ASSOCIATION STUDIES ARTICLE

Consortium-based genome-wide meta-analysis for childhood dental caries traits

Simon Haworth1,*†, Dmitry Shungin2,3,†, Justin T. van der Tas4, Strahinja Vucic4, Carolina Medina-Gomez5,6,7, Victor Yakimov8, Bjarke Feenstra9, John R. Shaffer9,10, Myoung Keun Lee10, Marie Standl11, Elisabeth Thiering11,12, Carol Wang13, Klaus Bønnelykke14, Johannes Waage14, Leon Eyrich Jessen14, Pia Elisabeth Nørrisgaard14, Raimo Joro15, Ilkka Seppälä16, Olli Raitakari17,18, Tom Dudding1, Olja Grgic4,5, Edwin Ongkosuwito5, Anu Vierola15, Aino-Maija Eloranta15, Nicola X. West19, Steven J. Thomas19, Daniel W. McNeil20, Steven M. Levy21, Rebecca Slayton22, Ellen A. Nohr23, Terho Lehtimäki16, Timo Lakka15,24,25, Hans Bisgaard14, Craig Pennell13, Jan Kühnisch26, Mary L. Marazita9,10, Mads Melbye8,27,28, Frank Geller8, Fernando Rivadeneira5,6,7, Eppo B. Wolvius4, Paul W. Franks29,30,31, Ingegerd Johansson2 and Nicholas J. Timpson1

1Medical Research Council Integrative Epidemiology Unit at Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK, 2Department of Odontology, Umeå University, Umeå 901 87, Sweden, 3Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA, 4Department of Oral and Maxillofacial Surgery, Special Dental Care and Orthodontics, 5The Generation R Study Group, 6Department of Internal Medicine, 7Department of Epidemiology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam 3015 CN, The Netherlands, 8Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark, 9Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA, 10Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA, 11Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg D-85764, Germany, 12Division of Metabolic and Nutritional Medicine, Dr von Hauner Children’s Hospital, University of Munich Medical Center, Munich 80337, Germany, 13Division of Obstetrics and Gynaecology, The University of Western Australia, Perth WA 6009, Australia, 14COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen 2730, Denmark, 15Institute of Biomedicine, School of Medicine, University of Eastern Finland

*The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

Received: March 1, 2018. Revised: May 29, 2018. Accepted: June 14, 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
important of genetic susceptibility for dental caries experience
preciation for the role of genetic influences in dental caries. The
rious (6–8). Over the last decades there has been increasing ap-
oral hygiene and low socio-economic status are the most noto-
toward dental caries, of which high sugar consumption, poor
(2–4). Dental caries results from reduced mineral satura-
nations (2–4). Dental caries results from reduced mineral satura-
was demonstrated in an animal model over 50 years ago, a find-
ing since substantiated in twin studies in humans (9–11). Of par-
relevance to caries traits in children and adolescents, Bretz et al. (10) analysed longitudinal rates of change in caries
status in children, and found that caries progression and sever-
ity were highly heritable in the primary and permanent
dentin. It has also been suggested that heritability for dental
caries does not depend entirely on genetic predisposition to
sweet food consumption (12). Despite evidence of a genetic con-
tribution to caries susceptibility, few specific genetic loci have
been identified.

Shaffer et al. (13) performed the first GWAS for dental caries
in 2011, studying the primary dentition of 1305 children. They
found evidence for association at novel and previously studied
candidate genes (ACTN2, MTR, EDARADD, MPPED2 and LPO), but
no individual single-nucleotide polymorphisms (SNPs) exceeded the genome-wide significance threshold ($P < 5.0e-08$), possibly as a consequence of the modest sample size (13). The first GWAS for dental caries in the permanent dentition in adults was performed at a similar time by Wang et al. (14). They included 7443 adults from five different cohorts and identified several suggestive loci ($P$-value $\leq 10e-05$) for dental caries (RPS6KA2, PTK2B, RHOU, FZD1, ADMTS3 and ISL1), different loci from those mentioned above for the primary dentition and again with no single variants reaching genome-wide significance.

The next wave of GWAS of caries suggested association at a range of different loci. Two GWAS used separate phenotype definitions for pit-and-fissure and smooth tooth surfaces and identified different loci associated with dental caries susceptibility in both primary and permanent dentition (15,16). The GWAS in primary dentition used a sample of approximately 1000 children and found evidence for association at loci reported in previous studies, including MMPED2, RPS6KA2 and AJAP1 (13–16). The largest GWAS for dental caries in permanent dentition was performed in a Hispanic and Latino sample of 11754 adults (17). This study identified unique genetic loci (NAMPT and BMP7) compared with previous GWAS in individuals of European ancestry. To date, it is unclear whether the variability in nominated loci reflects true variability in the genetic architecture of dental caries across different populations, age periods and sub-phenotypic definitions, or merely represent chance differences between studies given the modest power in the studies performed to date.

Dental caries is a complex and multifactorial disease, caused by a complex interplay between environmental, behavioural and genetic factors. Until now there has been a lack of large-scale studies of dental caries traits in children and the genetic basis of these traits remains poorly characterized. This investigation set out to examine the hypothesis that common genetic variants influence dental caries with modest effects on susceptibility. We anticipated that (a) caries in both primary and permanent teeth would be heritable in children and adolescents aged 2.5–18 years and (b) common genetic variants are likely to only have small effects on the susceptibility of a complex disease such as dental caries. Therefore, the aim of this large-scale, consortium-based GWAS is to examine novel genetic loci associated with dental caries in primary and permanent dentition in children and adolescents.

**Results**

**Single variant results**

Meta-analysis of caries in primary teeth in individuals of European ancestry included 17037 individuals (6922 affected) from 22 results files representing all nine coordinating centres. After final quality control (QC), this meta-analysis included 8 640 819 variants, with mild deflation (genomic inflation factor, $\lambda = 0.994$) (Supplementary Material, Fig. S1). Meta-analysis of caries in primary teeth which included individuals of multiple ethnicities in the Generation R (GENR) study included 19003 individuals (7530 affected) from 22 results files representing all 9 coordinating centres. There were 8 699 928 variants after final QC, with mild deflation in summary statistics ($\lambda = 0.986$) (Supplementary Material, Fig. S2). Analysis of caries status in permanent teeth included 13353 individuals (5875 affected) from 14 results files representing 7 coordinating centres. The sample size was smaller for permanent teeth as two coordinating centres did not have phenotype data for permanent teeth (RAIN and GENR), whilst the COPSAC group only had data for participants in the earlier birth cohort (COPSAC 2000). There were 8734121 variants after final QC, with mild deflation in summary statistics ($\lambda = 0.999$) (Supplementary Material, Fig. S3).

The strongest evidence for association with caries in primary teeth was seen at rs1594318 (odds ratio (OR) 0.85 for C allele, EAF 0.60, $P = 4.13e-08$) in the European ancestry meta-analysis (Figs 1, 2 and 3, Table 1). This variant is intronic within ALLC on 2p25, a locus which has not previously been reported for dental caries traits. In the meta-analysis combining individuals of all ancestries this variant no longer reached genome-wide significance, although suggestive evidence persisted at rs1594318 (OR 0.868 for C allele EAF 0.60, $P = 3.78e-07$) and other intronic variants within ALLC in high linkage disequilibrium (LD) (Fig. 3). For the permanent dentition the strongest statistical evidence for association was seen between caries status and rs7738851 (OR 1.28 for A allele, EAF 0.85, $P = 1.63e-08$) (Figs 1, 2 and 4, Table 1). This variant is intronic within NEDD9 on 6p24.

**Estimated heritability**

Using participant level data in ALSPAC heritability was estimated at 0.28 (95% CI 0.09: 0.48) and 0.17 (95% CI 0.02: 0.31) for primary and permanent teeth, respectively. Using summary statistics at the meta-analysis level produced point estimates near zero heritability, with wide confidence intervals (Table 2).

**Cross-phenotype comparisons**

Genome-wide mean chi-squared was too low to undertake genome-wide genetic correlation using the linkage disequilibrium score regression (LDSR) method for caries in either primary or permanent teeth. Hypothesis-free phenotype-wide lookup for rs1594318 included 885 GWAS where either rs1594318 or a proxy with $r^2 > 0.8$ was present. None of these traits showed evidence of association with rs1594318 at a Bonferroni-corrected alpha of 0.05. Lookup of rs7738851 and its proxies was performed against 662 traits, where similarly no traits reached a Bonferroni-corrected threshold. Hypothesis-driven lookup in adult caries traits revealed no strong evidence for persistent genetic effects into adulthood (Table 3).

**Gene prioritization, gene set enrichment and association with predicted gene transcription**

Gene-based tests identified association between caries status in the primary dentition and a region of 7q35 containing TCAF1, OR2F2 and OR2F1 ($P = 1.91e-06$, 1.58e-06 and 1.29e-06, respectively). There were insufficient independently associated loci to perform gene set enrichment analysis using DEPICT for either of the principal meta-analyses. Association with predicted gene transcription was tested but no genes met the threshold for association after accounting for multiple testing. The single greatest evidence for association was seen between increased predicted transcription of CDK5RAP3 and increased liability for permanent caries ($P = 3.94e-05$). CDK5RAP3 is known to interact with PAK4 and p14ARF, with a potential role in oncogenesis (18,19).

**Discussion**

Dental caries in children and adolescents has not been studied to date using a large-scale, consortium-based genome-wide
A meta-analysis approach. Based on previous knowledge of the heritability of caries in young populations and from our understanding of other complex diseases, we anticipated that common genetic variants would be associated with dental caries risk with consistent effects across different cohorts. We found evidence for association between rs1594318 and caries risk in primary teeth. This variant showed weaker evidence for association in the multi-ethnic meta-analysis, potentially relating to

Figure 1. Manhattan plots for each principal meta-analysis. (A) Caries in primary teeth (European ancestry), n samples = 17 036, n variants = 8 640 819, k = 0.9944. Variants within 500Kb of rs1594318 are highlighted in green. (B) Caries in primary teeth (multi-ethnic analysis), n samples = 19 003, n variants = 8 699 928, k = 0.9861. (C) Caries in permanent teeth (European ancestry), n samples = 13 353, n variants = 8 734 121, k = 0.9991. Variants within 500Kb of rs7738851 are highlighted in green.
different allele frequencies across the different ethnic groups included in analysis. Frequency of the G allele is reported to vary between 0.24 in Asian populations and 0.42 in populations of European ancestry based on 1KGP allele frequencies.

**ALLC** (Allantoicase) codes the enzyme allantoicase, which is involved in purine metabolism and whose enzymatic activity is believed to have been lost during vertebrate evolution. Mouse studies suggest that this loss of activity relates to low expression levels and low substrate affinity rather than total non-functionality (20). Although there is some evidence that **ALLC** polymorphisms are associated with response to asthma treatment (21), there is limited understanding of the implications of variation in **ALLC** for human health, and it is possible that rs1594318 tags functionality elsewhere in the same locus.

For permanent teeth, we found evidence for association between caries status and rs7738851, an intronic variant with **NEDD9** (neural precursor cell-expressed, developmentally down-regulated gene 9). **NEDD9** is reported to mediate integrin-initiated signal transduction pathways and is conserved from gnathostomes into mammals (22,23). **NEDD9** appears to play a number of functional roles in disease and normal development, including regulation of neuronal differentiation, development and migration (22,24-28). One such function involves regulation of neural crest cell migration (26). Disruption of neural crest signalling is known to lead to enamel and dentin defects in animal models (29,30) and might provide a mechanism for variation at rs7738851 to influence dental caries susceptibility.

Traditionally, risk assessment for dental caries in childhood has concentrated on dietary behaviours and other modifiable risk factors (31), with little focus on tooth quality. Although our understanding of the genetic risk factors for dental caries is incomplete, authors have noted that the evidence from previous genetic association studies tends to support a role for innate tooth structure and quality in risk of caries (32,33). If validated by future studies, the association with rs7738851 would provide further evidence for this argument, and may in the future enhance risk assessment in clinical practice.
The lookup of lead associated variants against adult caries traits provided no strong evidence for persistent association in adulthood. This might imply genetic effects which are specific to the near-eruption timepoint. An alternative explanation is that the variants identified in the present study represent false positive signals as the statistical evidence presented is not irrefutable and there is no formal replication stage in our study; yet, we see good consistency of effects across studies.

The meta-analysis heritability estimates were lower than anticipated from either previous within-study heritability estimates (34) or the new within-study heritability estimates obtained for this analysis. There are several possible explanations for this phenomenon. First, the methods used in the present analysis are SNP based which consistently underestimate heritability of complex traits relative to twin and family studies (35). Second, meta-analysis heritability represents the heritability of genetic effects which are consistent across populations. In the event of genuine differences in genetic architecture of dental caries across strata of age, geography, environmental exposure or subtly different phenotypic meanings, the meta-heritability estimate is not the same conceptually as the weighted average of heritability within each study.

More intuitively, genetic influences might be important within populations with relatively similar environments but not determine much of the overall differences in risk when comparing groups of people in markedly different environments. This view is consistent with existing literature from family based and candidate gene association studies suggesting the genetic

Figure 3. Forest plot for rs1594318 and caries in primary teeth. Effect sizes are expressed on a log OR scale, grouped by geographical location. The summary estimate is from the fixed-effect meta-analysis of participants of European ancestry.
architecture of dental caries is complex with multiple interactions. For example, gene-sex interactions are reported which change in magnitude between the primary and permanent dentition (36), genetic variants may have heterogeneous effects on the primary and permanent dentition (37) and environmental exposures such as fluoride may interact with genetic effects (38). Finally, the aetiological relevance of specific microbiome groups appears to vary between different populations (39), suggesting genetic effects acting through the oral microbiome might also vary between populations. Unfortunately, this study lacks statistical power to perform meta-analyses stratified on these exposures, so does not resolve this particular question.

In line with any consortium-based approach, the need to harmonize analysis across different collections led to some compromises. The phenotypic definitions used in this study do not contain information on disease extent or severity. Loss of information in creating these definitions may have contributed to the low statistical power of analysis. Our motivation for using simple definitions was based on the facts that (a) case-control status simply represents a threshold level of an underlying continuum of disease risk, (b) simple binary classifications facilitate comparison of studies with different assessment protocols and population risks and (c) simple classifications have been used successfully in a range of complex phenotypes.

Between participating centres there are differences in characteristics such as age at participation, phenotypic assessment and differences in the environment (such as nutrition, oral hygiene and the oral microbiome) which might influence dental
caries or its treatment, as reflected in the wide range of caries prevalence between different study centres. Varying phenotypic characteristics do not necessarily result in heterogeneous genetic effects, as this variability may be uncorrelated with genetic effects. There was little evidence for heterogeneity in the top associated loci reported, however, the test for heterogeneity (I²) is limited by the small number of participating studies in meta-analysis (40) and wide confidence intervals for within-study genetic effect estimates. Given these limitations, it is possible that heterogeneity contributed to low study power and prevented more comprehensive single variant findings.

In the ALSPAC study we made extensive use of questionnaire derived data. This will systematically under-report true

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Variant</th>
<th>Position</th>
<th>Effect allele</th>
<th>Other allele</th>
<th>EAF (SE)</th>
<th>Beta (SE)</th>
<th>Odds ratio</th>
<th>P-value</th>
<th>N</th>
<th>I²</th>
<th>P-value for heterogeneity</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caries in primary teeth (European ancestry analysis)</td>
<td>rs1594318</td>
<td>chr2: 3733944</td>
<td>C</td>
<td>G</td>
<td>0.60</td>
<td>-0.165 (0.030)</td>
<td>0.848</td>
<td>4.13e-08</td>
<td>16 994</td>
<td>0.0</td>
<td>0.69</td>
<td>Intronic, ALLC</td>
</tr>
<tr>
<td>Caries in primary teeth (multi-ethnic analysis)</td>
<td>rs1594318</td>
<td>chr2: 3733944</td>
<td>C</td>
<td>G</td>
<td>0.60</td>
<td>-0.142 (0.028)</td>
<td>0.868</td>
<td>3.78e-07</td>
<td>18 960</td>
<td>0.0</td>
<td>0.61</td>
<td>Intronic, ALLC</td>
</tr>
<tr>
<td>Caries in primary teeth (multi-ethnic analysis)</td>
<td>rs872877</td>
<td>chr2: 3738526</td>
<td>A</td>
<td>G</td>
<td>0.59</td>
<td>-0.142 (0.028)</td>
<td>0.868</td>
<td>4.18e-07</td>
<td>18 985</td>
<td>17</td>
<td>0.68</td>
<td>Intronic, ALLC</td>
</tr>
<tr>
<td>Caries in permanent teeth</td>
<td>rs7738851</td>
<td>chr6: 11241788</td>
<td>A</td>
<td>T</td>
<td>0.85</td>
<td>0.248 (0.044)</td>
<td>1.28</td>
<td>1.63e-08</td>
<td>13 353</td>
<td>13</td>
<td>0.20</td>
<td>Intronic, NEDD9</td>
</tr>
</tbody>
</table>

*No single variants were associated with dental caries status at the genome-wide level in the multi-ethnic analysis of primary teeth, however two variants are discussed in Results section and are included here for reference.

Table 2. Within-sample and meta-analysis heritability estimates

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Method</th>
<th>Estimated h² (95% CI)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caries in primary teeth</td>
<td>GCTA GREML</td>
<td>0.28 (0.09: 0.48)</td>
<td>7230</td>
</tr>
<tr>
<td>Caries in primary teeth</td>
<td>LDSR</td>
<td>0.01 (0.00: 0.06)</td>
<td>19 003</td>
</tr>
<tr>
<td>Caries in permanent teeth</td>
<td>GCTA GREML</td>
<td>0.17 (0.02: 0.31)</td>
<td>6657</td>
</tr>
<tr>
<td>Caries in permanent teeth</td>
<td>LDSR</td>
<td>0.06 (0.00: 0.12)</td>
<td>13 353</td>
</tr>
</tbody>
</table>

Table 3. Lookup of lead associated variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>Discovery trait</th>
<th>Risk increasing allele (discovery)</th>
<th>Cross trait lookup</th>
<th>P-value</th>
<th>Effect per caries risk increasing allele (se)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1594318</td>
<td>Caries in primary teeth (European ancestry meta-analysis)</td>
<td>G</td>
<td>Adult caries traits</td>
<td>DMFS (standard deviation of residuals of caries-affected surfaces)</td>
<td>0.87</td>
<td>−0.0015 (0.0092)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of teeth (inverse normal transformed residuals)</td>
<td>0.60</td>
<td>0.0051 (0.0098)</td>
<td>27 947</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standardized DFS (inverse normal transformed residuals)</td>
<td>0.033</td>
<td>−0.0195 (0.0091)</td>
<td>26 532</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(No traits meeting threshold for multiple testing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7738851</td>
<td>Caries in permanent teeth</td>
<td>A</td>
<td>Hypothesis free</td>
<td>Adult caries traits</td>
<td>DMFS (standard deviation of residuals of caries-affected surfaces)</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of teeth (inverse normal transformed residuals)</td>
<td>0.63</td>
<td>−0.0064 (0.013)</td>
<td>27 949</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Standardized DFS (inverse normal transformed residuals)</td>
<td>0.65</td>
<td>−0.0054 (0.012)</td>
<td>26 531</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(No traits meeting threshold for multiple testing)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adult caries traits were defined as follows. DMFS—a count of the number of decayed, missing or filled tooth surfaces. This count was residualized after regression on age and age-squared and standard deviations of residuals calculated. Number of teeth—a count of the number of teeth in the mouth. This count was residualized after regression on age and age-squared and residuals underwent inverse normal transformation. Standardized DFS. The number of decayed and filled surfaces was divided by the total number of tooth surfaces in the mouth. This ratio was residualized after regression on age and age-squared and residuals underwent inverse normal transformation.

Caries or its treatment, as reflected in the wide range of caries prevalence between different study centres. Varying phenotypic characteristics do not necessarily result in heterogeneous genetic effects, as this variability may be uncorrelated with genetic effects. There was little evidence for heterogeneity in the top associated loci reported, however, the test for heterogeneity in genetic effects (I²) is limited by the small number of participating studies in meta-analysis (40) and wide confidence intervals for within-study genetic effect estimates. Given these limitations, it is possible that heterogeneity contributed to low study power and prevented more comprehensive single variant findings. In the ALSPAC study we made extensive use of questionnaire derived data. This will systematically under-report true
caries exposure compared with other studies as children or their parents are unlikely to be aware of untreated dental caries which would be evident to a trained assessor. We have explored some of these issues previously and shown that self-report measures at scale can be used to make meaningful inference about dental health in childhood (41). We believe that misclassification and under-reporting in questionnaire data would tend to bias genetic effect estimates and heritability toward the null. Despite this we show evidence for heritability using these definitions and effect sizes at lead variants are comparable with effect sizes obtained using clinically assessed data (Figs 3 and 4).

As our power calculations showed, the sample size was sufficient to detect the identified variants associated at a genome wide significant level with caries in the primary teeth (rs1594318) and in permanent teeth (rs872877), where we observed relatively large effect sizes. For smaller effect sizes we were underpowered to identify association, and did not detect any variants with effect sizes (expressed as per-allele increased odds) smaller than 15% or 17% in the primary and permanent teeth, respectively. Caries is highly influenced by environmental factors and it is likely that its susceptibility is polygenic in nature (32) with individual genetic variants conferring small effect sizes, as seen in other comparable complex traits (42). Furthermore, some of the included studies had major differences in their caries prevalence, likely acting as a proxy for features affecting risk of caries. This may have introduced heterogeneity and reduced power to detect association, as discussed further below.

One area of interest in the literature is the ability of genetics to guide personalized decisions on risk screening or identifying treatment modalities, and this is also true in dentistry. The genetic variants identified in this study are unlikely to be useful on their own in this context, given the modest effect sizes and low total heritability observed in our meta-analysis. We would suggest clinicians should continue to consider environment and aggregate genetic effects (e.g. knowledge of disease patterns of close relatives) rather than specific genetic variants at this moment in time. Nevertheless, the findings of our study contribute to a better understanding of the genetic and biological mechanisms underlying caries susceptibility.

Materials and Methods

Study samples

We performed genome-wide association (GWA) analysis for dental caries case/control status in a consortium including nine coordinating centres. Study procedures differed between these centres. We use the term ‘clinical dental assessment’ to mean that a child was examined in person, whether this was in a dental clinic or a study centre. We use the term ‘examiner’ to refer to a dental professional, and use the term ‘assessor’ to refer to an individual with training who is not a dental professional, for example a trained research nurse.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal birth cohort which recruited pregnant women living near Bristol, UK with an estimated delivery date between 1991 and 1992. Follow-up has included clinical assessment and questionnaires and is ongoing (43). A subset of children attended clinics including clinical dental assessment by a trained assessor at age 31, 43 and 61 months of age. Parents were asked to complete questionnaires about their children’s health regularly, including comprehensive questions at a mean age of 5.4 and 6.4 years. Parents and children were asked to complete questionnaires about oral health at a mean age of 7.5, 10.7 and 17.8 years. Please note that the study website contains details of all the data that are available through a freely searchable data dictionary (www.bristol.ac.uk/alspac/researchers/access; date last accessed June 2018). Both clinical and questionnaire derived data were included in this analysis, with priority given to clinical data were available (Supplementary Material, Table S3).

The Copenhagen Prospective Studies on Asthma in Childhood includes two population-based longitudinal birth cohorts in Eastern Denmark. COPSAC2000 recruited pregnant women with a history of asthma between 1998 and 2001 (44). Children who developed wheeze in early life were considered for enrolment in a nested randomized trial for asthma prevention. COPSAC2010 recruited pregnant women between 2008 and 2010 and was not selected on asthma status. Both COPSAC2000 and COPSAC2010 studies included regular clinical follow-up. Within Denmark clinical dental assessment is routinely offered to children and adolescents until the age of 18 years and summary data from these examinations are stored in a national register. These data were obtained via index linkage for participants of COPSAC2000 and COPSAC2010 and used to perform joint analysis across both cohorts.

The Danish National Birth Cohort (DNBC) is a longitudinal birth cohort which recruited women in mid-pregnancy from 1996 onwards (45). For this analysis, index linkage was performed to obtain childhood dental records for mothers participating in DNBC. As with the COPSAC studies, these data were originally obtained by a qualified dentist and included surface level dental charting.

The Generation R study (GENR) recruited women in early pregnancy with expected delivery dates between 2002 and 2006 living in the city of Rotterdam, the Netherlands. The cohort is multi-ethnic with representation from several non-European ethnic groups. Follow-up has included clinical assessment visits and questionnaires and is ongoing (46). Intra-oral photography was performed as a part of their study protocol, with surface level charting produced by a dental examiner (a specialist in paediatric dentistry) (47). Analysis in GENR included (a) a multi-ethnic association study including all individuals with genetic and phenotypic data (48) and (b) analysis including only individuals of European ancestry.

The GENEVA consortium is a group of studies which undertake coordinated analysis across several phenotypes (49). Within GENEVA, the Center for Oral Health Research in rural Appalachia, West Virginia and Pennsylvania, USA (COHRA), the Iowa Fluoride Study in Iowa, USA (IFS) and the Iowa Head Start (IHS) study participated in analysis of dental traits in children (15). COHRA recruited families with at least one child aged between 1 and 18 years of age, with dental examination performed at baseline (50). IFS recruited mothers and new-born infants in Iowa between 1992 and 1995 with a focus on longitudinal fluoride exposures and dental and bone health outcomes. Clinical dental examination in IFS was performed by trained assessors aged 5, 9, 13 and 17 years (51). IHS recruited children participating in an early childhood education program which included a one-time clinical dental examination (13).

The ‘German Infant study on the influence of Nutrition Intervention plus air pollution and genetics on allergy development’ (GINIplus) is a multi-centre prospective birth cohort study which has an observational and interventional arm which conducted a nutritional intervention during the first 4 months of life. The study recruited new born infants with and without family history of allergy in the Munich and Wesel areas, Germany between 1995 and 1998 (52,53). The ‘Lifestyle-related
factors, Immune System and the development of Allergies in East and West Germany’ study (LISA) is a longitudinal birth cohort which recruited between 1997 and 1999 across four sites in Germany (52,54). For participants living in the Munich area, follow-up used similar protocols in both GINIplus and LISA, with questionnaire and clinic data including clinical dental examination by trained examiners at age 10 and 15 years. Analysis for caries in GINIplus and LISA was therefore performed across both studies for participants at the Munich study centre.

The Physical Activity and Nutrition in Children (PANIC) Study is an ongoing controlled physical activity and dietary intervention study in a population of children followed retrospectively since pregnancy and prospectively until adolescence. Altogether 512 children 6–8 years of age were recruited in 2008–2009 (55). The main aims of the study are to investigate risk factors and pathophysiological mechanisms for overweight, type 2 diabetes, atherosclerotic cardiovascular diseases, musculoskeletal diseases, psychiatric disorders, dementia and oral health problems and the effects of a long-term physical activity and dietary intervention on these risk factors and pathophysiological mechanisms. Clinical dental examinations were performed by a qualified dentist with tooth level charting.

The Cardiovascular Risk in Young Finns Study (YFS) is a multi-centre investigation which aimed to understand the determinants of cardiovascular risk factors in young people in Finland. The study recruited participants who were aged 3, 6, 9, 12, 15 and 18 years old in 1980. Eligible participants living in specific regions of Finland were identified at random from a national population register and were invited to participate. Regular follow-up has been performed through physical examination and questionnaires (56). Clinical dental examination was performed by a qualified dentist with tooth level charting.

The Western Australian Pregnancy Cohort (RAINE) study is a birth cohort which recruited women between 16th and 20th week of pregnancy living in the Perth area, Western Australia. Recruitment occurred between 1989 and 1991 with regular follow-up of mothers and their children through research clinics and questionnaires (57). The presence or absence of dental caries was recorded by a trained assessor following clinical dental examination at the year 3 clinic follow-up.

Further details of study samples are provided in Supplementary Material, Table S1.

Medical Ethics
Within each participating study written informed consent was obtained from the parents of participating children after receiving a full explanation of the study. Children were invited to give assent where appropriate. All studies were conducted in accordance with the Declaration of Helsinki.

Ethical approval for the ALSPAC study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee. Full details of ethical approval policies and supporting documentation are available online (http://www.bris tol.ac.uk/alspac/researchers/research-ethics/, date last accessed June 2018). Approval to undertake analysis of caries traits was granted by the ALSPAC executive committee (B2356).

The COPSAC2000 cohort was approved by the Regional Scientific Ethical Committee for Copenhagen and Frederiksberg (KF 01-289/96) and the Danish Data Protection Agency (2008-41-1574). The 2010 cohort (COPSAC2010) was approved by the Danish Ethics Committee (H-B-2008-093) and the Danish Data Protection Agency (2008-41-2599).

The DNBC study of caries was approved by the Scientific Ethics Committee for the Capital City Region (Copenhagen), the Danish Data Protection Agency and the DNBC steering committee.

Each participating site in the GENEVA consortium caries analysis received approval from the local university institutional review board (federal wide assurance number for GENEVA caries project: FWA00006790). Within the COHRA arm local approval was provided by the University of Pittsburgh (020703/0506048) and West Virginia University (15620B), whilst the IFS and IHS arms received local approval from the University of Iowa’s Institutional Review Board.

The GENR study design and specific data acquisition were approved by the Medical Ethical Committee of the Erasmus University Medical Center, Rotterdam, The Netherlands (MEC-2007-413).

The GINIplus and LISA studies were approved by the ethics committee of the Bavarian Board of Physicians (10 year follow-up: 05100 for GINIplus and 07098 for LISA, 15 year follow-up 10090 for GINIplus, 12067 for LISA).

The PANIC study protocol was approved by the Research Ethics Committee of the Hospital District of Northern Savo. All participating children and their parents gave informed written consent.

The YFS study protocol was approved by local ethics committees for contributing sites.

The RAINE study was approved by the University of Western Australia Human Research Ethics Committee.

Phenotypes
Primary teeth exfoliate and are replaced by permanent teeth between 6 and 12 years of age. We aimed to separate caries status in primary and permanent teeth wherever possible using clinical information or age criteria, in line with our expectation that the genetic risk factors for dental caries might differ between primary and permanent dentition. For children in the mixed dentition we created two parallel case definitions, whilst in younger or older children a single case definition was sufficient.

All study samples included a mixture of children with dental caries and children who were caries-free, with varying degrees of within-mouth or within-tooth resolution. To facilitate comparison across these differing degrees of resolution all analysis compared children who were caries-free (unaffected) or had dental caries (affected). Missing teeth could represent exfoliation or delayed eruption rather than the endpoint of dental caries and therefore missing teeth were not included in classifying children as caries-free or caries affected.

In children aged 2.50 years to 5.99 years, any individual with 1 or more decayed or filled tooth was classified as caries affected, with all remaining individuals classified as unaffected. In children aged 6.00 years to 11.99 years of age, parallel definitions were determined for the primary dentition and permanent dentition, respectively. Any individual with at least 1 decayed or filled primary tooth was classified as caries affected for primary teeth, while all remaining participants were classified as unaffected. In parallel, any individual with at least 1 decayed or filled permanent tooth was classified as caries affected for permanent teeth, while all remaining individuals were classified as unaffected. In children and adolescents aged 12.00 to 17.99 years of age, any individual with 1 or more decayed or filled tooth or tooth surface (excluding third molar teeth) was classified as caries affected, with remaining individuals classified as unaffected.
Analysis was conducted in cross-section, meaning a single participant could only be represented in a single phenotype definition once. Where multiple sources of dental data were available for a single participant within a single phenotype definition window, the first source of data was selected (reflecting the youngest age at participation), in line with our expectation that caries status would be mostheritable in the near-eruption period.

The sources of data used to create these phenotypic definitions are given in Supplementary Material, Table S3. Within ALSPAC only, questionnaire responses were used to supplement data from clinical examination. The questions asked did not distinguish between primary and permanent teeth. Based on the age at questionnaire response we derived variables which prioritized responses from questionnaires before 6.00 years of age (thought to predominantly represent caries in primary teeth), and responses after 10.00 years of age (which might predominantly represent caries in permanent teeth). The final data sweep considered in this analysis targeted adolescents at age 17.50 years. Some participants responded to this after their 18th birthday. Data derived from this final questionnaire sweep were not included in the principal meta-analyses but were included in the GCTA heritability analysis.

**Genotypes and imputation**

All participating studies used genetic data imputed to a comprehensive imputation panel. The 1000 genomes phase 1 version 3 panel (1KG phase 1 v3) was used as a common basis across six centres (GINIplus/LISA, GENR, GENEVA, YFS, PANIC, RAINIE) (Supplementary Material, Table S1). In ALSPAC, DNBC, COPSAc2000 and COPSAc 2010 the haplotype reference consortium (HRC v1.0 and v1.1) imputation panels were used (Supplementary Material, Table S1).

Each study performed routine QC measures during genotyping, imputation and association testing (Supplementary Material, Table S2). Further pre-meta-analysis QC was performed centrally using the EasyQC R package and accompanying 1KG phase1 v3 reference data (S8). Minor allele count (MAC) was derived as the product of minor allele frequency (MAF) and site-specific number of alleles (twice the site-specific sample size). Variants were dropped which had a per-file MAC of 6 or lower, a site-specific sample size of 30 or lower, or an impute INFO score of less than 0.4. Sites which reported effect and non-effect alleles other than those reported in 1KG phase 1 v3 reference data were dropped. Following meta-analysis, sites with a weighted MAF of less than 0.005 were dropped, along with variants present in less than 50% of the total sample.

**Statistical analysis**

**Association testing**

Each cohort preformed GWA analysis using an additive genetic model. Caries status was modelled against genotype dosage whilst accounting for age at phenotypic assessment, age squared, sex and cryptic relatedness. Sex was accounted for by deriving phenotypic definitions and performing analysis separately within male and female participants, or by including sex as a covariate in association testing. Each study adopted approaches to account for cryptic relatedness and population stratification, as described in Supplementary Material, Table S2.

In the GENR study parallel analyses were conducted for participants of European ancestry (using the approach described in Supplementary Material, Table S2) and the entire study population, using a previously published method (48). The software and exact approach used by each study is shown in Supplementary Material, Table S2.

**Meta-analysis**

Results of GWA analysis within each study were combined in two principal meta-analyses, representing caries status in primary teeth and caries status in permanent teeth. For primary teeth, parallel meta-analyses were performed, one using results of multi-ethnic analysis in the GENR study and the other using results of European ancestry analysis in the GENR study. The GENR study did not have phenotypic data for permanent teeth, therefore the analysis of permanent teeth contained only individuals of European ancestry. Fixed-effects meta-analyses was performed using METAL (59), with genomic control of input summary statistics enabled and $I^2$ test for heterogeneity. Meta-analysis was run in parallel in two centres and results compared. All available studies with genotype and phenotypic information were included in a one-stage design, therefore there was no separate replication stage.

**Meta-analysis heritability estimates**

For each principal meta-analysis population stratification and heritability were assessed using LDSR (60). Reference LD scores were taken from HapMap3 reference data accompanying the LDSR package.

**Within-sample heritability estimates**

For comparison, heritability within the ALSPAC study was assessed using the GREML method (61), implemented in the GCTA software package (62), using participant level phenotype data and a genetic relatedness matrix estimated from common genetic variants (with MAF > 5.0%) present in HapMap3.

**Hypothesis-free cross-trait lookup**

We used PLINK 2.0 (63) to clump meta-analysis summary statistics based on LD structure in reference data from the UK10K project. We then performed hypothesis-free cross-trait lookup of independently associated loci using the SNP lookup function in the MRBase catalogue (64). Proxies with an $r^2$ of 0.8 or higher were included where the given variant was not present in an outcome of interest. We considered performing hypothesis-free cross-trait genetic correlation analysis using bivariate LD score regression implemented in LDhub (65).

**Lookup in previously published paediatric caries GWAS**

Previously published caries GWAS was performed within the GENEVA consortium, which is also represented in our meta-analysis. We therefore did not feel it would be informative to undertake lookup of associated variants in previously published results.

**Lookup in GWAS for adult caries traits**

This analysis was planned and conducted in parallel with analysis of quantitative traits measuring lifetime caries exposure in adults (manuscript in draft). The principal trait studied in the adult analysis was an index of decayed, missing and filled tooth surfaces (DMFS). This index was calculated from results of clinical dental examination, excluding third molar teeth. The DMFS index was age-and-sex standardized within each participating adult study before GWAS analysis was undertaken. Study-specific results files were then combined in a fixed-effects meta-analysis. In addition to DMFS, two secondary
caries traits were studied in adults, namely number of teeth (a count of remaining natural teeth at time of study participation) and standardized DFS (derived as the number of decayed and filled surfaces divided by the number of natural tooth surfaces remaining at time of study participation). After age-and-sex standardization these secondary traits had markedly non-normal distribution and were therefore underwent rank-based inverse normal transformation before GWAS analysis and meta-analysis. We performed cross-trait lookup of lead associated variants in the paediatric caries meta-analysis against these three adult caries traits. As the unpublished analysis also contains samples which contributed to previously published GWAS, we did not feel it would be informative to undertake additional lookup in published data.

**Gene prioritization, gene set enrichment and association with predicted gene transcription**

Gene-based testing of summary statistics was performed using MAGMA (66) with reference data for LD correction taken from the UK10K project and gene definitions based on a 50 kb window either side of canonical gene start: stop positions. Gene set enrichment analysis was considered using the software package DEPICT (67). Tests for association between phenotype and predicted gene transcription were performed using S-PrediXcan (68), which is a summary-statistic-implementation of the PrediXcan method. This method aims to assess the effects of tissue-specific gene transcription on phenotypes. Gene transcription models are trained in datasets with transcriptomic data, then used to predict gene expression in datasets with phenotypic data. This method was applied using the MetaXcan standalone software (https://github.com/hakyimlab/MetaXcan; date last accessed June 2018) and a transcription prediction model trained in whole blood (obtained from the PedictDB data repository at http://predictdb.org; date last accessed June 2018). Bonferroni correction was applied on the basis of approximately 7000 independent gene-based tests for two caries traits, giving an experiment-wide significance level of approximately \( P < 3.6 \times 10^{-6} \).

**Power calculations**

Post-hoc power calculations were performed using the free, web-based tool Genetic Association Study (GAS) Power Calculator and the software utility Quanto (v1.2.4) (https://csg.sph.umich.edu/abecasis/gas_power_calculator/index.html, http://biostats.usc.edu/Quanto.html; date last accessed June 2018) (69). Using the sample size and caries prevalence of the final meta-analysis samples, we calculated the minimum effect size required to have 80% discovery power at a significance level of 5.0e-08 for variants with MAF between 0.05 and 0.50. For primary teeth (17 037 individuals, 6922 caries affected, prevalence 40.6%) we were able to detect variants with a minimal effect size (OR) between 1.13 and 1.37 for variants with MAF of 0.50 and 0.05, respectively (1.15 for MAF of 0.40) (Supplementary Material, Figs S4 and S5). For permanent teeth (13 353 individuals of which 5875 were caries-affected, prevalence 44.0%) we had 80% power to detect variants with a minimal effect size (OR) between 1.15 and 1.43 for variants with MAF of 0.50 and 0.05, respectively (1.17 for MAF of 0.40) (Supplementary Material, Figs S4 and S5).

**Supplementary Material**

Supplementary Material is available at HMG online.

**Acknowledgements**

This work was supported by Wellcome (grant number 202802/Z/16/Z to N.T., 201237/Z/16/Z to S.H.) and the UK Medical Research Council (grant number MC_UU_12013/3). N.T works in a unit which receives support from the University of Bristol, and in a biomedical research centre which receives support from the National Institute for Health Research. The Swedish Research Council provides support to D.S in the form of an International Fellowship (grant number 4.1-2016-00416). ALSPAC receives core support from the UK Medical Research Council and Wellcome (grant number 102215/2/13/Z) and the University of Bristol. This publication is the work of the authors and Nicholas Timpson will serve as guarantor for the contents of this article. A comprehensive list of grants funding available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). Collection of phenotype data was supported by Wellcome and the UK Medical Research Council (grant number 076467/Z/05/Z). GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

The Young Finns Study has been financially supported by the Academy of Finland (grant numbers 286284, 134309 [Eye], 126925, 121584, 124282, 129378 [Salve], 17787 [Gend], and 41071 [Skibili]) the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant number X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gylenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

Analysis within the GENEVA consortium was supported by the following USA National Institutes of Health (NIH) grants from the National Institute of Dental and Craniofacial Research (NIDCR): (grant numbers R01-DE014899, U01-DE018903, R03-DE024264, R01-DE09551, R01-DE12101, P60-DE-013076), and a National Institutes for Health contract (contract number HHSN268200782-096C). Analysis within Raine was supported by the National Health and Medical Research Council of Australia (grant numbers 572613 and 40398) and the Canadian Institutes of Health Research (grant number MOP-82893). The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the NH&MRC for their long-term funding to the study over the last 25 years and also the following institutes for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Curtin University, the Raine Medical Research Foundation, the UWA Faculty of Medicine, Dentistry and Health Sciences, the Telethon Kids Institute, the Women’s and Infant’s Research Foundation (King Edward Memorial Hospital), Murdoch University, The University of Notre Dame (Australia) and Edith Cowan University. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). We would also like to acknowledge the Raine Study participants for their ongoing participation in the study, and the Raine Study Team for study coordination and data collection. This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and Government of Western Australia.
We are very grateful to the children and families who agreed to participate in the contributing studies, without whom this research would not be possible. We would like to acknowledge the role of Mark McCarthy and the Early Growth Genetics consortium in recruiting studies which contributed to this analysis. For ALSPAC, we are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). We would also like to acknowledge the Raine Study participants for their ongoing participation in the study, and the Raine Study Team for study coordination and data collection.

The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosediens & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of GWAS genotype data for the Generation R Study was done at the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Manoushka Ganesh, Lizbeth Herrera and Marjolein Peters for their help in creating, managing and QC of the GWAS database. The general design of Generation R Study was made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organisation for Scientific Research (NWO), the Ministry of Health, Welfare and Sport and the Ministry of Youth and Families. Additionally, the Netherlands Organization for Health Research and Development supported the Generation R Study (ZonMw 907.00303, ZonMw 916.10159, ZonMw VIDI 016.136.361 and ZonMw VIDI 016.136.367) to F.R. and C.M.-G. of this manuscript. This project also received funding from the European Union’s Horizon 2020 research and innovation programme under the following grant agreements: [No. 633595 (DynaHEALTH) and No. 733206 (LIFECYCLE)]. Furthermore, Generation R received additional funding from the European Research Council (ERC Consolidator Grant, ERC-2014-CoG-648916).

Conflict of Interest statement. None declared.

Funding
Funding to pay the Open Access publication charges for this article was provided by the Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical Research Council and University of Bristol.

References


