**Supplementary Methods**

**Combined P-values calculated from the individual GWAS summary statistics**

The p-values of the SNP associations were transformed into the natural logarithm (LN). Transformed p-values were combined (-2\*SUM LN(p\_i) and the Meta-P was calculated with the distribution function of chi2 with 2\*k degrees of freedom (CHISQ.DIST.RT (transformed-p;LNp).

**Cell culture**

We cultured the human bone osteosarcoma cell line SaOS-2 in a complete growth medium of Alpha MEM Eagle (PAN-Biotech GmbH) supplemented with 10% fetal bovine serum and 50 mg/mL Gentamycin. SaOS-2 cells were seeded at 250,000 cells per 6-well before transfection with Lipofectamine 2000 (Thermo Fisher Scientific). HeLa cells were cultured as recently described (Chopra et al., 2021). In brief, cells were cultured in Earle’s MEM supplemented with 10% fetal bovine serum, 2mM L-glutamine, 1% non-essential amino acids and 1% penicillin-streptomycin. Prior to transfection, HeLa cells were seeded at 80,000 cells per well in 6-well plates (Techno Plastic Products, TPP) and were cultured to 50-60% confluence. HeLa-transfection was performed using jetPEI transfection reagent (Polyplus transfection) according to the manufacturer’s instructions.

**Luciferase reporter gene assays**

The putative regulatory DNA sequences (220 bp) spanning 110 bp up- and downstream of the individual alleles of SNP rs6056178 were cloned into the Hind III restriction site of the firefly luciferase vector pGL4.24 (Promega) as described in detail in the supplementary methods. HeLa cells were co-transfected in triplicates with 2.7 μg firefly luciferase reporter plasmid containing the putative regulatory sequence together with 0.3 μg renilla luciferase reporter vector (phRL-SV40, Promega) in 6-well plates for 24 hours. In parallel, HeLa cells were transfected with the empty pGL4.24 plasmid and 0.3 µg phRL-SV40 as control. Firefly and renilla luciferase activities were quantified using the Dual-Luciferase Stop & Glo Reporter Assay System (Promega) with the Orion II Microplate Luminometer (Berthold Technologies). Relative fold changes (FC) in activities were normalized according to the manufacturer’s instructions (Promega) and differences of transcript levels were calculated with a t-test using the software GraphPad Prism 6 (GraphPad Software, Inc.).

**Electrophoretic Mobility Shift Assay (EMSA)**

To characterize allele-specific DNA-protein interaction and TF GATA1 and -2 binding at rs6056178, we performed EMSAs with the Gelshift Chemiluminescent EMSA Kit (Active Motif) as recently described (Chopra et al., 2021). In brief, allele-specific oligonucleotide probes were synthesized (Metabiom International, **Supplementary Table 2**). Nuclear protein extract was prepared from SaOS-2 cells using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Pierce Biotechnology). The double-stranded oligonucleotides corresponding to both alleles of rs6056178 flanked by 21 bp in both cold and 3′-biotinylated form were obtained by annealing with their respective complementary primers. For supershift EMSA, 20 fmol biotin-labeled, double-stranded oligonucleotides were incubated for 20 min with nuclear extract (10 µg) in 1x binding buffer and 2 µL of a specific monoclonal antibody (GATA1 or GATA2, 10 µg/50 µL each (Santa Cruz Biotechnology Inc.) at room temperature. For competition assay, 4 pmol unlabeled double-stranded oligonucleotides were added to the binding reaction. The DNA-protein complexes were electrophoresed in a 5% native polyacrylamide gel in 0.5x TBE buffer at 100 V for 1 hour. After electric transfer of the products on a nylon membrane and cross-linking, the biotinylated probes were visualized by chemiluminescence detection (Chemostar Touch, INTAS). Band intensities were quantified by the absolute value area of the shifted antibody bands using the software ImageJ (Rueden et al., 2017). The band intensity peaks were selected and plotted as peaks. Subsequently, the area of each peak was measured and the peak size was calculated as a percent of the total size of the total area of plotted peaks.

**CRISPR/dCas9 activation (CRISPRa) & quantitative real-time PCR (qRT-PCR)**

We usedCRISPRa to induce *RSPO4* expression in the endogenous chromatin context for RNA-Sequencing (RNA-Seq) and to test if the associated haplotype block tagged by rs6056178 had cis-regulatory effects on *RSPO4* expression. This method allows activation of a specific genomic region within the endogenous chromosomal context, including naturally occurring genetic variants to quantify the regulatory potential of a chromatin sequence on the expression of candidate target genes (Simeonov et al., 2017). For RNA-Seq, we tiled a sgRNA to the promoter of *RSPO4*. For testing cis-regulatory effects, we tiled 2 sgRNAs 50bp downstream of rs114691134 and to position chr20:933,150-933,168 (hg19). SgRNAs were designed with the online tool E-CRISP (Heigwer, Kerr, & Boutros, 2014). A scrambled sgRNA with no genomic target was used as a control (Chopra et al., 2021). Cloning of the sgRNAs into the *Bbs*I site of sgRNA(MS2) cloning backbone vector (plasmid #61424) was performed according to the protocol described in (Ran et al., 2013). The sequences of the sgRNAs are listed in the **Supplementary Table 3**. All CRISPRa experiments were performed with HeLa cells in 3 independent replicates. sgRNA that targeted the *RSPO4* promoter, the SNP-associated and the scrambled control sgRNA were separately transfected. For CRISPRa, each 6-well was transfected with 1µg sgRNAs(MS2) (containing one specific sgRNA), 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). The level of gene activation was analysed by qRT-PCR.Total RNA was extracted using the RNeasy Mini Kit (Qiagen). Complementary DNA (cDNA) was synthesized from 500-1000 ng DNA-free total RNA, using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). SYBR Select Master Mix (Applied Biosystems) was applied with the following primers: *GAPDH*: 5’-CAAATTCCATGGCACCGTCA-3’, 5’-CCTGCAAATGAGCCCCAG-3’, *RSPO4*: 5’-CCAGGAGGTCAACAGGTGCA-3‘, 5‘-GGCAGACACTTCCCCTTGTACA-3‘. Fold changes of relative gene expression were calculated by the 2−ΔΔCt method. The cycle threshold (Ct) values of *RSPO4* were normalized to *GAPDH* Ct values. Differences of transcript levels were calculated with a t-test.

**RNA-Sequencing**

Total RNA was extracted from HeLa cells using the RNeasy Mini Kit (Qiagen) according to the manufacturer’s instructions. 500 - 1,000 ng total RNA of transfected 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). RNA-Seq was performed at the Berlin Institute of Health, Core Facility Genomics. Reads were aligned to the human genome sequences (build GRCh38.p7) 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), 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. 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 the 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 multiple testing using Benjamini-Hochberg correction. The corrected P values are given as q values (false discovery rate [FDR]).

**Supplementary Tables and Figures**

**Supplementary Table 1**. Predicted transcription factor binding sites and allele specific effects on binding affinities (sorted by p-value, inclusion threshold P<0.01)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **TRAP results\*****(single sequence)** | **sTRAP results\*\*****(difference between two sequences)** |
| **Haplo-block****(***r2*>0.8) | **SNP** | **Rank** | **P-value**  | **Matrix ID** | **Matrix name** | **Difference log(p)** **for 2 seq** | **pvalue** **(Seq1)** | **pvalue** **(Seq2)** |
| 1 | rs6086608 | 1 | <0.0052 | M01227 | V$MAFB\_01 | 0 | <0.0049 | <0.0049 |
| 2 | 0.0052 | M00319 | V$MEF3\_B | 0.980 | 0.00515 | 0.049 |
| 3 | 0.0090 | M01181 | V$NKX32\_01 | 0 | 0.009 | 0.009 |
| rs6056164 | 1 | <0.0022 | M01653 | V$HMGIY\_01 | 0 | <0.0022 | <0.0022 |
| 2 | 0.0022 | M00640 | V$HOXA4\_Q2 | 0 | 0.0022 | 0.002 |
| 3 | 0.0027 | M01294 | V$PROP1\_01 | 1.5832 | 0.0027 | 0.105 |
| rs59675440 | 1 | <0.0011 | M01292 | V$HOXA13\_01 | 0 | <0.0002 | <0.0002 |
| 2 | <0.0011 | M00980 | V$TBP\_Q6 | 0.8354 | 0.0106 | 0.072 |
| 3 | 0.0011 | M00081 | V$EVI1\_04 | 0.8448 | 0.0011 | 0.008 |
| 4 | 0.0016 | M01590 | V$SMAD1\_01 | 1.1716 | 0.0016 | 0.024 |
| rs6056178 | 1 | <0.0027 | M00789 | V$GATA\_Q6 | 1.5032 | <0.0025 | 0.080 |
| 2 | <0.0027 | M00076 | V$GATA2\_01 | 2.1045 | <0.0025 | 0.318 |
| 3 | 0.0027 | M00687 | V$ALPHACP1\_01 | 0.8607 | 0.0027 | 0.019 |
| 4 | 0.00542 | M00077 | V$GATA3\_01 | 1.3080 | 0.0054 | 0.110 |
| 5 | 0.0145 | M00128 | V$GATA1\_04 | 0.9806 | 0.0145 | 0.139 |
| 2 | rs6086607 | 1 | <0.0034 | M00630 | V$FOXM1\_01 | 1.1734 | <0.0034 | 0.050 |
| 2 | 0.0034 | M00792 | V$SMAD\_Q6 | 0.1703 | 0.0034 | 0.005 |
| rs11476402 | 1 | 0.0039 | M01009 | V$HES1\_Q2 | -0.0381 | 0.0038 | 0.004 |
| 2 | 0.0073 | M00770 | V$CEBP\_Q3 | -0.0455 | 0.0073 | 0.007 |
| rs6056182 | No motif at P<0.01 found |  |  |  |  |

\*informs which TFs will likely bind this DNA sequence

\*\*informs how potential TF binding sites are affected by sequence differences

**Supplementary Table 2**. Sequences of allele-specific oligonucleotide EMSA probes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Probe****(length:** **43 bp)** | **Forward (5'-3')** | **Reverse (5'-3')** |  | **3' Modification** |
| rs6056178-G\_Biotin | TTTATGGCTCTTAGCCAATCAGATAGGTGGAAAATGGTAGCTC | GAGCTACCATTTTCCACCTATCTGATTGGCTAAGAGCCATAAA |  | Biotin |
| rs6056178-T\_Biotin | TTTATGGCTCTTAGCCAATCATATAGGTGGAAAATGGTAGCTC | GAGCTACCATTTTCCACCTATATGATTGGCTAAGAGCCATAAA |  | Biotin |
| rs6056178-G | TTTATGGCTCTTAGCCAATCAGATAGGTGGAAAATGGTAGCTC | GAGCTACCATTTTCCACCTATCTGATTGGCTAAGAGCCATAAA |  | - |
| rs6056178-T | TTTATGGCTCTTAGCCAATCATATAGGTGGAAAATGGTAGCTC | GAGCTACCATTTTCCACCTATATGATTGGCTAAGAGCCATAAA |  | - |

**Supplementary Table 3**. sgRNA sequences used for CRISPRa

|  |  |  |  |
| --- | --- | --- | --- |
| **Probe & hg19 position****(oligo length: 19 bp)** | **Forward (5'-3') (overhangs in red)** | **Reverse (5'-3') (overhangs in red)** | **Description** |
| Promoterchr20:982,950-982,968 | CACCGGGTCTAGTGAGGGCGTTGG | AAACCCAACGCCCTCACTAGACCC | positive control |
| gRNA 1chr20:932,305-932,323 | CACCGGACTTAGGGTGCCAAGACC | AAACGGTCTTGGCACCCTAAGTCC | *RSPO4* |
| gRNA 2chr20:933,150-933,168 | CACCGGTCCAGACAGACTCAGGGG | AAACCCCCTGAGTCTGTCTGGACC | *RSPO4* |
| non-targeting scramble gRNA*\** | CACCGCACTACCAGAGCTAACTCA | AAACTGAGTTAGCTCTGGTAGTGC | negative control |

\*taken from 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.

**Supplementary Table 4.** Enriched gene sets with AUC ≥ 0.7

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **ID** | **Title** | **N** | **AUC** | **p.value** | **FDR** |
| **tmod** | LI.M127 | type I interferon response | 12 | 0.71 | 0.0000045 | 0.0006 |
| LI.M32.1 | platelet activation (II) | 21 | 0.72 | 0.000005 | 0.0006 |
| LI.M32.0 | platelet activation (I) | 22 | 0.73 | 0.000058 | 0.0059 |
| DC.M8.74 | Undetermined | 10 | 0.77 | 0.00011 | 0.0074 |
| LI.M136 | TBA | 16 | 0.78 | 0.00019 | 0.0095 |
| DC.M7.31 | Undetermined | 32 | 0.75 | 9.3e-8 | 0.000037 |
|  |  |  |  |  |  |  |
| **GO** | M11929 | RESPONSE TO INTERFERON ALPHA | 19 | 0.78 | 0.0000014 | 0.00017 |
| M15777 | GLUCOSE CATABOLIC PROCESS | 27 | 0.72 | 0.0000033 | 0.00029 |
| M24299 | GLYCOLYTIC PROCESS THROUGH FRUCTOSE 6 PHOSPHATE | 22 | 0.74 | 0.0000046 | 0.00037 |
| M12577 | RIBOSOMAL LARGE SUBUNIT ASSEMBLY | 26 | 0.75 | 0.000029 | 0.0015 |
| M29007 | CITRULLINE METABOLIC PROCESS | 7 | 0.82 | 0.000036 | 0.0018 |
| M22859 | VIRAL TRANSLATION | 16 | 0.72 | 0.000057 | 0.0024 |
| M23341 | CELLULAR RESPONSE TO INTERFERON ALPHA | 9 | 0.76 | 0.000081 | 0.0031 |
| M11702 | GASTRULATION WITH MOUTH FORMING SECOND | 17 | 0.75 | 0.00014 | 0.0048 |
| M11917 | GLUCOSE 6 PHOSPHATE METABOLIC PROCESS | 20 | 0.73 | 0.0002 | 0.0064 |
| M22424 | ARGININE CATABOLIC PROCESS | 6 | 0.82 | 0.00023 | 0.0071 |
| M13468 | CELL MIGRATION INVOLVED IN GASTRULATION | 9 | 0.76 | 0.00027 | 0.0078 |
|  |  |  |  |  |  |  |
| **Reactome** | M840 | CELL EXTRACELLULAR MATRIX INTERACTIONS | 18 | 0.80 | 0.0000027 | 0.00016 |
| M27216 | LAMININ INTERACTIONS | 30 | 0.75 | 0.0000034 | 0.00017 |
| M4974 | BASIGIN INTERACTIONS | 19 | 0.73 | 0.0000057 | 0.00025 |
| M27491 | RHO GTPASES ACTIVATE IQGAPS | 24 | 0.73 | 0.000018 | 0.00064 |
| M27221 | SCAVENGING BY CLASS A RECEPTORS | 9 | 0.76 | 0.000025 | 0.00084 |
| M27252 | HSF1 ACTIVATION | 27 | 0.76 | 0.00019 | 0.0043 |
|  |  |  |  |  |  |
| **hallmark** | M5928 | MYC TARGETS V2 | 58 | 0.70 | 1.8e-8 | 4.5e-7 |

**Supplementary Table 5.** Up-regulated genes with log2 Fold Change >1 in HeLa cells following *RSPO4* activation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ENSEMBL-ID** | **symbol** | **log2 Fold Change** | **lfcSE** | **pvalue** | **padj** |
| ENSG00000101282 | RSPO4 | 12.046 | 1.181 | 2.1e-24 | 4.2e-21 |
| ENSG00000186417 | GLDN | 8.777 | 1.209 | 3.8e-13 | 3.5e-10 |
| ENSG00000204544 | MUC21 | 7.475 | 1.251 | 2.3e-9 | 0.0000014 |
| ENSG00000267568 | LOC105371899 | 6.862 | 1.235 | 2.8e-8 | 0.000012 |
| ENSG00000267521 |  | 6.798 | 1.315 | 2.4e-7 | 0.00008 |
| ENSG00000074276 | CDHR2 | 6.504 | 0.505 | 5.8e-38 | 2.4e-34 |
| ENSG00000181626 | ANKRD62 | 6.287 | 1.342 | 0.0000028 | 0.00070 |
| ENSG00000165816 | VWA2 | 5.986 | 1.375 | 0.000013 | 0.0023 |
| ENSG00000230749 | MEIS1-AS2 | 5.949 | 1.084 | 4.1e-8 | 0.000017 |
| ENSG00000253520 |  | 5.457 | 1.323 | 0.000037 | 0.0050 |
| ENSG00000280122 |  | 5.211 | 1.501 | 0.00052 | 0.0382 |
| ENSG00000165810 | BTNL9 | 4.637 | 0.756 | 8.4e-10 | 5.1e-7 |
| ENSG00000126259 | KIRREL2 | 4.306 | 0.997 | 0.000016 | 0.0026 |
| ENSG00000099338 | CATSPERG | 4.251 | 1.236 | 0.0006 | 0.04 |
| ENSG00000257588 | LOC101927318 | 3.708 | 1.042 | 0.0004 | 0.0303 |
| ENSG00000105997 | HOXA3 | 3.145 | 0.382 | 1.9e-16 | 2.6e-13 |
| ENSG00000134571 | MYBPC3 | 3.113 | 0.751 | 0.000034 | 0.0048 |
| ENSG00000204421 | LY6G6C | 2.798 | 0.776 | 0.0003 | 0.0266 |
| ENSG00000267750 |  | 2.199 | 0.631 | 0.0005 | 0.0375 |
| ENSG00000198003 | CCDC151 | 2.009 | 0.434 | 0 | 0.0009 |
| ENSG00000108370 | RGS9 | 1.92 | 0.354 | 0 | 0 |
| ENSG00000186832 | KRT16 | 1.838 | 0.493 | 0.0002 | 0.0182 |
| ENSG00000265728 |  | 1.826 | 0.582 | 0.0017 | 0.0817 |
| ENSG00000126016 | AMOT | 1.69 | 0.331 | 0 | 0.0001 |
| ENSG00000213722 | DDAH2 | 1.631 | 0.086 | 0 | 0 |
| ENSG00000253552 |  | 1.604 | 0.513 | 0.0018 | 0.084 |
| ENSG00000182612 | TSPAN10 | 1.586 | 0.281 | 0 | 0 |
| ENSG00000117600 | PLPPR4 | 1.569 | 0.285 | 0 | 0 |
| ENSG00000123405 | NFE2 | 1.485 | 0.289 | 0 | 0.0001 |
| ENSG00000081803 | CADPS2 | 1.397 | 0.34 | 0 | 0.0053 |
| ENSG00000136274 | NACAD | 1.296 | 0.165 | 0 | 0 |
| ENSG00000180998 | GPR137C | 1.227 | 0.256 | 0 | 0.0004 |
| ENSG00000256148 |  | 1.16 | 0.36 | 0.0013 | 0.0678 |
| ENSG00000105855 | ITGB8 | 1.032 | 0.095 | 0 | 0 |
| ENSG00000135740 | SLC9A5 | 1.016 | 0.22 | 0 | 0.0009 |
| ENSG00000130813 | SHFL | 1.006 | 0.102 | 0 | 0 |

**Supplementary Table 6.** Down-regulated genes with log2 Fold Change <1 in HeLa cells following *RSPO4* activation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ENSEMBL-ID** | **symbol** | **log2 Fold Change** | **lfcSE** | **pvalue** | **padj** |
| ENSG00000214814 | FER1L6 | -7.088 | 1.228 | 7.9e-9 | 0.0000043 |
| ENSG00000176920 | FUT2 | -5.961 | 1.287 | 0.0000036 | 0.00086 |
| ENSG00000103522 | IL21R | -5.603 | 1.315 | 0.000020 | 0.0033 |
| ENSG00000173338 | KCNK7 | -4.26 | 0.651 | 6.2e-11 | 4.4e-8 |
| ENSG00000115155 | OTOF | -3.362 | 0.275 | 2.3e-34 | 7.7e-31 |
| ENSG00000260220 | CCDC187 | -3.027 | 0.727 | 0 | 0.0045 |
| ENSG00000155629 | PIK3AP1 | -2.082 | 0.388 | 8.2e-8 | 0.000030 |

****

****

****

**Supplementary Figure 1.** LD-maps forSNPs at *RSPO4* that were associated with PIII-IV/C.

The associated SNPs located on 3 haplotype blocks with strong linkage (*r2* > 0.8), the tagging SNPs are highlighted in bold letters below. Plots were generated using the LDproxy Tool (https://ldlink.nci.nih.gov); Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015). See **Table 1** for additional information.

**Panel A:** rs117629802, rs116969995, rs114691134, **rs114035391**, rs114674082, rs60994262, rs78230279, rs56037497

**Panel B:** rs6086608, rs6056164, rs59675440, **rs6056178**

**Panel C:** rs6086607, **rs11476402**, rs6056182

**Transcription Factor ChIP-seq peaks (338 factors) from ENCODE**

**Position (hg19):** chr20:934,302-934,705

**Factor:** GATA2

**Cluster Score (out of 1000):** 497

**Cell line:** SH-SY5Y (neuroepithelioma cell line)

**Position (hg19):** chr20:934,368-934,657

**Factor:** GATA3

**Cluster Score (out of 1000):** 247 and 517

**Cell line:** MCF-7 and T-47D (epithelial cell lines)

DNA-sequence that spans the reported GATA2 and GATA3 ChIP-peaks (yellow marking: GATA binding motif = position of SNP rs6056178 (G/T):

>chr20:934,289-934,723 (hg19human)

TAATtGGAgGAtGAaTCCCTAGAGgTGgATTGCTGGGTCAggggAcATGTGcgCTGTGATTTGAAtaGATGTCAacTCACCTTCTGTAgAGgCTGggTcaAGTCAccCTCTCCTGAGGAAgGTAGGGgAgGGtggcggcTTacccccaCcCAGCccaGGCCCtgtgGTTTcatTTTctttttttttcttttttAtGGCTCTTAGCCAaTcag**(t)atag**GtGGAAaATGgTAGCTCTGtgTGGctttattGtcATCTAATTTGAgTAAgGGTGTTTGGAgTTCCtTTTCTGGATgAgGcAtTtTTGCCACAcggggACAgcgCTGAGCcgacGAcAcgGCagTgCcgCTCTGGACAcgCtGATTCTCCCtgAGGcTCAtCTTcTcAGGCTgAGTGggAcTGTGCCTGTGGATGCCcgctcaTggagTAGTcGCCTGGT



**Supplementary Figure 2.** Transcription Factor ChIP-seq Clusters (338 factors, 130 cell types) from ENCODE 3 for the region chr20:934,293-934,736 (hg19) aligned to the positions of the periodontitis associated linked (*r2*>0.8) SNPs (blue vertical lines).



**Supplementary Figure 3**.The custom set of 12 predicted functional partners of *RSPO4* as indicated by the STRING database indicated differential expression after *RSPO4* CRISPR activation (P = 0.0045). Predicted functional partners were RSPO1, RSPO2, RSPO3, UBA52, ZNRF3, RNF43, UBB, RPS2, LGR4, UBC, LGR5, RPS27A.

**Luciferase reporter gene assay**

The reporter gene plasmids were generated in two cloning steps. First, the DNA sequence containing the reference G-allele was amplified by PCR using human genomic DNA as template. The primer sequences were: 5’-CCC AAG CTT GTC AAG TCA CCC TCT CCT GA-3‘ and 5‘-CCC AAG CTT TGT CCC CGT GTG GCA AAA-3‘ (*Hind*III restriction sites underlined). The PCR product was purified (QIAquick Gel Extraction Kit, Qiagen) and inserted into the *Hin*dIII restriction site of the vector upstream of the minimal promoter. The reporter plasmid was amplified in 5-alpha competent *E. coli* (New England Biolabs (NEB)) and extracted using the QiaPrep Plasmid Mini Kit, Qiagen. Cloning success was validated by restriction control and sequencing. For SNP allele exchange construction (G 🡪 T), the Q5 Site-Directed Mutagenesis Kit from NEB with the following primer sequenceswas applied: 5’-TAG CCA ATC ATA TAG GTG GAA AAT G-3’ and 5’-AGA GCC ATA AAA AAG AAA AAA AAA G-3’.

**Supplementary References**

Chopra, A., Mueller, R., Weiner, J., 3rd, Rosowski, J., Dommisch, H., Grohmann, E., & Schaefer, A. S. (2021). BACH1 Binding Links the Genetic Risk for Severe Periodontitis with ST8SIA1. *J Dent Res*, 220345211017510. doi:10.1177/00220345211017510

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

Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat Methods, 11*(2), 122-123. doi:10.1038/nmeth.2812

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

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

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

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