**Supplementaries**

**Supplementary Material**

**G x G interaction analysis**

For statistical modelling, genotypes (G) were encoded assuming a dominant effect of the minor allele, i.e. G=1 for homozygous or heterozygous carriers of the minor allele, G=0 for homozygous carriers of the major allele.

A dominant model was chosen here because it provides greater statistical power for low allele frequencies in G×G interaction analyses, where numerous zero counts can be expected for at least one of the two homozygous rare genotypes. In the standard CC design of G×G, disease status D is treated as the response variable whereas genotypes G1 and G2 are treated as predictor variables alongside an interaction effect G1G2, i.e.

In the CO design, the genotype of the first SNP (G1) is treated as a predictor variable, with respective regression coefficient δ1 representing the interaction effect, whereas the other genotype (G2) is treated as the response variable, i.e.

(1)

As described in (Piegorsch, Weinberg, & Taylor, 1994), no classical confounders such as age or sex were included in the CO model because their main effects on D cannot sensibly be modelled when using a CO design. Moreover, none of the SNPs that we analysed in the present study is known to be associated with age, sex or smoking. For the CO design to be valid, two key assumptions need to be fulfilled: (i) The disease of interest has a prevalence < 5% and (ii) G1 and G2 are uncorrelated in the general population. Here, the first requirement is met because the prevalence of the selected disease phenotype is < 0.1%. To ensure the second requirement, we assessed statistical independence of genotypes in our controls by using the Wald test of δ1 in (1). For further analyses we included only those pairs of SNPs with p>0.05 in the controls, which is an indicator that G1 and G2 are independent in the population. With the same test in the cases p-values < 0.05 indicated underlying interactions between G1 and G2. All statistical analyses were performed with either R (v. 3.5.0) (The-R-Development-Core-Team, 2008) or PLINK 21 (Purcell et al., 2007).

**Bioinformatical investigations of expressed quantitative trait loci (eQTLs), linkage disequilibrium (LD) and transcription factor binding sites (TFBS)**

To analyse potential correlations between a SNP allele and gene expression, we used the software tool QTLizer (<http://genehopper.de/qtlizer>) (Munz et al., 2020). We used the online tool LDproxy (Machiela and Chanock, 2015) to determine LD correlations (*r2* > 0.8) of SNPs that showed association with periodontitis at p < 5 x 10-6 in our previous GWAS (Munz et al., 2017) and GWAS meta-analysis (Munz et al., 2019). SNP sequence to motif alignments were performed with libraries of matrix descriptions for TFBS from Transfac professional (geneXplain), SNPInspector (Genomatix) and the open-access database Jaspar. The TFBS motif was confirmed using the web interface for Position Weight Matrix (PWM) model generation and evaluation, PWMTools (Ambrosini et al., 2018).

**5’- and 3’-Race-PCR and qRT-PCR**

RNA was extracted from human cells (Hela, B cell lines GM12878 and Raji), using the RNeasy Mini Kit (Qiagen). Complementary DNA (cDNA) was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). 5’ and 3’ RACE-PCR was performed using the RACE Marathon-System (Clontech) according to the manufacturer’s instructions. qRT-PCR was performed using SYBR Select Master Mix (Applied Biosystems). Gene expression of *CTD-2353F22.1*, *SLC1A3* and *GAPDH* were quantified using the 2-ΔΔCT method. The primer sequences are listed in Appendix Table 2.

**Cell culture**

In brief, the suspension B cell line GM12878 was cultivated in RPMI-1640 supplemented with 2mM L-Glutamin, 15% fetal bovine serum (FBS) and 1% penicillin-streptomycin (pen/strep). The B lymphocyte Raji cell line was cultivated in RPMI-1640 containing 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 4500 mg/L glucose and 1500 mg/L sodium bicarbonate supplemented with 10 % FBS and 1 % pem/strep and HeLa cells were grown in Earle’s MEM medium, containing 10% fetal bovine serum, 2mM L-Glutamine, 1% non-essential Amino acids and 1% pen/strep in T75 flasks and cultured at 37°C with 5% CO2. The suspension cell lines GM12878 and Raji cells were cultivated in an upright position. HeLa cells were seeded at 80,000 cells per well in 6-well plates (TPP Techno Plastic Products) and were cultured to 50-60% confluence.

**CRISPR-dCas9 activation**

CRISPR-dCas9 activation (CRISPRa) provides the possibility to test whether a genomic site serves as cis-regulatory element for a target gene of interest (Simeonov et al., 2017). It allows specific and efficient quantification of the regulatory potential of a chromatin sequence on the expression of candidate target genes in the endogenous context including naturally occurring variants. We used CRISPRa to analyze if the 2 periodontitis associated DNA elements, tagged by rs1122900 and rs6887423, had regulatory effects on *CTD-2353F22.*1 and *SLC1A3* expression. We tiled 9 individual short guide RNAs (sgRNAs) at the putative regulatory elements. To directly induce the expression of *CTD-2353F22.*1 and *SLC1A3*, we tiled two sgRNAs at the promoter of these genes. SgRNAs were designed with the online tool CRISPR-ERA (Liu et al., 2015) according to the protocol described in (Ran et al., 2013). As control, we used a scrambled sgRNA with no genomic target. The sgRNAs were synthesized (Metabion International AG) and cloned into the BbsI site of sgRNA(MS2) cloning backbone vector (Addgene Plasmid #61424). The CRISPRa system was transfected into HeLa cells using jetPEI transfection reagent. sgRNAs that targeted the regulatory regions were individually transfected in biological triplicates. The 2 sgRNAs that targeted the promoter and the scrambled control sgRNA were separately transfected in biological triplicates. For CRISPRa, each 6-well was transfected with 1μg sgRNAs(MS2), consisting of four sgRNAs, with 250 ng plasmid (addgene #61424) each, 1μg dCAS9-VP64\_GFP [#61422] and 1μg MS2-P65-HSF1\_GFP [#61423] and incubated for 44 hours. All plasmids were obtained from Addgene, gifted by Feng Zhang.

**RNA-Sequencing**

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 were 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, Magnusson, Lundin, & Kaller, 2016) summarizing on a number of approaches, including fastqc (available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc), dupradar (Sayols, Scherzinger, & Klein, 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, Huber, & Anders, 2014) version 1.26 was used. The only contrast fitted was the comparison between the *CTD-2353F22.*1 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, Wakefield, Smyth, & Oshlack, 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).

**Luciferase reporter gene assays**

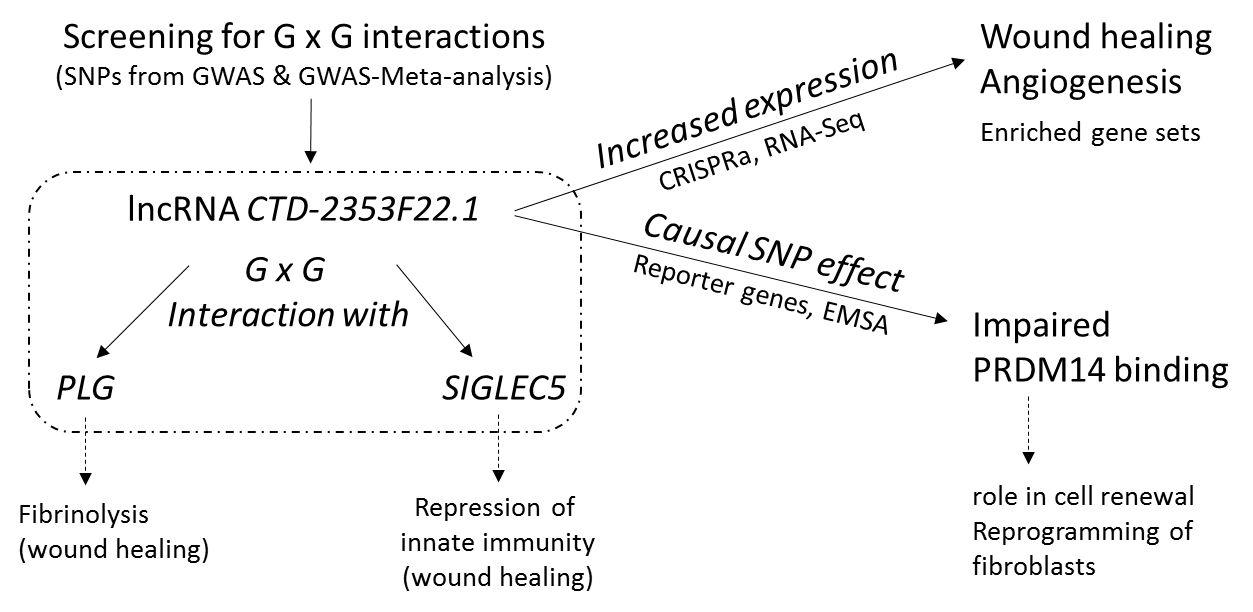
Genomic DNA (gDNA) was extracted from human cells (primary gingival fibroblast cells) using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) and the purified gDNA was used as a PCR-template. The DNA sequences that spanned the putative causal SNPs were amplified by PCR to subsequently test their regulatory potential on reporter gene expression. Specifically, 73 bp up- and downstream of rs1122900 and rs56038114 and 71 bp up- and downstream of rs6887423 were PCR amplified and cloned to the promoter of the firefly luciferase reporter gene of plasmid pGL4.24 (Promega). PCR products were amplified using Biozym Taq DNA Polymerase (Biozym) with forward and reverse primers containing the KpnI and XhoI restriction sites, respectively. The primer sequences are listed in the appendix.

HeLa cells were co-transfected in triplicates using jetPEI transfection reagent according to the manufacturer’s instruction with 2.7 µg luciferase reporter plasmid pGL4.24 carrying the allele specific SNP sequence cloned to the promoter of the luciferase reporter gene, and 0.3 µg *renilla* luciferase reporter vector (pRL-SV40, Promega). For control, HeLa cells were transfected with the empty pGL4.24 plasmid and phRL-SV40 for 24 h. After 24 hours, the firefly and *renilla* luciferase activities were quantified using the Dual Luciferase Stop & Glo Reporter Assay System (Promega) with a luminometer (Orion II Microplate Luminometer, Berthold).The activities of the reporter gene assays were quantified as relative light units that were normalized as the ratio of firefly luciferase activity to *renilla* luciferase activity. Relative fold changes of normalized reporter gene activities were calculated using a T-Test.

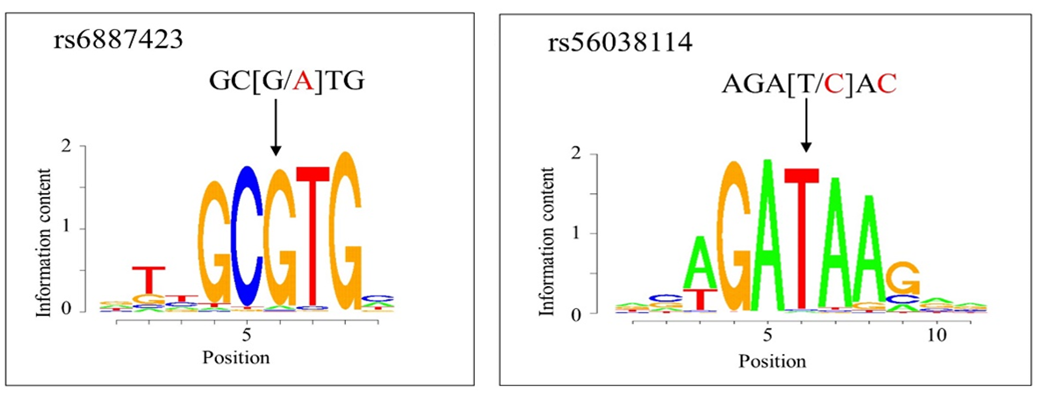
**Electrophoretic Mobility Shift Assay (EMSA)**

DNA-protein binding was determined by using the Gelshift Chemiluminescent EMSA Kit (Active Motif). The nuclear protein extract was prepared from the B lymphocyte cell line Raji using the Nuclear Extract Kit (Active Motif). All Biotin 3’ end-labeled and unlabeled oligonucleotides had a total length of 43 bp including each SNP allele. Cold and 3′-biotinylated forms were obtained by annealing with their respective complementary primers. For binding reaction, 20 fmol biotin-labeled, double-stranded oligonucleotides were incubated with nuclear extract (8-10 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 PRDM14 antibody (2 μL of 288 µg/mL (LSBio) 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 100 V for 2 hours. For the Supershift binding reaction, 10 µg nuclear extract and 20 fmol biotin-labeled double-stranded oligonucleotides were incubated at room temperature for 20 minutes with 1 x Binding Buffer and 2 µL of specific antibody. To verify the result of specific DNA-protein interaction, unlabeled oligonucleotides (4 pmol) were added to the binding reaction. The reactions were loaded onto a 5 % native polyacrylamide gel and run in 0.5 x TBE buffer at 100 V for 1-1.5 hours. After electric transfer of the products to a nylon membrane, the membrane was cross-linked at 120 mJ/cm2. The biotin-labeled oligonucleotides were detected by chemiluminescence an ECL ChemoStar Touch Imager (Intas, Germany). The allele-specific binding of the blotted antibody was quantified using the open-source image-processing program, ImageJ (Rueden et al., 2017). Sequences of oligonucleotide probes are given in the Appendix Table 6.

**Appendix Figures**

****

**Appendix Figure 1.** Workflow illustrating the experimental approach and main results

****

**Appendix Figure 2. PWMs of TFBS for AHR and GATA1.**

**Left**: The AHR TF binding motif shows a matrix similarity of 100% with the DNA sequence at the common G-allele of rs6887423.

**Right** The GATA1 TF binding motif shows a matrix similarity of 96% with the DNA sequence at the common T-allele of rs56038114.

The SNP alleles are shown in brackets. Letters that do not match to the PWM are given in red.

**Appendix Table 1.** qRT-PCR Primers

|  |  |  |
| --- | --- | --- |
| Designation | Forward (5’-3’) | Reverse (5’-3’) |
| GAPDH-RT | GCATCTTCTTTTGCGTCG | TGTAAACCATGTAGTTGAGGT |
| CTD-2353F22.1-1 | TGT ACC CCT TCT CCA CAA TGA | TAG GAG GGT TGA GAA CAG GC |
| CTD-2353F22.1-2 | CCA CAG GAA ACA AAT GCA TGG | TTT CAT CTC ACT CCC ACC CC |
| CTD-2353F22.1-3 | TGT GCA GCT AAT TCA TCC AGG | ACC AGT CAT ACA GCA AGG TCA |
| CTD-2353F22.1-2/3 | ACC AGG AAC CAA GGA TGC TA | CAC AGT GTG TCC CCA GAG G |

**Appendix Table 2.** Primer used for RACE PCR

CTD-201 = ENST00000510740.1, CTD-202 = ENST00000512329.1, CTD-203 = ENST00000508745.1

|  |  |  |
| --- | --- | --- |
| **Name** | **Sequence** | **Designation** |
| Marathon Adaptor Primer 1 | CCATCCTAATACGACTCACTATAGGGC | RACE mit Marathon Kit |
| Marathon Adaptor Primer 2 | ACTCACTATAGGGCTCAGCGGC | RACE mit Marathon Kit |
| 5-RACE-CTD\_202\_203 | CCCAGTGAATCTGCCTGGATGAATTA | RACE für *CTD* 202&203 |
| 3-RACE-CTD\_202\_203 | CTACATGTCACCACATTGCGATCCT | RACE für *CTD* 202&203 |
| 5-RACE-CTD\_201 | CTGGACCTGTTCCAGTGGCCCTTT | RACE für *CTD* 201 |
| 3-RACE-CTD\_201 | CCACTGGAACAGGTCCAGAAGGAA | RACE für *CTD* 201 |
| 5-RACE-nested-CTD\_202\_203 | CCCCAGAGGAACTAGATGAAGAAG | RACE für *CTD* 202&203 |
| 3-RACE-nested-CTD\_202\_203 | GATGTGGCCACAAACCAAGGATTGT | RACE für *CTD* 202&203 |
| 5-RACE-nested-CTD\_201 | CAGAGGTGAGTGAGGCCTCTCATT | RACE für *CTD* 201 |
| 3-RACE-nested-CTD\_201 | GGCAACTGCTGATTCAAGAAGATTGT | RACE für *CTD* 201 |
| 5-RACE-CTD\_202\_203 FOR | CTGTTATGGGATGGGCATCCTACC | RACE für *CTD* 202&203 |
| 3-RACE-CTD\_202\_203 REV | CACAGTTCTAGAGGGCAGAAGTCC | RACE für *CTD* 202&203 |
| 3-RACE-CTD\_201\_REV | TCCTTGCTTTCCTTGTAAGTTTGCCC | RACE für *CTD* 201 |
| 5-RACE-CTD\_201\_new | CCTTGCTTTCCTTGTAAGTTTGCCC | RACE für *CTD* 201 |
| 5-RACE-CTD\_202\_203\_new | CCTACTTAGAAGGATCGCAATGTGG | RACE für *CTD* 202&203 |

**Appendix Table 3.** Oligonucleotides for cloning reporter gene assay plasmids

Forward-oligonucleotides for cloning genomic DNA sequences including rs1122900, rs6887423 and rs56038114 have a *KpnI* restriction site and reverse-oligonucleotide sequences for the genomic DNA sequence including these SNPs have an *XhoI* restriction site.

|  |  |  |
| --- | --- | --- |
| Designation | Forward (5’-3’) | Reverse (5’-3’) |
| rs1122900 | ggccGGTACCTGTGGGGAAAGTTTTGTGCA | cgaCTCGAGTGCCTTCACAAAAGTTAGGGTG |
| rs6887423 | ggccGGTACCACGACTCCAGCAGAATTGAT | cgaCTCGAGGATCACGTGTTAGGCTGTGG |
| rs56038114 | ggccGGTACCAGGAGGAAGAGGTTTCAGTGA | cgaCTCGAGGATGCCTGACTTGCCATGTA |

**Appendix Table 4.** Oligonucleotides for SNP exchange (Q5) in reporter gene assay plasmids

|  |  |  |
| --- | --- | --- |
| Designation | Forward (5’-3’) | Reverse (5’-3’) |
| rs1122900-Q5 | CGGTTTCTCTaACAGATCTTAC | ATTGAAATTAAGAATTTTACCCC |
| rs6887423-Q5 | ACCTCTCCAtGCAGCTGGAC | AGGTGGTCCAGGAGAGGA |
| rs56038114-Q5 | CAGAGTGAGAcACTGTTTCAAAAAAAAAAAAAATCTCATTTACG | TTGCCCAGGCTGGAGTGC |

**Appendix Table 5.** Oligonucleotides for cloning sgRNAs for CRISPR activation using SAM-system.

|  |  |  |
| --- | --- | --- |
| Designation | Forward (5’-3’) | Reverse (5’-3’) |
| gRNA-rs1122900-1 | CACCGGGCAAATATTGATCTGCTC | AAACGAGCAGATCAATATTTGCCC |
| gRNA-rs1122900-2 | CACCGTCATCGCCCTTGAAGTTGT | AAACACAACTTCAAGGGCGATGAC |
| gRNA-rs1122900-3 | CACCGTACACCCTAACTTTTGTGA | AAACTCACAAAAGTTAGGGTGTAC |
| gRNA-rs1122900-P1-1 | CACCGGCACTTCCCTTAACCATAA | AAACTTATGGTTAAGGGAAGTGCC |
| gRNA-rs1122900-P1-2 | CACCGCCTCTTAGAGATCTCACTT | AAACAAGTGAGATCTCTAAGAGGC |
| gRNA-rs1122900-P1-3 | CACCGTGACGGCAGCCTGGGGTCG | AAACCGACCCCAGGCTGCCGTCAC |
| gRNA-rs6887423-1 | CACCGTTACCTGGGATCACGTGTT | AAACAACACGTGATCCCAGGTAAC |
| gRNA-rs6887423-2 | CACCGAACAATCTACATTGGGCAT | AAACATGCCCAATGTAGATTGTTC |
| gRNA-rs6887423-3 | CACCGCCTACCTTCTCCATGCAGC | AAACGCTGCATGGAGAAGGTAGGC |
| gRNA-rs56038114-1 | CACCGTCTTCTGATATACTAAGCA | AAACTGCTTAGTATATCAGAAGAC |
| gRNA-rs56038114-2 | CACCGCAGCGTATATGCTCTACGT | AAACACGTAGAGCATATACGCTGC |
| gRNA-rs56038114-3 | CACCGACTTGCCATGTAACACAAG | AAACCTTGTGTTACATGGCAAGTC |
| gRNA-SAM-CTD-2353F22.1-1 | CACCGATGAGTACGAACATGAAGG | AAACCCCTTCATGTTCGTACTCATC |
| gRNA-SAM-CTD-2353F22.1-6 | CACCGTGGGTTGTCTGTCTTGTCT | AAACAGACAAGACAGACAACCCAC |
| gRNA-SAM-CTD-2353F22.1-9 (Promoter) | CACCGGATCTCCTTCTTGAGGGAG | AAACCCTCCCTCAAGAAGGAGATCC |
| gRNA-scrambled | CACCGCACTACCAGAGCTAACTCA | AAACTGAGTTAGCTCTGGTAGTGC |

**Appendix Table 6.** Oligonucleotides for EMSA

|  |  |  |
| --- | --- | --- |
| Designation | Forward (5’-3’) | Reverse (5’-3’) |
| rs1122900-(A) | TTAATTTCAATCGGTTTCTCT(A)ACAGAT  CTTACGACTAAAAAA | TTTTTTAGTCGTAAGATCTGT(T)AGAG  AAACCGATTGAAATTAA |
| rs1122900-(C) | TTAATTTCAATCGGTTTCTCT(C)ACAGAT  CTTACGACTAAAAAA | TTTTTTAGTCGTAAGATCTGT(G)AGAG  AAACCGATTGAAATTAA |

**Supplementary References**

DeLuca, D. S., Levin, J. Z., Sivachenko, A., Fennell, T., Nazaire, M. D., Williams, C., . . . Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics, 28*(11), 1530-1532. doi:10.1093/bioinformatics/bts196

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

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. doi:10.1093/bioinformatics/btw354

Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Gotz, S., Tarazona, S., . . . Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics, 28*(20), 2678-2679. doi:10.1093/bioinformatics/bts503

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst, 1*(6), 417-425. doi:10.1016/j.cels.2015.12.004

Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X., & Qi, L. S. (2015). CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics, 31*(22), 3676-3678. doi:10.1093/bioinformatics/btv423

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Piegorsch, W. W., Weinberg, C. R., & Taylor, J. A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med, 13*(2), 153-162. doi:10.1002/sim.4780130206

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308. doi:10.1038/nprot.2013.143

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics, 18*(1), 529. doi:10.1186/s12859-017-1934-z

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. doi:10.1186/s12859-016-1276-2

Simeonov, D. R., Gowen, B. G., Boontanrart, M., Roth, T. L., Gagnon, J. D., Mumbach, M. R., . . . Marson, A. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature, 549*(7670), 111-115. doi:10.1038/nature23875

The-R-Development-Core-Team. (2008). <http://www.r-project.org>. In.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol, 11*(2), R14. doi:10.1186/gb-2010-11-2-r14

Zyla, J., Marczyk, M., Domaszewska, T., Kaufmann, S. H. E., Polanska, J., & Weiner, J. (2019). Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics, 35*(24), 5146-5154. doi:10.1093/bioinformatics/btz447