Appendix

BACH1 binding links the genetic risk for severe periodontitis with *ST8SIA1*

Authors: Avneesh Chopra[1], Ricarda Mueller[1], January Weiner[2], Jennifer Rosowski[3], Henrik Dommisch[1], Elisabeth Grohmann[4], Arne S. Schaefer[1].

Affiliations:

[1]Department of Periodontology, Oral Medicine and Oral Surgery, Institute for Dental and Craniofacial Sciences, Charité – University Medicine Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, 10117, Germany

[2]Core Unit Bioinformatics, Berlin Institute of Health, Berlin, 10178, Germany

[3]Department of Medical Biotechnology, Institute of Biotechnology, Technische Universität Berlin, Berlin, 13355, Germany

[4]Department of Microbiology, Faculty of Life Sciences and Technology, Beuth Hochschule für Technik Berlin, Berlin, 13353, Germany

Correspondence:

Arne Schaefer, Professor Dr. rer. nat.

Genetics of Oral Inflammatory Diseases group

Charité - University Medicine Berlin

Tel: ++49 (0)30 450 562-343

Email: arne.schaefer@charite.de

Materials and methods

Cell culture

Primary gingival fibroblasts and epithelial cells show very high lethality after transfection with CRISPRa plasmids and are not ideal for experiments using the CRISPR system. Therefore, for CRISPRa experiments we used HeLa cells. HeLa cells were cultured in a growth medium of Earle's MEM, containing 10% fetal bovine serum, 2mM L-Glutamine, 1% non-essential Amino acids and 1% penicillin-streptomycin. HeLa cells were seeded at 80,000 cells per well in 6-well plates (TPP Techno Plastic Products) and left overnight to reach around 50-60% confluence. Cells were transfected using jetPEI transfection reagent (Polyplus transfection) according to the manufacturer's instructions.

Immortalized human gingival fibroblasts (ihGFs, purchased from Applied Biological Materials Inc. [ABM]) were used for the reporter gene plasmids. ihGFs were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. ihGFs were seeded at 180,000 cells per 6-well before transfection with Lipofectamine 2000 (Thermo Fisher Scientific).

CRISPR-mediated gene activation

We tiled two individual sgRNAs at each of the two SNP-associated putative regulatory regions. To induce the expression of *ST8SIA1* directly, we tiled two sgRNAs at the promoter of *ST8SIA1*. SgRNAs were designed with the online tool CRISPR-ERA (Liu et al. 2015) according to the protocol described in (Ran et al. 2013). As controls, we used a scrambled sgRNA with no genomic target as well as a sgRNA that located to the microRNA hsa-miR-374b-5p on chromosome X, which did not affect *ST8SIA1* expression. The sgRNAs were synthesized **(**Metabion International AG) and cloned into the *Bbs*I site of sgRNA(MS2) cloning backbone vector (Plasmid #61424) (**Appendix Table 5**). The CRIPRa system was transfected into HeLa cells using jetPEI transfection reagent. SgRNAs that targeted the regulatory regions and the promoter of *ST8SIA1* were pooled and transfected into HeLa cells in biological triplicates. Separately, the control sgRNAs were pooled and transfected into HeLa cells in biological triplicates. For CRISPRa, each 6-well was transfected with 1µg sgRNAs(MS2) (consisting of four sgRNAs with 250 ng each), 1µg dCAS9-VP64_GFP [Plasmid #61422] and 1µg MS2-P65-HSF1_GFP [Plasmid #61423] and incubated for 44 hours. All plasmids were obtained from Addgene, gifted by Feng Zhang.

RNA-Seq

Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA integrity numbers (RIN) were >8 and were measured and calculated on the 2100 Bioanalyzer (Agilent) using the RIN software algorithm. The 500-1000 ng of total RNA of the 6 transfected independent HeLa cell cultures was sequenced with 16 million reads (75 bp single end) on a NextSeq 500 using the NextSeq 500/550 High Output Kit v2.5 (75 Cycles). Reads were aligned to the Genome Reference Consortium Human Build 38 patch release 7 (GRCh38.p7) genome using the STAR aligner v. 2.7.5a (Dobin et al. 2013). Quality control (QC) of the reads was inspected using the multiqc reporting tool (Ewels et al. 2016) summarizing on a number of approaches, including fastqc (available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc), dupradar (Sayols et al. 2016), qualimap (Garcia-Alcalde et al. 2012) and RNA-SeqC (DeLuca et al. 2012). Raw counts were extracted using the STAR program. For differential gene expression, the R package DESeq2 (Love et al. 2014) version 1.26 was used. The only contrast fitted was the comparison between the *ST8SIA1* induction and scramble controls. Gene set enrichment was performed using the CERNO test from the tmod package (Zyla et al. 2019) version 0.46.2 using the gene expression profiling based gene set included in the package as well as the MSigDB (Liberzon et al. 2015). For hypergeometric test and the Gene Ontology gene sets, the goseq package, version 1.38 (Young et al. 2010) was used. The p values of the differently expressed genes were corrected for multiply testing using Benjamini-Hochberg correction. The corrected p values are given as q values (false discovery rate, FDR).

Electrophoretic Mobility Shift Assay

DNA-protein binding was determined by using the Gelshift Chemiluminescent EMSA Kit (Active Motif). The protein extract was prepared from ihGFs using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Pierce Biotechnology). The double-stranded oligonucleotides corresponding to both alleles of rs2012722 flanked by 21 bp in both cold and 3′-biotinylated form were obtained by annealing with their respective complementary primers. For binding reaction, 20 fmol biotin-labeled, double-stranded oligonucleotides were incubated with nuclear extract (3-10 µg) in 1x binding buffer and 1 µg/µL Poly d(I-C) for 20 min at room temperature. For competition assay, unlabeled double-stranded oligonucleotides (200-fold molar excess) were added to the binding reaction. For supershift EMSA, a monoclonal antibody (2 µL of 10 µg/50 µL (Santa Cruz Biotechnology Inc.) was added to the binding reaction (without the Poly d(I-C)). The DNA–protein complexes were electrophoresed in a 5% non-denaturing

polyacrylamide gel in 0.5x TBE buffer at 100V for 2 hours and visualized by chemiluminescence detection after electric transfer of the products on a nylon membrane. The results were displayed on x-ray films.

Reporter gene assay

We used the firefly luciferase vector pGL4.24 (Promega), which was barcoded with unique DNA sequences to serve as templates for quantitative real-time PCR (qRT-PCR) primers, enabling parallel detection of individual reporter genes. The reporter gene plasmids were generated in two cloning steps. First, 80 bp non-human unique DNA sequences were flanked by two *Xba*I restriction sites and synthesized as barcodes for each reporter gene plasmid separately. Each barcode sequence was cloned between the luciferase open reading frame (ORF) and the SV40 poly(A) terminator sequence of the firefly luciferase vector pGL4.24 (Promega). The barcode sequences were designed to serve as templates of qRT-PCR primers of comparable efficiency and allowed parallel detection of individual reporter genes in a single qRT-PCR. The primer sequences are given in **Appendix Table 6**. We cloned an additional unique barcode as an internal reference control to normalize for basal expression of the reporter gene and to control for variation in transfection efficiency and cell death. Altogether, the plasmids contained the same plasmid backbone but differed in the barcode sequence. The putative regulatory DNA sequences (encompassing 567-1,012 bp) were amplified by PCR with human genomic DNA as template. The reporter plasmids containing either the reference (non-effect G-allele) or mutant allele (effect T-allele) within the binding motif of the TF BTB and CNC homology 1 (BACH1), differed in a single nucleotide (nt) and were amplified using single-stranded oligonucleotides (each 79 nt). The primer sequences are given in **Appendix Table 7**. All PCR products were purified (QIAquick Gel Extraction Kit, Qiagen) and inserted into the *Hind*III or *Kpn*I restriction sites of the barcoded reporter plasmids upstream of the minimal promoter. The control plasmids were not modified. All reporter plasmids were amplified in 5-alpha competent *E. coli* (NEB). After purification (QiaPrep Plasmid Mini Kit, Qiagen), the reporter plasmids were pooled at equal concentrations determined by Spectrophotometer (MultiScan FC, Thermo Fisher Scientific). Next, three subsamples of the pooled plasmids were separately transformed into *E. coli*. Following 1 hour recovery at 37°C, the three transformed subsamples were also pooled. Subsequently, the pooled transformed *E. coli* culture was transferred to 100 mL $LB_{AMP\,(50\,mg/mL)}$ medium and incubated at 37°C overnight. The bacterial culture was harvested and the plasmid library was extracted using the QiaPrep Plasmid Midi Kit (Qiagen). This procedure ensured equal ratios and concentrations of each

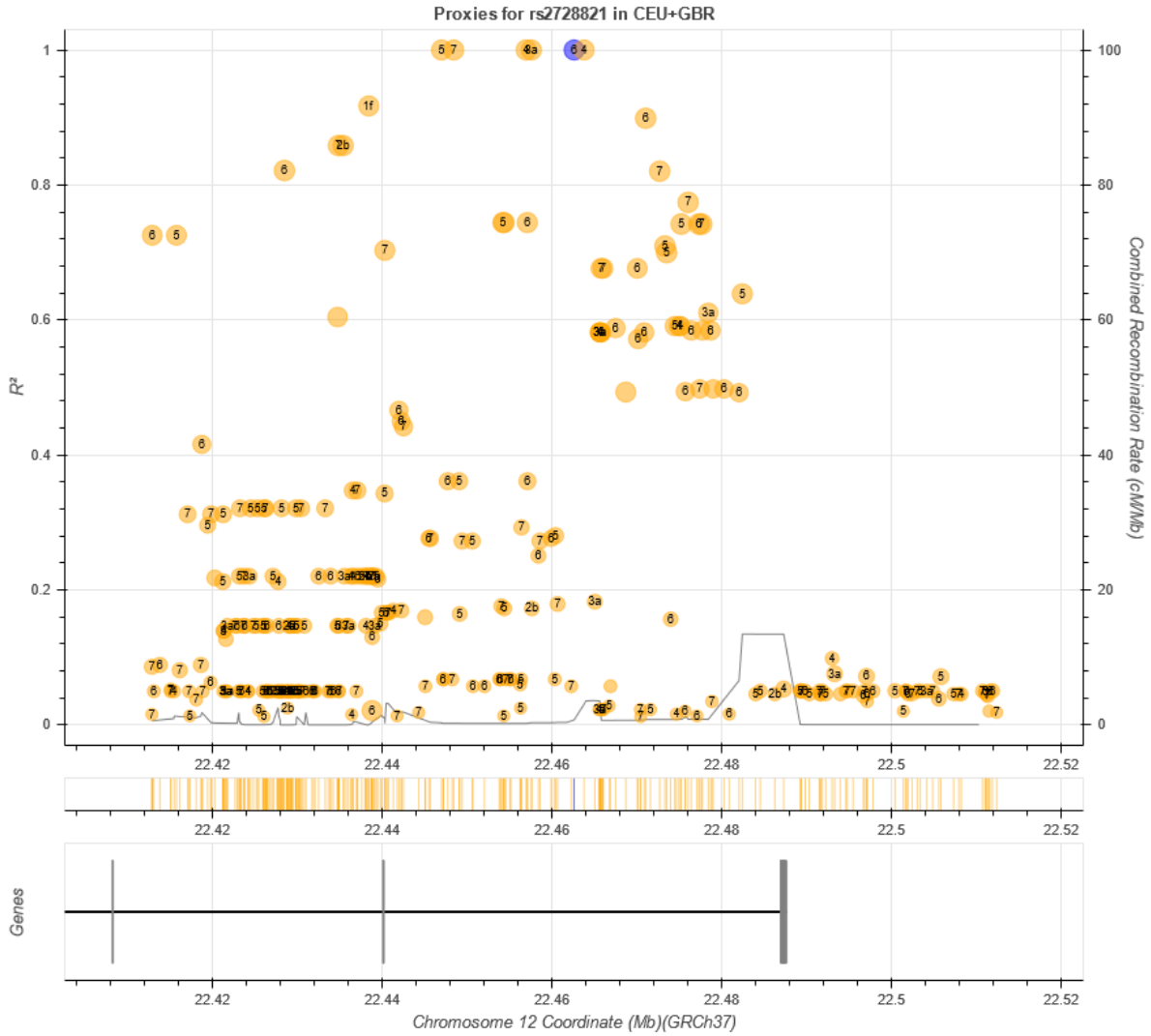
plasmid within the input library, which is a requisite for parallel transfection and quantification of the reporter gene activity. The plasmid library was transfected using 3 µg of input library for 24-48 hours. The expression levels of the reporter genes were quantified by qRT-PCR and normalized to the expression of the internal reference control (empty vector) by the $2^{-\Delta Ct}$ method. Relative fold changes (FC) in activities were normalized according to the manufacturer's instructions (Promega). Differences of transcript levels were calculated using a T-Test.

Preparation and induction of cigarette smoke extract (CSE)

For each CSE sample, smoke of five cigarettes was drawn through 15 mL cell culture medium. 24 h after transfection, the CSE was added to the cells (2 mL per 6-well) in three independent replicates with aliquots of the same CSE. 2 mL medium without CSE were added to the control cell plate. After CSE stimulation for 24 hours (HeLa cells) or 6 hours (ihGFs), the cells were harvested and washed twice with PBS. Subsequently, RNA extraction was performed.

Quantitative real-time PCR (qRT-PCR)

Human cells were harvested and total RNA was extracted from the cells using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. The RNAs of the reporter assays were subsequently cleaned with the *TURBO* DNase Kit (Ambion) to remove any traces of DNA. Complete removal of DNA was verified by PCR using plasmid backbone primers and a plasmid DNA template as a positive control. Complementary DNA (cDNA) was synthesized from 250-500 ng DNA-free total RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qRT-PCR was performed using SYBR Select Master Mix (Applied Biosystems) with the primers listed in the **Appendix Table 8**. Fold changes of relative gene expression were calculated by the $2^{-\Delta\Delta Ct}$ method. For the CRISPRa experiments, the threshold cycle (Ct) values of *ST8SIA1* were normalized to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) Ct values. Two sgRNA plasmids, each expressing an individual scrambled sgRNA were transfected as controls. For all qRT-PCR experiments, differences of transcript levels were calculated with a T-Test. Correction for multiple testing was performed using Welch's or Bonferroni–Holm's correction where applicable, using GraphPad Prism 6 (GraphPad Software, Inc.).



Appendix Figure 1. Proxy SNPs for rs2728821 in CEU and GBR. 12 SNPs are in LD (r^2) > 0.8 and span the second and first intron of *ST8SIA1* (*Genes*).

Appendix Tables

Appendix Table 1. R^2 proxy SNPs of rs2728821 in CEU and GBR populations.

RS_Number	chr.12_Position	alleles	Dprime	R2
rs3819872	22428485	(G/C)	0.9555	0.8214
rs1985103	22434841	(T/C)	0.9564	0.8584
rs2012722	22435394	(T/G)	0.9564	0.8584
rs4762901	22438424	(G/A)	0.9576	0.9171
rs2160536	22446994	(C/T)	1.0	1.0
rs2216230	22448429	(A/G)	1.0	1.0
rs2193179	22456961	(T/C)	1.0	1.0
rs2728818	22457581	(A/G)	1.0	1.0
rs2728821	**22462611**	**(A/G)**	**1.0**	**1.0**
rs2287169	22463786	(A/C)	1.0	1.0
rs2900502	22471047	(C/T)	0.989	0.8988
rs2728822	22472695	(A/C)	0.9449	0.8205

Appendix Table 2. eQTL effects of the associated SNPs annotated by the software tool QTLizer.

Index variant	LD-SNP ($r^2 > 0.8$)	LD value (r^2)	Affected Gene	Tissue	P-value	Beta	Effect Allele	Non-Effect Allele	Source
rs2728821	rs1985103	0.82	*ST8SIA1*	Adipose - Subcutaneous	5.1e-13	-0.17	C	T	GTEx v8
rs2728821	rs2012722	0.82	*ST8SIA1*	Adipose - Subcutaneous	5.1e-13	-0.17	G	T	GTEx v8
rs2728821	rs1985103	0.82	*ST8SIA1*	Artery - Tibial	1,00E-11	-0.21	C	T	GTEx v8
rs2728821	rs2012722	0.82	*ST8SIA1*	Artery - Tibial	1,00E-11	-0.21	G	T	GTEx v8
rs2728821	rs2900502	0.88	*ST8SIA1*	Brain - Cerebellum	0.000026	-0.25	T	C	GTEx v8
rs2728821	rs2160536	0.99	*ST8SIA1*	Brain - Temporal cortex in alzheimer's disease cases and controls	0.000088	-	-	-	GRASP 2 Catalog
rs2728821	rs4762901	0.89	*ST8SIA1*	Nerve - Tibial	6.4e-16	-0.28	A	G	GTEx v8
rs2728821	rs4762901	0.89	*ST8SIA1*	Skin - Sun exposed (Lower leg)	2.6e-9	-0.27	A	G	GTEx v8
rs2728821	rs1985103	0.82	*ST8SIA1*	Skin - Sun exposed (Lower leg)	9.4e-9	-0.26	C	T	GTEx v8
rs2728821	rs2012722	0.82	*ST8SIA1*	Skin - Sun exposed (Lower leg)	9.4e-9	-0.26	G	T	GTEx v8
rs2728821	rs4762901	0.89	*FAM156A*	Liver	2.5e-7	-	-	-	Haploreg v4.1
rs2728821	rs4762901	0.89	*NCOR1*	Liver	0.0000017	-	-	-	Haploreg v4.1
rs2728821	rs4762901	0.89	*ACYP2*	Liver	0.000002	-	-	-	Haploreg v4.1
rs2728821	rs4762901	0.89	*IARS2*	Liver	0.0000028	-	-	-	Haploreg v4.1

Appendix Table 3. Oligonucleotides of the *ST8SIA1* EMSA probes.

Probe & amplicon hg19 genomic coordinates (oligo length: 43 bp)	**Forward (5'-3')**	**Reverse (5'-3')**	**3' Modification**
BACH1 in Region tagged by rs3819872 chr21:30699076-30699329	CCTATTCCAGTACTGCTGTGAG TCAGGGGAATGATATGGAGGG	CCCTCCATATCATTCCCCTGAC TCACAGCAGTACTGGAATAGG	Biotin
BACH1 in Region tagged by rs3819872 chr21:30699076-30699329	CCTATTCCAGTACTGCTGTGAG TCAGGGGAATGATATGGAGGG	CCCTCCATATCATTCCCCTGAC TCACAGCAGTACTGGAATAGG	-
rs2012722-G in Region tagged by rs2012722 chr21:30699076-30699329	TAGACATGCCCATGTGACTCAG TCCTGATTAATGATACATGAG	CTCATGTATCATTAATCAGGAC TGAGTCACATGGGCATGTCTA	Biotin
rs2012722-T in Region tagged by rs2012722 chr21:30699076-30699329	TAGACATGCCCATGTGACTCAT TCCTGATTAATGATACATGAG	CTCATGTATCATTAATCAGGAA TGAGTCACATGGGCATGTCTA	Biotin
rs2012722-G in Region tagged by rs2012722 chr21:30699076-30699329	TAGACATGCCCATGTGACTCAG TCCTGATTAATGATACATGAG	CTCATGTATCATTAATCAGGAC TGAGTCACATGGGCATGTCTA	-
rs2012722-T in Region tagged by rs2012722 chr21:30699076-30699329	TAGACATGCCCATGTGACTCAT TCCTGATTAATGATACATGAG	CTCATGTATCATTAATCAGGAA TGAGTCACATGGGCATGTCTA	-

Appendix Table 4. *ST8SIA1* lead SNP rs2728821 and SNPs in LD ($r^2 > 0.8$).

SNP	Position at chr12 (hg19)	Distance to next SNP (kb)	Distance to nearest predictive regulatory pattern (kb)	SNP-specific TFBS (by Transfac professional)	LD ($r^2 > 0.8$) with rs2728821
rs3819872	22428485	6.4	0.1	-	0.8
rs1985103	22434841	0.6	0.5	-	0.8
rs2012722	**22435394**	**3.0**	**0**	**BACH1**	**0.8**
rs4762901	22438424	8.6	0.9	-	0.9
rs2160536	22446994	1.4	1.5	-	1.0
rs2216230	22448429	8.5	2.1	-	1.0
rs2193179	22456961	0.6	0.4	-	1.0
rs2728818	22457581	5.0	0.01	-	1.0
rs2728821	22462611	1.2	2.7	-	**Lead SNP**
rs2287169	22463786	7.3	1.6	-	1.0
rs2900502	22471047	1.6	0.1	-	0.9
rs2728822	22472695	NA	0.6	-	0.8

grey shade = SNPs mapped to DNA elements with features of regulatory function on gene expression, indicated by ENCODE data.

Appendix Table 5. Oligonucleotides of CRISPRa gRNA probes.

Probe & amplicon hg19 genomic coordinates/ upstream of the transcription start site (TSS) (sequence length: 19 nt each)	Forward (5'-3') (overhangs in red)	Reverse (5'-3') (overhangs in red)	Description
Promoter_-14 TSS chr12:22487663-22487681	CACCGGCGCAGAGAGCGCGTCTCG	AAACCGAGACGCGCTCTCTGCGCC	positive control
Promoter_-67 TSS chr12:22487716-22487734	CACCGGGGGCAGGATAGCGGTCCC	AAACGGGACCGCTATCCTGCCCCC	positive control
Region tagged by rs3819872_-11 TSS chr12:22429457-22429475	CACCGAGTCATGGAAGTGCCAAGG	AAACCCTTGGCACTTCCATGACTC	*ST8SIA1*
Region tagged by rs3819872_-15 TSS chr12:22428945-22428963	CACCGGTGAGTCAGGGGAATGATA	AAACTATCATTCCCCTGACTCACC	*ST8SIA1*
Region tagged by rs2012722_-1 TSS chr12:22435305-22435323	CACCGTTGCGTTTGTCAACTATAC	AAACGTATAGTTGACAAACGCAAC	*ST8SIA1*
Region tagged by rs2012722_-7 TSS chr12:22435718-22435736	CACCGAAGGGGTCTAATGTCTGGT	AAACACCAGACATTAGACCCCTTC	*ST8SIA1*
non-targeting scramble gRNA *taken from Liu et al. 2018**	CACCGCACTACCAGAGCTAACTCA	AAACTGAGTTAGCTCTGGTAGTGC	negative control
microRNA hsa-miR-374b-5p_-10 TSS chrX:73438697-73438715	CACCGACCTAATTCAACTGCTTGC	AAACGCAAGCAGTTGAATTAGGTC	negative control

*Liu, Y., Zhao, G., Xu, C. F., Luo, Y. L., Lu, Z. D., & Wang, J. (2018). Systemic delivery of CRISPR/Cas9 with PEG-PLGA nanoparticles for chronic myeloid leukemia targeted therapy. Biomaterials science, 6(6), 1592-1603.

Appendix Table 6. qRT-PCR detectable barcodes of the reporter gene assays and qRT-PCR program.

Function	Sequence (5' - 3')	qRT-PCR primer pairs (5' - 3')	Annealing temp (°C)	PCR cycle number	Amplification factor	Primer-Efficiency (%)
control	ACA CAG CCT CGG TCG TTT ACA CGC CGG CCA CGG GGC AGG TGA AGG TGG GGC ACG TTG AAG TCT TCT TGA ACA CGG GGC AC	GTGCCCCGTGTTCAAGAAG ACACAGCCTCGGTCGTTTA	64	30x	1.9	85.8
test (used for Region tagged by rs2012722, rs2012722-T)	TAG TTC AGC GGC CTC ACG CAC GCC GGC CAC GGG GCA GGT GAA GGT GGG GCA CGT TGA AGT CGA GGA GGG CGA CAG TAT TT	AAATACTGTCGCCCTCCTCG TAGTTCAGCGGCCTCACG				84.6
control	TTC CCC ACA CGA GCA GAA CAA GAC CAA CTC CGT TTT GAA TAG AAA ACC TTC TTG TTT GAA ATG GGT GTG AAT GTG GAG CC	GGCTCCACATTCACACCCA TTCCCCACACGAGCAGAAC	60	35x		85.4
test (used for Region tagged by rs3819872, rs2012722-G)	TGT CCC CAA ATC CCC AAG CAG ATT TGT CTG TTT GGT GAT TTT ATA AAG TAA AAA CAG TTA AGA ACA GAA GAG CCG CTG GA	TCCAGCGGCTCTTCTGTTC TGTCCCCAAATCCCCAAGC				85.1

Appendix Table 7. PCR and cloning primers used for barcode reporter gene assays.

Probe & Genomic Coordinates (construct length w/o restriction sites)	**Forward (5'-3')**	**Reverse (5'-3')**	**Description**
PCR_XbaI_Barcode-Set B_1 80 bp in *LOC542299 stress-induced protein 1* [Zea mays]	ATTTCTAGAGTGCCCCGTGTTCAAGAAG	CTGTCTAGAACACAGCCTCGGTCGTTTA	Barcode (pGL4.24)
PCR_XbaI_Barcode-Set B_2 80 bp in *LOC542299 stress-induced protein 1* [Zea mays]	ATTTCTAGAAAATACTGTCGCCCTCCTCG	CTGTCTAGATAGTTCAGCGGCCTCACG	Barcode (pGL4.24)
PCR_XbaI_Barcode-Set C_1 80 bp in *LOC542509 defective kernel 1* [Zea mays]	ATTTCTAGAGGCTCCACATTCACACCCA	CTGTCTAGATTCCCCACACGAGCAGAAC	Barcode (pGL4.24)
PCR_XbaI_Barcode-Set C_2 80 bp in *LOC100284365 frataxin* [Zea mays]	ATTTCTAGATCCAGCGGCTCTTCTGTTC	CTGTCTAGATGTCCCCAAATCCCCAAGC	Barcode (pGL4.24)
PCR_HindIII_near to rs3819872 567 bp in *ST8SIA1*	CCCAAGCTTACCAGATGGGGCTCAGTG	CCCAAGCTTCCCGAGTGTTCACACAGTTAG	*ST8SIA1* Region tagged by rs3819872
PCR_KpnI_rs1985103&rs2012722 1012 bp in *ST8SIA1*	CGGGGTACCGCCTGGTCAACATAACAAAACC	CGGGGTACCGGGTCTAATGTCTGGTGGGG	*ST8SIA1* Region tagged by rs2012722
PCR_HindIII_BACH1 motif 79 bp in *ST8SIA1*	CCCAAGCTTAAGCTGGACAGATTCCTG	CCCAAGCTTCCCAGGCTTTCTTGCAG	BACH1 motif
Oligonucleotide_BACH1_reference allele 79 nt in *ST8SIA1*	AAGCTGGACAGATTCCTGCTCATGTATCATTAATCAGGACTGAGTCACATGGGCATGTCT AACTGCAAGAAAGCCTGGG		BACH1 motif reference allele G
Oligonucleotide_BACH1_mutant allele 79 nt in *ST8SIA1*	AAGCTGGACAGATTCCTGCTCATGTATCATTAATCAGGACTGAGTAACATGGGCATGTCT AACTGCAAGAAAGCCTGGG		BACH1 motif mutant allele T
PCR_pGL4.24_Backbone 360 bp	AGAGCCTTCAACCCAGTCAG	GTTTCGCCACCTCTGACTTG	pGL4.24 Backbone

Appendix Table 8. Primers used for qRT-PCR.

Target Gene (UCSC Genes) & Barcode reporter gene assay	**Forward (5'-3')**	**Reverse (5'-3')**
GAPDH (taken from (Freitag-Wolf et al. 2019)	CAAATTCCATGGCACCGTCA	CCTGCAAATGAGCCCCAG
ST8SIA1 (taken from (Freitag-Wolf et al. 2019)	TGTGTCGTGGTCCTCTGTTG	CCCCTGCACGATCTCTTTCT

References

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. 2012. Rna-seqc: Rna-seq metrics for quality control and process optimization. Bioinformatics. 28(11):1530-1532.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. Star: Ultrafast universal rna-seq aligner. Bioinformatics. 29(1):15-21.

Ewels P, Magnusson M, Lundin S, Kaller M. 2016. Multiqc: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32(19):3047-3048.

Freitag-Wolf S, Munz M, Wiehe R, Junge O, Graetz C, Jockel-Schneider Y, Staufenbiel I, Bruckmann C, Lieb W, Franke A et al. 2019. Smoking modifies the genetic risk for early-onset periodontitis. J Dent Res. 98(12):1332-1339.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics. 28(20):2678-2679.

Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The molecular signatures database (msigdb) hallmark gene set collection. Cell Syst. 1(6):417-425.

Liu H, Wei Z, Dominguez A, Li Y, Wang X, Qi LS. 2015. Crispr-era: A comprehensive design tool for crispr-mediated gene editing, repression and activation. Bioinformatics. 31(22):3676-3678.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol. 15(12):550.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the crispr-cas9 system. Nat Protoc. 8(11):2281-2308.

Sayols S, Scherzinger D, Klein H. 2016. Dupradar: A bioconductor package for the assessment of pcr artifacts in rna-seq data. BMC Bioinformatics. 17(1):428.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for rna-seq: Accounting for selection bias. Genome Biol. 11(2):R14.

Zyla J, Marczyk M, Domaszewska T, Kaufmann SHE, Polanska J, Weiner J. 2019. Gene set enrichment for reproducible science: Comparison of cerno and eight other algorithms. Bioinformatics. 35(24):5146-5154.