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

Simon Haworth[1*], Dmitry Shungin[2,3], Justin T van der Tas[4], Strahinja Vucic[4], Carolina Medina-Gomez[5,6,7], Victor Yakimov[8], Bjarke Feenstra[8], John R Shaffer[9,10], Myoung Keun Lee[10], Marie Standl[11], Elisabeth Thiering[11,12], Carol Wang[13], Klaus Bønnelykke[14], Johannes Waage[14], Leon Eyrich Jessen[14], Pia Elisabeth Nørrisgaard[14], Raimo Joro[15], Ilkka Seppälä[16], Olli Raitakari[17,18], Tom Dudding[1], Olja Grgic[4,5], Edwin Ongkosuwito[5], Anu Vierola[15] Aino-Maija Eloranta[15], Nicola X West[19], Steven J Thomas[19], Daniel W McNeil[20], Steven M Levy[21], Rebecca Slayton[22], Ellen A Nohr[23], Terho Lehtimäki[16], Timo Lakka[15,24,25], Hans Bisgaard[14], Craig Pennell[13], Jan Kühnisch[26], Mary L Marazita[9,10], Mads Melbye[8,27,28], Frank Geller[8], Fernando Rivadeneira[5,6,7], Eppo B Wolvius[4], Paul W Franks[29,30,31], Ingegerd Johansson[2], Nicholas J Timpson[1].

[1]Medical Research Council Integrative Epidemiology Unit at Bristol Medical School, University of Bristol, Bristol BS8 2BN, United Kingdom

[2]Department of Odontology, Umeå University, Umeå 901 87, Sweden

[3]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, United States of America

[4]Department of Oral and Maxillofacial Surgery, Special Dental Care and Orthodontics, Erasmus Medical Center, Rotterdam 3015CN , The Netherlands

[5]The Generation R Study Group, Erasmus Medical Center, Rotterdam 3015CN, The Netherlands

[6]Department of Internal Medicine, Erasmus Medical Center, Rotterdam 3015CN , The Netherlands

[7]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, 3015 CN, Rotterdam, The Netherlands

[8]Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark

[9]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States of America

[10]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, United States of America

[11]Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg D-85764, Germany

[12]Division of Metabolic and Nutritional Medicine, Dr von Hauner Children's Hospital, University of Munich Medical Center, Munich 80337, Germany

[13]Division of Obstetrics and Gynaecology, The University of Western Australia, Perth WA6009, Australia

[14]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofe Hospital, University of Copenhagen, Copenhagen 2730, Denmark

[15]Institute of Biomedicine, School of Medicine, University of Eastern Finland Kuopio Campus, 70211 Kuopio, Finland

[16]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere - Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland

[17]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland

[18]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland

[19]Bristol Dental School, University of Bristol, Bristol BS1 2LY, United Kingdom

[20]Department of Psychology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, West Virginia 26506-6286, United States of America

[21]Department of Preventive and Community Dentistry, College of Dentistry, University of Iowa, Cedar Rapids, Iowa 52242-1010, United States of America

[22]Department of Pediatric Dentistry (Retired), School of Dentistry, University of Washington, Seattle, Washington 98195, United States of America

[23]Research Unit for Gynaecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense 5000, Denmark

[24]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio 70210, Finland

[25]Kuopio Research Institute of Exercise Medicine, Kuopio 70100, Finland

[26]Department of Conservative Dentistry and Periodontology, University Hospital, Ludwig-Maximilians-Universität München, Munich 80336, Germany

[27]Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200, Denmark

[28]Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, United States of America

[29]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö 202 13, Sweden

[30]Department of Public Health and Clinical Medicine, Umeå University, Umeå 901 85, Sweden

[31]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, United States of America

*Corresponding author Simon Haworth. MRC Integrative Epidemiology Unit, Oakfield House, Oakfield Grove, Bristol BS8 2BN, United Kingdom. e-mail: simon.haworth@bristol.ac.uk, Tel: +44 (0) 117 3310083 , Fax +44 (0)117 928 7325

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

Abstract

Prior studies suggest dental caries traits in children and adolescents are partially heritable, but there has been no large-scale consortium genome-wide association study (GWAS) to date. We therefore performed GWAS for caries in participants aged 2.5-18.0 years from 9 contributing centres. Phenotype definitions were created for the presence or absence of treated or untreated caries, stratified by primary and permanent dentition. All studies tested for association between caries and genotype dosage and results were combined using fixed–effects meta-analysis. Analysis included up to 19,003 individuals (7,530 affected) for primary teeth and 13,353 individuals (5,875 affected) for permanent teeth. Evidence for association with caries status was observed at rs1594318-C for primary teeth (intronic within *ALLC*, Odds Ratio (OR) 0.85, Effect Allele Frequency (EAF) 0.60, p 4.13e-8) and rs7738851-A (intronic within *NEDD9*, OR 1.28, EAF 0.85, p 1.63e-8) for permanent teeth. Consortium-wide estimated heritability of caries was low (h^2 of 1% [95% CI: 0%:7%] and 6% [95% CI 0%:13%] for primary and permanent dentitions, respectively) compared to corresponding within-study estimates (h^2 of 28%, [95% CI: 9%:48%] and 17% [95% CI:2%:31%]) or previously published estimates. This study was designed to identify common genetic variants with modest effects which are consistent across different populations. We found few single variants associated with caries status under these assumptions. Phenotypic heterogeneity between cohorts and limited statistical power will have contributed; these findings could also reflect complexity not captured by our study design, such as genetic effects which are conditional on environmental exposure.

Introduction

Dental caries remains a prevalent public health problem in both children and adults. Untreated dental caries was estimated to affect 621 million children worldwide in 2010, with little change in prevalence or incidence between 1990 and 2010 (1). This problem is not unique to lower income countries; around 50% of children have evidence of caries by age 5 in industrialized nations (2-4). Dental caries results from reduced mineral saturation of fluids surrounding teeth, driven by ecological shifts in the oral microbiome (5). Many different factors predispose towards dental caries, of which high sugar consumption, poor oral hygiene and low socio-economic status are the most notorious (6-8). Over the last decades there has been increasing appreciation for the role of genetic influences in dental caries. The importance of genetic susceptibility for dental caries experience was demonstrated in an animal model over 50 years ago, a finding since substantiated in twin studies in humans (9-11). Of particular relevance to caries traits in children and adolescents, Bretz et al. analyzed longitudinal rates of change in caries status in children, and found that caries progression and severity were highly heritable in the primary and permanent dentition (10). 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 contribution to caries susceptibility, few specific genetic loci have been identified.

Shaffer et al. performed the first GWAS for dental caries in 2011 (13), studying the primary dentition of 1,305 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 ($p \leq 5.0e-08$), 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 7,443 adults from five different cohorts and identified several suggestive loci (P-value ≤ 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 1,000 children and found evidence for association at loci reported in previous studies, including *MPPED2*, *RPS6KA2*, and *AJAP1* (13-16). The largest GWAS for dental caries in permanent dentition was performed in a Hispanic and Latino sample of 11,754 adults (17). This study identified unique genetic loci (*NAMPT* and *BMP7*) compared to 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, behavioral 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 – to - 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 17,037 individuals (6,922 affected) from 22 results files representing all 9 coordinating centres. After final quality control (QC), this meta-analysis included 8,640,819 variants, with mild deflation (genomic inflation factor (λ) = 0.994)(S1 Fig). Meta-analysis of caries in primary teeth which included individuals of multiple ethnicities in the Generation R (GENR) study included 19,003 individuals (7,530 affected) from 22 results files representing all 9 coordinating centres. There were 8,699,928 variants after final QC, with mild deflation in summary statustics (λ = 0.986)(S2 Fig). Analysis of caries status in permanent teeth included 13,353 individuals (5,875 affected) from 14 results files representing 7 coordinating centres. The sample size was smaller for permanent teeth as two coordinating centres did not have phentoype data for permanent teeth (RAINE and GENR), whilst the COPSAC group only had data for participants in the earlier birth cohort (COPSAC 2000). There were 8,734,121 variants afer final QC, with mild deflation in summary statistics (λ = 0.999)(S3 Fig).

The strongest evidence for association with caries in primary teeth was seen at rs1594318 (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 ancestories 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 (Figure 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 heritabiltiy. 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 LDSR method for caries in either primary or permanent teeth. Hypothesis-free phenome 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 $p14^{ARF}$, 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 meta-analysis aproach. 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 in primary teeth. This variant showed weaker evidence for association in the multi-ethnic meta-analysis, potentially relating to different alelle frequencies across the different ethnic groups included in analysis. Frequency of the G allele is reported to vary between 0.24 in Asian populations to 0.42 in populations of European ancestry based on 1KGP allele fequencies. *ALLC* (Allantoicase) codes the enzyme allantoicase, which is involved in purine metabolism and whose enzymatic activitiy is believed to have been lost during vertebrate evolution. Mouse studies suggest 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-regulard 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 susceptibiltiy.

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 heritabiltiy estimates were lower than anticipated from either previous within-study heritability estimates (34) or the the new within-study heritabiltiy estimates obtained for this analysis. There are several possible explanations for this phenomenom. 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 heritabiliy 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-heritabiltiy estimate is not the same conceptually as the weighted average of heritabiltiy within each study.

More intuitively, genetic influences might be important within populations with relatively similar environments but not determine much of the overall diffrences 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 architectre 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 defintions 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 continium 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 sucsessully in a range of complex phenotypes.

Between participating centres there are differences in characteristics such as age at participation, phenotypic assesment and differences in the environment (such as nutrition, oral hygiene and the oral microbime) 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 hetrogeneous 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 hetrogeneity in genetic effects (I^2) is limited by the small number of participting studies in meta-anlaysis (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 to other studies as children or their parents are unlikely to be aware of untreated dental caries which would be evident to a trained assesor. 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 towards the null. Despite this we show evidence for heritabiltiy using these definitions and effect sizes at lead variants are comparable with effect sizes obtained using clinically assessed data (Figs 3,4).

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 discused 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 (for example, knowledege 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 suceptibility.

Materials and Methods

Study samples

We performed genome-wide association (GWA) analysis for dental caries case/control status in a consortium including 9 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 age 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 fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Both clinical and questionnaire derived data were included in this analysis, with priority given to clinical data where available. (S3 Table).

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 enrollment 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 (50). COHRA recruited families with at least one child aged between 1 and 18 years of age, with dental examination performed at baseline (51). IFS recruited mothers and newborn 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 age 5, 9 13 and 17 years (52). 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 four 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 (53, 54) .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 (53, 55). 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 (56). 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 (57). 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 (58). 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 S1 Table.

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.bristol.ac.uk/alspac/researchers/research-ethics/.) 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 most heritable in the near-eruption period.

The sources of data used to create these phenotypic definitions are given in S3 Table. 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 eighteenth 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 6 centres (GINIplus/LISA, GENR, GENEVA, YFS, PANIC, RAINE, (S1 Table). In ALSPAC, DNBC, COPSAC2000 and COPSAC 2010 the haplotype reference consortium (HRC v1.0 and v1.1) imputation panels were used. (S1 Table)

Each study performed routine quality control measures during genotyping, imputation and association testing (S2 Table). Further pre-meta-analysis quality control was performed centrally using the EasyQC R package and accompanying 1KG phase1 v3 reference data (59). Minor allele count (MAC) was derived as the product of minor allele frequency 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 minor allele frequency (MAF) of less than 0.005were 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 S2 Table. In the GENR study parallel analyses were conducted for participants of European ancestry (using the approach described in S2 Table) and the entire study population, using a previously published method (48). The software and exact approach used by each study is shown in S2 Table.

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 (60), 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 linkage disequilibrium score regression (LDSR) (61). Reference linkage disequilibrium (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 (62), implemented in the GCTA software package (63), 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 (64) 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 (65). 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 (66).

Lookup in previously published pediatric 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 pediatric 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 (67) with reference data for LD correction taken from the UK10K project and gene definitions based on a 50 kilobase window either side of canonical gene start:stop positions. Gene set enrichment analysis was considered using the software package DEPICT (68). Tests for association between phenotype and predicted gene transcription were performed using S-PrediXcan (69), 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) and a transcription prediction model trained in whole blood (obtained from the PedictDB data repository at http://predictdb.org/) (70). Bonferroni correction was applied on the basis of approximately 7,000 independent gene-based tests for 2 caries traits, giving an experiment wide significance level of approximately $p < 3.6e-06$.

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) (71). 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, 6,922 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) (S4 Fig, S5 Fig). For permanent teeth (13,353 individuals of which 5,875 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). (S4 Fig, S5 Fig)

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 102215/2/13/2] 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 paper. 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 (Gendi), and 41071 (Skidi)] 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 Gyllenberg 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 co-ordination 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 Reserach 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 co-ordination 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 Trombosedienst & 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

The authors declare there are no conflicting interests.

References

1 Kassebaum, N.J., Bernabe, E., Dahiya, M., Bhandari, B., Murray, C.J.L. and Marcenes, W. (2015) Global Burden of Untreated Caries: A Systematic Review and Metaregression. *J. Dent. Res.*, **94**, 650-658.

2 Vernazza, C.R., Rolland, S.L., Chadwick, B. and Pitts, N. (2016) Caries experience, the caries burden and associated factors in children in England, Wales and Northern Ireland 2013. *Br. Dent. J.*, **221**, 315-320.

3 van der Tas, J.T., Kragt, L., Elfrink, M.E.C., Bertens, L.C.M., Jaddoe, V.W.V., Moll, H.A., Ongkosuwito, E.M. and Wolvius, E.B. (2017) Social inequalities and dental caries in six-year-old children from the Netherlands. *J. Dentistry*, **62**, 18-24.

4 Schuller, A.A., van Dommelen, P. and Poorterman, J.H.G. (2014) Trends in oral health in young people in the Netherlands over the past 20 years: a study in a changing context. *Community Dent. Oral Epidemiol.*, **42**, 178-184.

5 Philip, N., Suneja, B. and Walsh, L.J. (2018) Ecological Approaches to Dental Caries Prevention: Paradigm Shift or Shibboleth? *Caries Res.*, **52**, 153-165.

6 Peres, M.A., Sheiham, A., Liu, P., Demarco, F.F., Silva, A.E., Assunção, M.C., Menezes, A.M., Barros, F.C. and Peres, K.G. (2016) Sugar Consumption and Changes in Dental Caries from Childhood to Adolescence. *J. Dent. Res.*, **95**, 388-394.

7 Sheiham, A. and James, W.P. (2015) Diet and Dental Caries: The Pivotal Role of Free Sugars Reemphasized. *J. Dent. Res.*, **94**, 1341-1347.

8 Lynch, R.J. (2013) The primary and mixed dentition, post-eruptive enamel maturation and dental caries: a review. *Int. Dent. J.*, **63**, 3-13.

9 Boraas, J.C., Messer, L.B. and Till, M.J. (1988) A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart. *J. Dent. Res.*, **67**, 1150-1155.

10 Bretz, W.A., Corby, P.M., Schork, N.J., Robinson, M.T., Coelho, M., Costa, S., Melo Filho, M.R., Weyant, R.J. and Hart, T.C. (2005) Longitudinal analysis of heritability for dental caries traits. *J. Dent. Res.*, **84**, 1047-1051.

11 Chung, C.S. and Larson, R.H. (1967) Factors and inheritance of dental caries in the rat. *J. Dent. Res.*, **46**, 559-564.

12 Bretz, W.A., Corby, P.M.A., Melo, M.R., Coelho, M.Q., Costa, S.M., Robinson, M., Schork, N.J., Drewnowski, A. and Hart, T.C. (2006) Heritability estimates for dental caries and sucrose sweetness preference. *Arch. Oral Biol.*, **51**, 1156-1160.

13 Shaffer, J.R., Wang, X., Feingold, E., Lee, M., Begum, F., Weeks, D., Cuenco, K.T., Barmada, M.M., Wendell, S.K., Crosslin, D.R. *et al.* (2011) Genome-wide association scan for childhood caries implicates novel genes. *J. Dent. Res.*, **90**, 1457-1462.

14 Wang, X., Shaffer, J.R., Zeng, Z., Begum, F., Vieira, A.R., Noel, J., Anjomshoaa, I., Cuenco, K.T., Lee, M.K., Beck, J. *et al.* (2012) Genome-wide association scan of dental caries in the permanent dentition. *BMC Oral Health*, **12**, 57.

15 Zeng, Z., Feingold, E., Wang, X., Weeks, D.E., Lee, M., Cuenco, D.T., Broffitt, B., Weyant, R.J., Crout, R., McNeil, D.W. *et al.* (2014) Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. *Caries Res.*, **48**, 330-338.

16 Zeng, Z., Shaffer, J.R., Wang, X., Feingold, E., Weeks, D.E., Lee, M., Cuenco, K.T., Wendell, S.K., Weyant, R.J., Crout, R. *et al.* (2013) Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. *J. Dent. Res.*, **92**, 432-437.

17 Morrison, J., Laurie, C.C., Marazita, M.L., Sanders, A.E., Offenbacher, S., Salazar, C.R., Conomos, M.P., Thornton, T., Jain, D., Laurie, C.A. *et al.* (2016) Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL). *Hum. Mol. Genet.*, **25**, 807-816.

18 Mak, G.W.Y., Lai, W.L., Zhou, Y., Li, M.T., Ng, I.O.L. and Ching, Y.P. (2012) CDK5RAP3 Is a Novel Repressor of p14(ARF) in Hepatocellular Carcinoma Cells. *Plos One*, **7**.
19 Mak, G.W.Y., Chan, M.M.L., Leong, V.Y.L., Lee, J.M.F., Yau, T.O., Ng, I.O.L. and Ching, Y.P. (2011) Overexpression of a Novel Activator of PAK4, the CDK5 Kinase-Associated Protein CDK5RAP3, Promotes Hepatocellular Carcinoma Metastasis. *Cancer Res.*, **71**, 2949-2958.
20 Vigetti, D., Pollegioni, L., Monetti, C., Prati, M., Bernardini, G. and Gornati, R. (2002) Property comparison of recombinant amphibian and mammalian allantoicases. *Febs Letters*, **512**, 323-328.
21 Park, T.J., Park, J.S., Cheong, H.S., Park, B.L., Kim, L.H., Heo, J.S., Kim, Y.K., Kim, K.U., Uh, S.T., Lee, H.S. *et al.* (2014) Genome-wide association study identifies ALLC polymorphisms correlated with FEV1 change by corticosteroid. *Clinica Chimica Acta*, **436**, 20-26.
22 Tikhmyanova, N., Little, J.L. and Golemis, E.A. (2010) CAS proteins in normal and pathological cell growth control. *Cell. Mol. Life Sci.*, **67**, 1025-1048.
23 Singh, M.K., Dadke, D., Nicolas, E., Serebriiskii, I.G., Apostolou, S., Canutescu, A., Egleston, B.L. and Golemis, E.A. (2008) A novel Cas family member, HEPL, regulates FAK and cell spreading. *Mol. Biol. Cell*, **19**, 1627-1636.
24 Kumar, S., Tomooka, Y. and Noda, M. (1992) IDENTIFICATION OF A SET OF GENES WITH DEVELOPMENTALLY DOWN-REGULATED EXPRESSION IN THE MOUSE-BRAIN. *Biochem. Biophys. Res. Commun.*, **185**, 1155-1161.
25 Latasa, M.J., Jimenez-Lara, A.M. and Cosgaya, J.M. (2016) Retinoic acid regulates Schwann cell migration via NEDD9 induction by transcriptional and post-translational mechanisms. *Biochim. Biophys. Acta-Mol. Cell Res.*, **1863**, 1510-1518.
26 Aquino, J.B., Lallemend, F., Marmigere, F., Adameyko, II, Golemis, E.A. and Ernfors, P. (2009) THE RETINOIC ACID INDUCIBLE Cas-FAMILY SIGNALING PROTEIN Nedd9 REGULATES NEURAL CREST CELL MIGRATION BY MODULATING ADHESION AND ACTIN DYNAMICS. *Neurosci.*, **162**, 1106-1119.
27 Nikonova, A.S., Gaponova, A.V., Kudinov, A.E. and Golemis, E.A. (2014) CAS Proteins in Health and Disease: An Update. *Iubmb Life*, **66**, 387-395.
28 Riccomagno, M.M., Sun, L.O., Brady, C.M., Alexandropoulos, K., Seo, S., Kurokawa, M. and Kolodkin, A.L. (2014) Cas Adaptor Proteins Organize the Retinal Ganglion Cell Layer Downstream of Integrin Signaling. *Neuron*, **81**, 779-786.
29 Wang, S.K., Komatsu, Y. and Mishina, Y. (2011) Potential contribution of neural crest cells to dental enamel formation. *Biochem. Biophys. Res. Comms.*, **415**, 114-119.
30 Duverger, O., Zah, A., Isaac, J., Sun, H.W., Bartels, A.K., Lian, J.B., Berdal, A., Hwang, J. and Morasso, M.I. (2012) Neural Crest Deletion of Dlx3 Leads to Major Dentin Defects through Down-regulation of Dspp. *J. Biol. Chem*, **287**, 12230-12240.
31 Divaris, K. (2016) Predicting Dental Caries Outcomes in Children: A "Risky" Concept. *J. Dent. Res.*, **95**, 248-254.
32 Chapple, I.L.C., Bouchard, P., Cagetti, M.G., Campus, G., Carra, M.-C., Cocco, F., Nibali, L., Hujoel, P., Laine, M.L., Lingström, P. *et al.* (2017) Interaction of lifestyle, behaviour or systemic diseases with dental caries and periodontal diseases: consensus report of group 2 of the joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases. *J. Clin. Periodontol.*, **44**, S39-S51.
33 Nibali, L., Di Iorio, A., Tu, Y.-K. and Vieira, A.R. (2017) Host genetics role in the pathogenesis of periodontal disease and caries. *J. Clin. Periodontol.*, **44**, S52-S78.
34 Wang, X., Shaffer, J.R., Weyant, R.J., Cuenco, K.T., DeSensi, R.S., Crout, R., McNeil, D.W. and Marazita, M.L. (2010) Genes and Their Effects on Dental Caries May Differ between Primary and Permanent Dentitions. *Caries Res.*, **44**, 277-284.
35 Docherty, A.R., Moscati, A., Peterson, R., Edwards, A.C., Adkins, D.E., Bacanu, S.A., Bigdeli, T.B., Webb, B.T., Flint, J. and Kendler, K.S. (2016) SNP-based heritability estimates of the personality dimensions and polygenic prediction of both neuroticism and major depression: findings from CONVERGE. *Transl. Psychiatry*, **6**.

36 Shaffer, J.R., Wang, X.J., McNeil, D.W., Weyant, R.J., Crout, R. and Marazita, M.L. (2015) Genetic Susceptibility to Dental Caries Differs between the Sexes: A Family-Based Study. *Caries Res.*, **49**, 133-140.
37 Bayram, M., Deeley, K., Reis, M.F., Trombetta, V.M., Ruff, T.D., Sencak, R.C., Hummel, M., Dizak, P.M., Washam, K., Romanos, H.F. *et al.* (2015) Genetic influences on dental enamel that impact caries differ between the primary and permanent dentitions. *Eur. J. Oral Sci.*, **123**, 327-334.
38 Shaffer, J.R., Carlson, J.C., Stanley, B.O.C., Feingold, E., Cooper, M., Vanyukov, M.M., Maher, B.S., Slayton, R.L., Willing, M.C., Reis, S.E. *et al.* (2015) Effects of enamel matrix genes on dental caries are moderated by fluoride exposures. *Hum. Genet.*, **134**, 159-167.
39 Johansson, I., Witkowska, E., Kaveh, B., Holgerson, P.L. and Tanner, A.C.R. (2016) The Microbiome in Populations with a Low and High Prevalence of Caries. *J. Dent. Res.*, **95**, 80-86.
40 von Hippel, P.T. (2015) The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med. Res. Methodol.*, **15**, 35.
41 Haworth, S., Dudding, T., Waylen, A., Thomas, S.J. and Timpson, N.J. (2017) Ten years on: Is dental general anaesthesia in childhood a risk factor for caries and anxiety? *Br. Dent. J.*, **222**, 299-304.
42 Kemp, J.P., Morris, J.A., Medina-Gomez, C., Forgetta, V., Warrington, N.M., Youlten, S.E., Zheng, J., Gregson, C.L., Grundberg, E., Trajanoska, K. *et al.* (2017) Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis, *Nat. Genet.,* **49**, 1468.
43 Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Smith, G.D. (2013) Cohort Profile: The 'Children of the 90s'-the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, **42**, 111-127.
44 Bisgaard, H. (2004) The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann. Allergy Asthma Immunol.*, **93**, 381-389.
45 Olsen, J., Melbye, M., Olsen, S.F., Sorensen, T.I.A., Aaby, P., Andersen, A.M.N., Taxbol, D., Hansen, K.D., Juhl, M., Schow, T.B. *et al.* (2001) The Danish National Birth Cohort - its background, structure and aim. *Scand. J.Public Health*, **29**, 300-307.
46 Kooijman, M.N., Kruithof, C.J., van Duijn, C.M., Duijts, L., Franco, O.H., van Ijzendoorn, M.H., de Jongste, J.C., Klaver, C.C.W., van der Lugt, A., Mackenbach, J.P. *et al.* (2016) The Generation R Study: design and cohort update 2017. *Eur. J. Epidemiol.*, **31**, 1243-1264.
47 van der Tas, J.T., Kragt, L., Veerkamp, J.J.S., Jaddoe, V.W.V., Moll, H.A., Ongkosuwito, E.M., Elfrink, M.E.C. and Wolvius, E.B. (2016) Ethnic Disparities in Dental Caries among Six-Year-Old Children in the Netherlands. *Caries Res.*, **50**, 489-497.
48 Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., Duijts, L., Hofman, A., van Duijn, C.M., Uitterlinden, A.G. *et al.* (2015) Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur. J. Epidemiol.*, **30**, 317-330.
49 Cornelis, M.C., Agrawal, A., Cole, J.W., Hansel, N.N., Barnes, K.C., Beaty, T.H., Bennett, S.N., Bierut, L.J., Boerwinkle, E., Doheny, K.F. *et al.* (2010) The Gene, Environment Association Studies Consortium (GENEVA): Maximizing the Knowledge Obtained from GWAS by Collaboration Across Studies of Multiple Conditions. *Genet. Epidemiol.*, **34**, 364-372.
50 Zeng, Z., Feingold, E., Wang, X., Weeks, D.E., Lee, M., Cuenco, K.T., Broffitt, B., Weyant, R.J., Crout, R., McNeil, D.W. *et al.* (2014) Genome-Wide Association Study of Primary Dentition Pit-and-Fissure and Smooth Surface Caries. *Caries Res.*, **48**, 330-338.
51 Polk, D.E., Weyant, R.J., Crout, R.J., McNeil, D.W., Tarter, R.E., Thomas, J.G. and Marazita, M.L. (2008) Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health*, **8**, 18.
52 Levy, S.M., Warren, J.J., Broffitt, B., Hillis, S.L. and Kanellis, M.J. (2003) Fluoride, beverages and dental caries in the primary dentition. *Caries Res.*, **37**, 157-165.

53 Taal, H.R., St Pourcain, B., Thiering, E., Das, S., Mook-Kanamori, D.O., Warrington, N.M., Kaakinen, M., Kreiner-Moller, E., Bradfield, J.P., Freathy, R.M. *et al.* (2012) Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat. Genet.*, **44**, 532-+.
54 von Berg, A., Kramer, U., Link, E., Bollrath, C., Heinrich, J., Brockow, I., Koletzko, S., Grubl, A., Filipiak-Pittroff, B., Wichmann, H.E. *et al.* (2010) Impact of early feeding on childhood eczema: development after nutritional intervention compared with the natural course - the GINIplus study up to the age of 6 years. *Clin. Exp. Allergy*, **40**, 627-636.
55 Zutavern, A., Brockow, I., Schaaf, B., Bolte, G., von Berg, A., Diez, U., Borte, M., Herbarth, O., Wichmann, H.E., Heinrich, J. *et al.* (2006) Timing of solid food introduction in relation to atopic dermatitis and atopic sensitization: Results from a prospective birth cohort study. *Pediatrics*, **117**, 401-411.
56 Eloranta, A.M., Lindi, V., Schwab, U., Tompuri, T., Kiiskinen, S., Lakka, H.M., Laitinen, T. and Lakka, T.A. (2012) Dietary factors associated with overweight and body adiposity in Finnish children aged 6-8 years: the PANIC Study. *Int. J. Obesity*, **36**, 950-955.
57 Raitakari, O.T., Juonala, M., Ronnemaa, T., Keltikangas-Jarvinen, L., Rasanen, L., Pietikainen, M., Hutri-Kahonen, N., Taittonen, L., Jokinen, E., Marniemi, J. *et al.* (2008) Cohort Profile: The Cardiovascular Risk in Young Finns Study. *Int. J. Epidemiol.*, **37**, 1220-1226.
58 Straker, L., Mountain, J., Jacques, A., White, S., Smith, A., Landau, L., Stanley, F., Newnham, J., Pennell, C. and Eastwood, P. (2017) Cohort Profile: The Western Australian Pregnancy Cohort (Raine) Study-Generation 2. *Int. J. Epidemiol.*, **5**, 1384-1385.
59 Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E. *et al.* (2014) Quality control and conduct of genome-wide association meta-analyses. *Nat. Protocols*, **9**, 1192-1212.
60 Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190-2191.
61 Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M. and Schizophrenia Working, G. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291-+.
62 Yang, J.A., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**, 565-U131.
63 Yang, J.A., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.*, **88**, 76-82.
64 Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.
65 Hemani, G., Zheng, J., Wade, K.H., Laurin, C., Elsworth, B., Burgess, S., Bowden, J., Langdon, R., Tan, V., Yarmolinsky, J. *et al.* (2016) MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv*, in press.
66 Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., St Pourcain, B. *et al.* (2017) LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, **33**, 272-279.
67 de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: Generalized Gene-Set Analysis of GWAS Data. *Plos Comput. Biol.*, **11**.
68 Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T. *et al.* (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Comms.*, **6**, 9.
69 Barbeira, A.N., Dickinson, S.P., Torres, J.M., Bonazzola, R., Zheng, J., Torstenson, E.S., Wheeler, H.E., Shah, K.P., Edwards, T., Garcia, T. *et al.* (2017) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv*, in press.
70 Im Lab, G.M., Department of Medicine, The University of Chicago. (2017), in press.

71 Skol, A., Scott, L., Abecasis, G. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet*. **38**, 209-213.

Legends to Figures

Fig 1. Manhattan plots for each principal meta-analysis. 1a: Caries in primary teeth (European ancestry), n samples = 17,036, n variants = 8,640,819, $\lambda = 0.9944$. Variants within 500Kb of rs1594318 are highlighted in green. 1b: Caries in primary teeth (multi-ethnic analysis), n samples = 19,003, n variants = 8,699,928, $\lambda = 0.9861$. 2c: Caries in permanent teeth (European ancestry), n samples = 13,353, n variants = 8,734,121, $\lambda = 0.9991$. Variants within 500Kb of rs7738851 are highlighted in green.

Fig 2. Regional association plots. 2a: Regional association plot for rs1594318 and caries in primary teeth (European ancestry meta-analysis. 2b: Regional association plot for rs7738851 and caries in permanent teeth.

Fig 3. Forest plot for rs1594318 and caries in primary teeth. Effect sizes are expressed on a log odds ratio scale, grouped by geographical location. The summary estimate is from the fixed-effect meta-analysis of participants of European ancestry.

Fig 4. Forest plot for rs7738851 and caries in permanent teeth. Effect sizes are expressed on a log odds ratio scale, grouped by geographical location. The summary estimate is from fixed-effect meta-analysis.

Tables

Table 1. Lead associated single variants.

Phenotype	Variant	Position	Effect allele	Other allele	EAF	Beta (SE)	Odds ratio	P value	N	I^2	P value for heterogeneity	Annotation
Caries in primary teeth (European ancestry analysis)	rs1594318	chr2:3733944	C	G	0.60	-0.165 (0.030)	0.848	4.13e-08	16,994	0.0	0.69	Intronic, *ALLC*
Caries in primary teeth (multi-ethnic analysis)*	rs1594318	chr2:3733944	C	G	0.60	-0.142 (0.028)	0.868	3.78e-07	18,960	0.0	0.61	Intronic, *ALLC*
Caries in primary teeth(multi-ethnic analysis)*	rs872877	chr2:3735826	A	G	0.59	-0.142 (0.028)	0.868	4.18e-07	18,958	17.5	0.68	Intronic, *ALLC*
Caries in permanent teeth	rs7738851	chr6:11241788	A	T	0.85	0.248 (0.044)	1.28	1.63e-08	13,353	13.3	0.20	Intronic, *NEDD9*

*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 the results section and are included here for reference.

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

Phenotype	Method		Estimated h^2 (95% CI)	N
Caries in primary teeth	GCTA GREML		0.28 (0.09:0.48)	7,230
	LDSR	All participants	0.01 (0.00:0.06)	19,003
		European ancestry only	0.01 (0.00:0.07)	17,036
Caries in permanent teeth	GCTA GREML		0.17 (0.02:0.31)	6,657
	LDSR		0.06 (0.00:0.12)	13,353

Table 3 – Lookup of lead associated variants

Variant	Discovery	Risk	Cross trait lookup	P value	Effect per caries risk	N

	trait	increasing allele (discovery)				increasing allele (se)	
rs1594318	Caries in primary teeth (European ancestry meta-analysis)	G	Adult caries traits	DMFS (standard deviation of residuals of caries-affected surfaces)	0.87	-0.0015 (0.0092)	26,790
				Number of teeth (inverse normal transformed residuals)	0.60	0.0051(0.0098)	27,947
				Standardized DFS (inverse normal transformed residuals)	0.033	-0.0195(0.0091)	26,532
			Hypothesis free	(no traits meeting threshold for multiple testing)			
rs7738851	Caries in permanent teeth	A	Adult caries traits	DMFS (standard deviation of residuals of caries-affected surfaces)	0.57	-0.007 (0.011)	26,791
				Number of teeth (inverse normal transformed residuals)	0.63	-0.0064 (0.013)	27,949
				Standardized DFS (inverse normal transformed residuals)	0.65	-0.0054 (0.012)	26,531
			Hypothesis free	(no traits meeting threshold for multiple testing)			

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 residualzed 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.