GWAS Identifies New Loci for Painful Temporomandibular Disorder: Hispanic Community Health Study/Study of Latinos

A.E. Sanders, D. Jain, T. Sofer, K.F. Kerr, C.C. Laurie, J.R. Shaffer, M.L. Marazita, L.M. Kaste, G.D. Slade, R.B. Fillingim, R. Ohrbach, W. Maixner, T. Kocher, O. Bernhardt, A. Teumer, C. Schwahn, K. Sipilä, R. Lähdesmäki, M. Männikkö, P. Pesonen, M. Järvelin, C.M. Rizzatti-Barbosa, C.B. Meloto, M. Ribeiro-Dasilva, L. Diatchenko, P. Serrano, and S.B. Smith

Appendix

Description of Replication Cohorts

Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA)

The OPPERA case-control cohort included 3030 participants, of which 999 were TMD cases and 2031 TMD-free controls. Examiners determined classification of TMD according to the Research Diagnostic Criteria for Temporomandibular Disorder(Dworkin and LeResche, 1992). As reported previously (Ohrbach *et al.*, 2011), cases met all 3 of the following criteria: a) pain reported with sufficient frequency in the cheeks, jaw muscles, temples, or jaw joints during the preceding 6 months; b) pain reported in the examiner-defined orofacial region for at least 5 days out of the prior 30 days; and c) pain reported in at least 3 masticatory muscles or at least 1 temporomandibular joint in response to palpation of the orofacial muscles or maneuver of the jaw. Examination of the orofacial region included temporalis, preauricular, masseter, posterior mandibular, and submandibular areas bilaterally. TMD was defined as facial pain for at least the preceding 6 months including at least 5 days during the preceding month, and sufficient positive findings on examination.

Study of Health in Pomerania (SHIP)

The German cohort was derived from a large cross-sectional survey of a representative sample of Pomerania, Germany, the Study of Health in Pomerania (SHIP) study (Bernhardt *et al.*, 2004). Participants were aged 20-81 years, and included 51% females. Participants reported symptoms by questionnaire regarding pain in the temporomandibular joint and facial muscles; presence and frequency of pain were assessed. During a clinical exam, the examiner inquired about pain or discomfort upon palpation of masticatory tissues, including temporomandibuar joints (dorsocranial and lateral) at 2 kg/cm2, and masseter, temporalis, and medial pterygoid at 1 kg/cm2. Pain or discomfort during jaw movement, range of motion, and joint sounds were also assessed. TMD was defined as pain or discomfort during examination procedures in at least one muscle or TM joint. Participants with both phenotype and genotype data totaled 3651, including 607 cases (17%) and 3044 controls.

Northern Finland Birth Cohort (NFBC)

The Northern Finland Birth Cohort is a cohort study of all births in 1966 in the Oulu and Lapland provinces of northern Finland. An assessment for TMD was performed at the 46-year follow-up time-point. Participants (52% female) reported symptoms by responding to a questionnaire with the following questions: a) "Do you experience temple, temporomandibular joint, face, or jaw pain once a week or more often?" b) "Do you experience pain once a week or more often while opening your mouth wide" A clinical exam determined the presence of examiner-evoked pain in three

or more temporomandibular muscles and/or joints. Palpation sites included the temporalis (1 kg force), masseter (1 kg force), lateral temporomandibular joint (0.5 kg force), and temporomandibular joint around the pole (1 kg force). For each palpation, participants answered yes or no according to whether or not they experienced pain. To be classified with TMD, study participants were required to record the presence of both: a) reported pain according to the question a and/or b, b) palpation pain in three or more masticatory muscles and/or temporomandibular joints. Of 1940 participants who completed the TMD questionnaire and examination, 161 (11%) had TMD defined as a positive response to one or both symptom questions, and positive examination findings from palpation and/or jaw movement.

Brazilian cohort

The Brazilian participants were enrolled in a community based case-control study in Piracicaba, São Paulo, Brazil, and included females between the age of 18-44 years. Pain history was determined by asking participants the question: "Have you had pain in your head, face, jaw, or in front of the ears in the last 30 days?" The examiner manually palpated lateral and posterior temporomandibular joints (0.45 kg) and asked participants to report yes or no responses to the presence of pain. The examiner measured maximum unassisted opening, maximum assisted opening, and right and left excursion and protrusion of the jaw, also prompting the subject for pain response. Other symptoms assessed included range of motion and joint sounds. TMD case classification was determined by pain lasting greater than 3 months and pain on examination in at least one TM joint. Of 636 participants, 144 (22%) were classified as TMD cases. The remaining 492 controls were negative to both self-report and examination.

Functional annotation of replicated discovery loci

To identify potentially causal variants, we conducted functional annotation of four loci that were genome-wide significant in the discovery cohort. At each locus we determined if the lead SNP and its proximal correlated variants (r2 ≥ 0.8, calculated in the HCHS/SOL cohort) were positioned within putative regulatory regions. The regulatory regions were identified based on the enrichment of various histone modification ChIP-Seq (chromatin immunoprecipitation followed by sequencing) signals in skeletal muscle and neuronal cells from the Roadmap Epigenomics project (Kundaje *et al.*, 2015) or in periodontal fibroblasts from the ENCODE project (An integrated encyclopedia of DNA elements in the human genome, 2012; A user's guide to the encyclopedia of DNA elements (ENCODE), 2011). A genomic element enriched with the histone H3K4me1 ChIP-Seq signal was categorized as an enhancer while a genomic element enriched with the histone H3K4me3 ChIP-Seq signal was categorized as a promoter. Regulatory elements with H3k27ac mark and H3k9ac mark were annotated as active in the respective tissues of their enrichment. Enrichment was determined based on peak calls as well as visualization of ChIP-Seq signal tracks in the genome browser.

SNPs that were positioned within a putative promoter or enhancer and overlapped with a DNaseI hypersensitive site were prioritized as plausible functional variants. Regulatory elements often are bound by transcription factors, and hence we also report overlap with transcription factor ChIP-Seq peaks observed in the ENCODE project (An integrated encyclopedia of DNA elements in the human genome, 2012; A user's guide to the encyclopedia of DNA elements (ENCODE), 2011) to further support the functional role of putative regulatory elements. ChIP-Seq signal and peak call datasets from the ENCODE and Roadmap Epigenomics projects were accessed through the ENCODE analysis and Roadmap Epigenomics Hubs via the UCSC genome browser(Koscielny *et al.*, 2014; Raney *et al.*, 2014). To hypothesize likely modes of action by which causal variants influence phenotypes, we reported eQTL targets of prioritized variants

using HaploReg (Ward and Kellis, 2012) and gene expression status using data from the Genotype-Tissue Expression (GTEx) consortium (The Genotype-Tissue Expression (GTEx) project, 2013).

Mouse knockout results were reported from the International Mouse Phenotyping Consortium Web Portal (Koscielny *et al.*, 2014). All the datasets used for functional annotation were mapped to the Human GRCh37/hg19 assembly.

Dystrophin-glycoprotein pathway

Findings for the functional annotation of the implication regions are report in the online appendix. Functional annotation of the implicated regions sought to identify potential mechanisms that may influence the development or persistence of TMD pain. The lead SNP rs73460075 lies in the intronic region of the dystrophin-encoding *DMD* gene which is well known for its role in muscle related diseases. Mutations in the gene are causally implicated in painful muscle-related disorders, notably Duchenne dystrophy and Becker muscular dystrophy, two forms of inherited, progressive muscle wasting (Aartsma-Rus *et al.*, 2016; Emery, 2002). Functional annotation using epigenetic data from skeletal muscle cells, neuronal progenitor cells, and neuron cultured cells did not suggest that this lead SNP was located in a non-coding regulatory element, nor was it highly associated ($r^2 > 0.8$) with other assayed or imputed SNPs that have evidence for regulatory roles.

 The lead SNP rs4794106 is near the *SGCA* gene (the protein of which interacts with the protein from *DMD*), but located in the intronic region of the non-coding RNA gene RP11-893F2.13. We explored if rs4794106 overlapped with DNAse hypersensitive sites, an indicator of unpacked and accessible chromatin and a feature of active regulatory sequences. rs4794106 was found positioned within DNAse hypersensitive sites in normal human skeletal muscle myoblasts cells (HSMM) and neuronal progenitor cells suggesting that it lies within an active regulatory genomic element in these cell types. Further annotation with epigenetic marks indicate that rs4794106 is enriched with histone modification ChIP-Seq (chromatin immunoprecipitation followed by sequencing) signal typically associated with a promoter and its proximal region in skeletal muscles and neuron cultured cells (H3K4me3 and H3K9ac in skeletal muscle cells, H3k4me3 in neuron cultured cells). The putative promoter of RP11-893F2.13 also overlaps with ChIP-Seq peaks of RNA polymerase II, and transcription factors SMC3, P300, and CTCF in various ENCODE cell lines, an observation reinforcing promoter regulatory activity associated with this genomic element. Taken together the annotation suggests that rs4794106 is located in a non-coding promoter proximal regulatory element which may play a role in transcriptional regulation of non-coding gene RP11-893F2.13.

While the function of non-coding gene RP11-893F2.13 is not known, a plausible hypothesis is that RP11-893F2.13 or its transcription may regulate the expression of its upstream, anti-sense coding gene *SGCA* (Lepoivre *et al.*, 2013; Pandey *et al.*, 2008; Pelechano and Steinmetz, 2013).

SGCA is also reported as an eQTL of rs4794106 in lymphoblastoid cells, expressed at high levels in skeletal muscle cells and encodes the protein sarcoglycan, a component of the dystrophin-glycoprotein complex (DGC). This is particularly interesting in the context of the observation of a genome-wide significant hit in the dystrophin gene locus, *DMD*. Besides *SGCA*, *SAMD14* is another gene reported as an eQTL of rs4794106 in esophagus and whole blood.

Four SNPs, rs847688, rs112842682, rs847687 and rs202191526 are in high LD ($r^2 >= 0.8$ in HCHS/SOL samples) with the lead SNP related to SGCA in a region upstream of the *PPP1R9B* gene and we are predicting them to be causal SNPs at

this locus based on functional annotation. All four SNPs are positioned within a single genomic element which is DNAse hypersensitive in human gingival fibroblasts, normal human periodontal ligament fibroblasts, normal human skeletal muscle myoblasts, human skeletal muscle cells (SKMC) and neuronal progenitor cells. Further annotation with epigenetic marks indicate that the genomic element harboring these four SNPs is enriched for histone modification ChIP-Seq signals typically associated with promoter regulatory elements in skeletal muscles and neuron cultured cells (H3K4me3, H3k27ac and H3K9ac in skeletal muscle cells, H3k4me3 in neuron cultured cells and H3k27ac in neuronal progenitor cells). The putative promoter of *PPP1R9B* also overlaps with ChIP-Seq peaks of RNA polymerase II, and transcription factors SMC3, P300, ZNF143, CTCF in various ENCODE cell lines, an observation reinforcing promoter regulatory activity associated with this genomic element. Taken together the annotation suggests that rs847688, rs112842682, rs847687 and rs202191526 are located in a non-coding promoter regulatory element which may play a role in transcriptional regulation of *PPP1R9B*. Interestingly, homozygous knockout mice of Protein Phosphatase 1 Regulatory Subunit 9B (*PPP1R9B*) gene are reported to show significant deviations for various metrics of sleep behavior compared to normal mice [https://www.mousephenotype.org/ ,www.komp.org]. This association of abnormal sleep behavior associated with homozygous gene knock outs of *PPP1R9B* is particularly intriguing given that poor sleep quality is a strong predictor of TMD development (Sanders *et al.*, 2016). All four SNPs have been reported as eQTLs for downstream gene *SGCA* in lymphoblastoid cells and upstream gene SMAD14 in esophagus.

One suggestive locus, rs73271865, is upstream from *SP4*, which is a transcription factor for the capsaicin receptor, TRPV1. This is noteworthy as TRPV1 gene expression is implicated in nociception (Caterina *et al.*, 1997) and is expressed in the synovial tissue of the temporomandibular joint in humans with painful TMD. (Sato *et al.*, 2005)

The lead SNP rs1531554 lies in the intronic region of the *BAHCC1* gene. The lead SNP rs60249166 lies in the intergenic region between non-coding RNA gene RP11-326L9.1 and protein coding gene *RXFP2*. Neither of these lead SNPs has any proxy SNPs that have $r^2 \geq 0.8$ in the HCHS/SOL population. Functional annotation using epigenetic data indicated that there is no compelling evidence for regulatory function of these SNPs.
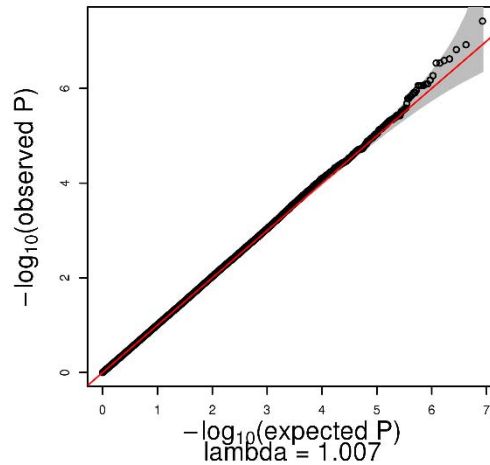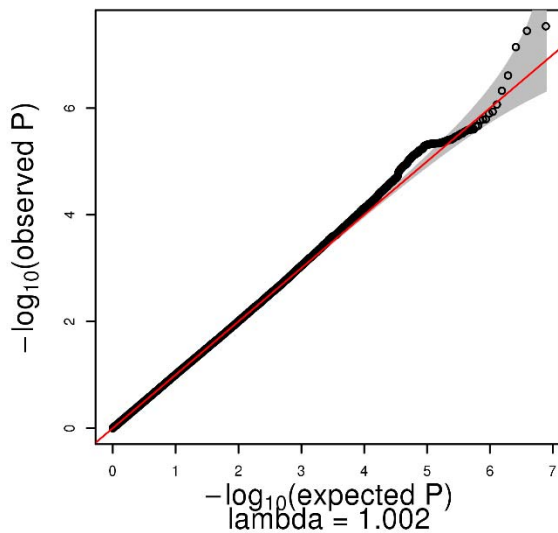
Appendix Figure 1. The QQ plot for performing association tests using generalized linear mixed model analyses use the p-values derived from the score test statistic, assuming a chi^2 distribution with one degree of freedom.
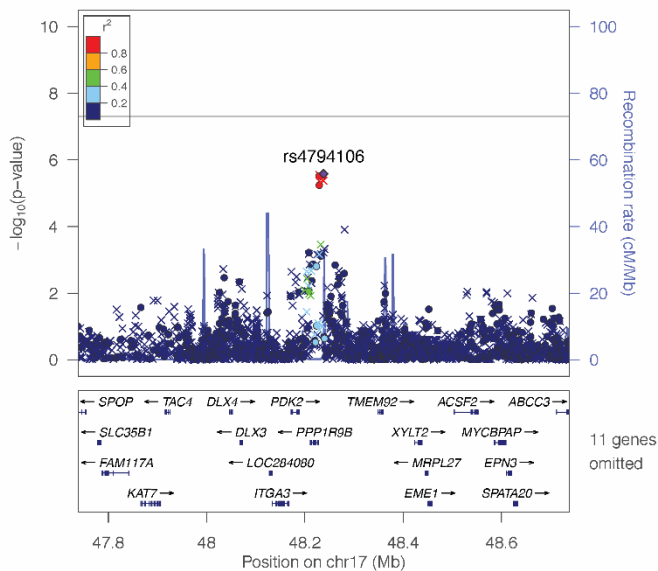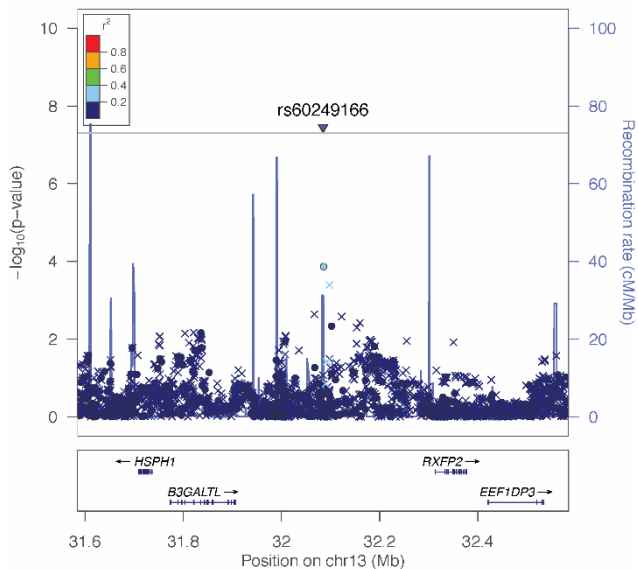

Appendix Figure 2A. Regional association plot for locus with suggestive evidence of association with TMD. Shows rs4794106 near the *SGCA* gene (p=2.6 x10^{-6}).
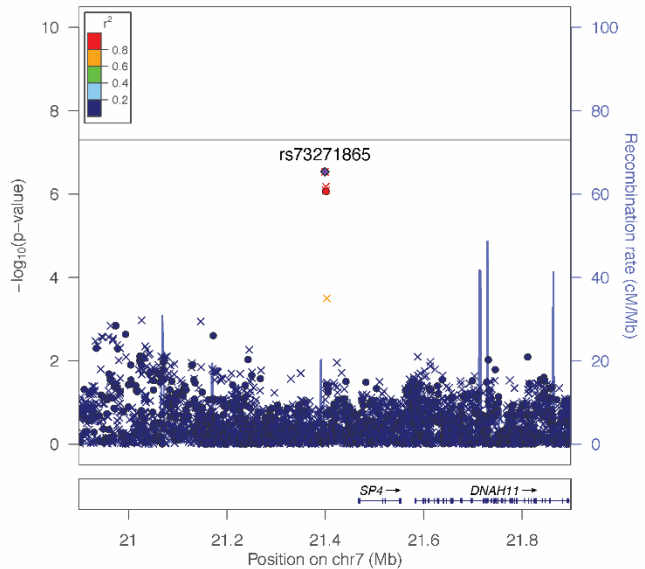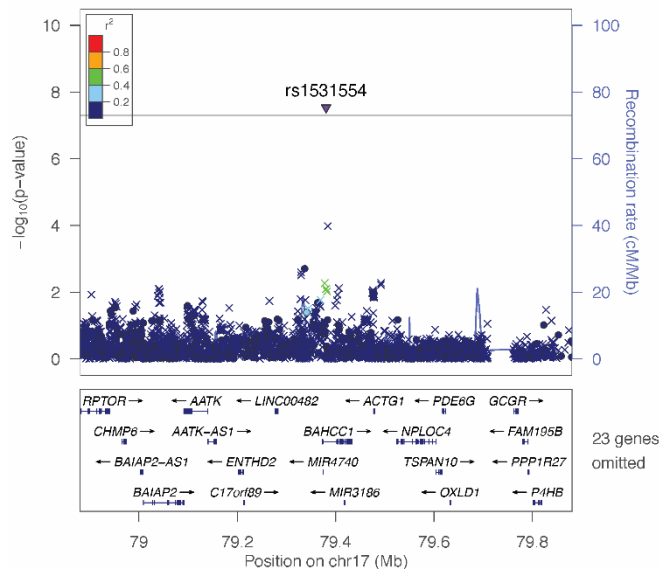
Appendix Figure 2B. Regional association plot locus with suggestive evidence of association with TMD. Shows rs73271865 (p=2.9 x10^{-7}), which lies upstream of the *SP4* gene.

Appendix Figure 3. The QQ plot for performing association tests using generalized linear mixed model analyses use the p-values derived from the score test statistic, assuming a chi^2 distribution with one degree of freedom.

Appendix Figure 4A. Regional association plot for genome-wide significant association identified in the HCHS/SOL discovery genome-wide association analysis restricted to females. Shows a region on chromosome 13 (lead SNP rs60249166, p=3.57 x10^{-8}).

Appendix Figure 4B. Regional association plot for genome-wide significant association identified in the HCHS/SOL discovery genome-wide association analysis restricted to females. Shows a region on chromosome 17 (lead SNP rs1531554, p=2.92 x 10^{-8}).

Appendix Table 1. Associations between characteristics of entire HCSH/SOL discovery study population and painful temporomandibular disorder (TMD), (n=15,344)

Study participant characteristics	Unweighted N,[a] weighted col %	TMD prevalence (95% confidence limits (CL))	P value	Site-adjusted odds ratios for TMD (95% CL)	P value [b]
Total	15,344 (100.0)	5.1 (4.7, 5.7)			
Hispanic/Latino background					
Dominican	1,360 (9.8)	4.4 (3.3, 5.9)	0.026	0.9 (0.6, 1.4)	0.630
Central American	1,637 (7.5)	5.5 (4.5, 6.7)		1.1 (0.8, 1.5)	0.438
Cuban	2,098 (19.3)	4.5 (3.3, 6.0)		0.9 (0.6, 1.3)	0.516
Mexican	6,291 (38.8)	5.0 (4.2, 5.9)		Referent	
Puerto Rican	2,449 (15.5)	7.0 (5.8, 8.4)		1.5 (1.0, 2.0)	0.026
South American	999 (4.9)	3.8 (2.7, 5.2)		0.7 (0.5, 1.1)	0.156
Mixed/other	472 (4.2)	4.9 (3.2, 7.4)		1.0 (0.6, 1.6)	0.927
Sex					
Female	9,178 (52.0)	6.7 (6.0, 7.4)	<0.001	2.0 (1.6, 2.5)	<0.001
Male	6,166 (48.0)	3.5 (2.9, 4.1)		Referent	
Age group, years					
18–34	3,640 (40.0)	4.2 (3.5, 5.1)	<0.001	Referent	
35–44	2,839 (21.7)	4.8 (4.0, 5.8)		1.1 (0.9, 1.5)	0.352
45–54	4,670 (19.1)	6.9 (5.9, 8.0)		1.7 (1.3, 2.2)	<0.001
55-64	3,100 (12.0)	6.3 (5.3, 7.5)		1.5 (1.2, 2.0)	0.002
≥65	1,095 (7.3)	4.6 (3.2, 6.6)		1.1 (0.7, 1.7)	0.642
Nativity					
Foreign born	12,607 (76.7)	5.4 (4.8, 6.0)	0.100	1.2 (1.0, 1.6)	0.090
U.S. born	2,712 (23.3)	4.4 (3.5, 5.4)		Referent	
Education					
No high school or equivalent	5,597 (31.7)	6.0 (5.2, 6.8)	0.026	1.2 (1.0, 1.5)	0.040
At most high school or equivalent	3,919 (28.7)	4.5 (3.8, 5.5)		0.9 (0.7, 1.2)	0.586
Greater than high school or equivalent	5,513 (39.6)	4.8 (4.1, 5.6)		Referent	
Cigarette use					
Lifetime non-smoker	9,449 (62.3)	4.9 (4.3, 5.5)	0.043	Referent	
Former smoker	3,002 (17.0)	4.8 (4.0, 5.8)		1.0 (0.8, 1.2)	0.942
Current smoker	2,858 (20.8)	6.2 (5.2, 7.4)		1.3 (1.0, 1.6)	0.026
Body mass index, kg/m2					
<25 (underweight/normal)	3,105 (23.3)	5.2 (4.3, 6.3)	0.125	Referent	
≥25 <30 (overweight)	5,720 (37.1)	4.5 (3.9, 5.3)		0.9 (0.7, 1.1)	0.263
≥30 (obese)	6,471 (39.6)	5.6 (4.9, 6.4)		1.1 (0.9, 1.4)	0.547
Health insurance					
Not insured	7,578 (50.2)	5.0 (4.4, 5.6)	0.517	1.0 (0.8, 1.2)	0.648
Insured	7,627 (49.8)	5.3 (4.6, 6.1)		Referent	
Depressive symptoms, CESD-10 [c]					
<10	10,785 (73.3)	3.7 (3.2, 4.2)	<0.001	Referent	<0.001
≥10 (at risk of depression)	4,407 (26.7)	9.1 (8.0, 10.3)		2.6 (2.2, 3.2)	
Trait anxiety, STAI [d]					
<20	11,322 (76.2)	4.1 (3.6, 4.7)	<0.001	Referent	<0.001
≥20 (most anxious quartile)	3,861 (23.8)	8.4 (7.3, 9.6)		2.1 (1.7, 2.6)	
Mental Health Composite Score, SF-12 [e]					
≤40 (poor mental health)	3,534 (21.8)	8.8 (7.6, 10.1)	<0.001	2.3 (1.9, 2.8)	<0.001
>40	11,660 (78.2)	4.1 (3.6, 4.6)		Referent	
Physical Health Composite Score, SF-12 [f]					
≤45 (poorest health quartile)	3,534 (25.0)	8.8 (7.7, 10.0)	<0.001	2.4 (2.0, 2.9)	<0.001
>45	11,660 (75.0)	3.9 (3.4, 4.4)		Referent	

(a) Responses unreported for: background (n=38); nativity (n=25); education (n=315); cigarette use (n=35); BMI (n=48); health insurance (n=139); depressive symptoms (n=152); trait anxiety (n=161); mental health score (n=150); physical health score (n=150).

(b) P-values are from a logistic regression model

(c) CESD-10, Center for Epidemiologic Studies Depression scale;

(d) Spielberger Trait Anxiety Inventory (STAI-Trait), Form Y-2

(e) Mental component summary of the Short-Form Health Survey; (f) Physical component summary of the Short-Form Health Survey

(e, f) For both components of the SF-12, scores are normalized, such that they are representative of the general U.S. population with a mean of 50, a standard deviation of 10, and a range of 0 to 100.

All estimates, except unweighted N, account for probability sampling

Appendix Table 2. Results for the individual studies used in replication analysis

Included	SNP	Chromosome	Base pair	Coded allele	OPPERA (a)		SHIP (b)		NFBC (c)		Brazil (d)	
					OR	p value	OR	p value	OR	p value	OR	p value
All	rs73460075	23	32283492	G	1.11	0.550*	NA		NA		NA	
All	rs4794106	17	48238294	T	1.01	0.414	1.02	0.376	0.80	0.064*	1.38	0.016
All	rs73271865	7	21399327	C	0.95	0.362	NA		NA		NA	
Females	rs60249166	13	32084901	C	0.88	0.104	0.85	0.106	NA		NA	
Females	rs1531554	17	79380547	T	0.84	0.0097	0.82	0.026	NA		NA	

* 2-tailed p-value, as the direction of effect differs from that of the discovery cohort

(a) Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA)
(b) Study of Health in Pomerania (SHIP)
(c) Northern Finland Birth Cohort 1966 (NFBC)
(d) Case control study of TMD conducted in Piracicaba, São Paulo, Brazil

References

Aartsma-Rus A, Ginjaar IB, Bushby K (2016). The importance of genetic diagnosis for Duchenne muscular dystrophy. *J Med Genet* 53(3)*:*145-151.

Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD, Julius D (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389(6653)*:*816-824.

Emery AE (2002). The muscular dystrophies. *Lancet* 359(9307)*:*687-695.

The Genotype-Tissue Expression (GTEx) project. (2013). *Nat Genet* 45(6)*:*580-585.

An integrated encyclopedia of DNA elements in the human genome. (2012). *Nature* 489(7414)*:*57-74.

Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J *et al.* (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res* 42(Database issue)*:*D802-809.

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A *et al.* (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539)*:*317-330.

Lepoivre C, Belhocine M, Bergon A, Griffon A, Yammine M, Vanhille L *et al.* (2013). Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* 14(914.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J *et al.* (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32(2)*:*232-246.

Pelechano V, Steinmetz LM (2013). Gene regulation by antisense transcription. *Nat Rev Genet* 14(12)*:*880-893.

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T *et al.* (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30(7)*:*1003-1005.

Sanders AE, Akinkugbe AA, Bair E, Fillingim RB, Greenspan JD, Ohrbach R *et al.* (2016). Subjective Sleep Quality Deteriorates Before Development of Painful Temporomandibular Disorder. *The journal of pain : official journal of the American Pain Society* 17(6)*:*669-677.

Sato J, Segami N, Yoshitake Y, Kaneyama K, Abe A, Yoshimura H *et al.* (2005). Expression of capsaicin receptor TRPV-1 in synovial tissues of patients with symptomatic internal derangement of the temporomandibular joint and joint pain. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 100(6)*:*674-681.

A user's guide to the encyclopedia of DNA elements (ENCODE). (2011). *PLoS Biol* 9(4)*:*e1001046.

Ward LD, Kellis M (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 40(Database issue)*:*D930-934.