations in which zinc status is highly compromised (Munger et al., 2009). There are also some data to support roles for vitamin A (Rothman et al., 1995) and cholesterol deficiency (Porter, 2006) in facial clefting.

In addition, rare exposures to some specific teratogens – including anti-convulsants (phenytoin, valproic acid), thalidomide, environmental estrogens or dioxins, ionizing radiation, retinoic acid – can cause orofacial clefts. The importance of such exposures is that they can suggest metabolic pathways whose disruption may play a role in the development of CL/P (Murray, 2002). Positive associations with maternal corticosteroid use in pregnancy have been reported (Park-Wyllie et al., 2000). Excessive vitamin A is well-known to be teratogenic (Rothman et al., 1995). Interestingly, an increased total vitamin A intake from food and supplements was shown to reduce the risk of cleft palate alone by 53% (Johansen et al., 2008). Anticonvulsant drugs increase risk of these birth defects (Shaw et al., 1995). Recently, evidence of association with cleft palate has been reported for valproic acid therapy (Jentink et al., 2010).

Besides nutrients and toxins, other environmental exposures could play a part in development of orofacial clefts. Maternal diabetes is associated with an increased risk of orofacial clefts (Spilson et al., 2001). In addition, maternal obesity in early pregnancy has been reported to be associated with an increased risk to have orofacial clefts in the offspring (Cedergren & Källén, 2005). Maternal occupational exposure to organic solvents and parental exposure to agricultural chemicals have been associated inconsistently with both CL/P and CP (Shaw et al., 2003; Garcia, 1998). Interferon regulatory transcription factors are activated after viral infection and association of *IRF6* with clefts raises the possibility that viral infection in the first trimester of pregnancy may enhance risk of a cleft (Acs et al., 2005). Nevertheless, there is no consensus yet on the harmful effects of environmental factors on orofacial clefting, and prospective studies in large cohorts may be required to determine these effects taking into account that many exposures apparently have both identifiable and un-identifiable coincident risks (Dixon et al., 2011).

1.1.5. Gene-environment interaction

Investigation of gene-environment interaction is important for better estimation of the main effects of genes and/or likely impact of environmental factors, and to improve our understanding of causal mechanisms and pathogenesis. Several studies have investigated many potential interactions of a range of common environmental factors, such as cigarette smoking, alcohol intake, multivitamin/ folic acid supplementation and the use of medication.

Maternal smoking and folic acid intake are the two main factors that appear to modify genetic risks for cleft lip and palate. The increased risk resulting from exposure to maternal smoking during the periconceptual period raises the possibility that deficiencies in detoxification pathways and genes in certain metabolic pathways may have a role in the development of CL/P (Dixon et al., 2011). Interaction between maternal smoking and fetal inheritance of a *GSTT1*-null deletion was shown to be significant in two independent CL/P samples (Shi et al., 2007). In a meta-analysis, interaction between maternal smoking and the infant's genotype at the *Taq1* marker in *TGFA* was reported to increase risk for CP, whereas among nonsmoking mothers such increased risk for CP was not observed. *TGFA* genotype in offspring did not increase risk to have CL/P, regardless of maternal smoking status (Zeiger et al., 2005). Moreover, markers in the *NAT1* and *NOS3* genes appear to influence risk of CL/P in the presence of maternal smoking (Lammer et al., 2004; Zhu et al., 2009). Smoking has also been recently associated with a joint risk with variants in *IRF6,* and the same study reported interactions between multivitamins and *IRF6* variants (Wu et al., 2010). Investigations with polymorphisms in genes involved in the phases I (*CYP1A1* and *EPHX1*) and II (*GSTM1*, *GSTT1*, and *NAT2*) detoxification pathways considered as genetic modifiers of smoking effects remain preliminary (Hartsfield et al., 2001; van Rooij et al., 2001; Lammer et al., 2004).

The observation that drinking high doses of alcohol in short periods of time will increase risk of orofacial clefts has been supported by association with variation in the alcohol dehydrogenase *ADH1C* gene (Boyles et al., 2010).

Genetic variants in *TGFA*, *TGFB3* and *MSX1* genes have been investigated for interactions with environmental risk factors such as smoking, alcohol consumption and vitamin supplements (Shaw et al., 1998; Christensen et al., 1999; Romitti et al., 1999; Mitchell et al., 2001; Beaty et al., 2002). In addition, interactions between polymorphisms in *RARA* and maternal vitamin A intake (Mitchell et al., 2003), polymorphisms in genes affecting folate and vitamin A metabolism (e.g., *MTHFR*, *MTR*, *SLC19A1*, *CBS*, *FOLH1*, *TGFBR2*) and maternal folate intake (Jugessur et al., 2003; Chevrier et al., 2007; Little et al., 2008a; Boyles et al., 2009), and polymorphisms in N-acetyltransferase genes and maternal medicinal drug use (van Rooij et al., 2002) have been studied. However, findings from these studies have been inconclusive. The possible reasons for uncertainty include: low statistical power to detect or exclude interaction; differences between studies in the individuals who have been genotyped (e.g., mother alone or with infant); research confined to populations in a few industrialized countries; and non-existent replication in other studies (Mossey et al., 2009).

1.2. Large-scale association studies of complex traits

1.2.1. Consensus, challenges and considerations

Genetic association studies analyze phenotypes (discrete or continuous traits) and genotypes among sample of individuals in order to identify relationships between DNA sequence variation and disease predisposition. Most common human diseases have a polygenic pattern of inheritance where multiple combinations of low-penetrance DNA sequence variants at many genetic loci are interacting with environmental exposures and lifestyle risk factors. A major drawback lies in the analysis of patients with heterogeneous etiology, since this decreases the chances of finding genuine gene-phenotype correlations, and an interplay of phenotypic and genetic heterogeneity may severely complicate the discovery of the true disease-causing variants. Numerous studies of non-syndromic clefts have favored a multifactorial model of inheritance in which genetic risk factors of modest individual effect size interact with environmental covariates, and therefore, identifying the key genes in human CL/P represents a major challenge.

The past years have witnessed substantial advances in the identification of low-penetrance, high-frequency susceptibility variants in common complex diseases, demonstrating the efficiency of case-control association mapping using sufficiently large sample sizes (thousands of individuals) and sufficiently dense set of SNP markers (from several hundreds up to hundreds of thousands of markers in GWA studies), and this approach can be applied to candidate genes or genome-wide strategies can be used. Hypothesis-driven candidate gene association studies have been a traditional approach to dissect the genetic basis of complex traits (Jorgensen et al., 2009). The most comprehensive candidate gene study in genetics of orofacial clefts included 357 genes related to cranio-facial development, being selected from published linkage and association studies on clefts, gene-knockout experiments in mice, extrapolations from the studies of syndromic forms of clefting, studies of chromosomal rearrangements in humans, and gene expression analyses in human and mouse embryonic tissues (Jugessur et al., 2009).

Genome-wide association studies (GWAS), in which several hundred thousands to more than a million SNPs are assayed in thousands of individuals, represent an important advance compared to candidate gene studies as a powerful and advantageous tool for detecting genetic variations throughout the human genome without prior knowledge of genes or underlying biological pathways potentially linked to the complex disease phenotype. They have been facilitated by the development of commercial microarrays that capture most, although not all, common variation in the genome. GWAS are also an important step beyond family-based linkage studies, in which inheritance patterns are related to several hundreds to thousands of genomic markers (Manolio et al., 2009).

GWA studies have successfully identified more than 600 genetic risk variants predisposing individuals to many common complex diseases (Hindorff

et al., 2009). Most common DNA variants individually or in combination confer relatively small increments in risk (1.1–1.5-fold) and explain only a small proportion of heritability – the component of phenotypic variance in a population attributable to additive genetic factors (Altshuler et al., 2008; Manolio et al., 2009). Many explanations for this missing heritability have been suggested, including much larger numbers of variants of smaller effect yet to be found; rarer variants (possibly with substantial effect sizes) that are poorly detected by available genotyping arrays that focus on variants present in 5% or more of the population; structural variants poorly captured by existing arrays; low power to detect gene-gene interactions; and inadequate accounting for shared environment among relatives (Manolio et al., 2009). This tenders a challenge for genetic studies of individual risk alleles because achieving sufficient statistical power in a genetic association study requires large case-control samples comprising of thousands individuals. The problem is amplified in patients of diverse ancestry and for clinically relevant endophenotypes within a given disease because creating subsets of patients further reduces sample size.

In a meta-analysis using 183,727 individuals, hundreds of genetic variants in at least 180 loci were shown to influence adult height, a classic polygenic trait with an estimated heritability of about 80%. These variants explain only approximately 10% of the phenotypic variation in height (Lango Allen et al., 2010). Similarly, an extended GWAS for plasma lipids in 100,000 individuals has identified 95 loci that collectively explain 10–12% of the total variance, representing ~25–30% of the genetic variance (Teslovich et al., 2010). In the case of Crohn's disease, over 30 significantly associated genomic markers account for less than 10% of the cumulative genetic variance (Barrett et al., 2008). There is a strong expectation that additional variance and biological mechanisms will be explained using complementary approaches that capture variants not examined in current GWA studies, such as lower frequency variants and short insertion-deletion polymorphisms. The large number of described loci reveals patterns with important implications for genetic studies of common human diseases and traits. Although disease-associated variants occur more frequently in protein-coding regions than expected from their representation on genotyping microarrays, in which over-representation of common and functional variants may introduce analytical biases, the vast majority (>80%) of associated variants fall outside coding regions, emphasizing the importance of including both coding and non-coding regions in the search for disease-associated variants (Hindorff et al., 2009).

GWAS will probably remain an efficient way of investigating the remaining heritability, because their association signals may well define the genomic regions where rare variants, structural variants, and other forms of underlying variation are likely to cluster. Near-term approaches for finding missing heritability which seem to have wide agreement include: use of expanded reference panels of genomic variation such as 1000 Genomes to enhance coverage of existing and produce arrays with even more comprehensive coverage for future GWAS, and to facilitate the investigation of the lower frequency spectrum

without the need for *de novo* sequencing; expanding studies to more diverse diseases (including less common diseases) and including measures of environmental exposures and more precisely ascertained phenotypes (as needed to reduce heterogeneity or explore pleiotropic effects); targeted or whole-genome sequencing in people with extreme phenotypes; mining of existing GWAS for associations with structural variants and evidence of gene-gene interactions; improved methods for detection of CNVs and other structural variants; and expansion of sample sizes for numerous complex diseases through larger individual studies and meta-analyses, including samples of non-European ancestry (Manolio et al., 2009).

Most published candidate gene and genome-wide association studies have featured case-control designs, which raises challenging methodological and study design issues related to the optimal selection of both case and control samples, selection of markers, study power, replication and population heterogeneity.

Marker selection. In candidate gene studies, haplotype-based methods represent the most recent approach to capture most of the common allelic variation in the regions of interest by applying computational approaches that improve the detection of associations that are attributable to variants that have not themselves been directly typed (Carlson et al., 2004; de Bakker et al., 2005). These methods are based on assumption that correlation between nearby variants (LD) allows to select the minimum number of informative tagSNPs that serve as proxies for neighboring variants, thereby substantially reducing the genotyping costs (Johnson et al., 2001). Traditionally, SNP ascertainment is performed on data from the reference populations in the International HapMap Consortium (International HapMap Consortium, 2005). In situations when haplotype-based analyses reveal evidence for association that exceeds that of any directly typed SNP in the vicinity, one can invoke either an effect that is directly attributable to the haplotype (that is, independent causal *cis* effects at multiple SNPs) or the explanation that the haplotype tags more efficiently than any individual genotyped SNP, an as yet untyped etiological variant (McCarthy et al., 2008). Importantly, the use of such methods is not restricted to samples drawn from HapMap reference populations (de Bakker et al., 2006).

Power and sample size. Power for studies of allelic association will depend primarily upon sample size, the effect size of the susceptibility locus, the strength of LD with a marker, and the frequencies of susceptibility and marker alleles (Zondervan & Cardon, 2004). In accordance with the results from candidate gene studies of complex traits, the initial wave of GWA studies has shown that, with rare exceptions, the effect sizes resulting from common SNP associations are modest, and that sample sizes in the thousands are essential because of low power to detect associations (McCarthy et al., 2008). Moreover, a single locus can harbor both common variants of weak effect and rare variants of large effect (Altshuler et al., 2008). Although family-based association methods provide a robust strategy for dealing with stratification as a merit in comparison with case-control study design, this typically results in the cost of reduced power (Laird et al., 2006).

Replication and heterogeneity. An appreciation of power and sample size is crucial to the design and interpretation of appropriate replication studies. Underpowered studies cannot confirm or refute the original finding, and may generate misleading inferences when considered in isolation. However, combinations of such studies might be of value provided that all suitable studies have been included (McCarthy et al., 2008). Calculations made about the required sample size for subsequent replication studies have to account for the so-called 'winner's curse' effect, whereby the original study yielding associations that pass desired thresholds of statistical significance will typically overestimate the true effect size (Zollner & Pritchard, 2007). The predictive ability of the described associations and the estimate of the risk variance explained by the associations are also inflated, and the magnitude of the winner's curse is inversely related to the power of the study. For small effects, even large meta-analyses could be largely underpowered and emerging associations could be considerably inflated. For rare variants, the power can be <1%, and therefore associations that are discovered for rare variants will have extremely inflated effects and the true effect size should await further replication (Ioannidis et al., 2009).

If well-performed replication studies confirm the original findings, then the evidence in favor of true association is enhanced. When even the well-powered replication studies indicate that there is genuine divergence between the effect size estimates, then there are two possible explanations: either the original finding was wrong, or the discrepancies between findings are attributable to some source of heterogeneity (Ioannidis et al., 2007). There are several potential causes of heterogeneity: it includes variable patterns of LD between the genotyped SNP and untyped causal alleles (although this is unlikely if the samples are of similar ancestry); differences in the distribution, frequency or effect size of the causal alleles at a given locus (possibly due to drift or selection, or differences in case ascertainment); and the impact of non-additive interactions with other genetic variants or environmental exposures (McCarthy et al., 2008).

Association studies can be confounded by population stratification, in which case misleading results can arise if individuals selected as disease cases have different ancestry than healthy controls (Price et al., 2008). In this context, population stratification refers to the significant differences in allele frequencies between cases and controls caused by the systematic ancestry differences across subpopulations within study population rather than by genuine association of gene(s) with the disease (Marchini et al., 2004). The propensity for latent population substructure (population stratification and cryptic relatedness) is related to the inflated type I error rate and generates spurious associations around variants that are informative for that substructure (Voight & Pritchard, 2005; Price et al., 2006). The importance of recognizing and adjusting for population structure is amplified when population controls are not closely matched to cases, but as long as cases and controls are well matched for broad ethnic background, and measures are taken to identify and exclude individuals whose GWA data reveal substantial differences in genetic background, the impact of residual substructure on type I error seems modest (McCarthy et al., 2008).

Population substructure can be explored and ascertained using a variety of algorithms that apply principal component analysis (Price et al., 2006) or non-hierarchical cluster analysis based on allele frequencies in individuals and groups (Pritchard et al., 2000). In contrast to other multi-locus adjustments such as genomic control method (Devlin et al., 2001), where correction for stratification is performed using an uniform overall inflation factor λ to adjust association statistics at each marker, these newer approaches take into account that some SNPs have large variations in allele frequencies across different ancestral populations. Multidimensional scaling (MDS) analysis based upon genome-wide identity-by-state (IBS) distances is also frequently used for population stratification visualization. With the recent availability of the vast amount of the genome-wide data, PCA and MDS methodologies have become increasingly popular in exploring the actual extent and relevance of genetic differences between populations because they are computationally less intensive and have higher discriminatory power than Bayesian analysis for closely related (e.g. European) populations (Li & Yu, 2008).

1.2.1. European genetic structure map

Differences in population genetic structure and substructure between cases and controls can miss real effects or lead to false positive association findings in large multicenter studies, which might inadvertently include some individuals with substantial but undetected levels of admixture (Marchini et al., 2004; Clayton et al., 2005). Disease prevalences often vary depending on geographic and ethnic origin (e.g. nonsyndromic CL/P), and allele frequencies of genetic variants (possibly) contributing to disease risk may differ widely in populations with different ancestries all over the world, and consequently, such variants may falsely appear to be related to disease. Therefore, it is important that ancestry differences – corresponding to ethnic groups – for study participants must be recognized prior to data analysis to disregard such erroneously positive associations with a trait.

Genotyping of hundreds of thousands autosomal markers in human genome has allowed the construction of genetic structure maps of populations that correlate partially with geographical maps. Since most association studies have focused on populations of European descent, the potential impact of European genetic substructure on association testing has particularly elicited interest. The studies in Americans of European descent have identified a clear gradient from northwest to southeast across Europe (Bauchet et al., 2007; Price et al., 2008; Tian et al., 2008). Additionally, the study discerning the ancestry of European Americans demonstrated that a small validated panel of ancestry-informative markers can reliably distinguish three clusters of individuals, which roughly correspond to northwest European, southeast European and Ashkenazi Jewish ancestry, permitting sufficient correction for most of the population stratification affecting genetic association studies (Price et al., 2008). These

findings were further extended in larger samples of individuals from multiple European populations (Lao et al., 2008; Novembre et al., 2008).

In a recent study aiming at the construction of a European-wide genetic map, Estonians, Latvians and Lithuanians were included for the first time among 3112 individuals comprising 19 different samples across 16 European countries (Nelis et al., 2009). Principal component analysis of more than 270,000 SNPs yielded a genetic structure map of Europe in which two first components of variation highlighted genetic diversity corresponding to a northwestern-southeastern gradient, and suggested that geographically adjacent populations overlap partly, forming a triangular structure with four distinct subgroups across Europe, with Central and Western European populations in its centre and Estonians, Latvians and Lithuanians clustering closely within one distinct region according to their geographic origin (Figure 4). These results allowed making conclusions that combining genotype data from neighboring populations is a relevant approach in large-scale association studies, and for example, data from the Baltic countries may be analyzed together with Western Russia and Poland because of their genetic similarity. Moreover, the results demonstrated that Estonian samples can be analyzed with most other European samples, except genetic isolates and southernmost Europeans, without significant loss of power (Nelis et al., 2009).

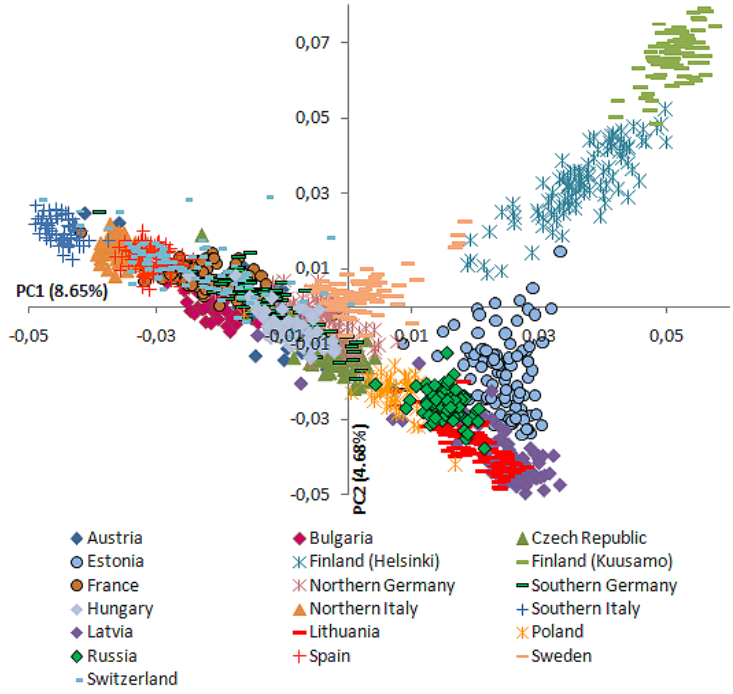

Figure 4. European genetic structure map as revealed by PC analysis performed on data from 273,464 SNPs (Nelis et al., 2009).

In summary, these studies provided further insights into European population genetic substructure and emphasize the importance of correction for stratification in determining genetic risk factors for complex diseases, and show that this approach can be applied for diminishing error rates in association testing of candidate genes and also in replication studies of genome-wide association scans.

2. MATERIALS AND METHODS

2.1. Study population

The entire study sample (Ref. I, II and III) consisted of 1010 individuals from Estonian, Latvian, and Lithuanian population. The subjects included 300 patients with nonsyndromic CL/P (66 patients with cleft lip and 234 patients with cleft lip and palate) and 104 patients with nonsyndromic cleft palate (CP). All 404 probands were clinically assessed by an experienced medical geneticist to identify anomalies suggestive of underlying syndromes known to be associated with any type of orofacial clefting. Patients with confirmed monogenic syndrome or chromosomal aberrations, associated malformations, and mental retardation were excluded from the study. The control sample was comprised of 606 unaffected unrelated individuals (268 males and 338 females) without a family history of clefting, collected as randomly selected population-based controls.

Ethical approvals for the study were obtained from the Ethics Review Committee on Human Research of the University of Tartu, the Central Medical Ethics Committee of Latvia, and the Lithuanian Bioethics Committee. All individuals gave written informed consent for participation in the study. In the case of patients who were under 18 years of age, consent was obtained from their parents.

Estonian patients (Ref. I, II and III)
The subjects included 153 patients (74 males and 79 females): 100 patients with CL/P (58 males and 42 females) and 53 patients with CP (16 males and 37 females). Patients were recruited from surgical clinics in North Estonia Medical Centre, Tallinn, and Tartu University Hospital.

Estonian patients (Ref. IV and V)
The subjects included 163 patients with nonsyndromic oral clefts: 105 patients with CL/P (62 males and 43 females) and 58 patients with cleft palate (15 males and 43 females). Additional patients were recruited from North Estonia Medical Centre, Tallinn.

Latvian and Lithuanian patients (Ref. II and III)
The Latvian subjects included 32 patients with CP (19 males and 13 females) and 108 patients with CL/P (62 males and 46 females). The Lithuanian subjects included 19 patients with CP (9 males and 10 females) and 92 patients with CL/P (59 males and 33 females). Latvian patients were recruited from the Riga Cleft Lip and Palate Centre, Institute of Stomatology, Riga Stradins University. Lithuanian patients were recruited from the Center for Medical Genetics, Vilnius University Hospital Santariškių Klinikos in collaboration with the largest orthodontic clinics in Lithuania.

Estonian controls (Ref. I, II and III)
Estonian controls (N=205) were selected from the Biobank of the Estonian Genome Center, University of Tartu (www.geenivaramu.ee), representing a random selection made proportionally from each of the 15 Estonian counties and with an equal male-to-female ratio.

Estonian controls (Ref. IV and V)
The population-based control sample was comprised of 254 unrelated individuals (Ref. V) and 1023 (including 254 individuals from Ref. V) unrelated individuals (Ref. IV) selected from the Biobank of the Estonian Genome Center, University of Tartu (www.geenivaramu.ee). These individuals represent a random selection made proportionally from each of the 15 Estonian counties with an equal male-to-female ratio.

Latvian and Lithuanian controls (Ref. I, II and III)
The Latvian control group consisted of 182 randomly selected individuals collected at the Latvian Biomedical Research and Study Center within the framework of the national project "Genome Database of Latvian Population". The Lithuanian control group consisted of 219 individuals selected from six ethnolinguistic groups of Lithuania with an equal male-to-female ratio.

Lithuanian patients and controls (Ref. IV and V)
The Lithuanian patients sample was comprised of 112 patients with CL/P. The population-based control sample consisted of 244 unrelated individuals selected from six ethnolinguistic groups with an equal male-to-female ratio.

2.2. Candidate genes and polymorphisms

40 candidate genes were selected on the basis of previously published findings from association and linkage studies on nonsyndromic orofacial clefts, gene expression patterns during craniofacial development, cleft phenotype in knockout or transgenic mouse models, genes that underlie Mendelian syndromic forms of clefting, and studies of chromosomal rearrangements associated with orofacial cleft phenotypes in humans.

The selected genes are encoding for a variety of molecules implicated in craniofacial morphogenesis: transcription factors (e.g. *IRF6, MSX1, TBX22, LHX8*), growth factors and their receptors (*FGF1, FGF2, FGFR1, TGFA, TGFB3*), polarizing signals (*BMP2, BMP4, SMAD2, SMAD4, WNT* genes) and cell adhesion molecules (*PVRL1, PVRL2*), and extracellular matrix molecules (*COL11A1, COL11A2, COL2A1, FN1, MMP2, MMP3, MMP9, MMP13, TIMP2*). Haplotype-tagging SNPs were selected to capture all the SNPs with minor allele frequencies (MAF) ≥ 0.05 and $r^2 \geq 0.8$ in the regions of interest based on the HapMap Phase II data (Release #21a, Jan 2007; http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B35/), using HapMap CEU as a reference population.

Multiple SNPs were selected for each gene, including 10–20 kb of both upstream and downstream genomic sequences. A list of selected genes and the number of genotyped SNPs (per gene) which were included in association analysis in Baltic clefting samples (both CP and CL/P) are shown in Table 1.

Table 1. Candidate genes & loci included in the Baltic clefting genetics study

Gene	Chromosome	Genotyped SNPs CP sample	Genotyped SNPs CLP sample
MTHFR	1p36.3	9	8
LHX8	1p31.1	9	9
COL11A1	1p21	42	42
SKI	1q22-q24	19	19
IRF6	1q32.3-q41	10	10
TGFA	2p13	36	35
FN1	2q34	27	27
MSX1	4p16.3-p16.1	15	15
FGF2	4q26-q27	18	18
FGF1	5q31	31	31
MSX2	5q34-q35	6	6
EDN1	6p24.1	15	15
COL11A2	6p21.3	19	19
FGFR1	8p11.2-p11.1	11	11
FOXE1	9q22	4	4
TBX10	11q13.2	10	10
MMP3	11q22.3	5	5
MMP13	11q22.3	19	19
PVRL1	11q23.3	17	17
COL2A1	12q13.11	32	32
BMP4	14q22-q23	4	4
TGFB3	14q24	8	8
JAG2	14q32	11	11
MMP25	16p13.3	6	6
MMP2	16q13-q21	20	20
CDH1	16q22.1	13	13
RARA	17q21	4	4
WNT3	17q21	16	16
WNT9B	17q21	11	11
TIMP2	17q25	25	25
'OFC11'	18q21[a]	25	25
BCL3	19q13.1-q13.2	3	2
PVRL2	19q13.2	12	12
CLPTM1	19q13.2-q13.3[b]	7	7
BMP2	20p12	25	25
MMP9	20q11.2-q13.1	6	5
TIMP3	22q12.3	36	36
TBX22	Xq21.1	5	5

[a] linkage locus (Beiraghi et al., 2007); includes *SMAD2* and *SMAD4* genes
[b] includes *APOC2* gene

In a GWAS replication study, six susceptibility loci no previously associated with CL/P (8q24.21, 17q22, 10q25.3, 13q31.1, 15q13.3, and 2p21) which were reported in a first GWAS conducted in a German and a mixed European sample (Birnbaum et al., 2009; Mangold et al., 2010) were investigated. The six most significantly associated SNPs (rs987525, rs227731, rs7078160, rs7590268, rs9574565, and rs1258763) in these loci were genotyped in CL/P patients and unrelated controls from Estonia and Lithuania.

2.3. Genotyping

Genomic DNA was extracted from peripheral blood lymphocytes according to standard high-salt extraction (Estonian sample) or phenol-chloroform methods (Latvian and Lithuanian samples) methods. SNP genotyping in a Baltic sample was performed according to the principles of an arrayed primer extension-based genotyping method (APEX-2). This method allows multiplex DNA amplification and detection of SNPs on microarrays via four-color single-base primer extension (Krjutškov et al., 2008).

In a GWAS replication study, genotyping in the Estonian and Lithuanian patients and controls was performed using genomic DNA that had been isolated from peripheral blood lymphocytes according to standard procedures. The Sequenom MALDI-TOF mass-spectrometer (MassArray® system) was used for genotyping and the data were analysed using the Spectrodesigner Software package (Sequenom^{TM}, San Diego, CA). Three distinct clusters were analyzed with the Sequenom^{TM} Typer Analyzer 4.0.1 software. Genotyping for SNP rs987525 was conducted in 1023 Estonian population-based controls using the Illumina HumanHap370CNV-Duo BeadChip according to the Infinium II protocol from Illumina (Illumina) in the Estonian Genome Center core facility (www.geenivaramu.ee).

2.4. Statistical analysis

All markers were tested for Hardy-Weinberg equilibrium in controls and affected individuals using a chi-square test for genotype proportions. LD measures were calculated and the haplotype blocks were defined using the confidence interval method, and haplotype frequencies were estimated with Haploview version 4.1 software. Association analyses were performed with SNPs which met the quality control (QC) criteria (MAF $\geq$ 1% in controls, genotype call rate > 95%, individual missingness < 0.10, HWE P > 0.001) adopted for the study. The alleles at each marker were tested for association separately for both clefting phenotypes: first, only individuals with CL/P were compared with controls; and, second, only individuals with CP were compared with controls. Allele frequency differences between cleft patients and control subjects were compared for each SNP by employing a standard χ^2-test with one

degree of freedom. Allelic odds ratios (ORs) and 95% confidence intervals (CIs) were estimated assuming a multiplicative model. The level of statistical significance was set at $\alpha = 0.05$ for nominal association. Haplotype-phenotype association tests were performed with the standard χ^2-test. Multiplicative interactions between two SNPs were tested using SNP × SNP epistasis analysis by assuming a general model. Statistical analyses were conducted using PLINK version 1.06 software (Purcell et al., 2007). Power analysis was performed with CaTS Power Calculator (http://www.sph.umich.edu/csg/abecasis/CaTS/) under a multiplicative model.

In a GWAS replication study, statistical analysis was performed using SAS software (version 9.1). The standard χ^2-test was used to test for deviations from Hardy-Weinberg equilibrium and the Cochran-Armitage trend test was performed to compare genotype distributions between cases and controls. Population attributable risk (PAR) was calculated for rs987525 as $(K - 1)/K$, where $K = c_2 \cdot \psi_2 + c_1 \cdot \psi_1 + c_0$, c_i is the population frequency of the i genotype, and ψ_i is the estimated genotype relative risk of the i genotype (Birnbaum et al., 2009).

3. RESULTS

3.1 Association analysis in Estonian clefting sample (Ref. I)

A total of 176 SNPs in 18 candidate genes related to orofacial clefting (*MTHFR, IRF6, SKI, MSX1, MSX2, EDN1, FOXE1, TBX10, PVRL1, LHX8, SPRY2, TGFB3, JAG2, RARA, CLPTM1, BCL3, PVRL2,* and *TBX22*) were analyzed in 153 patients with nonsyndromic oral clefts and 205 unrelated controls from the Estonian population as a pilot study within the framework of the Baltic clefting genetics project. The overall genotype call rate was 99.62% for the CL/P patients sample, 99.8% for the CP patients sample, and 99.72% for the control sample.

Twenty-six polymorphisms in 10 genes displayed nominal evidence of association with the CL/P phenotype in our study sample. Table 2 presents all 9 markers with allelic association test P-values below 0.01 (before correction for multiple testing) in the CL/P case-control sample.

The strongest association with CL/P was found for SNP rs6446693, which is located ~6 kb upstream of the muscle segment homeobox 1 (*MSX1*) gene, where the T allele was associated with an increased risk (OR = 1.934, 95% CI 1.366–2.738; $P = 1.82 \times 10^{-4}$). This association remained statistically significant after correcting for multiple testing (Bonferroni and FDR; $P_{corr} = 0.032$). The SNP rs1907998, located ~5 kb upstream of *MSX1*, was significantly associated with CL/P under the assumption of the dominant model of inheritance ($P = 8.53 \times 10^{-5}$).

Other markers of interest in the CL/P dataset included the *MTHFR, SKI*, *MSX1*, and *PVRL2* loci (Table 2). Among markers in or near the *MTHFR* and *SKI* genes, a variant allele had a protective effect, whereas the variant alleles of the *MSX1* and *PVRL2* markers were associated with an increased risk.

There was no evidence of a sex-specific component in the association between CL/P and the nine markers with the best P-values. The genotype distributions in male and female cases were similar for these SNPs ($P > 0.05$).

In the case-control analysis of the CP phenotype, eight common polymorphisms in five genes displayed nominal evidence of association in Estonian sample. Two SNPs showed evidence of an association with the CP phenotype: rs11624283 in the *JAG2* gene ($P = 0.0016$) and rs1106514 in the *MSX1* gene ($P = 0.0037$).

Haplotype analysis

Haplotype analysis was performed with the CL/P dataset only (cases + controls, N = 305). Case-control analysis was performed for haplotypes constructed from SNPs within linkage disequilibrium blocks identified in the *MSX1*, *MTHFR*, *IRF6*, *BCL3*, *TGFB3*, *PVRL2*, *EDN1*, *JAG2*, and *TBX10* genes.

Table 2. Most significant results from single-marker association analysis in the Estonian CL/P sample

Gene	Chr	SNP	Location	Alleles[a]	MAF		P	OR	95% CI
					cases	controls			
MTHFR	1	rs1476413	11774887	C/T	0.235	0.344	0.0062	0.586	0.399–0.861
		rs1801131	11777063	T/G	0.250	0.360	0.0067	0.594	0.406–0.867
		rs1994798	11777342	A/G	0.320	0.446	0.0028	0.584	0.409–0.833
		rs4846054	11791817	T/C	0.320	0.434	0.0070	0.614	0.430–0.877
SKI	1	rs1809822	2221455	A/C	0.040	0.107	0.0057	0.350	0.162–0.759
MSX1	4	rs6446693	4905981	C/T	0.542	0.379	0.000182	1.934	1.366–2.738
		rs1907998	4907480	A/G	0.450	0.337	0.0066	1.613	1.141–2.279
JAG2	14	rs1022431	104674460	C/A	0.125	0.063	0.0099	2.110	1.184–3.758
PVRL2	19	rs2075642	50069307	G/A	0.235	0.139	0.0031	1.902	1.237–2.925

[a] Major allele (in controls) is listed first

Table 3 presents the results from haplotype-based association analysis within an LD block at the *MTHFR* locus for all possible 2- to 5-SNP haplotypes with a frequency of > 5% among CL/P patients and with P ≤ 0.01. The best results, with P ≤ 0.01, for common (frequency > 5%) 2- to 4-SNP haplotypes in other five candidate loci are presented in Table 4.

Table 3. Case-control association analysis of haplotypes in the *MTHFR* gene

Haplo-type	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	Frequency		P
						cases	controls	
	rs1476413	rs1801131	rs1994798	rs4846054	rs17376328			
H1	C	T	A	T	G	0.639	0.531	0.0114
H1	*	T	A	T	G	0.660	0.541	0.0053
H1	*	*	A	T	G	0.660	0.544	0.0063
H2	T	G	G	C	*	0.215	0.324	0.0053
H1	C	T	A	*	*	0.659	0.536	0.0038
H2	T	G	G	*	*	0.214	0.326	0.0044
H1	*	T	A	T	*	0.660	0.541	0.0053
H2	*	G	G	C	*	0.235	0.342	0.0069
H1	*	*	*	T	G	0.680	0.567	0.0072
H1	C	T	*	*	*	0.730	0.629	0.0133
H2	T	G	*	*	*	0.215	0.331	0.0030
H1	*	*	T	G	*	0.660	0.544	0.0063
H2	*	*	G	C	*	0.300	0.424	0.0030
H1	*	T	A	*	*	0.680	0.546	0.0016
H2	*	G	G	*	*	0.250	0.352	0.0106
H1	C	*	A	*	*	0.658	0.537	0.0045
H2	T	*	G	*	*	0.213	0.327	0.0036
H1	C	*	*	T	*	0.659	0.554	0.0136
H2	T	*	*	C	*	0.214	0.330	0.0029
H1	*	T	*	T	*	0.664	0.553	0.0089
H2	*	G	*	C	*	0.234	0.346	0.0048

The strongest associations with CL/P in the *MTHFR* gene were found for common core (H1) haplotypes such as SNP 1-2-3 CTA (P = 0.0038), SNP 2-3 TA (P = 0.0016), and SNP 1-3 CA (P = 0.0045), which were all associated with higher risk of CL/P. The second most frequent (H2) haplotypes, such as SNP 1-2-3 TGG (P = 0.0044), SNP 1-2 TG (P = 0.0030), SNP 3-4 GC (P = 0.0030), SNP 1-3 TG (P = 0.0036), and SNP 1-4 TC (P = 0.0029), were associated with lower risk of CL/P.

The strongest haplotype-phenotype associations in other candidate loci were found for *MSX1* H2 TG, *PVRL2* H3 CA, *BCL3* H3 CATA, *EDN1* H2 CC, and *JAG2* H3 TA haplotypes, which were associated with higher risk of CL/P, whereas *MSX1* H1 CA haplotype, with the lowest P-value of 7×10^{-4}, was associated with a lower risk of CL/P. The P-values obtained from haplotype-based association analysis performed with markers in *MTHFR*, *BCL3, EDN1,* and *JAG2* were lower than analyses using individual SNPs.

Table 4. Case-control association analysis of haplotypes in five candidate genes

Haplo-type	SNP	SNP	SNP	SNP	Frequency		P
					cases	controls	
			BCL3				
	rs17728272	rs4803750	rs8100239	rs8103315			
H3	C	A	T	A	0.188	0.109	0.0071
H3	C	*	T	A	0.188	0.109	0.0071
			MSX1				
		rs6446693	rs1907998				
H1		C	A		0.444	0.590	7×10^{-4}
H2		T	G		0.431	0.308	0.0026
			EDN1				
		rs4714384	rs9471438				
H1		T	T		0.496	0.598	0.0179
H2		C	C		0.388	0.278	0.0062
			JAG2				
		rs2056860	rs1022431				
H3		T	A		0.125	0.061	0.0069
			PVRL2				
		rs519113	rs2075642				
H2		G	G		0.178	0.270	0.0125
H3		C	A		0.228	0.138	0.0054

The strongest evidence of gene-gene interactions from the SNP × SNP epistasis analysis of all pairwise combinations between 176 individual SNPs was found between two unlinked SNPs (rs4803766 and rs3745150) within the *PVRL2* gene region ($P = 4.56 \times 10^{-4}$), between SNP rs8103315 in the *BCL3* gene and SNP rs10807242 located ~10 kb upstream of the *EDN1* gene ($P = 4.75 \times 10^{-4}$), and between SNP rs17389541 located ~8 kb downstream of the *IRF6* gene and SNP rs884690 located ~9 kb upstream of the *MSX1* gene ($P = 7.67 \times 10^{-4}$). An interaction between *MSX1* and *BMP2* revealed a trend towards significance after Bonferroni correction ($P_{corr} = 0.080$).

3.2. Association analysis in Baltic cleft palate sample (Ref. II)

Association analysis of 591 tagSNPs was performed in 104 patients with nonsyndromic cleft palate and 606 unrelated controls from Estonian, Latvian and Lithuanian populations. The average genotype call rate for these SNPs was 99.25%.

Table 5 presents all markers with allelic association test P-values below 0.05 (before correction for multiple testing) in the CP case-control sample. Thirty-five polymorphisms in 17 genes displayed nominal evidence of association with CP, and 10 out of 35 SNPs had P-values less than 0.01.

The strongest association with CP was found for SNP rs17389541, which is located ~8 kb upstream of the interferon regulatory factor 6 (*IRF6*) gene, where the T allele was associated with higher risk (OR = 1.726, 95% CI 1.263-2.358; $P = 5.45 \times 10^{-4}$). This SNP was significantly associated with CP under the assumption of the recessive model of inheritance ($P = 9.87 \times 10^{-6}$). The A allele of the SNP rs1793949, located in intron 44 of the collagen type 2 alpha 1 (*COL2A1*) gene, was also associated with higher risk of CP (OR = 1.596, 95% CI 1.235-2.229; $P = 7.26 \times 10^{-4}$). Associations with these two SNPs did withstand correction for multiple testing after dividing by the number of genes analyzed ($P_{corr} = 0.018$ and 0.029; respectively).

Other markers of interest included the *FGF2, MSX1, FGFR1, WNT3*, and *TIMP3* loci (Table 5). In the case of the *MSX1, FGFR1, WNT3*, and *TIMP3* markers, a variant allele was associated with an increased risk, whereas the variant allele of the *FGF2* SNP had a protective effect.

There was no evidence for a sex-specific component in the association between CP and the 10 markers with the strongest association signals ($P < 0.01$). The genotype distributions in male and female cases were similar for these SNPs ($P > 0.05$).

Haplotype analysis

LD and haplotype analysis were performed with the whole dataset (non-syndromic CP cases + controls, N = 710). Case-control analysis was performed for haplotypes constructed from SNPs within LD blocks identified in the *IRF6, COL2A1, COL11A2*, *MSX1, CLPTM1, BMP2, WNT3, MMP2, FGF1, FGF2*, *FGFR1, EDN1*, *JAG2, PVRL2*, *SKI, TIMP2,* and *TIMP3* genes. The sliding-window approach was applied.

Table 5. Most significant results from single-marker association analysis in Baltic CP sample

Gene	Chr	SNP	Location	Alleles[a]	MAF		P	OR	95% CI
					cases	controls			
IRF6	1	rs17389541	208053795	A/G	0.361	0.246	**5.45×10^{-4}**	1.726	1.263–2.358
		rs9430018	208063165	G/T	0.514	0.440	0.0454	1.351	1.006–1.814
SKI	1	rs12562937	2219338	C/T	0.087	0.151	0.0143	0.534	0.321–0.889
TGFA	2	rs6743202	70537959	A/T	0.441	0.368	0.0467	1.356	1.004–1.831
FGF2	4	rs308434	123991278	C/T	0.054	0.119	0.0057	0.421	0.224–0.791
		rs308395	123966392	C/G	0.068	0.128	0.0147	0.499	0.283–0.881
		rs2034461	124003745	G/A	0.130	0.186	0.0493	0.652	0.424–1.001
		rs308379	124002346	T/A	0.471	0.386	0.0202	1.419	1.055–1.907
		rs308441	123993515	C/T	0.125	0.185	0.0360	0.629	0.407–0.973
MSX1	4	rs1106514	4926827	G/C	0.438	0.344	0.0095	1.482	1.100–1.998
FGF1	5	rs11167785	142043913	T/C	0.207	0.285	0.0197	0.655	0.456–0.936
		rs7722035	142036157	G/C	0.403	0.487	0.0262	0.712	0.527–0.961
		rs17208908	142040162	G/A	0.192	0.264	0.0280	0.664	0.459–0.959
COL11A2	6	rs213209	33284936	C/T	0.188	0.268	0.0138	0.630	0.435–0.912
		rs9277928	33236438	G/T	0.132	0.192	0.0451	0.637	0.408–0.993
FGFR1	8	rs7829058	38451252	G/C	0.164	0.098	0.0049	1.798	1.189–2.720
		rs2978083	38408060	C/T	0.019	0.061	0.0140	0.300	0.109–0.829
COL2A1	12	rs1793949	46657862	G/A	0.514	0.390	**7.26×10^{-4}**	1.659	1.235–2.229
		rs6823	46648679	G/C	0.572	0.469	0.0058	1.517	1.270–2.041

Gene	Chr	SNP	Location	Alleles[a]	MAF		P	OR	95% CI
					cases	controls			
		rs12228854	46683187	G/T	0.212	0.139	0.0067	1.663	1.148–2.409
		rs12368284	46659448	A/G	0.317	0.413	0.0093	0.661	0.483–0.904
		rs10875713	46650730	A/T	0.178	0.119	0.0197	1.596	1.074–2.370
		rs11168359	46702290	G/A	0.082	0.141	0.0203	0.544	0.323–0.916
JAG2	14	rs10134946	104725651	C/T	0.442	0.364	0.0318	1.384	1.028–1.864
TIMP2	17	rs7218237	74383233	G/T	0.058	0.104	0.0391	0.530	0.288–0.978
WNT3	17	rs11653738	42242117	T/C	0.414	0.317	0.0064	1.518	1.123–2.053
OFC11	18	rs328149	42424469	A/G	0.178	0.120	0.0213	1.586	1.068–2.356
CLPTM1	19	rs5127	50144534	G/T	0.308	0.229	0.0146	1.494	1.081–2.064
		rs16979595	50169221	G/A	0.264	0.198	0.0288	1.457	1.038–2.046
PVRL2	19	rs6859	50073874	G/A	0.476	0.402	0.0472	1.350	1.003–1.816
BMP2	20	rs7270163	6699316	A/G	0.087	0.153	0.0114	0.524	0.315–0.871
		rs235742	6727526	A/T	0.245	0.181	0.0287	1.473	1.039–2.088
		rs1980499	6694498	T/C	0.539	0.459	0.0347	1.375	1.022–1.850
		rs17731603	6720066	A/G	0.135	0.089	0.0375	1.598	1.024–2.493
TIMP3	22	rs242082	31554439	C/T	0.413	0.316	0.0068	1.518	1.121–2.055

[a] Major allele (in controls) is listed first

Table 6 Case-control association analysis of haplotypes in collagen family genes in Baltic CP sample

Haplo-type	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	Frequency		P
						cases	controls	
				COL11A2 – Block 1				
	rs12526336	rs1547387	rs213209	rs213208	rs2854028			
H1	G	C	C	G	C	0.445	0.342	0.0043
H1	*	C	C	G	C	0.581	0.472	0.0038
H1	G	C	C	*	*	0.577	0.473	0.0057
H1	*	C	C	G	*	0.709	0.603	0.0038
H1	*	*	C	G	C	0.593	0.488	0.0053
H4	*	C	C	*	*	0.712	0.603	0.0029
				COL11A2 – Block 2				
	rs9277928	rs3130165	*	*	*			
H1	G	C				0.606	0.477	9.85×10^{-4}
				COL2A1 – Block 1				
	rs1635527	rs12228854	rs1859443	rs6580647	*			
H4	G	T	G	C		0.212	0.138	0.0058
H4	G	T	G	*		0.211	0.136	0.0045
H4	*	T	G	C		0.212	0.139	0.0071
H4	G	T	*	*		0.214	0.138	0.0047
H4	*	T	G	*		0.212	0.139	0.0067
				COL2A1 – Block 2				
	rs12721428	rs1793949	rs12368284	*	*			
H1	C	A	A			0.514	0.387	5.76×10^{-4}
H2	C	G	G			0.317	0.413	0.0097
H1	C	G	*			0.431	0.534	0.0062
H2	C	A	*			0.513	0.388	6.72×10^{-4}
H1	*	A	A			0.514	0.389	6.78×10^{-4}
H2	*	G	G			0.317	0.412	0.0098

Haplo-type	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	Frequency		P
						cases	controls	
			COL2A1 – Block 3					
	rs12822608	rs6823	rs10875713	*	*			
H1	G	G	A			0.424	0.528	0.0058
H3	G	C	T			0.177	0.112	0.0090
H1	G	G	*			0.425	0.531	0.0047
H2	G	C	*			0.517	0.386	3.68×10^{-4}
H1	*	G	A			0.426	0.527	0.0073

The strongest association with CP was found for the second most frequent haplotype rs17389541-rs9430018 GT in the *IRF6* gene (with frequencies of 0.353 in NSCP cases and 0.233 in controls, respectively; $P = 2.23 \times 10^{-4}$). Additionally, multiple haplotypes in the *COL11A2* and *COL2A1* genes were associated with a higher risk of CP. Table 6 presents the results from haplotype-based association analysis within LD blocks identified at the *COL11A2* and *COL2A1* loci for multiple 2- to 5-SNP haplotypes with a frequency of > 5% in CP patients and with $P \leq 0.01$.

The lowest P-values were revealed for common core haplotypes CAA in *COL2A1* block 2 ($P = 5.76 \times 10^{-4}$) and GC in *COL11A2* block 2 ($P = 9.85 \times 10^{-4}$), and for the second most frequent (H2) haplotype GC in *COL2A1* block 3 ($P = 3.68 \times 10^{-4}$). The most frequent (H1) haplotypes in *COL2A1* block 3, such as GG (P = 0.0047) and GGA (P = 0.0058), and haplotype CG in *COL2A1* block 2 (P = 0.0062) were associated with lower risk of CP.

Table 7 presents the results from haplotype-phenotype association analysis of common haplotypes with $P \leq 0.01$ in the *CLPTM1, BMP2, WNT3, MMP2, FGFR1,* and *MSX1* genes.

Table 7. Case-control association analysis of haplotypes in six candidate genes

Haplotype	SNP 1	SNP 2	SNP 3	Frequency		P
				cases	controls	
			CLPTM1			
	rs5127	rs10413089	rs3760629			
H2	T	T	C	0.301	0.212	0.0045
			WNT3 – Block 1			
	rs199497	rs199496	rs11658976			
H2	T	G	A	0.443	0.339	0.0035
			WNT3 – Block 2			
	rs11653738	rs3933652	rs3933653			
H2	C	C	C	0.407	0.314	0.0087
			BMP2			
	rs7270163	rs1005464	rs235770			
H4	G	G	C	0.084	0.152	0.0094
			MMP2			
	rs837533	rs837535	rs12924764			
H2	A	A	A	0.318	0.232	0.0072
			FGFR1			
	rs7012413	rs6996321	rs7829058			
H3	C	G	G	0.156	0.249	0.0040
H1	C	*	G	0.517	0.616	0.0073
H3	C	*	C	0.157	0.095	0.0072
			MSX1			
	rs1106514	rs12501827	*			
H1	G	C		0.563	0.656	0.0092

The strongest haplotype-phenotype associations in other candidate loci were found for *WNT3* H2 TGA (P = 0.0035) and *CLPTM1* H2 TTC (P = 0.0045) haplotypes, which were associated with higher risk of CP, whereas the *FGFR1* H3 CGG haplotype (P = 0.0040) was associated with lower risk of CP. The P-values obtained from haplotype-based association analyses performed with *IRF6*, *COL11A2*, *COL2A1*, *CLPTM1*, *WNT3, BMP2*, and *MMP2* markers were lower than analyses using individual SNPs.

Gene-gene interaction

The best results from the SNP × SNP epistasis analysis of all pairwise combinations between 591 individual SNPs in Baltic CP sample are presented in Table 8.

Table 8. SNP × SNP epistasis analysis in Baltic CP sample

Chr	Gene	SNP1	Chr	Gene	SNP2	OR	P
4	*MSX1*	rs868257	20	*BMP2*	rs910141	3.240	1.296×10^{-4}
5	*FGF1*	rs250092	19	*PVRL2*	rs387976	3.151	1.890×10^{-4}
4	*FGF2*	rs1960669	12	*COL2A1*	rs6823	0.2825	3.030×10^{-4}
2	*TGFA*	rs12328204	17	*RARA*	rs506728	0.2311	3.200×10^{-4}
8	*FGFR1*	rs2288696	12	*COL2A1*	rs1793958	0.2733	3.239×10^{-4}
12	*COL2A1*	rs1635550	16	*MMP2*	rs11643630	3.020	4.396×10^{-4}
6	*COL11A2*	rs213209	14	*TGFB3*	rs2205181	2.778	4.468×10^{-4}
4	*FGF2*	rs2034461	6	*COL11A2*	rs2744507	5.289	4.512×10^{-4}
9	*FOXE1*	rs973473	14	*JAG2*	rs2091918	2.218	4.894×10^{-4}
1	*MTHFR*	rs1801131	17	*WNT3*	rs199497	3.601	5.039×10^{-4}
6	*COL11A2*	rs2855425	8	*FGFR1*	rs6987534	0.4755	5.572×10^{-4}
8	*FGFR1*	rs328300	17	*WNT3*	rs7218567	2.233	6.287×10^{-4}

The most significant results were found between SNPs rs868257 located ~9 kb upstream of the *MSX1* gene and SNP rs910141 located ~7 kb downstream of the *BMP2* gene, SNP rs250092 in the *FGF1* gene and SNP rs387976 in the *PVRL2* gene, and between SNP rs1960669 in the *FGF2* gene and SNP rs6823 located ~4 kb downstream of *COL2A1* within 3'UTR of the *TMEM106C* gene. An interaction between *MSX1* and *BMP2* revealed a trend towards significance after Bonferroni correction ($P_{corr} = 0.076$).

3.3. Association analysis in Baltic CL/P sample (Ref. III)

Following quality control, association analysis of 587 tagSNPs was performed in 300 CL/P patients and 606 unrelated healthy controls from Estonian, Latvian, and Lithuanian populations. The average genotype call rate for these SNPs was 99.1%. The genomic control inflation factor (λ_{GC}) was 1.008 for entire dataset.

All markers with allelic association test P-values below 0.05 (before correction for multiple testing) in the CL/P case-control sample are presented in Table 9. Thirty-three polymorphisms in 16 genes reached the nominal significance level. The strongest association with CL/P was found for SNP rs34010, which is located in intron 2 of the fibroblast growth factor 1 (*FGF1*) gene, where the T allele was associated with decreased risk (OR = 0.689, 95% CI 0.559–0.849; $P = 4.56 \times 10^{-4}$). The *WNT9B* rs4968282 minor allele G and the *FOXE1* rs7860144 minor allele A were both associated with decreased risk of CL/P (OR = 0.688, 95% CI 0.548–0.865, P = 0.0013; and OR = 0.723, 95% CI 0.589–0.889, P = 0.0021; respectively). Association with rs34010 did withstand correction for multiple testing after dividing by the number of independent SNPs, taken to be equal to the number of haploblocks (N = 82) within candidate genes ($P_{corr} = 0.037$) assuming that SNPs within one haploblock are not independent. The SNPs rs1907998 and rs6446693, both located 5' of *MSX1*, were associated with CL/P under the assumption of the dominant model of inheritance ($P = 3.97 \times 10^{-4}$ and 5.84×10^{-4}, respectively). Ten markers reached the 1% significance level, revealing *FOXE1, TIMP2, PVRL2,* and *MMP13* genes as additional loci of interest (Table 9).

Stratification of our case-control sample according to sex showed no evidence for a sex-specific component in the association between CL/P and the 10 markers with the strongest association signals ($P < 0.01$). The genotype distributions in male and female cases were similar for these SNPs ($P > 0.05$).

Haplotype analysis

LD and haplotype analysis were performed with the whole dataset (non-syndromic CL/P cases + controls, N = 906). Eighty-two haplotype blocks were described in selected regions. Case-control analysis was performed for haplotypes constructed from SNPs within LD blocks, and the sliding-window approach was applied. Two haplotypes in *FGF1* reached $P < 0.001$, and several haplotypes in seven genes – *FOXE1, FGF1, PVRL2, WNT9B, LHX8, MMP9,* and *TIMP2* showed $0.001 < P < 0.01$.

Table 10 presents the results from haplotype-based association analysis within LD blocks identified at the *FGF1, FOXE1,* and *TIMP2* loci for sliding windows of 2- to 4-SNP haplotypes with a frequency of > 5% in CL/P patients and with $P \leq 0.01$.

Table 9. Most significant results from single-marker association analysis in Baltic CL/P sample

Gene	Chr	SNP	Location	Alleles[a]	MAF[b]		*P*	OR	95% CI
					cases	controls			
IRF6	1	rs630065	208065285	T/C	0.285	0.238	0.0290	1.279	1.025–1.596
		rs17389541	208053795	A/G	0.201	0.246	0.0306	0.768	0.605–0.976
FN1	2	rs1404772	215918544	A/C	0.095	0.064	0.0169	1.542	1.078–2.204
		rs10498038	215964589	C/G	0.040	0.064	0.0385	0.612	0.383–0.978
TGFA	2	rs7605323	70637335	T/C	0.393	0.345	0.0471	1.230	1.002–1.509
MSX1	4	rs6446693	4905981	C/T	0.469	0.407	0.0135	1.286	1.053–1.570
		rs868257	4903074	C/G	0.322	0.274	0.0354	1.256	1.015–1.555
		rs2034461	4896678	G/A	0.130	0.186	0.0394	0.782	0.619–0.988
FGF2	4	rs11737764	124046230	C/T	0.047	0.077	0.0151	0.587	0.380–0.906
		rs308434	123991278	C/T	0.082	0.119	0.0173	0.661	0.469–0.931
		rs308441	123993515	C/T	0.143	0.185	0.0266	0.736	0.561–0.966
FGF1	5	rs34010	141961149	G/T	0.309	0.393	4.56×10^{-4}	0.689	0.559–0.849
EDN1	6	rs16872612	12416068	G/C	0.272	0.225	0.0272	1.290	1.029–1.618
FGFR1	8	rs7829058	38451252	G/C	0.137	0.098	0.0137	1.457	1.079–1.968
		rs6474354	38422122	C/T	0.293	0.247	0.0354	1.265	1.016–1.575
FOXE1	9	rs7860144	99666705	G/A	0.327	0.402	0.0021	0.723	0.589–0.889
		rs874004	99661939	C/G	0.510	0.437	0.0034	1.341	1.101–1.632
		rs973473	99660551	G/T	0.265	0.329	0.0054	0.735	0.591–0.913
MMP13	11	rs7119194	102312606	C/T	0.023	0.049	0.0095	0.465	0.258–0.840
MMP25	16	rs7188573	3051204	T/C	0.370	0.323	0.0463	1.233	1.003–1.516
WNT9B	17	rs4968282	42313936	A/G	0.223	0.295	0.0013	0.688	0.548–0.865
		rs1105127	42322721	G/C	0.392	0.342	0.0377	1.239	1.012–1.518
TIMP2	17	rs7502916	74413228	A/C	0.488	0.419	0.0050	1.325	1.088–1.613
		rs4789936	74409569	C/T	0.548	0.478	0.0052	1.323	1.087–1.611
		rs6501266	74418948	C/T	0.433	0.498	0.0092	0.770	0.632–0.937

Gene	Chr	SNP	Location	Alleles[a]	MAF[b]		*P*	OR	95% CI
					cases	controls			
		rs7211674	74410660	A/C	0.482	0.424	0.0209	1.261	1.036–1.535
		rs7212662	74429726	T/G	0.490	0.438	0.0398	1.233	1.010–1.505
PVRL2	19	rs519113	50068124	C/G	0.192	0.253	0.0039	0.702	0.552–0.894
		rs2075642	50069307	G/A	0.205	0.161	0.0206	1.347	1.046–1.733
BMP2	20	rs6085682	6719211	G/A	0.378	0.430	0.0374	0.808	0.661–0.988
		rs6054516	6719073	C/T	0.179	0.142	0.0383	1.322	1.015–1.721
MMP9	22	rs6094237	44062697	A/T	0.407	0.459	0.0362	0.809	0.664–0.987
		rs17576	44073632	A/G	0.363	0.414	0.0471	0.815	0.666–0.998

[a] Major allele (in controls) is listed first

Table 10. Case-control association analysis of haplotypes in in *FGF1, FOXE1* and *TIMP2* genes

Haplo-type	SNP 1	SNP 2	SNP 3	SNP 4	Frequency		P
					cases	controls	
			FGF1				
	rs34002	rs250092	rs34010	rs250103			
H2	T	G	T	A	0.288	0.349	0.00921
H1	T	G	T	*	0.308	0.391	5.42×10^{-4}
H4	T	G	G	*	0.173	0.121	0.00274[a]
H2	*	G	T	A	0.288	0.349	0.00914
H2	*	G	T	*	0.309	0.393	5.01×10^{-4}
H2	*	*	T	A	0.287	0.349	0.00794
			FOXE1				
	rs10984009	rs973473	rs874004	rs7860144			
H1	G	G	G	G	0.515	0.442	0.00402
H2	G	T	C	A	0.263	0.332	0.00303
H1	G	G	G	*	0.510	0.437	0.00318[b]
H2	G	T	C	*	0.261	0.330	0.00286
H1	*	G	G	G	0.508	0.438	0.00495
H2	*	T	C	A	0.263	0.328	0.00468
H1	G	G	*	*	0.688	0.605	0.00101
H2	G	T	*	*	0.259	0.330	0.00311
H1	*	G	G	*	0.509	0.438	0.00421
H2	*	T	C	*	0.263	0.329	0.00414
H1	*	*	G	G	0.508	0.437	0.00472
H2	*	*	C	A	0.323	0.402	0.00120
			TIMP2				
	rs4789936	rs7211674	rs7502916	rs6501266			
H1	C	A	A	T	0.407	0.478	0.00518
H1	C	A	A	*	0.402	0.480	0.00168[c]
H1	*	A	A	T	0.409	0.474	0.00886
H1	*	*	A	T	0.425	0.496	0.00436
H2	*	*	C	C	0.481	0.415	0.00892

[a]Likelihood ratio test for SNP1-3 sliding-window block haplotype effect: $\chi^2 = 16.47$ (3 df); P = 0.0009

[b]Likelihood ratio test for SNP1-3 sliding-window block haplotype effect: $\chi^2 = 14.55$ (3 df); P = 0.0022

[c]Likelihood ratio test for SNP1-3 sliding-window block haplotype effect: $\chi^2 = 13.69$ (5 df); P = 0.018

The strongest associations with CL/P were found for the second most frequent haplotype rs250092-rs34010 GT and the most frequent haplotype rs34002-rs250092-rs34010 TGT in the *FGF1* gene ($P = 5.01 \times 10^{-4}$ and 5.42×10^{-4}, respectively). Additionally, several common core haplotypes (H1) in *FOXE1* were associated with higher risk of CL/P, whereas the most frequent haplotypes in *TIMP2* were associated with lower risk of CL/P. The lowest P-values were revealed for common core haplotypes GG in *FOXE1* (P = 0.00101) and CAA in *TIMP2* (P = 0.00168), and for the second most frequent (H2) haplotype CA in *FOXE1* (P = 0.00120).

Table 11 shows the results from haplotype-phenotype association analysis for common haplotypes with P ≤ 0.01 in the *WNT9B, PVRL2, LHX8,* and *MMP9* genes.

Table 11. Case-control association analysis of haplotypes in four candidate genes

Haplo-type	SNP 1	SNP 2	SNP 3	Frequency		P
				cases	controls	
			WNT9B			
	rs17603901	rs4968282	rs1105127			
H1	C	A	*	0.776	0.705	0.00134[a]
H2	*	G	G	0.223	0.294	0.00155
			MMP9			
	rs13038175	rs6094237	rs17576			
H2	G	A	A	0.451	0.387	0.00969
			PVRL2			
	rs519113	rs2075642	*			
H3	G	G	*	0.188	0.251	0.00260
			LHX8			
	rs17565565	rs6593568	*			
H1	C	A	*	0.077	0.046	0.00610

[a]Likelihood ratio test for SNP1-2 sliding-window block haplotype effect: χ^2 = 11.55 (2 df); P = 0.0031

The strongest signals of association in other candidate genes were found for *WNT9B* CA (P = 0.00134) and *LHX8* CA (P = 0.00610) haplotypes, which were associated with an increased risk of CL/P. The *WNT9B* GG (P = 0.00155) and *PVRL2* GG (*P*=0.00260) haplotypes were associated with a decreased risk of CL/P.

Gene-gene interaction

The best results from the SNP × SNP epistasis analysis of all pairwise combinations between 587 individual SNPs are presented in Table 12.

The most significant results were found between SNPs rs1859443 located ~2.5 kb telomeric of the *COL2A1* gene and rs3933652 located in intron 1 of the *WNT3* gene, rs7703976 located ~10 kb centromeric of the *MSX2* gene and rs12373066 located ~10 kb centromeric of the *MMP2* gene, and between rs3732253 within 3'UTR of the *TGFA* gene and rs4673990 located ~19 kb centromeric of the *FN1* within the *ATIC* gene. An interaction between *COL2A1* and *WNT3* revealed a trend towards significance after Bonferroni correction (P_{corr} = 0.056).

Table 12. SNP × SNP epistasis analysis in Baltic CL/P sample

Chr	Gene	SNP 1	Chr	Gene	SNP 2	OR	*P*
12	*COL2A1*	rs1859443	17	*WNT3*	rs3933652	3.023	9.644×10^{-5}
5	*MSX2*	rs7703976	16	*MMP2*	rs12373066	2.006	1.072×10^{-4}
2	*TGFA*	rs3732253	2	*FN1*	rs4673990	1.843	1.231×10^{-4}
2	*FN1*	rs17518731	14	*TGFB3*	rs7156293	0.4295	1.453×10^{-4}
16	*MMP25*	rs2239301	19	*PVRL2*	rs387976	1.838	1.509×10^{-4}
4	*FGF2*	rs308434	5	*FGF1*	rs17208908	2.885	1.947×10^{-4}
2	*TGFA*	rs3771514	5	*FGF1*	rs11167785	2.033	1.977×10^{-4}
4	*MSX1*	rs10002530	17	*WNT9B*	rs12952746	3.040	2.564×10^{-4}
8	*FGFR1*	rs6474354	19	*PVRL2*	rs387976	1.981	2.653×10^{-4}
14	*TGFB3*	rs3917158	20	*BMP2*	rs235770	0.4669	2.771×10^{-4}

3.4. GWAS replication study (Ref. IV, V)

In the course of the first GWAS, performed in an extended German and a mixed European replication sample, three novel CL/P susceptibility loci that attained genome-wide significance (8q24.21, 17q22 and 10q25.3) were reported, and three loci (13q31.1, 15q13.3, 2p21) showed suggestive evidence of association (Birnbaum et al., 2009; Mangold et al., 2010). To investigate whether these six loci confer similar effects in North-East European population(s), the six most significantly associated SNPs (rs987525, rs227731, rs7078160, rs7590268, rs9574565, and rs1258763) were genotyped in CL/P patients and unrelated controls from Estonia and Lithuania. Genotype distributions among controls and cases were consistent with HWE expectations for all six SNPs. The results of the single-marker association analysis for rs987525 are summarized in Table 13.

A highly significant result was observed for rs987525 ($P = 5.97 \times 10^{-5}$) in Estonian CL/P sample. The OR was 1.58 (95% CI 1.02 – 2.44) for the heterozygous genotype and 5.56 (95% CI 2.52 – 12.30) for the variant homozygous genotype, values which are similar to those reported for the German CL/P patients. A highly significant result ($P = 1.6 \times 10^{-5}$) was also observed in Lithuanian CL/P sample, where the OR was 1.89 (95% CI 1.17 – 3.04) for the heterozygous genotype and 11.13 (95% CI 3.00 – 41.36) for the variant homozygous genotype. The risk allele frequency in Estonians and Lithuanians (~16%) was found to be lower than in the previously reported German control sample.

To estimate the contribution of rs987525 to CL/P risk in the investigated populations, the population attributable risk (PAR) was calculated for rs987525. The PAR was 0.204 for the Estonian CL/P dataset and 0.283 for the Lithuanian CL/P dataset, both lower than the PAR calculated from the German CL/P sample.

As described for the German sample, the observed genotype distributions for rs987525 in NSCL/P cases and controls are compatible with a multiplicative ($\psi_1 = 1.95$, $\psi_2 = \psi_1^2 = 3.80$, P = 0.138 in Estonians; $\psi_1 = 2.324$, $\psi_2 = \psi_1^2 = 5.40$, P = 0.119 in Lithuanians) mode of inheritance.

There was no evidence for a sex-specific component in the association between rs987525 and CL/P. The genotype distributions in male and female cases were nearly identical for this SNP in Estonians (P = 0.889) and similar in Lithuanians (P = 0.503). The result in the Estonian sample remained essentially unchanged after excluding the four CL/P patients with a positive family history from the analysis ($P = 3.40 \times 10^{-5}$).

No statistically significant association was observed between rs987525 and CP in the Estonian sample (P = 0.502).

In the further analysis of rs227731 (Chr 2: 43393629), rs7078160 (Chr 10: 118817550), rs7590268 (Chr 13: 79566875), rs9574565 (Chr 15: 30837715), and rs1258763 (Chr 17: 52128237), the Estonian patient sample comprised 101 CL/P patients. The results of the single-marker association analysis are shown in Table 14.

An association for rs7078160 CL/P sample remained significant after Bonferroni correction ($P_{corr} = 0.008$). The OR was 1.64 (95% CI 0.97–2.75) for the heterozygous genotype and 4.68 (95% CI 1.60–13.69) for the variant homozygous genotype. The risk allele frequency in Estonian controls (14.5%) was similar to that in the previously reported German controls (16.3%). To address the question whether the 10q25 locus possesses a larger impact in Estonian CL/P sample, we tested the hypothesis that the genotype-specific relative risks are identical between the German and the Estonian population. No significant differences in relative risks were observed (P = 0.41).

The other four SNPs showed no statistically significant association with CL/P in the Estonian sample. However, point estimates of genotypic relative risks for markers rs9574565 and rs1258763 were similar to the ORs observed in the previous German study. The power of the present sample to demonstrate a genotypic effect of the size reported for markers rs9574565 and rs1258763 was calculated as 0.59 and 0.67, respectively. This indicates that the present sample may have lacked sufficient power to demonstrate such effects at the required level of significance.

Table 13. Distribution of rs987525 (C/A) in the Estonian and Lithuanian case-control samples

Subsample	Genotypes Cases (AA, AC, CC)	MAF cases	Genotypes Controls (AA, AC, CC)	MAF controls	P	OR_{Het} (95% CI)	OR_{Hom} (95% CI)
CL/P – Estonia	10/36/59	0.267	22/279/722	0.158	5.97×10^{-5}	1.58 (1.02–2.44)	5.56 (2.52–12.30)
CL/P – Lithuania	11/46/55	0.304	3/74/167	0.164	1.60×10^{-5}	1.89 (1.17–3.04)	11.13 (3.00–41.36)
CP – Estonia	2/17/39	0.181	22/279/722	0.158	0.502	1.13 (0.63–2.03)	1.68 (0.38–7.42)

Table 14. Results of single marker association analysis in the Estonian case-control sample

SNP	Alleles[a]	Genotypes Cases	MAF cases	Genotypes Controls	MAF controls	P	OR_{Het} (95% CI)	OR_{Hom} (95% CI)
rs7590268	T/G	47/41/8	0.297	127/102/22	0.291	0.876	1.09 (0.66–1.78)	0.98 (0.41–2.36)
rs7078160	G/A	59/32/9	0.250	184/61/6	0.145	0.0016	1.64 (0.97–2.75)	4.68 (1.60–13.69)
rs9574565	C/T	60/31/7	0.230	127/96/25	0.294	0.099	1.15 (0.46–2.93)	1.69 (0.69–4.12)
rs1258763	A/G	54/40/7	0.267	114/115/25	0.325	0.130	1.24 (0.50–3.09)	1.69 (0.69–4.16)
rs222731	A/C	22/56/22	0.500	72/118/58	0.472	0.500	1.55 (0.88–2.76)	1.24 (0.63–2.46)

[a]major allele in controls listed first

4. DISCUSSION

Recent years have witnessed remarkable success in expanding the knowledge of genetic component contributing to orofacial clefts. As a general model, both genetic factors and environmental triggers, acting either independently or in combination, are contributing to the pathogenesis of nonsyndromic clefting. The available databases of variations in human DNA sequence and recent developments in genotyping technologies have enabled in-depth association analyses between genetic markers and disease phenotypes. A classical strategy in searching for genetic susceptibility factors of the disease of interest is relying on the comprehensive knowledge of underlying biological pathways combined with the candidate gene approach. Data from animal models, in which clefts arise either spontaneously or as a result of mutagenesis experiment, combined with an analysis how expression patterns correlate with gene function and examining the effects of gene-environment interactions have proven themselves as powerful tools for identifying candidate genes for complex traits, such as nonsyndromic clefts. Importantly, they also contribute to our knowledge of normal craniofacial development and the molecular pathogenesis of CL/P, taking into account that facial development in mice mirrors human craniofacial development. Several recent studies have also provided strong evidence that syndromic forms having Mendelian patterns of inheritance may provide insights into genetic etiology of nonsyndromic forms of clefting.

Furthermore, hypothesis-driven candidate gene studies can be complemented with hypothesis-free GWA studies, and the rapidly increasing number of GWAS provides an unprecedented opportunity to investigate the potential impact of common genetic variants on complex traits. The GWA approach represents an important advance compared to candidate gene studies, in which selection of variants to be analyzed often suffers from imperfect understanding of biological pathways, yielding associations that are difficult to replicate (Manolio et al., 2009). To date, GWA studies have provided hundreds of common susceptibility variants for human diseases and continuous traits of biomedical importance, and have provided valuable insights into the underlying genetic architecture of common complex diseases. However, common genetic variants have typically explained only a small fraction of the inherited risk, and therefore, further deep sequencing studies which include also rare single nucleotide variants, CNVs and other structural variants are required to identify additional genes or loci that may be implicated in complex traits.

On the basis of multiple reports on association between markers and haplotypes in various genes and orofacial clefts, we analyzed the role of 40 candidate gene regions on 17 chromosomes, previously suggested as risk factors for orofacial clefts, to determine their role in the etiology of CL/P and CP in a new clefting sample representing three populations from the Baltic region – Estonians, Latvians and Lithuanians. The genetic relatedness of Estonians, Latvians and Lithuanians, sharing the same geographic origin, has been recently confirmed using the principal component analysis, according to the pairwise

inflation factor λ and pairwise Fst values between samples (Nelis et al., 2009). We found no evidence of systematic bias due to population stratification as indicated by the genomic control inflation factor (λ_{GC}~1) and the quantile-quantile plots, which confirmed the high degree of homogeneity between all three population samples, allowing us to summarize the data of three Baltic countries to increase the study power.

4.1. Genetic susceptibility loci for CP

Results from this association analysis suggest that several regions contain genes predisposing to the development of CP. Among the 40 candidate genes analyzed, 35 SNPs in 17 genes showed nominal P values < 0.05. The strongest evidence of association was found for SNPs rs17389541 in the *IRF6* gene and rs1793949 in *COL2A1*. These associations were not significant after Bonferroni correction, but remained significant after correction by the number of genes evaluated. Haplotype analysis of the aforementioned genes supported these findings. Additional evidence from haplotype analysis demonstrating the possible involvement of *COL11A2*, *FGFR1*, *CLPTM1*, and *WNT3* variants in predisposition to CP was revealed. We can assume that the described polymorphisms are not functionally significant variants that actually contribute to disease susceptibility, and our findings need confirmation by replication in other independent cohorts or by resequencing the selected candidate genes in patients to identify the causal variants.

The most common syndromic form of orofacial clefts is Van der Woude syndrome (VWS), an autosomal dominant disorder characterized by the presence of CL/P or CP and/or lower lip pits. VWS is caused by mutations in the *IRF6* gene, which belongs to a family of transcription factors that share a highly conserved winged-helix DNA-binding domain and a less conserved protein interaction domain (Kondo et al., 2002). *IRF6* is expressed in the medial edge epithelia of the palatal shelves immediately before and during fusion (Knight et al., 2006). Interestingly, VWS is an example of an orofacial syndrome in which cases of CP and CL/P can occur in the same pedigree, suggesting that *IRF6* is likely involved in the fusion process that occurs in both primary and secondary palatogenesis. However, *IRF6* mutations are rare in families with nonsyndromic orofacial clefts (Jehee et al., 2009). Recently, it has been demonstrated that *IRF6* is essential for oral epithelial differentiation and that *IRF6* plays a key role in the control of palatal adhesion and fusion competence (Richardson et al., 2009). In the present study, the *IRF6* SNP rs17389541 showed evidence of association, supported by analysis of haplotypes including this polymorphism, which is a novel implication of *IRF6* in CP susceptibility. HapMap data from the CEU reference sample indicate that rs17389541 is not in strong LD with the common polymorphism *IRF6* rs642961 (r^2 = 0.057) that was significantly associated with CL/P and particularly with cleft lip in Europeans, but not with cleft palate (Rahimov et al., 2008). It is

likely that association between common variants in the *IRF6* locus and the risk of CP can be identified in other European populations and that the *IRF6* locus represents an important genetic modifier for this multifactorial malformation. Recently, a combination of experimental strategies has demonstrated that *IRF6* is a direct target of p63, and that p63 and IRF6 function within a regulatory loop to coordinate epithelial proliferation and differentiation during normal palate development, where p63 activates *IRF6* transcription through the *IRF6* enhancer element, variation within which increases susceptibility to cleft lip. Mutations in *TP63* or *IRF6* cause disruption of this loop and lead subsequently to several congenital malformations that include clefting as a hallmark feature (Thomason et al., 2010; Gritli-Linde et al., 2010).

Mutations in genes coding for cartilage collagens II and XI, *COL2A1*, *COL11A1*, and *COL11A2* cause syndromes that are often associated with Pierre-Robin sequence, cleft palate, or micrognathia, but it seems possible that sequence variations in collagen II and XI genes can cause or predispose to nonsyndromic conditions in some instances (Melkoniemi et al., 2003). A variety of mutations in the *COL2A1* gene have been identified in families with Stickler syndrome type 1 with systemic features involving Pierre-Robin sequence, severe myopia and/or vitreoretinal phenotype, and osteoarthritis later in life. This study presents the first report describing associations between CP and several common SNPs together with multiple risk and protective haplotypes in LD blocks spread throughout the entire *COL2A1* gene, with the strongest association signals found for the rs12822608-rs6823 GC haplotype downstream of the 3'UTR region and the intronic SNP rs1793949. Epistatic interactions were identified between SNPs in *COL2A1* and in *FGF2*, *FGFR1*, and *MMP2* genes. In addition, our data showed an association between CP and certain haplotypes in *COL11A2*, supporting recent findings describing the haplotype effect found for the *COL11A2* gene in a Norwegian sample of CP case-parent trios (Jugessur et al., 2009). Epistatic interactions were observed between SNPs in *COL11A2* and both *FGF2* and *TGFB3* genes. Moreover, one of the chromosomal regions for clefting in mice encompasses the H2 locus, which is homologous to the human HLA locus on 6p21 and includes susceptibility for both corticosteroid-induced and vitamin A-enhanced cleft palate (Murray, 1995).

The Wnt signaling pathway genes are involved in craniofacial development and upper lip fusion and are therefore plausible candidates for an etiologic role in nonsyndromic clefting. Wnt expression is observed in the upper lip and primary and secondary palates, and Wnt signaling mediates regional specification in the vertebrate face (Brugmann et al., 2007). In our study, the *WNT3* SNP rs11653738 and two risk haplotypes within different LD blocks showed evidence of association with CP. In addition, epistatic interactions were identified between SNPs in *WNT3* and both *FGFR1* and *MTHFR* genes. *WNT3*, a human homologue of the Drosophila spp. wingless gene, encodes a member of the WNT family known to play key roles in embryonic development, and *WNT3* is required at the earliest stages of human limb formation and for

craniofacial development. Furthermore, the clf1 locus mapped in clefting susceptible mice contains two *Wnt* genes, *Wnt3* and *Wnt9b* (Juriloff et al., 2005). In a recent study, the strongest association signals were found between SNPs and haplotypes in *WNT3A* and CL/P, as well as evidence of gene-gene interaction between *WNT3* and *WNT3A* (Chiquet et al., 2008). These findings were extended in a Norwegian sample of CP case-parent trios where a haplotype effect was reported for the *WNT3A* gene (Jugessur et al., 2009).

Several members of the fibroblast growth factor (FGF) and FGF receptor (FGFR) families are expressed during craniofacial development and can rarely harbor mutations that result in human clefting syndromes. In a study of Kallmann syndrome patients, CP and dental agenesis were exclusively found associated with *FGFR1* loss-of-function mutations (Albuisson et al., 2005). Animal models also support the involvement of FGFs and FGFRs in the pathogenesis of oral clefting; for example, Fgfr1 hypomorphic mice have cleft palate (Trokovic et al., 2003). We found associations between CP and SNPs rs7829058 in *FGFR1* and rs308434 in *FGF2* gene. Association with *FGFR1* was supported by haplotype analysis. Epistatic interaction was identified between SNPs in *FGF1* and *PVRL2* genes. Recently, borderline significant association between SNPs in several *FGF*/*FGFR* genes (including *FGFR1*) and CL/P was demonstrated, and it was suggested that impaired FGF signaling contributes to 3 to 5% of CL/P (Riley et al., 2007a). *FGFR1* encodes a trans-membrane receptor tyrosine kinase that transduces signals from secreted FGFs, and insufficient *FGFR1*-mediated signaling during embryonic development may affect palatogenesis in humans and/or mice.

Mutations in the T-box transcription factor gene *TBX22* are found in familial and sporadic patients with X-linked cleft palate (CPX) and ankyloglossia, which is inherited as a semidominant X-linked disorder affecting male patients and approximately one third of female carriers (Braybrook et al., 2001). The phenotypic variability in CPX ranges from a mild submucous cleft palate to a severe, complete cleft of the secondary palate. In addition to familial CPX cases, mutations in *TBX22* have been identified in approximately 5 to 8% of all patients with CP, and *TBX22* has been proposed to contribute significantly to the prevalence of CP across different populations (Marçano et al., 2004; Suphapeetiporn et al., 2007). In a recent study, a functional haplotype variant in the *TBX22* promoter was significantly associated with CP and ankyloglossia (CPA), but association was not significant between patients with cleft palate only (CPO) and controls (Pauws et al., 2009). Likewise, our results do not support an involvement of common *TBX22* variants in CP predisposition.

Multiple lines of evidence have implicated one or more genes at 19q13 in the etiology of nonsyndromic clefting. A study of a multiplex family in which CL/P segregated with a balanced translocation between 2q11.2 and 19q13.3 suggested that *CLPTM1* (cleft lip and palate transmembrane 1) gene, localized to this breakpoint might play a role in clefting (Yoshiura et al., 1998). A previous metaanalysis of 13 genome scans, involving combined datasets from multiple populations, supported the role of 19q13 in clefting (Marazita et al., 2004). In

our independent sample, haplotype TTC showed evidence of association, suggesting the involvement of *CLPTM1* in CP etiology. In addition, further evidence that the 19q13 region contributes to isolated clefting in heterogeneous populations of European descent has been found (Warrington et al., 2006).

The present study is the first to demonstrate an association between common SNPs and haplotypes in *IRF6*, *COL2A1*, and *WNT3* and nonsyndromic CP. Moreover, our analysis contributes to the evidence that variation in the *COL11A2*, *FGFR1*, and *CLPTM1* genes may influence the risk of cleft palate. The relatively small number of genes associated with this birth defect suggests the presence of causal variants not genotyped in this study. Our selection of haplotype-tagging SNPs was based on the HapMap Phase II data, and although analysis of haplotypes may protect against the loss of power to detect an association signal, insufficient SNP resolution may be a concern in the current study. Likewise, this finding may be a consequence of interactions with genetic variants in candidate genes overlooked in our selection and/or with environmental factors. It can be assumed that LD-based approach may have insufficient power to detect rare variants (e.g., CNVs) that could also be functionally relevant, and further studies are warranted to confirm reported associations.

The results of this study provide evidence that variation in cartilage collagen II and XI genes, *IRF6*, and the Wnt and FGF signaling pathway genes are likely involved in the etiology of CP in Northeastern European populations.

4.2. Genetic susceptibility loci for CL/P

Data from Estonian study provided further evidence implicating *MSX1* and *MTHFR* in the etiology of nonsyndromic CL/P across different populations.

Methylenetetrahydrofolate reductase (MTHFR) is an important enzyme involved in folate metabolism. Many studies have been undertaken to verify the association between two functional polymorphisms – C677T (Ala222Val) and A1298C (Glu429Ala) – in the *MTHFR* gene and an increased risk of CL/P. However, these studies have provided inconsistent results because of differences in the studied populations, including diverse genetic backgrounds and different exposure to varying environmental risk factors. It has been proposed that low periconceptional folate intake increases the risk of CL/P in offspring, and this risk is even more pronounced in mothers with *MTHFR* 677TT or 1298CC genotypes (van Rooij et al., 2003). In our study, the *MTHFR* SNPs rs1994798, rs1476413, and rs1801131 (A1298C), together with multiple risk and protective haplotypes within the same LD block, showed evidence of association with CL/P. In addition, we found evidence suggestive of an epistatic interaction between the SNPs rs1801131 and rs3760629 in the *CLPTM1* gene. It has been suggested that rs1801131 is not directly connected to the risk of developing CL/P, but this may be a result of near-complete LD between A1298C and C677T or disequilibrium with another mutation responsible for the

malformation (Pezzetti et al., 2004). Confirmation of an association or linkage between A1298C and CL/P risk, either separately or in combination with C677T, has not been found (Beaty et al., 2002; van Rooij et al., 2003). An association between A1298C and CP patients or mothers was not found in a previous study (Jugessur et al., 2003). Likewise, the A1298C variant was not found to be a risk factor for CL/P or CP (Mills et al., 2008). In conclusion, it appears likely that A1298C and the two other polymorphisms described here are not independent risk factors for CL/P, but our data do not exclude a possible involvement of the folate pathway in the development of CL/P.

MSX1 has been a plausible candidate gene for clefting. During palatogenesis, *Msx1* is a downstream target of BMP signalling in a number of embryonic tissues and *Msx1* is necessary for expression of *Bmp4* and/or *Bmp2* (Zhang et al., 2002). The complete sequencing of the *MSX1* gene demonstrated that rare point mutations in this gene appear to contribute to approximately 2% of all cases of CL/P (Jezewski et al., 2003). Association studies of CL/P (Lidral et al., 1998; Vieira et al., 2003; Beaty et al., 2002; Fallin et al., 2003; Suazo et al., 2004) and CP (Lidral et al., 1998) have supported a role for *MSX1* in nonsyndromic clefting in different populations. In our study, the most significant association with CL/P among all screened candidate genes was found for SNP rs6446693 in the *MSX1* region. Taken together with several other SNPs and haplotypes demonstrating an association with CL/P, it can be concluded that we have successfully replicated previous findings, showing an association between *MSX1* variants and CL/P. It has been suggested that an interaction between the *MSX1* and *TGFB3* loci is involved in the pathogenesis of CL/P (Lidral et al., 1998), and both genes have been reported as contributors to clefting (Vieira et al., 2003). Our data from *TGFB3* SNPs and haplotype association analyses yielded only borderline significance and no significant evidence of interaction between *MSX1* and *TGFB3* was found; therefore, our data do not support the involvement of *TGFB3* as a major locus predisposing to CL/P. However, our data provide the first evidence suggestive of an interaction between *IRF6* and *MSX1* which may be involved in the pathogenesis of CL/P. Significant evidence of an interaction between *IRF6* and *MSX1* has been reported in a study of human tooth agenesis, suggesting also that the described interaction may be relevant to CL/P (Vieira et al., 2007).

Baltic study

Results from this association analysis suggest that several regions may contain genes predisposing to the development of CL/P. Among the 40 candidate genes analyzed, 33 SNPs in 16 genes reached the 5% significance level. The most significant associations were found for SNPs in the *FGF1, FOXE1* and *WNT9B* genes. The strongest evidence of association was found for SNP rs34010 in the *FGF1* gene. This association remained significant after correction by the number of haploblocks evaluated. Conditional haplotype analysis of the aforementioned genes supported these findings. Additional evidence from

haplotype analysis demonstrating the involvement of *TIMP2, PVRL2, LHX8*, and *MMP9* variants in CL/P predisposition was revealed. The association findings in Estonian sample which did support common SNPs and haplotypes in *MSX1* and *MTHFR* as susceptibility factors for CL/P were not confirmed to be significant in all-Baltic dataset. Our approach allowed to detect only relatively common variants associated with CL/P. Efforts to find other etiologic variants in these genes will require in depth sequencing to characterize the possible regulatory function within the regions encompassing our putative disease haplotypes.

The FGF signaling pathway is known to have an important role in craniofacial development, and perturbation of the FGF signal is critical to palatogenesis (Nie et al., 2006). A systematic analysis of genes encoding different fibroblast growth factor (FGF) proteins and their receptors have highlighted the importance for FGF signaling in nonsyndromic CL/P, and it was suggested that impaired FGF signaling may contribute to 3–5% of CL/P (Riley et al., 2007a). In addition, FGF pathway has been suggested to be involved in interactions with environmental risk factors for CL/P (Pauws & Stanier, 2007). Recently, *FGF12* haplotypes were shown to be significantly associated with CL/P in two Scandinavian samples (Jugessur et al., 2009). In the present study, the *FGF1* SNP rs34010 showed evidence of association, supported by analysis of haplotypes including this polymorphism, which is a novel implication of *FGF1* in CL/P susceptibility. Epistatic interactions were observed between SNPs in *FGF1* and both *FGF2* and *TGFA* genes. Our results give additional evidence that disruptions in FGF signaling pathway contribute to the pathogenesis of CL/P, also suggesting that association between common variants in the *FGF1* locus and the risk of CL/P can be found in other European populations.

Fine mapping studies of the 9q22-q33 linkage region have implicated *FOXE1* on 9q21 as the third major locus along with *IRF6* and 8q24.21 in which common variants have a significant impact on the occurrence of CL/P in diverse populations (Marazita et al., 2009; Moreno et al., 2009). The forkhead box E1 gene (*FOXE1*) is a member of a transcription factor family that contains a DNA-binding forkhead domain and regulates diverse developmental processes in eukaryotes. *Foxe1* is expressed in the secondary palate epithelium in humans at embryonic week 11, and the specific expression pattern of *Foxe1* at the point of fusion between the medial nasal and maxillary processes strongly suggests *FOXE1* as an important player in primary palatogenesis (Trueba et al., 2005; Moreno et al., 2009). Twelve coding variants – missense or silent – in *FOXE1* have been described as rare causes of nonsyndromic CL/P (Vieira et al., 2005). Recent studies have implicated SNPs and/or haplotypes clustering in or near *FOXE1* as highly significant risk factors of CL/P, and the described association signals indicated the presence of causal mutations in nearby noncoding regions that regulate *FOXE1* expression (Jugessur et al., 2009; Marazita et al., 2009; Moreno et al., 2009). The three associated SNPs from the current study – rs7860144, rs874004 and rs973473 – are located within a 70 kb haplotype block

including SNPs rs3758249 and rs4460498, which were the most significantly associated with CL/P in multiple populations (Moreno et al., 2009). Our results support a substantial role for the *FOXE1* in the development of CL/P and it's likely that the *FOXE1* locus represents an important genetic modifier for this multifactorial malformation.

Wnt expression is observed in the upper lip and primary and secondary palates during craniofacial development and genes in the Wnt signaling pathway are therefore plausible candidates to have an etiologic role in nonsyndromic clefting. *WNT9B* lies ~32 kb telomeric from the start codon of *WNT3*, which is required at the earliest stages of human limb formation and for craniofacial morphogenesis. Furthermore, both Wnt9b and Wnt3 are mapped in the clf1 locus in clefting susceptible mice (Juriloff et al., 2005). The gene alterations outside *WNT9B* coding sequence are expected to cause reduced WNT9B gene or protein function in CL/P cases, and *WNT9B* has been suggested as a strong candidate gene for CL/P (Juriloff et al., 2006). This hypothesis has found further support from the observation that canonical WNT signaling is activated during midfacial morphogenesis in mice (Lan et al., 2006). In addition, genetic inactivation of low density *Lrp6* (lipoprotein receptor-related protein 6), a co-receptor of the WNT–β-catenin signalling pathway, causes CL/P. Moreover, both *Msx1* and *Msx2* are downstream targets of WnT–β-catenin signaling pathway during lip formation and fusion (Song et al., 2009). In the present study, the *WNT9B* SNP rs4968282 and two haplotypes including this polymorphism showed evidence of association with CL/P. Epistatic interaction was identified between common SNPs in *WNT9B* and *MSX1* genes. The SNP rs197915 downstream of *WNT9B* showed association with CL/P in the European American sample subgroup having positive family history, whereas the strongest association signals in entire dataset were reported for SNPs in *WNT3A*, *WNT5A*, and *WNT11* (Chiquet et al., 2008). In contrast, association with rs197915 was not replicated in Brazilian CL/P sample of Caucasian ancestry, while significant association with *WNT3* rs142167 was found (Menezes et al., 2010). Our data support an involvement of the Wnt signaling pathway in orofacial development, also suggesting that the 17q21 locus, containing *WNT9B* and *WNT3* genes, contributes to nonsyndromic clefting in populations of European descent.

Matrix metalloproteinases (MMPs) have been shown to be expressed in the developing secondary palate, associated with extracellular matrix (ECM) breakdown required for palatal fusion (Brown et al., 2002). MMPs are counteracted by the tissue inhibitors of metalloproteinases (TIMPs), which inhibit MMP activity and thereby reduce excessive proteolytic ECM degradation (Verstappen & von den Hoff, 2006). During embryonic craniofacial development, MMPs and TIMPs expression is temporally and spatially regulated to control tissue remodeling, and disruption of their balance can lead to occurrence of malformations, such as CL/P (Blavier et al., 2001). In this study, associations with CL/P were found for several SNPs and haplotypes in *TIMP2*, supporting initial findings in a Norwegian sample of CL/P case-parent

trios where a haplotype effect was reported for the *TIMP2* gene (Jugessur et al., 2009). The inactivation of MMPs by TIMP2 leads to a failure of palatal fusion in mice (Blavier et al., 2001). Therefore, strict regulation of MMP activity is necessary for a complete fusion of the palatal shelves, and an imbalance in favor of TIMP2 may increase the risk of a cleft palate (Verstappen & von den Hoff, 2006). The link between CL/P risk and variants in MMP and TIMP genes seems to rely also on alterations in ECM homeostasis and functions.

Our analysis contributes to the evidence that common haplotypes in *LHX8*, *MMP9* and *PVRL2* may influence the risk of CL/P. Members of the LIM homeobox gene family encode transcription regulators that are required for the specification and differentiation of different cell types during embryonic development, and point mutations in *LHX8* have been suggested as rare causes of CL/P (Vieira et al., 2005). MMPs have been shown to be expressed in the developing secondary palate, associated with ECM breakdown required for palatal fusion (Brown et al., 2002). Despite the recently reported lack of association between *MMP9* and CL/P (Letra et al., 2007), *MMP9* has been demonstrated to be selectively expressed in the ossification centres of the maxilla during palatal fusion (Blavier et al., 2001) and could be considered as candidate gene for clefting.

A previously performed meta-analysis of 13 genome scans was suggestive for the role of 19q13, containing *PVRL2*, in nonsyndromic clefting (Marazita et al., 2004). Interestingly, both rare and common mutations within *PVRL1*, closely related to *PVRL2*, have been reported to make a minor contribution to nonsyndromic CL/P by disrupting the initiation and regulation of cell-to-cell adhesion and downstream morphogenesis of the embryonic face (Avila et al., 2006; Scapoli et al., 2006). Mutations in *PVRL1* cause CLPED1 (Margarita Island ectodermal dysplasia and clefting) syndrome (Suzuki et al., 2000). An association study involving five populations did not find any association between CL/P and variations in the *PVRL2* gene. However, a significant association with an allelic variant in *PVR*, a gene homologus to *PVRL2*, was found (Warrington et al., 2006). In the same study, 16 *PVRL2* variants – 5 common and 11 rare – were identified in cleft patients. In the Baltic sample, the *PVRL2* SNP rs519113 showed evidence of association, supported by analysis of haplotypes including this polymorphism. In this study, no association signals were observed for markers or haplotypes in the *PVRL1* gene. Previous studies have provided inconclusive evidence for the *BCL3* gene, located on 19q13, to be implicated in the etiology of nonsyndromic clefting. In Estonian sample, we described associations between CL/P and *BCL3* markers and haplotypes, including SNP rs8100239, for which excess maternal transmission has been previously reported in patients with CL/P, probably reflecting an imprinting effect or a maternal genotype effect (Park et al., 2009). The association findings in Estonian sample were not confirmed to be significant in all-Baltic dataset, and the contribution of *BCL3* in the etiology of nonsyndromic clefts remains controversial.

GWAS replication study

Our follow-up study has successfully replicated the previously reported genome-wide significant association between rs987525 on chromosome 8q24.21 and CL/P in German discovery study in two independent samples – Estonians and Lithuanians. The risk allele frequency for rs987525 in Estonian and Lithuanian controls (~16%) was found to be lower than in the German discovery sample (~20%) and HapMap CEU reference population (~22%). The population attributable risks (PAR) for rs987525 calculated for the Estonian (0.204) and for the Lithuanian (0.283) CL/P datasets were both lower than the PAR calculated from the German CL/P sample (0.413). This reflects the lower frequency of the risk-associated genotypes in the two investigated populations.

No statistically significant association was observed between rs987525 and CP in the Estonian sample. This finding is in accordance with the findings from 295 German CP families (Birnbaum et al., 2009a).

Additionally, we have successfully replicated the previously reported genome-wide significant association between rs7078160 on chromosome 10q25.3 and CL/P in Estonian sample. The relative risks calculated in Estonian sample were higher than those reported for the German CL/P patients. However, there was no significant statistical evidence for heterogeneity of genotype-specific relative risks between the German and the Estonian population. Therefore, it may be premature to speculate about a different impact of the 10q25.3 locus in these populations.

The other four CL/P susceptibility loci (17q22, 13q31.1, 15q13.3, 2p21) were not replicated. This might reflect a true finding, however, we cannot exclude that the effect size of these four loci in Estonians may not be large enough to be detected given our sample size.

As reported previously, Estonians, the HapMap CEU reference population and several other European populations (including population-based samples from Germany) share a relatively common genetic background (Mueller et al., 2005; Montpetit et al., 2006). This could partially explain the similar results found in Estonians and Germans for both rs987525 and rs7078160. Further studies for above-mentioned SNPs in populations with a more distant genetic background are warranted.

No known genes have been reported within the 640-kb region of association on 8q24.21. According to the UCSC browser, the nearest flanking gene is *CCDC26*, which has been mapped approximately 420 kb telomeric of rs987525. *CCDC26* encodes a retinoic acid-dependent modulator of myeloid differentiation (Hirano et al., 2008). Because exposure to retinoic acid is known to induce orofacial clefts, *CCDC26* could be considered as a potential candidate gene. Two explanations have been proposed to describe the biological mechanism by which rs987525 exerts its effect on the development of CL/P. First, the observed association may mediate its effect by as-yet-unknown transcripts mapping within this region, while an alternative explanation is that the disease-associated region contains regulatory elements controlling the expression of more distant genes (Birnbaum et al., 2009a).

Two genes are located in close vicinity to the marker rs7078160: *KIAA1598* (40 kb centromeric) and *VAX1* (ventral anterior homeobox 1; 53 kb telomeric). *VAX1* encodes a transcriptional regulator with a DNA-binding homeobox domain. Mouse knockouts for *Vax1* display craniofacial malformations including cleft palate, and this gene is expressed widely in developing craniofacial structures (Hallonet et al., 1999), rendering variants in *VAX1* as strong candidates for contributing to CL/P.

The present study is the first to demonstrate an association between CL/P and common SNPs and haplotypes in *FGF1*, and provides new evidence that variation in the *TIMP2* and *WNT9B* genes contributes to nonsyndromic CL/P. Additionally, we have successfully replicated previous findings implicating *FOXE1* as a susceptibility locus for CL/P across different populations. Moreover, we identified associations between common SNPs and/or haplotypes in the *PVRL2*, *LHX8*, and *MMP9* genes. We emphasize that further studies in populations of different ethnic background are warranted to confirm reported associations. However, we could not demonstrate convincing evidence of an association between CL/P and *IRF6*, the candidate gene most consistently replicated. A recent collaborative study has identified a functionally relevant variant rs642961 in *IRF6*, which confers significant risk for CL/P, and the association was strongest in the cleft lip subgroup (Rahimov et al., 2008). This association was confirmed in the Central European CL/P sample (Birnbaum et al. 2009b). On the contrary, ethnic heterogeneity among populations was illustrated in a recent study where strong evidence of association with the rs2235371 was found, whereas very modest association with rs642961 was reported (Blanton et al., 2010). It is generally accepted that nonsyndromic clefts are caused by complex interactions between genetic and environmental variables, while it's becoming gradually more clear that these variables may have a different impact in distinct populations.

Furthermore, we replicated the recently reported significant associations between rs987525 and rs7078160 and CL/P in an independent Estonian case-control sample. Further studies are required to elucidate the biological mechanism by which the 8q24 and 10q25 loci exert their effect in the pathogenesis of CL/P.

The issue of multiple testing is tentative when a large number of SNPs in many genes is analyzed with multiple methodologies. However, most of the tests performed here are not independent in the context of multiple comparisons considering that strong LD between SNPs within inferred haplotype blocks generates significantly correlated P values, therefore applying Bonferroni correction as an adjustment strategy would be highly conservative and may result in too many false-negative findings. Our sample had 80% power to detect a disease-associated variant with an odds ratio $\geq$ 1.4 at the 5% significance level, assuming a frequency of $\geq$ 20% in healthy controls. Therefore, we elected to report empirical P values, and a stringent Bonferroni correction was applied only for epistasis analysis results. Previous studies have used the number of evaluated genes as independent analyses to adjust significance levels (Chiquet

et al., 2008), and such approach was adopted for single-marker association analysis performed in CP sample using a correction factor of 40 (0.05/40 = 0.00125) to account for the number of analyzed genes as independent analyses. Furthermore, for single-marker association analysis in CL/P sample we employed a correction factor of 82 (0.05/82 = 0.00061) to take the number of haploblocks in candidate genes as independent analyses into account.

4.3. Concluding remarks

The occurrence of nonsyndromic oral clefts seems to be influenced by the multifactorial combinations of genetic and environmental risk factors that can differ between CL/P and CP. The genetic mechanisms underlying lip and palate development may be due to the perturbation in important signaling pathways at consecutive levels that are required for the formation of specific anatomical structures. The studies of genes – and their molecular pathways – implicated in syndromic forms of these birth defects have been informative to improve our understanding of human craniofacial pathology, and have recently demonstrated an overlap between syndromic and nonsyndromic CL/P. Taking into account that similar developmental processes occur during the both primary and secondary palatogenesis, we cannot exclude a possibility that some genes may contribute to both types of clefting. Studies of the genetic etiology of clefts allow us to move beyond the attitude that CL/P and CP are only structural birth defects, and may further extend our understanding of other common complex diseases.

Numerous candidate gene studies have mostly failed to identify either major gene involvement or mutations exerting a major influence on risk of of developing nonsyndromic oral clefts. The failure to expound the molecular events leading to clefting in humans probably arises from imperfect knowledge of gene networks and regulation of gene expression during palatal development. More recently, association studies using LD-based approach have been more effective in detecting associations with multiple common variants that confer relatively small increments in risk, and analysis of haplotypes increases the potential to capture association signals from variants that have not been directly typed. The presence of true allelic heterogeneity, in which multiple mutations occur on different background haplotypes, would make it much more difficult to identify causal genes through association studies, and findings from recent GWA studies are increasingly demonstrating that allelic heterogeneity is a frequent feature of polygenic traits.

We acknowledge that study design based on LD between markers and unobserved etiological variant(s) will have limited power to detect multiple rare variants (e.g. copy number variants) that could also be functionally relevant. When investigating the association of single genes with the risk of a complex trait, likely to be governed by a considerable number of genes, ORs are almost always low to moderate. This reflects that a specific phenotype results of a

combination of different genes, each contributing only a small effect, and environmental factors. Therefore, it is appropriate to take in consideration the epistasis and gene-environment interactions to explain the remaining genetic risk to be identified for the nonsyndromic forms of orofacial clefts.

In conclusion, the results of this study suggest that several genes known to be involved in craniofacial morphogenesis and/or palatogenesis, each of which has a minor individual effect, may contribute to the incidence of orofacial clefts, but only few genes have a major role in the development of nonsyndromic clefts. The results of this study provide further evidence that variation in *FOXE1*, *TIMP2*, and the FGF and Wnt signaling pathway genes are likely involved in the development of nonsyndromic CL/P in North-Eastern European populations. Additionally, we replicated the recently reported genome-wide significant associations between 8q24 and 10q25 loci and CL/P in an independent Estonian case-control sample and this finding underlines that GWA studies are important mapping tools to highlight genomic regions and providing biological insights for further studies. Moreover, we have demonstrated that variation in cartilage collagen II and XI genes, *IRF6*, and the FGF and Wnt signaling pathway genes (both being different compared to CL/P) are likely involved in the etiology of CP in Northeastern European populations. The challenge is now to perform fine-mapping in these regions and identify genetic variants which are more likely to increase the susceptibility to clefting. However, the estimations of the total genetic contributions to the disease indicate that additional genetic factors involved in CL/P and CP remain to be identified, and both the functional effects of associated variants and the molecular mechanisms behind different pathways must still be ascertained. Further research using large, multicenter, collaborative studies is necessary to identify both genetic and environmental (including lifestyle) risk factors for nonsyndromic clefts and interactions between them.

CONCLUSIONS

Following conclusions can be drawn from the current PhD thesis:

- We have demonstrated that variation in *FGF1, FOXE1*, *TIMP2,* and *WNT9B* genes contributes susceptibility to nonsyndromic CL/P in North-Eastern European populations (Estonians, Latvians, Lithuanians).
- The genome-wide significant association between 8q24.21 locus and CL/P, reported in German population, was replicated in both independent Estonian and Lithuanian case-control samples.
- The genome-wide significant association between 10q25.3 locus and CL/P, recently reported in Central European populations, was replicated in an independent Estonian case-control sample.
- We have demonstrated that variation in *COL2A1*, *COL11A2*, *IRF6*, and *WNT3* genes contributes susceptibility to nonsyndromic CP in North-Eastern European populations (Estonians, Latvians, Lithuanians).
- Our results show that genetic variants in *MSX1* and *MTHFR* genes contribute risk to nonsyndromic CL/P in Estonians.
- The results of this study underline the importance of the FGF and Wnt signaling pathway genes in the etiology of both CL/P and CP.

These findings, together with previous studies have importance for enhancing our understanding of craniofacial development and will promote introduction of new approaches in the efforts to predict the regions in human genome implicated in genetic predisposition to nonsyndromic orofacial clefts. The majority of genetic variants predisposing to nonsyndromic oral clefts are to be discovered and larger samples from different ethnicities and precisely phenotyped cohorts of patients are required for comprehensive sequencing-based studies, which are targeting in addition to common SNPs also rare single nucleotide variants, CNVs and other structural variants, to account for more of the 'missing' heritability. The most benefitial approach in these studies, particularly in populations of European descent, will be collaboration between clefting research centers and population-based biobanks.

REFERENCES

Acs N, Bánhidy F, Puhó E, Czeizel AE. Maternal influenza during pregnancy and risk of congenital abnormalities in offspring. *Birth Defects Res A Clin Mol Teratol* 2005; 73: 989–996.

Albuisson J, Pêcheux C, Carel JC, Lacombe D, Leheup B, Lapuzina P, Bouchard P, Legius E, Matthijs G, Wasniewska M, Delpech M, Young J, Hardelin JP, Dodé C. Kallmann syndrome: 14 novel mutations in KAL1 and FGFR1 (KAL2). *Hum Mutat* 2005; 25: 98–99.

Alkuraya FS, Saadi I, Lund JJ, Turbe-Doan A, Morton CC, Maas RL. SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* 2006; 313: 1751.

Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; 322: 881–888.

Avila JR, Jezewski PA, Vieira AR, Orioli IM, Castilla EE, Christensen K, Daack-Hirsch S, Romitti PA, Murray JC. PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. *Am J Med Genet A* 2006; 140: 2562–2570.

Badovinac RL, Werler MM, Williams PL, Kelsey KT, Hayes C. Folic acid-containing supplement consumption during pregnancy and risk for oral clefts: a meta-analysis. *Birth Defects Res A Clin Mol Teratol* 2007; 79: 8–15.

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008; 40: 955–962.

Barrow LL, van Bokhoven H, Daack-Hirsch S, Andersen T, van Beersum SE, Gorlin R, Murray JC. Analysis of the p63 gene in classical EEC syndrome, related syndromes, and non-syndromic orofacial clefts. *J Med Genet* 2002; 39: 559–566.

Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesyan K, Deka R, Bradley DG, Shriver MD. Measuring European population stratification with microarray genotype data. *Am J Hum Genet* 2007; 80: 948–956.

Beaty TH, Hetmanski JB, Zeiger JS, Fan YT, Liang KY, VanderKolk CA, McIntosh I. Testing candidate genes for non-syndromic oral clefts using a case-parent trio design. *Genet Epidemiol* 2002; 22: 1–11.

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin SC, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF. A genome-wide association study of cleft lip with and

without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet* 2010; 42: 525–529.

Beiraghi S, Nath SK, Gaines M, Mandhyan DD, Hutchings D, Ratnamala U, McElreavey K, Bartoloni L, Antonarakis GS, Antonarakis SE, Radhakrishna U. Autosomal dominant nonsyndromic cleft lip and palate: significant evidence of linkage at 18q21.1. *Am J Hum Genet* 2007; 81: 180–188.

Bender PL. Genetics of cleft lip and palate. *J Pediatr Nurs* 2000; 15: 242–249.

Bille C, Olsen J, Vach W, Knudsen VK, Olsen SF, Rasmussen K, Murray JC, Andersen AM, Christensen K. Oral clefts and life style factors – a case cohort study based on prospective Danish data. *Eur J Epidemiol* 2007; 22: 173–181.

Bille C, Winther JF, Bautz A, Murray JC, Olsen J, Christensen K. Cancer risk in persons with oral cleft – a population-based study of 8,093 cases. *Am J Epidemiol* 161: 1047–1055.

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Bergé SJ, Reich RH, Schiefke F, Hemprich A, Pötzsch S, Steegers-Theunissen RP, Pötzsch B, Moebus S, Horsthemke B, Kramer FJ, Wienker TF, Mossey et al. PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nöthen MM, Mangold E. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet* 2009a; 41: 473–477.

Birnbaum S, Ludwig KU, Reutter H, Herms S, de Assis NA, Diaz-Lacava A, Barth S, Lauster C, Schmidt G, Scheer M, Saffar M, Martini M, Reich RH, Schiefke F, Hemprich A, Pötzsch S, Pötzsch B, Wienker TF, Hoffmann P, Knapp M, Kramer FJ, Nöthen MM, Mangold E. IRF6 gene variants in Central European patients with non-syndromic cleft lip with or without cleft palate. *Eur J Oral Sci* 2009b; 117: 766–769.

Birnbaum S, Reutter H, Mende M, de Assis NA, Diaz-Lacava A, Herms S, Scheer M, Lauster C, Braumann B, Schmidt G, Martini M, Hemprich A, Pötzsch S, Knapp M, Nöthen MM, Kramer FJ, Mangold E. Further evidence for the involvement of MYH9 in the etiology of non-syndromic cleft lip with or without cleft palate. *Eur J Oral Sci* 2009c; 117: 200–203.

Bjork BC, Turbe-Doan A, Prysak M, Herron BJ, Beier DR. Prdm16 is required for normal palatogenesis in mice. *Hum Mol Genet* 2010; 19: 774–789.

Blavier L, Lazaryev A, Groffen J, Heisterkamp N, Declerk YA, Kaartinen V. TGFB3-induced palatogenesis requires matrix metalloproteinases. *Mol Biol Cell* 2001; 12: 1457–1466.

Boyles AL, DeRoo LA, Lie RT, Taylor JA, Jugessur A, Murray JC, Wilcox AJ. Maternal alcohol consumption, alcohol metabolism genes, and the risk of oral clefts: a population-based case-control study in Norway, 1996–2001. *Am J Epidemiol* 2010; 172: 924–931.

Boyles AL, Wilcox AJ, Taylor JA, Shi M, Weinberg CR, Meyer K, Fredriksen A, Ueland PM, Johansen AM, Drevon CA, Jugessur A, Trung TN, Gjessing HK, Vollset SE, Murray JC, Christensen K, Lie RT. Oral facial clefts and gene polymorphisms in metabolism of folate/one-carbon and vitamin A: a pathway-wide association study. *Genet Epidemiol* 2009; 33: 247–255.

Braybrook C, Doudney K, Marçano AC, Arnason A, Bjornsson A, Patton MA, Goodfellow PJ, Moore GE, Stanier P. The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. *Nat Genet* 2001; 29: 179–183.

Brown NL, Yarram SJ, Mansell JP, Sandy JR. Matrix metalloproteinases have a role in palatogenesis. *J Dent Res* 2002; 81: 826–830.

Brugmann SA, Goodnough LH, Gregorieff A, Leucht P, ten Berge D, Fuerer C, Clevers H, Nusse R, Helms JA. Wnt signaling mediates regional specification in the vertebrate face. *Development* 2007; 134: 3283–3295.

Calzolari E, Bianchi F, Rubini M, Ritvanen A, Neville AJ; EUROCAT Working Group. Epidemiology of cleft palate in Europe: implications for genetic research. *Cleft Palate Craniofac J* 2004; 41: 244–249.

Calzolari E, Pierini A, Astolfi G, Bianchi F, Neville AJ, Rivieri F. Associated anomalies in multi-malformed infants with cleft lip and palate: An epidemiologic study of nearly 6 million births in 23 EUROCAT registries. *Am J Med Genet A* 2007; 143: 528–537.

Canfield MA, Collins JS, Botto LD, Williams LJ, Mai CT, Kirby RS, Pearson K, Devine O, Mulinare J; National Birth Defects Preventin Network. Changes in the birth prevalence of selected birth defects after grain fortification with folic acid in the United States: findings from a multi-state population-based study. *Birth Defects Res A Clin Mol Teratol* 2005; 73: 679–689.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; 74: 106–120.

Carter TC, Molloy AM, Pangilinan F, Troendle JF, Kirke PN, Conley MR, Orr DJ, Earley M, McKiernan E, Lynn EC, Doyle A, Scott JM, Brody LC, Mills JL. Testing reported associations of genetic risk factors for oral clefts in a large Irish study population. *Birth Defects Res A Clin Mol Teratol* 2010; 88: 84–93.

Cedergren M, Källén B. Maternal obesity and the risk for orofacial clefts in the offspring. *Cleft Palate Craniofac J* 2005; 42: 367–371.

Chevrier C, Perret C, Bahuau M, Nelva A, Herman C, Francannet C, Robert-Gnansia E, Cordier S. Interaction between the ADH1C polymorphism and maternal alcohol intake in the risk of nonsyndromic oral clefts: an evaluation of the contribution of child and maternal genotypes. *Birth Defects Res A Clin Mol Teratol* 2005; 73: 114–122.

Chevrier C, Perret C, Bahuau M, Zhu H, Nelva A, Herman C, Francannet C, Robert-Gnansia E, Finnell RH, Cordier S. Fetal and maternal MTHFR C677T genotype, maternal folate intake and the risk of nonsyndromic oral clefts. *Am J Med Genet A* 2007; 143: 248–257.

Chiquet BT, Blanton SH, Burt A, Ma D, Stal S, Mulliken JB, Hecht JT. Variation in WNT genes is associated with non-syndromic cleft lip with or without cleft palate. *Hum Mol Genet* 2008; 17: 2212–2218.

Chiquet BT, Lidral AC, Stal S, Mulliken JB, Moreno LM, Arcos-Burgos M, Valencia-Ramirez C, Blanton SH, Hecht JT. CRISPLD2: a novel NSCLP candidate gene. *Hum Mol Genet* 2007; 16: 2241–2248.

Choi SJ, Marazita ML, Hart PS, Sulima PP, Field LL, McHenry TG, Govil M, Cooper ME, Letra A, Menezes R, Narayanan S, Mansilla MA, Granjeiro JM, Vieira AR, Lidral AC, Murray JC, Hart TC. The PDGF-C regulatory region SNP rs28999109 decreases promoter transcriptional activity and is associated with CL/P. *Eur J Hum Genet 2009*; 17: 774–784.

Christensen K, Juel K, Herskind AM, Murray JC. Long term follow up study of survival associated with cleft lip and palate at birth. *BMJ* 2004; 328: 1405.

Christensen K, Mortensen PB. Facial clefting and psychiatric diseases: a follow-up of the Danish 1936–1987 Facial Cleft cohort. *Cleft Palate Craniofac J* 2002; 39: 392–396.

Christensen K, Olsen J, Nørgaard-Pedersen B, Basso O, Støvring H, Milhollin-Johnson L, Murray JC. Oral clefts, transforming growth factor alpha gene variants, and maternal smoking: a population-based case-control study in Denmark, 1991–1994. *Am J Epidemiol* 1999; 149: 248–255.

Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005; 37: 1243–1246.

Cobourne MT. The complex genetics of cleft lip and palate. *Eur J Orthod* 2004; 26: 7–16.

Cox TC. Taking it to the max: the genetic and developmental mechanisms coordinating midfacial morphogenesis and dysmorphology. *Clin Genet* 2004; 65: 163–176.

Croen LA, Shaw GM, Wasserman CR, Tolarová MM. Racial and ethnic variations in the prevalence of orofacial clefts in California. *Am J Med Genet* 1998; 79: 42–47.

de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, Onofrio RC, Lyon HN, Stram DO, Haiman CA, Freedman ML, Zhu X, Cooper R, Groop L, Kolonel LN, Henderson BE, Daly MJ, Hirschhorn JN, Altshuler D. Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 2006; 38: 1298–1303.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005; 37: 1217–1223.

DeRoo LA, Wilcox AJ, Drevon CA, Lie RT. First-trimester maternal alcohol consumption and the risk of infant oral clefts in Norway: a population-based case-control study. *Am J Epidemiol* 2008; 168: 638–646.

Devlin B, Roeder K, Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 60: 155–166.

Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 2011; 12: 167–178.

Dodé C, Levilliers J, Dupont JM, De Paepe A, Le Dû N, Soussi-Yanicostas N, Coimbra RS, Delmaghani S, Compain-Nouaille S, Baverel F, Pêcheux C, Le Tessier D, Cruaud C, Delpech M, Speleman F, Vermeulen S, Amalfitano A, Bachelot Y, Bouchard P, Cabrol S, Carel JC, Delemarre-van de Waal H, Goulet-Salmon B, Kottler ML, Richard O, Sanchez-Franco F, Saura R, Young J, Petit C, Hardelin JP. Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat Genet* 2003; 33: 463–465.

Druschel CM, Hughes JP, Olsen CL. First year-of-life mortality among infants with oral clefts: New York State, 1983–1990. *Cleft Palate Craniofac J* 1996; 33: 400–405.

Fallin MD, Hetmanski JB, Park J, Scott AF, Ingersoll R, Fuernkranz HA, McIntosh I, Beaty TH. Family-based analysis of MSX1 haplotypes for association with oral clefts. *Genet Epidemiol* 2003; 25: 168–175.

Fitzpatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H, Markham AF, Fantes JA, Bonthron DT. Identification of SATB2 as the cleft palate gene on 2q32-q33. *Hum Mol Genet* 2003; 12: 2491–2501.

Fitzpatrick DR, Denhez F, Kondaiah P, Akhurst RJ. Differential expression of TGF beta isoforms in murine palatogenesis. *Development* 1990; 109: 585–595.
Fraser FC, Gwyn A. Seasonal variation in birth date of children with cleft lip. *Teratology* 1998; 57: 93–95.
García AM. Occupational exposure to pesticides and congenital malformations: a review of mechanisms, methods, and results. *Am J Ind Med* 1998; 33: 232–240.
Gorlin RJ, Cohen MM, Levin LS. Syndromes of the head and neck (3rd ed.). New York: Oxford University Press, 1990.
Grant SF, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah HD, Kirschner RE, Hakonarson H. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr* 2009; 155: 909–913.
Greene RM, Nugent P, Mukhopadhyay P, Warner DR, Pisano MM. Intracellular dynamics of Smad-mediated TGFbeta signaling. *J Cell Physiol* 2003; 197: 261–271.
Gritli-Linde A. Molecular control of secondary palate development. *Dev Biol* 2007; 301: 309–326.
Gritli-Linde A. p63 and IRF6: brothers in arms against cleft palate. *J Clin Invest* 2010; 120: 1386–1389.
Grosen D, Bille C, Pedersen JK, Skytthe A, Murray JC, Christensen K. Recurrence risk for offspring of twins discordant for oral cleft: a population-based cohort study of the Danish 1936–2004 cleft twin cohort. *Am J Med Genet A* 2010; 152A: 2468–2474.
Hallonet M., Hollemann T., Pieler T., Gruss P. Vax1, a novel homeobox-containing gene, directs development of the basal forebrain and visual system. *Genes Dev* 1999; 13): 3106–3114.
Hartsfield JK Jr, Hickman TA, Everett ET, Shaw GM, Lammer EJ, Finnell RA. Analysis of the EPHX1 113 polymorphism and GSTM1 homozygous null polymorphism and oral clefting associated with maternal smoking. *Am J Med Genet* 2001; 102: 21–24.
Harville EW, Wilcox AJ, Lie RT, Vindenes H, Abyholm F. Cleft lip and palate versus cleft lip only: are they distinct defects? *Am J Epidemiol* 2005; 162: 448–453.
Hayes C. Environmental risk factors and oral clefts. In: Wyszynski DF, ed. Cleft lip and palate: from origin to treatment. New York: Oxford University Press, 2002: 159–169.
Hecht JT, Mulliken JB, Blanton SH. Evidence for a cleft palate only locus on chromosome 4 near MSX1. *Am J Med Genet* 2002; 110: 406–407.
Hernández-Díaz S, Werler MM, Walker AM, Mitchell AA. Folic acid antagonists during pregnancy and the risk of birth defects. *N Engl J Med* 2000; 343: 1608–1614.
Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009; 106: 9362–9367.
Hirano T, Ike F, Murata T, Obata Y, Utiyama H, Yokoyama KK. Genes encoded within 8q24 on the amplicon of a large extrachromosomal element are selectively repressed during the terminal differentiation of HL–60 cells. *Mutat Res* 2008; 640, 97–106.

Honein MA, Rasmussen SA, Reefhuis J, Romitti PA, Lammer EJ, Sun L, Correa A. Maternal smoking and environmental tobacco smoke exposure and the risk of orofacial clefts. *Epidemiology* 2007; 18: 226–233.

International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; 437: 1299–1320.

Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* 2007; 2: e841.

Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 2009; 10: 318–329.

Jehee FS, Burin BA, Rocha KM, Zechi-Ceide R, Bueno DF, Brito L, Souza J, Leal GF, Richieri-Costa A, Alonso N, Otto PA, Passos-Bueno MR. Novel mutations in IRF6 in nonsyndromic cleft lip with or without cleft palate: when should IRF6 mutational screening be done? *Am J Med Genet A* 2009; 149A: 1319–1322.

Jentink J, Loane MA, Dolk H, Barisic I, Garne E, Morris JK, de Jong-van den Berg LT; EUROCAT Antiepileptic Study Working Group. Valproic acid monotherapy in pregnancy and major congenital malformations. *N Engl J Med* 2010; 362: 2185–2193.

Jezewski PA, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE, Daack-Hirsch S, Schultz RE, Weber A, Nepomucena B, Romitti PA, Christensen K, Orioli IM, Castilla EE, Machida J, Natsume N, Murray JC. Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J Med Genet* 2003; 40: 399–407.

Johansen AM, Lie RT, Wilcox AJ, Andersen LF, Drevon CA. Maternal dietary intake of vitamin A and risk of orofacial clefts: a population-based case-control study in Norway. *Am J Epidemiol* 2008; 167: 1164–1170.

Johnson CY, Little J. Folate intake, markers of folate status and oral clefts: is the evidence converging? *Int J Epidemiol* 2008; 37: 1041–1058.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; 29: 233–237.

Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Hypothesis-driven candidate gene association studies: practical design and analytical considerations. *Am J Epidemiol* 2009; 170: 986–993.

Jugessur A, Murray JC. Orofacial clefting: recent insights into a complex trait. *Curr Opin Genet Dev* 2005; 15: 270–278.

Jugessur A, Shi M, Gjessing HK, Lie RT, Wilcox AJ, Weinberg CR, Christensen K, Boyles AL, Daack-Hirsch S, Trung TN, Bille C, Lidral AC, Murray JC. Genetic determinants of facial clefting: analysis of 357 candidate genes using two national cleft studies from Scandinavia. *PLoS One* 2009; 4: e5385.

Jugessur A, Wilcox AJ, Lie RT, Murray JC, Taylor JA, Ulvik A, Drevon CA, Vindenes HA, Abyholm FE. Exploring the effects of methylenetetrahydrofolate reductase gene variants C677T and A1298C on the risk of orofacial clefts in 261 Norwegian case-parent triads. *Am J Epidemiol* 2003; 157: 1083–1091.

Juriloff DM, Harris MJ, Dewell SL, Brown CJ, Mager DL, Gagnier L, Mah DG. Investigations of the genomic region that contains the clf1 mutation, a causal gene in multifactorial cleft lip and palate in mice. *Birth Defects Res A Clin Mol Teratol* 2005; 73: 103–113.

Juriloff DM, Harris MJ, McMahon AP, Carroll TJ, Lidral AC. Wnt9b is the mutated gene involved in multifactorial nonsyndromic cleft lip with or without cleft palate in A/WySn mice, as confirmed by a genetic complementation test. *Birth Defects Res A Clin Mol Teratol* 2006; 76: 574–579.

Kaartinen V, Cui XM, Heisterkamp N, Groffen J, Shuler CF. Transforming growth factor-beta3 regulates transdifferentiation of medial edge epithelium during palatal fusion and associated degradation of the basement membrane. *Dev Dyn* 1997; 209: 255–260.

Kaartinen V, Voncken JW, Shuler C, Warburton D, Bu D, Heisterkamp N, Groffen J. Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction. *Nat Genet* 1995; 11: 415–421.

Kimonis VE, Goldstein AM, Pastakia B, Yang ML, Kase R, DiGiovanna JJ, Bale AE, Bale SJ. Clinical manifestations in 105 persons with nevoid basal cell carcinoma syndrome. *Am J Med Genet* 1997; 69: 299–308.

Knight AS, Schutte BC, Jiang R, Dixon MJ. Developmental expression analysis of the mouse and chick orthologues of IRF6: the gene mutated in Van der Woude syndrome. *Dev Dyn* 2006; 235: 1441–1447.

Koillinen H, Lahermo P, Rautio J, Hukki J, Peyrard-Janvid M, Kere J. A genome-wide scan of non-syndromic cleft palate only (CPO) in Finnish multiplex families. *J Med Genet* 2005; 42: 177–184.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet* 2002; 32: 285–289.

Krapels IP, Rooij IA, Wevers RA, Zielhuis GA, Spauwen PH, Brussel W, Steegers-Theunissen RP. Myo-inositol, glucose and zinc status as risk factors for non-syndromic cleft lip with or without cleft palate in offspring: a case-control study. *BJOG* 2004; 111: 661–668.

Krjutškov K, Andreson R, Mägi R, Nikopensius T, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A. Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. *Nucleic Acids Res* 2008; 36: e75.

Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* 2006; 7: 385–394.

Lammer EJ, Shaw GM, Iovannisci DM, Van Waes J, Finnell RH. Maternal smoking and the risk of orofacial clefts: Susceptibility with NAT1 and NAT2 polymorphisms. *Epidemiology* 2004; 15: 150–156.

Lan Y, Ryan RC, Zhang Z, Bullard SA, Bush JO, Maltby KM, Lidral AC, Jiang R. Expression of Wnt9b and activation of canonical Wnt signaling during midfacial morphogenesis in mice. *Dev Dyn* 2006; 235: 1448–1454.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, /.../ O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. Hundreds of variants clustered in

genomic loci and biological pathways affect human height. *Nature* 2010; 467: 832–838.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Rüther A, Schreiber S, Becker C, Nürnberg P, Nelson MR, Krawczak M, Kayser M. Correlation between genetic and geographic structure in Europe. *Curr Biol* 2008; 18: 1241–1248.

Letra A, da Silva RA, Menezes R, de Souza AP, de Almeida AL, Sogayar MC, Granjeiro JM. Studies with MMP9 gene promoter polymorphism and nonsyndromic cleft lip and palate. *Am J Med Genet A* 2007; 143: 89–91.

Lettieri J. Lip and oral cavity. In: Stevenson RE, Hall JG, & Goodman RM, eds. Human malformations and related anomalies. New York: Oxford University Press, 1993: 367–374.

Li Q, Yu K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* 2008; 32: 215–226.

Lidral AC, Moreno LM. Progress toward discerning the genetics of cleft lip. *Curr Opin Pediatr* 2005; 17: 731–739.

Lidral AC, Murray JC. Genetic approaches to identify disease genes for birth defects with cleft lip/palate as a model. *Birth Defects Res A Clin Mol Teratol* 2004; 70: 893–901.

Lidral AC, Romitti PA, Basart AM, Doetschman T, Leysens NJ, Daack-Hirsch S, Semina EV, Johnson LR, Machida J, Burds A, Parnell TJ, Rubenstein JL, Murray JC. Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. *Am J Hum Genet* 1998; 63: 557–568.

Lie RT, Wilcox AJ, Skjaerven R. Survival and reproduction among males with birth defects and risk of recurrence in their children. *JAMA* 2001; 285: 755–760.

Little J, Cardy A, Munger RG. Tobacco smoking and oral clefts: a meta-analysis. *Bull World Health Organ* 2002; 82: 213–218.

Little J, Gilmour M, Mossey PA, Fitzpatrick D, Cardy A, Clayton-Smith J, Fryer AE; ITS MAGIC collaboration. Folate and clefts of the lip and palate – a U.K.-based case-control study: Part I: Dietary and supplemental folate. *Cleft Palate Craniofac J* 2008a; 45: 420–427.

Little J, Gilmour M, Mossey PA, Fitzpatrick D, Cardy A, Clayton-Smith J, Hill A, Duthie SJ, Fryer AE, Molloy AM, Scott JM; ITS MAGIC Collaboration. Folate and clefts of the lip and palate – a U.K.-based case-control study: Part II: Biochemical and genetic analysis. *Cleft Palate Craniofac J* 2008b; 45: 428–438.

López-Camelo JS, Castilla EE, Orioli IM; INAGEMP (Instituto Nacional de Genética Médica Populacional); ECLAMC (Estudio Colaborativo Latino Americano de Malformaciones Congénitas). Folic acid flour fortification: impact on the frequencies of 52 congenital anomaly types in three South American countries. *Am J Med Genet A* 2010; 152A: 2444–2458.

Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, Liang KY, Duffy DL, VanderKolk C. Application of transmission disequilibrium tests to nonsyndromic oral clefts: including candidate genes and environmental exposures in the models. *Am J Med Genet* 1997; 73: 337–344.

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, de Assis NA, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Pötzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann HE, Steegers-Theunissen RP, Kramer FJ, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey et al. PA, Hoffmann P, Nöthen MM. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet* 2010; 42: 24–26.

Mangold E, Reutter H, Birnbaum S, Walier M, Mattheisen M, Henschke H, Lauster C, Schmidt G, Schiefke F, Reich RH, Scheer M, Hemprich A, Martini M, Braumann B, Krimmel M, Opitz C, Lenz JH, Kramer FJ, Wienker TF, Nöthen MM, Diaz Lacava A. Genome-wide linkage scan of nonsyndromic orofacial clefting in 91 families of central European origin. *Am J Med Genet A* 2009; 149A: 2680–2694.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747–753.

Mansilla MA, Cooper ME, Goldstein T, Castilla EE, Lopez Camelo JS, Marazita ML, Murray JC. Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofac J* 2006; 43: 21–29.

Marazita ML, Lidral AC, Murray JC, Field LL, Maher BS, Goldstein McHenry T, Cooper ME, Govil M, Daack-Hirsch S, Riley B, Jugessur A, Felix T, Morene L, Mansilla MA, Vieira AR, Doheny K, Pugh E, Valencia-Ramirez C, Arcos-Burgos M. 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* 2009; 68: 151–170.

Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS, Daack-Hirsch S, Schultz R, Mansilla MA, Field LL, Liu YE, Prescott N, Malcolm S, Winter R, Ray A, Moreno L, Valencia C, Neiswanger K, Wyszynski DF, Bailey-Wilson JE, Albacha-Hejazi H, Beaty TH, McIntosh I, Hetmanski JB, Tunçbilek G, Edwards M, Harkin L, Scott R, Roddick LG. Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *Am J Hum Genet* 2004; 75: 161–173.

Marçano AC, Doudney K, Braybrook C, Squires R, Patton MA, Lees MM, Richieri-Costa A, Lidral AC, Murray JC, Moore GE, Stanier P. TBX22 mutations are a frequent cause of cleft palate. *J Med Genet* 2004; 41: 68–74.

Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet* 2004; 36: 512–517.

Martinelli M, Di Stazio M, Scapoli L, Marchesini J, Di Bari F, Pezzetti F, Carinci F, Palmieri A, Carinci P, Savoia A. Cleft lip with or without cleft palate: implication of the heavy chain of non-muscle myosin IIA. *J Med Genet* 2007; 44: 387–392.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; 9: 356–369.

Melkoniemi M, Koillinen H, Männikkö M, Warman ML, Pihlajamaa T, Kääriäinen H, Rautio J, Hukki J, Stofko JA, Cisneros GJ, Krakow D, Cohn DH, Kere J, Ala-

Kokko L. Collagen XI sequence variations in nonsyndromic cleft palate, Robin sequence and micrognathia. *Eur J Hum Genet* 2003; 11:265–270.

Menezes R, Letra A, Kim AH, Küchler EC, Day A, Tannure PN, Gomes da Motta L, Paiva KB, Granjeiro JM, Vieira AR. Studies with Wnt genes and nonsyndromic cleft lip and palate. *Birth Defects Res A Clin Mol Teratol* 2010; 88: 995–1000.

Meyer KA, Werler MM, Hayes C, Mitchell AA. Low maternal alcohol consumption during pregnancy and oral clefts in offspring: the Slone Birth Defects Study. *Birth Defects Res A Clin Mol Teratol* 2003; 67: 509–514.

Miettinen PJ, Chin JR, Shum L, Slavkin HC, Shuler CF, Derynck R, Werb Z. Epidermal growth factor receptor function is necessary for normal craniofacial development and palate closure. *Nat Genet* 1999; 22: 69–73.

Mills JL, Molloy AM, Parle-McDermott A, Troendle JF, Brody LC, Conley MR, Cox C, Pangilinan F, Orr DJ, Earley M, McKiernan E, Lynn EC, Doyle A, Scott JM, Kirke PN. Folate-related gene polymorphisms as risk factors for cleft lip and cleft palate. *Birth Defects Res A Clin Mol Teratol* 2008; 82: 636–643.

Mitchell LE, Murray JC, O'Brien S, Christensen K. Evaluation of two putative susceptibility loci for oral clefts in the Danish population. *Am J Epidemiol* 2001; 153: 1007–1015.

Mitchell LE, Murray JC, O'Brien S, Christensen K. Retinoic acid receptor alpha gene variants, multivitamin use, and liver intake as risk factors for oral clefts: a population-based case-control study in Denmark, 1991–1994. *Am J Epidemiol* 2003; 158: 69–76.

Mogass M, Bringas P Jr, Shuler CF. Characterization of desmosomal component expression during palatogenesis. *Int J Dev Biol* 2000; 44: 317–322.

Montpetit A, Nelis M, Laflamme P, Mägi R, Ke X, Remm M, Cardon L, Hudson TJ, Metspalu A. An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. *PLoS Genet* 2006; 2: e27.

Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S, L'heureux J, Valencia-Ramirez C, Rivera D, López AM, Moreno MA, Hing A, Lammer EJ, Jones M, Christensen K, Lie RT, Jugessur A, Wilcox AJ, Chines P, Pugh E, Doheny K, Arcos-Burgos M, Marazita ML, Murray JC, Lidral AC. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum Mol Genet* 2009; 18: 4879–4896.

Mossey PA, Castilla EE. Global registry and database on craniofacial anomalies. Geneva: World Health Organization, 2003.

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *Lancet* 2009; 374: 1773–1785.

Mossey PA, Little J. Epidemiology of oral clefts: an international perspective. In: Wyszynski DF, ed. Cleft lip and palate: from origin to treatment. New York: Oxford University Press, 2002: 127–158.

Mueller JC, Lõhmussaar E, Mägi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schreiber S, Pramstaller P, Pichler I, Romeo G, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T. Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am J Hum Genet* 2005; 76: 387–398.

Muenke, M. The pit, the cleft and the web. *Nat Genet* 2002; 32: 219–220.

Munger RG, Sauberlich HE, Corcoran C, Nepomuceno B, Daack-Hirsch S, Solon FS. Maternal vitamin B–6 and folate status and risk of oral cleft birth defects in the Philippines. *Birth Defects Res A Clin Mol Teratol* 2004; 70: 464–471.

Munger RG, Tamura T, Johnston KE, Feldkamp ML, Pfister R, Carey JC. Plasma zinc concentrations of mothers and the risk of oral clefts in their children in Utah. *Birth Defects Res A Clin Mol Teratol* 2009; 85: 151–155.

Murray JC, Daack-Hirsch S, Buetow KH, Munger R, Espina L, Paglinawan N, Villanueva E, Rary J, Magee K, Magee W. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J* 1997; 34: 7–10.

Murray JC, Schutte BC. Cleft palate: players, pathways, and pursuits. *J Clin Invest* 2004; 113: 1676–1678.

Murray JC. 1995. Face facts: genes, environment, and clefts. *Am J Hum Genet* 57: 227–232.

Murray JC. Gene/environment causes of cleft lip and/or palate. *Clin Genet* 2002; 61: 248–256.

Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskácková T, Balascák I, Peltonen L, Jakkula E, Rehnström K, Lathrop M, Heath S, Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgár N, Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V, Kasnauskiene J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet R, Marsal S, Julià A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M, Macek M, Krawczak M, Remm M, Metspalu A. Genetic structure of Europeans: a view from the North-East. *PLoS One* 2009; 4: e5472.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010a; 42: 30–35.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010b; 42: 790–793.

Nie X, Luukko K, Kettunen P. FGF signalling in craniofacial development and developmental disorders. *Oral Dis* 2006; 12: 102–111.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. Genes mirror geography within Europe. *Nature* 2008; 456: 98–101.

Osoegawa K, Vessere GM, Utami KH, Mansilla MA, Johnson MK, Riley BM, L'Heureux J, Pfundt R, Staaf J, van der Vliet WA, Lidral AC, Schoenmakers EF, Borg A, Schutte BC, Lammer EJ, Murray JC, de Jong PJ. Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. *J Med Genet* 2008; 45: 81–86.

Park BY, Sull JW, Park JY, Jee SH, Beaty TH. Differential parental transmission of markers in BCL3 among Korean cleft case-parent trios. *J Prev Med Public Health* 2009; 42: 1–4.

Park JW, McIntosh I, Hetmanski JB, Jabs EW, Vander Kolk CA, Wu-Chou YH, Chen PK, Chong SS, Yeow V, Jee SH, Park BY, Fallin MD, Ingersoll R, Scott AF, Beaty TH. Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. *Genet Med* 2007; 9: 219–227.

Park-Wyllie L, Mazzotta P, Pastuszak A, Moretti ME, Beique L, Hunnisett L, Friesen MH, Jacobson S, Kasapinovic S, Chang D, Diav-Citrin O, Chitayat D, Nulman I, Einarson TR, Koren G. Birth defects after maternal exposure to corticosteroids:

prospective cohort study and meta-analysis of epidemiological studies. *Teratology* 2000; 62: 385–392.

Pauws E, Moore GE, Stanier P. A functional haplotype variant in the TBX22 promoter is associated with cleft palate and ankyloglossia. *J Med Genet* 2009; 46: 555–561.

Pauws E, Stanier P. FGF signalling and SUMO modification: new players in the aetiology of cleft lip and/or palate. *Trends Genet* 2007; 23: 631–640.

Pezzetti F, Martinelli M, Scapoli L, Carinci F, Palmieri A, Marchesini J, Carinci P, Caramelli E, Rullo R, Gombos F, Tognon M. Maternal MTHFR variant forms increase the risk in offspring of isolated nonsyndromic cleft lip with or without cleft palate. *Hum Mutat* 2004; 24: 104–105.

Porter F. Cholesterol precursors and facial clefting. *J Clin Invest* 2006; 116: 2322–2325.

Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P, Seligsohn U, Waliszewska A, Schirmer C, Ardlie K, Ramos A, Nemesh J, Arbeitman L, Goldstein DB, Reich D, Hirschhorn JN. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 2008; 4: e236.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–909.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet* 2000; 67: 170–181.

Proetzel G, Pawlowski SA, Wiles MV, Yin M, Boivin GP, Howles PN, Ding J, Ferguson MW, Doetschman T. Transforming growth factor-beta 3 is required for secondary palate fusion. *Nat Genet* 1995; 11: 409–414.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; 81: 559–575.

Radhakrishna U, Ratnamala U, Gaines M, Beiraghi S, Hutchings D, Golla J, Husain SA, Gambhir PS, Sheth JJ, Sheth FJ, Chetan GK, Naveed M, Solanki JV, Patel UC, Master DC, Memon R, Antonarakis GS, Antonarakis SE, Nath SK. Genomewide scan for nonsyndromic cleft lip and palate in multigenerational Indian families reveals significant evidence of linkage at 13q33.1–34. *Am J Hum Genet* 2006; 79: 580–585.

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, Mossey PA, Little J, Steegers-Theunissen RP, Pennacchio LA, Schutte BC, Murray JC. Disruption of an AP–2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet* 2008; 40: 1341–1347.

Rice R, Spencer-Dene B, Connor EC, Gritli-Linde A, McMahon AP, Dickson C, Thesleff I, Rice DP. Disruption of Fgf10/Fgfr2b-coordinated epithelial-mesenchymal interactions causes cleft palate. *J Clin Invest* 2004; 113: 1692–1700.

Richardson RJ, Dixon J, Jiang R, Dixon MJ. Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. Hum Mol Genet 2009; 18: 2632–2642.

Richman JM, Lee SH. About face: signals and genes controlling jaw patterning and identity in vertebrates. *Bioessays* 2003; 25: 554–568.

Riley BM, Mansilla MA, Ma J, Daack-Hirsch S, Maher BS, Raffensperger LM, Russo ET, Vieira AR, Dodé C, Mohammadi M, Marazita ML, Murray JC. Impaired FGF

signaling contributes to cleft lip and palate. *Proc Natl Acad Sci USA* 2007a; 104: 4512–4517.

Riley BM, Murray JC. Sequence evaluation of *FGF* and *FGFR* gene conserved non-coding elements in non-syndromic cleft lip and palate cases. *Am J Med Genet A* 2007b; 143A: 3228–3234.

Rittler M, López-Camelo J, Castilla EE. Sex ratio and associated risk factors for 50 congenital anomaly types: clues for causal heterogeneity. *Birth Defects Res A Clin Mol Teratol* 2004; 70: 13–19.

Romitti PA, Lidral AC, Munger RG, Daack-Hirsch S, Burns TL, Murray JC. Candidate genes for nonsyndromic cleft lip and palate and maternal cigarette smoking and alcohol consumption: evaluation of genotype-environment interactions from a population-based case-control study of orofacial clefts. *Teratology* 1999; 59: 39–50.

Romitti PA, Sun L, Honein MA, Reefhuis J, Correa A, Rasmussen SA. Maternal periconceptional alcohol consumption and risk of orofacial clefts. *Am J Epidemiol* 2007; 166: 775–785.

Rothman KJ, Moore LL, Singer MR, Nguyen US, Mannino S, Milunsky A. Teratogenicity of high vitamin A intake. *N Engl J Med* 1995; 333: 1369–1373.

Satokata I, Maas R. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet* 1994; 6: 348–356.

Scapoli L, Palmieri A, Martinelli M, Vaccari C, Marchesini J, Pezzetti F, Baciliero U, Padula E, Carinci P, Carinci F. Study of the PVRL1 gene in Italian nonsyndromic cleft lip patients with or without cleft palate. *Ann Hum Genet* 2006; 70: 410–413.

Schliekelman P, Slatkin M. Multiplex relative risk and estimation of the number of loci underlying an inherited disease. *Am J Hum Genet* 2002; 71: 1369–1385.

Shaw GM, Nelson V, Iovannisci DM, Finnell RH, Lammer EJ. Maternal occupational chemical exposures and biotransformation genotypes as risk factors for selected congenital anomalies. *Am J Epidemiol* 2003; 157: 475–484.

Shaw GM, Wasserman CR, Lammer EJ, O'Malley CD, Murray JC, Basart AM, Tolarova MM. Orofacial clefts, parental cigarette smoking, and transforming growth factor-alpha gene variants. *Am J Hum Genet* 1996; 58: 551–561.

Shaw GM, Wasserman CR, Murray JC, Lammer EJ. Infant TGF-alpha genotype, orofacial clefts, and maternal periconceptional multivitamin use. *Cleft Palate Craniofac J* 1998; 35: 366–370.

Shaw GM, Wasserman CR, O'Malley CD, Lammer EJ, Finnell RH. Orofacial clefts and maternal anticonvulsant use. *Reprod Toxicol* 1995; 9: 97–98.

Shi M, Christensen K, Weinberg CR, Romitti P, Bathum L, Lozada A, Morris RW, Lovett M, Murray JC. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *Am J Hum Genet* 2007; 80: 76–90.

Shi M, Mostowska A, Jugessur A, Johnson MK, Mansilla MA, Christensen K, Lie RT, Wilcox AJ, Murray JC. Identification of microdeletions in candidate genes for cleft lip and/or palate. *Birth Defects Res A Clin Mol Teratol* 2009; 85: 42–51.

Shi M, Wehby GL, Murray JC. Review on genetic variants and maternal smoking in the etiology of oral clefts and other birth defects. *Birth Defects Res C Embryo Today* 2008; 84: 16–29.

Shiang R, Lidral AC, Ardinger HH, Buetow KH, Romitti PA, Munger RG, Murray JC. Association of transforming growth-factor alpha gene polymorphisms with nonsyndromic cleft palate only (CPO). *Am J Hum Genet* 1993; 53: 836–843.

Sivertsen A, Wilcox A, Johnson GE, Abyholm F, Vindenes HA, Lie RT. Prevalence of major anatomic variations in oral clefts. *Plast Reconstr Surg* 2008a; 121: 587–595.

Sivertsen A, Wilcox AJ, Skjaerven R, Vindenes HA, Abyholm F, Harville E, Lie RT. Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *BMJ* 2008b; 336: 432–434.

Song L, Li Y, Wang K, Wang YZ, Molotkov A, Gao L, Zhao T, Yamagami T, Wang Y, Gan Q, Pleasure DE, Zhou CJ. Lrp6-mediated canonical Wnt signaling is required for lip formation and fusion. *Development* 2009; 136: 3161–3171.

Sperber GH. Formation of the primary and secondary palate. In: Wyszynski DF, ed. Cleft lip and palate: from origin to treatment. New York: Oxford University Press, 2002: 5–24.

Spilson SV, Kim HJ, Chung KC. Association between maternal diabetes mellitus and newborn oral cleft. *Ann Plast Surg* 2001; 47: 477–481.

Stanier P, Moore GE. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet* 2004; 13 Spec No 1: R73-R81.

Suazo J, Santos JL, Carreño H, Jara L, Blanco R. Linkage disequilibrium between MSX1 and non-syndromic cleft lip/palate in the Chilean population. *J Dent Res* 2004; 83: 782–785.

Suazo J, Santos JL, Scapoli L, Jara L, Blanco R. Association between TGFB3 and nonsyndromic cleft lip with or without cleft palate in a Chilean population. *Cleft Palate Craniofac J* 2010; 47: 513–517.

Sull JW, Liang KY, Hetmanski JB, Fallin MD, Ingersoll RG, Park JW, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Park BY, Jee SH, Jabs EW, Redett R, Scott AF, Beaty TH. Excess maternal transmission of markers in TCOF1 among cleft palate case-parent trios from three populations. *Am J Med Genet A* 2008; 146A: 2327–2331.

Sun D, Baur S, Hay ED. Epithelial-mesenchymal transformation is the mechanism for fusion of the craniofacial primordia involved in morphogenesis of the chicken lip. *Dev Biol* 2000; 228: 337–349.

Suphapeetiporn K, Tongkobpetch S, Siriwan P, Shotelersuk V. TBX22 mutations are a frequent cause of non-syndromic cleft palate in the Thai population. *Clin Genet* 2007; 72: 478–483.

Suzuki K, Hu D, Bustos T, Zlotogora J, Richieri-Costa A, Helms JA, Spritz RA. Mutations of PVRL1, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nat Genet* 2000; 25: 427–430.

Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, Natsume N, Shimozato K, Ohbayashi N, Suzuki Y, Niimi T, Minami K, Yamamoto M, Altannamar TJ, Erkhembaatar T, Furukawa H, Daack-Hirsch S, L'heureux J, Brandon CA, Weinberg SM, Neiswanger K, Deleyiannis FW, de Salamanca JE, Vieira AR, Lidral AC, Martin JF, Murray JC. Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. *Am J Hum Genet* 2009; 84: 406–411.

Suzuki Y, Jezewski PA, Machida J, Watanabe Y, Shi M, Cooper ME, Viet le T, Nguyen TD, Hai H, Natsume N, Shimozato K, Marazita ML, Murray JC. In a Vietnamese population, MSX1 variants contribute to cleft lip and palate. *Genet Med* 2004; 6: 117–125.

Sözen MA, Hecht JT, Spritz RA. Mutation analysis of the PVRL1 gene in caucasians with nonsyndromic cleft lip/palate. *Genet Test Mol Biomarkers* 2009; 13: 617–621.

Sözen MA, Suzuki K, Tolarova MM, Bustos T, Fernández Iglesias JE, Spritz RA. Mutation of PVRL1 is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. *Nat Genet* 2001; 29: 141–142.

Zeiger JS, Beaty TH, Liang KY. Oral clefts, maternal smoking, and TGFA: a meta-analysis of gene-environment interaction. *Cleft Palate Craniofac J* 2005; 42: 58–63.

Zhang Z, Song Y, Zhao X, Zhang X, Fermin C, Chen Y. Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. *Development* 2002; 129: 4135–4146.

Zhao Y, Guo YJ, Tomac AC, Taylor NR, Grinberg A, Lee EJ, Huang S, Westphal H. (1999) Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene Lhx8. *Proc Natl Acad Sci USA* 1999; 96, 15002–15006.

Zhu H, Kartiko S, Finnell RH. Importance of gene-environment interactions in the etiology of selected birth defects. *Clin Genet* 2009; 75, 409–423.

Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* 2007; 80: 605–615.

Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004; 5: 89–100.

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, Natsume N, Yoshiura K, Vieira AR, Orioli IM, Castilla EE, Moreno L, Arcos-Burgos M, Lidral AC, Field LL, Liu YE, Ray A, Goldstein TH, Schultz RE, Shi M, Johnson MK, Kondo S, Schutte BC, Marazita ML, Murray JC. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip and palate. *N Engl J Med* 2004; 351: 769–780.

Taya Y, O'Kane S, Ferguson MW. Pathogenesis of cleft palate in TGF-beta3 knockout mice. *Development* 1999; 126: 3869–3879.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, /.../ Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; 466: 707–713.

Thomason HA, Dixon MJ, Dixon J. Facial clefting in Tp63 deficient mice results from altered Bmp4, Fgf8 and Shh signaling. *Dev Biol* 2008; 321: 273–282.

Thomason HA, Zhou H, Kouwenhoven EN, Dotto GP, Restivo G, Nguyen BC, Little H, Dixon MJ, van Bokhoven H, Dixon J. Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. *J Clin Invest* 2010; 120: 1561–1569.

Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK, Seldin MF. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet* 2008; 4: e4.

Tolarová MM, Cervenka J. Classification and birth prevalence of orofacial clefts. *Am J Med Genet* 1998; 75: 126–137.

Trokovic N, Trokovic R, Mai P, Partanen J. Fgfr1 regulates patterning of the pharyngeal region. *Genes Dev* 2003; 17: 141–153.

Trueba SS, Augé J, Mattei G, Etchevers H, Martinovic J, Czernichow P, Vekemans M, Polak M, Attié-Bitach T. PAX8, TITF1, and FOXE1 gene expression patterns during human development: new insights into human thyroid development and thyroid dysgenesis-associated malformations. *J Clin Endocrinol Metab* 2005; 90: 455–462.

Tudela C, Formoso MA, Martínez T, Pérez R, Aparicio M, Maestro C, Del Río A, Martínez E, Ferguson M, Martínez-Alvarez C. TGF-beta3 is required for the adhesion and intercalation of medial edge epithelial cells during palate fusion. *Int J Dev Biol* 2002; 46: 333–336.

van den Boogaard MJ, Dorland M, Beemer FA, van Amstel HK. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat Genet* 2000; 24: 342–343.

van Rooij IA, Groenen PM, van Drongelen M, Te Morsche RH, Peters WH, Steegers-Theunissen RP. Orofacial clefts and spina bifida: N-acetyltransferase phenotype, maternal smoking, and medication use. *Teratology* 2002; 66: 260–266.

van Rooij IA, Ocké MC, Straatman H, Zielhuis GA, Merkus HM, Steegers-Theunissen RP. Periconceptional folate intake by supplement and food reduces the risk of nonsyndromic cleft lip with or without cleft palate. *Prev Med* 2004; 39: 689–694.

van Rooij IA, Wegerif MJ, Roelofs HM, Peters WH, Kuijpers-Jagtman AM, Zielhuis GA, Merkus HM, Steegers-Theunissen RP. Smoking, genetic polymorphisms in biotransformation enzymes, and nonsyndromic oral clefting: a gene-environment interaction. *Epidemiology* 2001; 12: 502–507.

van Rooij IA, Vermeij-Keers C, Kluijtmans LA, Ocké MC, Zielhuis GA, Goorhuis-Brouwer SM, van der Biezen JJ, Kuijpers-Jagtman AM, Steegers-Theunissen RP. Does the interaction between maternal folate intake and the methylenetetrahydrofolate reductase polymorphisms affect the risk of cleft lip with or without cleft palate? *Am J Epidemiol* 2003; 157: 583–591.

Verstappen J, von den Hoff JW. Tissue inhibitors of metalloproteinases (TIMPs): their biological functions and involvement in oral disease. *J Dent Res* 2006; 85: 1074–1084.

Vieira AR, Avila JR, Daack-Hirsch S, Dragan E, Félix TM, Rahimov F, Harrington J, Schultz RR, Watanabe Y, Johnson M, Fang J, O'Brien SE, Orioli IM, Castilla EE, Fitzpatrick DR, Jiang R, Marazita ML, Murray JC. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet* 2005; 1: e64.

Vieira AR, McHenry TG, Daack-Hirsch S, Murray JC, Marazita ML. A genome wide linkage scan for cleft lip and palate and dental anomalies. *Am J Med Genet A* 2008a; 146A: 1406–1413.

Vieira AR, Modesto A, Meira R, Barbosa ARS, Lidral AC, Murray JC. Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. *Am J Med Genet A* 2007; 143: 538–545.

Vieira AR, Orioli IM, Castilla EE, Cooper ME, Marazita ML, Murray JC. MSX1 and TGFB3 contribute to clefting in South America. *J Dent Res* 2003; 82: 289–292.

Vieira AR. Unraveling human cleft lip and palate research. *J Dent Res* 2008b; 87: 119–125.

Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet* 2005; 1: e32.

Warrington A, Vieira AR, Christensen K, Orioli IM, Castilla EE, Romitti PA, Murray JC. Genetic evidence for the role of loci at 19q13 in cleft lip and palate. *J Med Genet* 2006; 43: e26.

Watanabe A, Akita S, Tin NT, Natsume N, Nakano Y, Niikawa N, Uchiyama T, Yoshiura K. A mutation in RYK is a genetic factor for nonsyndromic cleft lip and palate. *Cleft Palate Craniofac J* 2006; 43: 310–316.

Wehby GL, Murray JC. Folic acid and orofacial clefts: a review of the evidence. *Oral Dis* 2010; 16:11–19.

Weinberg SM, Brandon CA, McHenry TH, Neiswanger K, Deleyiannis FW, de Salamanca JE, Castilla EE, Czeizel AE, Vieira AR, Marazita ML. Rethinking isolated cleft palate: evidence of occult lip defects in a subset of cases. *Am J Med Genet A* 2008; 146A: 1670–1675.

Wilcox AJ, Lie RT, Solvoll K, Taylor J, McConnaughey DR, Abyholm F, Vindenes H, Vollset SE, Drevon CA. Folic acid supplements and risk of facial clefts: national population based case-control study. BMJ 2007; 334: 464.

Wilkie AO, Morriss-Kay GM. Genetics of craniofacial development and malformation. *Nat Rev Genet* 2001; 2: 458–468.

Wong WY, Eskes TK, Kuijpers-Jagtman AM, Spauwen PH, Steegers EA, Thomas CM, Hamel BC, Blom HJ, Steegers-Theunissen RP. Nonsyndromic orofacial clefts: association with maternal hyperhomocysteinemia. *Teratology* 1999; 60: 253–257.

Wu T, Liang KY, Hetmanski JB, Ruczinski I, Fallin MD, Ingersoll RG, Wang H, Huang S, Ye X, Wu-Chou YH, Chen PK, Jabs EW, Shi B, Redett R, Scott AF, Beaty TH. 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* 2010; 128: 401–410.

Xu X, Han J, Ito Y, Bringas P Jr, Urata MM, Chai Y. Cell autonomous requirement for Tgfbr2 in the disappearance of the medial edge epithelium during palatal fusion. *Dev Biol* 2006; 297: 238–248.

Yazdy MM, Honein MA, Xing J. Reduction in orofacial clefts following folic acid fortification of the U.S. grain supply. *Birth Defects Res A Clin Mol Teratol* 2007; 79: 16–23.

Yoshiura K, Machida J, Daack-Hirsch S, Patil SR, Ashworth LK, Hecht JT, Murray JC. Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q11.2;q13.3) in a family with cleft lip and palate. *Genomics* 1998; 54: 231–240.

SUMMARY IN ESTONIAN

Mittesündroomsete suulõhede geneetiline eelsoodumus

Suulõhed kuuluvad inimestel kõige sagedamini esinevate kaasasündinud väärarengute hulka ning nende hinnanguline esinemissagedus maailmas on 1/600 elussünni kohta. Suulõhede patogenees tuleneb kraniofatsiaalse morfogeneesi käigus tugevasti häirunud protsessidest molekulaarsel tasandil. Suulõhed jaotatakse anatoomiliste, geneetiliste ja embrüoloogiliste leidude põhjal kaheks suuremaks kompleksseks haigusfenotüübiks: huulelõhe koos huule- ja suulaelõhega (CL/P) ja suulaelõhe (CP).

Mittesündroomsed suulõhed on multifaktoriaalsed haigused, mille kujunemisel osalevad omavahel interakteeruvad geneetilised ning erinevad keskkonnategurite ja elustiiliga seostatud riskifaktorid. Mitmete suulõhede sündroomsete vormidega seostatud geenide uuringud on näidanud nende olulisust ka mittesündroomsete suulõhede etioloogias. Geneetiliste tegurite oluline osatähtsus mittesündroomsete suulõhede kujunemises on kinnitust leidnud perekondade ja kaksikute uuringutes, siiski on käesolevaks ajaks kirjeldatud ainult väiksem osa üldisest geneetiliste faktorite osakaalust ning kandidaatgeenidel põhinevate assotsiatsiooniuuringute ja aheldusanalüüside tulemuste vähene kokkulangevus on olulisel määral põhjustatud ka geneetilisest heterogeensusest ja väikesest uuringu võimsusest. Olulist lisaväärtust suulõhede kui komplekshaiguste geneetilise eelsoodumuse selgitamisel on paaril viimasel aastal andnud kogu genoomi assotsiatsiooniuuringud (GWAS) DNA mikrokiipidel.

Käesoleva doktoritöö eesmärgiks oli analüüsida kraniofatsiaalses morfogeneesis ja/või palatogeneesis osaleva 40 kandidaatgeeni järjestuse variantide võimalikku seost mittesündroomse CL/P ja CP geneetilises eelsoodumuses. Valitud geenidega teostati ca 590 ühenukleotiidse polümorfismi (SNP) *case-control* assotsiatsioonanalüüsid ühendatud Eesti, Läti ja Leedu uuringurühmas eraldi CL/P ja CP fenotüübiga. Kokku genotüpiseeriti 404 patsienti (300 CL/P ja 104 CP) ja 606 populatsioonipõhist kontrollindiviidi. Mainitud kolme geograafilises aspektis lähedase populatsiooni geneetiline lähedus on leidnud veenvat kinnitust PCA meetodiga. Lisaks viidi läbi replikatsiooniuuring Eesti ja Leedu CL/P uuringurühmades esimeses mittesündroomse CL/P GWAS uuringus kirjeldatud uutele genoomselt tähendusrikastele haigusseoselistele lookustele Kesk-Euroopa populatsioonides.

Uurimistöö olulisemad tulemused võib kokku võtta järgnevalt:

- *FGF1, FOXE1*, *TIMP2* ja *WNT9B* geenide järjestuse variandid on seotud mittesündroomse CL/P geneetilise eelsoodumusega Kirde-Euroopa populatsioonides (Eesti, Läti, Leedu).
- GWAS uuringutes kirjeldatud genoomselt tähendusrikas assotsiatsioon 8q24.21 kromosoomilookuse ning CL/P vahel replitseerus Eesti ja Leedu *case-control* uuringurühmades ning assotsiatsioon 10q25.3 kromosoomilookuse ja CL/P vahel replitseerus Eesti *case-control* uuringurühmas.

- *COL2A1, COL11A2, IRF6* ja *WNT3* geenide järjestuse variandid on seotud mittesündroomse CP geneetilise eelsoodumusega Kirde-Euroopa populatsioonides (Eesti, Läti, Leedu).
- *MSX1* ja *MTHFR* geenide järjestuse variandid on seotud mittesündroomse CL/P geneetilise eelsoodumusega Eesti populatsioonis.
- FGF ja Wnt signaalrajad ning nendes osalevad geenid omavad olulist rolli nii CL/P kui ka CP etioloogias.

Käesoleva uurimistöö tulemused näitavad, et paljud kraniofatsiaalse morfogeneesi ja/või palatogeneesiga seostatud geenid, sealjuures igaüks neist väikese individuaalse riskiefektiga, osalevad suulõhede kujunemises, kuid ainult üksikutel geenidel ja/või lookustel on oluline osatähtsus mittesündroomsete suulõhede geneetilises eelsoodumuses. Lisaks hüpoteesipõhistele kandidaatgeenide uuringutele on ka GWAS meetod tõendanud oma tähtsust suulõhede geneetilise eelsoodumusega seostatavate senikirjeldamata kromosoomiregioonide ning uute bioloogiliste hüpoteeside esiletoomisel edasisteks uuringuteks. Suurem osa mittesündroomsete suulõhede geneetilisest taustast on siiani avastamata ning uueks väljakutseks on läbi viia eelpoolkirjeldatud regioonide detailne geneetiline kaardistamine tegelike haigusseoseliste variantide funktsionaalsete efektide kindlakstegemiseks ning kraniofatsiaalses patoloogias osalevate erinevate signaalradade molekulaarsete mehhanismide täpsemaks selgitamiseks. Sellised suuremahulised resekveneerimisuuringud annavad parimaid tulemusi uurimiskeskuste ja biopankade rahvusvahelises koostöös, kus on kaasatud täpselt määratletud fenotüübiga patsientide kohordid ning suured kontrollindiviidide rühmad erineva etnilise taustaga populatsioonides.

ACKNOWLEDGEMENTS

I am most grateful to my supervisor Prof. Andres Metspalu for the opportunity to conduct this challenging research project in the Department of Biotechnology. His constructive criticism, support and encouragement to apply new methods and collaborate with other researchers have been valuable.

I thank all my good friends and colleagues from Department of Biotechnology for creating motivating and friendly atmosphere. I am especially thankful to Prof. Ants Kurg, Krista Liiv, Merike Leego, Heidi Saulep and Viljo Soo.

My deep gratitude goes to my distinguished colleagues Mare Saag and Triin Jagomägi for their valuable contribution. I appreciate the experimental and technical support given by Kaarel Krjutškov, Veronika Tammekivi and Tõnu Esko.

I would like to thank all the co-authors in Riga Stradins University and Vilnius University. I am very grateful to Prof. Markus Nöthen and Dr. Elisabeth Mangold together with Stefanie Birnbaum, Sven Cichon and all other people in the Institute of Human Genetics, University of Bonn, for the opportunity to perform a replication study in Estonian sample, for their warm welcome and for their help, support and discussions opening new horizons.

I would like to thank European Science Foundation for the grant which allowed me to carry out a collaborative study with researchers from University of Bonn.

My special thanks go to my dearest friends Piret, Kristel, Evelin, Julia, Marina, Triinu, Merli and Valentina for the inspiration.

I am grateful to Jaanis Kasesalu, my good friend through the years for being there for me in right time and right place.

I would also like to thank the gene donors in the Estonian Genome Center and all patients and their families for their cooperation.

PUBLICATIONS

CURRICULUM VITAE

Tiit Nikopensius

Date of birth:	24 October 1963, Tartu, Estonia
Address:	Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu Riia 23, 51010, Tartu, Estonia
Phone:	+372 7 375882
E-mail:	tiitn@ebc.ee

Education

1970–1981	Tartu 7th Secondary School
1981–1986	B.Sc. in biology, Faculty of Biology and Geography, Tartu State University
1994	Master's degree in life sciences (biotechnology), Institute of Biotechnology, Estonian Agricultural University

Professional employment

1986–1988	Estonian Academy of Sciences, Institute of Chemical and Biological Physics, senior engineer
1989–1994	Estonian Agrobiocentre, junior researcher
1995–1999	Estonian Agrobiocentre, researcher
2000–2002	University of Tartu, Centre of Technology, Department of Gene Technology, researcher
2003–2005	Estonian Biocentre, researcher
2006–	University of Tartu, Institute of Molecular and Cell Biology, researcher in biotechnology
2009–	Estonian Biocentre, researcher

My research interests

I have been participated as project leader and coordinator from IMCB/UT side in several collaborative (within Estonia and international) case-control association studies involving diverse fields in medical genetics where polymorphisms in candidate genes implicated in the etiology of affective and anxiety disorders, psoriasis, Parkinson's disease, myocardial infarction and coronary artery disease, and nonsyndromic orofacial clefts have been studied.

Publications

1. Kingo K, Kõks S, **Nikopensius T**, Silm H, Vasar E. Polymorphisms in the interleukin–20 gene: relationships to plaque-type psoriasis. *Genes Immun* 2004; 5: 117–121.
2. Koido K, Kõks S, **Nikopensius T**, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. Polymorphisms in wolframin (WFS1) gene are possibly related to increased risk for mood disorders. *Int J Neuropsychopharmacol* 2005; 8(2): 235–244.
3. Maron E, **Nikopensius T**, Kõks S, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Koido K, Kurg A, Metspalu A, Vasar E, Vasar V, Shlik J. Association study of 90 candidate gene polymorphisms in panic disorder. *Psychiatric Genet* 2005; 15(1): 17–24.
4. Kõks S, **Nikopensius T**, Koido K, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. Analysis of SNP profiles in patients with major depressive disorder. *Int J Neuropsychopharmacol* 2006; 9(2): 167–174.
5. Shadrina M, **Nikopensius T**, Slominsky P, Illarioshkin S, Bagyeva G, Markova E, Ivanova-Smolenskaia I, Kurg A, Limborska S, Metspalu A. Association study of sporadic Parkinson's disease genetic risk factors in patients from Russia by APEX technology. *Neurosci Lett* 2006; 405(3): 212–216.
6. Krjutškov K, Andreson R, Mägi R, **Nikopensius T**, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A. Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. *Nucleic Acids Res* 2008; 36(12): e75.
7. Khrunin A, Mihailov E, **Nikopensius T**, Krjutškov K, Limborska S, Metspalu A. Analysis of allele and haplotype diversity across 25 genomic regions in three Eastern European populations. *Hum Hered* 2009; 68(1): 35–44.
8. **Nikopensius T**, Ambrozaitytė L, Ludwig KU, Birnbaum S, Jagomägi T, Saag M, Matulevičienė A, Linkevičienė L, Herms S, Knapp M, Hoffmann P, Nöthen MM, Kučinskas V, Metspalu A, Mangold E. Replication of novel susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients. *Am J Med Genet A* 2009; 149(11): 2551–2553.
9. Theodoraki EV, **Nikopensius T**, Suhorutšenko J, Papamikos V, Kolovou GD, Peppes V, Panagiotakos D, Limberi S, Zakopoulos N, Metspalu A, Dedoussis GV. ROS1 Asp2213Asn polymorphism is not associated with coronary artery disease in a Greek case-control study. *Clin Chem Lab Med* 2009; 47(12): 1471–1473.
10. Krjutškov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, **Nikopensius T**, Kivisild T, Metspalu A.

Evaluation of the 124-plex SNP typing microarray for forensic testing. *Forensic Sci Int Genet* 2009; 4(1): 43–48.
11. Theodoraki EV, **Nikopensius T**, Suhorutšenko J, Peppes V, Fili P, Kolovou G, Papamikos V, Richter D, Zakopoulos N, Krjutškov K, Metspalu A, Dedoussis GV. Fibrinogen beta variants confer protection against coronary artery disease in a Greek case-control study. *BMC Med Genet* 2010; 11(1): 28.
12. Jagomägi T, **Nikopensius T**, Krjutškov K, Tammekivi V, Viltrop T, Saag M, Metspalu A. MTHFR and MSX1 contribute to the risk of non-syndromic cleft lip/palate. *Eur J Oral Sci* 2010; 118(3): 213–220.
13. **Nikopensius T**, Birnbaum S, Ludwig KU, Jagomägi T, Saag M, Herms S, Knapp M, Hoffmann P, Nöthen MM, Metspalu A, Mangold E. Susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 10q25 confers risk in Estonian patients. *Eur J Oral Sci* 2010; 118(3): 317–319.
14. **Nikopensius T**, Jagomägi T, Krjutškov K, Tammekivi V, Saag M, Prane I, Piekuse L, Akota I, Barkane B, Krumina A, Ambrozaitytė L, Matulevičienė A, Kučinskienė ZA, Lace B, Kučinskas V, Metspalu A. Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. *Birth Defects Res A Clin Mol Teratol* 2010; 88(9): 748–756.
15. **Nikopensius T**, Kempa I, Ambrozaitytė L, Jagomägi T, Saag M, Matulevičienė A, Utkus A, Krjutškov K, Tammekivi V, Piekuse L, Akota I, Barkane B, Krumina A, Klovins J, Lace B, Kučinskas V, Metspalu A. Variation in FGF1, FOXE1, and TIMP2 genes is associated with non-syndromic cleft lip with or without cleft palate. *Birth Defects Res A Clin Mol Teratol* 2011; 91(4): 218–225.

Fellowships

– Nordic Council of Ministers scholarship: visiting scientist at Danish Veterinary Laboratory, Copenhagen; October 1, 1996 – March 31, 1997
– Danish Rectors Conference scholarship: visiting scientist at Danish Veterinary Laboratory, Copenhagen; October 1 – November 30, 1998
– ESF “Frontiers of Functional Genomics” exchange grant 2182: collaboration study with Institute of Human Genetics & Department of Genomics/Life and Brain Center, University of Bonn, Germany; January 13 – March 12, 2009

ELULOOKIRJELDUS

Tiit Nikopensius

Sünniaeg:	24. oktoober 1963, Tartu, Eesti
Aadress:	Biotehnoloogia õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool Riia 23, 51010, Tartu, Eesti
Telefon:	+372 7 375882
E-mail:	tiitn@ebc.ee

Hariduskäik

1970–1981	Tartu 7. Keskkool
1981–1986	B.Sc. bioloogia erialal, Bioloogia-Geograafiateaduskond, Tartu Riiklik Ülikool
1994	loodusteaduste magister M.Sc. biotehnoloogia erialal, Biotehnoloogia Instituut, Eesti Põllumajandusülikool

Erialane teenistuskäik

1986–1988	ENSV TA Keemilise ja Bioloogilise Füüsika Instituut, vaneminsener
1989–1994	Eesti Agrobiokeskus, nooremteadur
1995–1999	Eesti Agrobiokeskus, teadur
2000–2002	Tartu Ülikool, Tehnoloogiakeskus, geenitehnoloogia osakond, teadur
2003–2005	Eesti Biokeskus, teadur
2006–	Tartu Ülikool, Molekulaar- ja Rakubioloogia Instituut, biotehnoloogia teadur
2009–	Eesti Biokeskus, teadur

Teadustegevus

Olen osalenud mitmetes Eesti-sisestes ja rahvusvahelistes koostööprojektides, mis on hõlmanud erinevaid meditsiinigeneetika valdkondi. Projektide koordinaatorina TÜMRI Biotehnoloogia õppetoolis olen läbi viinud kandidaatgeenide markerite analüüsil põhinevaid *case-control* assotsiatsiooniuuringuid meeleolu- ja ärevushäirete, psoriaasi, Parkinsoni haiguse, müokardi infarkti ja südame isheemiatõve ning mittesündroomsete suulõhede geneetilise eelsoodumuse selgitamiseks.

Publikatsioonid

1. Kingo K, Kõks S, **Nikopensius T**, Silm H, Vasar E. Polymorphisms in the interleukin–20 gene: relationships to plaque-type psoriasis. *Genes Immun* 2004; 5: 117–121.
2. Koido K, Kõks S, **Nikopensius T**, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. Polymorphisms in wolframin (WFS1) gene are possibly related to increased risk for mood disorders. *Int J Neuropsychopharmacol* 2005; 8(2): 235–244.
3. Maron E, **Nikopensius T**, Kõks S, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Koido K, Kurg A, Metspalu A, Vasar E, Vasar V, Shlik J. Association study of 90 candidate gene polymorphisms in panic disorder. *Psychiatric Genet* 2005; 15(1):17–24.
4. Kõks S, **Nikopensius T**, Koido K, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. Analysis of SNP profiles in patients with major depressive disorder. *Int J Neuropsychopharmacol* 2006; 9(2): 167–174.
5. Shadrina M, **Nikopensius T**, Slominsky P, Illarioshkin S, Bagyeva G, Markova E, Ivanova-Smolenskaia I, Kurg A, Limborska S, Metspalu A. Association study of sporadic Parkinson's disease genetic risk factors in patients from Russia by APEX technology. *Neurosci Lett* 2006; 405(3): 212–216.
6. Krjutškov K, Andreson R, Mägi R, **Nikopensius T**, Khrunin A, Mihailov E, Tammekivi V, Sork H, Remm M, Metspalu A. Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. *Nucleic Acids Res* 2008; 36(12): e75.
7. Khrunin A, Mihailov E, **Nikopensius T**, Krjutškov K, Limborska S, Metspalu A. Analysis of allele and haplotype diversity across 25 genomic regions in three Eastern European populations. *Hum Hered* 2009; 68(1): 35–44.
8. **Nikopensius T**, Ambrozaitytė L, Ludwig KU, Birnbaum S, Jagomägi T, Saag M, Matulevičienė A, Linkevičienė L, Herms S, Knapp M, Hoffmann P, Nöthen MM, Kučinskas V, Metspalu A, Mangold E. Replication of novel susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24 in Estonian and Lithuanian patients. *Am J Med Genet A* 2009; 149(11): 2551–2553.
9. Theodoraki EV, **Nikopensius T**, Suhorutšenko J, Papamikos V, Kolovou GD, Peppes V, Panagiotakos D, Limberi S, Zakopoulos N, Metspalu A, Dedoussis GV. ROS1 Asp2213Asn polymorphism is not associated with coronary artery disease in a Greek case-control study. *Clin Chem Lab Med* 2009; 47(12): 1471–1473.
10. Krjutškov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, **Nikopensius T**, Kivisild T, Metspalu A.

Evaluation of the 124-plex SNP typing microarray for forensic testing. *Forensic Sci Int Genet* 2009; 4(1): 43–48.
11. Theodoraki EV, **Nikopensius T**, Suhorutšenko J, Peppes V, Fili P, Kolovou G, Papamikos V, Richter D, Zakopoulos N, Krjutškov K, Metspalu A, Dedoussis GV. Fibrinogen beta variants confer protection against coronary artery disease in a Greek case-control study. *BMC Med Genet* 2010; 11(1): 28.
12. Jagomägi T, **Nikopensius T**, Krjutškov K, Tammekivi V, Viltrop T, Saag M, Metspalu A. MTHFR and MSX1 contribute to the risk of nonsyndromic cleft lip/palate. *Eur J Oral Sci* 2010; 118(3): 213–220.
13. **Nikopensius T**, Birnbaum S, Ludwig KU, Jagomägi T, Saag M, Herms S, Knapp M, Hoffmann P, Nöthen MM, Metspalu A, Mangold E. Susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 10q25 confers risk in Estonian patients. *Eur J Oral Sci* 2010; 118(3): 317–319.
14. **Nikopensius T**, Jagomägi T, Krjutškov K, Tammekivi V, Saag M, Prane I, Piekuse L, Akota I, Barkane B, Krumina A, Ambrozaitytė L, Matulevičienė A, Kučinskienė ZA, Lace B, Kučinskas V, Metspalu A. Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. *Birth Defects Res A Clin Mol Teratol* 2010; 88(9): 748–756.
15. **Nikopensius T**, Kempa I, Ambrozaitytė L, Jagomägi T, Saag M, Matulevičienė A, Utkus A, Krjutškov K, Tammekivi V, Piekuse L, Akota I, Barkane B, Krumina A, Klovins J, Lace B, Kučinskas V, Metspalu A. 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* 2011; 91(4): 218–225.

Stipendiumid

- Nordic Council of Ministers Scholarship: Danish Veterinary Laboratory, Copenhagen, külalisteadur; 1. oktoober 1996 – 31. märts 1997
- Danish Rectors Conference Scholarship: Danish Veterinary Laboratory, Copenhagen, külalisteadur; 1. oktoober – 30. november 1998
- ESF "Frontiers of Functional Genomics" exchange grant 2182: Institute of Human Genetics, University of Bonn, Saksamaa, külalisteadur koostööuuringus; 13. jaanuar – 12. märts 2009

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand.** The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous crassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

82. **Ivar Ilves.** Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003. 89 p.
83. **Andres Männik.** Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003. 109 p.
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003. 158 p.
85. **Gudrun Veldre.** Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003. 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004. 159 p.
87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
88. **Tiina Kahre.** Cystic fibrosis in Estonia. Tartu, 2004. 116 p.
89. **Helen Orav-Kotta.** Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004. 117 p.
90. **Maarja Öpik.** Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004. 175 p.
91. **Kadri Tali.** Species structure of *Neotinea ustulata*. Tartu, 2004. 109 p.
92. **Kristiina Tambets.** Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004. 163 p.
93. **Arvi Jõers.** Regulation of p53-dependent transcription. Tartu, 2004. 103 p.
94. **Lilian Kadaja.** Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004. 103 p.
95. **Jaak Truu.** Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004. 128 p.
96. **Maire Peters.** Natural horizontal transfer of the *pheBA* operon. Tartu, 2004. 105 p.
97. **Ülo Maiväli.** Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004. 130 p.
98. **Merit Otsus.** Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004. 103 p.
99. **Mikk Heidemaa.** Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004. 167 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004. 111 p.
101. **Lauri Saks.** Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004. 144 p.

102. **Siiri Rootsi.** Human Y-chromosomal variation in European populations. Tartu, 2004. 142 p.
103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.

122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007. 112 p.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.

142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.
143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2008, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast.** Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.

161. **Ain Vellak.** Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel.** Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe.** Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe.** Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa.** Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil.** Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik.** Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark.** Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap.** Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe.** Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi.** Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson.** Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts.** Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis.** Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov.** Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster.** Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap.** Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar.** Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.

181. **Kristjan Herkül.** Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk.** Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü.** Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla.** Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey.** The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp.** Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber.** The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro.** The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold.** Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert.** Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu.** Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik.** ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber.** Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper.** Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak.** Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo.** Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel.** Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus.** Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.