Received: 2 January 2020 | Revised: 12 June 2020 | Accepted: 15 July 2020

DOI: 10.1002/gepi.22343

RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

WILEY

Evaluating shared genetic influences on nonsyndromic cleft lip/palate and oropharyngeal neoplasms

Laurence J. Howe[1,2,3] | **Gibran Hemani**[1] | **Corina Lesseur**[4] | **Valérie Gaborieau**[4] | **Kerstin U. Ludwig**[5] | **Elisabeth Mangold**[5] | **Paul Brennan**[4] | **Andy R. Ness**[6,7] | **Beate St Pourcain**[1,3] | **George Davey Smith**[1] | **Sarah J. Lewis**[1]

[1]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, UK

[2]Institute of Cardiovascular Science, University College London, London, UK

[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[4]Section of Genetics, International Agency for Research on Cancer, Lyon, France

[5]Institute of Human Genetics, University of Bonn, Bonn, Germany

[6]NIHR Bristol Biomedical Research Centre, University Hospitals Bristol, Bristol, UK

[7]Weston NHS Foundation Trust, University of Bristol, Bristol, UK

Correspondence
Laurence J. Howe and Sarah J. Lewis, Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, United Kingdom.
Email: Laurence.Howe@bristol.ac.uk (L. J. H.) and s.j.lewis@bristol.ac.uk (S. J. L.)

Funding information
Wellcome Trust, Grant/Award Number: 208806/Z/17/Z; Medical Research Council, Grant/Award Number: MC_UU_00011/1; Cancer Research UK

Abstract

It has been hypothesised that nonsyndromic cleft lip/palate (nsCL/P) and cancer may share aetiological risk factors. Population studies have found inconsistent evidence for increased incidence of cancer in nsCL/P cases, but several genes (e.g., *CDH1*, *AXIN2*) have been implicated in the aetiologies of both phenotypes. We aimed to evaluate shared genetic aetiology between nsCL/P and oral cavity/oropharyngeal cancers (OC/OPC), which affect similar anatomical regions. Using a primary sample of 5,048 OC/OPC cases and 5,450 controls of European ancestry and a replication sample of 750 cases and 336,319 controls from UK Biobank, we estimate genetic overlap using nsCL/P polygenic risk scores (PRS) with Mendelian randomization analyses performed to evaluate potential causal mechanisms. In the primary sample, we found strong evidence for an association between a nsCL/P PRS and increased odds of OC/OPC (per standard deviation increase in score, odds ratio [OR]: 1.09; 95% confidence interval [CI]: 1.04, 1.13; $p = .000053$). Although confidence intervals overlapped with the primary estimate, we did not find confirmatory evidence of an association between the PRS and OC/OPC in UK Biobank (OR 1.02; 95% CI: 0.95, 1.10; $p = .55$). Mendelian randomization analyses provided evidence that major nsCL/P risk variants are unlikely to influence OC/OPC. Our findings suggest possible shared genetic influences on nsCL/P and OC/OPC.

KEYWORDS

birth defects, genetic epidemiology, Mendelian randomization, oral cancers, orofacial clefts, polygenic risk scores

1 | INTRODUCTION

Nonsyndromic cleft lip/palate (nsCL/P) is a birth defect characterised by lack of fusion of structures in the upper lip and palate, with a complex aetiology influenced by both genetic and environmental factors (Dixon, Marazita, Beaty, & Murray, 2011; Mossey, Little, Munger, Dixon, & Shaw, 2009). Evidence from epidemiological population-based studies that individuals with birth defects have increased incidence of cancer (Bjørge, Cnattingius, Lie, Tretli, & Engeland, 2008; Carozza, Langlois, Miller, & Canfield, 2012; Fisher et al., 2012; Johnson et al., 2017) suggests that birth defects, including nsCL/P, may have shared aetiology with cancer. However, for nsCL/P specifically, the evidence for increased incidence of cancers among cases and unaffected first degree relatives is largely inconsistent, with findings limited by available sample sizes (Bille et al., 2005; Bjørge et al., 2008; Carozza et al., 2012; Johnson et al., 2017; Vieira, Khaliq, & Lace, 2012; Zhu et al., 2002).

There are several practical limitations of comparing cancer incidence in nsCL/P cases with incidence in the general population. First, the co-occurrence of nsCL/P and cancer is relatively modest because of the modest prevalence of nsCL/P (around 1 in 700; Mossey et al., 2009), and the two phenotypes typically differ in timing of onset. Second, cancers are highly heterogeneous across subtypes (e.g., childhood leukaemia and colorectal cancer), are highly distinct and stratifying by cancer subtype would further reduce statistical power. Third, syndromic and nonsyndromic forms of CL/P have different aetiologies and can be difficult to distinguish, suggesting that combining them together could be problematic if they have differences in cancer risk.

An alternative approach to evaluating the co-occurrence of nsCL/P and cancer is to explore shared genetic risk factors. Previous studies have found several genes implicated in both nsCL/P and cancer (Dunkhase et al., 2016), notably *CDH1* (Hozyasz et al., 2014; Machado et al., 2017; Vogelaar et al., 2012) which is linked to gastric and breast cancer (Pharoah, Guilford, & Caldas, 2001), and *AXIN2* (Letra, Menezes, Granjeiro, & Vieira, 2009; Letra et al., 2012; Machado et al., 2017; Mostowska et al., 2012) which is associated with colorectal cancer and tooth agenesis (Liu et al., 2000; Lammi et al., 2004). These findings suggest that common biological pathways may influence nsCL/P and cancer risk, but the extent to which the two phenotypes share genetic risk factors is currently unclear.

Two methods for evaluating shared genetic risk factors between two traits are polygenic risk scores (PRS) and Mendelian randomization (Davey Smith & Ebrahim, 2003). PRS are scores consisting of multiple genetic variants associated with a phenotype that can be used to evaluate genetic overlap between two traits (Dudbridge, 2013). Similarly, Mendelian randomization uses genetic variants robustly associated with a trait, typically at genome-wide significance, in an instrumental variable framework to evaluate possible causal relationships (Davey Smith & Hemani, 2014). In this instance, Mendelian randomization can be applied to test the possibility that common nsCL/P genetic variants, a latent measure of an individual's underlying liability to nsCL/P, influence cancer risk. Previous studies have demonstrated the utility of nsCL/P PRS and Mendelian randomization for disentangling mechanistic relationships (Dardani et al., 2020; Howe et al., 2018, 2019). For example, these approaches were used to provide evidence of shared genetic influences between nsCL/P and facial morphology (Howe et al., 2018).

A group of cancers that are strong candidates for shared genetic aetiology with nsCL/P are cancers of the oral cavity and oropharynx (OC/OPC) because of the similar anatomical sites affected. Major risk factors for OC/OPC include alcohol consumption, tobacco use and human papillomavirus infection (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2007) but OC/OPC also has a substantial heritable component (Lesseur et al., 2016). The possibility of shared genetic risk factors between nsCL/P and OC/OPC has not been previously investigated, possibly because of the relative rarity of both phenotypes.

Here, we first constructed nsCL/P PRS at a number of different thresholds using nsCL/P genome-wide association study (GWAS) summary data. We then evaluated the association between the nsCL/P PRSs and OC/OPC using individual participant data on OC/OPC cases and controls using data from the largest OC/OPC GWAS (Lesseur et al., 2016) and UK Biobank. To differentiate between the two datasets, we refer to the samples as the International Agency for Research on Cancer (IARC) and UK Biobank samples, respectively. We then applied Mendelian randomization to both data sets to evaluate a potential causal effect of liability to nsCL/P, a latent measure of nsCL/P proxied by common nsCL/P variants, and OC/OPC. Finally, we used data from UK Biobank to explore associations between nsCL/P PRS and potential shared risk factors between nsCL/P and OC/OPC (alcohol/tobacco measures).

2 | METHODS

2.1 | Data sources

2.1.1 | Summary data from nsCL/P GWAS meta-analysis

We used data from two nsCL/P GWAS, which have previously been meta-analysed and published

(Beaty et al., 2010; Ludwig et al., 2012; Mangold et al., 2010). The combined summary statistics were not publicly available, so meta-analysis summary statistics were reconstructed by the authors, this has been described in detail previously (Howe et al., 2018, 2019). In brief, a Transmission Disequilibrium Test (TDT) was implemented in a genome wide association study of 638 parent-offspring trios and 178 offspring duos of European descent (Beaty et al., 2010). The TDT results were then meta-analysed with GWAS summary results on 399 cases and 1,318 controls from the Bonn-II study, also of European descent (Mangold et al., 2010).

2.1.2 | IARC OC and OPC cancer GWAS data set

In this study, we used a data set of OC/OPC cases and controls which were part of a previous OC/OPC GWAS (Lesseur et al., 2016). In brief, the data set includes 6,034 OC, oropharyngeal or hypopharyngeal cancer cases and 6,585 controls. More information on the study design, genotyping, and phenotyping is contained in the Supporting Information Methods.

For analyses, the data set was restricted to 5,048 cases and 5,450 controls of recent European ancestry (confirmed by principal components analysis), which were split into two subsamples, based on the continent of the study centre (North America and Europe).

2.1.3 | UK Biobank

UK Biobank is a large-scale cohort study of 502,655 participants aged between 40 and 69 years, who were recruited from 22 recruitment centres across the United Kingdom between 2006 and 2010. In UK Biobank, we identified OC/OPC cases and controls using secondary care data from Hospital Episode Statistics (HES) as well as the death and cancer registers. For secondary analyses, we used data on self-reported alcohol consumption and tobacco smoking which were collected at baseline using a questionnaire. More information on the study design, genotyping and phenotyping is contained in the Supporting Information Methods.

For analyses, we used a subset of the study of 750 OC/OPC cases and 336,319 controls, after restricting to individuals of self-reported "White British" descent and using kinship coefficients to remove individuals related to the greatest number of other individuals.

2.2 | Statistical analyses

2.2.1 | The association of nsCL/P PRS with OC/OPC in the IARC sample

nsCL/P PRS were defined using the nsCL/P meta-analysis GWAS summary statistics. The summary statistics were LD clumped ($r^2 < .1$ and 250 kb) at 11 different p value inclusion thresholds (.000001, .000005, .00001, 0.00004, .0001, .0005, .001, .005, .01, .05, .1) to generate sets of independent (in terms of LD) variants in each score. The PRS was then constructed using effect alleles and weightings based on the magnitude of effect on nsCL/P of each genetic variant taken from the meta-analysis GWAS summary statistics. LD clumping was performed using PLINK (Purcell et al., 2007) with the 1000 Genomes (Phase 3; Genomes Project Consortium, 2015) CEU samples used as the reference panel.

Next, the 11 different nsCL/P PRS were constructed separately in the European and North American OC/OPC case–control sample using PLINK (Purcell et al., 2007) with an individual's PRS defined as the sum of weights across all variants in the PRS of the nsCL/P multiplied by the number of alleles. The associations between the nsCL/P PRS and OC/OPC case–control status were then estimated in the two subsamples using logistic regression adjusting for the first 10 genetic principal components, sex and age. Analyses were run separately for; all cases, OC cases only and oropharyngeal cases only (OPC). As a sensitivity analysis, we additionally removed cases and controls with less than 70% CEU ancestry. The effect sizes, standard errors and p values from the European and North American subsamples were meta-analysed using a fixed-effects model.

2.2.2 | The association of nsCL/P PRS with OC/OPC in UK Biobank

The UK Biobank data set was used as a follow-up data set for analyses in the OC/OPC GWAS data set. For the analyses in UK Biobank, we used the nsCL/P PRS most strongly associated with OC/OPC in the OC/OPC GWAS data set, tested the association of this PRS with OC/OPC. The limited number of OC/OPC cases ($N = 750$) in our sample meant that it was not possible to stratify by OC/OPC subtype, so all cases were analysed together. All analyses were adjusted for the first 10 genetic principal components, age, and sex.

2.2.3 | Mendelian randomization analysis of liability to nsCL/P on OC/OPC

Mendelian randomization was then applied to evaluate a causal relationship between liability to nsCL/P and OC/OPC. As genetic instruments for liability to nsCL/P, we used six genome-wide significant single nucleotide polymorphisms (SNPs), as in a previous study (Howe et al., 2018; Ludwig et al., 2012; Table S2). Chosen variants have large effects on nsCL/P risk (Ludwig et al., 2012) suggesting that they can be utilised as effective genetic instruments. Although recent nsCL/P GWAS have identified additional genome-wide significant markers in European populations (Leslie et al., 2017; Ludwig et al., 2016), we were limited by available summary data. SNP data for liability to nsCL/P were extracted from the nsCL/P meta-analysis GWAS summary statistics.

SNP data for OC/OPC were extracted from a meta-analysis of the IARC GWAS sample (European and North American subsamples) and UK Biobank (Lesseur et al., 2016). The GWAS in the IARC sample has been previously described in detail (Lesseur et al., 2016). In UK Biobank, the GWAS of 750 cases and 336,319 controls was conducted using logistic regression in PLINK v2.00 (Purcell et al., 2007). Sex, age, and the first 10 principal components were fitted as covariates in the model. The GWAS summary data from all data sets were meta-analysed using METAL (Willer, Li, & Abecasis, 2010) to generate combined summary statistics for OC/OPC.

We then used the TwoSampleMR (Hemani et al., 2018) package in R to conduct Mendelian randomization analyses; reporting the inverse-variance weighted estimate as the primary analysis. We also used several sensitivity analyses; MR Egger, weighted median, weighted mode and Cochran's *Q* heterogeneity test (Hemani, Bowden, & Davey Smith, 2018).

2.2.4 | nsCL/P PRS, alcohol consumption, and tobacco smoking

We tested the nsCL/P PRS (including SNPs with $p < 0.1$) for association with alcoholic units consumed per week and tobacco smoking pack years. All analyses were adjusted for sex, age and the first 10 genetic principal components.

3 | RESULTS

3.1 | nsCL/P PRS and risk of OC/OPC in the IARC sample

We found strong evidence for an association between nsCL/P PRS and increased risk of OC/OPC in the IARC sample of 5,048 cases and 5,450 controls. The strongest association was for the nsCL/P PRS including SNPs with $p < .1$, where a 1 standard deviation (*SD*) increase in nsCL/P PRS was associated with increased odds of PC/OPC (odds ratio [OR]: 1.09; 95% confidence interval [CI]: 1.04, 1.13; $p = .000053$). PRS with more liberal inclusion thresholds (e.g., $p < .05$ and $p < .1$), which included thousands of SNPs, were more strongly associated with risk of OC/OPC than more conservative inclusion thresholds (Figure 1 and Table 1).

Similarly, there was strong evidence for an association between the nsCL/P PRS and the primary OC/OPC subtypes, oropharyngeal and OC cancer. A 1 *SD* increase in nsCL/P PRS ($p < 0.1$) was associated with increased odds of both oropharyngeal cancer (OR: 1.10; 95% CI: 1.04, 1.16; $p = .00079$) and OC cancer (OR: 1.12; 95% CI: 1.06, 1.17; $p = .000016$; Table S1).

3.2 | nsCL/P PRS and risk of OC/OPC in UK Biobank

In the UK Biobank sample of 750 cases and 336,319 controls, we attempted to replicate the association between the nsCL/P PRS most strongly associated with OC/OPC in the IARC sample (p value inclusion threshold $< .1$). Here, we found no clear evidence that this PRS is associated with increased risk of OC/OPC (OR: 1.02; 95% CI: 0.95, 1.10; $p = .55$). However, confidence intervals overlapped with the IARC sample estimate (OR: 1.09 95% CI: 1.04, 1.13) and the meta-analysis of the two estimates suggested evidence for an association (OR: 1.07; 95% CI: 1.03, 1.11; $p = .0009$).

3.3 | Mendelian randomization: Liability to nsCL/P and risk of OC/OPC

Using combined GWAS summary data from the IARC and UK Biobank samples, 5,798 cases and 341,769 controls, we found little evidence for an effect of liability to nsCL/P on OC/OPC; a 1-unit log odd increase in liability to nsCL/P was not strongly associated with OC/OPC (MR IVW: OR: 0.99; 95% CI: 0.95, 1.05; $p = .84$). Mendelian randomization sensitivity analyses generated consistent effect estimates (Table 2).

3.4 | nsCL/P PRS, alcohol consumption and tobacco smoking

To evaluate whether shared genetic effects could relate to interactions of the nsCL/P PRS with shared

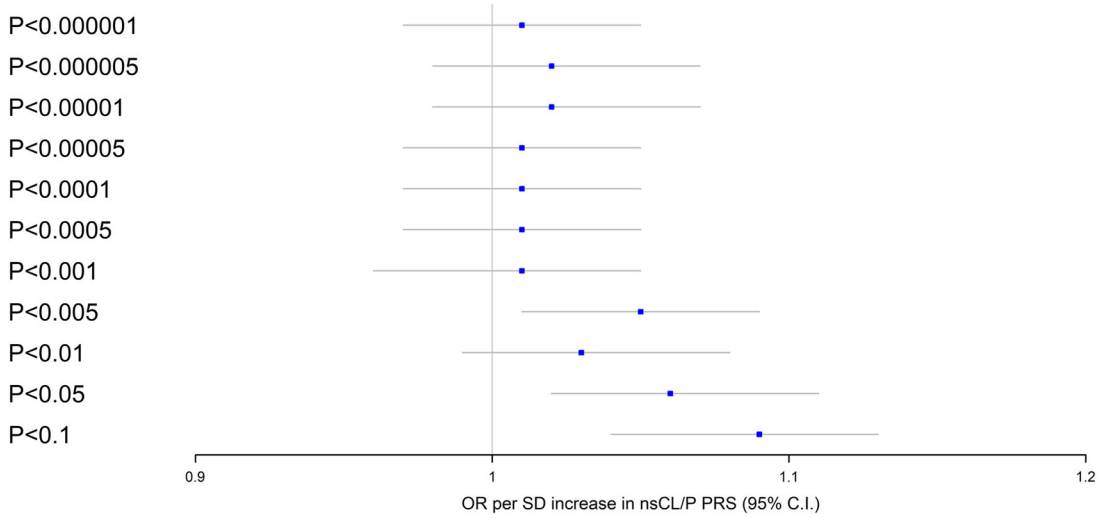

FIGURE 1 Association of nsCL/P PRS with OC/OPC in the International Agency for Research on Cancer sample. A forest plot showing the associations between different nsCL/P PRS, ranging from 10 SNPs to over 10,000 SNPs, and OC/OPC risk. nsCL/P, nonsyndromic cleft lip/palate; OC, oral cavity; OPC, oropharynx; PRS, polygenic risk score; SNP, single nucleotide polymorphism

environmental risk factors we investigated the genetic overlap between the PRS and measures of alcohol consumption and tobacco smoking in UK Biobank. Again, we used the nsCL/P PRS most strongly associated with OC/OPC in the IARC sample ($p < 0.1$). We did not find strong evidence that the nsCL/P PRS was associated with any of the measures tested. A 1 *SD* increase in nsCL/P PRS corresponded to a −0.02 pack year decrease in lifetime smoking (95% CI: −0.07, 0.03; $p = .46$; $N = 284{,}144$) and a −0.05 units per week decrease in alcohol use (95% CI: −0.11, 0.01; $p = .11$; $N = 336{,}026$).

TABLE 1 nsCL/P PRS and risk of OC/OPC

Polygenic risk score *p* value inclusion threshold	Number of SNPs in PRS	All OC/OPC subtypes against controls (5,048 cases and 5,450 controls)	
		OR (95% CI) per 1 *SD* increase in PRS	*p* Value
.000001	10	1.01 (0.97, 1.05)	.64
.000005	15	1.02 (0.98, 1.07)	.27
.00001	18	1.02 (0.98, 1.07)	.27
.00005	48	1.01 (0.97, 1.05)	.71
.0001	78	1.01 (0.97, 1.05)	.58
.0005	238	1.01 (0.97, 1.05)	.71
.001	424	1.01 (0.96, 1.05)	.79
.005	1,607	1.05 (1.01, 1.09)	.021
.01	2,777	1.03 (0.99, 1.08)	.10
.05	8,620	1.06 (1.02, 1.11)	.0026
.1	12,614	1.09 (1.04, 1.13)	.000053

Abbreviations: CI, confidence interval; nsCL/P, nonsyndromic cleft lip/palate; OC, oral cavity; OPC, oropharynx; PRS, polygenic risk score; *SD*, standard deviation; SNP, single nucleotide polymorphism.

4 | DISCUSSION

In this study, we found some evidence to suggest that nsCL/P PRS are associated with modestly increased risk of OC/OPC. Follow-up Mendelian randomization analyses using well established nsCL/P risk SNPs as a latent measure of nsCL/P suggested no consistent association between key common nsCL/P variants and OC/OPC. These findings are consistent with our initial hypothesis that nsCL/P and OC/OPC share some genetic risk factors as opposed to the possibility of a causal relationship, which has previously been demonstrated for nsCL/P and philtrum width (Howe et al., 2018).

The specific shared genetic influences and relevant biological pathways that may have induced the association between the nsCL/P PRS and OC/OPC are currently unclear. While the Mendelian randomization approach includes only replicated variants with strong associations, the PRS approach is much more liberal and will include many variants with weaker associations. The increased variation explained by the PRS approach comes with the caveat that many variants of small effects will be included which have less clear direct relevant to nsCL/P. The interpretation of what a PRS is proxying for becomes increasingly complex as the number of SNPs included in the

TABLE 2 Mendelian randomization analysis of liability to nsCL/P on OC/OPC risk

Test	Interpretation	OR (95% CI)	*p* Value
Inverse variance weighted	Primary result[a]	0.99 (0.95, 1.05)	.84
Heterogeneity of inverse variance weighted	Balanced pleiotropy	N/A	.59
MR-Egger	Intercept test for directional pleiotropy[b]	0.002 (−0.058, 0.063)	.94
	Regression estimate[a]	0.99 (0.87, 1.13)	.69
Weighted median	Consistency[a]	0.99 (0.93, 1.06)	.77
Weighted mode	Consistency[a]	0.99 (0.92, 1.06)	.71

Abbreviations: CI, confidence interval; nsCL/P, nonsyndromic cleft lip/palate; OC, oral cavity; OPC, oropharynx; OR, odds ratio.

[a]Units: odds ratio for OC/OPC perper 1-unit log odd increase in liability to nsCL/PnsCL/P.

[b]Units: average pleiotropic effect of a nsCL/P genetic variant on odds of OC/OPC.

score increases. Indeed, the nsCL/P PRS most strongly associated with OC/OPC included over 10,000 SNPs. One possibility is that the genetic overlap may be attributable to certain genes being involved both in early development and tumour suppression. For example, the *CDH1* gene, thought to be related to nsCL/P and cancer subtypes, has been shown to be related to both axonal growth and patterning in the developing murine brain (Konishi, Stegmüller, Matsuda, Bonni, & Bonni, 2004), and tumour suppression (Berx, Becker, Höfler, & Van Roy, 1998).

Another possibility is that the nsCL/P PRS may capture environmental influences relating to both nsCL/P and cancer risk. Maternal behaviour such as cigarette smoking and alcohol consumption have been hypothesised to influence nsCL/P risk and both are established risk factors for OC/OPC. Given that the maternal genotype is correlated with the offspring genotype, it is plausible that nsCL/P PRS could capture effects of maternal environmental factors. The foetal genotype may also play an important role in the effect of environmental exposures on risk of orofacial clefts. For example, there is some evidence that the maternal and foetal *ADH1C* haplotype may modify the association between maternal alcohol consumption and risk of orofacial clefts via alcohol metabolism (Boyles et al., 2010). We did not find strong evidence that nsCL/P PRS are associated with alcohol and cigarette use in UK Biobank but were unable to test pathways relating to metabolism.

The findings of this study are consistent with previous genetic studies that have found evidence of loci associated with both nsCL/P and various cancer subtypes (Dunkhase et al., 2016; Hozyasz et al., 2014; Letra et al., 2009, 2012; Machado et al., 2017; Mostowska et al., 2012; Vogelaar et al., 2012). Our results are also consistent with a previous study exploring adult-onset cancers in nsCL/P cases, which was underpowered to detect a modest effect (Bille et al., 2005).

This study is the first to investigate the genetic overlap between nsCL/P and OC/OPC. Previous epidemiological and genetic studies have explored the relationship between nsCL/P and all cancers, however, cancers arising from different organs may be aetiologically heterogeneous. Although similar results were found in analyses for the OC and OPC subtype analyses with these subtypes also aetiologically heterogeneous. In Western populations, smoking and alcohol are established risk factors for both OC and OPC, while HPV status is thought to be a risk factor for OPC only (Lesseur et al., 2016). Future work could evaluate genetic overlap between nsCL/P and other cancer subtypes to determine if the association is consistent across other cancer subtypes or localised specifically to oral tissues.

The use of a PRS as a genetic proxy for nsCL/P in our analyses is a considerable strength. PRS analyses have advantages over candidate gene or candidate SNP approaches used in previous studies (Dunkhase et al., 2016; Hozyasz et al., 2014; Letra et al., 2009, 2012; Machado et al., 2017; Mostowska et al., 2012; Vogelaar et al., 2012) because of the reduction in the number of statistical tests, and the potential to extend the evidence of genetic overlap at specific loci to examine genome-wide genetic overlap. Although the nsCL/P GWAS data set was modestly sized, the construction of nsCL/P PRS in a much larger OC/OPC GWAS data set meant that analyses were well-powered to detect a modest genetic overlap.

Nevertheless, there are several limitations of this study. First, the lack of a convincing replication of the PRS analysis in UK Biobank weakens the argument for genetic overlap between the two phenotypes. However, this may be because of the modest number of OC/OPC cases in the UK Biobank, with confidence intervals overlapping between the two studies. Second, the OC/OPC GWAS data set was highly ancestrally heterogeneous and included samples from 12 different

epidemiological studies, including a case-only study (the Head and Neck 5000). Therefore, it is possible that allele frequency differences between cases and controls relating to population differences could result in spurious associations with the nsCL/P PRS, although this is unlikely given the number of SNPs in the PRS. Third, with insufficient genetic instruments for OC/OPC, we did not perform a bidirectional Mendelian randomization analysis so cannot rule out the possibility that OC/OPC variants consistently affect nsCL/P risk. Fourth, we were unable to run stratified analysis investigating the possibility of heterogeneity between HPV and non-HPV driven tumours. Fifth, in this study we treated nsCL/P subtypes (cleft lip with cleft palate, cleft lip only) homogeneously despite growing evidence that they are aetiologically distinct (Leslie et al., 2017; Ludwig et al., 2016; Sharp et al., 2017). This complicates the interpretation of the results because the OPC includes the soft palate while the OC includes the lips and hard palate. It is therefore, possible that the different nsCL/P subtypes may have different mechanistic relationships with the OC/OPC subtypes. Finally, we were unable to explore the effects of somatic mutations in the oral tissues which could influence both orofacial cleft risk and localized cancer.

To conclude, we found some evidence of shared genetic influences on nsCL/P and OC/OPC unrelated to alcohol or tobacco intake. Follow-up analyses, potentially using additional datasets such as the Cleft Collective (www.bristol.ac.uk/dental/cleft-collective/), are required to investigate the possible common biological pathways between nsCL/P and OC/OPC, and to evaluate possible mechanistic relationships between nsCL/P and cancer subtypes affecting distinct areas to orofacial clefts.

ACKNOWLEDGEMENTS

We would like to thank all of the participants involved in this study as well as the funders and supporting organisations. L. J. H., G. H., G. D. S. and S. J. L. work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/1). S. J. L. is funded by the HEFCE and by a programme grant awarded by CRUK (ICEP, Principle Investigators: Martin and Relton), G. H. is funded by the Wellcome Trust (208806/Z/17/Z). Funding support for the nsCL/P ICC study was provided by several previous grants from the National Institute of Dental and Craniofacial Research (NIDCR). Funding for individual investigators include: R21-DE-013707 and R01-DE-014581 (Beaty); R37-DE-08559 and P50-DE-016215 (Murray, Marazita) and the Iowa Comprehensive Program to Investigate Craniofacial and Dental Anomalies (Murray); R01-DE-09886 (Marazita), R01-DE-012472 (Marazita), R01-DE-014677 (Marazita), R01-DE-016148 (Marazita), R21-DE016930 (Marazita); R01-DE-013939 (Scott). Parts of this study were supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Wilcox, Lie) and additional recruitment was supported by the Smile Train Foundation for recruitment in China (Jabs, Beaty, Shi) and a grant from the Korean government (Jee). The genome-wide association study, also known the Cleft Consortium, is part of the Gene Environment Association Studies (GENEVA) program of the trans-NIH Genes, Environment and Health Initiative (GEI) supported by U01-DE-018993. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University, contract number HHSN268200782096C. Funds for genotyping were provided by the NIDCR through CIDR's NIH contract. Assistance with genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01-HG-004446) and by the National Center for Biotechnology Information (NCBI). The Bonn nsCL/P GWAS was supported by the Deutsche Research Foundation (grant FOR 423, and individual grants MA 2546/3-1 to E.M, and LU 1944/3-1 to K.U.L. Genotyping for the OC/OPC GWAS data set was performed at the Center for Inherited Disease Research (CIDR), funded through the U.S. National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007780-0. Genotyping for shared controls with the Lung OncoArray initiative was funded through the grant X01HG007492-0. The University of Pittsburgh head and neck cancer case–control study is supported by National Institutes of Health grants P50 CA097190 and P30 CA047904. The Carolina Head and Neck Cancer Study (CHANCE) was supported by the National Cancer Institute (R01-CA90731). The HN5000 study was funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-300 0707-10034), the views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The Toronto study was funded by the Canadian Cancer Society Research Institute (020214) and the National Cancer Institute (U19 CA148127) and the Cancer Care Ontario Research Chair. The alcohol-related cancers and genetic susceptibility study in Europe (ARCAGE) was funded by the European Commission's 5th Framework Program (QLK1-306 2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte, and Padova University (CPDA057222). The Rome Study was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) IG 2011 10491 and IG2013 14220 to SB, and Fondazione Veronesi to SB.

The IARC Central Europe study was supported by European Commission's INCO-COPERNICUS Program (IC15-CT98-0332), NIH/National Cancer Institute grant CA92039, and the World Cancer Research Foundation grant WCRF 99A28. The IARC Oral Cancer Multicenter study was funded by: grant S06 96 202489 05F02 from Europe against Cancer; Grants FIS 97/0024, FIS 97/0662, and BAE 01/5013 from Fondo de Investigaciones Sanitarias, Spain; UICC Yamagiwa-Yoshida Memorial International Cancer Study; National Cancer Institute of Canada; Italian Association for Research on Cancer; and the Pan American Health Organization. The coordination of EPIC study is financially supported by the European Commission (DG SANCO) and the International Agency for Research on Cancer. UK Biobank received ethical approval from the Research Ethics Committee (11/NW/0383). This study was approved as part of application 15825 (PI: Dr Philip Haycock). The funders did not participate in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

AUTHOR CONTRIBUTIONS

L. J. H., A. R. N., and S. J. L. conceptualised and designed the project. L. J. H. and C. L. performed statistical analyses. A. R. N., E. M., K. L., V. G., and P. B. were involved in data acquisition. L. J. H. drafted the manuscript and interpreted results under the supervision of G. H., G. D. S., B. S. P., and S. J. L. All authors contributed to critically revising the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work. All authors declare no relevant conflicts of interest to the publication of this manuscript.

DATA AVAILABILITY STATEMENT

Research Data are not shareThe nsCL/P family data set used to construct summary data described in this manuscript were obtained from dbGaP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000094.v1.p1 through dbGaP accession number (phs000094.v1.p1). Individuals interested in accessing summary data from the nsCL/P Bonn GWAS are encouraged to contact Elisabeth Mangold (e.mangold@uni-bonn.de) or Kerstin Ludwig (kerstin.ludwig@uni-bonn.de). The OC/OPC GWAS data is available from dbGaP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001202.v1.p1 through dbGaP accession number (phs001202.v1.p1). Summary data from the UK Biobank GWAS of OC/OPC, described in this manuscript, will be made available in the MR Base platform and the GWAS catalogue before publication.

ORCID

Laurence J. Howe http://orcid.org/0000-0002-2819-9686

REFERENCES

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., ... Fallin, M. D. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics*, *42*(6), 525–529.

Berx, G., Becker, K. F., Höfler, H., & Van Roy, F. (1998). Mutations of the human E-cadherin (CDH1) gene. *Human Mutation*, *12*(4), 226–237.

Bille, C., Winther, J. F., Bautz, A., Murray, J. C., Olsen, J., & Christensen, K. (2005). Cancer risk in persons with oral cleft—a population-based study of 8,093 cases. *American Journal of Epidemiology*, *161*(11), 1047–1055.

Bjørge, T., Cnattingius, S., Lie, R. T., Tretli, S., & Engeland, A. (2008). Cancer risk in children with birth defects and in their families: A population based cohort study of 5.2 million children from Norway and Sweden. *Cancer Epidemiology and Prevention Biomarkers*, *17*(3), 500–506.

Boyles, A. L., DeRoo, L. A., Lie, R. T., Taylor, J. A., Jugessur, A., Murray, J. C., & Wilcox, A. J. (2010). Maternal alcohol consumption, alcohol metabolism genes, and the risk of oral clefts: A population-based case-control study in Norway, 1996–2001. *American Journal of Epidemiology*, *172*(8), 924–931.

Carozza, S. E., Langlois, P. H., Miller, E. A., & Canfield, M. (2012). Are children with birth defects at higher risk of childhood cancers?*American Journal of Epidemiology*, *175*(12), 1217–1224.

Dardani, C., Howe, L. J., Mukhopadhyay, N., Stergiakouli, E., Wren, Y., Humphries, K., ... Sharp, G. C. (2020). Cleft lip/palate and educational attainment: Cause, consequence or correlation? A Mendelian-randomization study. *International Journal of Epidemiology*, dyaa047. https://doi.org/10.1093/ije/dyaa047

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?*International Journal of Epidemiology*, *32*(1), 1–22.

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98.

Dixon, M. J., Marazita, M. L., Beaty, T. H., & Murray, J. C. (2011). Cleft lip and palate: Understanding genetic and environmental influences. *Nature Reviews Genetics*, *12*(3), 167–178.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics*, *9*(3), e1003348.

Dunkhase, E., Ludwig, K. U., Knapp, M., Skibola, C. F., Figueiredo, J. C., Hosking, F. J., ... Nakagawa, H. (2016). Nonsyndromic cleft lip with or without cleft palate and cancer: Evaluation of a possible common genetic background through the analysis of GWAS data. *Genomics Data*, *10*, 22–29.

Fisher, P. G., Reynolds, P., Von Behren, J., Carmichael, S. L., Rasmussen, S. A., & Shaw, G. M. (2012). Cancer in children with nonchromosomal birth defects. *The Journal of Pediatrics*, *160*(6), 978–983.

Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Hemani, G., Bowden, J., & Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*, *27*(R2), R195–R208. https://doi.org/10.1093/hmg/ddy163

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., ... Yarmolinsky, J. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, *7*, 078972.

Howe, L. J., Lee, M. K., Sharp, G. C., Davey Smith, G., St Pourcain, B., Shaffer, J. R., ... Stergiakouli, E. (2018). Investigating the shared genetics of non-syndromic cleft lip/palate and facial morphology. *PLOS Genetics*, *14*, 255901.

Howe, L. J., Richardson, T. G., Arathimos, R., Alvizi, L., Passos-Bueno, M. R., Stanier, P., ... Sharp, G. C. (2019). Evidence for DNA methylation mediating genetic liability to non-syndromic cleft lip/palate. *Epigenomics*, *11*(2), 133–145. https://doi.org/10.2217/epi-2018-0091

Hozyasz, K. K., Mostowska, A., Wójcicki, P., Lasota, A., Offert, B., Balcerek, A., ... Jagodziński, P. P. (2014). Nucleotide variants of the cancer predisposing gene CDH1 and the risk of non-syndromic cleft lip with or without cleft palate. *Familial cancer*, *13*(3), 415–421.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. (2007). *Human papilloma viruses*. World Health Organization.

Johnson, K. J., Lee, J. M., Ahsan, K., Padda, H., Feng, Q., Partap, S., ... Druley, T. E. (2017). Pediatric cancer risk in association with birth defects: A systematic review. *PLOS One*, *12*(7), e0181246.

Konishi, Y., Stegmüller, J., Matsuda, T., Bonni, S., & Bonni, A. (2004). Cdh1-APC controls axonal growth and patterning in the mammalian brain. *Science*, *303*(5660), 1026–1030.

Lammi, L., Arte, S., Somer, M., Järvinen, H., Lahermo, P., Thesleff, I., ... Nieminen, P. (2004). Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *The American Journal of Human Genetics*, *74*(5), 1043–1050.

Leslie, E. J., Carlson, J. C., Shaffer, J. R., Butali, A., Buxó, C. J., Castilla, E. E., ... Hecht, J. T. (2017). Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Human Genetics*, *136*(3), 275–286.

Lesseur, C., Diergaarde, B., Olshan, A. F., Wünsch-Filho, V., Ness, A. R., Liu, G., ... Lagiou, P. (2016). Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nature Genetics*, *48*(12), 1544–1550.

Letra, A., Bjork, B., Cooper, M., Szabo-Rogers, H., Deleyiannis, F., Field, L., ... Poletta, F. (2012). Association of AXIN2 with non-syndromic oral clefts in multiple populations. *Journal of Dental Research*, *91*(5), 473–478.

Letra, A., Menezes, R., Granjeiro, J. M., & Vieira, A. R. (2009). AXIN2 and CDH1 polymorphisms, tooth agenesis, and oral clefts. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *85*(2), 169–173.

Liu, W., Dong, X., Mai, M., Seelan, R. S., Taniguchi, K., Krishnadath, K. K., ... Christensen, E. (2000). Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. *Nature Genetics*, *26*(2), 146–147.

Ludwig, K. U., Ahmed, S. T., Böhmer, A. C., Sangani, N. B., Varghese, S., Klamt, J., ... Rubini, M. (2016). Meta-analysis reveals genome-wide significance at 15q13 for nonsyndromic clefting of both the lip and the palate, and functional analyses implicate GREM1 as a plausible causative gene. *PLOS Genetics*, *12*(3), e1005914.

Ludwig, K. U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., ... Nasser, E. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nature Genetics*, *44*(9), 968–971.

Machado, R. A., deFreitas, E. M., de Aquino, S. N., Martelli, D. R. B., Swerts, M. S. O., de Almeida Reis, S. R., ... Coletta, R. D. (2017). Clinical relevance of breast and gastric cancer-associated polymorphisms as potential susceptibility markers for oral clefts in the Brazilian population. *BMC Medical Genetics*, *18*(1), 39.

Mangold, E., Ludwig, K. U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., ... Mattheisen, M. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nature Genetics*, *42*(1), 24–26.

Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J., & Shaw, W. C. (2009). Cleft lip and palate. *The Lancet*, *374*(9703), 1773–1785.

Mostowska, A., Hozyasz, K. K., Wójcicki, P., Lasota, A., Dunin-Wilczyńska, I., & Jagodziński, P. P. (2012). Association of DVL2 and AXIN2 gene polymorphisms with cleft lip with or without cleft palate in a Polish population. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *94*(11), 943–950.

Pharoah, P. D., Guilford, P., & Caldas, C. (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*, *121*(6), 1348–1353.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Sharp, G. C., Ho, K., Davies, A., Stergiakouli, E., Humphries, K., McArdle, W., ... Relton, C. L. (2017). Distinct DNA methylation profiles in subtypes of orofacial cleft. *Clinical Epigenetics*, *9*(1), 63.

Vieira, A. R., Khaliq, S., & Lace, B. (2012). Risk of cancer in relatives of children born with isolated cleft lip and palate. *American Journal of Medical Genetics, Part A*, *158*(6), 1503–1504.

Vogelaar, I. P., Figueiredo, J., vanRooij, I. A., Simões-Correia, J., van der Post, R. S., Melo, S., ... Hoogerbrugge, N. (2012). Identification of germline mutations in the cancer predisposing gene CDH1 in patients with orofacial clefts. *Human Molecular Genetics*, *22*(5), 919–926.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191.

Zhu, J. L., Basso, O., Hasle, H., Winther, J., Olsen, J., & Olsen, J. (2002). Do parents of children with congenital malformations have a higher cancer risk? A nationwide study in Denmark. *British Journal of Cancer*, *87*(5), 524–528.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Howe LJ, Hemani G, Lesseur C, et al. Evaluating shared genetic influences on nonsyndromic cleft lip/palate and oropharyngeal neoplasms. *Genetic Epidemiology*. 2020;44:924–933. https://doi.org/10.1002/gepi.22343