Risk prediction models for oral clefts allowing for phenotypic heterogeneity

*Yalu Wen[1] and Qing Lu[2]**

[1] Department of Statistics, University of Auckland, Auckland, New Zealand, [2] Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA

Oral clefts are common birth defects that have a major impact on the affected individual, their family and society. World-wide, the incidence of oral clefts is 1/700 live births, making them the most common craniofacial birth defects. The successful prediction of oral clefts may help identify sub-population at high risk, and promote new diagnostic and therapeutic strategies. Nevertheless, developing a clinically useful oral clefts risk prediction model remains a great challenge. Compelling evidences suggest the etiologies of oral clefts are highly heterogeneous, and the development of a risk prediction model with consideration of phenotypic heterogeneity may potentially improve the accuracy of a risk prediction model. In this study, we applied a previously developed statistical method to investigate the risk prediction on sub-phenotypes of oral clefts. Our results suggested subtypes of cleft lip (CL) and palate have similar genetic etiologies (AUC = 0.572) with subtypes of CL only (AUC = 0.589), while the subtypes of cleft palate only (CPO) have heterogeneous underlying mechanisms (AUCs for soft CPO and hard CPO are 0.617 and 0.623, respectively). This highlighted the potential that the hard and soft forms of CPO have their own mechanisms despite sharing some of the genetic risk factors. Comparing with conventional methods for risk prediction modeling, our method considers phenotypic heterogeneity of a disease, which potentially improves the accuracy for predicting each sub-phenotype of oral clefts.

Keywords: sub-phenotype, multi-class likelihood-ratio ensemble method, cleft lip with or without palate, cleft palate only, genetic heterogeneity

Introduction

Oral clefts comprise a significant component of birth defects. Individuals born with orofacial clefts are likely to require subsequent dental, speech, and psychosocial therapies to correct for the craniofacial anomalies to various degrees (Strauss, 1999; Mossey et al., 2009; Wehby and Cassell, 2010). Traditionally, oral clefts are classified into cleft palate only (CPO) and cleft lip (CL) with or without palate (CL/P; Fraser, 1955, 1970). The birth prevalence of CL/P is about 1/700 with wide variability associated with geographic origin, whereas CPO affects 1/2500 births with less variability compared to CL/P (Mossey, 2003; Genisca et al., 2009; Beaty et al., 2011; Dixon et al., 2011). Oral clefts can also be divided into non-syndromic and syndromic forms, where approximately 70% of CL/P and 50% of CPO are non-syndromic (Jones, 1988; Stoll et al., 2000; Calzolari et al., 2007; Genisca et al., 2009; Jugessur et al., 2009; Dixon et al., 2011). With the advent of genomic era, major breakthroughs have been made into identifying genetic variants predisposing to the syndromic oral

clefts, while our knowledge of non-syndromic oral clefts genetic etiology were still far behind. This could be partially due to the heterogeneous etiology and the non-Mendelian inheritance pattern of non-syndromic oral clefts (Murray, 2002; Dixon et al., 2011). For this particular reason, for the rest of this paper we focus on non-syndromic oral clefts with the consideration of their phenotypic heterogeneity.

Compelling evidences suggest that genetic variants play a substantial role in the development of oral clefts (Little and Bryan, 1986; Wyszynski et al., 1998; Murray, 2002; Mossey et al., 2009; Grosen et al., 2010; Beaty et al., 2011; Grosen et al., 2011). Twin studies indicate that the concordance rates of CL, cleft lip and palate (CLP), and CPO are higher for monozygotic twins than for dizygotic twins (Little and Bryan, 1986; Murray, 2002; Grosen et al., 2010, 2011; Beaty et al., 2011). For example, in a nationwide study in Denmark, the proband-wise concordance rate is 33% for monozygotic twins while the rate is 7% for dizygotic twins, which is only slightly higher than the 3% recurrence risk observed for full siblings (Grosen et al., 2010, 2011). Moreover, the recurrence risk of oral clefts in families is greater than that predicted by the familial aggregation of environmental risk factors. Evidence also shows that the risk of oral clefts among first degree relatives of cases is much higher than that in the general population (Wyszynski et al., 1998; Sivertsen et al., 2008). Conventionally, CL/P and CPO are treated separately, because the developmental origins of these two defects are different during the embryonic stage. Through genetic linkage studies, various loci and genetic regions, such as *MTHFR*, *TGFA,* and *TGFB3*, were found to play a potential role in CL/P (Prescott et al., 2000; Zeiger et al., 2003). Genes related to growth factors [e.g., *TGFA* (Mitchell, 1997; Zeiger et al., 2005; Vieira, 2006)], transcription factors [e.g., *IRF6*(Zucchero et al., 2004; Park et al., 2007; Vieira et al., 2007a,b; Jugessur et al., 2008)], nutrient metabolism [e.g., *MTHFR* (Vieira et al., 2005)], and immune response [e.g., *PVRL1*(Sozen et al., 2001)] have also been examined through genetic association studies. As with many other candidate gene studies, rigorous confirmatory replication is not common, except for the gene *IRF6*, which is linked strongly to the isolated form of clefts. The association finding of *IRF6* with CL/P has been replicated in many different populations and ethnic groups (Zucchero et al., 2004; Park et al., 2007; Vieira et al., 2007a,b; Jugessur et al., 2008; Mossey et al., 2009). To date, much of the attention has been paid to CL/P rather than CPO among non-syndromic oral clefts (Beaty et al., 2011; Dixon et al., 2011). This may be explained by relatively large samples of CL/P that are available, better ascertainment and less confounding issues for CL/P as compared to CPO (Dixon et al., 2011). More studies will be needed to fill in the gap to shine light on the underlying biological mechanism of CPO. Despite these discoveries, the results from both linkage and association studies are largely inconsistent, indicating the challenge of identifying disease-associated genetic variants for complex diseases with heterogeneous etiology (Carter et al., 1982; Harville et al., 2005; Sivertsen et al., 2008; Mossey et al., 2009; Dixon et al., 2011).

With the increasing genetic and epidemiologic findings for oral clefts, the translation of these discoveries into health practice becomes one of the major challenges of the coming decades. It is hoped that the genetic risk prediction could help identify sub-population at high risk of oral clefts and then advanced disease prevention and intervention strategies can be used to reduce the risk. Despite such promise, the existing genetic findings are insufficient to explain the familial aggregation of oral clefts (Mossey et al., 2009; Dixon et al., 2011) and as the result the risk prediction models for oral clefts formed to date have lacked sufficient accuracy for clinical use. Part of this difficulty is due to the phenotypic heterogeneity, i.e., oral clefts with the same or similar clinical manifestations have different genetic etiologies. When heterogeneous sub-phenotypes were treated as a single entity, the predictive power of the disease-associated variants could be substantially reduced, leading to a prediction model with low accuracy (Morris et al., 2010; Dixon et al., 2011). The use of more refined sub-phenotypes defined based upon disease symptom, severity of illness, and age at onset, facilitates the identification of new genetic variants contributing to each sub-phenotype, and helps build a more accurate risk prediction model (Morris et al., 2010). The improved risk prediction model could be used to identify high risk sub-population deserving special attention so that more targeted prevention and intervention strategies can be used to reduce the mortality and morbidity. However, in the absence of a well-defined diagnosis criterion to classify oral clefts into more homogeneous sub-phenotype groups, a risk prediction model simply built on each sub-phenotype could be subject to low accuracy and high variability because of the small sample size for each sub-phenotype. Recently, a multi-class likelihood-ratio ensemble (MLRE) method has been proposed. It gradually combines sub-phenotype groups with similar genetic etiology into homogeneous subgroups, and in general does not require any prior knowledge of subgroup information (Wen and Lu, 2013). With explicitly accounting for phenotypic heterogeneity, the method has been shown to have greater power over the existing methods. By applying the new method to a large-scale oral clefts genetic data, we simultaneously consider nine refined sub-phenotypes of non-syndromic oral clefts defined primary based on clinical manifestations, and gradually combine sub-phenotypes with similar genetic etiology. We further build risk prediction models on each combined sub-phenotype group by considering 148 candidate single nucleotide polymorphisms (SNPs) and their potential interactions.

Materials and Methods

The International Consortium to Identify Genes and Interactions Controlling Oral Clefts (ICOCs) Study Dataset

The Interactions Controlling Oral Clefts (ICOCs) is one of the largest and most comprehensive family based studies conducted to date, aimed at discovering genes and interactions contributing to oral clefts. It was developed through the trans-*NIH* Genes, Environment, and Health Initiative (*GEI*) and pulled together a large collection of cases and their parents

from multiple populations based on similar research protocols (Beaty et al., 2010; Cornelis et al., 2010), which comprised nearly 1908 case-parent trios from different racial groups, including Caucasian, African Americans, and Asians. DNA samples were collected from both cases and their parents, and were genotyped using the Illumina Human 1M DNA Analysis Bead Chip. Cases were affected offspring identified through a treatment center of population based registry and were individuals diagnosed with an isolated, non-syndromic oral cleft that included soft CPO, hard CPO, left cleft lip, right cleft lip, bilateral cleft lip, left cleft lip, and palate, right cleft lip, and palate, bilateral cleft lip, and palate, microforms of oral cleft, and unknown types of oral clefts. Parents of the affected children were also recruited for the study. The DNA samples were genotyped with Illumina's 610 Quad platform at the Johns Hopkins University Center for Inherited Disease Research (CIDR). Based on existing literature, we have identified 148 SNPs potentially associated with oral cleft, which were available in the dataset. Due to the small sample size, individuals with microforms and bilateral cleft lip were excluded from the analyses. Individuals with unknown types of oral clefts were also excluded from the analyses, as the genetic causes for unknown types of oral clefts could be highly heterogeneous.

Genetic Risk Prediction Analysis Allowing for Phenotypic Heterogeneity

In this risk prediction analysis, we use a newly developed MLRE, method (Wen and Lu, 2013) to take the heterogeneous nature of oral clefts into account. The MLRE method gradually combines sub-phenotypes into homogeneous groups based on the genetic similarities among sub-phenotypes. It assumes no mode of inheritance and allows for the identification of high-order gene-gene interactions by using a computationally efficient forward selection algorithm. The MLRE starts with all sub-phenotype groups under investigation and treats each sub-phenotype as a distinct disease outcome. It then applies the receiver operating characteristic (ROC)-based forward selection algorithm to identify genetic risk factors for each sub-phenotype and then measures the pair-wise genetic similarities among all sub-phenotypes. It gradually combines the most similar sub-phenotypes and evaluates the overall classification accuracy of the model using a global ROC statistic, which addresses both the lack of accuracy due to the presence of heterogeneity among sub-phenotypes and the accuracy loss caused by reduced sample size when some sub-phenotypes indeed share the same genetic etiology. The process continues until all sub-phenotypes are clustered into one group. The optimum number of sub-phenotype groups with sufficient sample size of each group and least possible phenotype heterogeneity is determined through a K-fold cross validation procedure using the global ROC-statistic as a criterion. With the selected optimal number of sub-phenotypes, the method is then applied to all samples to build a risk prediction model for each sub-phenotype group. Through extensive simulation studies, Wen and Lu (2013) have demonstrated that MLRE attained higher accuracy compared with commonly adopted methods under various underlying numbers of sub-phenotype groups and disease models.

The major advantage of MLRE over existing methods (e.g., those analyzing sub-phenotype separately or analyzing all sub-phenotypes as a single entity) is that MLRE incorporates genetic information obtained from data and gradually combines sub-phenotype groups that share similar genetic etiology to improve both accuracy and precision of risk prediction models. Therefore, it overcomes the obstacles when the prior information of defining sub-phenotypes is lacking or inaccurate, and provides a powerful and flexible framework to search for genetic variants contributing to complex human diseases, allowing for heterogeneous genetic causes among sub-phenotypes of a disease.

Results

Descriptive Analysis

Table 1 describes the samples of all sub-phenotypes and controls in relation to the distribution of gender in the ICOC dataset. Noteworthy, while the distribution of gender among controls is roughly balanced, there is an excess of cleft palate cases in females and an excess of CL with/without palate cases in males.

Risk Prediction Modeling

While prior studies have focused on building risk prediction models for either CL with/without palate or CPO, we extended the risk prediction analysis by explicitly considering all sub-phenotypes of oral clefts, defined by their clinical manifestations (e.g., the pattern of symptoms). By using a more refined sub-phenotype, it is possible to detect genetic variants contributing to a refined sub-phenotype, leading to an improved risk prediction model for the sub-phenotype. The details of analyses are summarized in **Table 2** and the ROC curves of risk prediction models for all sub-phenotype groups are depicted in **Figure 1** (CL/P) and **Figure 2** (CPO). Consistent with our current understanding of oral clefts, left CLP, right CLP and bilateral CLP showed similar genetic etiologies and were combined into

TABLE 1 | Distribution of sub-phenotypes in the Interactions Controlling Oral Clefts (ICOCs) dataset.

Sub-phenotypes of oral clefts		All	Male	Female
Control		3692	1759	1933
Cleft palate only	CP-Soft only	223	85	138
	CP- Hard	172	76	96
Cleft lip only	Left CL	244	152	92
	Right CL	114	66	48
	Bilateral CL	30	15	15
Cleft lip and palate	Left CLP	485	314	171
	Right CLP	271	183	88
	Bilateral CLP	336	240	96
Unknown		615	361	254
Microforms		3	2	1

TABLE 2 | Risk prediction model for each sub-phenotype group of oral clefts.

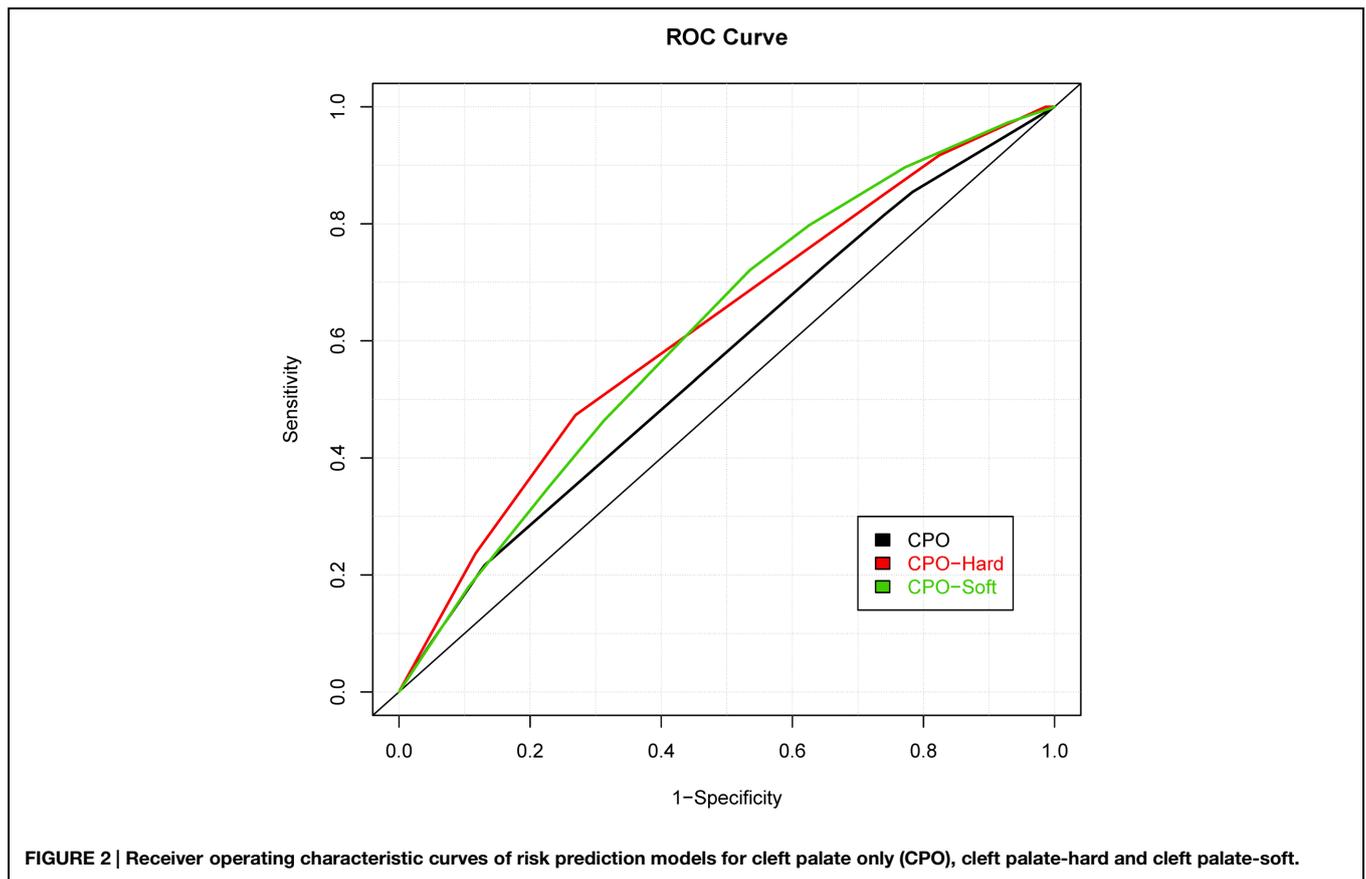| Selected SNPs	Allele	Chromosome	Gene	Position	AUC
Cleft lip and palate					
rs6072081	A/G	20	*MAFB*	39261054	0.572(0.556,0.589)[a]
rs2073485	A/G	1	*IRF6*	209962794	
rs560426	A/G	1	*ABCA4*	94553438	
Cleft lip only					
rs2073485	A/G	1	*IRF6*	209962794	0.589(0.562,0.617)
rs590223	A/G	1	*IRF6*	209946707	
Cleft palate – hard					
rs2073485	A/G	1	*IRF6*	209962794	0.623(0.583,0.663)
rs17389541	A/G	1	*IRF6*	208053795	
rs1530300	C/T	8	*8q24*	129988640	
Cleft palate-soft only					
rs227731	A/C	17	*17q22*	54773238	0.617(0.583,0.652)
rs861020	A/G	1	*IRF6*	209989270	
rs2514527	A/C	8	*GDF6*	97169326	

[a]*95% confidence interval (CI).*
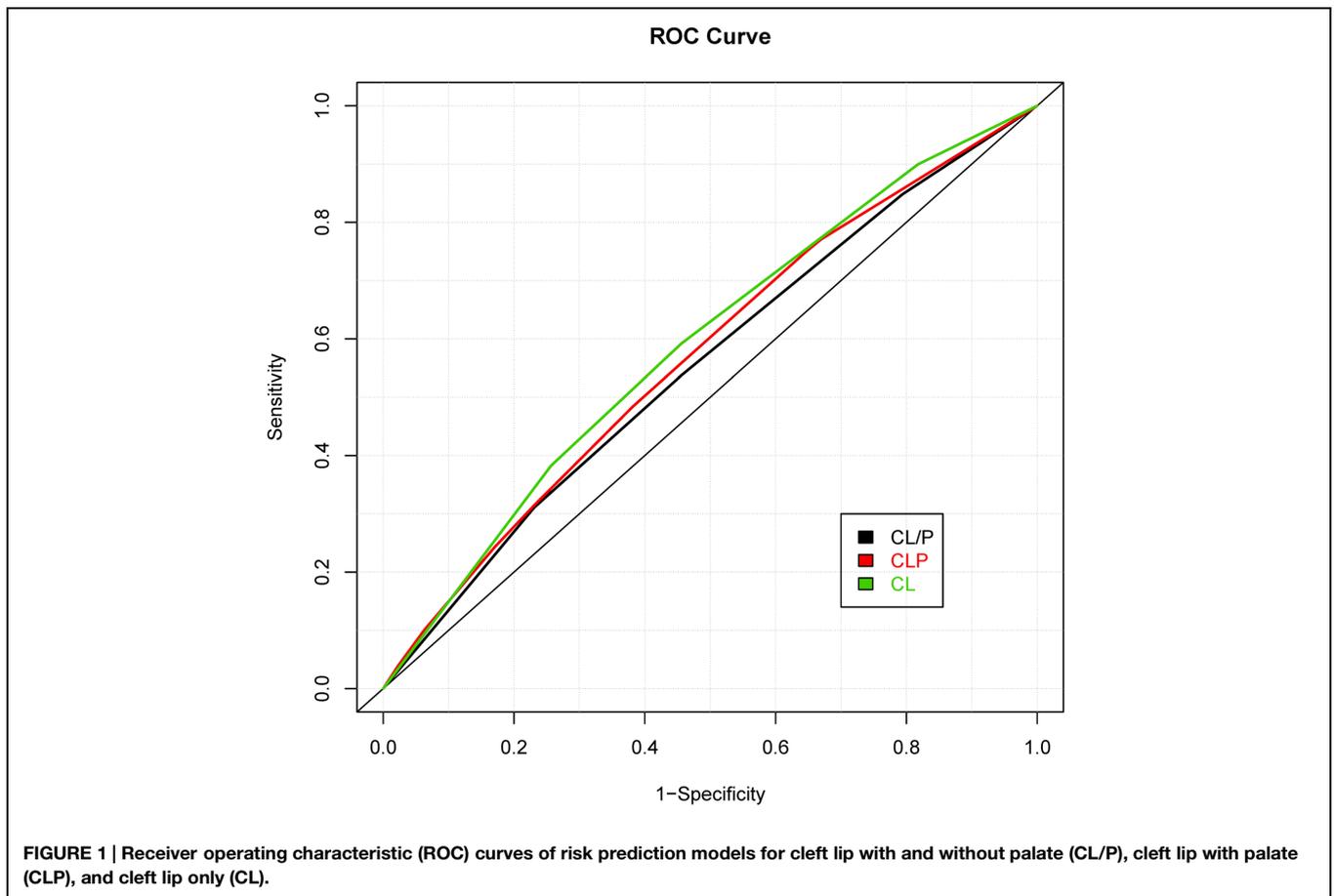
FIGURE 1 | Receiver operating characteristic (ROC) curves of risk prediction models for cleft lip with and without palate (CL/P), cleft lip with palate (CLP), and cleft lip only (CL).

one group by the MLRE method. The area under the ROC curve (AUC) of the CLP risk prediction model, which comprised of rs6072081 (*MAFB*), rs2073485 (*IRF6*) and rs560426 (*ABCA4*), was estimated to be 0.572 with 95% confidence interval (CI) of (0.556, 0.589). Left CL and right CL were also treated as a single entity by the MLRE method. The CL risk prediction model (**Figure 1**) with rs2073485 (*IRF6*) and rs590223 (*IRF6*) selected as risk predictors has an AUC value of 0.589 with 95% CI of (0.562, 0.617). Contrary to most of the current findings, our analyses treated cleft palate-soft only and cleft palate-hard

FIGURE 2 | Receiver operating characteristic curves of risk prediction models for cleft palate only (CPO), cleft palate-hard and cleft palate-soft.

TABLE 3 | Risk prediction models for cleft palate only and cleft lip with and without palate.

Selected SNPs	Allele	Chromosome	Gene	Position	AUC
Cleft palate only					
rs227731	A/C	17	*17q22*	54773238	0.604(0.577,0.631)[a]
rs861020	A/G	1	*IRF6*	209989270	
rs2514527	A/C	8	*GDF6*	97169326	
Cleft lip with and without palate					
rs2073485	A/G	1	*IRF6*	209962794	0.556(0.542,0.570)
rs7078160	A/G	10	*VAX1*	118827560	

[a]*95% CI.*

as two separate groups. For these two groups, rs227731 (*17q22*), rs8610209 (*IRF6*), and rs2514527 (*GDF6*) were selected as risk predictors for cleft palate-soft only, whereas rs2073485 (*IRF6*), rs17389541 (*IRF6*), and rs1530300 (*8q24*) jointly contributed to the risk of cleft palate-hard. The risk prediction model formed for cleft palate-soft only (**Figure 2**) achieved an AUC value of 0.617 with 95% CI of (0.583,0.652), while the risk prediction model formed for cleft palate-hard (**Figure 2**) had an AUC value of 0.623 with 95% CI of (0.583,0.663).

We also compared the risk prediction models formed by MLRE to those formed based on empirical knowledge. For such purposes, as suggested by previous literature, we classified the oral clefts into two sub-phenotype groups (i.e., CL/P and CPO). The same forward selection algorithm was then used to search

for the most parsimonious risk prediction models for CL/P and CPO. The details of the risk prediction models are summarized in **Table 3** and the ROC curves for CL/P and CPO are depicted in **Figures 1** and **2,** respectively. The risk prediction model for CL/P, which comprised of rs2073485 (*IRF6*) and rs7078160 (*VAX1*), achieved an AUC value of 0.556 with 95% CI of (0.542, 0.570). For CPO, rs227731 (*17q22*), rs8610209 (*IRF6*), and rs2514527 (*GDF6*) were selected as risk predictors, and the corresponding AUC value of the model was 0.604 with 95% CI of (0.577, 0.631). While there is a considerable overlap of predictors selected by MLRE and the conventional approach, AUCs of the risk prediction models formed by the convention approach without considering phenotypic heterogeneity were smaller than those built by the MLRE method (**Figures 1** and **2**).

Discussion

Genetic risk prediction studies have been recognized as an important step toward personalized genome medicine. Substantial evidences suggest that the oral clefts with the same or similar clinical manifestations have heterogeneous pathophysiological and etiological mechanisms (Harville et al., 2005; Mossey et al., 2009; Morris et al., 2010; Dixon et al., 2011). In the presence of phenotypic heterogeneity, methods treating all sub-phenotypes as the same outcome are subject to low predictive power, as the effects of genetic variants associated with one unique sub-phenotype are attenuated in the whole population (Mossey et al., 2009; Morris et al., 2010; Dixon et al., 2011). However, most of the genetic prediction studies of oral clefts do not explicitly explore the effect of phenotypic heterogeneity either because there is limited information to classify cases into homogeneous sub-phenotype groups or the sample size for each sub-phenotype is too small for valid statistical inference. Therefore, studies that infer homogeneous sub-phenotype groups based on empirical data could facilitate the identification of new genetic variants, leading to more accurate risk prediction models. In this study, we considered all sub-phenotypes of oral clefts and searched for genetic variants contributing to homogenous sub-phenotype groups of oral clefts. Among the sub-phenotypes considered, left CLP, right CLP, and bilateral CLP showed similar genetic etiologies, while left and right CL shared similar genetic causes. Different from traditional studies, where cleft palate-soft only and cleft palate-hard were treated as a single entity, our analysis suggested that they did not have exactly the same underlying genetic etiology. Although other criteria could be used to define the initial sub-phenotype groups, this study made use of the information from clinical diagnosis to define initial sub-phenotype groups. To the best of our knowledge, no prior risk prediction studies were conducted based on each sub-phenotype of oral clefts with the consideration of both phenotypic heterogeneity and gene-gene interactions. The approach used in this study could serve as the first step toward inferring homogeneous sub-phenotype groups of oral clefts, identifying new genetic variants associated with each sub-phenotype group, and exploring risk prediction models for each sub-phenotype group. With knowledge accumulated through further investigation and validation, a more accurate definition of oral clefts can be established, which might help us to better understand the process of the craniofacial anomalies in embryonic development and to build new diagnostic and therapeutic strategies to prevent the abnormality formed at embryonic stage.

Consistent with most of current findings, the interferon regulatory factor 6 (IRF6) has been selected as an important predictor for all forms of oral clefts (Zucchero et al., 2004; Ghassibe et al., 2005; Scapoli et al., 2005; Richardson et al., 2006; Park et al., 2007; Vieira et al., 2007a; Birnbaum et al., 2009; Grant et al., 2009; Beaty et al., 2010; Mangold et al., 2010). Starting from the 7th week of embryonic development, the palatal shelves rise to a horizontal position above the tongue and come into contact, and IRF6 is one of the most important factors to ensure the palatal shelves rise and adhesive

correctly (Sperber, 2002; Mossey et al., 2009; Dixon et al., 2011). IRF6 transcription is activated by p63, which underlies several malformation syndromes including oral clefts (Sun et al., 2000; Ashique et al., 2002; Thomason et al., 2008; Mossey et al., 2009). The protein encoded by IRF6 contribute to the development of van der Woude's syndrome and popliteal pterygium syndrome, both of which are characterized by various degrees of cleft lip, cleft lip and palate, and CPO (Kondo et al., 2002; Richardson et al., 2006). In addition to IRF6, v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (MAFB) and intron 6 of the ATP-binding cassette subfamily A member 4 (ABCA4) also contribute to the risk of cleft lip and palate (Mossey et al., 2009; Beaty et al., 2010; Huang et al., 2012). The genetic regions close to MAFB harbor numerous binding sites for transcription factors that are known to play a role in palate development (Beaty et al., 2010; Huang et al., 2012). Animal models further confirm the role of MAFB in oral clefts. In mouse, the MAFB mRNA and protein were highly expressed in the epithelium around the palatal shelves and in the medial edge epithelium during palatal fusion (Beaty et al., 2010). Although no evidence of ABCA4 expression has been observed in mouse palatal shelves and no apparent relationship between ABCA4 and oral clefts, ABCA4 may be served as a surrogate to a nearby disease-associated gene (Beaty et al., 2010). Traditionally, cleft lip only and cleft lip with palate are treated as a single entity with homogenous genetic etiology while CPO has its own mechanism, as the cleft lip and primary palate have different developmental originals from the secondary palate (Fraser, 1955, 1970; Dixon et al., 2011). Recently, several studies suggest that the etiology of cleft lip may also be different from that of cleft lip with palate (Harville et al., 2005). Nevertheless, our analysis is not supportive of the completely unified genetic etiology of CL and CLP, although CL only and CLP do share some risk factors. It remains controversial whether cleft lip should be treated the same as cleft lip with palate due to the complex and heterogeneous nature of the disease.

It is noteworthy that based on our analyses cleft palate-hard and cleft palate-soft only should be treated as two separate sub-phenotypes of isolated cleft palate. Our results highlighted the potential that the hard and soft forms of isolated cleft palate have their own mechanisms each though they share some of the genetic risk factors. Treating all forms of isolated CPO as a single entity may jeopardize our ability to uncover important genetic risk predictor predisposing to a specific sub-phenotype. Nevertheless, current studies considered hard and soft cleft palate as variants of the same defect without systematic investigation on each of them. This is partially due to limited knowledge of disease etiology and relatively small sample size of sub-phenotypes of CPO. Our analyses could be served as an initial step toward exploring sub-phenotypes of CPO, and subsequent explorations and validations of etiologies underlying various sub-phenotypes hold great promise to move forward our understanding of cleft palate, which may eventually lead to new diagnostic and therapeutic strategies for CPO.

One possible limitation of the study is that only genes previously reported to be associated with oral clefts were included. Consequently, new genetic variants contributing to a

specific sub-phenotype of oral clefts were not investigated in this study, which could lead to low performance of risk prediction models. A natural extension of current analysis is to conduct a genome-wide risk prediction analysis with the consideration of both gene–gene interactions and phenotypic heterogeneity. Another limitation of this study is related to the sub-phenotype definition. We used the sub-phenotypes defined in the data as our starting point and gradually combined the groups with similar genetic etiologies. The formed risk prediction models could have low performance if the initial sub-phenotype groups do not reflect the underlying genetic etiology. It is worthwhile to investigate other criteria of defining initial groups in future analysis.

Conclusion

We used a newly developed statistical method to form risk prediction models for oral clefts with the consideration of

heterogeneous disease etiology. The method first combined sub-phenotype groups shared similar genetic etiology, and then constructed risk prediction models for each newly formed sub-phenotype group. Further replication and follow-up studies are needed to validate the findings, but our analyses could serve as an initial step toward constructing risk prediction models for homogeneous sub-phenotype groups of oral clefts.

Acknowledgments

References

Ashique, A. M., Fu, K., and Richman, J. M. (2002). Endogenous bone morphogenetic proteins regulate outgrowth and epithelial survival during avian lip fusion. *Development* 129, 4647–4660.

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., et al. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat. Genet.* 42, 525–529. doi: 10.1038/ng.580

Beaty, T. H., Ruczinski, I., Murray, J. C., Marazita, M. L., Munger, R. G., Hetmanski, J. B., et al. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet. Epidemiol.* 35, 469–478. doi: 10.1002/gepi.20595

Birnbaum, S., Ludwig, K. U., Reutter, H., Herms, S., Steffens, M., Rubini, M., et al. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.* 41, 473–477. doi: 10.1038/ng.333

Calzolari, E., Pierini, A., Astolfi, G., Bianchi, F., Neville, A. J., and Rivieri, F. (2007). 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* 143, 528–537. doi: 10.1002/ajmg.a.31447

Carter, C. O., Evans, K., Coffey, R., Roberts, J. A., Buck, A., and Roberts, M. F. (1982). A three generation family study of cleft lip with or without cleft palate. *J. Med. Genet.* 19, 246–261. doi: 10.1136/jmg.19.4.246

Cornelis, M. C., Agrawal, A., Cole, J. W., Hansel, N. N., Barnes, K. C., Beaty, T. H., et al. (2010). The gene, environment association studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet. Epidemiol.* 34, 364–372. doi: 10.1002/gepi.20492

Dixon, M. J., Marazita, M. L., Beaty, T. H., and Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* 12, 167–178. doi: 10.1038/nrg2933

Fraser, F. C. (1955). Thoughts on the etiology of clefts of the palate and lip. *Acta Genet. Stat. Med.* 5, 358–369.

Fraser, F. C. (1970). The genetics of cleft lip and cleft palate. *Am. J. Hum. Genet.* 22, 336–352.

Genisca, A. E., Frias, J. L., Broussard, C. S., Honein, M. A., Lammer, E. J., Moore, C. A., et al. (2009). Orofacial clefts in the national birth defects prevention study, 1997-2004. *Am. J. Med. Genet. A* 149A, 1149–1158. doi: 10.1002/ajmg.a.32854

Ghassibe, M., Bayet, B., Revencu, N., Verellen-Dumoulin, C., Gillerot, Y., Vanwijck, R., et al. (2005). Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. *Eur. J. Hum. Genet.* 13, 1239–1242. doi: 10.1038/sj.ejhg.5201486

Grant, S. F., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C. E., et al. (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.* 155, 909–913. doi: 10.1016/j.jpeds.2009.06.020

Grosen, D., Bille, C., Petersen, I., Skytthe, A., Hjelmborg, J., Pedersen, J. K., et al. (2011). Risk of oral clefts in twins. *Epidemiology* 22, 313–319. doi: 10.1097/EDE.0b013e3182125f9c

Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Molsted, K., Sivertsen, A., et al. (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *J. Med. Genet.* 47, 162–168. doi: 10.1136/jmg.2009.069385

Harville, E. W., Wilcox, A. J., Lie, R. T., Vindenes, H., and Abyholm, F. (2005). Cleft lip and palate versus cleft lip only: are they distinct defects? *Am. J. Epidemiol.* 162, 448–453. doi: 10.1093/aje/kwi214

Huang, E., Cheng, H., Xu, M., Shu, S., and Tang, S. (2012). Association between single-nucleotide polymorphisms on chromosome 1p22 and 20q12 and nonsyndromic cleft lip with or without cleft palate: new data in Han Chinese and meta-analysis. *Birth Defects Res. A Clin. Mol. Teratol.* 94, 469–476. doi: 10.1002/bdra.23013

Jones, M. C. (1988). Etiology of facial clefts: prospective evaluation of 428 patients. *Cleft Palate J.* 25, 16–20.

Jugessur, A., Farlie, P. G., and Kilpatrick, N. (2009). The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral. Dis.* 15, 437–453. doi: 10.1111/j.1601-0825.2009.01577.x

Jugessur, A., Rahimov, F., Lie, R. T., Wilcox, A. J., Gjessing, H. K., Nilsen, R. M., et al. (2008). Genetic variants in IRF6 and the risk of facial clefts: single-marker and haplotype-based analyses in a population-based case-control study of facial clefts in Norway. *Genet. Epidemiol.* 32, 413–424. doi: 10.1002/gepi.20314

Kondo, S., Schutte, B. C., Richardson, R. J., Bjork, B. C., Knight, A., Watanabe, Y., et al. (2002). Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat. Genet.* 32, 285–289. doi: 10.1038/ng985

Little, J., and Bryan, E. (1986). Congenital anomalies in twins. *Semin. Perinatol.* 10, 50–64.

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.* 42, 24–26. doi: 10.1038/ng.506

Mitchell, L. E. (1997). Transforming growth factor alpha locus and nonsyndromic cleft lip with or without cleft palate: a reappraisal. *Genet. Epidemiol.* 14, 231–240. doi: 10.1002/(SICI)1098-2272(1997)14:3<231::AID-GEPI2>3.0.CO;2-8

Morris, A. P., Lindgren, C. M., Zeggini, E., Timpson, N. J., Frayling, T. M., Hattersley, A. T., et al. (2010). A powerful approach to sub-phenotype analysis

in population-based genetic association studies. *Genet. Epidemiol.* 34, 335–343. doi: 10.1002/gepi.20486

Mossey, P. (2003). Global strategies to reduce the healthcare burden of craniofacial anomalies. *Br. Dent. J.* 195, 613. doi: 10.1038/sj.bdj.4810738

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., and Shaw, W. C. (2009). Cleft lip and palate. *Lancet* 374, 1773–1785. doi: 10.1016/S0140-6736(09)60695-4

Murray, J. C. (2002). Gene/environment causes of cleft lip and/or palate. *Clin. Genet.* 61, 248–256. doi: 10.1034/j.1399-0004.2002.610402.x

Park, J. W., Mcintosh, I., Hetmanski, J. B., Jabs, E. W., Vander Kolk, C. A., Wu-Chou, Y. H., et al. (2007). Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. *Genet. Med.* 9, 219–227. doi: 10.1097/GIM.0b013e3180423cca

Prescott, N. J., Lees, M. M., Winter, R. M., and Malcolm, S. (2000). Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. *Hum. Genet.* 106, 345–350. doi: 10.1007/s004390051048

Richardson, R. J., Dixon, J., Malhotra, S., Hardman, M. J., Knowles, L., Boot-Handford, R. P., et al. (2006). Irf6 is a key determinant of the keratinocyte proliferation-differentiation switch. *Nat. Genet.* 38, 1329–1334. doi: 10.1038/ng1894

Scapoli, L., Palmieri, A., Martinelli, M., Pezzetti, F., Carinci, P., Tognon, M., et al. (2005). Strong evidence of linkage disequilibrium between polymorphisms at the IRF6 locus and nonsyndromic cleft lip with or without cleft palate, in an Italian population. *Am. J. Hum. Genet.* 76, 180–183. doi: 10.1086/427344

Sivertsen, A., Wilcox, A. J., Skjaerven, R., Vindenes, H. A., Abyholm, F., Harville, E., et al. (2008). Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives. *BMJ* 336, 432–434. doi: 10.1136/bmj.39458.563611.AE

Sozen, M. A., Suzuki, K., Tolarova, M. M., Bustos, T., Fernandez Iglesias, J. E., and Spritz, R. A. (2001). Mutation of PVRL1 is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. *Nat. Genet.* 29, 141–142. doi: 10.1038/ng740

Sperber, G. (2002). "Formation of the primary and secondary palate," in *Cleft lip and Palate: From Origin to Treatment*, ed. D. Wyszynski (New York, NY: Oxford University Press), 5–24.

Stoll, C., Alembik, Y., Dott, B., and Roth, M. P. (2000). Associated malformations in cases with oral clefts. *Cleft. Palate. Craniofac. J.* 37, 41–47. doi: 10.1597/1545-1569(2000)037<0041:AMICWO>2.3.CO;2

Strauss, R. P. (1999). The organization and delivery of craniofacial health services: the state of the art. *Cleft. Palate. Craniofac. J.* 36, 189–195. doi: 10.1597/1545-1569(1999)036<0189:TOADOC>2.3.CO;2

Sun, D., Baur, S., and Hay, E. D. (2000). Epithelial-mesenchymal transformation is the mechanism for fusion of the craniofacial primordia involved in morphogenesis of the chicken lip. *Dev. Biol.* 228, 337–349. doi: 10.1006/dbio.2000.9946

Thomason, H. A., Dixon, M. J., and Dixon, J. (2008). Facial clefting in Tp63 deficient mice results from altered Bmp4, Fgf8 and Shh signaling. *Dev. Biol.* 321, 273–282. doi: 10.1016/j.ydbio.2008.06.030

Vieira, A. R. (2006). Association between the transforming growth factor alpha gene and nonsyndromic oral clefts: a HuGE review. *Am. J. Epidemiol.* 163, 790–810. doi: 10.1093/aje/kwj103

Vieira, A. R., Cooper, M. E., Marazita, M. L., Orioli, I. M., and Castilla, E. E. (2007a). Interferon regulatory factor 6 (IRF6) is associated with oral-facial cleft in individuals that originate in South America. *Am. J. Med. Genet. A* 143A, 2075–2078. doi: 10.1002/ajmg.a.31884

Vieira, A. R., Modesto, A., Meira, R., Barbosa, A. R., Lidral, A. C., and Murray, J. C. (2007b). Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. *Am. J. Med. Genet. A* 143, 538–545. doi: 10.1002/ajmg.a.31620

Vieira, A. R., Murray, J. C., Trembath, D., Orioli, I. M., Castilla, E. E., Cooper, M. E., et al. (2005). Studies of reduced folate carrier 1 (RFC1) A80G and 5,10-methylenetetrahydrofolate reductase (MTHFR) C677T polymorphisms with neural tube and orofacial cleft defects. *Am. J. Med. Genet. A* 135, 220–223. doi: 10.1002/ajmg.a.30705

Wehby, G. L., and Cassell, C. H. (2010). The impact of orofacial clefts on quality of life and healthcare use and costs. *Oral. Dis.* 16, 3–10. doi: 10.1111/j.1601-0825.2009.01588.x

Wen, Y., and Lu, Q. (2013). A multiclass likelihood ratio approach for genetic risk prediction allowing for phenotypic heterogeneity. *Genet. Epidemiol.* 37, 715–725. doi: 10.1002/gepi.21751

Wyszynski, D. F., Zeiger, J., Tilli, M. T., Bailey-Wilson, J. E., and Beaty, T. H. (1998). Survey of genetic counselors and clinical geneticists regarding recurrence risks for families with nonsyndromic cleft lip with or without cleft palate. *Am. J. Med. Genet.* 79, 184–190. doi: 10.1002/(SICI)1096-8628(19980923)79:3<184::AID-AJMG6>3.0.CO;2-N

Zeiger, J. S., Beaty, T. H., and Liang, K. Y. (2005). Oral clefts, maternal smoking, and TGFA: a meta-analysis of gene-environment interaction. *Cleft. Palate. Craniofac. J* 42, 58–63. doi: 10.1597/02-128.1

Zeiger, J. S., Hetmanski, J. B., Beaty, T. H., Vanderkolk, C. A., Wyszynski, D. F., Bailey-Wilson, J. E., et al. (2003). Evidence for linkage of nonsyndromic cleft lip with or without cleft palate to a region on chromosome 2. *Eur. J. Hum. Genet.* 11, 835–839. doi: 10.1038/sj.ejhg.5201052

Zucchero, T. M., Cooper, M. E., Maher, B. S., Daack-Hirsch, S., Nepomuceno, B., Ribeiro, L., et al. (2004). Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N. Engl. J. Med.* 351, 769–780. doi: 10.1056/NEJMoa032909