The NIH Consensus Development Program released a statement in 2001 (http://consensus.nih.gov/2001/2001DentalCaries115html.htm) and listed six major clinical caries research directions. One of these directions was the need for genetic studies to identify genes and genetic markers of diagnostic, prognostic and therapeutic value. This last decade has seen a steep increase in studies investigating the presence of genetic factors influencing individual susceptibility to caries. This review revisits recent caries human genetic studies and provides a perspective for future studies in order to fulfil their promise of revolutionizing our understanding of and the standard of care for the most prevalent bacteria-mediated non-contagious disease in the world.

Despite more than 100 years of accumulated knowledge on the pathogenesis of this disease, caries is still a major oral health problem in most industrialized countries, affecting 60–90% of schoolchildren and the vast majority of adults. According to the World Health Organization, caries is the most prevalent oral disease in several Asian and Latin American countries, while it appears to be less common and less severe in most African countries [World Health Organization, 2003]. In light of changing living conditions, however, it is expected that the incidence of caries will increase in many developing countries in Africa, due particularly to growing consumption of sugars and inadequate exposure to fluorides.

The interest in understanding the mechanisms underlying individual susceptibility to caries coincides with the development of feasible approaches to understand genetic susceptibility to complex human disease, thanks in large part to tools developed by the Human Genome Project (www.genome.gov). The combination of the high prevalence of caries among certain groups and the evidence that fluoride exposures do not protect all individuals [Slade et al., 2013] has specifically motivated research towards identifying genetic contributors to caries. A combination of candidate gene and genome-wide studies has arisen in the literature, with some notable successes, but also notably exemplifying the difficulties of developing a study with robust phenotype definitions and sample sizes that allow enough statistical power to detect genetic effects.
Further, note that the human genome is not the only genome involved in caries, i.e. the oral microbiome clearly plays a role in pathogenesis, and a number of groups are simultaneously working on the metagenomic contributions to caries. The oral cavity contains more than 700 species of bacteria, which grow primarily in multi-species biofilms (for example, dental plaque) that play vital roles in homeostasis of the mouth and in diseases of the oral cavity [Zarco et al., 2012]. Although caries is strongly associated with differences in diet and environmental factors, a study of the taxonomy of the salivary microbiome of 120 individuals selected from 12 different locations worldwide (10 per location) found little geographic structure; only 13.5% of the variance in genera composition observed between populations was explained by differences among individuals [Nasidze et al., 2009]. This suggests that microbial taxonomy alone – even after adjusting for environmental exposures – is insufficient to explain individual, familial and regional variations in oral microbiota and particularly caries susceptibility.

In this paper, we focus on recent human genetic studies, provide an overview of where the field stands and summarize the promise of these studies in devising new strategies to prevent and manage caries.

**Pathogenesis**

The term caries is thought to derive from words meaning rottenness, decay, injure, break, death, destruction, withered or faded. It alludes to a disease process, but does not describe it. According to Fejerskov et al. [2008], caries describes the signs and symptoms of a disease. In other words, it is the result of a localized chemical dissolution of the tooth surface caused by metabolic events that take place in the biofilm (dental plaque) that covers the affected area. Carious lesions result from a shift in the ecology and metabolic activity of the biofilm, whereby an imbalance in the equilibrium between tooth mineral and biofilm fluid has developed. Although the biofilm is a prerequisite for carious lesions to occur, its presence on solid surfaces does not necessarily result in the development of clinically visible carious lesions. Some chemical modifications are so subtle that they can only be recorded microscopically. When cumulative numerous pH fluctuations result in a net loss of calcium and phosphate of an extent that makes the enamel porous and visible, it is considered a 'white spot' lesion.

Featherstone [2006] explains the caries process as a balance between pathological and protective factors. Pathological factors, which include cariogenic bacteria, frequent ingestion of fermentable carbohydrates and salivary dysfunction, drive the caries process toward demineralization. Protective factors, which include salivary components, fluoride and remineralization, and targeted biofilm control drive the caries process toward remineralization [Featherstone, 2004]. If pathological factors outweigh protective factors, caries progresses as a consequence [Featherstone, 2006]. Behaviors related to oral hygiene practices, dietary choices and decisions involving seeking professional oral health care also contribute to caries, and defining these factors and including them along analyses of biological underpinnings is a major challenge.

The American Academy of Pediatric Dentistry [2012] currently recommends that caries risk assessment based on a child’s age, biological factors, protective factors and clinical findings should be a routine component of new and periodic examinations by oral health and medical providers.

There is no single test that takes into consideration all these factors and accurately predicts an individual’s susceptibility to caries. The caries risk may be determined by analyzing and integrating several factors such as caries experience (initial caries lesions and established caries defects, secondary caries and present caries activity), fluoride use, extent of plaque present, diet, bacterial and salivary activity and social and behavioral factors [Reich et al., 1999]. Although the best tool to predict future caries is past caries experience, this risk factor is not applicable for young children due to the need to determine caries risk before the disease is present [American Academy of Pediatric Dentistry, 2012].

Host susceptibility is underlined by a potential genetic contribution for caries risk. The understanding of genetic contributions to caries can be highly valuable for dental practitioners. In the future, clinicians might be able to explain to patients that some forms of caries are more strongly associated with inherited risk. This could offer an explanation for both the patient and dental practitioner why persons with similar behavioral risks (i.e. tooth brushing frequency or dietary habits) have different caries risk and/or caries activity [Bretz et al., 2003].

**Early Evidence for a Genetic Component of Caries**

Although a genetic contribution to caries in humans has been controversial, even the earliest studies of family patterns in twins [Boraas et al., 1988; Conry et al., 1993],

---

*Vieira/Modesto/Marazita*
families [Klein and Palmer, 1940] and animal breeding [Hunt et al., 1944] were consistent with a genetic component. The most compelling early evidence for a genetic component to caries comes from studies of twins reared apart. In two related studies, investigators found significant resemblance for percentage of teeth and surfaces restored or carious within monozygotic but not dizygotic twin pairs reared apart and estimated the genetic contribution to caries as 40% [Boraas et al., 1988; Conry et al., 1993]. More recent studies of twins reared together [Bretz et al., 2003, 2005] or of families [Wang et al., 2010] estimated the heritability for caries measured by the DMFT/S as ranging from 45 to 64%, with primary dentition caries in general showing higher heritability than permanent dentition caries. Heritability studies alone are not sufficient to demonstrate a genetic component in caries because shared behavior and other environmental factors can contribute to covariance between relatives and mimic genetic correlation. Notably however, genetic studies in animal models (see e.g. Nariyama et al. [2004]) and recent human molecular genetic studies bear out the heritability results. A full summary of the long history of animal model studies in caries is beyond the scope of this review, but note that many of the candidate gene studies done in humans (see table 1 for a summary) were motivated by the findings in animal models. This review focuses in large part on the recent strides in caries molecular genetic studies in human populations.

Caries Phenotypes

Shared behavior, practice and habits within families can be expected to contribute to covariance between relatives and to mimic genetic correlation if not controlled by the experimental design [Potter, 1990]. This is particularly true for caries, which can be profoundly affected by dietary sugar intake and/or oral hygiene practices within families. This effect can explain results that suggest a major gene effect [Werneck et al., 2011], despite the fact that one would not expect that to be the case for caries.

Successful genetic studies require careful measurement of the phenotype of interest (i.e. the disorder or physical characteristic under study), ideally in a biologically meaningful way. Genetic studies of caries are thus inherently difficult because the usual disease measures originally developed for clinical and epidemiological studies more directly assess the consequence of the disease rather than its pathogenesis. The DMFT/DMFS score (i.e. a count of the number of Decayed/Missing due to caries/Filled Teeth or Surfaces) is the most widely used index and gives a good estimate of the prevalence of the disease, as well as an estimate of the severity of the affection based on the number of teeth (or surfaces) affected by caries. The DMFT distribution in 12-year-olds across many countries shows that there is a skewed distribution of caries prevalence [Nishi et al., 2002]. While a proportion of 12-year-olds are caries-free, a considerable number still have DMFT values higher than the goal of 3 set by the Health Assembly of the World Health Organization (Bratthall [2000]; see also fig. 1). Caries prevalence assessed in the third of the population with higher DMFT scores (Significant Caries Index) brings to attention the individuals that are experiencing most severe disease. It has been proposed that the Significant Caries Index for countries should be less than 3 DMFT in 12-year-olds by the year 2015 [Bratthall, 2000].

Over the life course, populations tend to have increased DMFT/DMFS scores (fig. 2), in part due to untreated caries disease, in part due to restorative procedures and other reasons. Therefore, genetic studies defining the phenotype based on caries experience scores will have to deal with the limitation that these scores provide an estimate of the prevalence of the disease, but not necessarily directly determine the individual factors involved in its pathogenesis. DMFT scores have limited ability to provide insight into the severity or pathogenesis of the caries experience in an individual. With the widespread utilization of fluorides, carious lesions have become more difficult to detect and many are incipient. Epidemiological assessments usually have study participants at only one time and judgment of lesion status needs to be made at the visit. Dentists in their practices have the ability to follow up cases and make decisions based on how an apparent lesion progresses or stays arrested, and treatment decisions can be planned accordingly. To sophisticate the power of discrimination of caries experience evaluations at the population level, another index was proposed, the International Caries Detection and Assessment System (ICDAS; https://www.icdas.org/; Pitts [2004]). This index provides a more sophisticated assessment of the caries experience, but at the same time brings to attention the difficulties of discerning between specific codes of early alteration of the enamel (codes 0–3) and the feasibility of applying this approach to large settings.

Both DMFT/DMFS and ICDAS are used to obtain information on primary dentition, which is ideally examined after complete eruption, after 4 and before 6 years of age. Caries experience data have been generated for individuals older than 12 years of age as well, but in these
### Table 1. Summary of the candidate genes studied for caries in humans

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Summary of reported results</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enamel formation genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuftelin (TUFT1)</td>
<td>expressed in the developing and mineralized tooth and nonmineralizing soft tissues [Deutsch et al., 2002]</td>
<td>associated with higher caries experience; this association can be dependent of the presence of <em>Streptococcus mutans</em> no evidence of association</td>
<td>Slayton et al. [2005]; Deeley et al. [2008]; Patir et al. [2008]; Shimizu et al. [2012]</td>
</tr>
<tr>
<td>Amelogenin (AMELX)</td>
<td>involved in the mineralization during tooth enamel development</td>
<td>associated with higher caries experience no evidence of association</td>
<td>Deeley et al. [2008]; Patir et al. [2008]; Kang et al. [2011]; Shimizu et al. [2012]; Jeremias et al. [2013]</td>
</tr>
<tr>
<td>Enamelin (ENAM)</td>
<td>involved in the mineralization and structural organization of the enamel</td>
<td>associated with higher caries experience no evidence of association</td>
<td>Patir et al. [2008]; Shimizu et al. [2012]; Jeremias et al. [2013]</td>
</tr>
<tr>
<td>Tuftelin-interacting protein 11 (TFIP11)</td>
<td>thought to interact with tuftelin and can play a role in spliceosome disassembly in Cajal bodies [Stanek et al., 2008]</td>
<td>associated with initiation of carious lesions and higher caries experience no evidence of association when DMFT scores are analyzed</td>
<td>Shimizu et al. [2012]; Jeremias et al. [2013]</td>
</tr>
<tr>
<td>Ameloblastin (AMBN)</td>
<td>involved in enamel matrix formation and mineralization</td>
<td>associated with higher caries experience no evidence of association</td>
<td>Patir et al. [2008]; Shimizu et al. [2012]; Ergöz et al. [2013]</td>
</tr>
<tr>
<td>Matrix metalloproteinase 20 (MMP20)</td>
<td>degrades amelogenin</td>
<td>associated with higher caries experience in Whites with poor oral health habits no evidence of association</td>
<td>Tannure et al. [2012b]</td>
</tr>
<tr>
<td>Kallikrein-related peptidase 4 (KLK4)</td>
<td>degrades enamel proteins</td>
<td>associated with lower caries experience no evidence of association</td>
<td>Wang et al. [2012b]</td>
</tr>
<tr>
<td><strong>Immune response genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD14 molecule (CD14)</td>
<td>mediates innate immune response to bacterial lipopolysaccharide</td>
<td>absent in the saliva of individuals with active carious lesions</td>
<td>Bergandi et al. [2007]</td>
</tr>
<tr>
<td>Gene</td>
<td>Function</td>
<td>Summary of reported results</td>
<td>References</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>-----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Human leukocyte antigen; major histocompatibility complex, class II, DR beta 1 (HLA-DRB1) and DQ beta 1 (HLA-DQB1)</td>
<td>presents peptides derived from extracellular proteins</td>
<td>frequency of allele 4 of DRB1 is increased in children with early childhood caries; also allele 2 of DQB1 is increased in adolescents affected by caries; DRB1 allele 1 and DQB1 allele 3 frequencies are increased in the presence of <em>Streptococcus mutans</em></td>
<td>Lehner et al. [1981]; Altun et al. [2008]; Bagherian et al. [2008]; Valarini et al. [2012]</td>
</tr>
<tr>
<td>Beta defensin 1 (DEFB1)</td>
<td>antimicrobial peptide implicated in the resistance of epithelial surface to microbial colonization</td>
<td>distinct <em>DEFB1</em> haplotypes are associated with low and high caries experience</td>
<td>Ozturk et al. [2010]; Krasone et al. [2013]</td>
</tr>
<tr>
<td>Lactotransferrin (LTF)</td>
<td>major iron-binding protein in milk and body secretions with antimicrobial activity</td>
<td>associated with lower caries experience</td>
<td>Azevedo et al. [2010]; Fine et al. [2013]</td>
</tr>
<tr>
<td>Mucin 7 (MUC7)</td>
<td>facilitates the clearance of bacteria in the oral cavity</td>
<td>associated with higher caries experience with poor oral hygiene</td>
<td>Pol [2011]</td>
</tr>
<tr>
<td>Mannose-binding lectin (protein C) 2, soluble (MBL2)</td>
<td>recognizes mannose and N-acetylglucosamine on many microorganisms</td>
<td>associated with higher caries experience</td>
<td>Olszowski et al. [2012]</td>
</tr>
<tr>
<td>Mannan-binding lectin serine peptidase 2 (MASP2)</td>
<td>bactericidal factor that binds to the Ra and R2 polysaccharides expressed by certain enterobacteria</td>
<td>no evidence of association</td>
<td>Olszowski et al. [2012]</td>
</tr>
<tr>
<td>Saliva genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aquaporin 5 (AQP5)</td>
<td>water channel protein that plays a role in the generation of saliva, tears and pulmonary secretions</td>
<td>associated with higher caries experience</td>
<td>Wang et al. [2012b]</td>
</tr>
<tr>
<td>Protein-rich protein HaeIII subfamily 1 (PRH1)</td>
<td>provide protective environment for the teeth</td>
<td>associated with higher caries experience and colonization by <em>Streptococcus mutans</em></td>
<td>Zakhary et al. [2007]</td>
</tr>
<tr>
<td>Other genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix metalloproteinase 2 (MMP2)</td>
<td>degrades type IV collagen</td>
<td>no evidence of association</td>
<td>Tannure et al. [2012a]</td>
</tr>
<tr>
<td>Matrix metalloproteinase 9 (MMP9)</td>
<td>degrades type IV and V collagens</td>
<td>no evidence of association</td>
<td>Tannure et al. [2012a]</td>
</tr>
</tbody>
</table>
cases, total caries experience tends to increase due to other factors that cannot be easily discerned from caries: sound surfaces may have been prepared to receive tooth restorations or are bases for fixed prosthodontics, teeth can be lost due to trauma, periodontal disease or orthodontic indications, and depending on the study, these factors cannot be easily detected during the clinical assessment.

For human genetic study purposes, research groups have begun to use tooth- and/or surface-specific scores, for example dividing the surfaces into pit-and-fissure versus smooth surfaces [Zeng et al., 2013]. An intriguing recent study used cluster analysis to divide the tooth surfaces into groups with similar caries experience [Shaffer et al., 2013b], resulting in five primary clusters summarized in footnote ‘e’ of Table 2. Of note some clusters of

### Table 1 (continued)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Summary of reported results</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix metalloproteinase 13 (MMP13)</td>
<td>involved in endochondral ossification and bone remodeling; this gene is physically close to MMP20 and maybe evolved at the same time before the divergence of ray-finned and lobe-finned fish [Kawasaki and Suzuki, 2011]</td>
<td>associated with lower caries experience</td>
<td>Tannure et al. [2012a]</td>
</tr>
<tr>
<td>TIMP metallopeptidase inhibitor 2 (TIMP2)</td>
<td>possibly critical to the maintenance of tissue homeostasis by suppressing the proliferation of quiescent tissues in response to angiogenic factors, and by inhibiting protease activity in tissues undergoing remodeling of the extracellular matrix</td>
<td>no evidence of association</td>
<td>Tannure et al. [2012a]</td>
</tr>
<tr>
<td>Taste receptor, type 2, member 38 (TAS2R38)</td>
<td>controls the ability to taste glucosinolates</td>
<td>associated with lower caries experience</td>
<td>Wendell et al. [2010]</td>
</tr>
<tr>
<td>Taste receptor, type 1, member 2 (TAS1R2)</td>
<td>sweet taste receptor</td>
<td>associated with both lower and higher caries experience</td>
<td>Wendell et al. [2010]; Kulkarni et al. [2013]</td>
</tr>
<tr>
<td>Guanine nucleotide-binding protein, alpha transducing 3 (GNAT3)</td>
<td>believed to be involved in dietary preferences</td>
<td>no evidence of association</td>
<td>Wendell et al. [2010]</td>
</tr>
<tr>
<td>Solute carrier family 2 (facilitated glucose transporter), member 2 (SLC2A2)</td>
<td>mediates facilitated bidirectional glucose transport</td>
<td>associated with higher caries experience</td>
<td>Kulkarni et al. [2013]</td>
</tr>
<tr>
<td>Dentin sialophosphoprotein (DSPP)</td>
<td>involved in the mineralization process of dentin</td>
<td>associated with lower caries experience</td>
<td>Wang et al. [2012b]</td>
</tr>
<tr>
<td>Secreted phosphoprotein 1 (SPP1)</td>
<td>involved in the attachment of osteoclasts to the mineralized bone matrix</td>
<td>no evidence of association</td>
<td>Wang et al. [2012b]</td>
</tr>
<tr>
<td>Carbonic anhydrase IV (CA4)</td>
<td>involved in respiration, calcification, acid-base balance, bone resorption and formation of aqueous humor, cerebrospinal fluid, saliva and gastric acid</td>
<td>no evidence of association</td>
<td>Yarat et al. [2011]</td>
</tr>
</tbody>
</table>
surfaces did not show significant heritability (e.g. maxillary incisors, see table 2), indicating that perhaps some clusters are not under genetic control while other clusters do have genetic contributions to risk of caries.

**Human Genetic Studies of Caries**

There are two major ways to study the genetics of a complex trait such as caries, and both have been applied to caries: candidate gene studies (summarized below and in table 1) and genome-wide studies (summarized below and in table 2). Candidate gene studies test hypotheses regarding association between specific genes or gene variants and caries experience. Genome-wide studies test either linkage or association between anonymous DNA variants with known locations throughout the genome, and thus represent a hypothesis-generating procedure. DNA variants with genome-wide significant statistical signals imply that the anonymous variants are near caries etiologic variants; such positive signals then require follow-up studies for identification of the etiologic variants. Since genome-wide studies are hypothesis-generating, one typically targets for follow-up both strictly genome-wide significant results and also suggestive results, i.e. those results that are near strict significance.

**Candidate Gene Studies**

Most of the genetic studies of caries to date approach the problem of detecting a genetic factor contributing to caries by testing genetic variation in specific genes, based on their assumed or known function which is thought to be relevant to the disease, using the standard statistical approach of testing for association between specific vari-

---

Fig. 1. Graphs showing caries data for two populations of 12-year-olds, expressed as frequency distribution. a Data (n = 539) from Curitiba, Brazil (principal investigator Paula C. Trevilatto) included in Shimizu et al. [2012]. b Data (n = 368) from the Center for Oral Health Research in Appalachia, USA (principal investigator Mary L. Marazita) included in Wang et al. [2012a]. 48% of children from the Brazilian site (n = 258) and 28% of the children from the US site (n = 102) were caries-free. The mean DMFT was 1.45 (Brazilian site) and 3.45 (US site), respectively. The Significant Caries Index calculated from the third of the population with the highest caries scores was 3.64 (Brazilian site) and 7.82 (US site), respectively. The third of the population with the highest caries experience likely includes individuals who accumulate the greatest number of genetic factors increasing susceptibility to caries. Among the caries-free individuals are probably some individuals that have a greater number of genetic factors protecting against caries risk. Environmental factors such as dietary habits, fluoride exposure and access to dental care can easily overcome genetic factors and also explain differences in caries experience in the two cohorts.
Are individuals with DMFT = 1 the same as individuals with at least one affected tooth. The obvious question: Are individuals with DMFT = 10? The studies that take into consideration age when evaluating the DMFT likely provide a better estimate of genetic associations.

Another interesting phenotype is a surrogate of the carious lesion severity. Some individuals can have a more dramatic progression of the lesion into dentin than others, and these lesions can involve the pulp to a point that these cases can be more susceptible to developing periapical lesions detected radiographically. MMP2 expression is higher in dentin affected by caries [Toledano et al., 2010]. When the presence of periapical lesions associated with deep carious lesions in dentin was used as a phenotype (in comparison to absence of periapical lesions despite the presence of deep carious lesions in dentin), associations were found with MMP2 and MMP3 [Menezes-Silva et al., 2012], demonstrating the promise of exploring phenotypes related to the severity of the carious lesions.

**Genome-Wide Linkage Studies**

With early molecular genetic tools such as restriction fragment polymorphisms and single nucleotide polymorphisms (SNPs), genome-wide studies began to identify regions in the genome likely to harbor caries risk genes. The first genome-wide attempt to identify genetic contributors to caries used the linkage approach [Vieira et al., 2008] (see also Table 2). Linkage studies utilize the recombination that occurs between genetic loci that are near each other on the same chromosome during crossing over of homologous chromosomes during meiosis I. The recombination frequency is a function of distance between loci, and the larger the estimated recombination frequency, the lower the likelihood that the loci are linked (i.e. close to each other). If a genetic marker is physically close or in the vicinity of the genetic variant causing the disease, there will be statistically significant evidence of linkage, as estimated by a statistic termed LOD score, which has a value of 3.4 for genome-wide statistically significant linkage or between 2 and 3.4 for suggestive evidence of linkage.

The genome-wide linkage scan performed for caries controlled for several key environmental factors that influence caries. The 46 families studied shared similar cultural habits, were all from a specific geographic area in the Philippines and had very limited access to dental care. The analysis was done twice, once utilizing a definition of the phenotype as discussed earlier. Most studies compare individuals who are caries-free to individuals with at least one affected tooth. The obvious question is ‘Are individuals with DMFT = 1 the same as individuals with DMFT = 10?’. The studies that take into consideration age when evaluating the DMFT likely provide a better estimate of genetic associations.

### Table 1: Candidate Gene Studies

<table>
<thead>
<tr>
<th>Study Descriptions</th>
<th>LOD score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brancher et al., 2011; Yarat et al., 2011; Buczewska-Radlińska et al., 2012; Oliszewski et al., 2012; Gasse et al., 2013; Yang et al., 2013</td>
<td></td>
</tr>
</tbody>
</table>

The most studied group of candidate genes includes the enamel formation genes. The aggregate association data [Slayton et al., 2005; Deley et al., 2008; Patir et al., 2008; Shimizu et al., 2012; Ergöz et al., 2013; Jeremias et al., 2013] have only one mostly consistent result, the lack of association between TFIP11 and caries (Jeremias et al. [2013] is the exception). However, an association with this gene was found when the phenotype tested was microhardness of enamel after the creation of an artificial caries lesion [Shimizu et al., 2012]. These data suggest that the bulk of candidate gene studies can succeed from the definition of the phenotype as discussed earlier. Most studies compare individuals who are caries-free to individuals with at least one affected tooth. The obvious question is ‘Are individuals with DMFT = 1 the same as individuals with DMFT = 10?’. The studies that take into consideration age when evaluating the DMFT likely provide a better estimate of genetic associations.

![Fig. 2. DMFT scores tend to grow as individuals grow older. The effect of age in any caries experience score will potentially add undetectable bias since it is not possible to account for all factors contributing to increasing levels of tooth loss and affection, and these likely impact genetic analyses (data from Jindal et al. [2011]).](image-url)
Table 2. Summary of genome-wide studies of dental caries

<table>
<thead>
<tr>
<th>Reference and study population</th>
<th>Phenotype</th>
<th>Results</th>
<th>Follow-up studies: locus (gene), reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genome-wide linkage studies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vieira et al. [2008]; 46 Filipino extended kindreds, 642 individuals</td>
<td>DMFT, categoricald</td>
<td>low caries experience: 5q13.3, 14q11.2, Xq27.1</td>
<td>(1) 5q13.3 (BTF3), Shimizu et al. [2013]; (2) 14q11.2 (TRAV4), Briseño-Ruiz et al. [2013]; (3) 13q31.1 (intergenic), Küchler et al. [2013]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>high caries experience: 13q31.1, 14q24.3</td>
<td></td>
</tr>
<tr>
<td><strong>Genome-wide association studies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shaffer et al. [2012]; 1,305 US White children aged 3 – 12 years</td>
<td>primary dentition, categorical, dftr ≥1</td>
<td>1q42–q43 (ACTN2, MTR, EDARADD), 11p13 (MPPED2), 17q23.1 (LPO)</td>
<td>(1) no significant replication in 1,695 Danish children aged 2 – 7 years [Shaffer et al., 2012]; (2) significant replication utilizing gene set enrichment [Wang et al., 2013]; see below in this table</td>
</tr>
<tr>
<td>Wang et al. [2012a]; 7,443 US White adults</td>
<td>permanent dentition, categorical, DMFT ≥1</td>
<td>4q32 (TLR2), 5q11.1 (ISL1), 6q27 (RP56KA2), 7q21 (FZD1), 14q22 (CDKN3, CNIH, GMFB, CGRRF1, BMP4), 1q42.11–q42.3 (RHOU), 4q13.3 (ADAMTS3), 8p21.1 (PTK2B)</td>
<td></td>
</tr>
<tr>
<td>Shaffer et al. [2013a]; 920 US White adults (17 – 75 years; a subset of subjects from Wang et al. [2012a])</td>
<td>five quantitative permanent dentition DMFS scores (DMFS1 – 5) derived from hierarchical cluster analysis</td>
<td>DMFS2: 10p11.23 (LYZL2) DMFS5: 1p36 (AJAP1)</td>
<td></td>
</tr>
<tr>
<td>Zeng et al. [2013]; 1,017 US White adults (14 – 56 years; a subset of subjects from Wang et al. [2012a])</td>
<td>two quantitative permanent dentition DMFS scores (pit/ fissure, smooth)f</td>
<td>smoothf: 8q21.3 (no gene) pit/fissuref: 7p15 – 13 (INHBA), 8q21.3 (no gene), Xp11.4 (BCOR) smoothf: 2q35 (CXCR1, CXCR2), Xq26.1 (BCORL1)</td>
<td></td>
</tr>
</tbody>
</table>
Reference and study population

Phenotype

Results

Follow-up studies: locus (gene), reference

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Results</th>
<th>locus (gene)</th>
</tr>
</thead>
<tbody>
<tr>
<td>two quantitative primary dentition dfs scores (pit/fissure, smooth)</td>
<td>pit/fissure: 3q26.1 (KPNA4)</td>
<td>multiple suggestive gene sets, see Wang et al. [2013]</td>
</tr>
<tr>
<td></td>
<td>pit/fissure: 11p14.1 (MPPED2) smooth: 1p36 (AJAP1), 6q27 (RP56KA2), 16p11.2 (ITGAL), 20q11.21 (PLUNC)</td>
<td></td>
</tr>
</tbody>
</table>

Genome-wide association studies with gene set enrichment

| Wang et al. [2013]; 1,142 US White children (3–12 years; a subset of children included in Shaffer et al. [2012]) | categorical primary dentition dfs score, dfs ≥1 adjusted for age | gene sets: (1) cytokine secretion (18, including INS); (2) ligase activity forming carbon nitrogen bonds (68, including WWP2, RNF217); (3) ubiquitin protein ligase activity (49); (4) protein secretion (32); (5) regulation of protein secretion (22); (6) regulation of cytokine secretion (16); (7) regulation of axonogenesis (10); (8) regulation of neurogenesis (14, including ROBO2, SLIT2); (9) central nervous system development (122); (10) small conjugating protein ligase activity (51); (11) glycoprotein catabolic process (12); (12) axonogenesis (43); (13) cell matrix junction (16; includes ACTN2) | multiple suggestive gene sets, see Wang et al. [2013] |

---

a Genome-wide linkage utilizes techniques of genetic linkage analysis and a panel of multi-allelic markers evenly spaced throughout the genome; genome-wide association utilizes standard association analysis methods to detect the relationship between a trait and anonymous bi-allelic SNP variants.

b 'Significant' = genome-wide significant, i.e. for genome-wide linkage: LOD scores ≥3.4; for genome-wide association: p value ≤10−7, additional adjustments for multiple phenotypes are also done.

c 'Suggestive': for genome-wide linkage: LOD scores between 2 and 3.4; for genome-wide association: p value between 10−5 and 10−7.

d Categories (see Vieira et al. [2008]) in children (≤13 years): caries experience very low (DMFT = 0–1), low (DMFT = 2), moderate (DMFT = 3–4), high (DMFT ≥5); in teenagers (13–18 years): caries experience very low (DMFT = 0–2), low (DMFT = 3–5), moderate (DMFT = 6–8), high (DMFT ≥9); in adults: caries experience very low (DMFT = 0–4), low (DMFT = 5–8), moderate (DMFT = 9–13), high (DMFT ≥14).

e Hierarchical clusters (see Shaffer et al. [2013b]): DMFS1 (DMFS scored from the occlusal surfaces of molars; heritability (h²) = 27%, p = 0.06), DMFS2 (mandibular incisors, canines and first premolars; h² = 54%, p = 0.002), DMFS3 (molars and maxillary premolars excluding occlusal molar surfaces; h² = 43%, p = 0.004), DMFS4 (maxillary incisors; h² = 0%, p = 0.5) and DMFS5 (canines and premolars; h² = 40%, p = 0.008).

f DMFS calculated on pit and fissure surfaces, and DMFS calculated on smooth surfaces, with each score adjusted for age, sex, presence of Streptococcus mutans and home water fluoride level.

g dfs calculated on pit and fissure surfaces adjusted for age and age², dfs on smooth surfaces adjusted for age.

h Gene sets are combinations of genes that are functionally related (terms used here are from the Gene Ontology of Ashburner et al. [2000]); the number in parentheses is the number of genes tested within the set [Wang et al., 2013]; selected gene names with known/suspected involvement in oral or dental phenotypes are also listed. Terms included in this table are those significant after adjustment for multiple testing (FDR <0.05) under at least one of four statistical methods utilized for the gene set analyses (GenGen, Alligator, SNP ratio test, mixed model; see Wang et al. [2013]). Bold terms are those that were significant under more than one statistical method. FDR = False discovery rate.
Dry tooth for long enough to dehydrate any possible lesion (approximately 5 seconds).

Is there undermining discoloration of the dentin? This is often seen better when wet.

Yes

Code 4

No

Yes

Code 3

Is there an opacity/discoloration at the entrance to the fissure?

No

Code 0

Sound

Yes

Code 1

First visible sign of caries

Are there any visible signs of caries when the tooth has been cleaned and when viewed wet?

No

Yes

Is there a microcavity or obvious cavity?

No

Yes

Code 2

Is dentin visible at the base of the cavity?

No

Yes

Code 5

Is more than half of the crown involved in caries?

Yes

Code 6

Is there a distinct opacity at the entrance to the fissure or slightly beyond?

No

Yes

Microcavitation

Fig. 3. Questionable carious lesions can be managed with minimal intervention. a A 9-year-old patient with discoloration on an occlusal fissure of the right maxillary first molar. Radiographic exam showed no signs of dentine affection. The dental explorer was used gently and suggested the fissure was retentive. The mother of the child was instructed to carefully brush that tooth with an over the counter fluoridated toothpaste every day at least before the child going to bed. b After 5 years, there were no signs of lesion progression, both clinically and radiographically, which suggests the decision of not performing restorative treatment was for the benefit of the child. c Diagram of the decision making tree of the ICDAS (modified from https://www.icdas.org). The surface of the tooth above would have been possibly coded as 3 in a larger-scale study, even though it was sound.
The results suggested five loci linked to caries experience (LOD scores above 2.0 or non-parametric \( p \) values <0.0009), three for low caries experience (10p11.23 and Xq27.1) and two for high caries experience (13q31.1 and 14q24.3). Follow-up fine mapping of these regions using association tests was completed for 5q13.3 [Shimizu et al., 2013], 14q11.2 [Briseño-Ruiz et al., 2013] and 13q31.1 [Küchler et al., 2013] in 72 Filipino families and additional population data sets from multiple sources. In chromosome 5, BTF3 was a gene in the region flanking association signals that showed gene expression levels in whole saliva associated with caries experience. Similarly, in chromosome 14, genetic markers flanking TRAV4 were associated with low caries experience, and TRAV4 expression in whole saliva of individuals with low caries experience was higher in children and teenagers in comparison to adults. An intergenic SNP at 13q31.1 was associated with high caries experience and was predicted to disrupt the binding sites of two different transcription factors. Fine mapping of the two remaining regions is ongoing.

**Genome-Wide Association Studies**

Genome-wide association studies generally utilize large panels of SNPs, with typically \( \geq 600,000 \) SNPs actually genotyped and millions more SNPs imputed from the genotyped data, all of which are used for analysis. A typical threshold for genome-wide statistical significance for such studies is a \( p \) value \( \leq 10^{-7} \) for a 600,000 SNP panel, and a \( p \) value between \( 10^{-5} \) and \( 10^{-7} \) is considered suggestive. Caries genome-wide association studies are summarized in table 2.

The first genome-wide association studies for caries, one for the primary dentition [Shaffer et al., 2012] and one for the permanent dentition [Wang et al., 2012a], suggested different loci than the ones reported in the earlier genome-wide linkage study, which is not surprising given the relative strengths of the two approaches (see below). The analysis of the primary dentition [Shaffer et al., 2012] did not unveil any formal statistically significant association if multiple testing corrections are implemented, but had three loci with suggestive results: 1q42–q43, 11p13 and 17q23.1. This study was done in 1,305 US children 3–12 years of age (table 2). The analysis was repeated taking into consideration home fluoride exposure data, which were available for 720 children; suggestive results were found for 22q12.1 when fluoride exposure was suboptimal and for 1p34 and 6q16.1 when fluoride exposure was optimal. However, when these initial results were tested in an independent sample of 1,695 Danish children aged 2–7 years, no associations were replicated with statistical significance [Shaffer et al., 2012].

The genome scan of caries in the permanent dentition [Wang et al., 2012a] included 7,443 subjects from five studies from the US that were analyzed separately because caries assessments were done differently, DNA sources were different and genotyping was done at different times in distinct platforms. Individuals having at least one affected tooth in any of the studies were considered affected. Similar to the caries in the primary dentition scan, the results for the permanent dentition yielded a few loci with nominal genome-wide significance (i.e. \( p \) value \( \leq 10^{-7} \)) after adjustment of multiple testing. Given the multiple populations included in this study, these results were appropriately considered suggestive, as were several other results with \( p \) values between \( 10^{-5} \) and \( 10^{-7} \) (table 2).

A portion of the data included in the permanent dentition analyses [Wang et al., 2012a] was re-analyzed twice under surface-specific phenotypic definitions that might be more biologically meaningful than the overall DMFS. The first re-analysis defined affection status by hierarchical cluster analysis of tooth-surface level data [Shaffer et al., 2013a, b]. These analyses yielded five distinct phenotype definitions (DMFS1–5, see footnote ‘e’ in table 2) and resulted in genome-wide statistically significant results for two loci: one for caries in anterior mandibular teeth (DMFS2) and a marker close to \( LYZL2 \) (10p11.23; \( p \) value = 9 \( \times \) 10\(^{-8} \)), and one for caries in canines and premolars (DMFS5) and a marker close to \( A J A P I \) (1p36.32; \( p \) value = 2 \( \times \) 10\(^{-8} \)). Additional suggestive loci are summarized in table 2.

A second re-analysis of the permanent dentition data included two affection definitions: the presence of at least one molar occlusal surface affected (pit and fissure) or the presence of at least one smooth surface affected [Zeng et al., 2013]. These analyses yielded a few genome-wide significant or suggestive associations for each phenotype definition. For occlusal caries, markers flanking \( B C O R \) (Xp11.4; \( p \) value = 1.8 \( \times \) 10\(^{-7} \)) and \( I N H B A \) (7p15–p13; \( p \) value = 6.5 \( \times \) 10\(^{-6} \)) showed a trend for association. For smooth surface caries, markers flanking \( B C O R L 1 \) (Xq26.1; \( p \) value = 1.0 \( \times \) 10\(^{-7} \)) and \( C X C R 1 \) and \( C X C R 2 \) (2q35; \( p \) value = 1.9 \( \times \) 10\(^{-9} \)) had suggestive results. Similarly, re-analysis of the primary dentition data in Shaffer et al. [2012] included two affection definitions: the presence of at least one molar occlusal surface affected (pit and fissure) or the presence of at least one smooth surface affected [Zeng et al., 2013, in press]. The full results are summarized in table 2. Genome-wide significant association was observed with \( K P N A 4 \) (3q26.1) and pit and fisch
sure caries, and suggested associations with three genes that were observed in previous studies: MPPED2 (p value = 6.9 × 10^{-6}), AJAP1 (p value = 1.6 × 10^{-6}) and RPS6KA2 (p value = 7.3 × 10^{-6}).

In addition to the above association analyses of individual SNPs with caries, a recent analysis [Wang et al., 2013] applied a new statistical methodology to the primary dentition association results from Shaffer et al. [2012]. Instead of analyzing SNPs one by one, they used gene set-based analysis, which has recently emerged as a useful approach to examine the joint effects of multiple risk loci in complex human diseases or phenotypes such as caries. Wang et al. [2013] used four complementary gene set statistical analysis methods and analyzed 1,331 gene sets under Gene Ontology terms [Ashburner et al., 2000]. Identified were 13 significantly associated Gene Ontology terms/gene sets (table 2). 17 additional sets were further identified as marginally relevant. The identified gene sets encompass broad functions that potentially interact and contribute to the oral immune response related to caries development, which have not yet been reported in any standard single marker-based analysis. These approaches are under development, but it appears that the gene set enrichment analysis approach can provide complementary insights into the molecular mechanisms and polygenic interactions in caries.

As summarized by Wang et al. [2013], five genes from either ligase activity (WWP2 and RNF217), neuronal development (ROBO2 and SLIT2) or cytokine/protein secretion (INS) gene sets listed in table 2 might be potentially associated with dental traits. WWP2 is a member of ligase activity pathways and functions as a ligase for and mediates degradation of PTEN, whose gene is expressed in mouse oral development. RNF217 is located at 6q22.31, a genomic region reported to be associated with oral clefts. ROBO2 is a receptor for SLIT2 and possibly SLIT1 genes, which appear to work cooperatively to establish anatomical midlines during neuronal development and establishment of olfactory organization. SLIT1 is also expressed in the primary and secondary enamel knots during molar tooth cusp formation. INS can impact caries through insulin sensitivity. Insulin receptor binding sites are present on rat incisors. None of these relationships are apparently relevant for caries development, but the gene sets and the subset of tooth-related genes raise interesting possible mechanisms for caries.

**Summary of Genome-Wide Analyses**

When we compare the results across all these genome-wide analyses there are three notable conclusions:

(1) Very little overlap exists across studies, but the promising associations with 1p36.32 and 10p11.23 should be prioritized for future studies for the identification of genetic factors contributing to caries. (2) Phenotype definitions for caries warrant further refinement because suggestive but not formally significant results were obtained for the traditional DMFS/T caries definitions, but genome-wide significant results were found for surface-specific caries scores. (3) Utilizing approaches that look for joint effects of DNA variants across sets of related genes (such as the gene set enrichment approach of Wang et al. [2013]) appear to hold promise for identifying possibly functional relationships between caries-associated genetic factors.

Interestingly, as mentioned above, the significant or suggestive signals to date from genome-wide association studies of caries are different from the significant signals found from genome-wide linkage analyses. This fact emphasizes that the two approaches have different strengths: association studies are more sensitive in detecting common variants of small effect size than are linkage studies, but linkage is more robust in detecting etiologic genes that exhibit allelic heterogeneity – if multiple different variants (especially rare variants) within a gene can lead to caries, linkage is much more likely to detect such genes [Risch and Merikangas, 1996]. Thus further follow-up studies of the results from both types of studies are necessary to confirm genetic loci for caries.

**Final Remarks**

Advances in molecular genetics, engineering, management, research and education offer many new ways to treat patients regarding caries management and prevention. Advancement is underway in caries risk assessment, carious lesion detection, genetic predisposition technologies and restorative techniques [Berg, 2013]. The understanding of genetic contributions to caries can be highly valuable for dental practitioners as the starting point of host susceptibility. In the future, clinicians might be able to explain to patients that some forms of caries are more strongly associated with inherited risk, offering an explanation for both the patient and dental practitioner why similar behavioral risks (i.e. tooth brushing frequency or dietary habits) have different caries risk [Bretz et al., 2003]. Individuals at higher genetic risk could then be monitored more closely and provided with more aggressive caries management and prevention programs [Bretz et al., 2003].

---

Caries: Review of Human Genetics Research

Caries Res 2014;48:491–506
DOI 10.1159/000358333

503
There has been much interest and progress recently in identifying the genetic contributions to caries. The disease is still highly prevalent despite more than 100 years of studies clearly defining its pathogenesis. New strategies that can protect individuals at higher risk even when exposed to fluorides through drinking water or other sources are needed. Recent studies have shown promise in identifying promising biomarkers that have the potential to help in determining and personalizing treatment for individuals at higher risk to develop the disease. Initial findings suggest differential expression of genes and proteins in whole saliva (BTF3 [Shimizu et al., 2013], TRA V4 [Briseño-Ruiz et al., 2013] and total protein count [Roa et al., 2008]) bring the promise that personalized treatment for caries is on the horizon. Future studies should carefully consider phenotypic definitions and incorporate environmental exposures and demographic variables such as age and sex in the analysis. Models that take into consideration longitudinal evaluations of caries [Isaksson et al., 2013] can be the most promising. Taking advantage of clinical studies such as randomized trials [Bader et al., 2013] also can provide novel insight. In future work, an integration of other genetic and genomic information (such as metagenomics, gene expression, protein-protein interaction network, evidence from multiple species and multi-dimensional functional module analysis) can also open new avenues to understand the etiology of caries.

Acknowledgments

A.R. Vieira and M.L. Marazita are supported by NIH Grants R01-DE018914 and R01-DE014899. Sarah Vinski provided administrative support.

Disclosure Statement

The authors have no conflicts to disclose.

References


