

Genome-wide interaction study implicates *VGLL2* and alcohol exposure and *PRL* and smoking in orofacial cleft risk

Jenna C. Carlson¹, John R. Shaffer², Fred Deleyiannis³, Jacqueline T. Hecht⁴, George L. Wehby⁵, Kaare Christensen⁶, Eleanor Feingold^{1,2}, Seth M. Weinberg^{2,6}, Mary L. Marazita^{2,7}, Elizabeth J. Leslie⁸

- 1 Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
- 2 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
- 3 UHealth Medical Group, Colorado Springs, CO. USA
- 4 Department of Pediatrics, University of Texas Health Science Center at Houston, Houston, TX, USA
- 5 Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA, USA
- 6 Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark
- 7 Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, 15219 USA
- 8 Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

ABSTRACT

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is a common birth defect, affecting approximately 1 in 700 births. NSCL/P has complex etiology including several known genes and environmental factors; however, known genetic risk variants only account for a small fraction of the heritability of NSCL/P. It is commonly suggested that gene-by-environment (G×E) interactions may help explain some of the “missing” heritability of NSCL/P. We conducted a genome-wide G×E interaction study in European cases and controls with three common maternal exposures during pregnancy: alcohol, smoking, and vitamin use using a two-stage design. After selecting 127 loci with suggestive 2df tests for gene and G x E effects, 40 loci showed significant G x E effects after correcting for multiple tests. Notable interactions included SNPs of 6q22 near *VGLL2* with alcohol and 6p22.3 near *PRL* with smoking. These interactions could provide new insights into the etiology of CL/P and new opportunities to modify risk through behavioral changes.

INTRODUCTION

Interest in identifying the causal factors for birth defects, including orofacial clefts (OFCs) can be traced back centuries (1) and has often involved a debate as to the contribution of genetic versus environmental risk factors(2). Evidence for a multifactorial model for OFC etiology originated with Fraser’s early studies exposing pregnant mice to cortisone showing that the incidence of corticosteroid induced cleft palate varied by strain. The multifactorial model is favored in human nonsyndromic OFCs, where twin and family studies provide evidence for a strong, but incomplete, genetic component. Many environmental factors have been postulated to modify risk of OFCs including maternal medications, smoking or alcohol consumption(3, 4), nutrition (5), obesity (6), gestational diabetes (7), and occupational exposures. Of these, cigarette smoking, alcohol consumption, and folic acid supplementation are the most widely studied, but, except for smoking(8), these exposures have inconsistent results in epidemiological studies. Maternal smoking, on the other hand, has been shown to increase risk consistently across many studies with similar effect sizes(9).

Over the last ten years, 20 independent genome-wide association studies (GWAS) or meta-analyses have identified at least 50 loci associated with OFCs (10). Cumulatively, these loci are estimated to account for 25-30% of the heritable risk attributed to additive genetic effects. Other approaches are needed to identify the “missing” heritability, some of which may be attributed to interaction effects. Leveraging published data from GWAS studies, gene-environment interaction (G×E) studies have become a popular approach to further elucidate OFC risk. Following one such GWAS study in European and Asian case-parent trios (Beaty 2010), candidate gene and pathway-based G×E analyses have suggested interactions between smoking and *RUNX2* (11), and between multiple exposures and *BMP4* (12). Genome-wide G×E using a variety of statistical approaches have identified interactions for maternal smoking, alcohol, or folate supplementation in cleft palate (13-15) and cleft lip with or without cleft palate (16, 17).

In this study, we performed genome-wide G×E analyses in a European case-control sample from the Pittsburgh Orofacial Cleft Study to identify interactions between genetic variants and maternal smoking, alcohol consumption, or vitamin supplementation during the periconceptual period that influence risk of OFCs.

MATERIALS AND METHODS

Study Samples and Genotyping

The sample for these analyses was derived from a larger multiethnic OFC cohort, which has been previously described (18). Briefly, participants were recruited from 18 sites worldwide as part of ongoing genetic studies conducted by the University of Iowa and the University of Pittsburgh Center for Craniofacial and Dental Genetics. All sites obtained Institutional Review Board approval both locally and at the University of Iowa or the University of Pittsburgh. All participants gave informed consent. These data are available through dbGaP (accession number: phs000774.v2.p1).

This large multiethnic cohort contains OFC-affected probands and their unaffected relatives in addition to controls without a family history of OFC or other craniofacial anomalies. Additionally, data were obtained on three maternal periconceptional exposures: alcohol use, smoking, and vitamin use. Periconceptional exposures were gathered from maternal self-report of alcohol use, personal smoking, or vitamin use during the three months prior to pregnancy and during the first trimester for each child.

From this larger cohort, a subset of unrelated cases and controls of European ancestry and with complete data for the three environmental exposure measures was drawn. It consisted of 344 NSCL/P cases and 194 controls of European descent from Denmark, Hungary, and the United States (Table 1).

The genotyping, quality control procedures, imputation, and generation of principal components of ancestry (PCAs) have been previously described (18). Briefly, samples were genotyped using the Illumina HumanCore + Exome chip (Illumina Inc., San Diego, CA), with approximately 97% of the genotyped SNPs passing quality control filters, resulting in a total of 539,473 genotyped SNPs. Genotype imputation for an additional 34,985,077 unobserved polymorphisms was performed using IMPUTE2 software with the multiethnic 1,000 Genomes Project Phase 3 reference panel (19). Imputed genotype probabilities were converted to most-likely genotypes using GTOOL; only most-likely genotypes with probabilities >0.9 were retained for statistical analysis. Imputed SNPs with INFO scores <0.5 or those deviating from Hardy–Weinberg equilibrium ($p < 1 \times 10^{-4}$ in a set of unrelated European controls) were also excluded from the analysis. The University of Washington, Genetics Coordinating Center released a full online report on the data cleaning, quality assurance, ancestry analyses, and imputation for this study (http://www.ccdg.pitt.edu/docs/Marazita_ofc_QC_report_feb2015.pdf, last accessed April 25, 2016). Additionally, we excluded variants with a minor allele frequency $< 10\%$, as the power to detect interaction effects among low-frequency variants is very limited and analysis of such variants is prone to type 1 errors. After these filtering steps, 5,165,675 were included in the genome-wide interaction analyses. Population structure was measured through generation of principal components of ancestry using the R package SNPRelate following the approach introduced by Patterson, Price, and Reich (20).

Statistical Analyses

To assess the association between genetic variants and NSCL/P that may or may not be modified by an environmental factor, we used the strategy laid out in Kraft et al 2007 for each of the three environmental factors (21). First, a 2 degree of freedom (2df) joint test of the gene (G) and gene-environment interaction (GE) effects was employed genome-wide using logistic

regression, adjusting for five principal components of ancestry. This provides a sensitive test for two scenarios simultaneously: (1) where there is a genetic effect that is similar across all environmental strata, and (2) where there is a genetic effect that is specific to a specific environmental stratum. Then, the GE effect was interrogated for any variants demonstrating at least suggestive results from the G-GE joint test (i.e., $p < 1 \times 10^{-4}$) to identify regions in which there is a genetic effect differing across environmental stratum. We used a Bonferroni-adjusted significance threshold for examining the GE effect, adjusting for the number of regions with suggestive results from the G-GE joint test. All analyses were conducted in PLINK v1.9 (22).

RESULTS

As a follow-up study to previous GWAS identifying marginal gene effects (G) increasing risk for NSCL/P, this study focused on identifying G x E interactions. We carried out a two-stage analysis, where we first examined the 2df joint test of G and G x E (GE). We observed suggestive evidence of association in the joint G-GE test for 127 loci. Several loci had multiple SNPs with $p < 1 \times 10^{-4}$ in each of the three analyses. These included 8q24, one of the most reliably associated loci with NSCL/P in European populations (10). As expected, this locus showed strong G effects, but did not have a significant GE effect ($p > 0.05$). A locus on 2p23 with similarly strong G effects but no GE effects was identified in all three analyses but to our knowledge has not been reported before in GWAS of NSCL/P.

To identify gene-environment interactions, we next examined the GE effect alone for each variant with $p < 1 \times 10^{-4}$ in the joint test. 40 regions (212 SNPs) had a GE p-value less than 3.9×10^{-4} (i.e., Bonferroni correction for 127 loci) (Table 2), although none had a genome-wide significant GE p-value. Among these regions, we observed a statistically significant interaction effect between rs706954, a variant <5kb downstream of *VGLL2*, and periconceptual alcohol use ($p_{GE} = 4.62 \times 10^{-6}$). The minor allele (G) of rs706945 was associated with higher odds of NSCL/P within individuals with periconceptual alcohol use, but lower odds with individuals without periconceptual alcohol use (Figure 2).

Two additional loci had more than 5 SNPs per locus with significant GE p-values for alcohol exposures. These included 6q21 (lead SNP rs6929494, $p_{GE} = 1.18 \times 10^{-5}$) and 11q21 (lead SNP rs7116990, $p_{GE} = 1.80 \times 10^{-5}$). At the 6q21 locus, which included genes *REVL3* and *TRAF3IP2*, the major allele (A) of rs6929494 was associated with higher odds of NSCL/P within individuals with periconceptual alcohol use, but had an opposite (protective) effect for individuals without periconceptual alcohol use (Figure S1). The associated SNPs at the 11q21 locus, are located in an intergenic region between *CNTN5* and *JRKL*; for the lead SNP, rs7116990, the major allele (A) was associated with greater odds of NSCL/P in the individuals with periconceptual alcohol use and lower odds of NSCL/P for individuals without periconceptual alcohol use (Figure S1).

In the periconceptual smoking and vitamin use analyses, only one locus in each study had more than 5 significant SNPs. Thirty-six SNPs on 3q27.3 showed a significant interaction with vitamin use (lead SNP rs61069053, $p_{GE} = 1.05 \times 10^{-5}$). The minor allele (G) of rs61069053, located in an intergenic region between *CRYGS* and *DGKG*, had higher estimated odds of NSCL/P in the individuals without periconceptual vitamin use, but had lower odds of NSCL/P for the periconceptual vitamin user group (Figure S1). However, this result may be confounded by the distribution of vitamin use and allele frequency; only one individual was homozygous for the

minor allele and did not have vitamin use. At 6p22.3, 47 SNPs showed significant GE p-values for smoking (lead SNP rs35097027, $p_{GE}=1.96\times 10^{-5}$). These variants are upstream of *PRL*; the minor allele (C) of rs35097027 conferred higher odds of CL/P for individuals with maternal smoking, but lower odds of NSCL/P for individuals without maternal smoking (Figure S1).

DISCUSSION

The primary goal of this paper was to identify additional gene-environment interactions for NSCL/P and build on our previous work in this dataset (18). These analyses were motivated by multiple studies reporting associations between OFCs and periconceptional smoking, alcohol use, and vitamin supplementation (10, 23). Genome-wide interaction scans, however, have been limited for NSCL/P (13, 16). Our analyses add to a growing list of loci for which interactions between SNPs and maternal environmental exposures influence risk of OFCs.

The 2df joint G-GE test has statistical power above 80% for variants with minor allele frequency > 0.25 with strong environmental and/or genetic effects ($OR>1.5$), even with a relatively small sample size (21). Consistent with these calculations, the loci in these analyses with statistically significant G-GE associations that were driven by the GE effect had markedly strong effect estimates. Notably, the estimated odds of NSCL/P were 2.73 times higher with each copy of the minor allele G of rs117083 for individuals with periconceptional alcohol use and was 35% lower for individuals without periconceptional alcohol use. For the 5 loci with more than 5 SNPs per locus with significant GE p-values (i.e., 6q21 [alcohol], 6q22.1 [alcohol], 11q21 [alcohol], 6p22.3, [smoking], and 3q27.3 [vitamin use]), the genotypic odds ratio ranged from 60% lower odds of CL/P to 4.47 times higher odds. For alcohol and smoking exposures, we primarily identified scenarios described as “cross-over” interactions where the genotype group at lowest risk for NSCL/P without the exposure had high risk for NSCL/P with the exposure. In each of these interactions, the influence of genotype entirely depends on the exposure. The method we used to detect interactions is better suited for these “cross-over” interactions. However, it is limited in its ability to detect associations that are only present in one environmental context. For example, it is quite plausible that a genetic variant may only modify risk in the presence of a specific environment. This method has lower statistical power to detect such an association; additional methods are needed to explicitly search for these association signals.

We identified multiple loci with significant interactions. Of these, two have notable biological plausibility which we discuss in detail. One of the loci interacting with periconceptional alcohol use was on 6q22.1, approximately 5kb upstream of *VGLL2*, encoding vestigial-like 2. *VGLL2* is one of several Vestigial-like factors that plays a role in muscle fiber differentiation and the distribution of skeletal muscle fibers (24). Although its expression in adult tissues is restricted to skeletal muscle, during development, *Vgll2* is more broadly expressed. In mice, its expression is enriched in the mandibular and maxillary prominences at embryonic day 10.5 and is localized to the epithelia of these structures(25). In the zebrafish, *Vgll2* is expressed in the pharyngeal pouches and somites (26, 27). A role for *VGLL2* in craniofacial development is further supported by *vgll2a* zebrafish morphants, which induce death of neural crest cells in the pharyngeal arches causing hypoplasia of the Meckel’s and palatoquadrate cartilages and truncation of the ethmoid plate (25). Models for fetal alcohol syndrome indicate that ethanol exposure enhances cell death of cranial neural crest cells (28, 29). Although a specific interaction between *VGLL2* and ethanol has not been described in vitro or in vivo, these data, combined with our statistical interaction suggest this may be a fruitful area for future studies.

At the 6p22.3 locus associated with maternal smoking, the lead SNP was located 60kb upstream of *PRL*, which encodes prolactin. Prolactin is a reproductive hormone primarily known for its ability to stimulate mammary gland development and lactation, but it also has a variety of functions in reproduction, growth, and development (30). Prolactin receptor transcripts and proteins have been detected in non-lactogenic tissues notably including the facial cartilage and olfactory epithelium in mammals (31). Prolactin levels are known to be altered by both active smoking and exposure to secondhand tobacco smoke (32-35). Although smoking is associated with lower prolactin levels in pregnant women, the same is not true for fetal prolactin levels (36). A role for prolactin in craniofacial development and OFCs is unclear as knockouts of mouse *PRL* receptors are viable and lack gross morphological defects (37). However, in amphibian models, prolactin signaling has been suggested as the pathway underlying the ability of pre-metamorphic *Xenopus laevis* tadpoles to correct craniofacial defects induced by thioridazine (38). However, the same group also found that increased prolactin signaling during development alone does not cause craniofacial defects. Nonetheless, there is evidence linking smoking to prolactin levels and for prolactin signaling in craniofacial development; additional research will be needed to connect these results and determine a role for *PRL* and smoking in OFCs.

Our previous studies using this dataset have made use of the ancestral diversity that resulted from the 13 different recruitment sites worldwide. However, this study was limited to individuals of European descent as the number of unrelated participants with complete data from the other ancestry groups was quite small. In previous GWASs for gene-environment interactions, a lack of observations precluded the analysis of smoking and alcohol exposures in Asian trios (16). The authors speculated that this was consistent with general trends of low alcohol and cigarette use among Asian women (39). This may also be true for other populations recruited into our study, but the incomplete data may also be due to changing study designs over the recruitment period. We also note that the recruitment of these participants was not population-based and thus does not reflect the population prevalence of NSCL/P nor the three environmental exposures we considered. In addition, a binary exposure variable does not capture or represent the range of possible environmental exposures that may influence development of NSCL/P. Therefore, the conclusions drawn here are limited to the study sample used in these analyses and cannot be used to draw conclusions at a broader level.

To summarize, we performed a genome-wide analysis to detect gene-environment interactions influencing risk of NSCL/P in Europeans. We identified two notable interactions: near *VGLL2* and *PRL* with periconceptional alcohol use and smoking, respectively that are plausibly associated with risk of NSCL/P. These interaction effects are novel and warrant further investigations. If confirmed, these interactions provide new insights into the etiology of NSCL/P and could provide opportunities to modify risk through behavioral changes.

ACKNOWLEDGEMENTS

This project would not have been possible without the participation of families, field staff, and collaborators around the world. Special recognition to our colleague Dr. Andrew Czeizel (deceased). This work was supported by grants from the National Institutes of Health (NIH): R00-DE025060 [EJL], X01-HG007485 [MLM, EF], R01-DE016148 [MLM, SMW]. Genotyping and data cleaning were provided via an NIH contract to the Johns Hopkins Center for Inherited Disease Research: HHSN268201200008I.

CONFLICT OF INTEREST

None.

REFERENCES

1. Marazita ML. The evolution of human genetic studies of cleft lip and cleft palate. *Annu Rev Genomics Hum Genet.* 2012;13:263-83. Epub 2012/06/19. doi: 10.1146/annurev-genom-090711-163729. PubMed PMID: 22703175; PMCID: PMC3760163.
2. Beames TG, Lipinski RJ. Gene-environment interactions: aligning birth defects research with complex etiology. *Development.* 2020;147(21). Epub 2020/07/19. doi: 10.1242/dev.191064. PubMed PMID: 32680836; PMCID: PMC7375468.
3. Grewal J, Carmichael SL, Ma C, Lammer EJ, Shaw GM. Maternal periconceptional smoking and alcohol consumption and risk for select congenital anomalies. *Birth Defects Res A Clin Mol Teratol.* 2008;82(7):519-26. Epub 2008/05/17. doi: 10.1002/bdra.20461. PubMed PMID: 18481814; PMCID: PMC2861577.
4. Romitti PA, Sun L, Honein MA, Reefhuis J, Correa A, Rasmussen SA. Maternal periconceptional alcohol consumption and risk of orofacial clefts. *Am J Epidemiol.* 2007;166(7):775-85. Epub 2007/07/05. doi: 10.1093/aje/kwm146. PubMed PMID: 17609516.
5. Kelly D, O'Dowd T, Reulbach U. Use of folic acid supplements and risk of cleft lip and palate in infants: a population-based cohort study. *Br J Gen Pract.* 2012;62(600):e466-72. Epub 2012/07/12. doi: 10.3399/bjgp12X652328. PubMed PMID: 22781994; PMCID: PMC3381272.
6. Blomberg MI, Kallen B. Maternal obesity and morbid obesity: the risk for birth defects in the offspring. *Birth Defects Res A Clin Mol Teratol.* 2010;88(1):35-40. Epub 2009/08/28. doi: 10.1002/bdra.20620. PubMed PMID: 19711433.
7. Figueiredo JC, Ly S, Magee KS, Ihenacho U, Baurley JW, Sanchez-Lara PA, Brindopke F, Nguyen TH, Nguyen V, Tangco MI, Giron M, Abrahams T, Jang G, Vu A, Zolfaghari E, Yao CA, Foong A, DeClerk YA, Samet JM, Magee W, 3rd. Parental risk factors for oral clefts among Central Africans, Southeast Asians, and Central Americans. *Birth Defects Res A Clin Mol Teratol.* 2015;103(10):863-79. Epub 2015/10/16. doi: 10.1002/bdra.23417. PubMed PMID: 26466527; PMCID: PMC5049483.
8. Hackshaw A, Rodeck C, Boniface S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Hum Reprod Update.* 2011;17(5):589-604. Epub 2011/07/13. doi: 10.1093/humupd/dmr022. PubMed PMID: 21747128; PMCID: PMC3156888.
9. Little J, Cardy A, Munger RG. Tobacco smoking and oral clefts: a meta-analysis. *Bull World Health Organ.* 2004;82(3):213-8. Epub 2004/04/28. PubMed PMID: 15112010; PMCID: PMC2585921.
10. Beaty TH, Marazita ML, Leslie EJ. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000Res.* 2016;5:2800. doi: 10.12688/f1000research.9503.1. PubMed PMID: 27990279; PMCID: PMC5133690.
11. Wu T, Fallin MD, Shi M, Ruczinski I, Liang KY, Hetmanski JB, Wang H, Ingersoll RG, Huang S, Ye X, Wu-Chou YH, Chen PK, Jabs EW, Shi B, Redett R, Scott AF, Murray JC, Marazita ML, Munger RG, Beaty TH. Evidence of gene-environment interaction for the RUNX2 gene and environmental tobacco smoke in controlling the risk of cleft lip with/without cleft palate. *Birth Defects Res A Clin Mol Teratol.* 2012;94(2):76-83. Epub 2012/01/14. doi: 10.1002/bdra.22885. PubMed PMID: 22241686; PMCID: PMC3608124.
12. Chen Q, Wang H, Schwender H, Zhang T, Hetmanski JB, Chou YH, Ye X, Yeow V, Chong SS, Zhang B, Jabs EW, Parker MM, Scott AF, Beaty TH. Joint testing of genotypic and gene-environment interaction

- identified novel association for BMP4 with non-syndromic CL/P in an Asian population using data from an International Cleft Consortium. *PLoS One*. 2014;9(10):e109038. Epub 2014/10/11. doi: 10.1371/journal.pone.0109038. PubMed PMID: 25303326; PMCID: PMC4193821.
13. Wu T, Schwender H, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, Parker MM, Wang P, Murray T, Taub M, Li S, Redett RJ, Fallin MD, Liang KY, Wu-Chou YH, Chong SS, Yeow V, Ye X, Wang H, Huang S, Jabs EW, Shi B, Wilcox AJ, Jee SH, Scott AF, Beaty TH. Evidence of gene-environment interaction for two genes on chromosome 4 and environmental tobacco smoke in controlling the risk of nonsyndromic cleft palate. *PLoS One*. 2014;9(2):e88088. Epub 2014/02/12. doi: 10.1371/journal.pone.0088088. PubMed PMID: 24516586; PMCID: PMC3916361.
 14. 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. Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol*. 2011;35(6):469-78. Epub 2011/05/28. doi: 10.1002/gepi.20595. PubMed PMID: 21618603; PMCID: PMC3180858.
 15. Haaland OA, Jugessur A, Gjerdevik M, Romanowska J, Shi M, Beaty TH, Marazita ML, Murray JC, Wilcox AJ, Lie RT, Gjessing HK. Genome-wide analysis of parent-of-origin interaction effects with environmental exposure (PoOxE): An application to European and Asian cleft palate trios. *PLoS One*. 2017;12(9):e0184358. Epub 2017/09/13. doi: 10.1371/journal.pone.0184358. PubMed PMID: 28898263; PMCID: PMC5595425.
 16. Haaland OA, Lie RT, Romanowska J, Gjerdevik M, Gjessing HK, Jugessur A. 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*. 2018;9:60. Epub 2018/03/15. doi: 10.3389/fgene.2018.00060. PubMed PMID: 29535761; PMCID: PMC5834486.
 17. Haaland OA, Romanowska J, Gjerdevik M, Lie RT, Gjessing HK, Jugessur A. 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*. 2019;8:960. Epub 2019/08/06. doi: 10.12688/f1000research.19571.2. PubMed PMID: 31372216; PMCID: PMC6662680.
 18. Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, Sanchez C, Jacobs J, Emanuele B, Vieira AR, Neiswanger K, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos-Burgos M, Czeizel AE, Field LL, Padilla CD, Cutiongco-de la Paz EM, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Castilla EE, Mereb JC, Poletta FA, Orioli IM, Carvalho FM, Hecht JT, Blanton SH, Buxo CJ, Butali A, Mossey PA, Adeyemo WL, James O, Braimah RO, Aregbesola BS, Eshete MA, Abate F, Koruyucu M, Seymen F, Ma L, de Salamanca JE, Weinberg SM, Moreno L, Murray JC, Marazita ML. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum Mol Genet*. 2016;25(13):2862-72. Epub 2016/04/02. doi: 10.1093/hmg/ddw104. PubMed PMID: 27033726; PMCID: PMC5181632.
 19. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529. Epub 2009/06/23. doi: 10.1371/journal.pgen.1000529. PubMed PMID: 19543373; PMCID: PMC2689936.
 20. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190. Epub 2006/12/30. doi: 10.1371/journal.pgen.0020190. PubMed PMID: 17194218; PMCID: PMC1713260.
 21. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111-9. Epub 2007/02/07. doi: 10.1159/000099183. PubMed PMID: 17283440.

22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. Epub 2015/02/28. doi: 10.1186/s13742-015-0047-8. PubMed PMID: 25722852; PMCID: PMC4342193.
23. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet*. 2011;12(3):167-78. Epub 2011/02/19. doi: 10.1038/nrg2933. PubMed PMID: 21331089; PMCID: PMC3086810.
24. Honda M, Hidaka K, Fukada SI, Sugawa R, Shirai M, Ikawa M, Morisaki T. Vestigial-like 2 contributes to normal muscle fiber type distribution in mice. *Sci Rep*. 2017;7(1):7168. Epub 2017/08/05. doi: 10.1038/s41598-017-07149-0. PubMed PMID: 28769032; PMCID: PMC5540913.
25. Johnson CW, Hernandez-Lagunas L, Feng W, Melvin VS, Williams T, Artinger KB. Vgll2a is required for neural crest cell survival during zebrafish craniofacial development. *Dev Biol*. 2011;357(1):269-81. Epub 2011/07/12. doi: 10.1016/j.ydbio.2011.06.034. PubMed PMID: 21741961; PMCID: PMC3519931.
26. Maeda T, Chapman DL, Stewart AF. Mammalian vestigial-like 2, a cofactor of TEF-1 and MEF2 transcription factors that promotes skeletal muscle differentiation. *J Biol Chem*. 2002;277(50):48889-98. Epub 2002/10/12. doi: 10.1074/jbc.M206858200. PubMed PMID: 12376544.
27. Mielcarek M, Gunther S, Kruger M, Braun T. VITO-1, a novel vestigial related protein is predominantly expressed in the skeletal muscle lineage. *Mech Dev*. 2002;119 Suppl 1:S269-74. Epub 2003/10/01. doi: 10.1016/s0925-4773(03)00127-8. PubMed PMID: 14516696.
28. Smith SM, Garic A, Flentke GR, Berres ME. Neural crest development in fetal alcohol syndrome. *Birth Defects Res C Embryo Today*. 2014;102(3):210-20. Epub 2014/09/16. doi: 10.1002/bdrc.21078. PubMed PMID: 25219761; PMCID: PMC4827602.
29. Cartwright MM, Smith SM. Stage-dependent effects of ethanol on cranial neural crest cell development: partial basis for the phenotypic variations observed in fetal alcohol syndrome. *Alcohol Clin Exp Res*. 1995;19(6):1454-62. Epub 1995/12/01. doi: 10.1111/j.1530-0277.1995.tb01007.x. PubMed PMID: 8749810.
30. Bole-Feysot C, Goffin V, Edery M, Binart N, Kelly PA. Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocr Rev*. 1998;19(3):225-68. Epub 1998/06/17. doi: 10.1210/edrv.19.3.0334. PubMed PMID: 9626554.
31. Freemark M, Driscoll P, Maaskant R, Petryk A, Kelly PA. Ontogenesis of prolactin receptors in the human fetus in early gestation. Implications for tissue differentiation and development. *J Clin Invest*. 1997;99(5):1107-17. Epub 1997/03/01. doi: 10.1172/JCI119239. PubMed PMID: 9062370; PMCID: PMC507920.
32. Xue Y, Morris M, Ni L, Guthrie SK, Zubieta JK, Gonzalez K, McConnell DS, Domino EF. Venous plasma nicotine correlates of hormonal effects of tobacco smoking. *Pharmacol Biochem Behav*. 2010;95(2):209-15. Epub 2010/01/27. doi: 10.1016/j.pbb.2010.01.007. PubMed PMID: 20100507; PMCID: PMC2832319.
33. Ng SP, Steinetz BG, Lasano SG, Zelikoff JT. Hormonal changes accompanying cigarette smoke-induced preterm births in a mouse model. *Exp Biol Med (Maywood)*. 2006;231(8):1403-9. Epub 2006/09/02. doi: 10.1177/153537020623100814. PubMed PMID: 16946409.
34. Muraki T, Tokunaga Y, Nakadate T, Kato R. Inhibition by cholinergic agonists of the prolactin release induced by morphine. *Naunyn Schmiedebergs Arch Pharmacol*. 1979;308(3):249-54. Epub 1979/09/01. doi: 10.1007/BF00501389. PubMed PMID: 503252.
35. Benedict MD, Missmer SA, Ferguson KK, Vitonis AF, Cramer DW, Meeker JD. Secondhand tobacco smoke exposure is associated with prolactin but not thyroid stimulating hormone among nonsmoking women seeking in vitro fertilization. *Environ Toxicol Pharmacol*. 2012;34(3):761-7. Epub 2012/10/11. doi: 10.1016/j.etap.2012.09.010. PubMed PMID: 23046534; PMCID: PMC3514562.

36. Andersen AN, Ronn B, Tjonneland A, Djursing H, Schioler V. Low maternal but normal fetal prolactin levels in cigarette smoking pregnant women. *Acta Obstet Gynecol Scand.* 1984;63(3):237-9. Epub 1984/01/01. doi: 10.3109/00016348409155504. PubMed PMID: 6730939.
37. Ormandy CJ, Camus A, Barra J, Damotte D, Lucas B, Buteau H, Edery M, Brousse N, Babinet C, Binart N, Kelly PA. Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. *Genes Dev.* 1997;11(2):167-78. Epub 1997/01/15. doi: 10.1101/gad.11.2.167. PubMed PMID: 9009200.
38. Pinet K, Deolankar M, Leung B, McLaughlin KA. Adaptive correction of craniofacial defects in pre-metamorphic *Xenopus laevis* tadpoles involves thyroid hormone-independent tissue remodeling. *Development.* 2019;146(14). Epub 2019/06/30. doi: 10.1242/dev.175893. PubMed PMID: 31253636.
39. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, Wollum A, Sanman E, Wulf S, Lopez AD, Murray CJ, Gakidou E. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA.* 2014;311(2):183-92. Epub 2014/01/09. doi: 10.1001/jama.2013.284692. PubMed PMID: 24399557.

TABLES AND FIGURES

Table 1. Characteristics of Sample. Sample size (column percent) comparing environmental exposure status between cases and controls within each recruitment site.

Table 2. Results for top-associated variants. Columns: CHR, BP, SNP, A2 (major), A1 (minor), MAF, imputed, P.2df, P.gxe, OR (G when E=0), OR (G when E=1), scan, cytoband

Figure 1. Manhattan plots for maternal periconceptional (A) alcohol use, (B) smoking, (C) vitamin use.

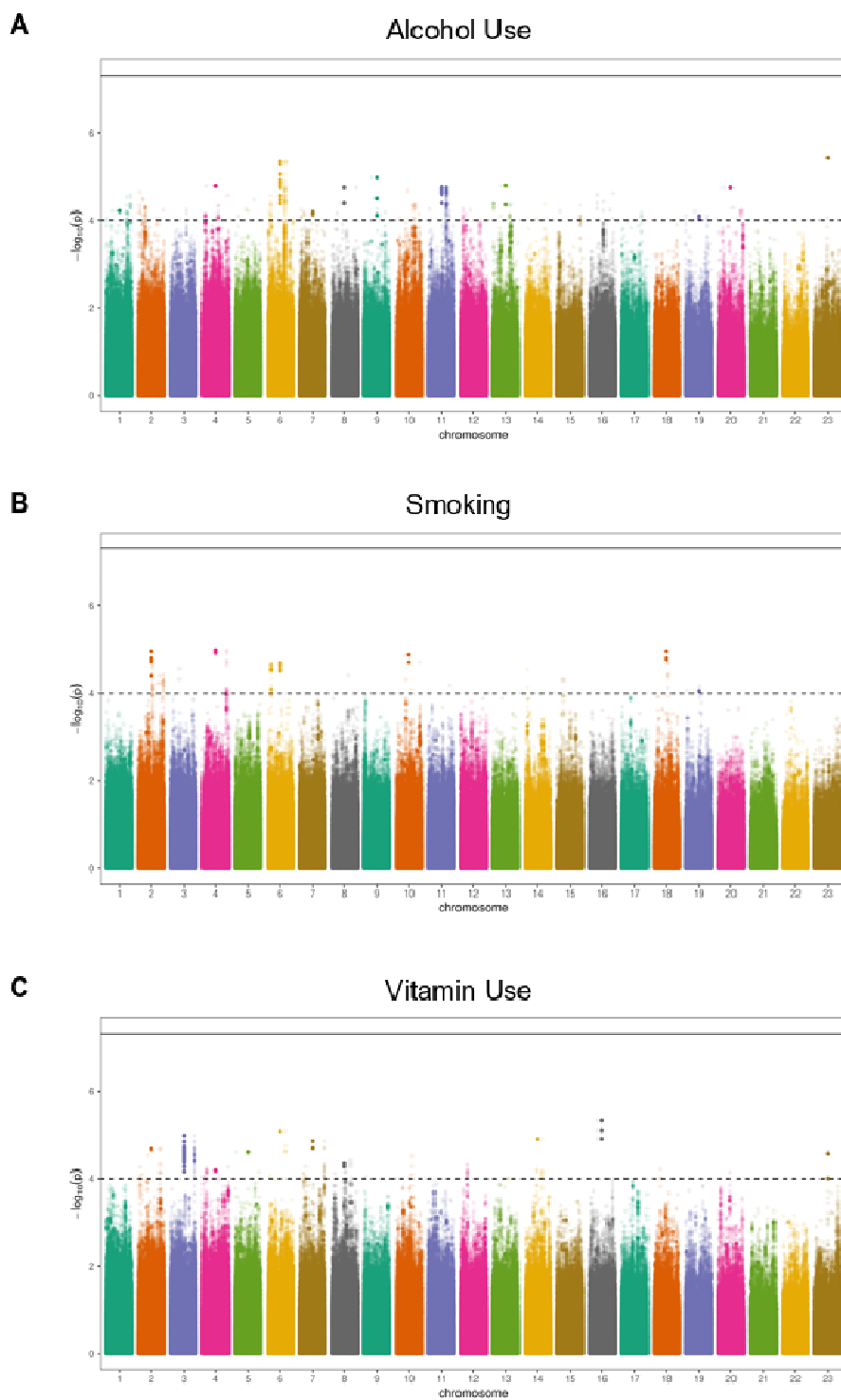


Figure 2. Regional association plots of the GE test with interaction plots (A) 6q22 (rs706954) interaction with alcohol use. (B) 6p22.3 (rs35097027) interaction with smoking. (C) 3q27.3 (rs61069053) with vitamin use. Note, only one individual is homozygous for the G allele in the no vitamin use strata

