



RESEARCH ARTICLE

REVISED 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 [version 2; peer review: 2 approved]

Øystein Ariansen Haaland ¹, Julia Romanowska^{1,2}, Miriam Gjerdevik^{1,3}, Rolv Terje Lie^{1,4}, Håkon Kristian Gjessing ^{1,4}, Astanand Jugessur^{1,3,4}

¹Department of Global Public Health and Primary Care, University of Bergen, Bergen, N-5020, Norway

²Computational Biology Unit, University of Bergen, Bergen, N-5020, Norway

³Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Skøyen, Oslo, Skøyen, N-0213, Norway

⁴Centre for Fertility and Health (CeFH), Norwegian Institute of Public Health, Skøyen, Oslo, N-0213, Norway

v2 First published: 24 Jun 2019, 8:960 (<https://doi.org/10.12688/f1000research.19571.1>)
 Latest published: 19 Jul 2019, 8:960 (<https://doi.org/10.12688/f1000research.19571.2>)

Abstract

Background: Although both genetic and environmental factors have been reported to influence the risk of isolated cleft lip with or without cleft palate (CL/P), the exact mechanisms behind CL/P are still largely unaccounted for. We recently developed new methods to identify parent-of-origin (PoO) interactions with environmental exposures (PoOxE) and now apply them to data from a genome-wide association study (GWAS) of families with children born with isolated CL/P.

Methods: Genotypes from 1594 complete triads and 314 dyads (1908 nuclear families in total) with CL/P were available for the current analyses. Of these families, 1024 were Asian, 825 were European and 59 had other ancestries. After quality control, 341,191 SNPs remained from the original 569,244. The exposures were maternal cigarette smoking, use of alcohol, and use of vitamin supplements in the periconceptional period. Our new methodology detects if PoO effects are different across environmental strata and is implemented in the *R*-package Haplin.

Results: Among Europeans, there was evidence of a PoOxSmoke effect for *ANK3* with three SNPs (rs3793861, $q=0.20$, $p=2.6e-6$; rs7087489, $q=0.20$, $p=3.1e-6$; rs4310561, $q=0.67$, $p=4.0e-5$) and a PoOxAlcohol effect for *ARHGEF10* with two SNPs (rs2294035, $q=0.32$, $p=2.9e-6$; rs4876274, $q=0.76$, $p=1.3e-5$).

Conclusion: Our results indicate that the detected PoOxE effects have a plausible biological basis, and thus warrant replication in other independent cleft samples. Our demonstration of the feasibility of identifying complex interactions between relevant environmental exposures and PoO effects offers new avenues for future research aimed at unravelling the complex etiology of cleft lip defects.

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
REVISED version 2 published 19 Jul 2019	 report	
	↑	
version 1 published 24 Jun 2019	? report	 report

- 1 **Evie Stergiakouli** , University of Bristol, Bristol, UK
MRC Integrative Epidemiology at the University of Bristol, Bristol, UK
- 2 **Yongchu Pan** , Nanjing Medical University (NMU), Nanjing, China

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Orofacial cleft, cleft lip with or without cleft palate, case-parent triads, gene-environment interaction, parent-of-origin, PoOxE, Haplin

Corresponding author: Øystein Ariansen Haaland (oystein.haaland@uib.no)

Author roles: **Haaland ØA:** Conceptualization, Data Curation, Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Romanowska J:** Conceptualization, Formal Analysis, Methodology, Software, Visualization, Writing – Review & Editing; **Gjerdevik M:** Conceptualization, Formal Analysis, Methodology, Software, Visualization, Writing – Review & Editing; **Lie RT:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Writing – Review & Editing; **Gjessing HK:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Software, Writing – Review & Editing; **Jugessur A:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was supported by the Bergen Medical Research Foundation [807191], by the Research Council of Norway (RCN) through its Centres of Excellence funding scheme [grant 262700], and by the Biobank Norway II from the RCN [245464]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2019 Haaland ØA *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Haaland ØA, Romanowska J, Gjerdevik M *et al.* **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 [version 2; peer review: 2 approved]** F1000Research 2019, 8:960 (<https://doi.org/10.12688/f1000research.19571.2>)

First published: 24 Jun 2019, 8:960 (<https://doi.org/10.12688/f1000research.19571.1>)

REVISED Amendments from Version 1

In this version we have considered the remarks of Dr. Evie Stergiakouli. We have addressed some limitations that were pointed out and generally edited the manuscript for clarity.

Abstract:

Removed reference to previously published work and added information about the methodology.

Introduction:

Rephrased sentence that overemphasized how intriguing the problem of missing heritability is in clefting.

Methods:

Clarified that 59 individuals of other ethnicities than Asian or European were included in the pooled sample and added a sentence about the American Statistical Association's negative attitudes towards using a fixed p-value as a threshold for statistical significance.

Results:

Revised titles of Table 3, Table 4 and Table 7 to include information about which analyses were conducted.

Discussion:

Added a paragraph about genomic imprinting as a possible mechanism for PoO effects, clarified that our findings should be treated with caution until they have been replicated, and addressed the limitation that we have not imputed genotypes.

In general:

Corrected minor errors and slightly altered text to improve flow and precision.

See referee reports

Introduction

Cleft lip with or without cleft palate (CL/P) appears in approximately 3.4 to 22.9 per 10,000 live births¹. Based on the severity of the cleft, patients undergo varying degrees of medical, dental, speech and psychosocial interventions over the first two decades of their lives, a long-term multidisciplinary treatment that not only imposes a heavy burden on patients and their families^{2,3}, but also accounts for a substantial outlay in national healthcare budgets^{4,5}.

Multiple genetic and environmental factors have been reported to influence the risk of CL/P, individually and through complex interactions in relevant biological pathways⁶⁻¹⁰. Major advances in high-throughput genotyping technologies, coupled with a boost in international collaborations, have led to substantial progress in gene-mapping for orofacial clefts, and the first wave of genome-wide association studies (GWAS) identified and replicated several key genes and loci associated with clefting¹¹⁻¹⁶. Despite this success, the genetic variants identified so far collectively explain only a minor fraction of the total variance attributable to additive genetic effects, even though the heritability of CL/P among Europeans is more than 70%¹⁷⁻²⁰. This has spurred renewed interest in investigating disease mechanisms other than fetal or maternal effects²¹. One example is parent-of-origin (PoO), where the effect of a particular allele in the offspring differs according to its parental origin²²⁻²⁴, and another is gene-environment interaction (GxE), where fetal effects differ across strata of environmental exposures²⁵. Identifying GxE effects may not only provide new insights into the causes of CL/P, but

may also provide an opportunity to intervene on environmental risk factors alone, particularly in subgroups of the population that are genetically more susceptible to these environmental effects.

Recently, we went one step further and developed new methods for a genome-wide screening for PoO interactions with environmental exposures (i.e., PoOxE) in the case-parent triad setting²². We applied the new methodology, implemented in the R-package Haplin²⁶, to isolated cleft palate only (CPO)²⁷, using genotypes and exposure data from the largest published GWAS dataset on case-parent triads of orofacial clefts¹¹. Epidemiological and embryological findings have previously shown that CL/P and CPO may have distinct etiologies. Therefore, we used the same GWAS dataset and methodology to perform a genome-wide scan for PoOxE effects in the larger sample of isolated CL/P.

Methods

Study participants

The study participants were mainly of Asian or European origin and were recruited as part of an international cleft collaboration¹¹. Information was available on genotypes as well as maternal vitamin use, cigarette smoking and alcohol consumption in the periconceptional period (three months before and three months after pregnancy). The information on environmental exposures was based on interviews and questionnaires. More detailed characteristics of the study participants can be found in our recent work²⁸.

Table 1 shows the distribution of the CL/P families according to ethnicity, triad completeness and maternal exposure. There were 1908 families in the pooled sample (5424 individuals in total), which included all the participants. Of these, 825 families were in the European sample, 1024 families were in the Asian sample, and 59 families were in the sample consisting of other ethnicities (Table 1). We performed three main sets of analyses on the following samples: All participants (denoted as "pooled analysis"), only Asians ("Asian analysis"), and only Europeans ("European analysis"). The 59 families with other ethnicities were not analyzed as a group due to the small sample size, but they were included in the pooled analysis. In the pooled and European analyses, we examined all exposures. As cigarette smoking and alcohol consumption were rare among Asian mothers, we were only able to conduct PoOxVitamin analyses for this ethnicity.

Quality control for excluding single-nucleotide polymorphisms (SNPs) and individuals were conducted as described in Haaland *et al.* (2017)²⁷. That is, we included SNPs with a missing call rate less than 5%, a minor allele frequency (MAF) greater than 5%, a p-value of less than 0.001 for the test for Hardy-Weinberg equilibrium presented by Wigginton *et al.* (2005)²⁹, and a Mendelian error rate greater than 10%. Further, if two or more SNPs were in perfect linkage disequilibrium ($r^2=1$) with each other, we only included one in the analyses. After applying these same criteria here, 341,191 were left for the current analyses from a total of 569,244 SNPs (Table 2).

Table 1. Number of isolated cleft lip with or without cleft palate families according to ethnicity, triad completeness and maternal exposure to alcohol, smoking, and vitamin.

Ethnicity	Complete + incomplete triads		Total		Mother exposed (missing)		
	Individuals	Families	Individuals	Families	Alcohol	Smoking	Vitamin
European	2024+310	670+155	2334	825	325 (8)	249 (6)	462 (98)
Asian ^a	2670+268	890+134	2938	1024	-	-	142 (155)
Other ^b	102+ 50	34+ 25	152	59	-	-	-
Pooled	4796+628	1594+314	5424	1908	350 (22)	284 (9)	638 (255)

^aNo analyses of parent-of-origin interactions with alcohol (PoOxAlcohol) or parent-of-origin interactions with smoking (PoOxSmoke) were conducted for this group because of a lack of observations for these exposures. ^bOwing to the small sample size, no analysis of parent-of-origin interactions with environmental exposures (PoOxE) was conducted for this group. Note that a subset of the complete triads included more than one offspring. Incomplete triads are parent-offspring dyads.

Table 2. Quality control.

Total number of single-nucleotide polymorphisms (SNPs)	569,244
Criteria:	
Failed HWE test ($p < 0.001$)	173,955
More than 5% missing calls	1934
MAF less than 5%	61,167
$r^2 = 1$ with flanking SNPs	2880
Mendelian errors detected ($> 1\%$)	349
Number of SNPs remaining after quality control ^a	341,191

^aSome SNPs failed several criteria. Hence, the remaining number of SNPs (341,191) plus the ones that failed the different criteria do not add up to the total number of SNPs (569,244). HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency.

Statistical analysis

For statistical analysis, we used the statistical software **Haplin**²⁶, which is written in the *R* statistical programming language³⁰. Haplin is based on log-linear modeling in a maximum likelihood framework and is well-suited for the analysis of offspring-parent triads. Because Haplin uses the expectation-maximization (EM) algorithm to account for missing parental genotypes²⁶, we were able to include the 314 case-parent dyads in the analyses beside the complete triads (Table 1). Haplin also uses the EM algorithm to reconstruct haplotypes, which enabled haplotype analyses for different combinations of SNPs in the genes that showed a plausible PoOxE effect.

A detailed description of the method for PoOxE analysis has been provided in our previous works^{22,27,31}. Briefly, PoOxE effects were calculated as follows:

- 1) Calculate the relative risk (RR) for an allele inherited from the mother (RR_{mat}) and do the same for the father (RR_{pat}).
- 2) Calculate the relative risk ratio ($RRR_{PoO} = RR_{mat} / RR_{pat}$) between the RRs in (1). RRR_{PoO} is thus an estimate of the parent-of-origin (PoO) effect.

- 3) Calculate RRR_{PoOxE} as $RRR_{PoO}(\text{Exposed}) / RRR_{PoO}(\text{Unexposed})$, where $RRR_{PoO}(\text{Exposed})$ and $RRR_{PoO}(\text{Unexposed})$ are RRR_{PoO} among triads with exposed and unexposed mothers.

Haplin uses a Wald test to test the null hypothesis of $RRR_{PoOxE} = 1$.

In order to control for multiple testing (one test for each of 341,191 SNPs), we obtained q-values using the false discovery rate (FDR) method described by Storey & Tibshirani (2003)³². Specifically, the q-values were calculated from the p-values with the *R*-function `qvalue()`³³. A q-value of 0.2 corresponds to an FDR of 20%, which means that at least 80% of SNPs with a q-value less than 0.2 would be expected to be truly associated with the outcome. As in our previous work on isolated CPO²⁷, we identified the top 20 SNPs for each of the analyses performed (see Results for details) and calculated relative risk ratios (RRRs) with 95% confidence intervals (95% CI). We paid more attention to a given gene if SNPs in that gene showed up multiple times in one set or across different sets of analyses. In accordance with recent recommendations by the

American Statistical Association and others^{34,35}, we did not consider a fixed p-value as a threshold for statistical significance.

To illustrate the general ability of the PoOxE analyses to detect true associations, power analyses for a wide range of PoOxE scenarios were performed using the Haplin function hapPowerAsymp(), as described in our recent works^{22,36}.

We focused on the regions flanking SNPs in the most interesting genes and constructed regional plots based on R-scripts developed by the Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University and Novartis Institutes of BioMedical Research³⁷. Such plots capture the extent of linkage disequilibrium between a lead SNP and neighboring SNPs, while also providing information on recombination patterns and the position of genes.

R-scripts used to conduct the statistical analyses and create figures are available (see *Software availability*)³⁸.

Bioinformatics analyses

To contextualize the findings, we searched for connections among a selection of genes in the STRING database³⁹, as well as for enrichment of these genes in expression patterns using ExpressionAtlas⁴⁰ and BGe (R package BgeeDB_2.10.0)⁴¹. Further, using Hetionet (Ver.1.0)⁴², we searched for indirect links between the genes highlighted by our analyses, the exposures and the phenotype (“cleft lip”). Hetionet is a heterogeneous network of various relationships among various data types, such as interactions between genes, or regulation of gene expression between a drug and a gene. The data used in Hetionet were carefully curated from 29 publicly available databases. To simplify the query output, the number of relationships between any two of the input query nodes (i.e., exposure, cleft lip, and the genes) was set to at most two. The exact queries together with their output are available (see *Software availability*)³⁸.

Ethical statement

The individual institutional review boards of the members of the International Cleft Consortium provided ethical approval, which can be found in the online supplementary material of the original publication⁴³. Written informed consent was provided by all participating families. Please refer to the dbGaP database for more information.

Results

For clarity, this section is structured as follows: We present the results of the PoOxE analyses of the pooled sample first (Table 3), followed by those of the European (Table 4) and Asian (Table 5) samples. We used the integrative database GeneCards and the gene-centric links therein to collate information on the genes in these tables. The 1000 Genomes browser was used to determine the chromosomal band location of a SNP. In the following sections, we focus on q-values, but all the corresponding p-values can also be found in Table 3–Table 5.

Table 6 provides a reference for the full names of all the genes mentioned in Table 3–Table 5. Table 7 shows the results of the haplotype analyses of SNPs in the most interesting genes. Figure 1 and Figure 2 present visualizations of the results from the bioinformatics analyses, and regional plots for the most important regions from Table 3–Table 5 are shown in Figure 3 and Figure 4. Figure 5 illustrates power calculations to detect different PoOxE effects in single-SNP analyses under different parameters, such as different sample sizes and minor allele frequencies. Quantile-quantile (QQ) plots for each set of analyses are shown in Figure 6–Figure 8.

Pooled sample

All the top 20 SNPs in the PoOxAlcohol analysis had the same q-value of 0.99 and are therefore not considered here as they are probably false positives (Table 3). All the SNPs in the PoOxSmoke analysis had q-values of around 0.6. Even though these q-values are still quite large, they indicate that around 40% of the SNPs are potentially true PoOxE associations. Among the top 20 SNPs in the PoOxSmoke analysis, two are in the gene for ‘Focadhesin’ (*FOCAD*), two are in ‘Platelet derived growth factor D’ (*PDGFD*), two are in ‘Ankyrin 3’ (*ANK3*), and one is in ‘Fraser syndrome 1’ (*FRASI*). Note that associations with *PDGFD* and *ANK3* were also detected in the European analyses (see below). In the PoOxVitamin analysis, only three SNPs had q-values below 0.99, and none of the genes linked to these SNPs have previously been associated with orofacial clefts.

European sample

Among the SNPs with the lowest q-values in the PoOxAlcohol analysis, rs2294035 (q=0.32, p=2.9e-6) and rs4876274 (q=0.76, p=1.3e-5) are in ‘Rho guanine nucleotide exchange factor 10’ (*ARHGEF10*; GeneCards identifier [GCID]: GC08P001823) (Table 4). The remaining SNPs had q-values above 0.76 and are not considered any further. *ARHGEF10* has not previously been linked with orofacial clefts. In the PoOxSmoke analysis, three of the SNPs were in *ANK3* (rs3793861: q=0.20, p=2.6e-6; rs7087489: q=0.20, p=3.1e-6; and rs4310561: q=0.67, p=4.0e-5). PoOxE effects in *ANK3* were also detected in the analysis of the pooled sample above. To our knowledge, *ANK3* has not previously been linked with orofacial clefts, and the same applies to SNP rs10763707 in ‘Lysosome like 1’ (*LYZL1*; GCID: GC10P029297), which had a q-value of 0.20. In the PoOxVitamin analysis, several of the SNPs shared the same q-value of 0.87 and are not considered any further. The top six SNPs had q-values of 0.44–0.65. Again, none of these genes appear to have any previous connections to clefting. For example, ‘cAMP responsive element binding protein 5’ (*CREB5*; GCID: GC07P028305) and its network of genes are involved in colorectal cancer⁴⁴, while ‘R-Spondin 3’ (*RSPO3*; GCID: GC06P127118) is implicated in tumor development. That said, Park and co-workers reported that *RSPO3* acts as an agonist in the canonical Wnt/β-catenin signaling⁴⁵, a pathway known to be implicated in a wide range of developmental processes, including craniofacial development and homeostasis^{46–49}.

Table 3. The top 20 single-nucleotide polymorphisms (SNPs) sorted by p-value in the pooled PoOxE analysis.

Exposure	SNP	Chromosomal band location ^a	P-value	Q-value	RRR (95% CI)	Gene symbol ^b	Shared ^c
ALCOHOL	rs7964474	12p13.31	7.4e-06	0.99	0.34 (0.22-0.55)	<i>ANO2</i>	
	rs999783	16q23.3-q24.1	1.8e-05	0.99	2.63 (1.69-4.10)	<i>MBTPS1</i>	
	rs4982619	14q11.2	2.1e-05	0.99	2.44 (1.62-3.68)	<i>TRA</i>	
	rs7945550	11p13	2.1e-05	0.99	2.46 (1.62-3.72)	<i>EHF</i>	Europe
	rs880813	2p12	2.5e-05	0.99	2.36 (1.58-3.51)	<i>CTNNA2</i>	
	rs2280025	16q23.3-q24.1	2.7e-05	0.99	2.59 (1.66-4.03)	<i>MBTPS1</i>	
	rs11584506	1q42.1	3.4e-05	0.99	0.39 (0.25-0.61)	<i>NC</i>	
	rs10897066	11q12.2	3.8e-05	0.99	2.29 (1.54-3.40)	<i>~MS4A5 and MS4A1</i>	
	rs2032442	14q11.2	3.9e-05	0.99	2.37 (1.57-3.59)	<i>TRA</i>	
	rs163684	12q14.1-q14.2	4.2e-05	0.99	3.23 (1.84-5.65)	<i>PPM1H</i>	
	rs8025763	15q26.3	5.6e-05	0.99	2.31 (1.54-3.47)	<i>NC</i>	
	rs13008096	2p15	6.1e-05	0.99	2.26 (1.52-3.36)	<i>NC</i>	
	rs4699228	4q24	6.2e-05	0.99	2.80 (1.69-4.63)	<i>NC</i>	
	rs2723057	4q24	6.2e-05	0.99	2.76 (1.68-4.54)	<i>NC</i>	
	rs7201659	16p12.3	6.5e-05	0.99	0.43 (0.29-0.65)	<i>XYLT1</i>	
	rs2151225	9q21.3	6.8e-05	0.99	2.54 (1.61-4.02)	<i>NC</i>	
	rs7197476	16p12.3	6.9e-05	0.99	0.44 (0.29-0.66)	<i>XYLT1</i>	
	rs2367283	9q21.3	7.0e-05	0.99	0.42 (0.28-0.65)	<i>GPR98</i>	
	rs2914354	19q13.42	7.7e-05	0.99	0.47 (0.32-0.68)	<i>~VN1R4</i>	
	rs7209652	17p12	8.7e-05	0.99	0.45 (0.30-0.67)	<i>LINC00670</i>	
SMOKE	rs10097386	8q22.1	2.6e-06	0.57	2.86 (1.85-4.43)	<i>NC</i>	
	rs2383162	9p21.3	8.5e-06	0.57	2.73 (1.75-4.24)	<i>FOCAD</i>	
	rs10738571	9p21.3	1.3e-05	0.57	2.67 (1.72-4.16)	<i>FOCAD</i>	
	rs7419201	1q43	1.4e-05	0.57	3.21 (1.90-5.44)	<i>NC</i>	Europe
	rs7541537	1q43	1.4e-05	0.57	2.52 (1.66-3.81)	<i>NC</i>	Europe
	rs7042192	9p21.3	1.5e-05	0.57	2.70 (1.72-4.22)	<i>FOCAD</i>	
	rs4977848	9p21.3	1.5e-05	0.57	2.71 (1.72-4.24)	<i>FOCAD</i>	
	rs7920088	10p14	1.6e-05	0.57	2.94 (1.80-4.81)	<i>SFMBT2</i>	
	rs12740826	1q25.2	1.9e-05	0.57	0.35 (0.22-0.57)	<i>NPHS2</i>	
	rs13173741	5q14.1	2.5e-05	0.57	2.51 (1.63-3.84)	<i>NC</i>	
	rs10757168	9p21.3	2.5e-05	0.57	2.60 (1.67-4.05)	<i>FOCAD</i>	
	rs8181543	11q22.3	2.6e-05	0.57	0.34 (0.21-0.56)	<i>PDGFD</i>	Europe
	rs168283	4q21.21	2.7e-05	0.57	0.36 (0.22-0.58)	<i>FRAS1</i>	
	rs17408603	1p31.1	2.7e-05	0.57	3.19 (1.86-5.49)	<i>NC</i>	
	rs11624380	14q22.3	2.8e-05	0.57	0.37 (0.23-0.59)	<i>PELI2</i>	
	rs2177971	8p21.2	2.8e-05	0.57	3.43 (1.93-6.11)	<i>NC</i>	Europe
	rs7943401	11q22.3	2.9e-05	0.57	0.34 (0.20-0.56)	<i>PDGFD</i>	
	rs3793861	10q21.2	3.0e-05	0.57	2.71 (1.70-4.34)	<i>ANK3</i>	Europe
	rs4394682	1p36.13	3.4e-05	0.58	0.34 (0.20-0.57)	<i>~CAPZB</i>	
	rs7087489	10q21.2	3.5e-05	0.58	2.69 (1.68-4.30)	<i>ANK3</i>	Europe

Exposure	SNP	Chromosomal band location ^a	P-value	Q-value	RRR (95% CI)	Gene symbol ^b	Shared ^c
VITAMIN	rs2302304	19p13.3	1.3e-06	0.46	3.12 (1.97-4.94)	<i>TJP3</i>	
	rs2689128	1q43	4.2e-06	0.71	3.28 (1.98-5.43)	<i>NC</i>	Europe
	rs9572250	13q21.33	7.8e-06	0.88	0.44 (0.31-0.63)	<i>KLHL1</i>	
	rs4875398	8p23.2	1.4e-05	0.99	2.08 (1.49-2.89)	<i>CSMD1</i>	
	rs3909551	13q21.33	1.7e-05	0.99	0.46 (0.32-0.65)	<i>KLHL1</i>	
	rs9371494	6q25.1	2.4e-05	0.99	2.23 (1.54-3.24)	<i>MTHFD1L</i>	
	rs8101981	19p13.12	2.9e-05	0.99	0.48 (0.34-0.68)	<i>LINC00905</i>	Europe
	rs7939975	11.p12	3.6e-05	0.99	2.08 (1.47-2.94)	<i>NC</i>	
	rs10495767	2p23.2	3.6e-05	0.99	2.28 (1.54-3.36)	<i>NC</i>	
	rs11673884	2q36.3	4.2e-05	0.99	0.51 (0.37-0.70)	<i>~SLC19A3</i>	
	rs6489630	12p13.31	4.6e-05	0.99	2.23 (1.52-3.28)	<i>NTF3</i>	
	rs3815311	17p12	5.3e-05	0.99	3.19 (1.82-5.59)	<i>ARHGAP44</i>	Europe
	rs358017	3p21.1-p14.3	5.4e-05	0.99	2.25 (1.52-3.34)	<i>~CACNA2D3</i>	
	rs7082286	10q21.1	5.8e-05	0.99	4.03 (2.04-7.96)	<i>NC</i>	
	rs921743	10p13	6.0e-05	0.99	2.18 (1.49-3.19)	<i>RSU1</i>	
	rs10764037	10p12.31	6.3e-05	0.99	0.50 (0.36-0.70)	<i>MALRD1</i>	
	rs8112256	19p13.11	6.8e-05	0.99	2.13 (1.47-3.10)	<i>FAM129C</i>	
	rs4569521	2q21.1	8.1e-05	0.99	0.42 (0.27-0.65)	<i>ARHGEF4</i>	
	rs6830509	4q28	8.7e-05	0.99	1.96 (1.40-2.73)	<i>NC</i>	
	rs9503155	6p25.3	8.8e-05	0.99	0.49 (0.34-0.70)	<i>GMDS-AS1</i>	

^aThe 1000 Genomes browser was used to determine the chromosomal band location of a SNP. ^bIf a SNP is located within a gene itself, the gene symbol is provided (the full names of the genes are provided in Table 6). SNPs located within 40 kb of a gene have the prefix '~', and those not located within a 40 kb-distance of a gene are denoted as NC (for 'not close'). Note that pseudogenes and non-coding RNAs are excluded. ^cShared: Also featured in Table 4 or Table 5. SNP, single-nucleotide polymorphism; RRR, relative risk ratio; CI, confidence interval; NC, not close.

Table 4. The top 20 single-nucleotide polymorphisms (SNPs) sorted by p-value in the European PoOxE analysis.

Exposure	SNP	Chromosomal band location ^a	P-value	Q-value	RRR (95% CI)	Gene symbol ^b	Shared ^c
ALCOHOL	rs10496410	2q12	7.5e-07	0.15	6.04 (2.96-12.32)	<i>NC</i>	
	rs7579926	2q12	9.3e-07	0.15	5.95 (2.92-12.13)	<i>NC</i>	
	rs2294035	8p23.3	2.9e-06	0.32	0.31 (0.19-0.51)	<i>ARHGEF10</i>	
	rs6975650	7q33	1.1e-05	0.76	0.31 (0.19-0.52)	<i>NC</i>	
	rs4876274	8p23.3	1.3e-05	0.76	2.99 (1.83-4.90)	<i>ARHGEF10</i>	
	rs2245225	12q14	1.4e-05	0.76	3.46 (1.98-6.05)	<i>NC</i>	
	rs927318	9p24.2	2.0e-05	0.76	0.36 (0.22-0.57)	<i>GLIS3</i>	
	rs10735337	12q23.1	2.0e-05	0.76	0.36 (0.23-0.58)	<i>CCDC38</i>	
	rs6427247	1q24	2.1e-05	0.76	2.87 (1.77-4.67)	<i>NC</i>	
	rs12669493	7p21.1	2.4e-05	0.79	3.11 (1.84-5.26)	<i>LRRC72</i>	
	rs13255561	8p23.3	3.6e-05	0.88	0.30 (0.17-0.53)	<i>DLGAP2</i>	
	rs12242535	10q21.2	3.9e-05	0.88	3.94 (2.05-7.57)	<i>NC</i>	
	rs943881	14q32.2	4.3e-05	0.88	0.36 (0.22-0.59)	<i>CYP46A1</i>	
	rs10491327	5q34	4.4e-05	0.88	0.28 (0.15-0.52)	<i>NC</i>	
	rs7945550	11p13	4.5e-05	0.88	2.82 (1.71-4.64)	<i>EHF</i>	Pooled
	rs7232492	18p11.31	5.3e-05	0.88	0.27 (0.15-0.51)	<i>DLGAP1</i>	
	rs11242213	5q31.1	5.4e-05	0.88	4.71 (2.22-10.00)	<i>UBE2B</i>	
	rs34352212	5q34	6.0e-05	0.88	0.32 (0.18-0.55)	<i>NC</i>	
	rs1990185	17q24	6.1e-05	0.88	3.16 (1.80-5.54)	<i>NC</i>	
	rs521419	17p12	6.5e-05	0.88	2.80 (1.69-4.64)	<i>NC</i>	

Exposure	SNP	Chromosomal band location ^a	P-value	Q-value	RRR (95% CI)	Gene symbol ^b	Shared ^c
SMOKE	rs10763707	10p12.1-p11.23	1.5e-06	0.20	4.08 (2.30-7.23)	<i>LYZL1</i>	
	rs7541537	1q43	2.0e-06	0.20	3.31 (2.02-5.43)	NC	Pooled
	rs7419201	1q43	2.1e-06	0.20	4.77 (2.50-9.11)	NC	Pooled
	rs3793861	10q21.2	2.6e-06	0.20	3.67 (2.13-6.32)	<i>ANK3</i>	Pooled
	rs7087489	10q21.2	3.1e-06	0.20	3.63 (2.11-6.25)	<i>ANK3</i>	Pooled
	rs814518	19q13.2	4.5e-06	0.25	3.35 (2.00-5.62)	<i>SHKBP1</i>	
	rs4693142	4q21.3	6.4e-06	0.30	0.26 (0.15-0.47)	<i>MAPK10</i>	
	rs4454616	10p14	9.2e-06	0.38	3.06 (1.86-5.00)	NC	
	rs2904096	4q21.3	1.2e-05	0.40	0.27 (0.15-0.49)	<i>MAPK10</i>	
	rs2290682	19q13.2	1.3e-05	0.40	3.22 (1.90-5.44)	<i>SHKBP1</i>	
	rs6532013	4q22	1.4e-05	0.40	3.04 (1.84-5.02)	NC	
	rs1868368	8q24.2	2.2e-05	0.61	0.29 (0.17-0.52)	NC	
	rs2177971	8p21.2	2.6e-05	0.64	4.02 (2.10-7.68)	NC	Pooled
	rs6807522	3q22.1	3.4e-05	0.67	2.91 (1.75-4.81)	<i>TMEM108</i>	
	rs17604550	15q25.3	3.6e-05	0.67	0.33 (0.20-0.56)	<i>AGBL1</i>	
	rs12883776	14q22.3	3.7e-05	0.67	0.34 (0.20-0.57)	<i>PELI2</i>	
	rs7234787	18q21.1	3.8e-05	0.67	0.22 (0.11-0.45)	<i>ZBTB7C</i>	
	rs8181543	11q22.3	3.8e-05	0.67	0.28 (0.16-0.52)	<i>PDGFD</i>	Pooled
	rs4310561	10q21.2	4.0e-05	0.67	2.90 (1.75-4.83)	<i>ANK3</i>	
	rs3800036	6p25.3	4.1e-05	0.67	0.35 (0.22-0.58)	<i>GMDS</i>	
VITAMIN	rs2689128	1q43	2.2e-06	0.44	4.82 (2.52-9.25)	NC	Pooled
	rs2237360	7p15.1	4.0e-06	0.44	0.29 (0.18-0.50)	<i>CREB5</i>	
	rs7793050	7p21	4.0e-06	0.44	3.94 (2.20-7.05)	<i>RPA3-AS1</i>	
	rs7766106	6q22.33	6.4e-06	0.53	0.31 (0.19-0.52)	<i>RSPO3</i>	
	rs2809964	1p36.11	1.2e-05	0.65	3.05 (1.85-5.03)	<i>~RCAN3, NCMAP and RPL26P8</i>	
	rs3859121	16q12.1	1.2e-05	0.65	0.16 (0.07-0.37)	<i>N4BP1</i>	
	rs1092733	3p26	2.4e-05	0.87	0.32 (0.19-0.54)	NC	
	rs7559678	2q11.2	2.7e-05	0.87	0.35 (0.22-0.57)	<i>VWA3B</i>	
	rs2366837	5p13.2	3.0e-05	0.87	0.35 (0.21-0.57)	NC	
	rs10084852	4q28.3	3.0e-05	0.87	7.26 (2.86-18.45)	<i>PCDH10</i>	
	rs6446389	4p16.2	4.2e-05	0.87	3.38 (1.89-6.06)	<i>EVC2</i>	
	rs2242909	21q22.1	4.2e-05	0.87	2.92 (1.75-4.87)	NC	
	rs595536	1q42.2	4.4e-05	0.87	3.21 (1.83-5.60)	<i>~SIPA1L2</i>	
	rs6726527	2q37.1	4.5e-05	0.87	0.24 (0.12-0.48)	<i>~SP140 and SP140L</i>	
	rs12733019	1p32.1	4.8e-05	0.87	0.24 (0.12-0.47)	NC	
	rs8101981	19p13.12	4.8e-05	0.87	0.34 (0.21-0.58)	<i>LINC00905</i>	Pooled
	rs17793145	8p22	4.9e-05	0.87	3.39 (1.88-6.12)	<i>DLC1</i>	
	rs4973310	2q37.1	5.1e-05	0.87	0.24 (0.12-0.48)	<i>~SP140 and SP140L</i>	
	rs3815311	17p12	5.5e-05	0.87	4.57 (2.18-9.57)	<i>ARHGAP44</i>	Pooled
	rs8072885	17q25.3	5.5e-05	0.87	0.22 (0.10-0.46)	<i>RBFOX3</i>	

^aThe 1000 Genomes browser was used to determine the chromosomal band location of a SNP. ^bIf a SNP is located within a gene itself, the gene symbol is provided (the full names of the genes are provided in Table 6). SNPs located within 40 kb of a gene have the prefix '~', and those not located within a 40 kb-distance of a gene are denoted as NC (for 'not close'). Note that pseudogenes and non-coding RNAs are excluded. ^cShared: Also featured in Table 3 or Table 5. SNP, single-nucleotide polymorphism; RRR, relative risk ratio; CI, confidence interval; NC, not close.

Table 5. The top 20 SNPs sorted by p-value in the Asian parent-of-origin interactions with vitamins (PoOxVitamin) analysis.

SNP	Chromosomal band location ^a	P-value	Q-value	RRR (95% CI)	Gene symbol ^a
rs1889976	1q25.3	8.8e-06	0.86	3.88 (2.13-7.05)	<i>SWT1</i>
rs259395	6q24.3	1.1e-05	0.86	0.23 (0.12-0.45)	<i>ADGB</i>
rs10798004	1q25.3	1.5e-05	0.86	3.70 (2.04-6.68)	<i>~IVNS1ABP and SWT1</i>
rs12431484	14q11.2	2.2e-05	0.86	0.24 (0.12-0.46)	<i>TRA</i>
rs10518981	15q15.3-q21.1	2.3e-05	0.86	0.22 (0.11-0.45)	<i>~CTDSPL2 and EIF3J-AS1 and EIF3J</i>
rs1940698	11q23.2	2.4e-05	0.86	0.21 (0.10-0.43)	<i>NCAM1</i>
rs171477	21q21	2.5e-05	0.86	0.23 (0.12-0.46)	<i>C21orf91-OT1</i>
rs9862866	3p14.1	3.1e-05	0.86	0.24 (0.12-0.47)	<i>~RPL21P41</i>
rs865585	6p21.1	3.6e-05	0.86	0.19 (0.09-0.42)	<i>NC</i>
rs17591732	11q23.2	3.8e-05	0.86	0.22 (0.11-0.45)	<i>NCAM1</i>
rs12630106	3q13.1	5.6e-05	0.86	3.56 (1.92-6.61)	<i>NC</i>
rs7316350	12q15	6.0e-05	0.86	0.22 (0.10-0.46)	<i>NC</i>
rs7336296	13q31	6.1e-05	0.86	3.39 (1.87-6.17)	<i>NC</i>
rs1499916	2q22	6.4e-05	0.86	0.22 (0.10-0.46)	<i>NC</i>
rs7153574	14q11.2	6.5e-05	0.86	0.26 (0.13-0.50)	<i>TRA</i>
rs6439772	3q22	7.0e-05	0.86	0.26 (0.14-0.51)	<i>NC</i>
rs1348564	3q22	7.1e-05	0.86	0.27 (0.14-0.52)	<i>NC</i>
rs2360838	11p15.4	7.3e-05	0.86	3.37 (1.85-6.14)	<i>~OR10A3 and NLRP10 and OR10A6</i>
rs12204808	6q14.1	7.3e-05	0.86	4.63 (2.17-9.87)	<i>IMPG1</i>
rs1407555	1q25.3	7.5e-05	0.86	3.30 (1.83-5.96)	<i>TRMT1L</i>

^aThe 1000 Genomes browser was used to determine the chromosomal band location of a SNP. ^bIf a SNP is located within a gene itself, the gene symbol is provided (the full names of the genes are provided in Table 6). SNPs located within 40 kb of a gene have the prefix '~', and those not located within a 40 kb-distance of a gene are denoted as NC (for 'not close'). Note that pseudogenes and non-coding RNAs (ncRNA) are excluded. There is no column for "shared" here, as none of these SNPs featured among those listed in Table 3 or Table 4. SNP, single-nucleotide polymorphism; RRR, relative risk ratio; CI, confidence interval; NC, not close.

Asian sample

In the only analysis possible for this ethnic group (PoOxVitamin), all the SNPs had the same q-value of 0.86 (Table 5). They are thus most likely to be false positives and will not be considered any further.

Haplotype analysis of SNPs in *ANK3* and *ARHGEF10*

We chose to focus here on the PoOxE effects detected with SNPs in *ANK3* and *ARHGEF10*. As mentioned above, *ANK3* showed up several times among the top PoOxSmoke hits both in the pooled and European analyses, and strong signals for SNPs in *ARHGEF10* were detected twice in the European analysis of PoOxAlcohol. We conducted stratified analyses of the effect of the child's allele, GxE effects, PoO effects, and PoOxE effects for each SNP and haplotype (with haplotypes analyzed both in two-SNP and three-SNP combinations) in these two genes (Table 7). Specifically, we analyzed rs3793861, rs7087489 and rs4310561 in *ANK3* that showed PoOxSmoke effects in the European sample, and rs2294035 and rs4876274 in *ARHGEF10* that showed PoOxAlcohol effects in the same sample (Table 4). The results

did not show any child effects or GxSmoke effects for single SNPs in *ANK3*. By contrast, the p-values were low for all three SNPs in the PoO analyses or PoOxSmoke analyses. This was also the case with the 't-a' allele in the two-SNP combination rs3793861-rs7087489 and the 'c-t-a' allele in the three-SNP-combination rs3793861-rs7087489-rs4310561. The other alleles were only associated with PoOxSmoke effects.

For *ARHGEF10*, we analyzed the two SNPs that showed PoOxAlcohol effects in the European sample (rs2294035 and rs4876274) but did not discover any effects in either single-SNP or haplotype analyses.

Bioinformatics analysis

Because of their low q-values, the genes appearing in the PoOxSmoke and PoOxAlcohol analyses in Table 4 were selected for further analyses using the STRING database, ExpressionAtlas and BGee. However, none of the searches for direct links among the genes yielded any evidence to explain why those genes appeared in our results together.

Table 6. Full names of all the genes and loci mentioned in Table 3–Table 5.

Gene symbol	Full gene/locus name
<i>ADGB</i>	Androglobin
<i>AGBL1</i>	ATP/GTP binding protein like 1
<i>ANK3</i>	Ankyrin 3, node of Ranvier (ankyrin G)
<i>ANO2</i>	Anoctamin 2
<i>ARHGAP44</i>	Rho GTPase activating protein 44
<i>ARHGEF10</i>	Rho guanine nucleotide exchange factor 10
<i>ARHGEF4</i>	Rho guanine nucleotide exchange factor (GEF) 4
<i>C21orf91-OT1</i>	NA
<i>CACNA2D3</i>	Calcium channel, voltage-dependent, alpha 2/delta subunit 3
<i>CAPZB</i>	Capping protein (actin filament) muscle Z-line, beta
<i>CCDC38</i>	coiled-coil domain containing 38
<i>CREB5</i>	cAMP responsive element binding protein 5
<i>CSMD1</i>	CUB and Sushi multiple domains 1
<i>CTDSP2</i>	CTD small phosphatase like 2
<i>CTNNA2</i>	Catenin (cadherin-associated protein), alpha 2
<i>CYP46A1</i>	cytochrome P450 family 46 subfamily A member 1
<i>DLC1</i>	DLC1 Rho GTPase activating protein
<i>DLGAP1</i>	DLG associated protein 1
<i>DLGAP2</i>	DLG associated protein 2
<i>EHF</i>	Ets homologous factor
<i>EIF3J</i>	eukaryotic translation initiation factor 3 subunit J
<i>EIF3J-AS1</i>	EIF3J divergent transcript
<i>EVC2</i>	EvC ciliary complex subunit 2
<i>FAM129C</i>	Family with sequence similarity 129, member C
<i>FOCAD</i>	Focadhesin
<i>FRAS1</i>	Fraser syndrome 1
<i>GLIS3</i>	GLIS family zinc finger 3
<i>GMDS</i>	GDP-mannose 4,6-dehydratase
<i>GMDS-AS1</i>	GMDS antisense RNA 1 (head to head)
<i>GPR98</i>	G protein-coupled receptor 98
<i>IMPG1</i>	interphotoreceptor matrix proteoglycan 1
<i>IVNS1ABP</i>	influenza virus NS1A binding protein
<i>KLHL1</i>	Kelch-like family member 1
<i>LINC00670</i>	Long intergenic non-protein coding RNA 670
<i>LINC00905</i>	Long intergenic non-protein coding RNA 905
<i>LRRC72</i>	leucine rich repeat containing 72
<i>LYZL1</i>	Lysozyme like 1
<i>MAPK10</i>	Mitogen-activated protein kinase 10
<i>MALRD1</i>	MAM and LDL receptor class A domain containing 1
<i>MBTPS1</i>	Membrane-bound transcription factor peptidase, site 1

Gene symbol	Full gene/locus name
<i>MS4A1</i>	Membrane-spanning 4-domains, subfamily A, member 1
<i>MS4A5</i>	Membrane-spanning 4-domains, subfamily A, member 5
<i>MTHFD1L</i>	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like
<i>N4BP1</i>	NEDD4 binding protein 1
<i>NCAM1</i>	neural cell adhesion molecule 1
<i>NCMAP</i>	non-compact myelin associated protein
<i>NLRP10</i>	NLR family pyrin domain containing 10
<i>NPHS2</i>	Nephrosis 2, idiopathic, steroid-resistant (podocin)
<i>NTF3</i>	Neurotrophin 3
<i>OR10A3</i>	olfactory receptor family 10 subfamily A member 3
<i>OR10A6</i>	olfactory receptor family 10 subfamily A member 6 (gene/pseudogene)
<i>PCDH10</i>	protocadherin 10
<i>PDGFD</i>	Platelet derived growth factor D
<i>PELI2</i>	Pellino E3 ubiquitin protein ligase family member 2
<i>PPM1H</i>	Protein phosphatase, Mg ²⁺ /Mn ²⁺ dependent, 1H
<i>RBFOX3</i>	RNA binding fox-1 homolog 3
<i>RCAN3</i>	RCAN family member 3
<i>RPL21P41</i>	ribosomal protein L21 pseudogene 41
<i>RPL26P8</i>	ribosomal protein L26 pseudogene 8
<i>RSPO3</i>	R-spondin 3
<i>RSU1</i>	Ras suppressor protein 1
<i>SFMBT2</i>	Scm-like with four mbt domains 2
<i>SHKBP1</i>	SH3KBP1 binding protein 1
<i>SIPA1L2</i>	signal induced proliferation associated 1 like 2
<i>SLC19A3</i>	Solute carrier family 19 (thiamine transporter), member 3
<i>SP140</i>	SP140 nuclear body protein
<i>SP140L</i>	SP140 nuclear body protein like
<i>SWT1</i>	SWT1, RNA endoribonuclease homolog
<i>TJP3</i>	Tight junction protein 3
<i>TMEM108</i>	Transmembrane protein 108
<i>TRA</i>	T cell receptor alpha locus
<i>TRMT1L</i>	tRNA methyltransferase 1 like
<i>UBE2B</i>	ubiquitin conjugating enzyme E2 B
<i>UMAD1</i>	UBAP1-MVB12-associated (UMA) domain containing 1
<i>VN1R4</i>	Vomer nasal 1 receptor 4
<i>VWA3B</i>	von Willebrand factor A domain containing 3B
<i>XYLT1</i>	Xylosyltransferase I
<i>ZBTB7C</i>	zinc finger and BTB domain containing 7C

Table 7. Stratified analyses of the top single-nucleotide polymorphisms (SNPs) and haplotypes in ANK3 (PoOxSmoke) and ARHGEF10 (PoOxAlcohol).

Gene name	SNP/haplotype	^a Target allele/Reference	^b Frequency	Effect type	^c RRR (95% CI)	p-value
ANK3	rs3793861	c/G	0.30	Child	1.06 (0.91-1.24)	0.45
				GxSmoke	1.20 (0.83-1.60)	0.41
				PoO	1.39 (1.09-1.76)	0.007
				PoOxSmoke	3.67 (2.13-6.32)	2.6e-6
	rs7087489	t/A	0.30	Child	1.06 (0.91-1.24)	0.44
				GxSmoke	1.10 (0.82-1.60)	0.42
				PoO	1.40 (1.10-1.77)	0.006
				PoOxSmoke	3.63 (2.11-6.25)	3.1e-6
	rs4310561	a/T	0.34	Child	1.12 (0.97-1.24)	0.13
				GxSmoke	1.20 (0.87-1.60)	0.27
				PoO	1.31 (1.03-1.64)	0.02
				PoOxSmoke	2.90 (1.75-4.83)	4.0e-5
	rs3793861- rs7087489	c-t/G-A	0.30	Child	1.07 (0.92-1.24)	0.39
				GxSmoke	1.10 (0.81-1.60)	0.45
				PoO	1.38 (1.08-1.75)	0.008
				PoOxSmoke	3.71 (2.16-6.39)	2.2e-6
	rs7087489- rs4310561	A-a/A-T	0.04	Child	1.32 (0.95-1.83)	0.10
				GxSmoke	1.50 (0.71-3.10)	0.29
				PoO	0.93 (0.60-1.45)	0.74
				PoOxSmoke	1.57 (0.62-3.96)	0.34
		t-a/A-T	0.30	Child	1.09 (0.94-1.28)	0.26
				GxSmoke	1.20 (0.83-1.60)	0.37
				PoO	1.34 (1.06-1.70)	0.02
				PoOxSmoke	3.65 (2.13-6.28)	2.7e-6
rs3793861- rs7087489- rs4310561	G-A-a/G-A-T	0.04	Child	1.32 (0.95-1.83)	0.10	
			GxSmoke	1.50 (0.71-3.10)	0.29	
			PoO	0.93 (0.60-1.45)	0.75	
			PoOxSmoke	1.56 (0.62-3.94)	0.35	
	c-t-a/G-A-T	0.30	Child	1.09 (0.94-1.28)	0.26	
			GxSmoke	1.20 (0.83-1.60)	0.37	
			PoO	1.35 (1.06-1.71)	0.01	
			PoOxSmoke	3.62 (2.10-6.21)	3.3e-6	

Gene name	SNP/haplotype	^a Target allele/Reference	^b Frequency	Effect type	^c RRR (95% CI)	p-value
ARHGGEF10	rs2294035	a/T	0.49	Child	0.94 (0.82-1.08)	0.38
				GxAlcohol	1.20 (0.87-1.50)	0.32
				PoO	0.95 (0.75-1.20)	0.67
				PoOxAlcohol	0.32 (0.19-0.51)	2.9e-6
	rs4876274	t/A	0.47	Child	1.04 (0.90-1.20)	0.57
				GxAlcohol	0.90 (0.67-1.20)	0.47
				PoO	1.02 (0.80-1.29)	0.90
				PoOxAlcohol	2.99 (1.83-4.90)	1.3e-5
	rs2294035-rs4876274	T-A/a-A	0.04	Child	1.15 (0.80-1.68)	0.44
				GxAlcohol	0.73 (0.33-1.60)	0.45
				PoO	1.41 (0.85-2.37)	0.19
				PoOxAlcohol	1.56 (0.49-4.93)	0.45
T-t/a-A		0.47	Child	1.04 (0.90-1.20)	0.63	
			GxAlcohol	0.90 (0.67-1.20)	0.46	
			PoO	1.00 (0.80-1.27)	0.98	
			PoOxAlcohol	3.20 (1.97-5.21)	2.8e-6	

^aEffect allele or haplotype against the reference. Lowercase indicates the minor allele at the SNP. ^bMinor allele frequency for a given SNP. In haplotype analyses, this corresponds to the frequencies of haplotypes other than the reference. ^cRR for child effects; RRR for GxSmoke or GxAlcohol, PoO and PoOxSmoke or PoOxAlcohol. All p-values <0.05 are highlighted in bold. Note that in a two-SNP-haplotype, there are four possible combinations, and in a three-SNP-haplotype there are eight. However, only two or three of these combinations were actually observed in the data.

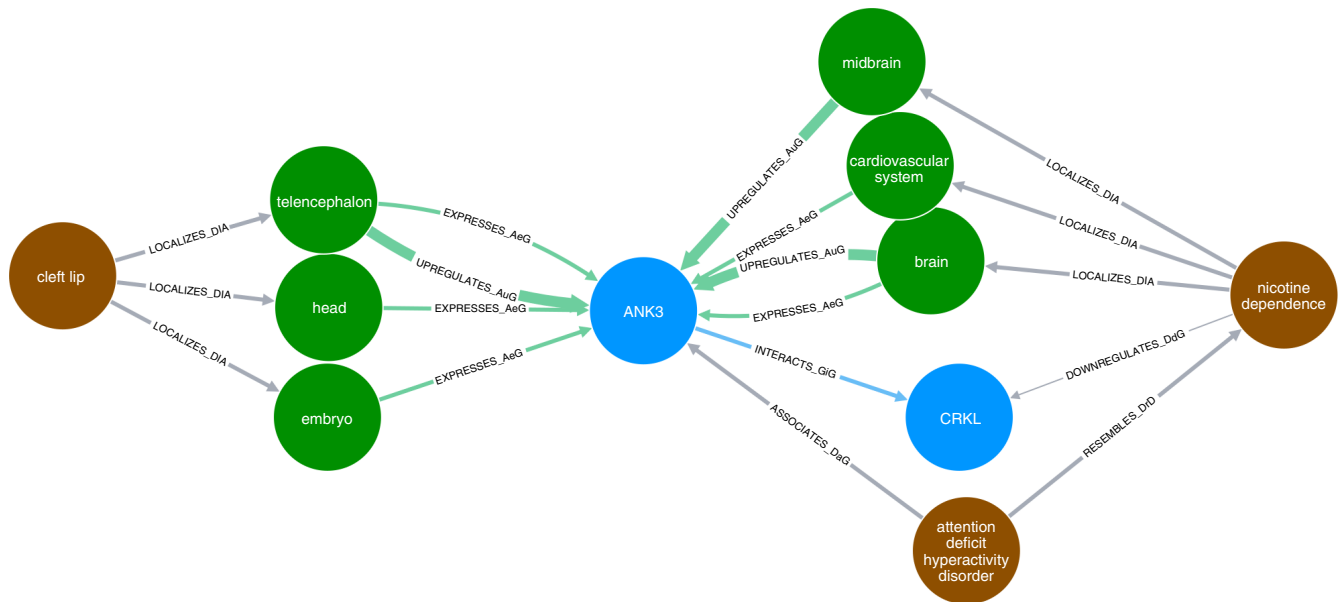


Figure 1. Indirect relationships between ANK3 and nicotine dependence, and between ANK3 and cleft lip. The brown nodes represent diseases, blue nodes show genes/proteins, and green nodes represent organs (anatomy). Each arrow represents a specific relationship between nodes: “LOCALIZES_DiA” = disease was found to be localized in an anatomy (organ); “EXPRESSES_AuG”, “UPREGULATES_AuG”, “DOWNREGULATES_AuG” mean that the gene is expressed, upregulated, or downregulated in the anatomy (organ), respectively; “INTERACTS_GiG” means that the two genes were found to interact with each other (physically, as proteins); “ASSOCIATES_DaG” means that the gene was found to be associated with the disease; “RESEMBLES_DrD” means that the two diseases were found to occur significantly more often together in MEDLINE articles than would be expected by chance alone. Note that in this setting, the term “disease” includes any adverse medical condition, like syndromes, mental disorders, congenital anomalies, and so on.

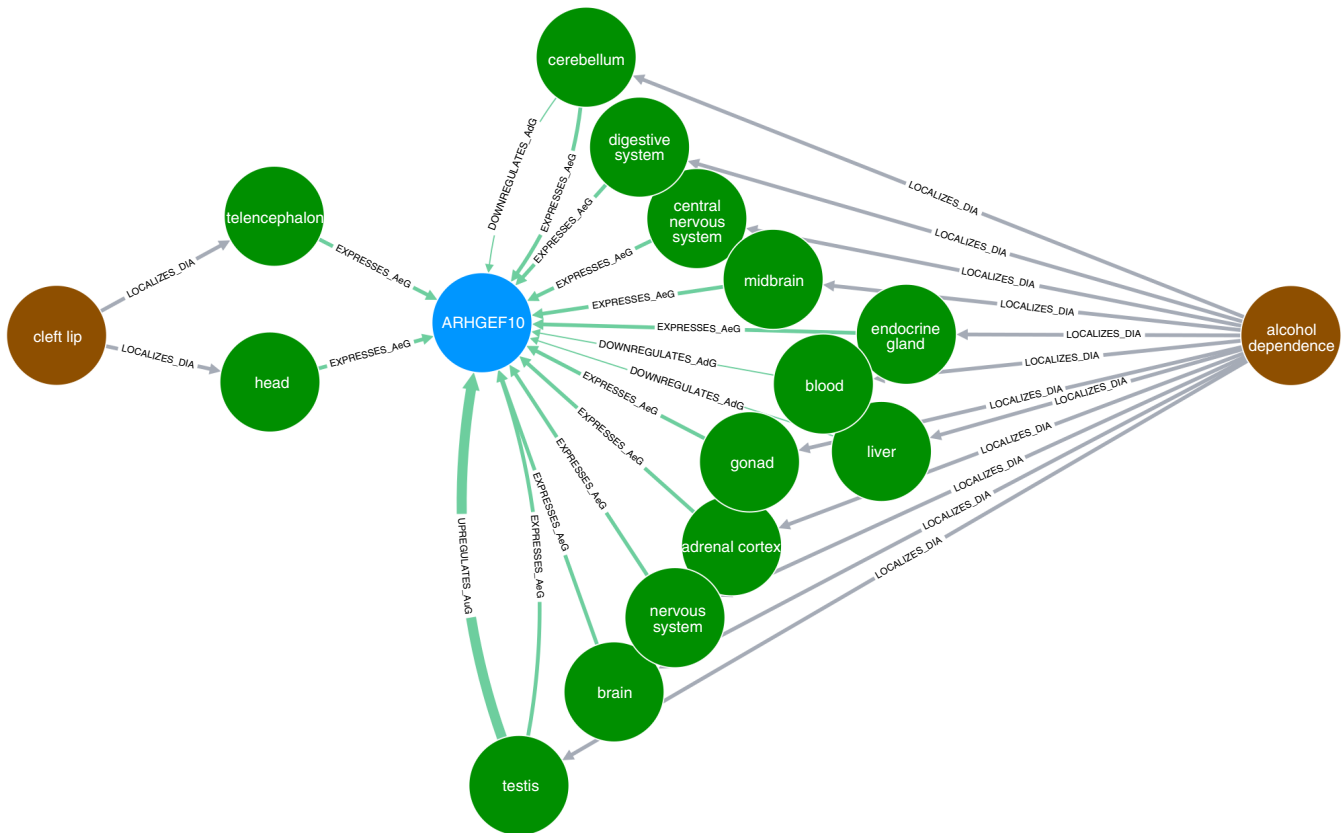


Figure 2. Indirect relationships between *ARHGEF10* and alcohol dependence, and between *ARHGEF10* and cleft lip. The brown nodes represent diseases, blue nodes show genes/proteins, and green nodes represent organs (anatomy). Each arrow represents a specific relationship between nodes: "LOCALIZES_DiA" = disease was found to be localized in an anatomy (organ); "EXPRESSES_AuG", "UPREGULATES_AuG", "DOWNREGULATES_AuG" mean that the gene is expressed, upregulated, or downregulated in the anatomy (organ), respectively; "INTERACTS_GiG" means that the two genes were found to interact with each other (physically, as proteins); "ASSOCIATES_DaG" means that the gene was found to be associated with the disease; "RESEMBLES_DrD" means that the two diseases were found to occur significantly more often together in MEDLINE articles than would be expected by chance alone. Note that in this setting, the term "disease" includes any adverse medical condition, like syndromes, mental disorders, congenital anomalies, and so on.

Regarding the indirect relationships, these are visualized in [Figure 1](#) for relationships between *ANK3* and cleft lip (Disease Ontology ID [DOID]: 9296), and, simultaneously, between *ANK3* and nicotine dependence (DOID: 0050742). As Hetionet does not include information about smoking, we chose "nicotine dependence" as a proxy. *ANK3* is connected to nicotine dependence through several nodes, two of which are particularly noteworthy. First, *ANK3* has been reported to be strongly associated with attention-deficit/hyperactivity disorder (ADHD)⁵⁰, and a connection between ADHD and nicotine dependence has been reported. The connection was calculated based on articles listed in MEDLINE, where this pair of conditions co-occurred significantly more frequently than would be expected by chance^{51,52}. The second path goes through the gene 'CRK Like Proto-Oncogene, Adaptor Protein' (*CRKL*). It interacts with *ANK3* and is downregulated in nicotine dependence. Furthermore, *ANK3* is expressed in the telencephalon (the most highly developed

part of the forebrain), the embryo, and the head – all of which are all relevant to CL/P.

[Figure 2](#) shows the indirect relationships between cleft lip, *ARHGEF10* and alcohol dependence. Like nicotine dependence in the above analyses, alcohol dependence (DOID: 0050741) was used here as a proxy for maternal alcohol consumption. The only relationships found between *ARHGEF10* and cleft lip are the expression of *ARHGEF10* in the head and telencephalon. By contrast, there were twelve different relationships between *ARHGEF10* and alcohol dependence. However, there were no shared paths connecting cleft lip and alcohol dependence via any of the 12 organs.

Regional plots

The regional plot for rs3793861 ([Figure 3](#)) shows that several SNPs in *ANK3* that were not in linkage disequilibrium with

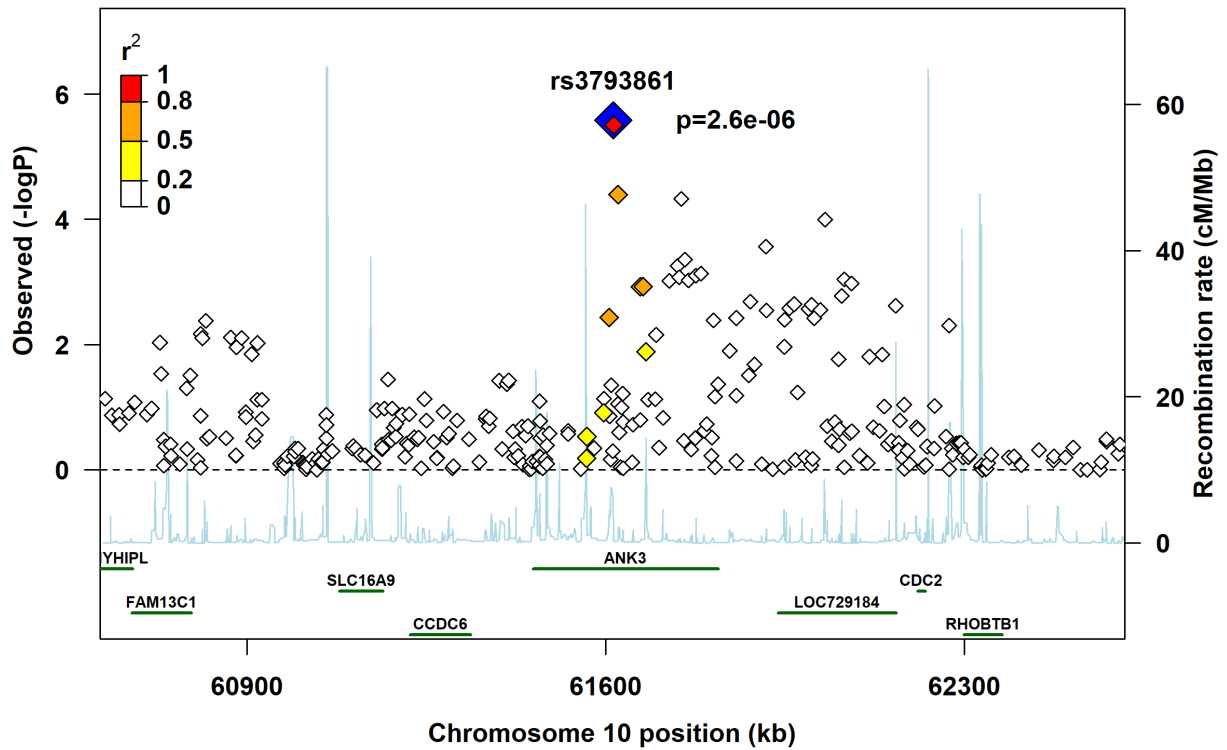


Figure 3. Regional association plot for rs3793861 in ANK3. The plot provides information on the recombination rate and linkage disequilibrium between the lead SNP (blue diamond) and other SNPs in the region.

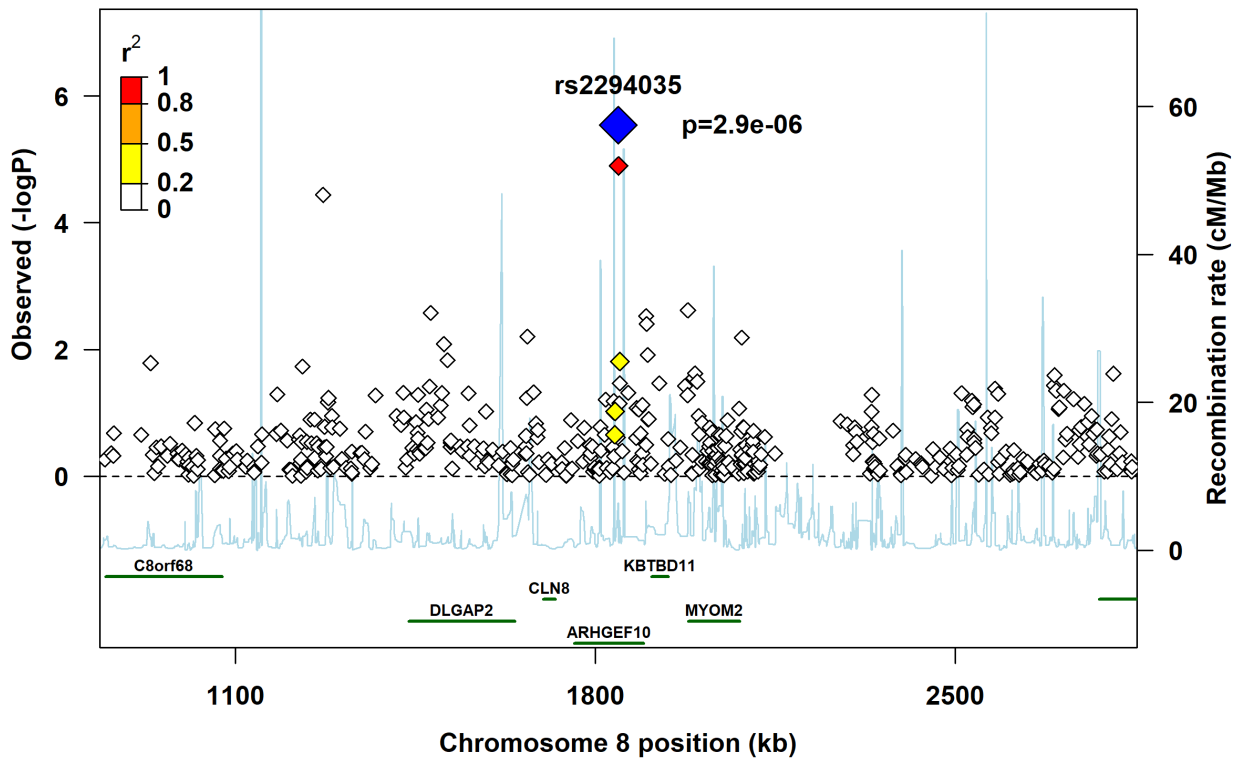


Figure 4. Regional association plot for rs2294035 in ANK3. The plot provides information on the recombination rate and linkage disequilibrium between the lead SNP (blue diamond) and other SNPs in the region.

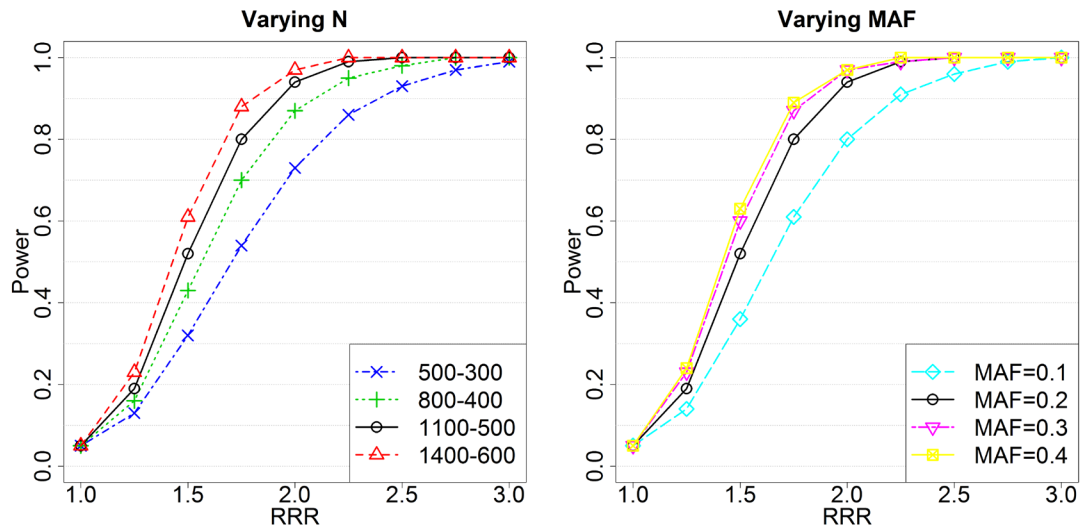


Figure 5. Power vs. RRR. Left panel: Setting the minor allele frequency to 0.2 while varying the number of unexposed and exposed triads (unexposed-exposed). Right panel: Setting the number of unexposed and exposed triads to 1100 and 500, respectively, while varying the minor allele frequency. In all analyses, the significance level was 0.05. We varied the maternal RR in exposed triads, so that $RRR = RR_{mat}(Exposed)$. The black curve is the same in both panels because of shared parameters. RRR, relative risk ratio; RR_{mat} , relative risk for an allele inherited from the mother; MAF, minor allele frequency.

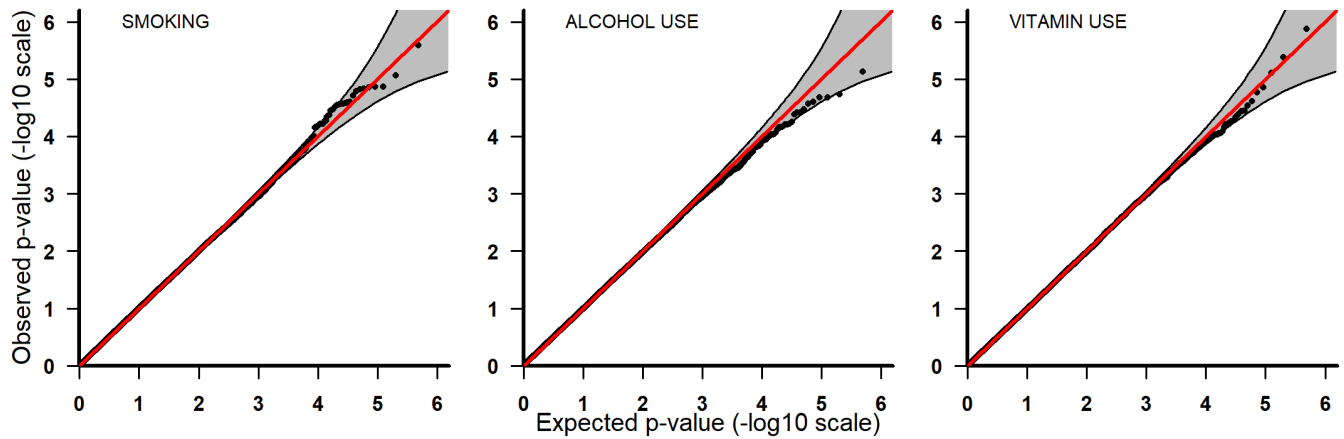


Figure 6. Pooled analyses. Q-Q plots for PoOxSmoke (left) PoOxAlcohol (middle) and PoOxVitamin (right) with 95% pointwise confidence bands. Q-Q, quantile-quantile; PoOxAlcohol, parent-of-origin interactions with alcohol; PoOxSmoke, parent-of-origin interactions with smoking; PoOxVitamin, parent-of-origin interactions with vitamins.

rs3793861 had p-values in the range 10^{-4} to 10^{-3} , which lends support to either *ANK3* itself or genes in its vicinity influencing the risk of clefting. However, we did not observe a similar pattern in the regional plot for rs2294035 (Figure 4).

Power analysis

Figure 5 shows that the power does not increase appreciably when the minor allele frequency increases beyond 0.2. However, there is a lot to be gained by increasing the sample size from 500 unexposed and 300 exposed (European, smoke/alcohol) to 1400-600 (pooled, vitamin). Further, the RRRs in the plots are based on changing only the effect of the maternal allele in the

exposed triads. The same RRRs could have been achieved in a number of ways, which complicates the interpretation of the RRRs in Table 3–Table 5. Still, if a strong effect is detected with a SNP in a gene, this strengthens the case for its contribution to clefting.

Discussion

The main aim of this paper was to identify genome-wide PoOxE effects in the larger sample of isolated CL/P, based on the same methodology and GWAS dataset we had previously used in a similar analysis of the smaller sample of isolated CPO²⁷. As with the CPO study, the current analyses benefitted from being

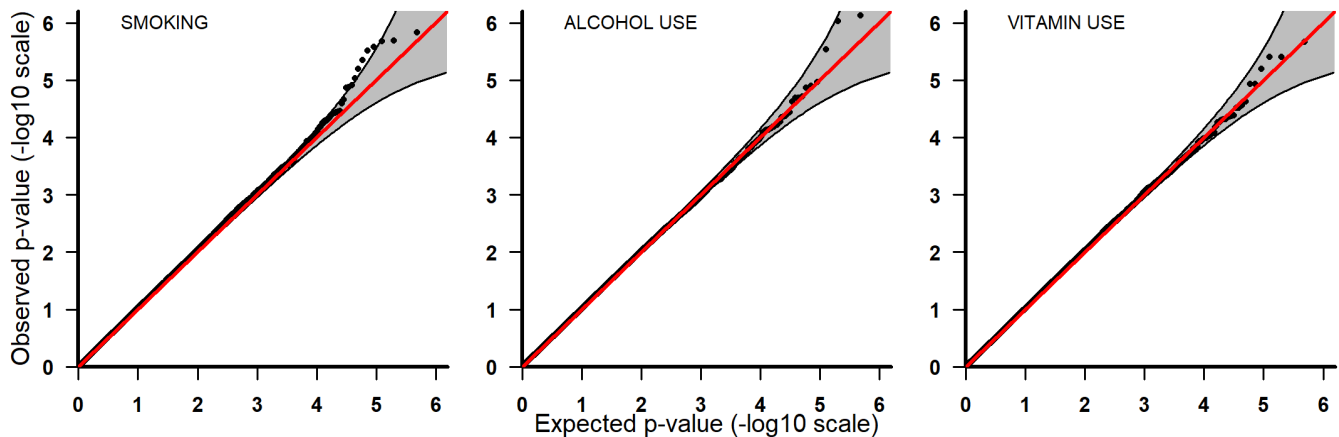


Figure 7. European analyses. Q-Q plots for PoOxSmoke (left) PoOxAlcohol (middle) and PoOxVitamin (right) with 95% pointwise confidence bands. Q-Q, quantile-quantile; PoOxAlcohol, parent-of-origin interactions with alcohol; PoOxSmoke, parent-of-origin interactions with smoking; PoOxVitamin, parent-of-origin interactions with vitamins.

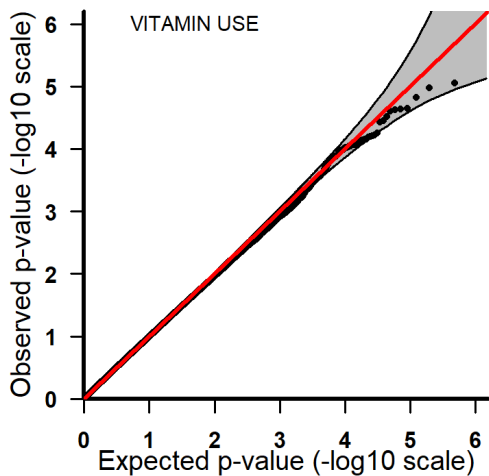


Figure 8. Asian analyses. Q-Q plots for PoOxVitamin with 95% pointwise confidence bands. Q-Q, quantile-quantile; PoOxVitamin, parent-of-origin interactions with vitamins.

based on the largest available GWAS dataset of case-parent triads of orofacial clefts to date. Moreover, data were available for two major ethnicities, European and Asian, which is useful in assessing the generalizability of the findings across different ethnic groups. In the current dataset, however, very few of the Asian mothers reported smoking cigarettes or consuming alcohol during the periconceptional period, thus preventing a comparison of PoOxE effects for these exposures across these two ethnic groups. This is a common impediment to GxE studies, where the number of exposed individuals needs to be large enough for a meaningful analysis⁵³.

A possible mechanism for a PoO effect is genomic imprinting²¹. This occurs when DNA methylation in the germline causes the expression of alleles to be silenced depending on their parental origin. Maternal environmental exposures that affect

methylation patterns may also affect paternally and maternally inherited alleles differently. Furthermore, a PoO effect that is not affected by an environmental exposure will not be detected in our PoOxE analysis. Hence, a detected PoOxE effect may have a better chance than a PoO or GxE effect of uncovering a true causal relationship involving genomic imprinting.

Relying on the q-values for assessing the false positive rate, our analyses detected possible PoOxSmoke effects with SNPs in *LYZLI*, *ANK3*, *PDGFD*, *FOCAD* and *FRAS1*, and possible PoOxAlcohol effects with two SNPs in *ARHGEF10*. Without formal validation in a comparable and independent replication cohort, it would be premature to accept these associations as true PoOxE effects. Not having previously been linked with orofacial clefts does not necessarily imply that the identified gene is not relevant for clefting. This applies to several genes in our analyses; for example, *SHKBP1* and *MAPK10* in the PoOxSmoke analysis of the European sample and *TJP3* in the PoOxVitamin analysis of the pooled sample. The current study was primed to explore new hypotheses for disease mechanisms and to provide as many of the results as possible so that other researchers with access to similar GWAS datasets would be able to validate the findings presented here. To avoid being overly stringent, we thus presented all the results for the top 20 SNPs in [Table 3–Table 5](#).

Despite an exhaustive literature search, we were unable to find any obvious evidence linking *ANK3* and orofacial clefts. The Hetionet results confirmed this lack of a direct connection ([Figure 1](#)). *ANK3* encodes a member of the Ankyrin family of proteins, whose function is to bind the integral membrane proteins to the spectrin-actin cytoskeleton. This is important for cell motility, activation, proliferation, contact and the maintenance of specialized membrane domains; cellular activities that are also relevant for the proper development of craniofacial structures. For example, Stankewich and colleagues⁵⁴ showed that the spectrin–ankyrin scaffold is important for cell migration,

tissue patterning and organogenesis. Homozygous deletion of the gene encoding α II-spectrin in mice (*Spna2*) resulted in craniofacial, neural tube and cardiac anomalies, in addition to retarded intrauterine growth. **Figure 3** indicates that several SNPs in *ANK3* are potentially associated with clefting.

Like *ANK3*, *ARHGEF10* has not previously been associated with orofacial clefts, and the Hetionet results are consistent with this observation (**Figure 2**). *ARHGEF10* encodes a Rho guanine nucleotide exchange factor that may be involved in neural morphogenesis⁵⁵. *LYZL1* belongs to the family of lysozyme-like proteins that are implicated in sperm function and innate immunity⁵⁶. According to GeneCards (GCID: [GC09P020659](#)), *FOCAD* encodes a tumor suppressor gene that is highly expressed in the brain. It has also been linked to Alzheimer's disease⁵⁷. Furthermore, germline deletions in *FOCAD* are associated with polyposis and colorectal cancer⁵⁸. Again, as with *ANK3* and *ARHGEF10* above, there do not seem to be any obvious connections between *LYZL1* or *FOCAD* with clefting.

In contrast to the above genes, *FRASI* and several members of the platelet-derived growth factor (PDGF) gene family are known to be implicated in orofacial clefts. *PDGFD* is a member of the PDGF gene family and plays a central role in the PDGF receptor-alpha (PDGFR- α) signaling pathway. More specifically, disruption of *Pdgf* signaling results in clefting of the palate⁵⁹. *FRASI* (GCID: [GC04P078056](#)) encodes an extracellular matrix protein that plays a critical role in epithelial-mesenchymal interactions during embryonic development⁶⁰. Loss-of-function mutations in *FRASI* underlie Fraser syndrome, which is characterized by craniofacial, urogenital and respiratory system abnormalities⁶¹. Both of these genes are therefore worthy of further investigations in other isolated orofacial cleft cohorts.

A limitation of this study is that genotypes were not imputed. To avoid Mendelian inconsistencies, the imputation procedure would have had to account for the full triads, as opposed to imputing each sample independently, which has not been done for our data. Instead we conducted *post hoc* haplotype analyses for combinations of top SNPs located close to each other. This is akin to imputation, in that such an analysis takes into account information about a whole area of DNA, instead of just one SNP. **Table 7** shows that in the PoOxSmoke analysis of the two-SNP combination rs3793861-rs7087489 in *ANK3*, the p-value was slightly lower and the RRR slightly higher than in the corresponding analyses of each individual SNP. A similar pattern was observed in the PoOxAlcohol analyses of the rs2294035-rs4876274 haplotype in *ARHGEF10*. This indicates that the two-SNP combinations may be driving the effects observed with the individual SNPs.

The genes *PDGFD*, *CSMD1* and *RSUI* detected here had previously showed up in a study focusing on identifying GxE effects in the same CL/P triads²⁸. In that study, a possible GxVitamin effect was detected with *PDGFD* and *RSUI*, and a possible GxAlcohol effect was detected with *CSMD1*. In the current study, a PoOxSmoke effect was detected with *PDGFD* and PoOxVitamin effects were detected with *RSUI* and *CSMD1*. In

other words, only *RSUI* had the same exposure (vitamin) across the studies. *RSUI* stands for 'Ras suppressor protein 1' and is localized to chromosome 10p13. Its protein product is found at cell-extracellular matrix adhesion sites and has been reported to be involved in suppressing v-Ras transformation in the Ras signal transduction pathway⁶². *CSMD1* stands for 'CUB and Sushi multiple domains 1' and is localized to chromosome 8p23.2 (GCID: [GC08M002953](#)). It is involved in tumor suppression, as it has frequently been found to be deleted in many types of cancers^{63,64}. Again, there does not seem to be any obvious connections to clefting.

None of the top SNPs identified in our previous study focusing on PoOxE effects in CPO triads overlapped with SNPs identified in this study of CL/P²⁷. This is consistent with the observation that CPO and CL/P are etiologically distinct, so that the lead SNPs may be subtype-specific and differ between the two conditions. However, we detected associations with the 'cytochrome P450 family 4 subfamily F member 3' gene (*CYP4F3* on chr 19p13.12) in the previous CPO analyses, and with the 'cytochrome P450 family 46 subfamily A member 1' gene (*CYP46A1* on chr 14q32.2) in the present study. These two genes are members of the cytochrome P450 superfamily of enzymes that are primarily found in liver cells and whose function is to catalyze many reactions involved in the biotransformation of xeno- and endobiotics, and the biosynthesis of cholesterol and lipids, among others⁶⁵. It is therefore not surprising that these genes would appear in an analysis focusing on smoking, alcohol and vitamin intake.

We searched Hetionet for indirect links between *ANK3* or *ARHGEF10* and cleft lip, as well as between *ANK3* or *ARHGEF10* and nicotine or alcohol dependence, respectively. This approach has several limitations. First, using "nicotine dependence" and "alcohol dependence" in lieu of the actual smoking and alcohol consumption status may introduce some bias. Second, Hetionet is built from a curated set of database information, which means that not all the information, especially the newest, would be available. However, when interpreting our results, we used the source databases to make sure that the connections between the nodes are reliable.

To conclude, our search for an interaction between a PoO-effect and an environmental exposure for CL/P identified possible relationships between SNPs in *ANK3* and maternal smoking, and SNPs in *ARHGEF10* and maternal intake of alcohol. There is a possibility that these interactions have a biological basis, although without replication they remain speculative. Our demonstration of the feasibility of identifying complex interactions between relevant environmental exposures and PoO-effects opens new possibilities in the search for the genetic etiology of CL/P.

Data availability

Underlying data

The **GWAS data** are available in the **dbGaP database**. Additional information regarding the inclusion/exclusion criteria of the study, the ethics statements, data variables, study history,

publications, and other documentation related to the study is provided on the dbGaP website.

The dbGaP database at the National Center for Biotechnology Information, U.S. National Library of Science (NCBI/NLM) provides an extensive overview of the cleft dataset used in this study. Entering the dbGaP accession number phs000094.v1.p1 provides access to information regarding the variables, study documents, and datasets. For example, detailed information about the mother's exposure to alcohol, vitamins, and smoke is provided under the header "Variable Selection". Information on study questionnaires, institutional review boards and consent forms from each participating cohort can be found under the header "Documents".

Controlled-access data from dbGaP is available only through the [dbGaP authorized access portal](#). There are separate procedures for accessing individual data, depending on whether the researcher is NIH-affiliated (intramural) or not (extramural). In addition, other restrictions apply; e.g., the principal investigator from the applying institution needs to be a permanently employed professor, senior scientist, or equivalent, to submit a data access request. A valid eRA Commons account for logging in to the dbGaP system is also mandatory.

Software availability

- Source code available from: <https://github.com/oeH041/A-genome-wide-scan-of-cleft-lip-triads-identifies-parent-of-origin-interaction-effects-between-ANK3->
- Archived source code at time of publication: <https://doi.org/10.5281/zenodo.3241319>³⁸
- License: CC-BY 4.0

Grant information

This research was supported by the Bergen Medical Research Foundation [807191], by the Research Council of Norway (RCN) through its Centres of Excellence funding scheme [grant 262700], and by the Biobank Norway II from the RCN [245464].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We extend our warmest gratitude to the families from so many different countries in the world who opted to contribute to this study. We are also grateful to everybody who contributed to the collection of the data, from recruiters in the field to technicians in the lab.

References

1. Mossey P, Castilla E: **Global registry and database on craniofacial anomalies.** In: World Health Organization, Geneva; 2003.
[Reference Source](#)
2. Berg E, Haaland ØA, Feragen KB, *et al.*: **Health Status Among Adults Born With an Oral Cleft in Norway.** *JAMA Pediatr.* 2016; **170**(11): 1063–1070.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Berg E, Sivertsen Å, Ariansen AM, *et al.*: **Socio-Economic Status and Reproduction among Adults Born with an Oral Cleft: A Population-Based Cohort Study in Norway.** *PLoS One.* 2016; **11**(9): e0162196.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Strauss RP: **The organization and delivery of craniofacial health Services: the state of the art.** *Cleft Palate Craniofac J.* 1999; **36**(3): 189–195.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Wehby GL, Cassell CH: **The impact of orofacial clefts on quality of life and healthcare use and costs.** *Oral Dis.* 2010; **16**(1): 3–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Rahimov F, Jugessur A, Murray JC: **Genetics of nonsyndromic orofacial clefts.** *Cleft Palate Craniofac J.* 2012; **49**(1): 73–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Dixon MJ, Marazita ML, Beaty TH, *et al.*: **Cleft lip and palate: understanding genetic and environmental influences.** *Nat Rev Genet.* 2011; **12**(3): 167–178.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Jugessur A, Farlie PG, Kilpatrick N: **The genetics of isolated orofacial clefts: from genotypes to subphenotypes.** *Oral Dis.* 2009; **15**(7): 437–453.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Marazita ML: **The evolution of human genetic studies of cleft lip and cleft palate.** *Annu Rev Genomics Hum Genet.* 2012; **13**: 263–283.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Kousa YA, Schutte BC: **Toward an orofacial gene regulatory network.** *Dev Dyn.* 2016; **245**(3): 220–232.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Beaty TH, Murray JC, Marazita ML, *et al.*: **Erratum: A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4.** *Nat Genet.* 2010; **42**(8): 727.
[Publisher Full Text](#)
12. Birnbaum S, Ludwig KU, Reutter H, *et al.*: **Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24.** *Nat Genet.* 2009; **41**(4): 473–477.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Camargo M, Rivera D, Moreno L, *et al.*: **GWAS reveals new recessive loci associated with non-syndromic facial clefting.** *Eur J Med Genet.* 2012; **55**(10): 510–514.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Grant SF, Wang K, Zhang H, *et al.*: **A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24.** *J Pediatr.* 2009; **155**(6): 909–913.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Mangold E, Ludwig KU, Birnbaum S, *et al.*: **Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate.** *Nat Genet.* 2010; **42**(1): 24–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Ludwig KU, Mangold E, Herms S, *et al.*: **Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci.** *Nat Genet.* 2012; **44**(9): 968–971.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Grosen D, Bille C, Pedersen JK, *et al.*: **Recurrence risk for offspring of twins discordant for oral cleft: a population-based cohort study of the Danish 1936–2004 cleft twin cohort.** *Am J Med Genet A.* 2010; **152A**(10): 2468–2474.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Grosen D, Chevrier C, Skytthe A, *et al.*: **A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance.** *J Med Genet.* 2010; **47**(3): 162–168.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Grosen D, Bille C, Petersen I, *et al.*: **Risk of oral clefts in twins.** *Epidemiology.* 2011; **22**(3): 313–319.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Sivertsen A, Wilcox AJ, Skjaerven R, *et al.*: **Familial risk of oral clefts by morphological type and severity: population based cohort study of first degree relatives.** *BMJ.* 2008; **336**(7641): 432–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Lawson HA, Cheverud JM, Wolf JB: **Genomic imprinting and parent-of-origin effects on complex traits.** *Nat Rev Genet.* 2013; **14**(9): 609–617.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Gjerdevik M, Haaland ØA, Romanowska J, *et al.*: **Parent-of-Origin-Environment Interactions in Case-Parent Triads With or Without Independent Controls.** *Genet Epidemiol.* 2017; **41**(7): 664–665.

23. Guilmatre A, Sharp AJ: **Parent of origin effects.** *Clin Genet.* 2012; **81**(3): 201–209. [PubMed Abstract](#) | [Publisher Full Text](#)
24. Peters J: **The role of genomic imprinting in biology and disease: an expanding view.** *Nat Rev Genet.* 2014; **15**(8): 517–530. [PubMed Abstract](#) | [Publisher Full Text](#)
25. Thomas D: **Gene–environment-wide association studies: emerging approaches.** *Nat Rev Genet.* 2010; **11**(4): 259–272. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Gjessing HK, Lie RT: **Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes.** *Ann Hum Genet.* 2006; **70**(Pt 3): 382–396. [PubMed Abstract](#) | [Publisher Full Text](#)
27. Haaland OA, Jugessur A, Gjerdevik M, et al.: **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. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Haaland OA, Lie RT, Romanowska J, et al.: **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. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Wigginton JE, Cutler DJ, Abecasis GR: **A note on exact tests of Hardy-Weinberg equilibrium.** *Am J Hum Genet.* 2005; **76**(5): 887–893. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. R Core Team: **R: A Language and Environment for Statistical Computing.** *R Foundation for Statistical Computing, Vienna, Austria.* 2017. [Reference Source](#)
31. Skare O, Jugessur A, Lie RT, et al.: **Application of a novel hybrid study design to explore gene–environment interactions in orofacial clefts.** *Ann Hum Genet.* 2012; **76**(3): 221–236. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A.* 2003; **100**(16): 9440–9445. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Storey JD, Bass AJ, Dabney A, et al.: **qvalue: Q-value estimation for false discovery rate control.** 2015. [Reference Source](#)
34. Amrhein V, Greenland S, McShane B: **Retire statistical significance.** *Nature.* 2019; **567**: 305–307. [Reference Source](#)
35. Wasserstein RL, Schirm AL, Lazar NA: **Moving to a World Beyond “p < 0.05”.** *Am Stat.* 2019; **73**: 1–19. [Publisher Full Text](#)
36. Gjerdevik M, Jugessur A, Haaland OA, et al.: **Haplin power analysis: a software module for power and sample size calculations in genetic association analyses of family triads and unrelated controls.** *BMC Bioinformatics.* 2019; **20**(1): 165. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Pruim RJ, Welch RP, Sanna S, et al.: **LocusZoom: regional visualization of genome-wide association scan results.** *Bioinformatics.* 2010; **26**(18): 2336–2337. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Haaland OA, Romanowska J: **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.** *Zenodo.* 2019. <http://www.doi.org/10.5281/zenodo.3241319>
39. Szklarczyk D, Franceschini A, Wyder S, et al.: **STRING v10: protein-protein interaction networks, integrated over the tree of life.** *Nucleic Acids Res.* 2015; **43**(Database issue): D447–452. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Petryszak R, Keays M, Tang YA, et al.: **Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants.** *Nucleic Acids Res.* 2016; **44**(D1): D746–752. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Komljenovic A, Roux J, Wollbrecht J, et al.: **BgeeDB, an R package for retrieval of curated expression datasets and for gene list expression localization enrichment tests [version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved].** *F1000Res.* 2016; **5**: 2748. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Himmelstein DS, Lizée A, Hessler C, et al.: **Systematic integration of biomedical knowledge prioritizes drugs for repurposing.** *eLife.* 2017; **6**: pii: e26726. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Beaty TH, Murray JC, Marazita ML, et al.: **A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4.** *Nat Genet.* 2010; **42**(6): 525–529. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Qi L, Ding Y: **Involvement of the CREB5 regulatory network in colorectal cancer metastasis.** *Yi Chuan.* 2014; **36**(7): 679–684. [PubMed Abstract](#)
45. Park S, Cui J, Yu W, et al.: **Differential activities and mechanisms of the four R-spondins in potentiating Wnt/β-catenin signaling.** *J Biol Chem.* 2018; **293**(25): 9759–9769. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Jin YR, Han XH, Taketo MM, et al.: **Wnt9b-dependent FGF signaling is crucial for outgrowth of the nasal and maxillary processes during upper jaw and lip development.** *Development.* 2012; **139**(10): 1821–1830. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. He F, Xiong W, Wang Y, et al.: **Epithelial Wnt/β-catenin signaling regulates palatal shelf fusion through regulation of Tgfβ3 expression.** *Dev Biol.* 2011; **350**(2): 511–519. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Hu X, Gao J, Liao Y, et al.: **Retinoic acid alters the proliferation and survival of the epithelium and mesenchyme and suppresses Wnt/β-catenin signaling in developing cleft palate.** *Cell Death Dis.* 2013; **4**: e898. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Jiang Z, Pan L, Chen X, et al.: **Wnt6 influences the viability of mouse embryonic palatal mesenchymal cells via the β-catenin pathway.** *Exp Ther Med.* 2017; **14**(6): 5339–5344. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Cross-Disorder Group of the Psychiatric Genomics Consortium: **Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.** *Lancet.* 2013; **381**(9875): 1371–1379. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Himmelstein DS, Pankov A: **Mining knowledge from MEDLINE articles and their indexed MeSH terms.** 2015. [Publisher Full Text](#)
52. Zhou XZ, Menche J, Barabasi AL, et al.: **Human symptoms-disease network.** *Nat Commun.* 2014; **5**: 4212. [PubMed Abstract](#) | [Publisher Full Text](#)
53. McAllister K, Mechanic LE, Amos C, et al.: **Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases.** *Am J Epidemiol.* 2017; **186**(7): 753–761. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. Stankewich MC, Cianci CD, Stabach PR, et al.: **Cell organization, growth, and neural and cardiac development require all-spectrin.** *J Cell Sci.* 2011; **124**(Pt 23): 3956–3966. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Verhoeven K, De Jonghe P, Van de Putte T, et al.: **Slowed conduction and thin myelination of peripheral nerves associated with mutant rho Guanine-nucleotide exchange factor 10.** *Am J Hum Genet.* 2003; **73**(4): 926–932. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Narmadha G, Yenugu S: **In Silico and Biochemical Characterization of Lysozyme-Like Proteins in the Rat.** *PLoS One.* 2016; **11**(9): e0161909. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Brockschmidt A, Trost D, Peterziel H, et al.: **KIAA1797/FOCAD encodes a novel focal adhesion protein with tumour suppressor function in gliomas.** *Brain.* 2012; **135**(Pt 4): 1027–1041. [PubMed Abstract](#) | [Publisher Full Text](#)
58. Weren RD, Venkatchalam R, Cazier JB, et al.: **Germline deletions in the tumour suppressor gene FOCAD are associated with polyposis and colorectal cancer development.** *J Pathol.* 2015; **236**(2): 155–164. [PubMed Abstract](#) | [Publisher Full Text](#)
59. Eberhart JK, He X, Swartz ME, et al.: **MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis.** *Nat Genet.* 2008; **40**(3): 290–298. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
60. Short K, Wiradajaja F, Smyth I: **Let's stick together: the role of the Fras1 and Frem proteins in epidermal adhesion.** *IUBMB Life.* 2007; **59**(7): 427–435. [PubMed Abstract](#) | [Publisher Full Text](#)
61. McGregor L, Makela V, Darling SM, et al.: **Fraser syndrome and mouse blebbed phenotype caused by mutations in FRAS1/Fras1 encoding a putative extracellular matrix protein.** *Nat Genet.* 2003; **34**(2): 203–208. [PubMed Abstract](#) | [Publisher Full Text](#)
62. Cutler ML, Bassin RH, Zanoni L, et al.: **Isolation of Rsp-1, a Novel Cdna Capable of Suppressing V-Ras Transformation.** *Mol Cell Biol.* 1992; **12**(9): 3750–3756. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
63. Ma CQ, Quesnelle KM, Sparano A, et al.: **Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues.** *Cancer Biol Ther.* 2009; **8**(10): 907–916. [PubMed Abstract](#) | [Publisher Full Text](#)
64. Kamal M, Shaaban AM, Zhang L, et al.: **Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma.** *Breast Cancer Res Treat.* 2010; **121**(3): 555–563. [PubMed Abstract](#) | [Publisher Full Text](#)
65. Nebert DW, Russell DW: **Clinical importance of the cytochromes P450.** *Lancet.* 2002; **360**(9340): 1155–1162. [PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 26 July 2019

<https://doi.org/10.5256/f1000research.21886.r51378>

© 2019 Stergiakouli E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Evie Stergiakouli 

School of Oral and Dental Sciences, University of Bristol, Bristol, UK

The authors have improved the manuscript substantially, although not having imputed their data is a significant limitation. I would encourage them to attempt this in the future.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetic epidemiology of childhood disorders

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 25 July 2019

<https://doi.org/10.5256/f1000research.21459.r50328>

© 2019 Pan Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Yongchu Pan 

Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University (NMU), Nanjing, China

The authors conducted a genome-wide interaction analysis with smoking, alcohol, and vitamin on cleft lip triads. The analysis was based on 1,908 nuclear families from dbGaP database with both genotyping and environmental exposure status. Their results suggested that three SNPs in *ANK3* has a PoOxSmoke effect and two SNPs in *ARHGEF10* has a PoOxAlcohol effect. The results may provide evidence for

identifying interactions between relevant environmental exposures and PoO effects. Totally, this paper was very well written; however, there are some concerns that should be addressed to improve the paper as detailed below:

1. The GWAS data are available in the dbGaP database (the dbGaP accession number: phs000094.v1.p1). Since the imputed data were provided from dbGaP, it is more appropriate to use that data for interaction analysis to find more meaningful results in this study.
2. It seems the author only presented the process on quality control for SNPs, and the process on quality control for individuals should be added in the part of Methods.
3. In the present study, the author focused the q-value rather than p-value, which is more widely accepted as known. It is confused that the cut off of q-value chosen in each interaction analysis on different population is unclear.
4. It is unclear that the cutoff of q-value in each interaction analysis on different population. The author needs to determine a clear threshold for screening, which should be more strict.
5. There are a lot of tables in the article. It is recommended to add the forest plots of the interaction analysis in order to make the results more intuitive.
6. I am not sure if it is appropriate to state 'After quality control, 341,191 SNPs remained from the original 569,244.' in Methods section from abstract, because that seems to be part of the result.
7. All the population and analysis were based on nuclear families with cleft lip with/without cleft palate, it is proper to replace 'cleft lip' with 'cleft lip with/without cleft palate' in the title of this paper.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics of craniofacial diseases

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 04 July 2019

<https://doi.org/10.5256/f1000research.21459.r50329>

© 2019 Stergiakouli E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Evie Stergiakouli 

School of Oral and Dental Sciences, University of Bristol, Bristol, UK

This is a very interesting paper focusing on identifying parent of origin interaction effects in CL/P using a genome-wide approach. The environmental factors selected are biologically plausible to be involved in cleft and the sample, although modest in size, involved full trios and individuals of non-European ancestry. The authors identify suggestive evidence of interaction for 3 SNPs in ANK3 with smoking and 10 SNPs for ARHGEF10 with alcohol consumption. This study provides proof of principle for their approach and the authors highlight the need for replication in independent samples. The authors follow the results with bioinformatics analysis and the paper is well written and detailed.

My main concern with the statistical analysis is with the use of directly genotyped SNPs in a genome-wide scan. Can the authors explain why there is no imputation of the directly genotyped SNPs? Imputation would have provided a much larger number of SNPs to work with.

Also, what is the rationale for not using the established genome-wide significance level (5×10^{-8})? If anything, the authors should have been more stringent than that since they effectively performed 3 genome-wide analyses (one for each environmental factor).

The abstract should be made for concise: The 59 individuals of other ancestries were not used for any analyses, so they should not be mentioned in the abstract. The methods section of the abstract does not contain any description of how they performed the analysis and relies heavily on their previous study.

I am not sure if the following statement should be called intriguing: 'Despite this success, the genetic variants identified so far collectively explain only a minor fraction of the total variance attributable to additive genetic effects, which is intriguing considering the more than 70% heritability of CL/P among Europeans 17–20'. This is not unusual, and it is the case for many other complex disorder and traits, so I would suggest rephrasing.

The table titles should include more information to make the paper easier to follow. For example, Table 3 should include that it is a PoOxE analysis.

Discussion: The authors need to highlight that none of their findings reached genome-wide significance level and given that they have performed more tests than in a standard genome-wide study, the evidence for association is weak. So, their findings require replication and/or meta-analysis in a larger sample. I am not sure I agree with their statement: 'Without formal validation in a comparable and independent

replication cohort, it would be premature to dismiss the veracity of the remaining associations merely on the basis of their having too high q-values.' All of their findings need to be replicated and should be treated with caution until then.

Another point that I need to make is that I would have expected the discussion to be more about result interpretation and even trying to suggest possible mechanisms. At the moment the discussion is focused on detailed bioinformatics analysis which although interesting, would be more suited to the results section.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetic epidemiology of childhood disorders

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 09 Jul 2019

Øystein Ariansen Haaland, University of Bergen, Bergen, Norway

We would like to thank Dr. Stergiakouli for her insightful comments, and have addressed her concerns below. We have also made minor changes throughout the manuscript to correct grammatical errors or improve clarity. These are not addressed below, but are highlighted in the new version of the manuscript.

Comment:

My main concern with the statistical analysis is with the use of directly genotyped SNPs in a genome-wide scan. Can the authors explain why there is no imputation of the directly genotyped SNPs? Imputation would have provided a much larger number of SNPs to work with.

Reply:

This is a very good observation, and something which we should have addressed in the text.

Added to Discussion:

“A limitation of this study is that genotypes were not imputed. To avoid Mendelian inconsistencies, the imputation procedure would have had to account for the full triads, as opposed to imputing each sample independently, which has not been done for our data. Instead we conducted *post hoc* haplotype analyses for combinations of top SNPs located close to each other. This is akin to imputation, in that such an analysis takes into account information about a whole area of DNA, instead of just one SNP.”

Comment:

Also, what is the rationale for not using the established genome-wide significance level (5×10^{-8})? If anything, the authors should have been more stringent than that since they effectively performed 3 genome-wide analyses (one for each environmental factor).

Reply:

In recent years, more and more statisticians have argued for a move away from the term “statistically significant” (e.g., see Amrhein, Greenland & McShane (2019). Retire statistical significance. *Nature* **567**, 305-307 doi: 10.1038/d41586-019-00857-9, and Ronald L. Wasserstein & Nicole A. Lazar (2019) Moving to a World Beyond “ $p < 0.05$ ”. *Am Stat* **73**, 1-19 doi: 10.1080/00031305.2019.1583913), which is why we have avoided using a fixed p-value as a threshold for significance. Therefore, we focused on the q-value. However, for transparency we still report p-values, as well as effect estimates with 95% confidence intervals.

Added to Methods:

“In accordance with recent recommendations by the American Statistical Association and others^{34,35}, we did not consider a fixed p-value as a threshold for statistical significance.”

Comment:

The abstract should be made for concise: The 59 individuals of other ancestries were not used for any analyses, so they should not be mentioned in the abstract.

Reply:

We mention these 59 individuals because they were part of the pooled analyses, although they were not analyzed as a group. This is perhaps not clear from the text. In response, we have added a sentence to clarify this in the text.

Added to Methods (in the paper, not the abstract):

“The 59 families with other ethnicities were not analyzed as a group due to the small sample size, but they were included in the pooled analysis.”

Comment:

The methods section of the abstract does not contain any description of how they performed the analysis and relies heavily on their previous study.

Reply:

Thank you for pointing this out.

Text changed in Abstract:

“We recently developed new methods to identify parent-of-origin (PoO) interactions with environmental exposures (PoOxE) and now apply them to data from a genome-wide association study (GWAS) of families with children born with isolated CL/P.”

“The methodology applied in the analyses detects if PoO effects are different across environmental strata, and is implemented in the *R*-package Haplin.”

Comment:

I am not sure if the following statement should be called intriguing: 'Despite this success, the genetic variants identified so far collectively explain only a minor fraction of the total variance attributable to additive genetic effects, which is intriguing considering the more than 70% heritability of CL/P among Europeans^{17–20}'. This is not unusual, and it is the case for many other complex disorder and traits, so I would suggest rephrasing.

Reply:

Thank you for pointing this out.

Text changed in Introduction:

“Despite this success, the genetic variants identified so far collectively explain only a minor fraction of the total variance attributable to additive genetic effects, even though the heritability of CL/P among Europeans is more than 70%¹⁷⁻²⁰.”

Comment:

The table titles should include more information to make the paper easier to follow. For example, Table 3 should include that it is a PoOxE analysis.

Reply:

Thank you for this comment.

Title changed in Table 3:

“The top 20 single-nucleotide polymorphisms (SNPs) sorted by p-value in the pooled PoOxE analysis.”

Title changed in Table 4:

“The top 20 single-nucleotide polymorphisms (SNPs) sorted by p-value in the European PoOxE analysis.”

Title changed in Table 7:

“Stratified analyses of the top single-nucleotide polymorphisms (SNPs) and haplotypes in *ANK3* (PoOxSmoke) and *ARHGEF10* (PoOxAlcohol).”

Comment:

Discussion: The authors need to highlight that none of their findings reached genome-wide

significance level and given that they have performed more tests than in a standard genome-wide study, the evidence for association is weak. So, their findings require replication and/or meta-analysis in a larger sample. I am not sure I agree with their statement: 'Without formal validation in a comparable and independent replication cohort, it would be premature to dismiss the veracity of the remaining associations merely on the basis of their having too high q-values.' All of their findings need to be replicated and should be treated with caution until then.

Reply:

Thank you for pointing this out.

Text changed in Discussion:

"Relying on the q-values for assessing the false positive rate, our analyses detected possible PoOxSmoke effects with SNPs in *LYZL1*, *ANK3*, *PDGFD*, *FOCAD* and *FRAS1*, and possible PoOxAlcohol effects with two SNPs in *ARHGEF10*. Without formal validation in a comparable and independent replication cohort, it would be premature to accept these associations as true PoOxE effects."

Comment:

Another point that I need to make is that I would have expected the discussion to be more about result interpretation and even trying to suggest possible mechanisms. At the moment the discussion is focused on detailed bioinformatics analysis which although interesting, would be more suited to the results section.

Reply:

Thank you for this suggestion.

Added to Discussion:

"A possible mechanism for a PoO effect is genomic imprinting²¹. This occurs when DNA methylation in the germline causes the expression of alleles to be silenced depending on their parental origin. Maternal environmental exposures that affect methylation patterns may also affect paternally and maternally inherited alleles differently. Furthermore, a PoO effect that is not affected by an environmental exposure will not be detected in our PoOxE analysis. Hence, a detected PoOxE effect may have a better chance than a PoO or GxE effect of uncovering a true causal relationship involving genomic imprinting."

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research