Maternal and Parent-of-Origin Effects on the Etiology of Orofacial Clefting

Author: Nikola Rasevic

August 30th, 2021

Supervisor: Dr. Marie-Hélène Roy Gagnon

Co-supervisor: Dr. Julian Little

TAC Member: Dr. Kelly Burkett

School of Epidemiology and Public Health

University of Ottawa

A thesis submitted in partial fulfillment of the requirements for the
Master's degree in Epidemiology

© Nikola Rasevic, Ottawa, Canada, 2021

Table of Contents

Preface III
Author's Contributions IV
Abstract V
Acknowledgements VI
Chapter 1: Introduction 1
Chapter 2: Literature Review 4
Chapter 3: Study Aims 17
Chapter 4: Methods 18
Chapter 5: Manuscript 28
Chapter 6: Additional Results 56
Chapter 7: Discussion 65
References 72

Preface

The thesis is the original work of Nikola Rasevic. The thesis includes a manuscript in chapter 5. The manuscript is titled “Maternal and parent-of-origin gene-environment effects on the etiology of orofacial clefting” and is planned on being submitted to the American Journal of Medical Genetics Part A or the European Journal of Medical Genetics. The thesis investigates the association between genetic effects and gene-environment interaction effects related to orofacial clefts.

Author's Contributions

Study conception and design were done by Marie-Hélène Roy-Gagnon, Michele Rubini, Julian Little and Nikola Rasevic. Nikola Rasevic performed the statistical analysis under the supervision of Marie-Hélène Roy-Gagnon and Kelly Burkett. Nikola Rasevic interpreted the results and wrote the manuscript under the supervision of Marie-Hélène Roy-Gagnon and Julian Little. Michele Rubini and Julian Little were involved in the design of the original EUROCRAN study, and Michele Rubini was responsible for the design of the ITALCLEFT study and for the acquisition of the data used in this study.

Abstract

Objective: To investigate the association of previously reported single nucleotide polymorphisms (SNPs) in relation to orofacial clefts and assess their interaction with environmental factors.

Methods: Genome-wide SNP genotypes were obtained for case-parent triads from the EUROCRAN and ITALCLEFT studies. Candidate SNPs were selected from a previous genome-wide association study (Shi et al., 2012) along with surrounding SNPS for a total of 2142 genotyped and imputed SNPs. A total of 411 case-parent triads and 25 case-parent dyads were analyzed using log-linear models to test for maternal and parent-of-origin effects along with their interaction with maternal smoking and maternal folic acid consumption.

Results: A significant association (q = 0.025) was detected for a region in the *ATXN3* gene. This significance refers to the interaction between maternal periconceptional smoking and maternal genetic effects. Nominally significant associations in genes relating to the brain were also detected.

Conclusion: SNPs in the *ATXN3* region warrant further investigation.

Acknowledgements

I would like to thank my supervisors: Dr. Roy-Gagnon and Dr. Little, for their guidance. I would also like to thank the University of Ottawa for their financial support.

Chapter 1: Introduction

Orofacial clefts (OC) are a subset of congenital anomalies that are classified as cleft lip with or without cleft palate or cleft palate only. These types of OC are thought to have different etiological mechanisms (Murray, 2002; Spritz, 2001). Globally, it is estimated that the prevalence of OC among live births is 1.47 per 1000, which makes it common compared to other congenital anomalies (Panamonta et al., 2015). The prevalence of OC in live births in China is 1.4/1000 (Wang et al., 2017). In low to middle income countries, the prevalence of cleft lip with or without palate is about 1.4/1000 (Kadir et al., 2017).

It is widely understood that OC have a complex etiology that is associated with the interaction of genetic and environmental risk factors. There are a number of genetic syndromes that involve clefting, all likely having their own biological mechanisms. These syndromes include van der Woude's syndrome, Kallman's syndrome, ectrodactyly, ectodermal dysplasia, clefting syndrome, x-linked clefting and ankyloglossia, Gorlin's syndrome and Margarita Island ectodermal dysplasia (Mossey et al., 2009). Furthermore, several environmental risk factors are associated with OC. The attributable risk for these environmental risk factors, which includes maternal smoking and lack of any folic acid supplementation, is 50% for cleft lip with or without palate and 43% for cleft palate only (Raut et al., 2019). In twin studies, there is a higher concordance rate for OC among monozygotic twins compared to dizygotic twins, which indicates genetic predisposition (Beaty et al., 2016). Based on these twin studies, it has been estimated that the heritability is 91% for cleft lip with or without palate and 90% for cleft palate only; however, gene-environment effects were not considered, and the heritability estimates may have been inflated by shared environmental risk (Beaty et al., 2016). About 40 genetic risk regions associated with non-syndromic orofacial clefts (nsOC) have been identified through

genome-wide association (GWA) studies (Thieme and Ludwig, 2017), but as has generally been the case for chronic diseases, these account for only a small proportion of the heritability of OC (Manolio et al., 2009). Therefore, the question of missing heritability arises. Some studies suggest that the missing heritability is explained by rare variants that are not assayed in a GWAS (Basha et al., 2018; Saleem et al., 2019). Another possible explanation is the role of the parent's alleles in the risk of orofacial clefts, which includes maternal and parent-of-origin genetic effects. In maternal genetic effects, the mother's alleles influence the risk on the child's phenotype, and in parent-of-origin genetic effects, the transmission of the allele from mother or father affects the phenotype of the child through epigenetic mechanisms. In fact, there are a wide range of epigenetic effects that influence the risk of OC which include distorted gene expression through DNA methylation as well as histone modification and misexpression of gene-encoding microRNAs (Garland et al., 2020b). As such, it is anticipated that two possible mechanisms that may have an influence on orofacial clefts are maternal genetic effects and parent-of-origin genetic effects.

So far as we are aware, the case-parent triad study of OC of Shi et al. (2012) is the largest genome-wide study to test for both maternal and POO effects. Whilst the study did not identify any loci that passed the genome wide level of significance, there is likely to have been limited statistical power to detect significant effects if these effects are modified by environmental factors. In the present study, the most significant associations from Shi et al. (2012) will be further investigated with consideration of possible interaction with two environmental factors: maternal smoking and maternal use of supplements containing folic acid. The reason that these two risk factors were chosen is because they are consistently associated with OC (Avşar et al., 2021; Hackshaw et al., 2011; Little et al., 2004; Sabbagh et al., 2015; Zhou et al., 2020) and are

known to be associated with other types of congenital anomalies including congenital heart defects (Hackshaw et al., 2011) and neural tube defects (Czeizel and Dudás, 1992; MRC Vitamin Study Research Group, 1991). Therefore, one of the goals of this study is to replicate the parent-of-origin and maternal genetic results of Shi et al. (2012) with a new dataset. Additionally, gene-environment interactions will be examined to help unweave the complex network that is thought to underlie the etiology of OC.

The data used in the present study were obtained in the EUROCRAN and ITALCLEFT studies. In these studies, case-parent triads were recruited across Europe. In general, information on demographics, pregnancy history and lifestyle, including exposure to environmental risk factors during the periconceptional period (three months before to three months after conception) were collected at the time of primary surgery for OC in both studies.

Chapter 2: Literature Review

Classification of OC

Orofacial clefts can be classified in four different ways: CL (cleft lip), CLP (cleft lip and palate), CL/P (cleft lip without cleft palate) and CP (cleft palate) (Dixon et al., 2011). Furthermore, OC can be classified as syndromic or non-syndromic, as well as complete and incomplete. Syndromic OC occurs when there are further malformations with the orofacial clefts. Syndromes associated with teratogens, chromosomal anomalies or single gene syndromes occur in 30% of cases of cleft lip with or without cleft palate and 50% of cases for cleft palate only (Dixon et al., 2011). Complete clefting is described as the reduced thickness of the upper lip, reduced height of the upper face and broadening of the nasal cavity. For incomplete clefting, this does not occur (Smahel and Brejcha, 1983). A variety of other different classification systems have been proposed, none of which is universally accepted (Kernahan and Stark, 1958; McBride et al., 2016). This thesis will be focused on non-syndromic OC.

Known Molecular Pathogenesis

Cell migration, growth, apoptosis and differentiation are required for the development of the cleft and palate (Mossey et al., 2009). The lip and primary palate have different ways of development which allows clefts to be classified as cleft lip with or without cleft palate or isolated cleft palate. In the molecular level, it is considered that the formation of clefts can be influenced by environmental factors such as folic acid metabolism and exposure to tobacco smoke.

Folate is a compound that is critical in the methylation process of many biomolecular reactions. Some of these needed reactions include the methylation of DNA, RNA and proteins

which are essential for the cell proliferation, differentiation and repair (Loenen, 2006). Folate enters the complex folic acid cycle where it is eventually converted into 5-methyl-THF. 5-methyl-THF is transported to the cell membrane where it begins the one-carbon cycle. 5-methyl-THF is then converted to SAM (S-adenosylmethionine). SAM is responsible for the methylation of biomolecules that may influence cell proliferation, differentiation and repair (Garland et al., 2020a).

Tobacco smoke is another complicated environmental exposure as there are many different compounds within that can give rise to OC. A study illustrated that gene expression involved with cell cycle regulation, DNA repair and oxidative stress response are affected by tobacco smoke in mice (Izzotti et al., 2003). Additionally, nicotine and carbon monoxide can induce fetal hypoxia which results in an increased risk of OC. Nicotine harms the vascular function of the uterus which would disturb oxygen concentrations in the fetus. Carbon monoxide forms carboxyhemoglobin in red blood cells which halts the transportation of oxygen to the fetus' cells (Garland et al., 2020a).

Impact of OC on the Index patient, the Family and Society

OC can affect the mental and physical health of a child in several different ways, including effects on psychology, speech, hearing, appearance and cognition. As a result, multidisciplinary care throughout childhood and the teenage years for those affected from OC is needed. Children diagnosed with OC have been reported to have higher morbidity and mortality rates (Berg et al., 2016; Christensen et al., 2004; Malic et al., 2020; Mossey et al., 2009; Ngai et al., 2005; Stock et al., 2015). The care that may be required for children born with OC includes nursing, plastic surgery, maxillofacial surgery, otolaryngology, speech therapy, audiology, counselling, psychology, genetics, orthodontics and dentistry. Despite all the available resources,

these different types of care are often disjointed, which results in variation in management. OC not only affects the child; it can also have psychological and financial effects on the family. It has been estimated that the direct cost of treatment of OC is $10,000-$13,000 in European populations (Galloway et al., 2017). For children below the age of 10, the difference between the annual mean cost of raising a child for those with orofacial clefts and those without is $13,405 (Boulet et al., 2009).

Epidemiology of OC

In Canada, the prevalence of OC at birth was 1.63 per 1000 from 1998 to 2007 for total births. Of 5599 cases of OC, 70 had another congenital anomaly and 95 were stillborn (Little and Nelson, 2013). From 2002 to 2008, the prevalence of OC in Canada ranged from 1.10 to 1.53 per 1000 live births. The distribution of cleft type was 17% for cleft lip, 41% for cleft palate and 42% for cleft lip and palate. 62% of cleft lip cases, 66% of cleft lip and palate cases, and 44% of cleft palate cases were comprised of males (Pavri and Forrest, 2013).

It is estimated that the worldwide prevalence of OC among live births is 1.47 per 1000 (Panamonta et al., 2015). Populations derived from Europe have cleft lip and palate prevalence rates of 1 in 1000 (Mossey et al., 2009). Italian and Hungarian registries from 2012 indicated that 0.6% of babies with OC were stillborn and 11% occurred from terminated pregnancies (ICBDSR, 2015). Seasonal patterns and time trends of the prevalence of OC were too inconsistent to form any conclusions (Castilla et al., 1990; Fraser and Gwyn, 1998). Isolated cleft palate is more frequent in females, whereas cleft lip with or without cleft palate are more common in males. The sex ratio for cleft lip with or without cleft palate is about 2:1 for males to females in white populations (Mossey et al., 2009). OC are also associated with other congenital anomalies. A study was conducted that observed that out of roughly 4000 individuals with cleft

palate without cleft lip from Europe, 18% of cases had another congenital anomaly. For cleft lip, a study with over 5000 patients found that 29% of the cases had another anomaly (Calzolari et al., 2007).

Biological Mechanisms of Maternal and Parent-of-Origin Genetic Effects

It is important to consider the role of the parent's genes in regard to an early onset disease/disorder. Two genetic effects will be considered: parent-of-origin and maternal.

Parent-of-origin genetic effects occur when the phenotypic effect of an allele is dependent on whether it is obtained by the mother or father (Lawson et al., 2013). There are many different mechanisms that can result in parent-of-origin genetic effects. One of these ways is genomic imprinting, which can be seen in figure 2.1. Genomic imprinting is a mode of inheritance where the regulation of a gene is dependent on whether the offspring obtains the gene from the mother or father. Through various mechanisms (DNA methylation, histone modification, etc.), the gene is marked according to which parent the gene was received from. This results in an expression of the gene that is parent specific (Macdonald, 2012).

Maternal genetic effects have been defined as "the causal influence of the maternal genotype or phenotype on the offspring phenotype" (Wolf and Wade, 2009). Half of the offspring's genome is contributed by the mother. However, this is not the only form of influence that the mother has on the child, as the mother also provides an environment in which the fetus develops. This is important to consider when investigating congenital anomalies such as OC, since the intra-uterine environment can play a role in affecting the risk of a child having OC. Variations in the mother's genome, as well as her exposures, can impact the intra-uterine environment which can give rise to OC.

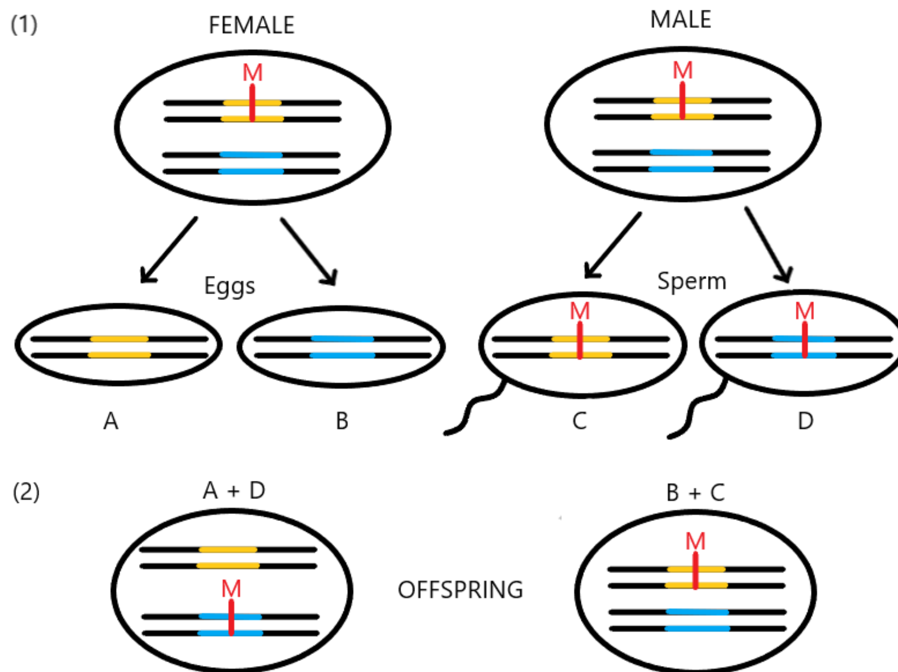

Figure 2.1. An example of a genomic imprinting mechanism that can be a cause of parent-of-origin effects (Alberts et al., 2015). (1) A somatic cell from a female and a somatic cell from a male contain a pair of homologous chromosomes. For both somatic cells, the top homolog is inherited by the father and as such, the allele is subject to imprinting by methylation (note: there are other forms of imprinting) as indicated by the red-letter M. This imprinting silences the expression of the allele. The bottom homolog is inherited by the mother and the allele is not subject to imprinting, so it is expressed. During germ cell formation, the imprinting is erased for both the egg (A and B) and sperm (C and D) and is re-established for the sperm. (2) Pictured are possible offspring of the combination of egg A and sperm D as well as egg B and sperm C. The cell on the left is imprinted on one allele while the cell on the right is imprinted on another allele for the same gene. Thus, if the two alleles differ, then the two cells may express differences in

phenotype despite having the same DNA sequences of the gene. This is dependent on whether the gene was inherited by the father or mother (Alberts et al., 2015).

Lifestyle and Environmental Risk Factors

Several maternal risk factors have been identified to be associated with OC including tobacco smoke, folic acid supplementation, alcohol, weight, poor nutrition, teratogens, medicinal drugs and viruses (Chen et al., 2007; Izedonmwen et al., 2015; Kutbi et al., 2017; Mossey et al., 2007; Stothard et al., 2009). Associations related to maternal alcohol consumption in OC are inconsistent, as positive associations were reported for some studies (Bille et al., 2007; DeRoo et al., 2016; Romitti et al., 1999; Yin et al., 2019) but not for others (Meyer et al., 2003; Romitti et al., 2007). This was the case for both heavy and moderate alcohol consumption. Maternal obesity and low weight have been both associated with OC (Stothard et al., 2009; Waller et al., 2007). These conditions could influence the amount and balance of nutrients consumed by women. Children with orofacial clefts and their mothers have been reported with higher levels of homocysteine in the blood (van Rooij et al., 2003; Wong et al., 1999). A cofactor to homocysteine is vitamin B6 and has shown a reduction of OC in animals (Miller, 1972). Additionally, biomarkers indicating low vitamin B6 levels in people increase the risk of OC in populations in the Netherlands (Wong et al., 1999) and Philippines (Munger et al., 2004). Another compound associated with OC is Zinc. Mothers of children affected with OC have lower concentrations of zinc than mothers with children not affected by OC (Krapels et al., 2004). A dose-response relationship of zinc and OC was observed in the Philippines, with high concentrations of zinc in the plasma being associated with a low risk of OC (Tamura et al., 2005). Retinoids such as retinol, or vitamin A may play a role in OC, as exposure to these compounds have given rise to instances of craniofacial abnormalities (Lammer et al., 1985;

Mitchell et al., 2003; Rothman et al., 1995; Warkany and Nelson, 1940). Exposure of mothers to organic solvents and agricultural chemicals, suggested to be risk for OC in some studies, have not been found to show a clear association with OC in larger studies with detailed assessment of exposure (Shaw et al., 2003). Exposure to anticonvulsant drugs has been reported to be associated with craniofacial abnormalities (Abrishamchian et al., 1994; Dravet et al., 1992; Shaw et al., 1995). Finally, there have also been associations reported with maternal corticosteroid use during pregnancy (Park-Wyllie et al., 2000).

The environmental factors that will be investigated include maternal periconceptional use of supplements containing folic acid and maternal smoking. It has been shown that folic acid supplementation during the periconceptional period reduces the risk of neural tube defects and may lower the risk of other types of congenital anomalies (Czeizel and Dudás, 1992; MRC Vitamin Study Research Group, 1991). In a meta-analysis, Zhou et al. (2020), found an inverse association between the use of folic acid containing supplements and cleft lip with or without palate with an Odds Ratio of 0.6 (0.51 - 0.69). There may be a possible interaction with hypothermia during pregnancy and folic acid supplements where the supplements reduce the increased risk of OC during hypothermia (Botto et al., 2002). Dosage may also play a factor when evaluating the risk of OC with supplements. A randomised control trial randomly allocated a low dose of folic acid (0.4 mg) and high dose (4.0 mg) to mothers who were at risk of recurrence of OC. The recurrence rate was similar between the low dose group and the high dose group, but there was a suggestion that it was reduced in the high dose group (Wehby et al., 2013). However, the recruitment of participants in this trial did not reach the desired sample size to detect a clinically important difference between the exposure groups. The effects of dietary

folate intake are still unknown. Results are inconsistent across countries, in part because some countries have mandatory folic acid fortification programs.

The other environmental factor to be considered is maternal smoking. A consistent positive association has been observed between OC and smoking (Avşar et al., 2021; Hackshaw et al., 2011; Little et al., 2004; Sabbagh et al., 2015). Passive maternal smoking during pregnancy increases the risk of a child being born with non-syndromic OC by 1.5 times (Sabbagh et al., 2015). Smoking during pregnancy has shown an increased risk of OC with a population-attributable risk of 20% (Honein et al., 2007; Little et al., 2004). Additionally, maternal smoking raises the possibility of genes in metabolic pathways, e.g. xenobiotic metabolism, contributing to the development of OC.

Genetic Risk Factors

Offspring Genetic Effects

Approaches to investigating offspring genetic effects in relation to OC have included linkage studies; investigation of candidate genes selected on the basis of (a) putative effects on metabolism of exposures that may be of etiological importance, (b) involvement in chromosome breaks associated with OC, and (c) involvement in specific genetic syndromes in which an OC is a component; and genome wide association studies. Linkage studies identified various risk loci as having a causal role in cleft lip and palate (Mossey et al., 2009). These genes include *MTHFR, TGFA, D4S175, F13A1, TGFB3, D17S250,* and *APOC2* (Prescott et al., 2000; Zeiger et al., 2003). Mossey et al. (2017) reported a study that focused on candidate genes putatively affecting metabolism or etiological pathways and gene-gene and gene-environment effects in EUROCRAN data, which recruited trios in which the index individuals had non-syndromic cleft

lip with or without palate and cleft palate only. A strong association was found for the *MTHFR-TGFA* gene-gene effect in cleft palate only. Specifically, an inverse association was observed between cleft palate and *MTHFR* in the presence of a *TGFA* Taq1 deletion. One strategy of identifying potential candidate genes is by looking at chromosomal aberrations (Mossey et al., 2009), which led to the discovery that *SATB2* has a key role in the development of the secondary palate (FitzPatrick et al., 2003). Other genes that were considered on the basis of chromosomal aberrations include *PVRL1* and *GAD1* (Jakobsen et al., 2006). These aberrations give rise to the Pierre-Robin sequence, which is a congenital anomaly that is usually characterized by cleft lip (Baxter and Shanks, 2020). OC are listed to be a feature of more than 200 genetic syndromes, and it is estimated that the percentage of OC that are associated with these syndromes is 5-7% (Mossey et al., 2009). Thus, common variants of genes that are involved in Mendelian syndromes with an OC may be worth investigating as candidates involved in non-syndromic orofacial clefts. Mutations of *IRF6* are causally associated with van der Woude's syndrome, and this led to consideration of whether common variants of *IRF6* might be associated with nsOC; such an association was found and replicated in several populations (Mossey et al., 2009; Zucchero et al., 2004). An example of a genome wide association study is the investigation by Yu et al. (2017) of solely offspring genetic effects on an additive scale, who found multiple loci associated with non-syndromic OC (mapping to *TAF1B, MSX1, SPRY1, FGF10, OFCC1/TFAP2A, FGFR1, RAD54B, PTCH1, KRT18, RPS26, TMEM19, LINC00640, GSC/DICER1, WNT9B, ARHGAP29, IRF6, FAM49A, MMP16, MYC, GADD45G, VAX1, SPRY2, CREBBP, NTN1, NOG, MAFB*).

Maternal and Parent-of-Origin Genetic effects

Gjerdevik et al. (2018) reanalyzed some of the most significant SNPs from Shi et al. (2012) by testing solely parent-of-origin genetic effects on cleft palate (a multiplicative model was used for the maternal effects: a partly co-varying effect). They found a strong signal associated with the rs7516430 variant of *CHD1L* ($p = 5.6 \times 10^{-5}$). Haaland et al. (2019), in a genome wide scan (using an additive dose model for maternal effects) on cleft lip with or without palate, reported signals for SNPs in the *LYZL1* gene and the *NC* gene for the parent-of-origin genetic effects solely. Jugessur et al. (2010) tested for maternal genetic effects for several candidate genes using an additive risk model and found weak evidence for variants of *FLNB*, *HIC1* and *ZNF189* for individuals with cleft palate only. Of the three genes, *FLNB* was the only one to have appeared as one of the most significant in two different sets of analysis. Mossey et al. (2017) found some evidence of a maternal homozygous *MSX1* genotype resulting in a doubling of risk for female offspring as it pertains to nsOC. Finally, Shi et al. (2012) also performed a GWAS for maternal and parent-of-origin effects using an additive risk model for maternal effects. They investigated cleft lip with or without palate or cleft palate only. No significant association was reported.

Gene-Environment Interactions

With regard to gene-environment interactions for OC, there are three main environmental factors studied: smoking, use of supplements containing folic acid (including multivitamin) and alcohol consumption.

For the maternal gene-smoking interaction, Beaty et al. (2011) conducted a genome-wide analysis on cleft palate only and found an increased risk of OC with the SNPs in maternal genes:

TBK1 and *ZNF236*. Genetic effects were modelled using an additive model. Haaland et al. (2019) found three weakly associated SNPs in the *ANK3* gene for an interaction between the parent-of-origin genetic effects and smoking for cleft lip with or without palate with a minimum p-value of 2.6^{-06}. Interestingly, Halaand et al. (2017) performed a similar genome-wide analysis (using a multiplicative model for maternal effects) to investigate parent-of-origin gene-smoking interaction effects for cleft palate only but reported different genes that were nominally significant: *ICE1*, with a minimum q-value of 0.14 and minimum p-value of 6^{-07} as well as *NAALDAL2* with multiple hits, although none being very significant. A q-value is a correction for multiple testing using the false discovery rate (FDR) and can be interpreted similarly to p-values (Benjamini and Hochberg, 1995). In Haaland et al. (2018), a genome-wide scan using a multiplicative dose response model, no associations were observed for child gene-smoking interactions for cleft lip with or without cleft palate. In Lie et al. (2008), a candidate gene study, there were no observed interaction effects for smoking and maternal genetic effect for cleft lip with or without cleft palate and cleft palate only.

For folic acid/multivitamin supplement consumption, SNP alleles in *BAALC* displayed a risk reduction in mothers who took multivitamins only (Beaty et al., 2011). Halaand et al. (2017) reported no significant interaction effects for the parent-of-origin gene-multivitamin interaction effect. For offspring gene-multivitamin interaction, Haaland et al. (2018) discovered three SNPs in the *ESRPG* gene with relatively large associations in a pooled analysis of cleft lip with or without palate, with the smallest q-value being 0.011. Skare et al. (2012), observed a relatively significant SNP for the child gene-multivitamin interaction in the *TBX4* gene for cleft palate only (q = 0.0496), but no significant SNPs were found between the child gene-folic acid supplements interaction. As the authors stated, this could be due to a lack of statistical power. Boyles et al.

(2009), a candidate SNP study, did not find evidence of association for interaction effects for multivitamin use and maternal genetic effect for cleft lip with or without cleft palate and cleft palate only.

A genome wide search for interactions between gene and alcohol consumption was conducted (Haaland et al., 2018). Potential effects with SNPs: rs17734557 and rs1316471 were observed, however, these SNPs were not located in a region with a strong association to OC (Haaland et al., 2018). Beaty et al. (2011) displayed an increased risk in the SNPs located in genes: *MLLT3* and *SMC2* for maternal alcohol consumption. Furthermore, it was observed that mothers carrying a certain *ADH1C* haplotype were at an increased risk for cleft lip and palate (Boyles et al., 2010). This haplotype is associated with decreased alcohol metabolism. For parent-of-origin gene and alcohol consumption interaction, Haaland et al. (2017) did not observe any significant SNPs for cleft palate only. The most significant SNP that they did observe was in the *NC* gene with a q-value of 0.5.

In short, previous literature indicates that there are inconsistent signals that occur throughout the genome with multiple different effects having a role. Although this may seem to be unconvincing, this is not surprising when considering that OC is a condition of complex etiology in which multiple genes, environmental exposures and different types of genetic effect may play a role (Shi and Weinberg, 2011). In general, the sample size of these studies did not have enough power to detect very small relative risks.

Shi et al. (2012)

Shi et el. (2012) conducted a genome-wide association study that investigated the parent-of-origin and maternal genetic effects on the etiology of OC using a multi-ethnic sample of 2458 case-parent trios. Results of the tests for the maternal genetic effects and parent-of-origin genetic

effects displaying associations at the $p < 10^{-5}$ level is shown in tables 4.1 and 4.2. No single nucleotide polymorphisms were deemed to be significant at the genome wide level ($p = 5x10^{-8}$). However, some SNPs were close to significance with large effect sizes. Thus, it is worthwhile to investigate the most significant SNPs from this study using a different data set. For less common conditions such as orofacial clefts, the application of stringent GWAS thresholds may result in false negative associations with specific genetic loci that are etiologically relevant (Ioannidis et al., 2011; Kaler and Purcell, 2019). One concept to consider for replication studies is the Winner's curse phenomenon, first described in studies referring to the statistical outcomes of auctions (Bazerman and Samuelson, 1983). Thus, the most significant SNPs from the Shi et al. (2012) results are likely to have inflated estimated effect sizes, and when conducting a replication, the observed effect sizes may be smaller. In the present study, the most significant SNPs from Shi et al. (2012) will be tested under a correction for multiple testing using a false discovery rate approach. Additionally, the consideration of gene-environment interactions may result in detection of additional genetic susceptibility loci (McAllister et al., 2017). It is important to study gene-environment effects to (i) understand the biomolecular mechanisms of the disease, (ii) use the gene-environment effects to improve prognosis and treatment and (iii) identify groups of individuals that may be at higher risk (McAllister et al., 2017).

Chapter 3: Study Aims

<u>Objective</u>

The outcome of the study is OC, and the exposures are target SNPs identified from Shi et al. (2012) and their possible interactions with maternal active smoking or use of supplements containing folic acid. The present study will only include European triads, whereas Shi et al. (2012) included European and Asian triads. There are many different European countries represented in the present study that cover Northern, Central, Southern and Eastern Europe, whereas the European based sample in Shi et al. (2012) was predominantly based in Northern Europe and United States. As such, the present study may be more generalizable for European populations.

Aim 1. Replicate the most significant associations between SNPs and OC observed in Shi et al. (2012) using new data on European case-parent triads. This is the replication portion of the thesis, and the results of this replication are presented in Chapter 6.

Aim 2. Identify whether the environmental factors of smoking and use of supplements containing folic acid modify the associations identified in Aim 1. The results are presented in Chapter 5.

Chapter 4: Methods

Study Population and Data Collection

The sample is comprised of individuals from the European Collaboration on Craniofacial Anomalies study (EUROCRAN) and ITALCLEFT (Mossey et al., 2017). Individuals recruited consisted of children with non-syndromic OC and their mothers and fathers for a total of 1536 (523 families) individuals. In both studies, mothers answered a questionnaire during the visit when the child was to be assessed for primary surgery. Information on socio-demographic factors, pregnancy history, clinical diagnosis and exposure to environmental risk factors during the periconceptional period were collected using a questionnaire. Particularly, the environmental risk factors of folic acid supplementation and smoking are of interest. A mother was identified as using folic acid supplementation if she took at least 0.4 mg/day of folic acid or folic acid-containing supplements for a minimum of one month during the periconceptional period. Mothers were also identified as smokers if they smoked at least one cigarette/day during the periconceptional period. Peripheral blood specimens or buccal cell samples were collected from the children and their parents in both studies. Samples were genotyped using an Illumina GSA array: GSAMD V 1.0 chip and recommended genotype calling procedures.

The design of the study utilizes case-parent triads. The case-parent triad study is potentially a particularly appropriate design in relation to congenital anomalies and other disorders in childhood. In the case-parent triad design, a child diagnosed with a specific disorder (OC in this case) and his/her parents are genotyped. Thus, when testing for a genetic effect of the offspring's genes, the controls are effectively the non-transmitted alleles of the parents (Ainsworth et al., 2011).

The outcome of the study is non-syndromic OC. All forms of OC: CL/P, CLP and CP will be used in a combined analysis. Exposure factors include the target SNPs for maternal and parent-of-origin genetic effects in the Shi et al. (2012) study and environmental factors: maternal smoking and maternal use of supplements containing folic acid.

Quality Control and Selection of SNPs for Analysis

Genotype data was received in three different batches of individuals across all chromosomes. Using Illumina Genome Studio, three different files were created and merged. Refer to the manuscript methods section for information regarding the data processing and quality control. Additionally, duplicate individuals in the data were removed (thus, one individual was removed per duplicate).

Phasing was accomplished using SHAPEIT (Delaneau et al., 2013a; Delaneau et al., 2011; Delaneau et al., 2013b; O'Connell et al., 2014) software, and the outputted phased haplotypes underwent SNP imputation using the Michigan imputation server (Das et al., 2016). Imputation is achieved using a reference panel, in this case: HRC 1.1 (2016) (McCarthy et al., 2016), to observe haplotypes containing genotyped and non-genotyped markers. Using the haplotypes, probabilities of the ungenotyped variants in the haplotypes are computed. A random sample is taken, and the results are used to be imputed into the genotyped dataset. PLINK (Purcell et al., 2007) was used for the data processing and quality control steps outlined below. SNPs that were within 10,000 base pairs of the target SNPs from Shi et al. (2012) were extracted. The target SNPs were the 25 most significant SNPs for maternal genetic effects (Table 4.1) and 28 most significant SNPs for parent-of-origin genetic effects (Table 4.2) reported by Shi et al. (2012). After quality control, the final sample size for analysis was 1283 individuals (411 triads + 25 dyads) with 2142 SNPs (949 SNPs for maternal genetic effects and 1193 SNPs for parent-

of-origin genetic effects). Of the 2142 SNPs, 1999 were imputed (872 SNPs for maternal genetic effects and 1127 SNPs for parent-of-origin genetic effects). Python scripts (Van Rossum and Drake Jr., 1995) were written by the author in order to prepare the data format for statistical analysis.

Table 4.1. Most significant SNPs in testing maternal genetic effects (Shi, 2012). The 25 most significant SNPs were selected.

Marker	Cytogenetic region	Phenotype	P-value	Gene
rs17138064	17q12	CP	5.03-07	Intergenic
rs10174126	2q37.2	CP	6.69-07	*SH3BP4*
rs745080	14q22.1	CP	7.72-07	*TXNDC16*
rs527589	11q14.1	CL/P	2.73-06	*PAK1*
rs4505466	2q37.2	CP	3.13-06	*SH3BP4*
rs11228719	11q12.1	CP	3.67-06	Intergenic
rs4703822	5q14.1	CP	3.69-06	*RASGFR2*
rs3764628	19p13.11	CL	4.01-06	*KLHL26*
rs1417437	1p31.1	CLP	4.39-06	*LRRC7*
rs2068361	6p24.3	CL	4.87-06	*BMP6*
rs3006564	10p11.23	CLP	5.15-06	Intergenic
rs1329189	10q26.2	CLP	5.26-06	Intergenic
rs17079928	13q12.12	CL/P	5.81-06	*SPATA13*
rs17807815	14q32.12	CP	6.31-06	*ATXN3*
rs212016	3p14.2	CL	7.22-06	*FHIT*
rs1450100	3p26.1	CL	8.16-06	*GRM7*
rs10905099	10p14	CP	8.39-06	Intergenic
rs1417437	1p31.1	CL/P	8.65-06	*LRRC7*
rs16948813	16q12.1	CL	1.04-05	*CYLD*
rs7197376	16p12.3	CL/P	1.06-05	Intergenic
rs1921817	2q14.3	CL	1.07-05	Intergenic
rs4791959	17p13.1	CL	1.08-05	Intergenic
rs1397056	11q12.1	CP	1.09-05	Intergenic
rs9445243	6q16.1	CP	1.10-05	Intergenic
rs10886648	10q26.12	CP	1.12-05	Intergenic
rs11228710	11q12.1	CP	1.12-05	*OR5AR1*

Table 4.2. Most significant SNPs in testing parent-of-origin genetic effects (Shi, 2012). The 28 most significant SNPs were selected.

Marker	Cytogenetic region	Phenotype	P-value	Gene
rs2523375	17q23.1	CL/P	1.13-12	*TBC1D3P1-DHX40P1*
rs9630717	17p11.2	CLP	2.11-10	Intergenic
rs9630717	17p11.2	CL/P	1.19-07	Intergenic
rs2523375	17q23.1	CL	2.97-07	*TBC1D3P1-DHX40P1*
rs2383658	1q31.1	CLP	1.29-06	Intergenic
rs1339083	1q31.1	CLP	1.44-06	Intergenic
rs1834570	8p23.2	CP	1.48-06	*CSMD1*
rs7129890	11p11.12	CL	2.09-06	Intergenic
rs10469858	2q12.1	CP	2.85-06	Intergenic
rs10811366	9p24.3	CLP	3.03-06	*SMARCA2*
rs92833	5p15.33	CLP	3.31-06	Intergenic
rs3769528	2p22.3	CP	3.49-06	*LTBP1*
rs4921798	8p22	CL/P	4.42-06	*MTUS1*
rs10150163	14q31.3	CL/P	4.48-06	*KCNK10*
rs11153238	6q21	CL/P	4.66-06	Intergenic
rs960925	4q28.1	CL/P	4.77-06	Intergenic
rs2523375	17q23.1	CLP	4.81-06	TBC1D3P1-DHX40P1
rs2196457	12q23.1	CL/P	5.07-06	*CCDC38*
rs2591089	14q24.3	CL	5.08-06	Intergenic
rs6834028	4q28.1	CL/P	5.19-06	Intergenic
rs8054408	16p12.3	CL	5.30-06	*SYT17*
rs12144639	1q32.3	CP	5.70-06	Intergenic
rs92833	5p15.33	CL/P	5.80-06	Intergenic
rs1662695	9q21.33	CLP	5.91-06	*NTRK2*
rs1437903	2q21.2	CL/P	6.60-06	*NCKAP5*
rs2401756	14q31.3	CL/P	6.84-06	*KCNK10*
rs7516430	1q21.1	CP	7.66-06	*CHD1L*
rs4765855	12p13.33	CP	1.04-05	*CACNA2D4*
rs16833488	1q23.3	CLP	1.22-05	*PBX1*
rs960722	1p33	CL	1.28-05	Intergenic
rs3774581	3p21.1	CLP	1.31-05	*CACNA1D*
rs10490051	2p22.1	CL	1.35-05	*SLC8A1*

Statistical Analysis

All statistical analysis was performed using R (R Core Team, 2019) software. Exploratory data analysis was conducted on the phenotype and covariate data, which involved evaluating distributions of key variables like smoking and folic acid consumption and considering missing data and erroneous values.

Tests of Association for Maternal and Parent-of-Origin Genetic Effects

Haplin (Gjessing and Lie, 2006) is an R package that uses a log-linear model and maximum likelihood to test for both parent-of-origin and maternal genetic effects. Log-linear modelling is an adaptable method that can take various genetic effects under consideration such as offspring, maternal and parent-of-origin. Log-linear modelling also has the capability of handling incomplete data in an efficient manner with the use of the expectation-maximization algorithm (Weinberg, 1999). Haplin returns a set of relative risks, their confidence intervals, and p-values for each individual SNP (Weinberg et al., 1998). Formulas 1 to 6 illustrate how the log-linear model estimates the maternal and parent-of-origin genetic effects.

The log-linear model is used to describe the frequencies of all the different possible triads dependent on the offspring having the disease (Weinberg et al., 1998). We assume a single locus with $K = 2$ alleles: A_1 and A_2. Their respective population frequencies are p_1 and p_2. Given this, the frequency of the triad type (A_iA_j, A_kA_l, A_jA_l) is given by n_{ijkl} such that $1 \leq i,j,k,l \leq 2$. Genotype A_iA_j represents the mother's genotype, A_kA_l represents the father's genotype and A_jA_l represents the child's genotype. It is expected that the observed triad type frequencies: n_{ijkl} have expected cell values proportional to $P(M, F, C/D)$, where M represents the mother's genotype, F represents the father's genotype, C represents the child's genotype and D represents disease.

Using Bayes theorem, this can be expressed in formula 1. $P(M,F,C)$ is equal to $p_i p_j p_k p_l$ and $p(D)$ is a normalizing constant. This leaves $P(D/M,F,C)$ as the only un-defined expression.

$$\text{Formula 1: } P(M,F,C|D) = P(D|M,F,C) * P(M,F,C)/P(D)$$

The simplest probability of disease is estimated through child effects. With this effect, the probability of disease given all the genotypes of a triad is equivalent to the probability of disease given the child's genotype only, which is expressed in formula 2.

$$\text{Formula 2: } P(D|M,F,C) = P(D|C) = P(D|A_j A_l)$$

For child effects, the probability of the disease given the child's genotype can be expressed in formula 3. R_j and R_l are single dose relative risks for alleles j and l respectively, and R_{jl}^* provides the double dose relative risk of the expression when $j = l$. With this, the log-linear model can be formed.

$$\text{Formula 3: } P\left(D\middle|A_j A_l\right) = nR_j R_l R_{jl}^*$$

Using formulas 1 and 3, the expression in formula 4 is obtained. $\mathcal{E}$ is defined as a normalizing constant. Formula 4 can then be adjusted to take on the shape of a log-linear model, but this form will be shown due to ease of displaying the other effects. In essence, the log-linear model for estimating child effects solely is based on formula 4.

$$\text{Formula 4: } \mathcal{E}_{ijkl} = \mathcal{E} * p_i p_j p_k p_l * R_j R_l R_{jl}^*$$

The expression used for the basis of the log-linear model for estimating maternal genetic effects is given in formula 5. This expression would be the equivalent to formula 3 but it also considers the maternal genetic effects. M is the equivalent to R, but for the mother's genotype. This model assumes mating symmetry which is defined as $P(A_i A_j, A_k A_l) = P(A_k A_l, A_i A_j)$.

$$\text{Formula 5: } P\left(D\middle|A_jA_l\right) = n * R_jR_lR_{jl}^* * M_iM_jM_{ij}^*$$

The expression used for the basis of the log-linear model for estimating parent-of-origin genetic effects is given in formula 6. As seen in the expression, maternal genetic effects are accounted for. The main difference is observed in the single dose relative risk. R_j and R_l are separated into $R_j^{(M)}$ and $R_l^{(F)}$ depending on whether the allele is transmitted from the father or mother.

$$\text{Formula 6: } P\left(D\middle|A_jA_l\right) = n * R_j^{(M)}R_l^{(F)}R_{jl}^* * M_iM_jM_{ij}^*$$

Log-linear models implemented in Haplin assume Hardy-Weinberg equilibrium and random mating. The Wald test used in Haplin compares the relative risks/relative risk ratios of each SNP between triads where mothers were exposed to the environment variable versus those who have not been exposed to the environment variable. The maternal gene relative risks were estimated by maximum likelihood from log-linear models using the mother's genotype. The parent-of-origin relative risk ratios were estimated by considering the transmission of an allele from a particular parent to child. The relative risk ratio is calculated as the relative risk of an allele inherited by the mother divided by the relative risk of an allele inherited by the father as seen in formulas 7, 8 and 9 (Gjerdevik et al., 2018). When testing for parent-of-origin genetic effects, maternal genetic effects were included in the model, however, when testing for maternal genetic effects, parent-of-origin genetic effects were not included.

$$\text{Formula 7: } RR_{mother} = \frac{P(D|pat = a, mat = a_2)}{P(D|pat = a, mat = a_1)}$$

$$\text{Formula 8: } RR_{father} = \frac{P(D|pat = a_2, mat = a)}{P(D|pat = a_1, mat = a)}$$

$$\text{Formula 9: } RRR_{Parent-of-Origin} = RR_{mother}/RR_{father}$$

We applied the log-linear model described above to test for maternal and parent-of-origin genetic effects and their interactions with maternal smoking and folic acid supplementation. This can be seen in figure 4.1. For the maternal effect, we estimated relative risks using a multiplicative dose response. This means that relative risks were modelled such that the relative risk of a genotype having two minor alleles is the square of a genotype having one minor allele. Haplin handles missing genotype data using an expectation-maximization algorithm. Nominal p-values from the different SNP and region association tests are displayed using "Manhattan plots" generated using Python. Regions were tested in two different ways. The first way used a false discovery rate correction on the most significant SNP. The second method conducted adaptive sum of power (aSPU) tests on all the SNPs within the region (Kwak and Pan, 2016). The aSPU test estimates the power of several sum of power (SPU) score tests and selects the most powerful test. The SPU tests contain many different tests such as the burden, variance component and univariate tests (Pan et al., 2014). The false discovery rate method was used to account for multiple testing, and thus, q-values were reported. Q-values can be defined as the expected proportion of false positives incurred when a test is statistically significant. While the p-value is a measure of the false positive rate, the q-value is a measure of the false discovery rate, which is defined as the proportion of significant tests that are not truly significant (Benjamini and Hochberg, 1995). In essence, q-values are corrected p-values, and can be interpreted similarly. For example, let us say that a region has a q-value of 0.05. This means that 5% of regions that are as significant, or more than this region are false positives. Unlike the Bonferroni correction, which controls for the Family Wise Error Rate (FWER), the q-value provides more liberal estimations of the significance of each test as a result of the correlation structure.

Mathematically, the q-value is computed as the minimum false discovery rate value of a test. The false discovery rate of a test is computed as the expected number of false positives divided by the expected number of significant tests (Storey and Tibshirani, 2003).

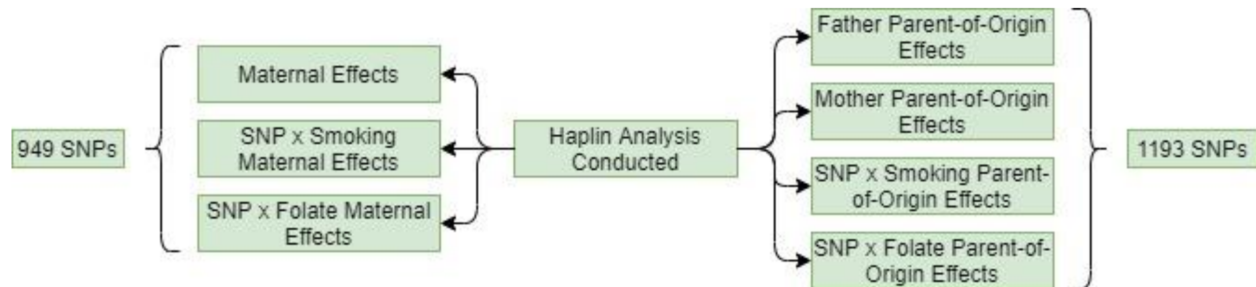

Figure 4.1. Analysis was conducted using the Haplin R package. Maternal gene-environment interaction effects were tested for 949 SNPs while parent-of-origin gene-environment interaction effects were tested for 1193 SNPs.

Power

The power calculated pertains to offspring effects. Depending on the minor allele frequency, a power of 0.8 is obtained for relative risks as low as 1.62 to 2.55 for parent-of-origin genetic effects and 1.61 to 2.53 for maternal genetic effects using a sample size of 411 case-parent triads, 15 mother-child dyads and 10 father-child dyads. This is the number of triads/dyads that were used in the analysis. The power may be underestimated in this study, since the alpha used to calculate it underwent a Bonferroni correction. Bonferroni is a conservative correction that aims to reduce the family wise error rate to 0.05 for independent testing. However, many SNPs have large linkage disequilibrium values meaning that they are very correlated, and as such, the Bonferroni correction may not be appropriate to apply.

Chapter 5: Manuscript

This manuscript is still in preparation to be submitted to a journal such as the American Journal of Medical Genetics Part A or the European Journal of Medical Genetics.

Maternal and parent-of-origin gene-environment effects on the etiology of orofacial clefting

Nikola Rasevic[1], Michele Rubini[2], Kelly M Burkett[3], Peter A Mossey[4], Borut Peterlin[5], Mohammad Faisal J Khan[2], Amin Ravaei[2], Luca Autelitano[6], Maria C Meazzini[6], Julian Little[1], Marie-Hélène Roy-Gagnon[1]

[1] School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada;

[2] Department of Biomedical and Specialty Surgery Sciences, University of Ferrara;

[3] Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada;

[4] World Health Organization–collaborating Centre for Oral and Craniofacial Research, Dental Hospital and School, University of Dundee, Dundee, Scotland;

[5] Clinical Institute of Medical Genetics, Department of Obstetrics & Gynecology, University Medical Center Ljubljana, Ljubljana, Slovenia;

[6] Smile House Milan, Regional Centre for Orofacial Clefts and Craniofacial Anomalies, Department of Cranio-Maxillo-Facial Surgery, San Paolo Hospital, University of Milan, Milan, Italy.

Abstract Word Count: 232
Word Count Main Text: 2860
Number of Tables: 5
Number of Figures: 6

Acknowledgements

Funding was provided by the Government of Canada through the CIHR under grant PJT-148911. The authors wish to thank the EUROCRAN and ITALCLEFT participants.

Conflict of interest statement

The authors declare no conflict of interest.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Canadian Institutes of Health Research Grant: 148911

Corresponding Author:

Marie-Hélène Roy-Gagnon, School of Epidemiology and Public Health, University of Ottawa, 600 Peter Morand Crescent, Office 101E, Ottawa, Ontario, Canada, K1G 5Z3, mroygagn@uottawa.ca, 613-562-5800 x8098

Abstract

Objective: Our goal was to investigate maternal and parent-of-origin gene-environment effects on the risk of orofacial clefts for candidate single nucleotide polymorphisms (SNPs).

Methods: Genome-wide SNP genotypes were obtained for case-parent triads from the EUROCRAN and ITALCLEFT studies. Candidate SNPs were selected from a previous genome-wide association study (Shi et al., 2012), along with surrounding SNPs, resulting in a total of 2142 SNPs. Non-genotyped SNPs were imputed. A total of 411 case-parent triads and 25 case-parent dyads were analyzed using log-linear models to test for maternal and parent-of-origin effects, and their interaction with maternal smoking and maternal use of supplements containing folic acid.

Results: An association was detected for a region in the *ATXN3* gene as it pertains to the interaction between maternal periconceptional smoking and maternal genetic effects (q = 0.025). Some regions displayed nominally significant associations in the interaction between maternal use of folic acid supplements and parent-of-origin genetic effects, specifically those located in the *CSMD1*, *KCNK10* and *SYT17* gene loci.

Conclusion: Investigation of interaction between exposure and maternal genetic and parent-of-origin genetic effects has detected involvement of regions in the *ATXN3*, *CSMD1*, *KCNK10* and *SYT17* genes. A mutation of the *ATXN3* gene has been involved in a neurological disorder, and the latter three genes have high mRNA expression levels in the brain, which suggest that further investigation of the link between orofacial clefts and brain development would be warranted.

Key Words: orofacial clefts, maternal genetic effects, parent-of-origin genetic effects, gene-environment interaction, case-parent trios

Introduction

Orofacial clefts (OC) are a subgroup of congenital anomalies that are categorized as cleft lip with or without cleft palate or cleft palate only. It is estimated that the prevalence of OC among live births in Europe is 1.55 per 1000 (Panamonta et al., 2015). OC can impact the mental and physical health of a child in a number of different ways, including but not limited to psychology, speech, hearing, appearance and cognition. As a result, multidisciplinary care through child and teenage hood for those affected from OC is needed. Children diagnosed with OC have been reported to have elevated morbidity and mortality rates (Christensen et al., 2004; Ngai et al., 2005).

It is understood that OC have a complex etiology that is associated with the interaction of genetic and environmental risk factors. Two possible genetic mechanisms that could have an influence on orofacial clefts are maternal genetic effects and parent-of-origin genetic effects. Maternal genetic effects are defined as the effect that a mother's genotype can have on a child's phenotype (Wolf and Wade, 2009). The mechanism by which this operates is that the mother's genotype influences the intra-uterine environment in which the offspring develops. This causes the offspring to be at greater risk of a congenital anomaly if a risk allele is present in the mother's genotype. Parent-of-origin genetic effects are present when the phenotype of an offspring's allele is dependent on whether the allele was transmitted by the mother or the father (Lawson et al., 2013). This operates through a biological mechanism called genomic imprinting. Dependent on which parent a child receives an allele from, the allele could be chemically marked which would affect expression.

Shi et al. (2012) performed a genome-wide association study to identify risk loci in OC case-parent triads. The study considered the effects of the mother's genotype as well as parent-of-origin genetic effects. No loci passed the genome wide level of significance. The etiology of

orofacial clefts is rather complex, and as such, when testing solely genetic effects, there may be reduced power if the genetic effects are modified by environmental factors. Use of supplements containing folic acid during early pregnancy have been found to be associated with a lowered risk for OC. For mothers who took supplements containing folic acid in pregnancy, a meta-analysis has shown a 40% reduction in risk of cleft lip with or without palate and 12% reduction in risk of cleft palate only (Zhou et al., 2020). There is a positive-dose response association between OC and maternal smoking during the first trimester (Lie et al., 2008; Shi et al., 2007). Passive maternal smoking during pregnancy is associated with a 1.5 fold increase in risk of a child being born with non-syndromic OC (Sabbagh et al., 2015).

In this study, the most significant associations from Shi et al. (2012) were further investigated while taking into consideration the interaction of two environmental factors: maternal smoking and maternal use of supplements containing folic acid. A total of 411 case-parent triads and 25 dyads were recruited from various populations in Europe as part of the EUROCRAN and ITALCLEFT studies (Mossey et al., 2017). Log-linear models implemented in the R package Haplin were applied to investigate the interaction effects of maternal and parent-of-origin genetic effects with maternal folic acid consumption and smoking.

Samples and Methods

Ethical Considerations and Reporting

The EUROCRAN and ITALCLEFT studies were approved by ethics boards at each collection centre. Ethical permission was sought and obtained from parents for themselves and their children at surgical centres in each participating countries at the time of first surgical intervention on the index infant. The use of data and DNA samples from EUROCRAN and

ITALCLEFT biobanks was approved by MREC Scotland (Dec 7th 2011, #MREC/1/0/7) and S. Paolo Hosp. E.C. (Mar 2nd, 2012, #3503) respectively. Ethics approval of the analysis reported in this article was obtained from the Ottawa Health Science Network Research Ethics Board. Informed consent was obtained from the parents. The methods and results are reported following the STREGA guideline (Little et al., 2009).

Study Population and Data Collection

The study population consisted of case-parent triads from Europe. Children with non-syndromic OC and their mothers and fathers were recruited between 2001 and 2005 through the European Collaboration on Craniofacial Anomalies study (EUROCRAN; the Netherlands, United Kingdom, Spain, Hungary, Bulgaria, Estonia and Slovakia) and the ITALCLEFT study (Italy) (Mossey et al., 2017). Diagnosis of non-syndromic OC was confirmed at surgical centres. Infants with recognized syndromes or Pierre Robin sequence were excluded. For the ITALCLEFT study, data on demographics, clinical diagnosis, pregnancy details and complications as well as exposure to environmental risk factors during the periconceptional period (three months before to three months after conception) were collected by a clinician at the time of primary surgery.

EUROCRAN/ITALCLEFT individuals were genotyped using an Illumina GSA array: GSAMD V 1.0 chip and recommended genotype calling procedures. In the EUROCRAN study, mothers answered a questionnaire that was administered by personal interview when the child was brought for primary surgery. The EUROCRAN questionnaire included questions on demographics, pregnancy history and lifestyle, including exposure to environmental risk factors during the periconceptional period. We focused on the environmental risk factors of maternal use of supplements containing folic acid and smoking. A mother was classified as using a

supplement containing folic acid if she specifically reported having taken at least 0.4 mg/day of folic acid or a folic acid-containing supplements for a minimum of one month during the periconceptional period. A mother was also classified as a smoker if she reported having smoked at least one cigarette/day during the periconceptional period. The outcome considered in this analysis includes all forms of non-syndromic OC combined. Both studies included collection of peripheral blood specimens or buccal cell samples from children and their parents. The study sample includes 1283 individuals (411 triads, 25 dyads). Of the 436 offspring, 108 had cleft lip only, 197 had cleft lip with cleft palate, 125 had cleft palate only and 6 sub-phenotypes were not available. Table 1 describes the study characteristics of this sample.

Genotype Quality Control and Imputation

Genotype data processing and quality control were done using Python (Van Rossum and Drake Jr., 1995) and PLINK (Purcell et al., 2007). Supplementary Figures 1 and 2 show the data processing steps for the initial 1536 individuals (523 families) and 699,911 SNPs. All SNP positions followed the GRCH37/hg19 genome assembly. Supplementary Figure 1 illustrates the data processing steps before phasing and imputation. Individual sex was imputed. SNPs were removed if they (i) were not in an autosome, (ii) had > 10% missing genotype across all triads, (iii) had a minor allele frequency < 0.05, (iv) deviated from Hardy-Weinberg equilibrium at a Bonferroni corrected p-value of p = 0.00001 or (v) had a Mendel error rate > 0.1. Individuals were excluded if they (i) had > 10% missing genotype across all SNPs (ii) were a part of a triad that contained a mother/father sex mismatch, or no sex was assigned to the mother/father or (iii) had a Mendelian transmission error rate > 0.05.

Supplementary Figure 2 represents the data processing steps for phasing, imputation and afterwards. Phasing was accomplished using SHAPEIT (Delaneau et al., 2013a; Delaneau et al.,

2011; Delaneau et al., 2013b; O'Connell et al., 2014). Then, the Michigan Imputation Server was used to assign non-genotyped SNPs into the phased haplotypes identified by SHAPEIT (Das et al., 2016). Imputation was conducted using the Haplotype Reference Consortium (McCarthy et al., 2016) reference panel. Target candidate SNPs were selected as the 25 SNPs for maternal genetic effects and 28 SNPs for parent-of-origin genetic effects for which $p < 10^{-5}$ in the study of Shi et al. (2012). SNPs that were within a region of 10,000 base pairs from the target SNPs were extracted for analysis from the imputed data. Individuals were removed if they were the only individuals in their triad. Once again, SNPs with a minor allele frequency < 0.05 were omitted. All 10 SNPs from a selected region in chromosome 17 had a minor allele frequency < 0.05 and were removed. As a result, the final sample size for analysis was 1283 individuals (411 triads + 25 dyads) with 2142 SNPs (949 SNPs for maternal genetic effects and 1193 SNPs for parent-of-origin genetic effects).

Statistical Analysis

Analysis was performed using log-linear models to test for parent-of-origin and maternal genetic effects, as implemented in the Haplin package (Gjessing and Lie, 2006) in the R environment (R Core Team, 2019). In this model, a statistically significant result would be interpreted as suggesting that the relative risk of a maternal or parent-of-origin genetic effect would significantly differ between smokers and non-smokers or between folic acid consumers and non-folic acid consumers. Models in Haplin assume Hardy-Weinberg equilibrium and random mating. For the maternal effect, relative risks were estimated using a multiplicative dose response. For parent-of-origin effects, relative risk ratios were estimated as the ratio of the relative risk of a maternal minor allele being transmitted to the relative risk of a paternal minor allele being transmitted. Haplin includes an expectation-maximization algorithm to handle

missing genotype data. Thus, Haplin conducts a test for each SNP. Nominal p-values from the region and SNP association tests were displayed using "Manhattan" plots generated using Python by plotting the $-\log_{10}$(p-values) of each SNP or region by their chromosome. False discovery rates were used to adjust for multiple testing and q-values, the expected proportion of false positives incurred when a test is statistically significant, are reported (Storey and Tibshirani, 2003). The preferred method for testing a region is by conducting a false discovery rate correction of the most significant SNP of the region. The adaptive sum of power (aSPU) method was another method we used to test the p-value of a specific region of SNPs (Kwak and Pan, 2016).

Results

Table 1 describes characteristics of the study sample. Just over half of the triads were from Italy (53.2%), just under 10% from elsewhere in western Europe (Spain, U.K), and the remainder from Central/Eastern Europe (Hungary, Bulgaria, Estonia and Slovakia). There were 108 infants with cleft lip only, 197 with cleft lip and cleft palate, and 125 with cleft palate only. The overall male proportion was 60.8%; there was a female preponderance for cleft palate only. This is consistent with other large epidemiological studies of OC (Raut et al., 2019). The proportion of mothers who reported that they had smoked in the periconceptual period was 22%. Almost half of the mothers reported periconceptual use of supplements containing folic acid.

Figures 3 to 6 are four different Manhattan plots that represent the significance of the tested genetic effects. The minimum SNP q-value will be referenced when discussing the significance of a region. Figures 3 and 4 display the significance of the interaction tests of folic acid supplementation with maternal and parent-of-origin genetic effects, respectively. Similarly,

figures 5 and 6 display significance of the interaction tests of smoking with maternal and parent-of-origin genetic effects, correspondingly. A significant region in the test between the interaction of maternal genetic effects and smoking was detected. This region is located in chromosome 14 and is within the *ATXN3* gene (nominal p = 0.00071, q = 0.025). A nominally significant association was also detected for a region in the *SYT17* gene in chromosome 16 as it pertains to the interaction between parent-of-origin genetic effects and folic acid consumption (nominal p = 0.0018).

Tables 2-5 show relative risk/relative risk ratio estimates for the nominally significant SNPs. Genotyped SNPs that were close to being the most significant SNP were selected. Otherwise, the most significant non-genotyped SNP was chosen. Relative risk estimates for the maternal effects stratified by folic acid supplementation status and smoking status are shown in tables 2 and 3, respectively. For the interaction between maternal genetic effects and smoking, the region containing the *ATXN3* gene was significant after correcting for multiple testing (q = 0.025). Interestingly, the estimates of relative risk indicate a protective effect of minor alleles for SNPS in *ATXN3* only when maternal smoking is present. There were no SNPs trending towards significance in the maternal gene interaction with folic acid supplementation.

Relative risk estimates for the parent-of-origin effects stratified by folic acid supplementation status and smoking status are shown in tables 4 and 5, respectively. In terms of the region that these SNPs reside in, *KCNK10* displayed one of the highest significances (p = 0.011, q = 0.64) for the interaction between parent-of-origin genetic effects and smoking. After adjusting for multiple testing, these genes did not include any significant regions. Regions in the *CSMD1* (p = 0.0081, q = 0.57), *SYT17* (p = 0.0018, q = 0.57) and *KCNK10* (p = 0.0056, q = 0.57) genes showed some of the lowest p-values for the parent-of-origin gene interaction with

folic acid supplementation. The estimates of relative risk ratio indicate a lower risk of the minor alleles at the SNPs in *CSMD1* and *SYT17* when maternal periconceptional folic acid supplementation is not present and an increase of risk when folic acid supplementation is present.

Discussion

We found significant maternal-gene smoking interactions in the region of the *ATXN3* gene (minimum p = 0.00071 and q = 0.025). Machado-Joseph disease, a neurodegenerative disorder, is characterized by the expansion of the $(CAG)_n$ motif in this gene (Bettencourt et al., 2010). However, it is important to keep in mind that mutations in this gene giving rise to Machado-Joseph disease are rarer than the SNP variant identified in the present study. This gene encodes a protein that deubiquitinates other proteins and thus, is involved in the regulation of the degradation of proteins (Zeng et al., 2018). The relative risk for the smoking stratum is lower than the non-smoking stratum in the presence of the minor allele in this gene.

There were some other interesting results. Some of the most significant regions observed in the test between parent-of-origin genetic effects and maternal periconceptional folic acid use are located in *CSMD1* (p = 0.0081, q = 0.57), *KCNK10* (p = 0.0056, q = 0.57) and *SYT17* (p = 0.0018, q = 0.57). There have been additional studies that do illustrate that an association does exist for *CSMD1* and OC for offspring genes (Camargo et al., 2012). Outside of this, there have been few other studies that identified associations between OC and the other genes highlighted in the present paper. The reason that *CSMD1*, *KCNK10* and *SYT17* are genes of interest is because they are all highly expressed in the brain which could open up future avenues of research. Specifically, a gene-gene interaction can be explored. In terms of biological function, *CSMD1* is involved in regulating inflammation in the nervous system (Kraus et al., 2006) and has been

found to be associated with schizophrenia (Athanasiu et al., 2017). *KCNK10* encodes a protein located in a potassium ion channel and also plays a role in adipogenesis (Nishizuka et al., 2014). The estimates of relative risk ratios suggest a heightened risk of the minor alleles in *CSMD1* and *SYT17* when maternal folic acid supplementation is present. For the interaction between smoking and parent-of-origin genetic effects, *KCNK10* contained the most nominally significant region with a p-value of 0.011. Of all the parent-of-origin genetic effects discussed, the regions were not significant for child effects after correcting for multiple testing, suggesting that the parent-of-origin effect may not be confounded by the child genetic effect.

In essence, some of the most significant results in the present study have some form of relation to the brain. A mutation in *ATXN3* gives rise to a neurological disorder, *CSMD1* is expressed primarily in the brain and is associated with schizophrenia. *KCNK10* and *SYT17* are also expressed in the brain. This is interesting since a study has shown that children with isolated cleft lip with or without palate had abnormal brain structures (Nopoulos et al., 2007). Perhaps variations of genes with biological functions in the brain can increase the risk of OC within certain environmental contexts.

There are some limitations in this study. When testing for associations between gene-exposure interactions in case-parent triads, independence of the child genotype and the exposure conditional on parental mating type needs to be assumed. Further assumptions that need to be assumed include transmission ratio symmetry for the estimation of parent-of-origin genetic effects, mating symmetry for the estimation of maternal genetic effects, and the Haplin models assume Hardy-Weinberg equilibrium. Additionally, when calculating the interaction effects between parent-of-origin and environment, a loss of power is expected as statistical testing is essentially split into four categories: maternal transmission with environment, maternal

transmission without environment, paternal transmission with environment and paternal transmission without environment.

This study aimed to examine the effects of maternal periconceptional smoking and folic acid on maternal and parent-of-origin genetic effects using a log-linear model. Maternal gene-smoking interaction effects were found to be significant for the region located in the *ATXN3* gene. Parent-of-origin gene-folic acid interaction effects were found to be significant for SNPs located in the *CSMD1*, *KCNK10* and *SYT17* genes. These genes are connected to the brain through high mRNA expression and/or associations with neurological disorders such as Machado-Joseph disease and schizophrenia. These results provide indication that the brain could be involved in the etiology of orofacial clefts.

References

Athanasiu L, Giddaluru S, Fernandes C, Christoforou A, Reinvang I, Lundervold AJ, Nilsson LG, Kauppi K, Adolfsson R, Eriksson E, Sundet K, Djurovic S, Espeseth T, Nyberg L, Steen VM, Andreassen OA, Le Hellard S. 2017. A genetic association study of CSMD1 and CSMD2 with cognitive function. Brain Behav Immun 61:209-216.

Bettencourt C, Santos C, Montiel R, Kay T, Vasconcelos J, Maciel P, Lima M. 2010. The (CAG)n tract of Machado-Joseph Disease gene (ATXN3): a comparison between DNA and mRNA in patients and controls. Eur J Hum Genet 18(5):621-623.

Camargo M, Rivera D, Moreno L, Lidral AC, Harper U, Jones M, Solomon BD, Roessler E, Vélez JI, Martinez AF, Chandrasekharappa SC, Arcos-Burgos M. 2012. GWAS reveals new recessive loci associated with non-syndromic facial clefting. Eur J Med Genet 55(10):510-514.

Christensen K, Juel K, Herskind AM, Murray JC. 2004. Long term follow up study of survival associated with cleft lip and palate at birth. Bmj 328(7453):1405.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. 2016. Next-generation genotype imputation service and methods. Nat Genet 48(10):1284-1287.

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013a. Haplotype estimation using sequencing reads. Am J Hum Genet 93(4):687-696.

Delaneau O, Marchini J, Zagury JF. 2011. A linear complexity phasing method for thousands of genomes. Nat Methods 9(2):179-181.

Delaneau O, Zagury JF, Marchini J. 2013b. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. United States. p 5-6.

Gjessing HK, Lie RT. 2006. Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. Ann Hum Genet 70(Pt 3):382-396.

Kraus DM, Elliott GS, Chute H, Horan T, Pfenninger KH, Sanford SD, Foster S, Scully S, Welcher AA, Holers VM. 2006. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. J Immunol 176(7):4419-4430.

Kwak IY, Pan W. 2016. Adaptive gene- and pathway-trait association testing with GWAS summary statistics. Bioinformatics 32(8):1178-1184.

Lawson HA, Cheverud JM, Wolf JB. 2013. Genomic imprinting and parent-of-origin effects on complex traits. Nat Rev Genet 14(9):609-617.

Lie RT, Wilcox AJ, Taylor J, Gjessing HK, Saugstad OD, Aabyholm F, Vindenes H. 2008. Maternal smoking and oral clefts: the role of detoxification pathway genes. Epidemiology 19(4):606-615.

Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N. 2009. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6(2):e22.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG,

Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R. 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48(10):1279-1283.

Mossey PA, Little J, Steegers-Theunissen R, Molloy A, Peterlin B, Shaw WC, Johnson C, FitzPatrick DR, Franceschelli P, Rubini M. 2017. Genetic Interactions in Nonsyndromic Orofacial Clefts in Europe-EUROCRAN Study. Cleft Palate Craniofac J 54(6):623-630.

Ngai CW, Martin WL, Tonks A, Wyldes MP, Kilby MD. 2005. Are isolated facial cleft lip and palate associated with increased perinatal mortality? A cohort study from the West Midlands Region, 1995-1997. J Matern Fetal Neonatal Med 17(3):203-206.

Nishizuka M, Hayashi T, Asano M, Osada S, Imagawa M. 2014. KCNK10, a tandem pore domain potassium channel, is a regulator of mitotic clonal expansion during the early stage of adipocyte differentiation. Int J Mol Sci 15(12):22743-22756.

Nopoulos P, Langbehn DR, Canady J, Magnotta V, Richman L. 2007. Abnormal brain structure in children with isolated clefts of the lip or palate. Arch Pediatr Adolesc Med 161(8):753-758.

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, McQuillan R, Fraser RM, Campbell H, Polasek O, Asiki G, Ekoru K, Hayward C, Wright AF, Vitart V, Navarro P, Zagury JF, Wilson JF, Toniolo D, Gasparini P, Soranzo N, Sandhu MS, Marchini J. 2014. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet 10(4):e1004234.

Panamonta V, Pradubwong S, Panamonta M, Chowchuen B. 2015. Global Birth Prevalence of Orofacial Clefts: A Systematic Review. J Med Assoc Thai 98 Suppl 7:S11-21.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559-575.

R Core Team. 2019. R: A language and environment for statistical computing.

Raut JR, Simeone RM, Tinker SC, Canfield MA, Day RS, Agopian AJ. 2019. Proportion of Orofacial Clefts Attributable to Recognized Risk Factors. Cleft Palate Craniofac J 56(2):151-158.

Sabbagh HJ, Hassan MH, Innes NP, Elkodary HM, Little J, Mossey PA. 2015. Passive smoking in the etiology of non-syndromic orofacial clefts: a systematic review and meta-analysis. PLoS One 10(3):e0116963.

Shi M, Christensen K, Weinberg CR, Romitti P, Bathum L, Lozada A, Morris RW, Lovett M, Murray JC. 2007. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. Am J Hum Genet 80(1):76-90.

Shi M, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Wu T, Murray T, Redett RJ, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Yeow V, Chong SS, Shi B, Christensen K, Scott AF, Patel P, Cheah F, Beaty TH. 2012. Genome wide study of maternal and parent-of-origin effects on the etiology of orofacial clefts. Am J Med Genet A 158a(4):784-794.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100(16):9440-9445.

Van Rossum G, Drake Jr. F. 1995. Python Tutorial. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Wolf JB, Wade MJ. 2009. What are maternal effects (and what are they not)? Philos Trans R Soc Lond B Biol Sci 364(1520):1107-1115.

Zeng L, Zhang D, McLoughlin HS, Zalon AJ, Aravind L, Paulson HL. 2018. Loss of the Spinocerebellar Ataxia type 3 disease protein ATXN3 alters transcription of multiple signal transduction pathways. PLoS One 13(9):e0204438.

Zhou Y, Sinnathamby V, Yu Y, Sikora L, Johnson CY, Mossey P, Little J. 2020. Folate intake, markers of folate status and oral clefts: An updated set of systematic reviews and meta-analyses. Birth Defects Res 112(19):1699-1719.

Table 1. Characteristics of the study sample (n = 436 triads/dyads).

	Number of triads (%)			
	All	**Cleft lip without palate (24.8)**	**Cleft lip with palate (45.2)**	**Cleft palate only (28.7)**
Maternal periconceptual exposure:				
Smoking	97 (22.2)	19 (4.4)	52 (11.9)	26 (6.0)
Supplements Containing Folic Acid	210 (48.2)	59 (13.5)	84 (19.3)	67 (15.4)
Child sex:				
Male	265 (60.8)	69 (15.8)	138 (31.7)	54 (12.4)
Female	171 (39.2)	39 (8.9)	59 (13.5)	71 (16.3)
Country of origin:				
Italy	232 (53.2)	63 (14.4)	115 (26.4)	51 (11.7)
UK	33 (7.6)	8 (1.8)	9 (2.1)	13 (3.0)
Spain	8 (1.8)	1 (0.2)	7 (1.6)	0 (0)
Hungary	81 (18.6)	14 (3.2)	39 (8.9)	28 (6.4)
Bulgaria	29 (6.7)	10 (2.3)	14 (3.2)	5 (1.1)
Estonia	24 (5.5)	4 (0.9)	6 (1.4)	14 (3.2)
Slovakia	29 (6.7)	8 (1.8)	7 (1.6)	14 (3.2)

Table 2. Maternal gene-folic acid supplementation interaction tests for the most significant SNP in nominally significant regions.

SNP	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk[a] (No Folic Acid / Folic Acid)
rs2026959	10	A/G (0.19)	*RPL21P16*[b]	0.0037	0.77	0.069	0.73	1.52/0.74
rs212016	3	T/C (0.28)	*FHIT*	0.0070	0.77	0.11	0.73	1.37/0.77
rs12974161	19	G/C (0.13)	*KLHL26*	0.016	0.77	0.17	0.73	1.88/0.92
rs1921819[c]	2	T/C (0.46)	*CNTNAP5*[b]	0.018	0.77	0.25	0.73	1.35/0.85
rs2068361[c]	6	C/T (0.31)	*BMP6*	0.019	0.77	0.12	0.73	1.40/0.86
rs10489554	1	C/T (0.24)	*LRRC7*	0.028	0.77	0.21	0.73	1.39/0.84
rs4505466	2	C/T (0.19)	*SH3BP4*	0.045	0.77	0.69	0.99	0.71/1.17
rs12413650[c]	10	G/A (0.16)	*MKI67*[b]	0.047	0.77	0.27	0.73	1.32/0.78

[a] Estimate of the relative risk associated with the presence of one copy of the minor allele. The estimate of the relative risk associated with two copies is the square of that for one copy.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for the interaction between maternal genetic effects and folic acid supplementation.
[e] False discovery rate corrected q-value of the interaction between maternal genetic effects and folic acid supplementation.
[f] aSPU test p-value for the interaction between maternal genetic effects and folic acid supplementation.
[g] False discovery rate corrected q-value of the aSPU test.

Table 3. Maternal gene-smoking interaction for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk[a] (No Smoking / Smoking)
rs8013669	14	C/G (0.39)	*ATXN3*	0.00071	0.025	0.0040	0.096	1.22/0.54
rs1921827	2	T/C (0.11)	*CNTNAP5*[b]	0.013	0.25	0.19	0.62	1.09/0.42
rs9870680	3	C/T (0.20)	*GRM7*	0.016	0.29	0.28	0.62	1.06/0.52
rs4938835	11	A/G (0.14)	*P2RX3*[b]	0.019	0.29	0.047	0.56	0.82/1.86
rs11868591	17	G/C (0.33)	*MYH13*[b]	0.027	0.35	0.50	0.75	0.97/0.56
rs4238650	16	C/T(0.10)	*XYLT1*[b]	0.029	0.36	0.34	0.64	1.53/0.60
rs7326323[c]	13	G/T (0.30)	*SPATA13*	0.041	0.42	0.28	0.62	1.05/0.62
rs141806038	10	G/T (0.14)	*MKI67*[b]	0.042	0.42	0.38	0.65	0.81/1.63
rs12107803	3	C/T (0.09)	*FHIT*	0.043	0.42	0.94	0.91	0.76/1.75
rs17131017[c]	1	T/C (0.17)	*LRRC7*	0.045	0.44	0.71	0.90	0.80/1.51

[a] Estimate of the relative risk associated with the presence of one copy of the minor allele. The estimate of the relative risk associated with two copies is the square of that for one copy.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for the interaction between maternal genetic effects and smoking.
[e] False discovery rate corrected q-value of the interaction between maternal genetic effects and smoking.
[f] aSPU test p-value for the interaction between maternal genetic effects and smoking.
[g] False discovery rate corrected q-value of the aSPU test.

Table 4. Parent-of-origin gene-folic acid supplementation interaction tests for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk Ratio[a] (No Folic Acid / Folic Acid)
rs62025397	16	T/C (0.43)	*SYT17*	0.0018	0.57	0.011	0.23	0.82/3.79
rs7154539	14	C/T (0.50)	*KCNK10*	0.0056	0.57	0.070	0.26	2.08/0.50
rs116529928	5	C/T (0.07)	*IRX2*[b]	0.0057	0.57	0.15	0.33	0.50/5.34
rs62481577[c]	8	T/C (0.27)	*CSMD1*	0.0081	0.57	0.18	0.33	0.30/1.28
rs2543347	14	G/A (0.36)	*GPATCH2L*[b]	0.012	0.57	0.17	0.33	2.93/0.72
rs2789405	1	T/C (0.17)	*PLA2G4A*[b]	0.020	0.57	0.061	0.26	0.45/1.94
rs72687592	9	A/T (0.36)	*SMARCA2*	0.021	0.57	0.23	0.39	1.52/0.45
rs2906893	12	C/T (0.25)	*CACNA2D4*	0.027	0.57	0.26	0.39	3.01/0.88
rs72674763	4	T/C (0.08)	*TUBAP10*[b]	0.031	0.57	0.069	0.26	2.24/0.42
rs139714205	1	A/G (0.31)	*SPATA6*[b]	0.037	0.57	0.41	0.49	2.31/0.62
rs1443440	9	C/T (0.25)	*NTRK2*	0.041	0.57	0.37	0.48	1.29/0.38

[a] Estimate of relative risk ratio associated with the maternal allele being transmitted.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for the interaction between parent-of-origin genetic effects and folic acid supplementation.
[e] False discovery rate corrected q-value of the interaction between parent-of-origin genetic effects and folic acid supplementation.
[f] aSPU test p-value for the interaction between parent-of-origin genetic effects and folic acid supplementation.
[g] False discovery rate corrected q-value of the aSPU test.

Table 5. Parent-of-origin gene-smoking interaction tests for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk Ratio[a] (No Smoking / Smoking)
rs1437897	2	A/G (0.24)	*NCKAP5*	0.00096	0.64	0.18	0.67	0.52/4.59
rs11624297	14	G/A (0.43)	*GPATCH2L*[b]	0.0018	0.64	0.091	0.62	0.84/5.88
rs55653268[c]	8	G/T (0.49)	*MTUS1*	0.0050	0.64	0.17	0.67	1.32/0.22
rs11625905	14	T/G (0.22)	*KCNK10*	0.011	0.64	0.056	0.62	0.47/2.36
rs79187764	12	A/G (0.05)	*CACNA2D4*	0.016	0.64	0.27	0.79	0.35/9.91
rs36120312	12	T/C (0.13)	*CCDC38*	0.018	0.64	0.97	0.99	1.98/0.29
rs461131	5	T/A (0.39)	*IRX2*[b]	0.020	0.64	0.093	0.62	1.06/0.26
rs12144639	1	A/G (0.47)	*PROX1-AS1*[b]	0.021	0.64	0.32	0.80	0.70/3.10
rs12623060	2	T/C (0.19)	*LTBP1*	0.021	0.64	0.37	0.83	2.03/0.42
rs1155245[c]	2	T/C (0.35)	*SLC8A1*	0.031	0.64	0.20	0.67	0.81/3.24
rs55984928	17	G/A (0.40)	*MPRIP*[b]	0.039	0.64	0.78	0.69	1.15/0.32
rs2269787	16	G/A (0.16)	*SYT17*	0.047	0.66	0.73	0.99	0.96/4.63

[a] Estimate of relative risk ratio associated with the maternal allele being transmitted.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for the interaction between parent-of-origin genetic effects and smoking.
[e] False discovery rate corrected q-value of the interaction between parent-of-origin genetic effects and smoking.
[f] aSPU test p-value for the interaction between parent-of-origin genetic effects and smoking.
[g] False discovery rate corrected q-value of the aSPU test.

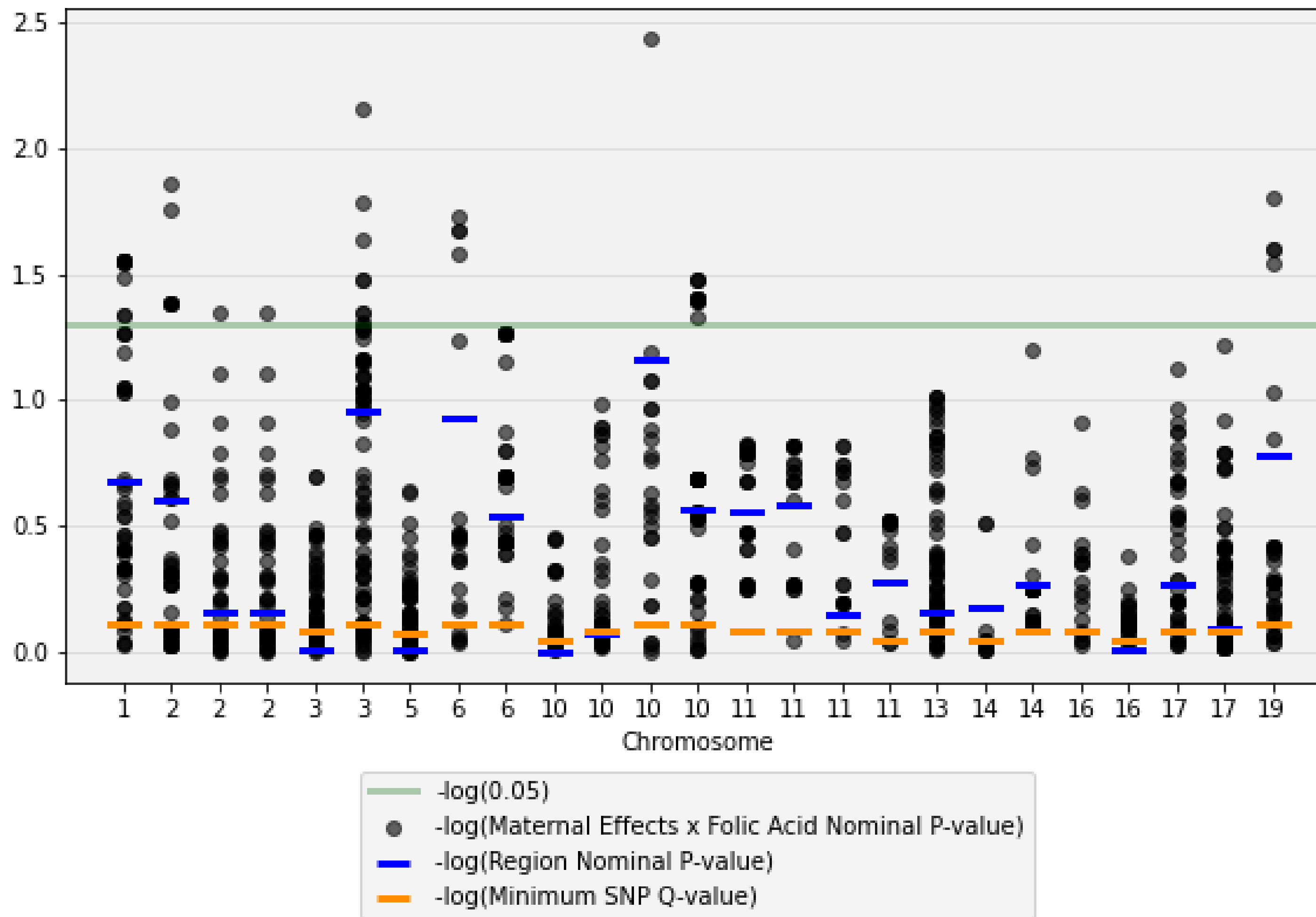

Figure 3. Manhattan plot of the Wald test for the interaction between maternal genetic effects and folic acid supplementation.

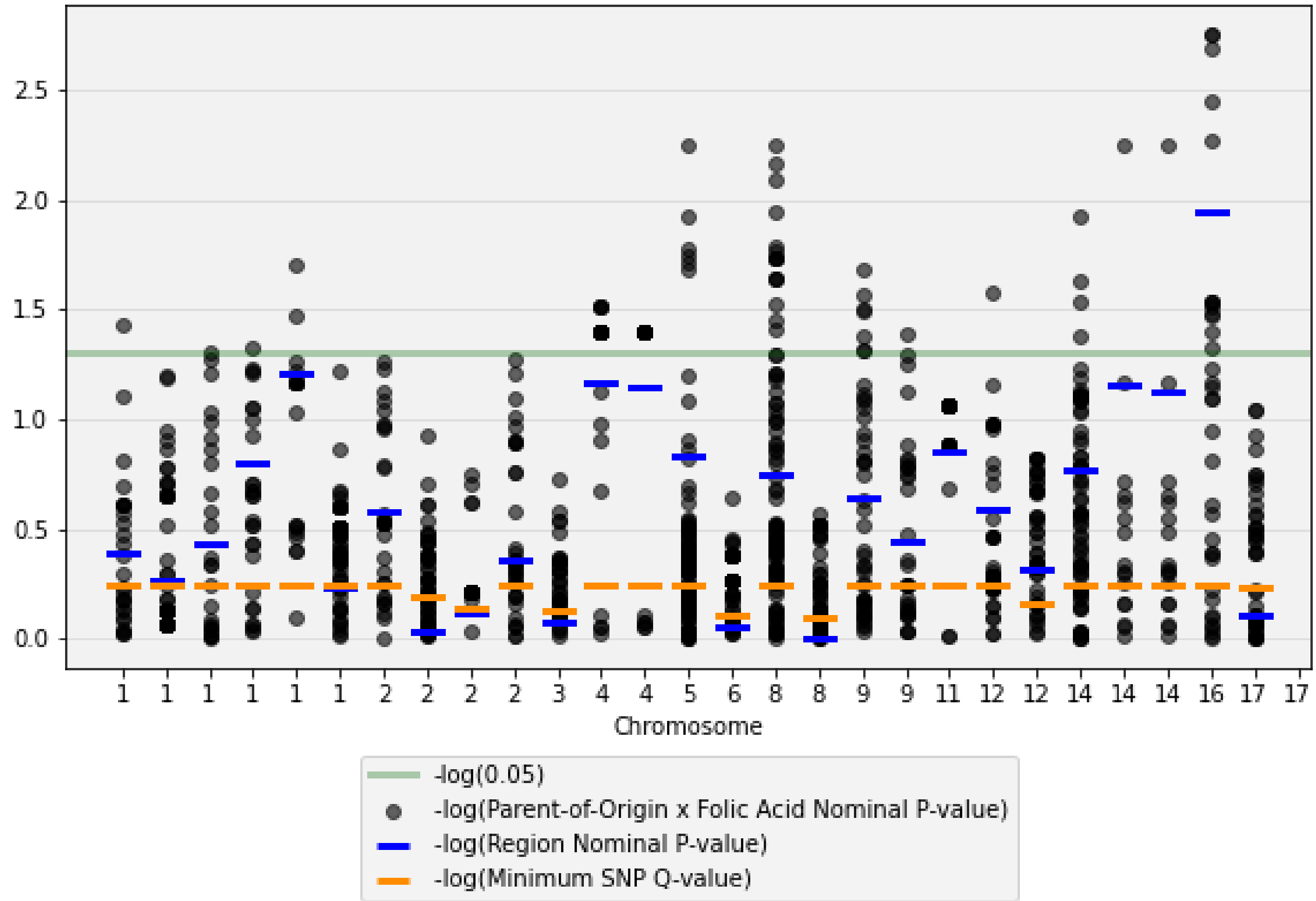

Figure 4. Manhattan plot of the Wald test for the interaction between parent-of-origin genetic effects and folic acid supplementation.

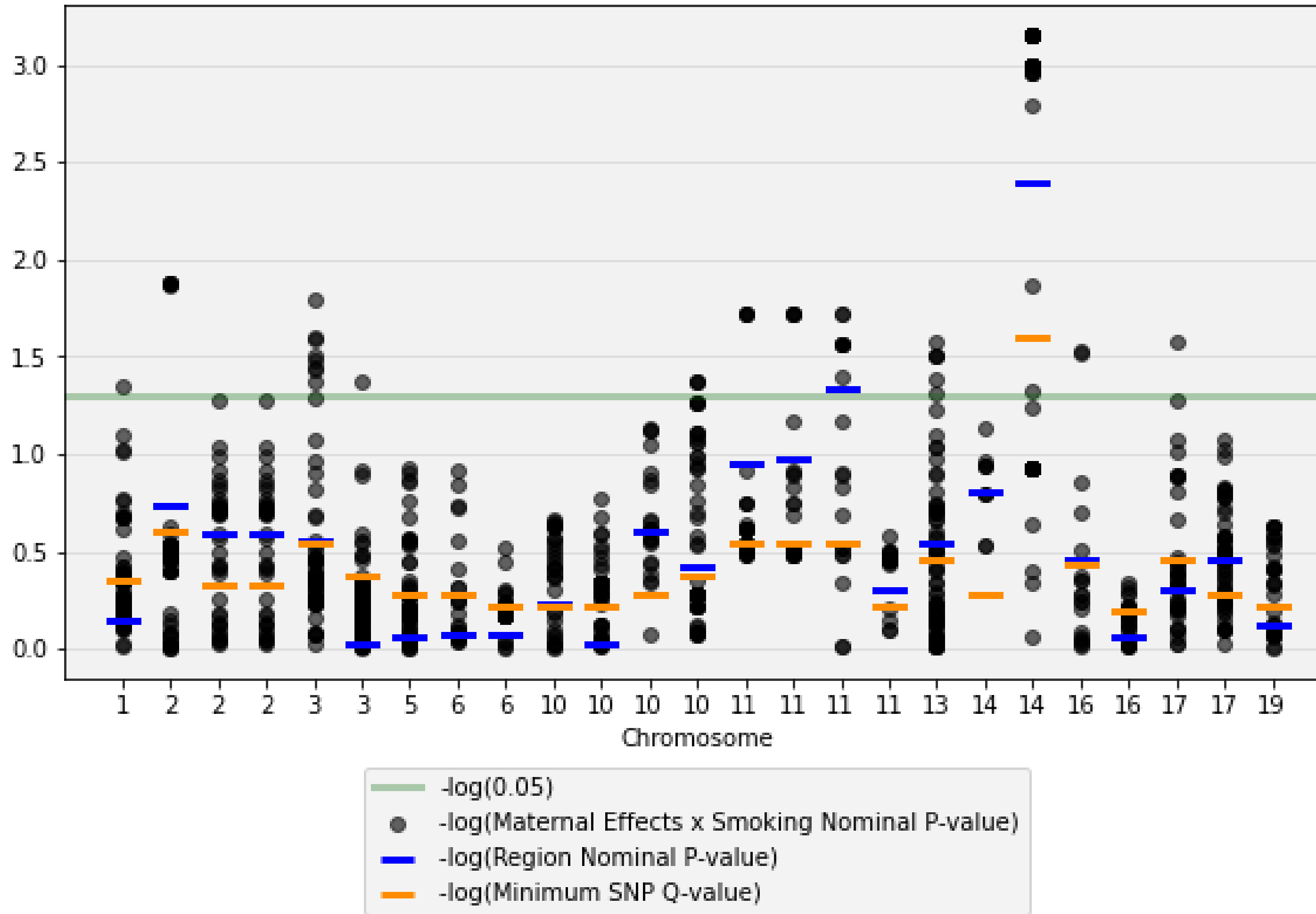

Figure 5. Manhattan plot of the Wald test for the interaction between maternal genetic effects and smoking.

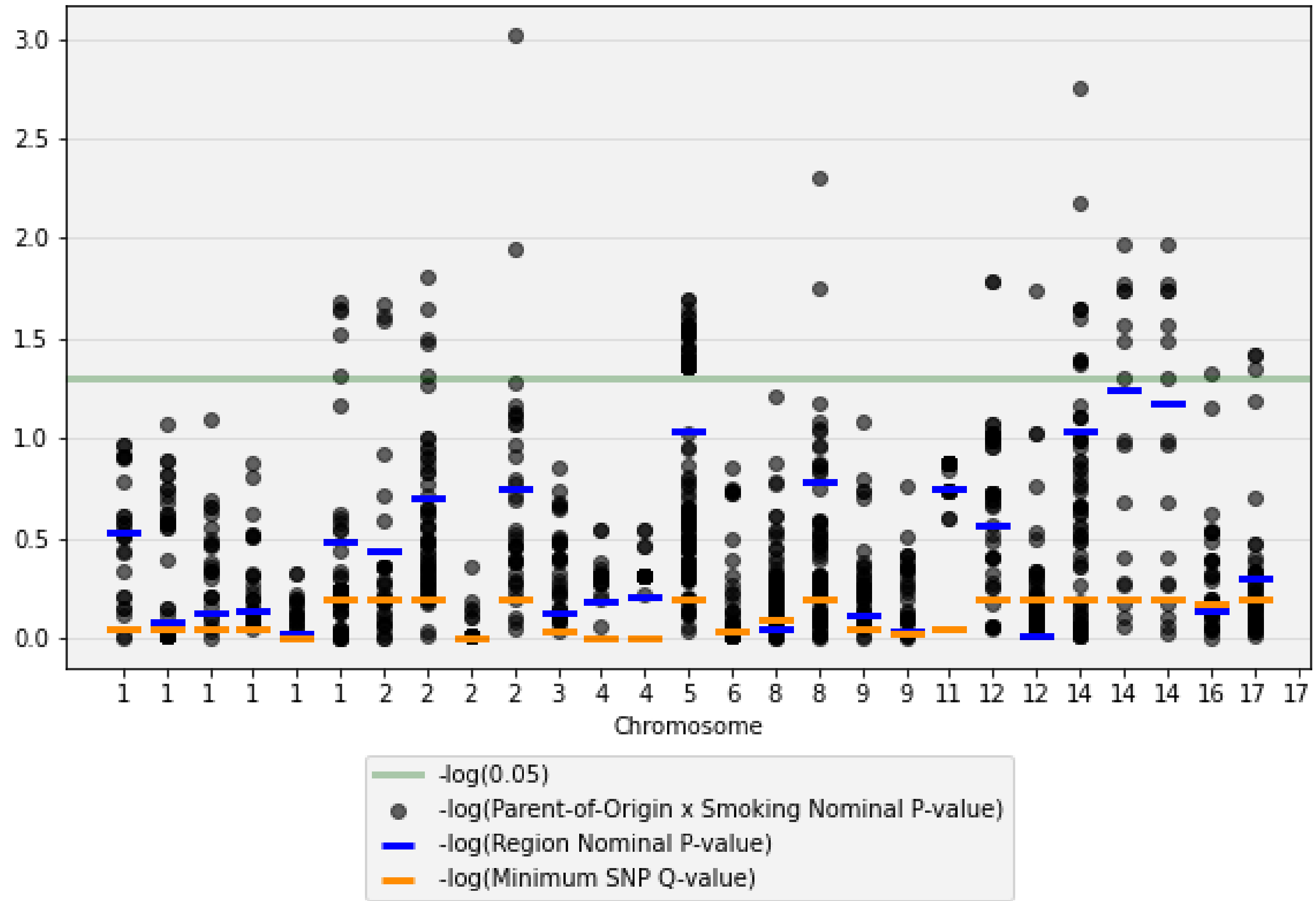

Figure 6. Manhattan plot of the Wald test for the interaction between parent-of-origin genetic effects and smoking.

Supplementary Material

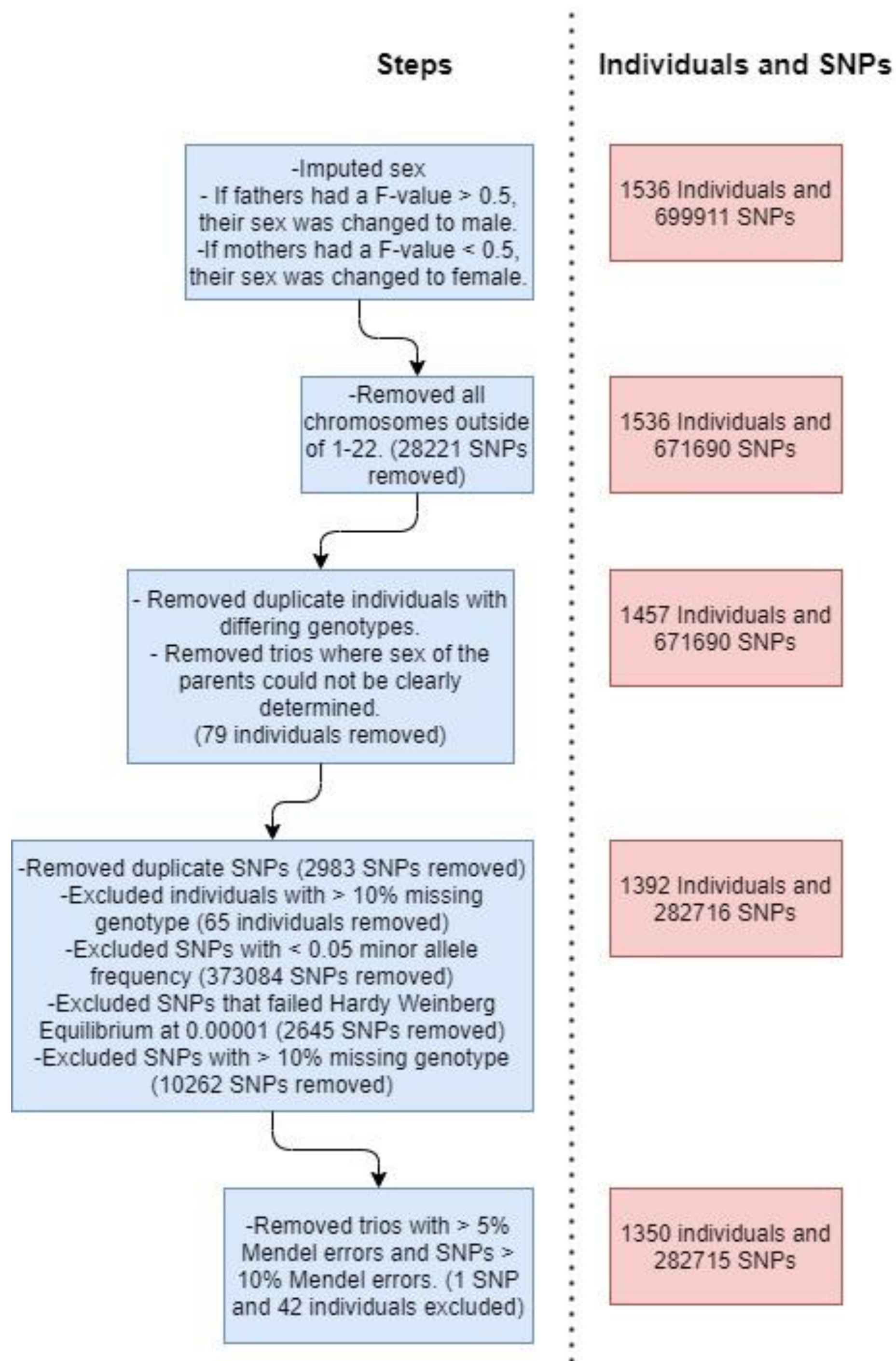

Figure 1. Summary of the pre-phasing steps. The analysis began with a total of 1536 individuals and 699911 SNPs. Sex was assigned based on genotypes using the F-statistic which is defined as the X-chromosome homozygosity estimate (Purcell et al., 2007). SNPs were removed if they (i) were not in an autosomal chromosome, (ii) had > 10% missing genotype across all triads, (iii) had a minor allele frequency < 0.05, (iv) deviated from Hardy-Weinberg equilibrium at p = 0.00001 or (v) had a Mendel error rate > 0.1. Individuals were excluded if they (i) had > 10% missing genotype across all SNPs, (ii) were a part of a triad that contained a mother/father sex mismatch, or no sex was assigned to the mother/father or (iii) had a Mendelian transmission error rate > 0.05. After these quality control steps, 1350 individuals and 282715 SNPs remained for the analysis.

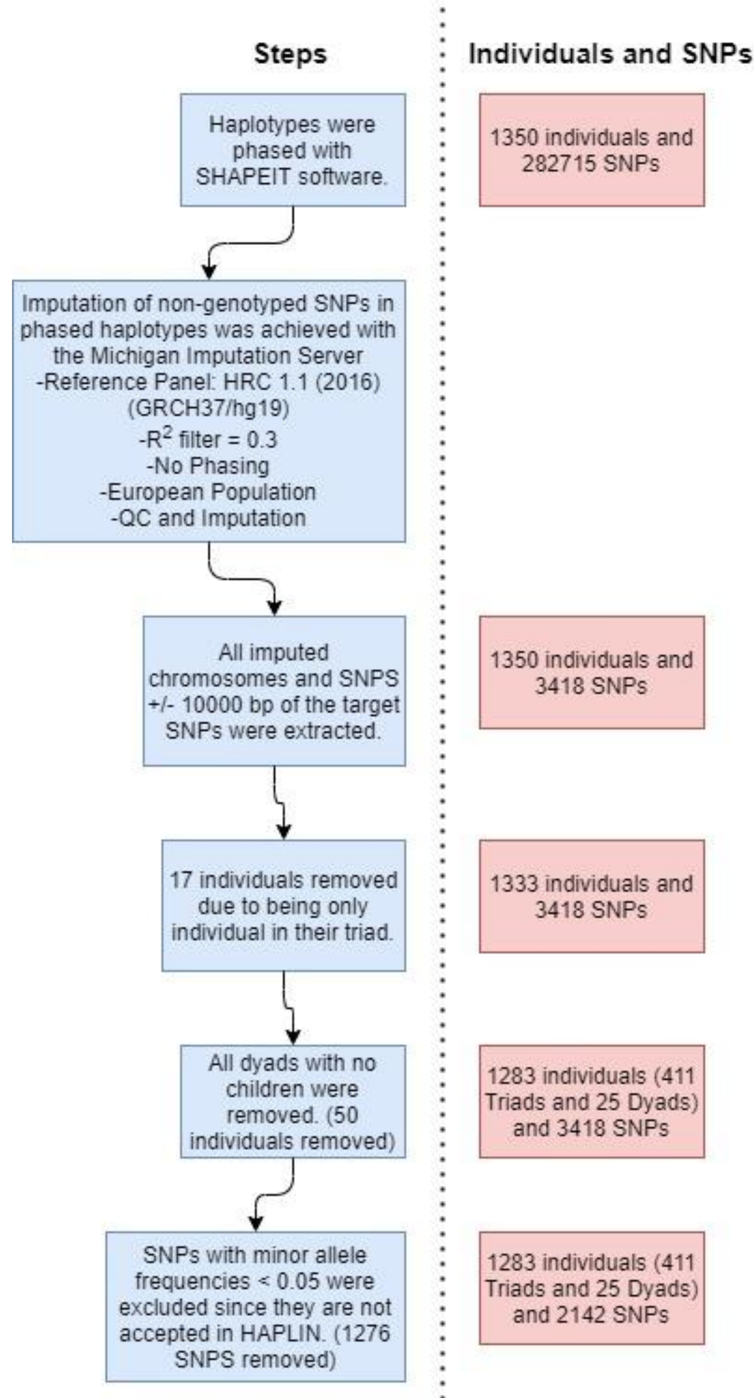

Figure 2. Summary of the steps for phasing and imputation. Before phasing commenced, a total of 1350 individuals and 282715 SNPs were accounted for. Phasing was accomplished using the SHAPEIT (Delaneau et al., 2013a; Delaneau et al., 2011; Delaneau et al., 2013b; O'Connell et al., 2014) software, and the outputted phased haplotypes were then imputed using the Michigan imputation server (Das et al., 2016). Imputation was achieved using the HRC 1.1 (McCarthy et al., 2016) reference panel. Target candidate SNPs were selected as the twenty five most significant SNPs for maternal genetic effects and twenty eight most significant SNPs for parent-of-origin genetic effects ($p < 10^{-5}$) from the genome-wide association study of Shi et al. (2012). SNPs that were within 10,000 base pairs of the target SNPs were extracted for analysis from the genotyped and imputed data. Individuals were removed if they were the only individual in their triad. SNPs with a minor allele frequency < 0.05 were omitted. All 10 SNPs from a region in chromosome 17 had a minor allele frequency < 0.05 and were removed. 1283 individuals (411 triads + 25 dyads) and 2142 SNPs remained for the duration of the analysis.

Chapter 6: Additional Results

Replication Results and Power Calculations

The additional results consist of the replication of the Shi et al. (2012) study. In this replication, figures 6.1, 6.2 and 6.3 represent all tested SNPs for maternal genetic effects, mother parent-of-origin genetic effects and father parent-of-origin genetic effects, respectively, indicating the significance of the region. Tables 6.1, 6.2 and 6.3 represent the most nominally significant SNPs for maternal genetic effects, mother parent-of-origin genetic effects and father parent-of-origin genetic effects, respectively. When multiple testing was corrected for, there were no significant regions remaining. Regions in/near genes *RASGRF2, JCAD* and *CNTNAP5* displayed nominal significance for maternal genetic effects. For father parent-of-origin genetic effects, regions near the genes: *SLC22A16* and *TMEM182* displayed nominal significance. Two regions in the mother parent-of-origin genetic effects displayed nominal significance. These regions are in/near *LTBP1* and *PBX1*.

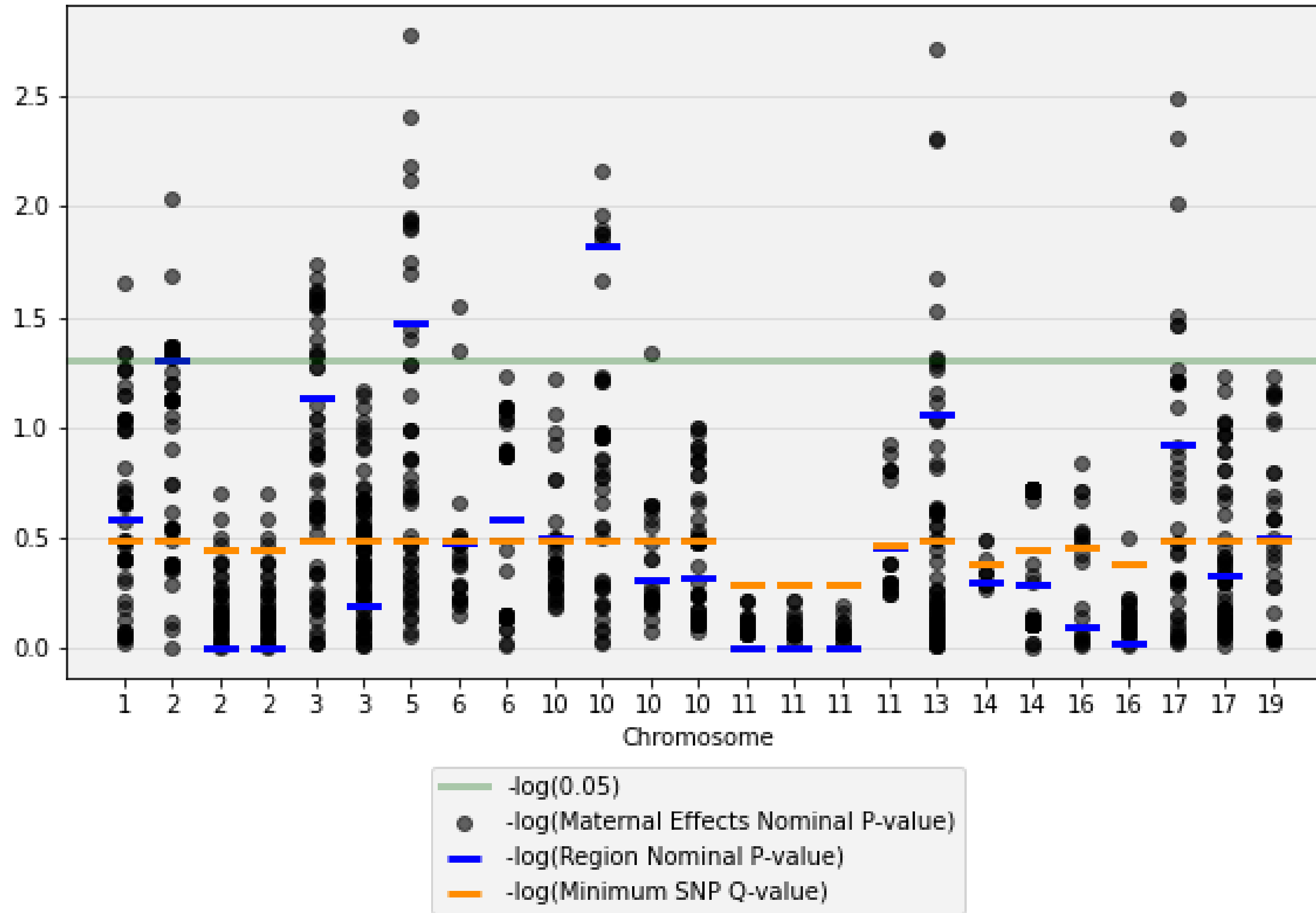

Figure 6.1. Manhattan plot of the maternal genetic effects.

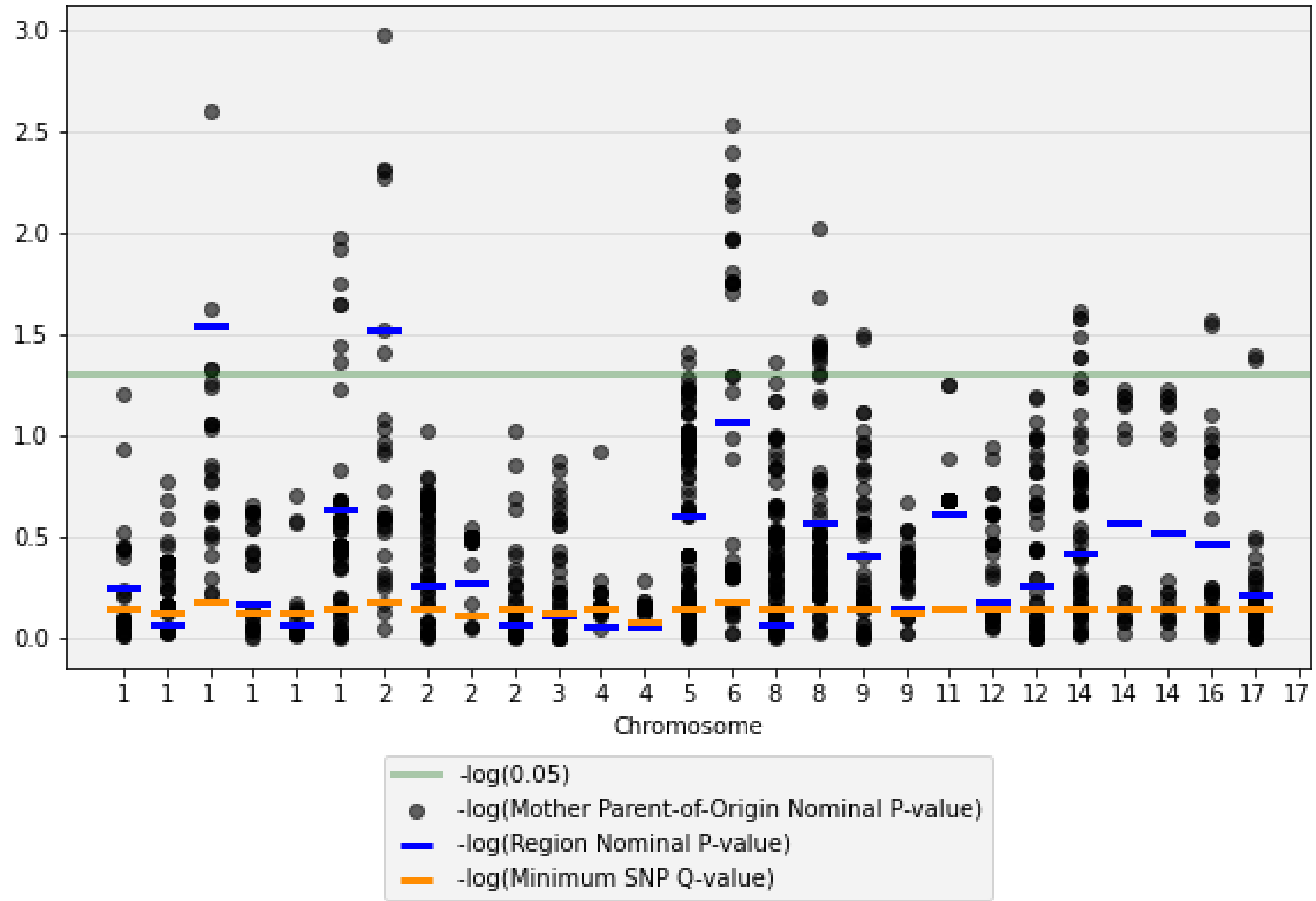

Figure 6.2. Manhattan plot of the mother parent-of-origin genetic effects.

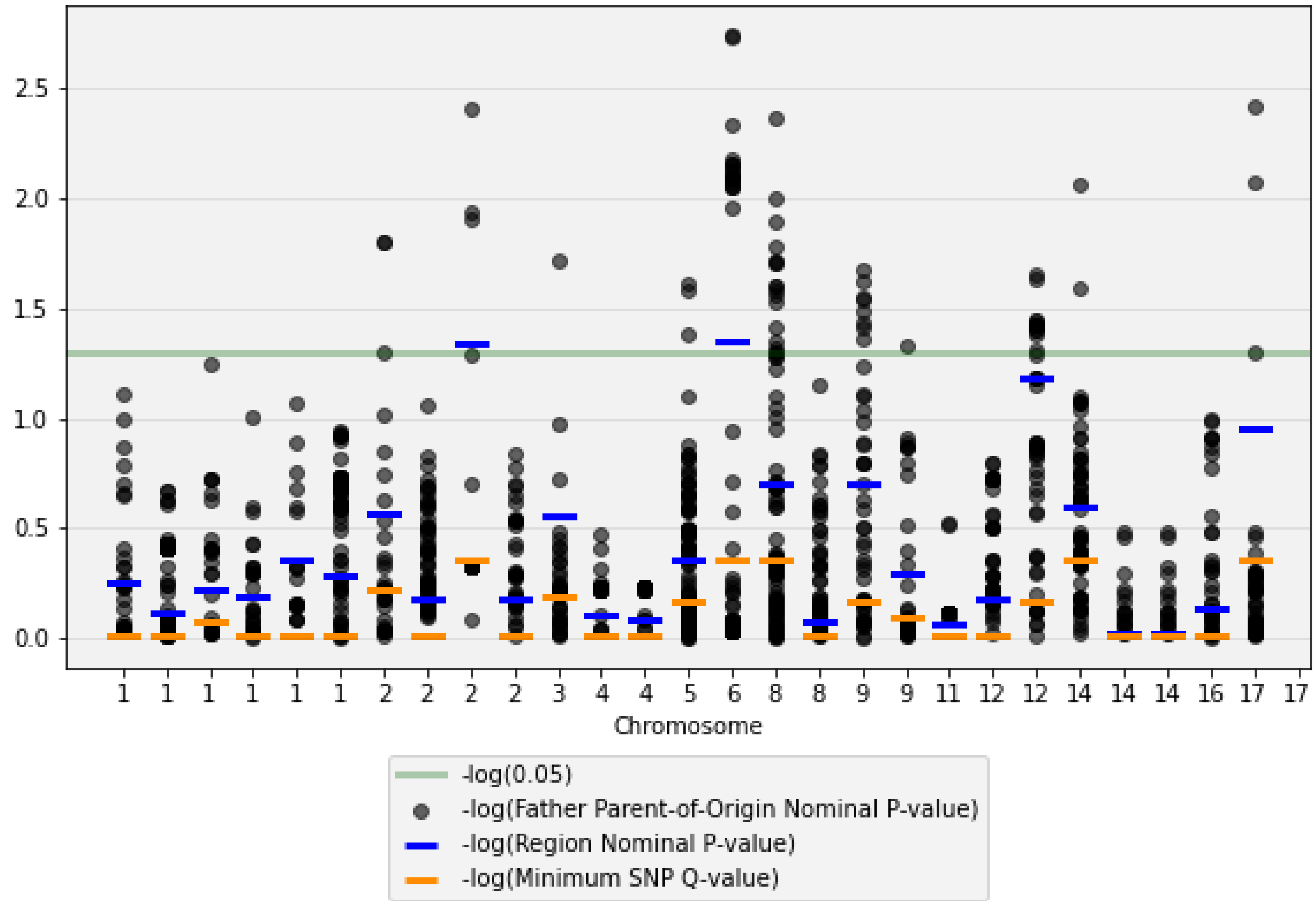

Figure 6.3. Manhattan plot of the father parent-of-origin genetic effects.

Table 6.1: Maternal genetic effect tests for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk[a]
rs153229[c]	5	G/A (0.40)	*RASGRF2*	0.0017	0.33	0.034	0.41	0.73
rs9805786[c]	13	G/T (0.39)	*SPATA13*	0.0050	0.33	0.087	0.44	1.32
rs7077430	10	A/G (0.08)	*JCAD*[b]	0.0068	0.33	0.015	0.38	0.62
rs11123015	2	C/G (0.22)	*CNTNAP5*[b]	0.0093	0.33	0.049	0.41	1.36
rs9900188[c]	17	A/G (0.16)	*MYH13*[b]	0.010	0.33	0.12	0.50	1.41
rs1450101[c]	3	A/G (0.30)	*GRM7*	0.025	0.33	0.073	0.44	1.27
rs1040735	1	G/T (0.44)	*LRRC7*	0.022	0.33	0.26	0.73	0.80
rs6900447	6	T/C (0.33)	*BMP6*	0.028	0.33	0.84	0.73	1.26
rs1999414	10	T/G (0.32)	*PLPP4*[b]	0.046	0.33	0.50	0.76	0.81

[a] Estimate of the relative risk associated with the presence of one copy of the minor allele. The estimate of the relative risk associated with two copies is the square of that for one copy.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for maternal genetic effects.
[e] False discovery rate corrected q-value for maternal genetic effects.
[f] aSPU test p-value for maternal genetic effects.
[g] False discovery rate corrected q-value of the aSPU test.

Table 6.2. Mother parent-of-origin genetic effect tests for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk[a]
rs817537	2	T/G (0.31)	*LTBP1*	0.0011	0.66	0.030	0.41	0.57
rs12048281	1	A/C (0.33)	*PBX1*	0.0025	0.66	0.029	0.41	1.64
rs2428173	6	G/A (0.34)	*SLC22A16*[b]	0.0030	0.66	0.085	0.77	1.63
rs10089607	8	C/G (0.36)	*MTUS1*	0.0095	0.71	0.27	0.88	0.63
rs11120177	1	A/G (0.10)	*PROX1-AS1*[b]	0.011	0.71	0.23	0.88	0.55
rs2543356	14	T/C (0.10)	*GPATCH2L*[b]	0.025	0.71	0.38	0.88	0.59
rs111451719	16	A/G (0.05)	*SYT17*	0.027	0.71	0.34	0.88	2.11
rs11999634	9	T/C (0.18)	*SMARCA2*	0.032	0.71	0.39	0.88	1.54
rs2628163	5	T/C (0.33)	*IRX2*[b]	0.039	0.71	0.25	0.88	1.41
rs57958042	17	T/C (0.21)	*MPRIP*[b]	0.040	0.71	0.61	0.89	1.44
rs2449817[c]	8	G/T (0.47)	*CSMD1*	0.043	0.71	0.86	0.89	1.40

[a] Estimate of relative risk associated with the maternal allele being transmitted.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for mother parent-of-origin genetic effects.
[e] False discovery rate corrected q-value for mother parent-of-origin genetic effects.
[f] aSPU test p-value for the mother parent-of-origin genetic effects.
[g] False discovery rate corrected q-value of the aSPU test.

Table 6.3. Father parent-of-origin genetic effect tests for the most significant SNP in nominally significant regions.

Marker	Chromosome	Minor/Other Allele (Minor Allele Frequency)	Gene	Uncorrected P-Value[d]	Q-Value[e]	aSPU uncorrected P-Value[f]	aSPU Q-value[g]	Relative Risk[a]
rs112946488	6	G/T (0.38)	*SLC22A16*[b]	0.0018	0.43	0.044	0.59	0.59
rs9907708	17	T/A (0.27)	*TNFRSF13B*[b]	0.0038	0.43	0.11	0.74	0.61
rs2375763	2	C/T (0.15)	*TMEM182*[b]	0.0039	0.43	0.046	0.59	1.80
rs1834569	8	A/C (0.34)	*CSMD1*	0.0043	0.43	0.20	0.83	0.62
rs12886269	14	T/G (0.34)	*GPATCH2L*[b]	0.0086	0.43	0.25	0.83	0.64
rs74565093	2	C/T (0.06)	*LTBP1*	0.016	0.60	0.27	0.83	2.08
rs3774577[c]	3	C/T (0.34)	*CACNA1D*	0.019	0.65	0.28	0.83	0.66
rs10811361	9	A/G (0.27)	*SMARCA2*	0.021	0.68	0.20	0.83	1.50
rs34716529	12	T/C (0.14)	*CCDC38*	0.022	0.68	0.066	0.59	0.61
rs468727[c]	5	G/A (0.37)	*IRX2*[b]	0.026	0.69	0.44	0.93	0.68
rs1187286	9	G/T (0.27)	*NTRK2*	0.047	0.81	0.51	0.93	1.39

[a] Estimate of relative risk associated with the paternal allele being transmitted.
[b] SNP is in close proximity but not in the gene.
[c] Genotyped SNP.
[d] Wald test p-value for father parent-of-origin genetic effects.
[e] False discovery rate corrected q-value for father parent-of-origin genetic effects.
[f] aSPU test p-value for the father parent-of-origin genetic effects.
[g] False discovery rate corrected q-value of the aSPU test.

Power Calculations

The power calculations are based on offspring genetic effects. A power of 0.8 is obtained for relative risks as low as 1.62 to 2.55 for the sample size used in parent-of-origin genetic effects and 1.61 to 2.53 for the sample size used in maternal genetic effects, depending on the minor allele frequency. This calculation is based on a sample size of 411 case-parent triads, 15 mother-child dyads and 10 father-child dyads. The alpha in the power calculation underwent a Bonferroni correction.

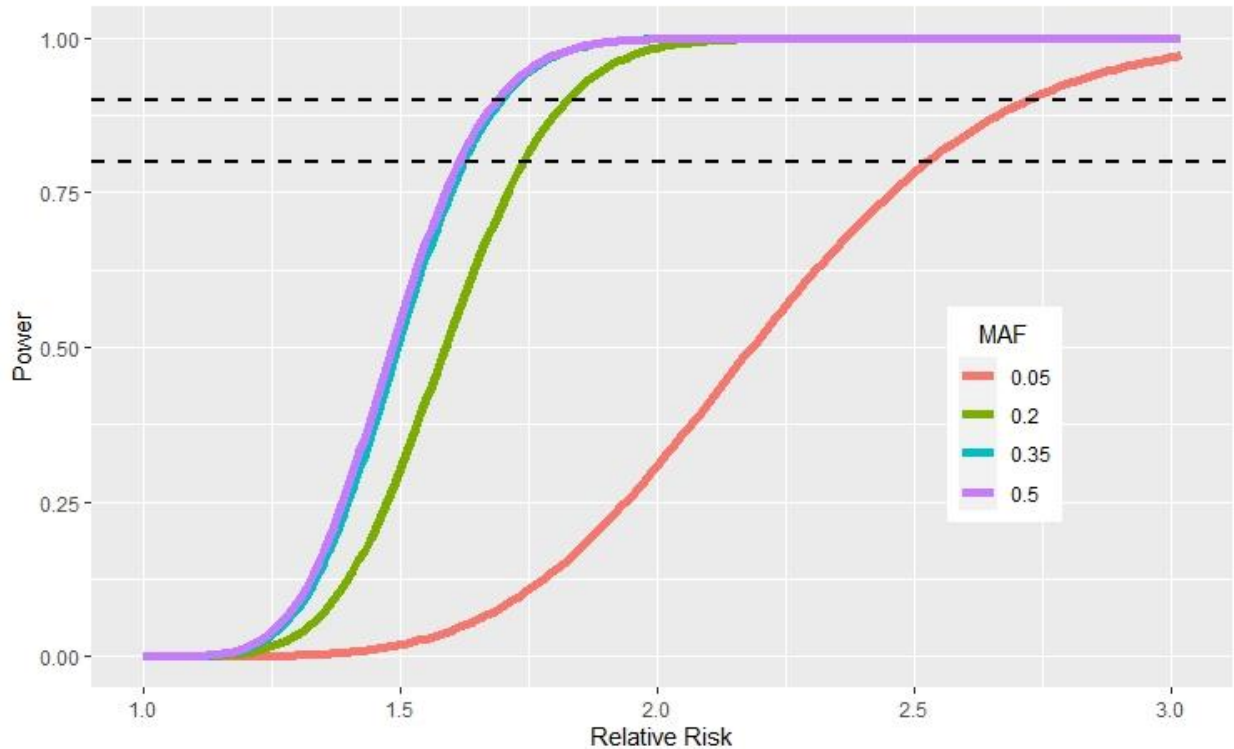

Figure 6.4. Maternal Power Calculation. 97% of the observed relative risks are below RR = 1.5. Alpha was set to 0.05/949. The solid lines represent the power calculation for the number of triads after quality control (411 Triads, 15 Mother-Child Dyads and 10 Father-Child Dyads).

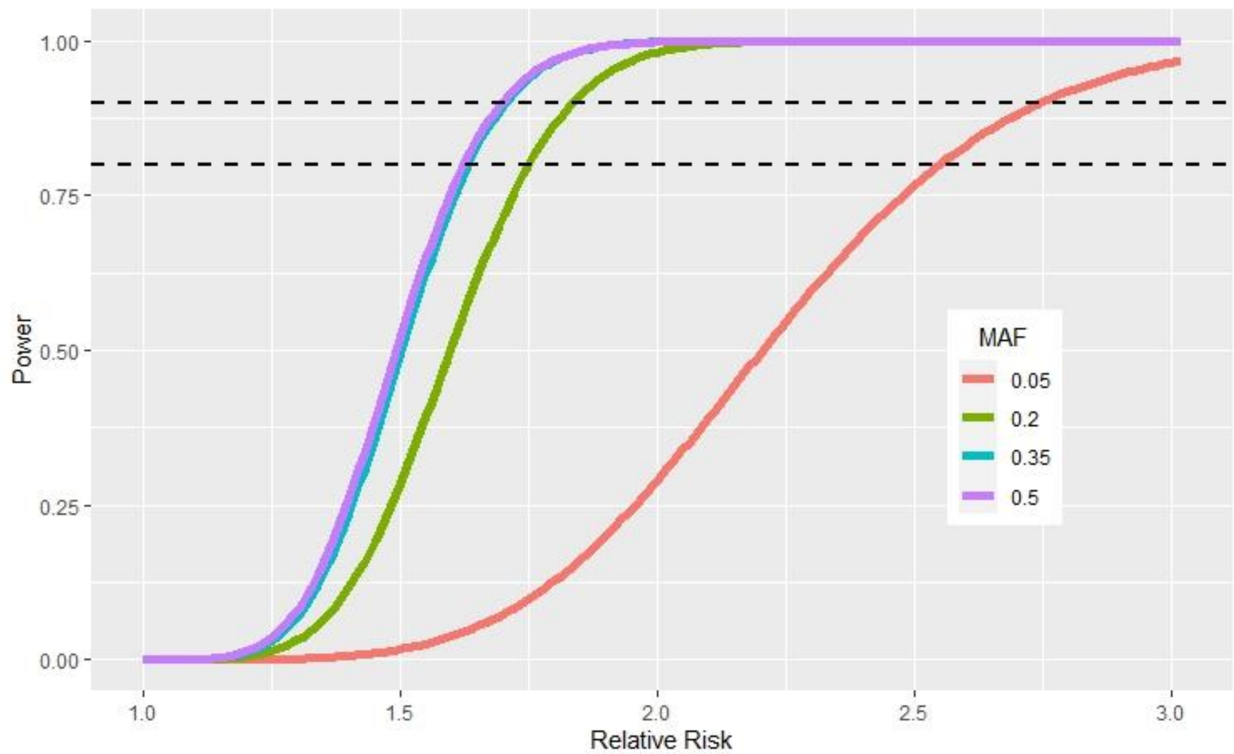

Figure 6.5. Parent-of-Origin Power Calculation. 97% of the observed relative risks are below RR = 1.5. Alpha was set to 0.05/1193. The solid lines represent the power calculation for the number of triads after quality control (411 Triads, 15 Mother-Child Dyads and 10 Father-Child Dyads).

Chapter 7: Discussion

Discussion of Results

One of the goals of this study was to replicate the parent-of-origin and maternal effects observed in Shi et al. (2012). Shi et al. (2012) tested for maternal genetic effects and parent-of-origin genetic effects (tables 4.1 and 4.2). In their study, none of the SNPs passed the GWAS level of significance, however some did get close. The replication results are shown in tables 6.1 to 6.3. Instead of reporting the relative risk ratio for parent-of-origin effects, the relative risks of the mother and father transmission were reported. For maternal genetic effects, the lowest p-value observed from Shi et al. (2012) is in an intergenic region in chromosome 17 with a p-value of 5.03^{-07}. Other genes from this study that showcased a relatively high significance are *GRM7, FHIT, BMP6, LRR7, SPATA13, SH3BP4, RASGFR2, TXNDC16* and *ATNX3*. In our replication, *RASGFR2* contained one of the most significant regions for maternal genetic effects, although the p-value was only 0.0017. Secondly, Shi et al. (2012) tested for parent-of-origin genetic effects with the smallest observed p-value of 1.29^{-06} in an intergenic region in chromosome 1. The genes that exhibited a relatively high level of significance in Shi et al. (2012) included *BCYRN1, SYT17, SMARCA2, NTRK2, NAP5, CCDC38, KCNK10, CHD1L* and *CSMD1*. In the present study, one of the most significant regions for the mother parent-of-origin effect is in *LTBP1* (p = 0.0011). This is a gene that codes for a protein belonging to a family of TGF-β binding proteins. Specifically, the protein is responsible for targeting latent complexes of TGF-β to the extracellular matrix where TGF-β is then activated (Tritschler et al., 2009). TGF-β is a family of cytokines that play important roles in growth and development, particularly in utero (Clark and Coker, 1998). As for expression, *LTBP1* is primarily expressed in the gallbladder, placenta, endometrium and urinary bladder (Fagerberg et al., 2014).

A significant minimum q-value was observed for the region of the *ATXN3* gene (p = 0.00071 and q = 0.025) for the interaction between maternal periconceptional smoking and maternal effects. The relative risk for the smoking strata is lower than the non-smoking strata in the presence of the minor allele in this gene. It is worth noting that this gene does not have any nominally significant SNPs within the testing of only maternal genetic effects. This suggests that maternal smoking may have a real influence on the effect of the minor allele of this gene. Still, there is no literature that links the gene to OC. The mRNA of this gene is expressed across many different types of tissue (Fagerberg et al., 2014). A mutation of this gene is associated with Machado-Joseph disease, a neurodegenerative disorder (Bettencourt et al., 2010). In terms of function, *ATXN3* encodes for a protein that deubiquitinates proteins. In essence, it is involved in the regulation of the degradation of proteins (Zeng et al., 2018).

As for parent-of-origin interaction effects, there were no significant associations observed after correcting for multiple testing. However, an interesting pattern was observed with the genes *CSMD1*, *SYT17* and *KCNK10*. These genes displayed the highest nominal significance in the parent-of-origin interaction with folic acid supplementation. These genes also happen to be primarily expressed within the brain (Fagerberg et al., 2014). *CSMD1* has previously displayed associations with schizophrenia (Athanasiu et al., 2017). Specifically, *CSMD1* is involved in regulating inflammation in the nervous system (Kraus et al., 2006). *KCNK10* encodes for a protein located in a potassium ion channel. This gene plays a role in adipogenesis: the formation of fat cells. However, the exact mechanism of this process is not yet known (Nishizuka et al., 2014). A region in *KCNK10* also had one of the most significant interaction between smoking and parent-of-origin genetic effects with a region p-value of 0.011. Little is known about the function of the *SYT17* gene. Of all the parent-of-origin genetic effects discussed, the regions

were insignificant for child effects after correcting for multiple testing, indicating that the transmittance of the alleles from parent to child may have an unconfounded impact.

Out of all the genes discussed, only *CSMD1* has shown a significant association for OC in previous literature, specifically pertaining to offspring genetic effects (Camargo et al., 2012). In the current study, a sensitivity analysis was done where a free-dose response model was implemented instead of a multiplicative dose response model. Another sensitivity analysis was also done where only complete triad data was used. There was little difference in the results presented and the sensitivity analysis results for the most significant regions (results not shown).

Biological and Public Health Implications

Gene environment effects are studied to (i) understand the biomolecular processes of disease, (ii) use the gene-environment effects to improve prognosis and treatment and (iii) identify groups of individuals that may be at higher risk (McAllister et al., 2017).

In terms of understanding the biomolecular processes of OC, *ATXN3* is an interesting gene to observe. As mentioned previously, mutation of this gene is associated with Machado-Joseph disease, and further research can be conducted to identify the possible mechanism by which this gene may affect the etiology of OC. Interestingly, onset of Machado-Joseph disease is characterized as an expansion of the $(CAG)_n$ motif in cytogenic position 14q32.1 (Bettencourt et al., 2010). The most significant SNP in the region identified in this study is fairly close, as it is located in 14q32.12. As such, it is interesting that this SNP is located in a gene that is involved in a neurological disorder. Another interesting observation is that many genes that are highly expressed in the brain showed nominally significant associations with folate supplementation. A study concluded that children with isolated cleft lip with or without palate had an abnormal brain

structure (Nopoulos et al., 2007). Since 60% of the brain is made of fat (Chang et al., 2009), it may be worth taking a closer look at the *KCNK10* gene, since it does play a role in adipogenesis. Additionally, there is evidence that non-syndromic OC is associated with schizophrenia (Pedersen et al., 2016). One of the genes identified in this study: *CSMD1*, is also associated with schizophrenia. As such, *CSMD1* may be involved in the molecular mechanism of OC. Due to the biological plausibility of some of these results, follow up research should be conducted through possible replications with different populations and more refined exposure information. Additional research can be conducted in the form of gene candidate studies, investigation of possible gene-gene interactions and gene-knockout studies.

If any of these genes are demonstrated to be involved in the molecular processes of OC, this can improve prognosis and treatment. Especially since all of these genes have some form of expression or function in the brain. It can also help identify individuals at higher risk. As such, recommendations such as consumption of supplements containing folic acid or reduction of smoking can be advised.

Strengths

The combination of the EUROCRAN and ITALCLEFT studies gives rise to one of the largest case-parent triad sample sizes for OC. Furthermore, the studies include populations from across Europe, which increases the diversity of exposures like tobacco smoking and consumption of supplements containing folic acid. This increased diversity is expected to increase the power of the study (Mossey et al., 2017). Unlike Canada and the United States, Europe does not have a folate fortification program (Berry et al., 2010). This results in lower baseline folate levels in mothers and therefore a ceiling effect associated with the consumption of supplements containing folic acid is unlikely.

Limitations

One limitation is that there are some weaknesses in the case-parent study design. Mating symmetry is assumed when estimating maternal genetic effects. Mating symmetry is defined as the occurrence of equal allele frequencies for males and females in a population. When calculating maternal genetic effects, formula 5 (Chapter 4) is dependent on the contrast of allele frequencies between the mother and the father (Gjessing and Lie, 2006). Additionally, when observing associations between gene-exposure interactions in case-triad data, independence of child genotype and exposure conditional on parental mating type needs to be assumed. However, case-parent data is not enough for testing some of these needed assumptions (Weinberg and Umbach, 2005). Another issue is that offspring gene, maternal gene, parent-of-origin gene or maternal gene-offspring gene interaction effects are all connected to one-another, and it can often be difficult to disentangle the true effects of each genetic mechanism (Buyske, 2008). It is also important to take note of confounding by population substructure. Confounding by population substructure can be described as artificial associations observed between markers and disease, due to the differing frequencies of the markers in the subpopulations. All log-linear models implemented in Haplin assume Hardy-Weinberg equilibrium, which makes the tests of association sensitive to confounding by population stratification (Ainsworth et al., 2011). Assuming Hardy-Weinberg equilibrium reduces the number of parameters to estimate in the model and allows more efficient computation and the testing of hypotheses that could not be tested otherwise. Although Hardy-Weinberg equilibrium was respected for all the SNPs showing significance in this study, further investigation of the significant results is needed to ensure that they are not spurious results due to population stratification.

A stronger design would be a case-control design with parents. This type of study design would combine the strengths of both the case-control and case-parent triad designs. For example, when estimating child or parent-of-origin genetic effects in case-parent triads, transmission ratio distortion is a bias where the transmission of a specific parent's allele is not related to the disease in question but is over or under transmitted due to another biological cause, thus giving an artificial association (Huang et al., 2013). Adding a control triad, would provide protection against this bias. Protection against other biases would also be present in this type of study design, such as mating asymmetry bias for maternal genetic effects.

Lastly, all the sub-phenotypes were pooled together for an overall analysis in the study. Cleft palate and cleft lip with or without palate are thought to have different etiologies, and thus may present different estimates of significance (Murray, 2002; Spritz, 2001). As such, it would be worth obtaining a larger sample size to appropriately stratify sub-phenotypes.

Future Research

A region in the gene *ATXN3* displayed significance after correcting for multiple testing. This significance was observed in the interaction between maternal genetic effects and maternal periconceptional smoking. Mutations in *ATXN3* is involved with the onset of a neurological disorder. A gene candidate study is warranted given the significance observed with this association. An interesting observation is that *CSMD1, KCNK10* and *SYT17* all had their largest significance in the interaction between folic acid supplementation and parent-of-origin genetic effects while also having predominant expression in the brain. As a result, investigations connecting the role of the brain to orofacial clefts, as it pertains to the genes listed here, should be conducted. A replication can be conducted with different populations and more detailed exposure information. Novel studies like possible gene-gene interactions of the genes listed here

can also be performed. Once consistent associations appear for these genes, possible gene-knockout studies can also be conducted. Additionally, future research pertaining to the investigation of solely cleft palate and cleft lip with or without palate should be conducted with larger sample sizes, as these phenotypes have distinct etiologies.

Conclusion

In conclusion, log-linear models were used to estimate the parent-of-origin and maternal genetic effects, along with their interaction with maternal consumption of supplements containing folic acid and maternal smoking, using case-parent triad data. Maternal gene-smoking interaction effects were found to be significant for a region located in the *ATXN3* gene. Parent-of-origin gene-folic acid interaction effects were found to be nominally significant for regions located in the *CSMD1*, *KCNK10* and *SYT17* genes. These genes are all expressed in the brain. Some of these genes are involved with neurological disorders as well. These results give further evidence that biological mechanisms in the brain may play a key role in the etiology of OC.

References

Abrishamchian AR, Khoury MJ, Calle EE. 1994. The contribution of maternal epilepsy and its treatment to the etiology of oral clefts: a population based case-control study. Genet Epidemiol 11(4):343-351.

Ainsworth HF, Unwin J, Jamison DL, Cordell HJ. 2011. Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. Genet Epidemiol 35(1):19-45.

Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. 2015. Molecular Biology of the Cell. New York, NY: Garland Science.

Athanasiu L, Giddaluru S, Fernandes C, Christoforou A, Reinvang I, Lundervold AJ, Nilsson LG, Kauppi K, Adolfsson R, Eriksson E, Sundet K, Djurovic S, Espeseth T, Nyberg L, Steen VM, Andreassen OA, Le Hellard S. 2017. A genetic association study of CSMD1 and CSMD2 with cognitive function. Brain Behav Immun 61:209-216.

Avşar TS, McLeod H, Jackson L. 2021. Health outcomes of smoking during pregnancy and the postpartum period: an umbrella review. BMC Pregnancy Childbirth 21(1):254.

Basha M, Demeer B, Revencu N, Helaers R, Theys S, Bou Saba S, Boute O, Devauchelle B, Francois G, Bayet B, Vikkula M. 2018. Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. J Med Genet 55(7):449-458.

Baxter D, Shanks A. 2020. Pierre Robin Syndrome.

Bazerman MH, Samuelson WF. 1983. I won the auction but don't want the prize. Journal of Conflict Resolution 27(4):618-634.

Beaty TH, Marazita ML, Leslie EJ. 2016. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Res 5:2800.

Beaty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, Murray T, Redett RJ, Fallin MD, Liang KY, Wu T, Patel PJ, Jin SC, Zhang TX, Schwender H, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Ye X, Wang H, Huang S, Jabs EW, Shi B, Wilcox AJ, Lie RT, Jee SH, Christensen K, Doheny KF, Pugh EW, Ling H, Scott AF. 2011. Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet Epidemiol 35(6):469-478.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57(1):289-300.

Berg E, Haaland Ø A, Feragen KB, Filip C, Vindenes HA, Moster D, Lie RT, Sivertsen Å. 2016. Health Status Among Adults Born With an Oral Cleft in Norway. JAMA Pediatr 170(11):1063-1070.

Berry RJ, Bailey L, Mulinare J, Bower C. 2010. Fortification of flour with folic acid. Food Nutr Bull 31(1 Suppl):S22-35.

Bettencourt C, Santos C, Montiel R, Kay T, Vasconcelos J, Maciel P, Lima M. 2010. The (CAG)n tract of Machado-Joseph Disease gene (ATXN3): a comparison between DNA and mRNA in patients and controls. Eur J Hum Genet 18(5):621-623.

Bille C, Olsen J, Vach W, Knudsen VK, Olsen SF, Rasmussen K, Murray JC, Andersen AM, Christensen K. 2007. Oral clefts and life style factors--a case-cohort study based on prospective Danish data. Eur J Epidemiol 22(3):173-181.

Botto LD, Erickson JD, Mulinare J, Lynberg MC, Liu Y. 2002. Maternal fever, multivitamin use, and selected birth defects: evidence of interaction? Epidemiology 13(4):485-488.

Boulet SL, Grosse SD, Honein MA, Correa-Villaseñor A. 2009. Children with orofacial clefts: health-care use and costs among a privately insured population. Public Health Rep 124(3):447-453.

Boyles AL, DeRoo LA, Lie RT, Taylor JA, Jugessur A, Murray JC, Wilcox AJ. 2010. 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 172(8):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. 2009. Oral facial clefts and gene polymorphisms in metabolism of folate/one-carbon and vitamin A: a pathway-wide association study. Genet Epidemiol 33(3):247-255.

Buyske S. 2008. Maternal genotype effects can alias case genotype effects in case-control studies. Eur J Hum Genet. England. p 783-785.

Calzolari E, Pierini A, Astolfi G, Bianchi F, Neville AJ, 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 143a(6):528-537.

Camargo M, Rivera D, Moreno L, Lidral AC, Harper U, Jones M, Solomon BD, Roessler E, Vélez JI, Martinez AF, Chandrasekharappa SC, Arcos-Burgos M. 2012. GWAS reveals new recessive loci associated with non-syndromic facial clefting. Eur J Med Genet 55(10):510-514.

Castilla EE, Orioli IM, Lugarinho R, Dutra GP, Lopez-Camelo JS, Campana HE, Spagnolo A, Mastroiacovo P. 1990. Monthly and seasonal variations in the frequency of congenital anomalies. Int J Epidemiol 19(2):399-404.

Chang CY, Ke DS, Chen JY. 2009. Essential fatty acids and human brain. Acta Neurol Taiwan 18(4):231-241.

Chen XK, Wen SW, Fleming N, Yang Q, Walker MC. 2007. Teenage pregnancy and congenital anomalies: which system is vulnerable? Hum Reprod 22(6):1730-1735.

Christensen K, Juel K, Herskind AM, Murray JC. 2004. Long term follow up study of survival associated with cleft lip and palate at birth. Bmj 328(7453):1405.

Clark DA, Coker R. 1998. Transforming growth factor-beta (TGF-beta). Int J Biochem Cell Biol 30(3):293-298.

Czeizel AE, Dudás I. 1992. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. N Engl J Med 327(26):1832-1835.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. 2016. Next-generation genotype imputation service and methods. Nat Genet 48(10):1284-1287.

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013a. Haplotype estimation using sequencing reads. Am J Hum Genet 93(4):687-696.

Delaneau O, Marchini J, Zagury JF. 2011. A linear complexity phasing method for thousands of genomes. Nat Methods 9(2):179-181.

Delaneau O, Zagury JF, Marchini J. 2013b. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. United States. p 5-6.

DeRoo LA, Wilcox AJ, Lie RT, Romitti PA, Pedersen DA, Munger RG, Moreno Uribe LM, Wehby GL. 2016. Maternal alcohol binge-drinking in the first trimester and the risk of orofacial clefts in offspring: a large population-based pooling study. Eur J Epidemiol 31(10):1021-1034.

Dixon MJ, Marazita ML, Beaty TH, Murray JC. 2011. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet 12(3):167-178.

Dravet C, Julian C, Legras C, Magaudda A, Guerrini R, Genton P, Soulayrol S, Giraud N, Mesdjian E, Trentin G, et al. 1992. Epilepsy, antiepileptic drugs, and malformations in children of women with epilepsy: a French prospective cohort study. Neurology 42(4 Suppl 5):75-82.

Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjöstedt E, Lundberg E, Szigyarto CA, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson A, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlén M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics 13(2):397-406.

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. 2003. Identification of SATB2 as the cleft palate gene on 2q32-q33. Hum Mol Genet 12(19):2491-2501.

Fraser FC, Gwyn A. 1998. Seasonal variation in birth date of children with cleft lip. Teratology 57(2):93-95.

Galloway J, Davies G, Mossey P. 2017. International Knowledge of Direct Costs of Cleft Lip and Palate Treatment. Arch Pediatr Surg. p 10-25.

Garland MA, Reynolds K, Zhou CJ. 2020a. Environmental mechanisms of orofacial clefts. Birth Defects Res 112(19):1660-1698.

Garland MA, Sun B, Zhang S, Reynolds K, Ji Y, Zhou CJ. 2020b. Role of epigenetics and miRNAs in orofacial clefts. Birth Defects Res 112(19):1635-1659.

Gjerdevik M, Haaland Ø A, Romanowska J, Lie RT, Jugessur A, Gjessing HK. 2018. Parent-of-origin-environment interactions in case-parent triads with or without independent controls. Ann Hum Genet 82(2):60-73.

Gjessing HK, Lie RT. 2006. Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. Ann Hum Genet 70(Pt 3):382-396.

Haaland A, Lie RT, Romanowska J, Gjerdevik M, Gjessing HK, Jugessur A. 2018. A Genome-Wide Search for Gene-Environment Effects in Isolated Cleft Lip with or without Cleft Palate Triads Points to an Interaction between Maternal Periconceptional Vitamin Use and Variants in ESRRG. Front Genet 9:60.

Haaland Ø A, Jugessur A, Gjerdevik M, Romanowska J, Shi M, Beaty TH, Marazita ML, Murray JC, Wilcox AJ, Lie RT, Gjessing HK. 2017. Genome-wide analysis of parent-of-origin interaction effects with environmental exposure (PoOxE): An application to European and Asian cleft palate trios. PLoS One 12(9):e0184358.

Haaland Ø A, Romanowska J, Gjerdevik M, Lie RT, Gjessing HK, Jugessur A. 2019. A genome-wide scan of cleft lip triads identifies parent-of-origin interaction effects between ANK3 and maternal smoking, and between ARHGEF10 and alcohol consumption. F1000Res 8:960.

Hackshaw A, Rodeck C, Boniface S. 2011. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. Hum Reprod Update 17(5):589-604.

Honein MA, Rasmussen SA, Reefhuis J, Romitti PA, Lammer EJ, Sun L, Correa A. 2007. Maternal smoking and environmental tobacco smoke exposure and the risk of orofacial clefts. Epidemiology 18(2):226-233.

Huang LO, Labbe A, Infante-Rivard C. 2013. Transmission ratio distortion: review of concept and implications for genetic association studies. Hum Genet 132(3):245-263.

ICBDSR. 2015. ICBDSR (International Clearinghouse for Birth Defects Surveillance and Research) (2014) Annual Report Ed. Rome.

Ioannidis JP, Tarone R, McLaughlin JK. 2011. The false-positive to false-negative ratio in epidemiologic studies. Epidemiology 22(4):450-456.

Izedonmwen OM, Cunningham C, Macfarlane TV. 2015. What is the Risk of Having Offspring with Cleft Lip/Palate in Pre-Maternal Obese/Overweight Women When Compared to Pre-Maternal Normal Weight Women? A Systematic Review and Meta-Analysis. J Oral Maxillofac Res 6(1):e1.

Izzotti A, Balansky RM, Cartiglia C, Camoirano A, Longobardi M, De Flora S. 2003. Genomic and transcriptional alterations in mouse fetus liver after transplacental exposure to cigarette smoke. Faseb j 17(9):1127-1129.

Jakobsen LP, Knudsen MA, Lespinasse J, García Ayuso C, Ramos C, Fryns JP, Bugge M, Tommerup N. 2006. The genetic basis of the Pierre Robin Sequence. Cleft Palate Craniofac J 43(2):155-159.

Jugessur A, Shi M, Gjessing HK, Lie RT, Wilcox AJ, Weinberg CR, Christensen K, Boyles AL, Daack-Hirsch S, Nguyen TT, Christiansen L, Lidral AC, Murray JC. 2010. Maternal genes and facial clefts in offspring: a comprehensive search for genetic associations in two population-based cleft studies from Scandinavia. PLoS One 5(7):e11493.

Kadir A, Mossey PA, Blencowe H, Moorthie S, Lawn JE, Mastroiacovo P, Modell B. 2017. Systematic Review and Meta-Analysis of the Birth Prevalence of Orofacial Clefts in Low- and Middle-Income Countries. Cleft Palate Craniofac J 54(5):571-581.

Kaler AS, Purcell LC. 2019. Estimation of a significance threshold for genome-wide association studies. BMC Genomics 20(1):618.

Kernahan DA, Stark RB. 1958. A new classification for cleft lip and cleft palate. Plast Reconstr Surg Transplant Bull 22(5):435-441.

Krapels IP, Rooij IA, Wevers RA, Zielhuis GA, Spauwen PH, Brussel W, Steegers-Theunissen RP. 2004. 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 111(7):661-668.

Kraus DM, Elliott GS, Chute H, Horan T, Pfenninger KH, Sanford SD, Foster S, Scully S, Welcher AA, Holers VM. 2006. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. J Immunol 176(7):4419-4430.

Kutbi H, Wehby GL, Moreno Uribe LM, Romitti PA, Carmichael S, Shaw GM, Olshan AF, DeRoo L, Rasmussen SA, Murray JC, Wilcox A, Lie RT, Munger RG. 2017. Maternal underweight and obesity and risk of orofacial clefts in a large international consortium of population-based studies. Int J Epidemiol 46(1):190-199.

Kwak IY, Pan W. 2016. Adaptive gene- and pathway-trait association testing with GWAS summary statistics. Bioinformatics 32(8):1178-1184.

Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, Curry CJ, Fernhoff PM, Grix AW, Jr., Lott IT, et al. 1985. Retinoic acid embryopathy. N Engl J Med 313(14):837-841.

Lawson HA, Cheverud JM, Wolf JB. 2013. Genomic imprinting and parent-of-origin effects on complex traits. Nat Rev Genet 14(9):609-617.

Lie RT, Wilcox AJ, Taylor J, Gjessing HK, Saugstad OD, Aabyholm F, Vindenes H. 2008. Maternal smoking and oral clefts: the role of detoxification pathway genes. Epidemiology 19(4):606-615.

Little J, Cardy A, Munger RG. 2004. Tobacco smoking and oral clefts: a meta-analysis. Bull World Health Organ 82(3):213-218.

Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N. 2009. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6(2):e22.

Little J, Nelson C. 2013. CONGENITAL ANOMALIES IN CANADA 2013. Orofacial Clefts. p 42-49.

Loenen WA. 2006. S-adenosylmethionine: jack of all trades and master of everything? Biochem Soc Trans 34(Pt 2):330-333.

Macdonald WA. 2012. Epigenetic mechanisms of genomic imprinting: common themes in the regulation of imprinted regions in mammals, plants, and insects. Genet Res Int 2012:585024.

Malic CC, Lam M, Donelle J, Richard L, Vigod SN, Benchimol EI. 2020. Incidence, Risk Factors, and Mortality Associated With Orofacial Cleft Among Children in Ontario, Canada. JAMA Netw Open 3(2):e1921036.

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. 2009. Finding the missing heritability of complex diseases. Nature 461(7265):747-753.

McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, Conti D, Gauderman WJ, Hsu L, Hutter CM, Jankowska MM, Kerr J, Kraft P, Montgomery SB, Mukherjee B, Papanicolaou GJ, Patel CJ, Ritchie MD, Ritz BR, Thomas DC, Wei P, Witte JS. 2017. Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases. Am J Epidemiol 186(7):753-761.

McBride WA, McIntyre GT, Carroll K, Mossey PA. 2016. Subphenotyping and Classification of Orofacial Clefts: Need for Orofacial Cleft Subphenotyping Calls for Revised Classification. Cleft Palate Craniofac J 53(5):539-549.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R. 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48(10):1279-1283.

Meyer KA, Werler MM, Hayes C, Mitchell AA. 2003. Low maternal alcohol consumption during pregnancy and oral clefts in offspring: the Slone Birth Defects Study. Birth Defects Res A Clin Mol Teratol 67(7):509-514.

Miller TJ. 1972. Cleft palate formation: a role for pyridoxine in the closure of the secondary palate in mice. Teratology 6(3):351-356.

Mitchell LE, Murray JC, O'Brien S, Christensen K. 2003. 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 158(1):69-76.

Mossey PA, Davies JA, Little J. 2007. Prevention of orofacial clefts: does pregnancy planning have a role? Cleft Palate Craniofac J 44(3):244-250.

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. 2009. Cleft lip and palate. Lancet 374(9703):1773-1785.

Mossey PA, Little J, Steegers-Theunissen R, Molloy A, Peterlin B, Shaw WC, Johnson C, FitzPatrick DR, Franceschelli P, Rubini M. 2017. Genetic Interactions in Nonsyndromic Orofacial Clefts in Europe-EUROCRAN Study. Cleft Palate Craniofac J 54(6):623-630.

MRC Vitamin Study Research Group. 1991. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. Lancet 338(8760):131-137.

Munger RG, Sauberlich HE, Corcoran C, Nepomuceno B, Daack-Hirsch S, Solon FS. 2004. Maternal vitamin B-6 and folate status and risk of oral cleft birth defects in the Philippines. Birth Defects Res A Clin Mol Teratol 70(7):464-471.

Murray JC. 2002. Gene/environment causes of cleft lip and/or palate. Clin Genet 61(4):248-256.

Ngai CW, Martin WL, Tonks A, Wyldes MP, Kilby MD. 2005. Are isolated facial cleft lip and palate associated with increased perinatal mortality? A cohort study from the West Midlands Region, 1995-1997. J Matern Fetal Neonatal Med 17(3):203-206.

Nishizuka M, Hayashi T, Asano M, Osada S, Imagawa M. 2014. KCNK10, a tandem pore domain potassium channel, is a regulator of mitotic clonal expansion during the early stage of adipocyte differentiation. Int J Mol Sci 15(12):22743-22756.

Nopoulos P, Langbehn DR, Canady J, Magnotta V, Richman L. 2007. Abnormal brain structure in children with isolated clefts of the lip or palate. Arch Pediatr Adolesc Med 161(8):753-758.

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, McQuillan R, Fraser RM, Campbell H, Polasek O, Asiki G, Ekoru K, Hayward C, Wright AF, Vitart V, Navarro P, Zagury JF, Wilson JF, Toniolo D, Gasparini P, Soranzo N, Sandhu MS, Marchini J. 2014. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet 10(4):e1004234.

Pan W, Kim J, Zhang Y, Shen X, Wei P. 2014. A powerful and adaptive association test for rare variants. Genetics 197(4):1081-1095.

Panamonta V, Pradubwong S, Panamonta M, Chowchuen B. 2015. Global Birth Prevalence of Orofacial Clefts: A Systematic Review. J Med Assoc Thai 98 Suppl 7:S11-21.

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. 2000. Birth defects after maternal exposure to corticosteroids: prospective cohort study and meta-analysis of epidemiological studies. Teratology 62(6):385-392.

Pavri S, Forrest CR. 2013. Demographics of orofacial clefts in Canada from 2002 to 2008. Cleft Palate Craniofac J 50(2):224-230.

Pedersen DA, Wehby GL, Murray JC, Christensen K. 2016. Psychiatric Diagnoses in Individuals with Non-Syndromic Oral Clefts: A Danish Population-Based Cohort Study. PLoS One 11(5):e0156261.

Prescott NJ, Lees MM, Winter RM, 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(3):345-350.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559-575.

R Core Team. 2019. R: A language and environment for statistical computing.

Raut JR, Simeone RM, Tinker SC, Canfield MA, Day RS, Agopian AJ. 2019. Proportion of Orofacial Clefts Attributable to Recognized Risk Factors. Cleft Palate Craniofac J 56(2):151-158.

Romitti PA, Lidral AC, Munger RG, Daack-Hirsch S, Burns TL, Murray JC. 1999. 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 59(1):39-50.

Romitti PA, Sun L, Honein MA, Reefhuis J, Correa A, Rasmussen SA. 2007. Maternal periconceptional alcohol consumption and risk of orofacial clefts. Am J Epidemiol 166(7):775-785.

Rothman KJ, Moore LL, Singer MR, Nguyen US, Mannino S, Milunsky A. 1995. Teratogenicity of high vitamin A intake. N Engl J Med 333(21):1369-1373.

Sabbagh HJ, Hassan MH, Innes NP, Elkodary HM, Little J, Mossey PA. 2015. Passive smoking in the etiology of non-syndromic orofacial clefts: a systematic review and meta-analysis. PLoS One 10(3):e0116963.

Saleem K, Zaib T, Sun W, Fu S. 2019. Assessment of candidate genes and genetic heterogeneity in human non syndromic orofacial clefts specifically non syndromic cleft lip with or without palate. Heliyon 5(12):e03019.

Shaw GM, Nelson V, Iovannisci DM, Finnell RH, Lammer EJ. 2003. Maternal occupational chemical exposures and biotransformation genotypes as risk factors for selected congenital anomalies. Am J Epidemiol 157(6):475-484.

Shaw GM, Wasserman CR, O'Malley CD, Lammer EJ, Finnell RH. 1995. Orofacial clefts and maternal anticonvulsant use. Reprod Toxicol. United States. p 97-98.

Shi M, Christensen K, Weinberg CR, Romitti P, Bathum L, Lozada A, Morris RW, Lovett M, Murray JC. 2007. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. Am J Hum Genet 80(1):76-90.

Shi M, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Wu T, Murray T, Redett RJ, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Yeow V, Chong SS, Shi B, Christensen K, Scott AF, Patel P, Cheah F, Beaty TH. 2012. Genome wide study of maternal and parent-of-origin effects on the etiology of orofacial clefts. Am J Med Genet A 158a(4):784-794.

Shi M, Weinberg CR. 2011. How much are we missing in SNP-by-SNP analyses of genome-wide association studies? Epidemiology 22(6):845-847.

Skare O, Jugessur A, Lie RT, Wilcox AJ, Murray JC, Lunde A, Nguyen TT, Gjessing HK. 2012. Application of a novel hybrid study design to explore gene-environment interactions in orofacial clefts. Ann Hum Genet 76(3):221-236.

Smahel Z, Brejcha M. 1983. Differences in craniofacial morphology between complete and incomplete unilateral cleft lip and palate in adults. Cleft Palate J 20(2):113-127.

Spritz RA. 2001. The genetics and epigenetics of orofacial clefts. Curr Opin Pediatr 13(6):556-560.

Stock NM, Feragen KB, Rumsey N. 2015. "It Doesn't All Just Stop at 18": Psychological Adjustment and Support Needs of Adults Born With Cleft Lip and/or Palate. Cleft Palate Craniofac J 52(5):543-554.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100(16):9440-9445.

Stothard KJ, Tennant PW, Bell R, Rankin J. 2009. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. Jama 301(6):636-650.

Tamura T, Munger RG, Corcoran C, Bacayao JY, Nepomuceno B, Solon F. 2005. Plasma zinc concentrations of mothers and the risk of nonsyndromic oral clefts in their children: a

case-control study in the Philippines. Birth Defects Res A Clin Mol Teratol 73(9):612-616.

Thieme F, Ludwig KU. 2017. The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts. J Dent Res 96(11):1238-1247.

Tritschler I, Gramatzki D, Capper D, Mittelbronn M, Meyermann R, Saharinen J, Wick W, Keski-Oja J, Weller M. 2009. Modulation of TGF-beta activity by latent TGF-beta-binding protein 1 in human malignant glioma cells. Int J Cancer 125(3):530-540.

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. 2003. 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 157(7):583-591.

Van Rossum G, Drake Jr. F. 1995. Python Tutorial. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Waller DK, Shaw GM, Rasmussen SA, Hobbs CA, Canfield MA, Siega-Riz AM, Gallaway MS, Correa A. 2007. Prepregnancy obesity as a risk factor for structural birth defects. Arch Pediatr Adolesc Med 161(8):745-750.

Wang M, Yuan Y, Wang Z, Liu D, Sun F, Wang P, Zhu H, Li J, Wu T, Beaty TH. 2017. Prevalence of Orofacial Clefts among Live Births in China: A Systematic Review and Meta-Analysis. Birth Defects Res 109(13):1011-1019.

Warkany J, Nelson RC. 1940. APPEARANCE OF SKELETAL ABNORMALITIES IN THE OFFSPRING OF RATS REARED ON A DEFICIENT DIET. Science 92(2391):383-384.

Wehby GL, Félix TM, Goco N, Richieri-Costa A, Chakraborty H, Souza J, Pereira R, Padovani C, Moretti-Ferreira D, Murray JC. 2013. High dosage folic acid supplementation, oral cleft recurrence and fetal growth. Int J Environ Res Public Health 10(2):590-605.

Weinberg CR. 1999. Allowing for missing parents in genetic studies of case-parent triads. Am J Hum Genet 64(4):1186-1193.

Weinberg CR, Umbach DM. 2005. A hybrid design for studying genetic influences on risk of diseases with onset early in life. Am J Hum Genet 77(4):627-636.

Weinberg CR, Wilcox AJ, Lie RT. 1998. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 62(4):969-978.

Wolf JB, Wade MJ. 2009. What are maternal effects (and what are they not)? Philos Trans R Soc Lond B Biol Sci 364(1520):1107-1115.

Wong WY, Eskes TK, Kuijpers-Jagtman AM, Spauwen PH, Steegers EA, Thomas CM, Hamel BC, Blom HJ, Steegers-Theunissen RP. 1999. Nonsyndromic orofacial clefts: association with maternal hyperhomocysteinemia. Teratology 60(5):253-257.

Yin X, Li J, Li Y, Zou S. 2019. Maternal alcohol consumption and oral clefts: a meta-analysis. Br J Oral Maxillofac Surg 57(9):839-846.

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, Zhou F, Chen G, Zheng X, Wang X, Liang B, Zhu Z, Fu X, Sheng Y, Hao J, Liu Z, Yan H, Mangold E, Ruczinski I, Liu J, Marazita ML, Ludwig KU, Beaty TH, Zhang X, Sun L, Bian Z. 2017. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat Commun 8:14364.

Zeiger JS, Hetmanski JB, Beaty TH, VanderKolk CA, Wyszynski DF, Bailey-Wilson JE, de Luna RO, Perandones C, Tolarova MM, Mosby T, Bennun R, Segovia M, Calda P, Pugh EW, Doheny K, McIntosh I. 2003. Evidence for linkage of nonsyndromic cleft lip with or without cleft palate to a region on chromosome 2. Eur J Hum Genet 11(11):835-839.

Zeng L, Zhang D, McLoughlin HS, Zalon AJ, Aravind L, Paulson HL. 2018. Loss of the Spinocerebellar Ataxia type 3 disease protein ATXN3 alters transcription of multiple signal transduction pathways. PLoS One 13(9):e0204438.

Zhou Y, Sinnathamby V, Yu Y, Sikora L, Johnson CY, Mossey P, Little J. 2020. Folate intake, markers of folate status and oral clefts: An updated set of systematic reviews and meta-analyses. Birth Defects Res 112(19):1699-1719.

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. 2004. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med 351(8):769-780.