9-1-2016

Extending genome-wide association study (GWAS) results to test classic anthropological hypotheses: Human third molar agenesis and the 'probable mutation effect'

Alexis Sullivan
aps216@psu.edu

Adrijana Vukelic
adrijana713@gmail.com

Jacob Cohen
jac6188@psu.edu

George (PJ) Perry
Pennsylvania State University, ghp3@psu.edu

Invited Commentary

Extending genome-wide association study (GWAS) results to test classic anthropological hypotheses: Human third molar agenesis and the 'probable mutation effect'

Adrijana Vukelic[1,2], Jacob A. Cohen[2], Alexis P. Sullivan[2], George H. Perry[1,2]

[1] Department of Anthropology, Pennsylvania State University, University Park, PA 16802

[2] Department of Biology, Pennsylvania State University, University Park, PA 16802

Date Submitted: May 10 2016
Date Accepted: Feb 24 2017

Running title: Human M3 agenesis evolutionary genetics

Keywords: Anthropometric traits; evolutionary genomics; Human dental formula evolution; wisdom tooth

Corresponding author: George H. Perry
513 Carpenter Building
Pennsylvania State University
University Park, PA 16802 USA
E-mail: ghp3@psu.edu
Phone: +1-814-863-7654

ACKNOWLEDGMENTS

The computational resource instrumentation used in this study was funded by the National Science Foundation (OCI–0821527). A.P.S. is supported by the National Science Foundation Graduate Research Fellowship Program (DGE1255832). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ABSTRACT

A genome-wide association study (GWAS) identifies regions of the genome that likely affect the variable state of a phenotype of interest. These regions can then be studied with population genetic methods to make inferences about the evolutionary history of the trait. There are increasing opportunities to use GWAS results – even from clinically-motivated studies – for tests of classic anthropological hypotheses. One such example, presented here as a case study for this approach, involves tooth development variation related to dental crowding. Specifically, more than 10% of humans fail to develop one or more permanent third molars (M3 agenesis). M3 presence/absence variation within human populations has a significant genetic component (heritability estimate $h^2 = 0.47$). The evolutionary significance of M3 agenesis has a long history of anthropological speculation. First, the modern frequency of M3 agenesis could reflect a relaxation of selection pressure to retain larger and more teeth following the origins of cooking and other food-softening behaviors (i.e., the genetic drift hypothesis, or classically, the “probable mutation effect”). Alternatively, commensurate with increasing hominin brain size and facial shortening, M3 agenesis may have conferred an adaptive fitness advantage if the risk of M3 impaction and potential health complications was reduced (i.e., the positive selection hypothesis). A recent GWAS identified 70 genetic loci that may play a role in human M3 presence/absence variation. To begin evaluating the contrasting evolutionary scenarios for M3 agenesis, we used the integrated haplotype score (iHS) statistic to test whether those 70 genetic regions are enriched for genomic signatures of recent positive selection. None of our findings are inconsistent with the null hypothesis of genetic drift to explain the high prevalence of human M3 agenesis. This result might suggest that M3 impaction rates for modern humans don’t accurately retrodict those of the pre-agricultural past. Alternatively, the absence of support for the positive selection hypothesis could reflect a lack of power; this analysis should be repeated following the completion of more comprehensive GWAS analyses for human M3 agenesis.

INTRODUCTION

The current workhorse approach for identifying genetic variants that underlie human complex traits – phenotypes influenced by combinations of multiple genetic loci and environmental factors – is the genome-wide association study, or GWAS (Bush and Moore 2012). Here, genotype data for large numbers of single nucleotide polymorphisms (SNPs; often for ~500,000 to ~2 million SNP loci across the genome per individual) are considered against the presence/absence status or quantitative values of a human trait (phenotype) for 100s to 100,000s of individuals per analysis. Outputs from a carefully designed and executed GWAS include the identification of genomic regions containing variants that are significantly associated with the expression of the trait of interest and the typical genotype effect sizes.

GWAS results provide a wealth of data for subsequent studies of the evolutionary histories of traits of interest. For example, one may test whether phenotype-associated genetic loci are characterized by or enriched for signatures of a past history of positive selection, with examples already published for stature, pigmentation, body mass index, markers of arsenic metabolism, and other variable human traits (Berg and Coop 2014; Field et al. 2016; Minster et al. 2016; Perry et al. 2014; Pickrell et al. 2009; Schlebusch et al. 2015; Turchin et al. 2012). The evolutionary analysis of trait-associated loci can sometimes become even more powerful and direct with the benefit of time-stamped ancient DNA data (Allentoft et al. 2015; Fehren-Schmitz and Georges 2016; Gamba et al. 2014; Lindo et al. 2016; Mathieson et al. 2015; Olalde et al. 2014; Pickrell and Reich 2014; Sams et al. 2015; Wilde et al. 2014).

This general approach can be used to test longstanding anthropological hypotheses, even with currently available data. While most GWAS published to date have understandably been motivated by medical rather than anthropological concerns, there is of course substantial intersection among traits of clinical and evolutionary interest given the relationship between health and fitness. In addition, GWAS results for non-clinical, anthropometric traits are increasingly available (e.g., Adhikari et al. 2016; Adhikari et al. 2015; Cole et al. 2016; Shaffer et al. 2016), in some cases facilitated by analyses of self-reported phenotype survey data for direct-to-consumer genomic data participants (Eriksson et al. 2010; Hu et al. 2016). As we look forward to increasing numbers of anthropologically-relevant GWAS datasets becoming available for evolutionary hypothesis testing, we present an initial test of a classic evolutionary hypothesis for an anthropometric trait, third molar agenesis, as a case study example that highlights some of the considerations, challenges, and opportunities of this approach.

Third Molar Agenesis

Approximately 10-25% of modern human adults fail to develop one or more wisdom teeth, or permanent third molars (M3 agenesis) (Carter and Worthington 2015; Crispim et al. 1972; Elomaa and Elomaa 1973; Garn et al. 1963; Grahnen 1956; Keene 1964; Lavelle and Moore 1973; Legovic et al. 1998; Nanda 1954; Rozkovcova et al. 1999; Thompson et al. 1974). In contrast, M3 agenesis almost never occurs in non-human apes (Colyer 1936; Lavelle and Moore 1973). Among all non-human primates, only marmoset and tamarin monkeys do not have permanent third molars (Swindler 2002). Human M3 presence/absence variation within populations has a significant genetic component, with a heritability estimate (h^2) = 0.47 (Grahnen 1956).

Two evolutionary scenarios – positive selection vs. genetic drift – compete to explain the high frequency of human M3 agenesis. Central to the positive selection hypothesis are observations of a gradual hominin jaw size decrease over the past 2 million years, from average maxillary and mandibular arch lengths of 84.2 and 84.8mm, respectively, in early representatives of the genus *Homo*, to only 65.6 and 61.6mm in Neolithic modern humans (Calcagno and Gibson 1991). Individuals with smaller teeth and M3 agenesis may have thus experienced a relative fitness advantage (Brothwell et al. 1963; Calcagno and Gibson 1988; Darwin 1874; Lavelle and Moore 1973) via protection from third molar impaction, otherwise caused by a lack of adequate eruption space for the late-developing third molar (Song et al. 2000). Among modern adults with non-extracted third molars, an average of 24% suffer from impaction (Carter and Worthington 2016) and serious associated medical complications including systemic infection (Berge 1996; Saglam and Tuzum 2003). Prior to recent advances in medicine and dental practice, these complications could be fatal (Mead 1928).

Alternatively, M3 agenesis may reflect a relaxation of selective constraint, if this tooth (perhaps more so than others in the dental arcade) was no longer necessary for adequate mastication following cultural changes in food production and preparation, such as cooking with fire (e.g., Gowlett and Wrangham 2013; Organ et al. 2011). That is, evolutionary pressure to *maintain* larger teeth and third molars may have been removed or reduced with the dietary transition to softened foods, such that any mutations resulting in smaller teeth or M3 agenesis might have been maintained and gradually increased in frequency through random neutral evolution processes (i.e., genetic drift). In the anthropological literature, this scenario has been discussed as the “Probable Mutation Effect”, a model of relaxed selective constraint combined with the

notion that mutations are more likely to erode rather than regenerate a function or structure (Brace 1963; Brace and Mahler 1971; Graber 1978; McKee 1984). For this paper we will refer to this general evolutionary scenario simply as genetic drift.

To begin evaluating these evolutionary scenarios for human M3 agenesis, in this study we test whether regions of the genome previously associated with M3 presence/absence variation (Haga et al. 2013) are enriched for genomic signatures of recent positive selection (Voight et al. 2006).

MATERIALS AND METHODS

Haga and colleagues (2013) performed a GWAS to identify genetic variants associated with M3 agenesis in a population sample of 149 adult Korean and Japanese individuals with one or more missing M3 teeth and 338 control individuals from the same populations with all four M3 present. They generated genome-wide genotype data for ~700,000 single nucleotide polymorphisms (SNPs) for each individual using an Illumina OmniExpress Bead Chip, and identified 21, 26, and 30 SNPs that were associated ($P < 1 \times 10^{-4}$) with agenesis of 1-4 M3 teeth from any part of the mouth, agenesis of 1-2 maxillary M3, and agenesis of 1-2 mandibular M3, respectively (Haga et al. 2013). As expected, some of the 77 total identified SNPs were identified multiple times for the different types of M3 agenesis; our analyses of the Haga et al. (2013) results considered the set of 70 unique M3 agenesis-associated SNPs from the combined dataset (**Supplementary Table 1**).

To evaluate the biological plausibility of the Haga et al. (2013) M3 agenesis GWAS results, we tested whether the genes within or nearby their 70 SNPs are significantly enriched for those with known roles in tooth development processes. For this analysis, we first identified the RefSeq gene overlapping or nearest each of the 70 M3 agenesis-associated SNPs, based on gene location data for the human genome (hg18) obtained from the UCSC genome table browser. There were 51 unique genes for the 70 M3 agenesis-associated SNPs, as expected for GWAS results due to linkage disequilibrium among SNPs. We also downloaded a curated database of 254 genes whose expression has been recorded in developing teeth tissues (http://bite-it.helsinki.fi; downloaded on 20 October 2015; **Supplementary Table 2**).

To test the hypothesis of positive selection for M3 agenesis-associated SNPs, we obtained integrated haplotype score (iHS) values (Voight et al. 2006; http://haplotter.uchicago.edu)

computed from dense genome-wide SNP genotype data generated by the International HapMap Phase II (International HapMap Project Consortium 2007) for 90 unrelated individuals of Asian ancestry (45 individuals from Tokyo, Japan and 45 individuals from Bejing, China). Briefly, the iHS statistic can be used to identify genetic variants with relatively high frequencies given their inferred ages, based on patterns of surrounding haplotype variation, which may reflect a recent history of positive selection on those identified variants (Voight et al. 2006). For each of the M3 agenesis-associated loci from Haga et al. (2013), we considered a 5,000 bp window around the peak-associated SNP (collapsing to one 5,000 bp window for nearby SNPs), and computed the mean and median absolute iHS ($|iHS|$) values for all of the HapMap Phase II SNPs within these regions.

We used permutation schemes to estimate the probabilities i) that the number of tooth development genes among the set of 51 genes within or nearest the M3 agenesis-associated SNPs could be observed by chance, and ii) that the mean and median absolute $|iHS|$ values for the 70 M3 agenesis-associated SNP loci could be observed by chance. Specifically, for the first analysis, we randomly identified 70 autosomal SNPs from the set of SNPs included on the Illumina 700k Human OmniExpress Bead Chip and used the identical procedure as described above to identify the RefSeq gene overlapping or nearest each SNP. From the unique set of nearest RefSeq genes to these 70 SNPs, we randomly selected 51 genes and counted the number that were also in the database of 254 tooth development genes. We then repeated this procedure 10,000 times, and computed an empirical *P*-value as the proportion of permutated datasets with an equal or greater number of tooth development genes as our observation for the Haga et al. (2013) M3 agenesis-associated SNPs. The second analysis used a similar procedure, but with computation of the mean and median $|iHS|$ values for the HapMap SNPs within the 5k windows surrounding each permutation of the 70 random SNP positions, and the empirical *P*-value as the proportion of the 10,000 permutated results with $|iHS|$ values equal to or greater than that observed for the original dataset.

RESULTS

The goal of our study was to test whether a null hypothesis of neutral evolution could be rejected for a set of 70 genetic loci that were previously associated with human M3 agenesis (Haga et al. 2013). However, we needed to first confirm the biological plausibility of the 70 regions identified in the Haga et al. (2013) study. Specifically, most recent GWAS analyses on complex (e.g., polygenic) traits were conducted with thousands to tens of thousands of

individuals in order to confidently identify phenotype-associated genetic loci (e.g., Hu et al. 2016; Hyde et al. 2016; Locke et al. 2015; Wood et al. 2014), whereas Haga et al. (2013) performed their analyses with genotype data from 149 individuals with M3 agenesis and 338 controls (total n = 487). Moreover, a common GWAS standard is to conduct a similar analysis with a replication cohort (McCarthy et al. 2008), which has not yet been performed for the M3 agenesis trait.

Therefore, before proceeding with the evolutionary analysis of the Haga et al. (2013) loci, we tested whether the 70 SNPs they identified are found within or nearby genes with known tooth development roles significantly more often than expected by chance. While we would not expect all genes that influence M3 presence/absence to have yet been identified as tooth development genes, we would still expect enrichment for such loci nearby genetic variants identified by a successful GWAS for M3 agenesis.

Of the 51 unique genes overlapping or nearest the 70 M3 agenesis-associated SNPs, four genes (7.8%; *TGA4*, *ROBO2*, *IRX1*, and *APP*) were also recorded in the database of 254 tooth development genes. Based simply on the total number of 18,963 RefSeq genes in the autosomal genome and not considering the size of those genes, we can calculate that by chance alone we would expect to observe an intersection of only 0.68 of the 51 M3 agenesis genes (1.3%) with the tooth development gene dataset (Fisher's Exact Test; $P = 0.005$). We also conducted a permutation analysis (see *Methods*) to better account for gene size variation and the unequal distribution of genotyped SNPs across the genome ($P = 0.06$; **Figure 1**). Together, these results suggest the genetic regions identified by Haga et al. (2013) are likely at least enriched for those that do underlie M3 agenesis, encouraging us to proceed to the evolutionary analysis.

To evaluate the evidence for a history of recent positive selection on the M3 agenesis phenotype, we considered $|\text{iHS}|$ values computed for two populations (Chinese and Japanese; combined into one population) of Southeast Asian ancestry (Voight et al. 2006) for SNPs in 5 kb windows surrounding the M3 agenesis-associated loci (Haga et al. 2013). We selected these populations for our evolutionary analysis because the M3 agenesis GWAS was also performed in Southeast Asian populations (Japan and Korea; Haga et al. 2013). iHS is a haplotype-based statistic that aims to identify genomic variants whose frequencies may have been increased, at least in part, by a past history of positive selection (Voight et al. 2006). SNPs with higher $|\text{iHS}|$

values are more unusual relative to the genome-wide distribution, and more likely to have been affected by positive selection, than SNPs with lower |iHS| values. The mean and median |iHS| values in the windows surrounding the 70 M3 agenesis-associated loci were 0.74 and 0.61, respectively. These values were not exceptional compared to those obtained by permutation (*P* = 0.94 and *P* = 0.95, respectively; **Figure 2**). Similar results were obtained when considering only the highest |iHS| value within each 5 kb window surrounding the M3 agenesis-associated loci, rather than the values for all SNPs within each window (**Supplemental Figure 1**).

In addition, we conducted a *post-hoc* analysis, considering only the subset of four M3 agenesis-associated SNPs located within or nearby known tooth development genes (*TGA4*, *ROBO2*, *IRX1*, and *APP*; see above). The mean and median |iHS| values for the 5 kb windows surrounding these loci were 0.73 and 0.64, also not unexpected by chance based on our permutations (*P* = 0.64 and P = 0.73, respectively; **Supplemental Figure 2**). Thus, our analysis reveals no evidence for a history of recent positive selection on genetic regions associated with human M3 agenesis.

DISCUSSION

We performed a population genetic analysis to evaluate contrasting evolutionary hypotheses for the high frequency of human M3 agenesis. Specifically, we tested whether values from a commonly used statistic to identify signatures of recent positive selection (Voight et al. 2006) were significantly elevated at genetic regions previously associated with M3 presence/ absence (Haga et al. 2013) relative to the remainder of the genome. They were not. Thus, at the present time we cannot reject the null hypothesis of neutral evolution/genetic drift to explain the high prevalence of human M3 agenesis.

The possibility that there might not have been a selective advantage to M3 agenesis for archaic hominin and prehistoric human populations might seem difficult to reconcile with recent statistics on M3 impaction and infection (Berge 1996; Quek et al. 2003; Saglam and Tuzum 2003) and potential mortality prior to the availability of modern medical treatment (Mead 1928). However, recent morphometric analyses have shown that human jaw development is affected by subsistence strategy and diet, with hunter-gatherers having longer mandibles than agriculturalists (Pinhasi et al. 2015; von Cramon-Taubadel 2011). Thus, current degrees and rates of tooth crowding and impaction-related infection may not be representative of more than the relatively recent evolutionary history of human agriculturalists. While previous studies have

detected signatures of positive selection that acted within a similar timeframe on genetic variants associated with many other human traits (Laland et al. 2010), we might expect reductions in the power of these tests under relatively shorter timeframes of moderate selection pressures.

That said, we must emphasize that our failure to reject the null hypothesis of genetic drift in this study cannot be equated to a demonstration of the absence of positive selection. In order to provide a more definitive result, it will be important to return to this hypothesis following the completion of more powerful GWAS analyses for M3 agenesis, alongside improved methods for identifying genomic signatures of recent adaptation on polygenic traits (e.g., Berg and Coop 2014; Field et al. 2016; Perry et al. 2014; Schweizer et al. 2016; Stephan 2016; Wellenreuther and Hansson 2016). It will additionally be important – and potentially quite interesting from an evolutionary perspective, given the substantial variation observed in M3 agenesis frequencies among human populations (Carter and Worthington 2015; Rozkovcova et al. 1999) – to conduct similar analyses in global population samples outside of Southeast Asia.

LITERATURE CITED

Adhikari K., M. Fuentes-Guajardo, M. Quinto-Sanchez et al. 2016. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat. Commun.* 7:11616.

Adhikari K., G. Reales, A. J. Smith et al. 2015. A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat. Commun.* 6:7500.

Allentoft, M. E., M. Sikora, K. G. Sjogren et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167-172.

Berg, J. J. and G. Coop. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet.* 10:e1004412.

Berge, T. I. 1996. Incidence of infections requiring hospitalization associated with partially erupted third molars. *Acta Odontol. Scand.* 54:309-313.

Brace, C. L. 1963. Structural reduction in evolution. *Am. Nat.* 97:39-49.

Brace, C. L. and P. E. Mahler. 1971. Post-Pleistocene changes in the human dentition. *Am. J. Phys. Anthropol.* 34:191-203.

Brothwell, D. R., V. M. Carbonell, and D. H. Goose. 1963. Congenital absence of teeth in human populations. In: Brothwell, DR, editor. Dental anthropology. New York: Pergamon Press.

Bush, W. S. and J. H. Moore. 2012. Genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822.

Calcagno, J. M. and K. R. Gibson. 1988. Human dental reduction: natural selection or the probable mutation effect. *Am. J. Phys. Anthropol.* 77:505-517.

Calcagno, J. M. and K. R. Gibson. 1991. Selective compromise: evolutionary trends and mechanisms in hominid tooth size. In: Kelley, MA, and CS Larsen, editors. Advances in Dental Anthropology. New York: Wiley-Liss. p 59-76.

Carter, K. and S. Worthington. 2015. Morphologic and Demographic Predictors of Third Molar Agenesis: A Systematic Review and Meta-analysis. *J. Dent. Res.* 94:886-894.

Carter, K. and S. Worthington. 2016. Predictors of Third Molar Impaction: A Systematic Review and Meta-analysis. *J. Dent. Res.* 95:267-276.

Cole, J. B., M. Manyama, E. Kimwaga et al. 2016. Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. *PLoS Genet.* 12:e1006174.

Colyer, F. 1936. Variations and diseases of the teeth of animals. London: Bale and Danielsson.

Crispim, J., K. G. Trigueiro, F. Sa e Benevides et al. 1972. Third molar agenesis in a trihybrid Brazilian population. *Am. J. Phys. Anthropol.* 37:289-292.

Darwin, C. 1874. The descent of man and selection in relation to sex. New York: American Publishers Corp.

Elomaa, M., and E. Elomaa. 1973. Third molar aplasia and formation in orthodontic patients. *Proc. Finn. Dent. Soc.* 69:141-146.

Eriksson, N., J. M. Macpherson, J. Y. Tung et al. 2010. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* 6:e1000993.

Fehren-Schmitz, L. and L. Georges. 2016. Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. *Sci. Rep.* 6:23485.

Field, Y., E. A. Boyle, N. Telis et al. 2016. Detection of human adaptation during the past 2000 years. *Science* 354:760-764.

Gamba, C., E. R. Jones, M. D. Teasdale et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5:5257.

Garn, S., A. Lewis, and J. Vicinus. 1963. Third molar polymorphism and its significance to dental genetics. *J. Dent. Res.* 42:1344-1363.

Gowlett, J. A. J., and R. W. Wrangham. 2013. Earliest fire in Africa: towards the convergence of archaeological evidence and the cooking hypothesis. *Azania* 48:5-30.

Graber, L. W. 1978. Congenital absence of teeth: a review with emphasis on inheritance patterns. *J. Am. Dent. Assoc.* 96:266-275.

Grahnen, H. 1956. Hypodontia in the permanent dentition. *Odontol. Revy.* 7:1-101.

Haga, S., H. Nakaoka, T. Yamaguchi et al. 2013. A genome-wide association study of third molar agenesis in Japanese and Korean populations. *J. Hum. Genet.* 58:799-803.

Hu, Y., A. Shmygelska, D. Tran et al. 2016. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* 7:10448.

Hyde, C. L., M. W. Nagle, C. Tian et al. 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet*.

International HapMap Project Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-861.

Keene, H. J. 1964. Third molar agenesis, spacing and crowding of teeth, and tooth size in caries-resistant naval recruits. *Am. J. Orthod.* 50:445-451.

Laland, K. N., J. Odling-Smee, and S. Myles. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* 11:137-148.

Lavelle, C. L., and W. J. Moore. 1973. The incidence of agenesis and polygenesis in the primate dentition. *Am. J. Phys. Anthropol.* 38:671-679.

Legovic, M., L. Mady, M. Slaj et al. 1998. Development of third molars in children in Istria. *Coll. Antropol.* 22 Suppl:127-132.

Lindo, J., E. Huerta-Sanchez, S. Nakagome et al. 2016. A time transect of exomes from a Native American population before and after European contact. *Nat. Commun.* 7:13175.

Locke, A. E., B. Kahali, S. I. Berndt et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518:197-206.

Mathieson, I., I. Lazaridis, N. Rohland et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499-503.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon et al. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9:356-369.

McKee, J. K. 1984. A genetic model of dental reduction through the probable mutation effect. *Am. J. Phys. Anthropol.* 65:231-241.

Mead, S. V. 1928. Diseases of the mouth. St. Louis: The C.V. Mosby Company.

Minster, R. L., N. L. Hawley, C. T. Su et al. 2016. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat. Genet.* 48:1049-1054.

Nanda, R. S. 1954. Agenesis of the third molar in man. *Am. J. Orthod.* 40:698-706.

Olalde, I., M. E. Allentoft, F. Sanchez-Quinto et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507:225-228.

Organ, C., C. L. Nunn, Z. Machanda et al. 2011. Phylogenetic rate shifts in feeding time during the evolution of Homo. *Proc. Natl. Acad. Sci. U. S. A.* 108:14555-14559.

Perry, G. H., M. Foll, J. C. Grenier et al. 2014. Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc. Natl. Acad. Sci. U. S. A.* 111:E3596-3603.

Pickrell, J. K., G. Coop, J. Novembre et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19:826-837.

Pickrell, J. K., and D. Reich. 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* 30:377-389.

Pinhasi, R., V. Eshed, and N. von Cramon-Taubadel. 2015. Incongruity between affinity patterns based on mandibular and lower dental dimensions following the transition to agriculture in the Near East, Anatolia and Europe. *PLoS One* 10:e0117301.

Quek, S. L., C. K. Tay, K. H. Tay et al. 2003. Pattern of third molar impaction in a Singapore Chinese population: a retrospective radiographic survey. *Int. J. Oral Maxillofac. Surg.* 32:548-552.

Rozkovcova, E., M. Markova, and J. Dolejsi. 1999. Studies on agenesis of third molars amongst populations of different origin. *Sb. Lek.* 100:71-84.

Saglam, A. A., and M. S. Tuzum. 2003. Clinical and radiologic investigation of the incidence, complications, and suitable removal times for fully impacted teeth in the Turkish population. *Quintessence Int.* 34:53-59.

Sams, A. J., J. Hawks, and A. Keinan. 2015. The utility of ancient human DNA for improving allele age estimates, with implications for demographic models and tests of natural selection. *J. Hum. Evol.* 79:64-72.

Schlebusch, C. M., L. M. Gattepaille, K. Engstrom et al. 2015. Human adaptation to arsenic-rich environments. *Mol Biol Evol* 32:1544-1555.

Schweizer, R. M., J. Robinson, R. Harrigan et al. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Mol. Ecol.* 25:357-379.

Shaffer, J. R., E. Orlova, M. K. Lee et al. 2016. Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet.* 12:e1006149.

Song, F., S. O'Meara, P. Wilson et al. 2000. The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth. *Health Technol. Assess.* 4:1-55.

Stephan, W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol. Ecol.* 25:79-88.

Swindler, D. R. 2002. Primate Dentition: an Introduction to the Teeth of Non-human Primates. Cambridge: Cambridge Univ. Press.

Thompson, G. W., F. Popovich, and D. L. Anderson. 1974. Third molar agenesis in the Burlington Growth Centre in Toronto. *Community Dent. Oral Epidemiol.* 2:187-192.

Turchin, M. C., C. W. Chiang, C. D. Palmer et al. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44:1015-1019.

Voight, B. F., S. Kudaravalli, X. Wen et al. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.

von Cramon-Taubadel, N. 2011. Global human mandibular variation reflects differences in agricultural and hunter-gatherer subsistence strategies. *Proc. Natl. Acad. Sci. U. S. A.* 108:19546-19551.

Wellenreuther, M., and B. Hansson. 2016. Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends Genet.* 32:155-164.

Wilde, S., A. Timpson, K. Kirsanow et al. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U. S. A.* 111:4832-4837.

Wood, A. R., T. Esko, J. Yang et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46:1173-1186.

FIGURES

Figure 1

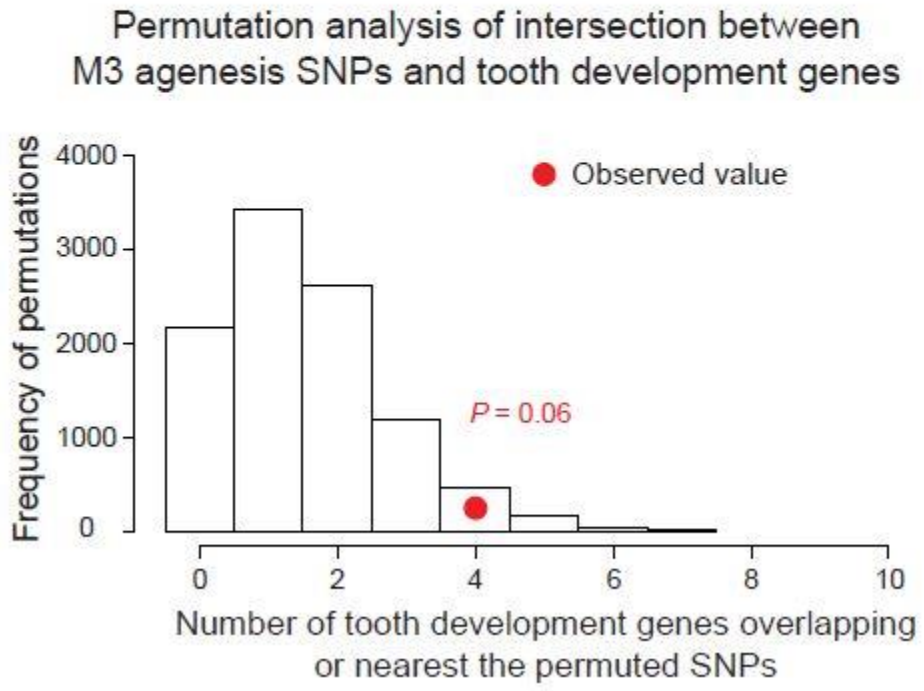


Figure 1. Permutation analysis of the intersection between M3 agenesis candidate SNPs and tooth development genes. To assess the functional plausibility of the 70 candidate M3 agenesis SNPs identified by Haga et al. (2013) we created 10,000 sets of 70 randomly-selected autosomal SNPs and identified the human genes overlapping or nearest each SNP (up to 70 unique genes). For each set of up to 70 unique genes, we randomly selected 51 genes (the number of unique genes overlapping or nearest the 70 candidate SNPs from the actual dataset) and computed the number of those genes that were also included in the database of 254 tooth development genes. The frequency distribution of those results for the 10,000 permutated datasets are shown. The observed number of tooth development genes among the 51 total genes overlapping or nearest the 70 M3 agenesis-associated SNPs from the original Haga et al. (2013) dataset and the proportion of permutated datasets in which this result or a greater number of tooth development genes were observed (as an empirical *P*-value) are also indicated.

Figure 2

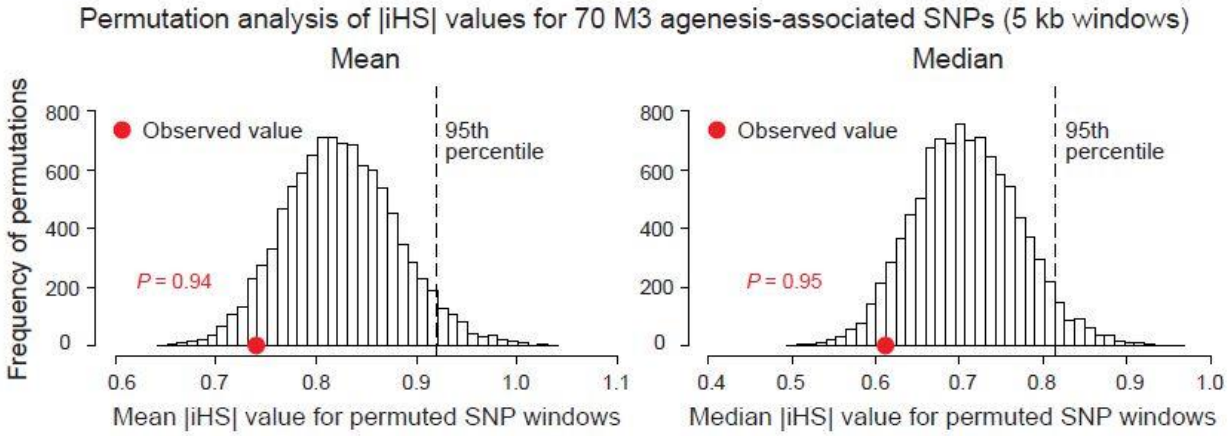
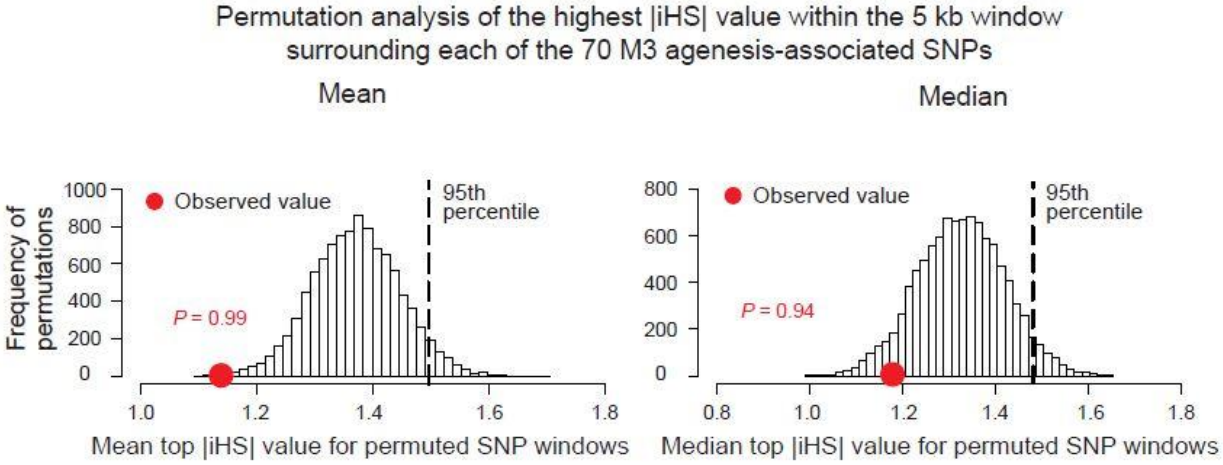


Figure 2. Permutation analysis of |iHS| values for 5 kb windows surrounding M3 agenesis candidate SNPs. We randomly generated 10,000 sets of 70 autosomal SNPs, the number of M3 agenesis-associated SNPs from Haga et al. (2013). For each permuted set, we computed the mean and median |iHS| values for all SNPs located in 5 kb windows surrounding the 70 SNPs. The frequency distribution of those results for the 10,000 permuted datasets are shown, along with the 95th percentile. The observed mean and median |iHS| values for the SNPs located in 5 kb windows surrounding the actual 70 SNPs from the Haga et al. (2013) dataset are shown. The indicated empirical P-values represent the probability that the observed values from the actual dataset are equal to or greater than those from a randomly selected set of 70 SNPs.

SUPPLEMENTARY MATERIAL

Supplemental Figure 1



Supplemental Figure 2

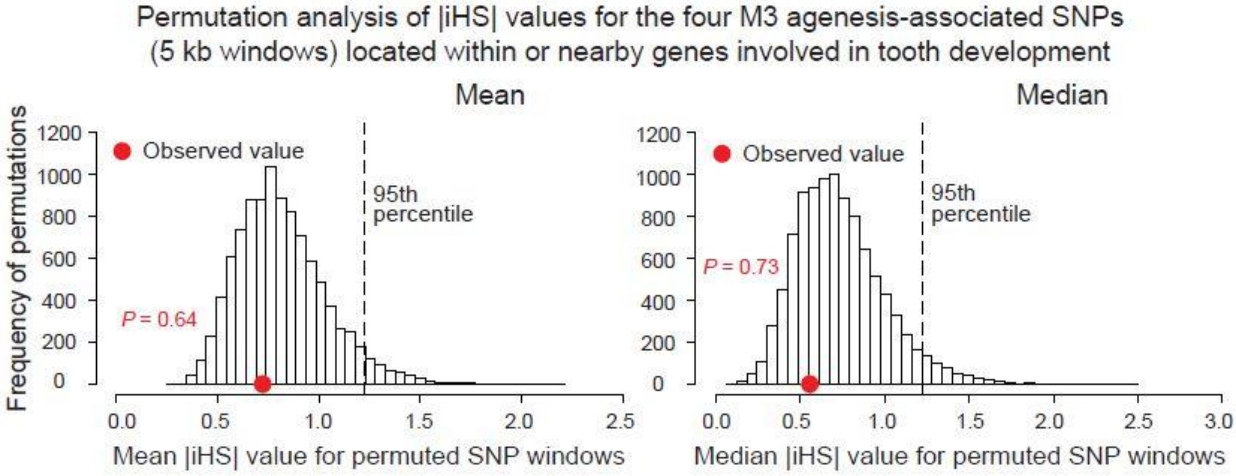


Table S1. M3 agenesis-associated SNPs from Haga et al. (2013)

Chromosome	SNP ID	hg18 position
chr1	rs11165783	97569544
chr2	rs13398007	21469336
chr2	rs1469622	137875875
chr2	rs1432205	137896264
chr2	rs11693055	182033294
chr2	rs7577873	208133793
chr2	rs4675658	208145220
chr2	rs2218249	235628051
chr2	rs6730767	235635006
chr2	rs13404290	237991777
chr2	rs7577630	237992613
chr3	rs935523	74710819
chr3	rs7614879	78132623
chr3	rs1127343	122128394
chr3	rs9839782	122287289
chr4	rs1454158	177290945
chr5	rs2398624	3714522
chr5	rs12110039	79368392
chr5	rs6873059	121939965
chr5	rs9327261	121990722
chr5	rs10045568	158949047
chr6	rs6459434	9823784
chr6	rs484754	109042459
chr6	rs3734652	109786980
chr6	rs2236084	109797304
chr6	rs1054607	109814411
chr8	rs7836791	70859624
chr9	rs12683929	127613385
chr10	rs1327249	6651563
chr10	rs12572318	43483490
chr10	rs10509178	64725655
chr10	rs938036	65658744
chr11	rs4553350	12759834
chr11	rs10834449	24739804
chr11	rs1638586	67174597
chr11	rs3802912	120738397
chr11	rs4936562	120741937

chr11	rs492761	134117015
chr11	rs12786906	134684914
chr11	rs906628	134738270
chr12	rs10845623	12898699
chr12	rs12369725	13914407
chr12	rs1075010	13918341
chr12	rs10879232	71377563
chr12	rs10774588	121652541
chr12	rs25644	121666646
chr12	rs1169717	121668684
chr14	rs10132731	59458350
chr14	rs4902843	71209025
chr14	rs2158530	71246995
chr14	rs10141804	71249230
chr14	rs4899377	71317212
chr14	rs12879624	71319317
chr15	rs17355029	47447605
chr15	rs289156	62702643
chr16	rs7204283	6836312
chr16	rs12444808	80053766
chr17	rs2035262	4499121
chr17	rs210837	32735169
chr17	rs1801200	37879588
chr18	rs523436	4295384
chr18	rs1511937	35994996
chr18	rs885033	36001458
chr18	rs11082076	36012486
chr18	rs17819948	71103611
chr19	rs12972158	46158473
chr20	rs6024403	54347434
chr20	rs6015829	59345638
chr21	rs7282042	15932610
chr21	rs928261	26456343

Table S2. List of genes with known involvement in tooth development

Gene symbol	Human gene NMID	Chromosome
INHBA	NM_002192	7
AGC1	NM_001135	15
AHR	NM_001621	7
ALPL	NM_000478	1
AMBN	NM_016519	4
OAZ1	NM_004152	19
APP	NM_201414	21
ARNT	NM_001668	1
AXIN1	NM_181050	16
AXIN2	NM_004655	17
BARX1	NM_021570	9
BAX	NM_001291428	19
BCL2	NM_000633	18
BEND3	NM_001080450	6
BMP2	NM_001200	20
BMP3	NM_001201	4
BMP4	NM_130851	14
BMP5	NM_021073	6
BMP6	NM_001718	6
BMP7	NM_001719	20
KAZALD1	NM_030929	10
BPAG1	NM_183380	6
BCAN	NM_021948	1
BTRC	NM_003939	10
NEU	NM_004448	17
CDH1	NM_004360	16
CALB1	NM_004929	8
CTNNB1	NM_001904	3
CD44	NM_001202556	11
CDKN1A	NM_000389	6
GPC2	NM_152742	7
PCDHA6	NM_018909	5
COL1A1	NM_000088	17
COL2A1	NM_001844	12
COL3A1	NM_000090	2
COL4A1	NM_001845	13
COL5A1	NM_000093	9

COL6A1	NM_001848	21
CX43	NM_000165	6
CRABP1	NM_004378	15
CSPG4	NM_001897	15
CCNA1	NM_003914	13
CCND1	NM_053056	11
CYP26C1	NM_183374	10
DAB1	NM_021080	1
DCN	NM_133503	12
DMP1	NM_001079911	4
DSPP	NM_014208	4
TWIST2	NM_001271893	2
DSG1	NM_001942	18
DLL1	NM_005618	6
DLX1	NM_178120	2
DLX2	NM_004405	2
DLX3	NM_005220	17
DLX4	NM_138281	17
DLX5	NM_005221	7
DLX6	NM_005222	7
ALCAM	NM_001627	3
SOSTDC1	NM_015464	7
EDAR	NM_022336	2
EGF	NM_001963	4
EGR1	NM_001964	5
ENAM	NM_031889	4
MMP20	NM_004771	11
EDN1	NM_001955	6
EPHA7	NM_004440	6
ERBB3	NM_001982	12
ERBB4	NM_005235	2
FADD	NM_003824	11
FAS	NM_000043	10
FGF1	NM_000800	5
FGF10	NM_004465	5
FGF2	NM_002006	4
FGF3	NM_005247	11
FGF4	NM_002007	11
FGF7	NM_002009	15

FGF8	NM_033165	10
FGF9	NM_002010	13
FGFR1	NM_023110	8
FGFR2	NM_000141	10
FGFR3	NM_000142	4
FGFR4	NM_002011	5
FMOD	NM_002023	1
FN1	NM_212482	2
FST	NM_013409	5
GAS1	NM_002048	9
GDNF	NM_000514	5
GFRA1	NM_005264	10
GFRA2	NM_001495	8
GLI1	NM_005269	12
GLI2	NM_005270	2
GLI3	NM_000168	7
GPC1	NM_002081	2
HAND1	NM_004821	5
HAND2	NM_021973	4
HES1	NM_005524	3
HES5	NM_001010926	1
HGF	NM_000601	7
HIP1	NM_005338	7
HSPG2	NM_001291860	1
IGF1	NM_001111283	12
IKBA	NM_020529	14
CHUK	NM_001278	10
IKBKB	NM_001556	8
ITGA6	NM_001079818	2
ITGAV	NM_002210	2
ITGB1	NM_133376	10
ITGB4	NM_000213	17
ITGB5	NM_002213	3
IRF6	NM_006147	1
IRX1	NM_024337	5
IRX2	NM_033267	5
IRX3	NM_024336	16
IRX4	NM_001278632	5
IRX5	NM_005853	16

IRX6	NM_024335	16
ISL1	NM_002202	5
JAG1	NM_000214	20
JAG2	NM_002226	14
KLK4	NM_004917	19
LAMA3	NM_198129	18
LAMB3	NM_000228	1
LAMC2	NM_005562	1
LAMA1	NM_005559	18
LAMA2	NM_000426	6
LAMA4	NM_001105206	6
LAMA5	NM_005560	20
LEF1	NM_016269	4
LFNG	NM_001040167	7
LUM	NM_002345	12
MME	NM_000902	3
MET	NM_001127500	7
MFNG	NM_002405	22
MDK	NM_001012334	11
MMP14	NM_004995	14
MMP2	NM_004530	16
MMP9	NM_004994	20
MSX1	NM_002448	4
MSX2	NM_002449	5
NTN1	NM_004822	17
NTN3	NM_006181	16
CSPG5	NM_006574	3
NRP1	NM_003873	10
NRP2	NM_201266	2
NGF	NM_002506	1
NGFR	NM_002507	17
LRRN3	NM_001099660	7
NOG	NM_005450	17
NOTCH1	NM_017617	9
NOTCH2	NM_024408	1
NOTCH3	NM_000435	19
NRG1	NM_004495	8
NTF3	NM_001102654	12
NTF4	NM_006179	19

NTRK1	NM_001012331	1
NTRK3	NM_001012338	15
CREB3L1	NM_052854	11
OCLN	NM_002538	5
ODC1	NM_002539	2
OSR2	NM_001142462	8
BGLAP	NM_199173	1
SPP1	NM_001040058	4
CDH3	NM_001793	16
PCSK6	NM_002570	15
PTCH1	NM_001083602	9
PTCH2	NM_003738	1
PAX9	NM_006194	14
PCDHGA	NM_018914	5
PCNA	NM_002592	20
PTPRZ1	NM_002851	7
FUS	NM_004960	16
PITX2	NM_153427	4
JUP	NM_002230	17
PRRX1	NM_006902	1
PRRX2	NM_016307	9
PTHLH	NM_198965	12
PTH1R	NM_000316	3
RAB23	NM_016277	6
NR1B1	NM_000964	17
NR1B2	NM_001290276	3
NR1B3	NM_000966	12
RBP1	NM_002899	3
RELN	NM_005045	7
RET	NM_020975	10
RFNG	NM_002917	17
ROBO1	NM_002941	3
ROBO2	NM_002942	3
ROR1	NM_005012	1
ROR2	NM_004560	9
RUNX1	NM_001754	21
RUNX2	NM_001024630	6
RUNX3	NM_001031680	1
RXRA	NM_002957	9

RXRB	NM_001291989	6
RXRG	NM_006917	1
SC1	NM_007109	6
SPP1	NM_001040058	4
SEMA3A	NM_006080	7
SEMA3B	NM_004636	3
SEMA3C	NM_006379	7
SEMA3F	NM_004186	3
SLIT1	NM_003061	10
SLIT2	NM_004787	4
SLIT3	NM_001271946	5
SLITRK6	NM_032229	13
SMO	NM_005631	7
SNAI1	NM_005985	20
SHH	NM_000193	7
SOX9	NM_000346	17
SP6	NM_001258248	17
SPOCK1	NM_004598	5
SPRY1	NM_001258038	4
SPRY2	NM_005842	13
SPRY4	NM_030964	5
SDC1	NM_001006946	2
SDC2	NM_002998	8
SDC3	NM_014654	1
SDC4	NM_002999	20
NKNA	NM_013996	7
TNC	NM_002160	9
TGFB1	NM_000660	19
TIMP2	NM_003255	17
TIMP3	NM_000362	22
TJP1	NM_003257	15
TJP2	NM_004817	9
TJP3	NM_001267560	19
TNFRSF19	NM_018647	13
FASLG	NM_000639	1
TRAF1	NM_005658	9
TRAF2	NM_021138	9
TRAF3	NM_145725	14
TRAF4	NM_004295	17

TRAF6	NM_145803	11
TRKB	NM_006180	9
KET	NM_003722	3
TUFT1	NM_020127	1
VCAN	NM_004385	5
WNT10A	NM_025216	2
WNT10B	NM_003394	12
WNT3	NM_030753	17
WNT4	NM_030761	1
WNT5A	NM_003392	3
WNT6	NM_006522	2
WNT7B	NM_058238	22
WT1	NM_000378	11
MYB	NM_001130173	6
CLU	NM_001831	8
FBLN1	NM_006487	22
FBLN2	NM_001004019	3
HS3ST1	NM_005114	4
HSPB1	NM_001540	7
ITGA4	NM_000885	2
PVRL1	NM_002855	11
TBX1	NM_005992	22