Identification and characterization of causal genetic variants and non-coding RNAs that contribute to the risk of periodontitis

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to

the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Ricarda Müller

2021

This work was carried out from July 2017 to July 2021 at the research group Molecular Genetics of Oral Diseases at the Charité – University Medicine Berlin under the supervision of Prof. Dr. Arne Schäfer.

First Referee

Prof. Dr. Arne Schäfer
Department of Periodontology, Oral Medicine and Oral Surgery
Institute for Dental and Craniofacial Sciences
Charité – University Medicine Berlin
Aßmannshauser Str. 4-6
14197 Berlin, Germany
E-mail: arne.schaefer@charite.de

Second Referee

Prof. Dr. Sigmar Stricker
Department of Biology, Chemistry and Pharmacy
Biochemistry and Genetics
Freie Universität Berlin
Thielallee 63
14195 Berlin, Germany
E-mail: sigmar.stricker@fu-berlin.de

Date of oral examination: 15.09.2021

Summary

Periodontitis is a common complex inflammatory disease of the oral cavity and it is characterized by inflammation of gingival tissue and alveolar bone loss. Recently, a genome-wide association study (GWAS) and two GWAS-meta-analyses identified two associated regions at the inhibitory immune receptor gene *SIGLEC5* to increase the risk for periodontitis. The GWAS also identified two associated regions at the long non-coding RNA *CTD-2353F22.1*. Additionally, a gene-gene interaction study identified gene-gene interaction between risk variants at *SIGLEC5* and at *CTD-2353F22.1*. The aims of the current thesis were the identification of the putative causal variants underlying these associations, characterization of their molecular biological effects and validation of *SIGLEC5* and *CTD-2353F22.1* as the target genes and to identify genes and genetic pathways regulated by *CTD-2353F22.1*. In addition, this thesis aims to identify the target genes of a miRNA that was differentially expressed in periodontitis patients compared to healthy controls. Employed methods included SNP mapping to DNA elements with predictive features of regulatory functions, screening for transcription factor (TF) binding sites, antibody electrophoretic mobility shift assays (EMSAs), luciferase reporter gene assays, CRISPR/dCas9 gene activation and RNA-sequencing to identify genetic interaction partners of *CTD-2353F22.1*. Genome-wide mRNA expression was quantified (Clarion Array) to identify the target gene of hsa-miR-374b-5p in primary gingival fibroblasts (pGF) and gene specific regulation was validated by qRT-PCR and Luciferase activity. In summary, EMSA in B lymphocytes showed that TF MAF bZIP (MAFB) binds at the common G-allele of rs4284742, whereas the minor A-allele reduced TF binding by 69%, corresponding to 9-fold reduction of luciferase reporter gene activity by the A-allele ($p = 0.01$). EMSA in peripheral blood mononuclear cells revealed that E-26 transformation-specific TF related gene (ERG) binds at rs11084095, with almost complete loss of binding at the minor A-allele. Allele-specific reporter genes showed enhancer function of the DNA-sequence at rs11084095, which was abrogated in the background of the A-allele. For the associations at *CTD-2353F22.1*, allele-specific binding of the TF PR Domain Zinc Finger Protein 14 (PRDM14) at rs1122900 was confirmed using antibody EMSA. Binding was reduced by 76% at the rare C-allele. Luciferase reporter gene assays revealed an allele-specific enhancer effect with a 5.5-fold increase of expression in the background of the common A-allele ($p = 0.0003$). Using CRISPR-dCas9, the enhancer at rs4284742 showed strong activation of *SIGLEC5* expression, validating this gene as the target gene of the association. The enhancer at rs11084095 and rs1122900 demonstrated moderately activated *SIGLEC5* and *CTD-2353F22.1* expression. RNA sequencing of *CTD-2353F22.1* CRISPRa cells indicated significant

upregulation of the gene sets “angiogenesis” and “TNFalpha signaling via NFκB”. *UHMK1* was the top one downregulated gene in pGFs ($p = 2.5x10^{-04}$, fold change (FC) = -1.8) in genome-wide expression analysis after hsa-miR-374b-5p overexpression. Using reporter gene assays with the cloned 3’UTR of *UHMK1*, the downregulation of mRNA level ($p = 0.02$; FC = -1.5) and reduction of luciferase protein activity ($p = 0.013$, FC = -1.3) was validated. In conclusion, rs4284742, rs11084095 and rs1122900 were identified as putative causal variants for the genome-wide significant associations with periodontitis that impair MAFB, ERG and PRDM14 binding, respectively and *SIGLEC5* and *CTD-2353F22.1* were validated as the target genes of the associations. hsa-miR-374b-5p was shown to regulate *UHMK1*, which has a role in osteoclast differentiation.

Data generated from this project can help to improve the understanding of the underlying disease biology, to point to a regulatory genetic pathway and hopefully leading to new treatment options.

Zusammenfassung

Parodontitis ist eine weit verbreitete komplexe Entzündungskrankheit der Mundhöhle. Charakteristisch für Parodontitis ist eine Entzündung des Zahnfleisches und ein alveolarer Knochenverlust. In einer genomweiten Assoziationsstudie (GWAS) und zwei GWAS Metaanalysen wurden kürzlich zwei assoziierte Regionen am hemmenden Immun-Rezeptorgen *SIGLEC5* identifiziert, die das Risiko für Parodontitis erhöhen. Die GWAS fand auch zwei assoziierte Regionen bei der langen nicht-codierenden RNA *CTD-2353F22.1*. Eine Gen-Gen Interaktionsstudie identifizierte zusätzlich eine Gen-Gen-Interaktion zwischen den Risiko-Varianten bei *SIGLEC5* und bei *CTD-2353F22.1*. Ziel dieser Dissertation war die Identifizierung der putativen kausalen Varianten, die diesen Assoziationen zugrunde liegen, die Charakterisierung ihrer molekularen biologischen Effekte sowie die Validierung von *SIGLEC5* und *CTD-2353F22.1* als Zielgene der Assoziationen und die Identifizierung von Genen mit denen *CTD-2353F22.1* interagiert. Zusätzlich hatte die vorliegende Arbeit das Ziel, die Zielgene einer miRNA zu identifizieren, von der gezeigt wurde, dass sie in Parodontitis Patienten signifikant stärker als in gesunden Kontrollprobanden exprimiert ist. Die im Rahmen dieser Studien eingesetzten Methoden umfassten das SNP Mapping bezogen auf DNA Elemente mit prädiktiven Eigenschaften für regulatorische Funktionen, das Screening nach Transkriptionsfaktor-Bindestellen (TF binding sites), EMSAs (Antibody Electrophoretic Mobility Shift Assays), Luciferase Reporter-Gen Assays, CRISPR dCas9 Genaktivierung und RNA-Sequenzierung um Gen-Interaktionspartner von *CTD-2353F22.1* zu bestimmen. Die genomweite mRNA Expression wurde in primären Gingiva-Fibroblasten (pGF) quantifiziert

(Clarion Array), um das Zielgen von hsa-miR-374b-5p zu identifizieren und die Gen-spezifische Regulation wurde mittels qRT-PCR und Luciferase-Aktivität validiert. Mit einem EMSA in B Lymphozyten konnte gezeigt werden, dass TF MAF bZIP (MAFB) an das häufige G-Allel von rs4284742 bindet, wohingegen das seltenere A-Allel die TF Bindung um 69% reduzierte, was einer 9-fachen Reduktion der Aktivität des Luciferase-Reportergens durch das A-Allel entspricht ($p = 0.01$). Hingegen konnte mittels EMSA in peripheren mononuklearen Blutzellen nachgewiesen werden, dass ERG (E-26 transformation-specific TF related gene) an rs11084095 bindet, bei gleichzeitig nahezu komplettem Verlust der Bindung an das A-Allel. Allel-spezifische Reportergene zeigten eine Enhancer-Funktion der DNA-Sequenz bei rs11084095, welche vor dem Hintergrund des A-Allels außer Kraft gesetzt wurde. Für die Assoziationen bei *CTD-2353F22.1* wurde die Allel-spezifische Bindung des TF PRDM14 (PR Domain Zinc Finger Protein 14) an rs1122900 mittels Antikörper-EMSA bestätigt. Die Bindung wurde in diesem Fall um 76% am seltenen C-Allel reduziert. Luciferase Reportergen-Assays wiesen einen Allel-spezifischen Enhancer-Effekt mit einer 5.5-fachen Zunahme der Expression vor dem Hintergrund des häufigen A-Allels ($p = 0.0003$) auf. Mittels CRISPR-dCas9 konnte gezeigt werden, dass der Enhancer bei rs4284742 eine deutliche Aktivierung der *SIGLEC5* Expression bewirkt, was belegt, dass dieses Gen das Zielgen der Assoziierung ist. Der Enhancer bei rs11084095 und rs1122900 bewirkte eine moderate Aktivierung der *SIGLEC5* und der *CTD-2353F22.1* Expression. Mittels RNA-Sequenzierung von *CTD-2353F22.1* CRISPR aktivierter Zellen konnte eine signifikante Hochregulierung der Gensets "Angiogenese" und "TNFalpha Signalweg via NFκB" festgestellt werden. *UHMK1* erwies sich als das am deutlichsten herunterregulierte Gen in pGFs ($p = 2.5x10^{-04}$, fold change (FC) = -1.8) in einer genomweiten Expressionsanalyse nach hsa-miR-374b-5p Überexpression. Unter Verwendung von Reportergen-Assays mit der klonierten 3'UTR von *UHMK1*, wurde die Herunterregulierung des mRNA-Levels ($p = 0.02$; FC = -1.5) und die Reduktion der Luciferase Proteinaktivität ($p = 0.013$, FC = -1.3) validiert. Zusammenfassend wurden rs4284742, rs11084095 und rs1122900 als putative kausale Varianten für die genomweiten signifikanten Assoziationen in Bezug auf Parodontitis identifiziert, die die Bindung von MAFB, ERG beziehungsweise PRDM14 vermindern. *SIGLEC5* und *CTD-2353F22.1* konnten als Zielgene dieser Assoziationen validiert werden. Zusätzlich wurde gezeigt, dass hsa-miR-374b-5p *UHMK1* reguliert, welches für die Differenzierung von Osteoklasten relevant ist. Die aus dieser Dissertation gewonnen Daten könnten das Verständnis der zugrundeliegenden Krankheitsbiologie verbessern, auf einen genetischen Regulationsweg hinweisen und hoffentlich zu neuen Behandlungsmöglichkeiten führen.

Table of Contents

List of Figures

List of Tables

0 Abbreviations

3C	chromatin conformation capture
AgP	aggressive periodontitis
AHR	aryl hydrocarbon receptor
ALL	acute lymphoblastic lymphoma
Amp.	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
AUC	area under curve
BACH2	BTB domain and CNC homolog 2
bp	basepairs
BSA	bovine serum albumin
CAL	clinical attachment loss
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
cm	centimeter
CP	chronic periodontitis
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CRISPRa	CRISPR activation
crRNA	CRISPR RNA
DAP12	Transmembrane immune signaling adaptor TYROBP
dCas9	dead Cas9
DHS	DNase I hypersensitive site
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphate
DAMPs	danger-associated molecular patterns
DPBS	Dulbecco‘s phosphate-buffered saline
DMEM	Dulbecco’s Modified Eagle Medium
EBV	Epstein-Barr virus
ECM	extracellular matrix

EDTA	ethylenediaminetetraacetic acid
e.g.	*exempli gratia* (for example)
EMSA	electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA Elements
eQTL	expression quantitative trait loci
ERG	E-26 transformation-specific TF related gene
EtBr	ethidiumbromide
EtOH	ethanol
FBS	fetal bovine serum
FC	fold change
fmol	femtomol
fw	forward
g	gram
GADD45A	Growth arrest and DNA damage inducible alpha
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GATA1	GATA binding protein 1
GCF	gingival crevicular fluid
GFP	green fluorescent protein
GO	Gene Ontology
GLIDE	Gene-Lifestyle interactions dental endpoints
gRNA	guide RNA
GTEx	Genotype-Tissue Expression
GWAS	genome-wide association study
h	hours
HRP	horseradish peroxidase
i.a.	*inter alia* (among other things)
ICAM-1	intracellular adhesion molecule
i.e.	*id est* (that is)
ITIM	immunoreceptor tyrosine-based inhibitory motif
kb	kilobases
LD	linkage disequilibrium

lncRNA	long non-coding RNA
LPS	lipopolysaccharide
LED	light-emitting diode
LPL	lipoprotein lipase
mA	milli ampere
MAF	minor allele frequency
MAFB	MAF bZIP
mg	milligram
µm	micrometer
miRNA	microRNA
min	minutes
mJ	millijoule
mL	milliliters
mM	millimolar
MMPs	matrix metalloproteinases
MPRA	massive parallel reporter assays
mRNA	messenger RNA
miRNA	microRNA
NaCl	sodium chloride
NaOH	sodium hydroxide
ncRNA	non-coding RNA
NEA	non-essential amino acids
NFATc1	nuclear factor of activated T cells 1
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
Ng	nanogram
NLR	Nod-like receptor
nm	nanometer
NOTCH1	Notch receptor 1
nt	nucleotides
OR	odds ratio
OSCAR	osteoclast-associated receptor

P_{adj}	adjusted p-value
PAM	protospacer adjacent motif
PAMPs	pathogen-associated molecular patterns
PBMC	Peripheral blood mononuclear cell
PCR	polymerase chain reaction
PD	periodontitis
pen/strep	penicillin/streptomycin
pGF	primary gingival fibroblast cells
pmol	picomole
PNK	Polynucleotide Kinase
PTK9	Protein Tyrosine Kinase 9
PRDM14	TF PR Domain Zinc finger protein
PRRs	pattern recognition receptors
PWM	position weight matrix
qRT-PCR	quantitative real-time PCR
RACE	Rapid Amplification of cDNA Ends
RANKL	receptor activator of NF-κB ligand
Rev	reverse
RNA	ribonucleic acid
RNP	ribonucleoprotein complex
rpm	rounds per minute
RT	room temperature
SAM	Synergistic Activation Mediator
SDS	sodium dodecyl sulfate
SERPINE1	endothelial plasminogen activator inhibitor
sgRNA	single guide RNA
sec	seconds
SHP-1	Protein tyrosine phosphatase non-receptor type 6
SHP-2	Protein tyrosine phosphatase non-receptor type 11
SIGLEC5	Sialic acid Binding Ig like lektin 5
SIGLEC14	Sialic acid Binding Ig like lektin 14

SNP	single nucleotide polymorphism
TBE	Tris borate EDTA
TEMED	Tetramethylethylenediamine
TF	transcription factor
TFBS	transcription factor binding site
TLR	Toll-like receptor
TNF	Tumor necrosis factor
tracrRNA	trans-activating crRNA
Tris	Tris(hydroxymethyl)aminomethane
TSS	transcription start site
U	units
UHMK1	U2AF Homology motif kinase 1
UK Biobank	United Kingdom Biobank
UPW	ultrapure water (RNase- and DNase-free)
UTR	untranslated region
UV	ultraviolet
V	volt
VEGF	vascular endothelial growth factor
VPR	VP64-p65-Rta
vWF	von Willebrand factor
YT-medium	medium with yeast extract and tryptone

1 Introduction

The overall scientific objective of this thesis was to identify putative causal variants of periodontitis associated haplotype blocks and to identify a microRNA that is significantly upregulated in blood and cells of the oral mucosa upon exposure to the periodontitis main risk factors oral inflammation and tobacco smoke. This thesis targeted to identify three putative causal variants of the association with periodontitis by demonstrating allele-specific enhancer function of the DNA-sequence, identification and validation of transcription factors that bind allele-specific to the associated variants and validation of the predicted target genes *SIGLEC5* and *CTD-2353F22.1*. Based on the results, a functional context of the associations with impaired regulation of endothelial homeostasis and healing of aseptic tissue injuries will be suggested. In addition, this thesis will show a microRNA (hsa-miR-374b-5p) to be upregulated in blood and gingiva of periodontitis patients and identifying *UHMK1* as the target gene of its regulation.

1.1 General aspects of periodontitis

Periodontitis (PD) is a common complex disease of the periodontium and a major cause of alveolar bone loss in adults above 40 years (Eke et al. 2012; Nesse et al. 2008). It is one of the sixth most common chronic inflammatory diseases worldwide with a prevalence of 11% (Kassebaum et al. 2014; Marcenes et al. 2013; Tonetti et al. 2017). Its main symptoms are inflammation of the oral mucosa and the tooth-bearing tissue leading to gingival bleeding, loss of connective tissue and formation of gingival pockets, alveolar bone loss and subsequently tooth loss. A strong risk factor for PD is long-term gingivitis which is characterized by swollen and bleeding gums. Bleeding on probing for 25 years or more was shown to result in an almost 50% increased tooth loss (Lang et al. 2009). The causes of PD and its precise molecular mechanisms in the pathogenesis of PD are not yet fully understood. The healthy oral cavity is inhabited by more than 700 bacteria species and microorganisms such as *Entamoeba ginigvalis* (Bao et al. 2020) which establish a complex local biofilm (Dewhirst et al. 2010; Verma et al. 2018). It is assumed that oral inflammatory diseases such as PD result from a dysbiosis in this biofilm (**Figure 1**). However, it is not the biofilm presence that causes the destruction of the periodontium, but rather a shift from a biofilm that is accepted by the immune system to a biofilm that causes an excessive immune response against it (Hajishengallis 2014; Kinane et al. 2017).

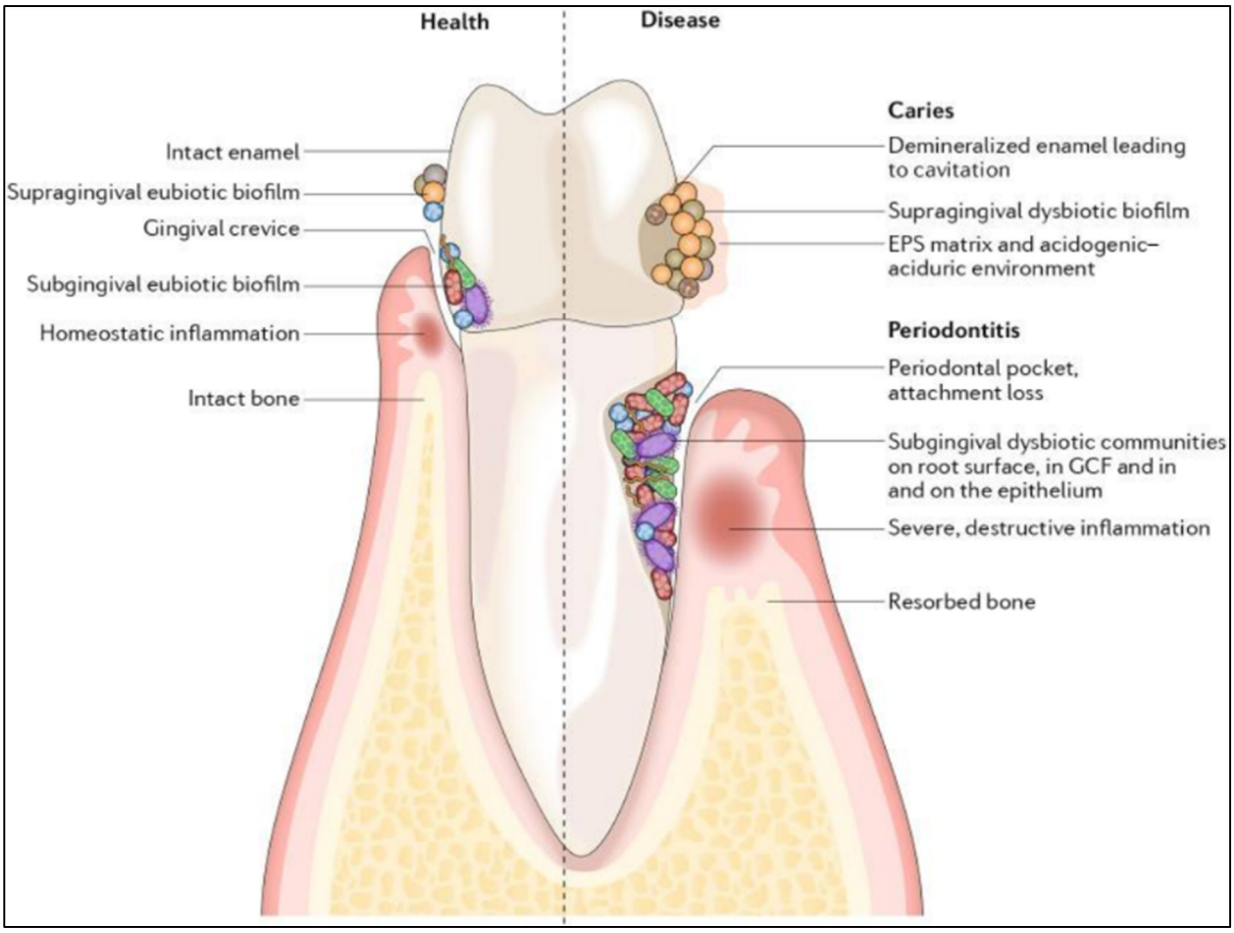


Figure 1: Schematic of healthy oral cavity and PD. On the left side, healthy gingival tissue and alveolar bone are shown. On the right side, inflammation of the gingival tissue, destruction of alveolar bone caused by a pathogenic dysbiotic biofilm (PD), is illustrated (taken from (Lamont et al. 2018)).

A subgingival growth of the biofilm is commonly prevented by the presence of an innate immune response at the intact epithelial cell barrier level. The onset of this response leads to an increase in salivary flow and sulcus fluid, which can mechanically detach bacterial colonies and also contain antibacterial substances (Darveau et al. 1997; Kinane 2001). In case bacteria, adhere to the subgingival surface and oral cavity, the innate immune response ensures recognition of the bacteria by host cells, mediated by the Toll-like receptors (TLRs) (Di Benedetto et al. 2013). TLRs belong to the pattern recognition receptors (PRRs) which are expressed by leukocytes and resident cells. TLRs can recognize pathogen-associated molecular patterns (PAMPs) including *e.g.* bacterial DNA, lipoproteins and lipopolysaccharide (LPS).

Once TLRs have recognized these PAMPs, they initiate the activation of several transcription factors. Activation of these transcription factors results in the release of pro-inflammatory cytokines, chemokines, proteases and prostaglandins, and other inflammatory mediators and the recruitment of phagocytes and lymphocytes (Di Benedetto et al. 2013; Trindade et al. 2014).

Although these inflammatory cytokines, prostaglandins and proteases act against microbial inflammation, they also promote connective tissue loss and alveolar bone resorption, leading to several histopathological changes, namely deepening of periodontal pockets, bone loss and eventually tooth loss.

Not only the host-specific immune response, but also the development of PD, the severity and the course of the disease are influenced by risk factors. These risk factors include not only external circumstances such as oral hygiene, smoking (Freitag-Wolf et al. 2019) and diet, but also periodontitis-associated risk genes (Genco and Borgnakke 2013; Schaefer 2018).

1.1.1 Classification of periodontitis

Until 2018, PD used to be classified into chronic PD (CP) and aggressive PD (AgP) (Armitage 1999). Whereas CP progresses slowly, occurs mainly in people over 35 years of age and long-standing external factors such as smoking are thought to play an important role in its etiology, AgP is characterized by rapid disease progression and occurs in young people under 35 years of age (Roshna and Nandakumar 2012). Segregation analyses showed that genetic causes play an important role in AgP.

The new classification from 2018 is a stage (stage I-IV)- and grade (grade A-C)-based system that is based on the severity, the clinical attachment loss (CAL) and the evidence of rapid progression and effects on systemic health (Tonetti et al. 2018).

However, as most of the publications that this thesis is based on were published prior to 2018, the old classification will be used in this thesis.

1.1.2 Heritability of PD

Heritability describes the hereditary proportion of a phenotypic variation that is attributable to genetic variation. To determine the heritability of AgP and CP, segregation analyses are used. A segregation analysis is a method of formal genetic analysis used to determine how frequently a heritable trait is present in a family tree. This method also allows for the determination of inheritance, i.e., whether the genetic factor has a dominant or recessive effect on the phenotype. The largest AgP family study (segregation analysis) included 527 subjects, 227 of whom had AgP, and showed autosomal dominant inheritance with 70 % inheritance of the original genetic

factors (Marazita et al. 1994). Both, this study and another from 2009 with 475 subjects (de Carvalho et al. 2009) show that genetic factors may have an influence on the inheritance of AgP. However, results of segregation analyses should be interpreted with caution, as they may also reflect external factors such as a shared lifestyle (oral hygiene, diet, and smoking). Therefore, a preferred alternative method is to perform a segregation analysis with identical twins, as they are genetically identical, and the influence of external factors can be investigated as well. The so-called agreement test is used to indicate the probability of both twins showing the disease if one of them already has the disease. In a study of monozygotic and dizygotic twins, the match rate on this test was determined for an increased risk of developing AgP. The agreement rate for risk of developing AgP was twice as high in monozygotic twins as in dizygotic twins (Corey et al. 1993). Twin studies of adult twins also demonstrated a genetic influence on the onset of the disease for CP with a heritability of 38% to 82% (110 pairs of adult twins) and 50% (117 pairs of adult twins) (Michalowicz 1994; Michalowicz et al. 1991; Michalowicz et al. 2000).

1.2 The role of genetic variation in complex diseases

Complex diseases such as PD are caused by environmental and genetic factors. Diverse genetic risk variants can cause the development of a complex disease whereby the effect size of each variant often has only a minor influence on the pathogenesis. The individual risk variants also occur in healthy individuals and only contribute to the disease risk through a specific combination of several risk alleles in conjunction with environmental and lifestyle factors like smoking and age (Page et al. 2003). Identification of genetic variants that contribute to disease susceptibility (Timpson et al. 2018; Yong et al. 2020) can improve the understanding of the pathogenic mechanisms underlying diseases like PD.

1.2.1 Single nucleotide polymorphism

The most common form of human genetic variation are single nucleotide polymorphisms (SNPs) which are point mutations of individual base pairs (bp) in the DNA sequence. These are individual positions in the genome at which alternative nucleotides (alleles) can occur in individuals of a population. With a minor allele frequency (MAF) > 0.01, SNPs are denoted as frequent SNPs, whereas SNPs with a MAF < 0.01 are referred as rare (Karki et al. 2015). Although SNPs can in principle be bi-, tri-, or even tetra-allelic, biallelic SNPs are predominant

in humans making them well-suited for genotyping (Brookes 1999). SNPs are distributed over the entire human genome, with an average of 1 SNP per 300 base pairs (Feuk et al. 2006; Genomes Project et al. 2015). In total, there are about 4.1 million to 5 million SNPs, most of which do not trigger any effect on genetic diseases (Genomes Project et al. 2015). In general, it is assumed that individual SNPs can be disease-associated in very specific combinations with each other and are thought to influence these disease manifestations of complex diseases like PD along with environmental factors (Boyle et al. 2017; Wray et al. 2018).

A distinction of genetic diseases is made between complex and monogenic diseases. In monogenic diseases, the disease onset is caused by the occurrence of the corresponding risk allele, whereas in complex diseases, the disease onset is caused by multiple risk alleles in combination with environmental influences (Risch 2000). The individual risk variants often occur in healthy individuals and only lead to the onset of the disease through a specific combination of several risk alleles (Page et al. 2003). SNPs are classified into different categories depending on whether they interfere in a direct way with the pathogenesis of a genetic disease or are only associated with that disease. When the disease onset is directly affected by the corresponding allele of a SNP, it is called a causal SNP. Causal SNPs can be located in non-protein-coding regions, protein-coding regions, and regulatory regions such as promoters. Non-causal disease associated SNPs have no direct influence on the phenotype but are in strong linkage disequilibrium (LD) with the actual causal variant. LD is the statistical association between alleles that occur more often together than can be accounted for by chance between two or more loci (Slatkin 2008). Chromosomal regions of SNPs in strong LD are named haplotype blocks. These haplotype blocks are separated from each other by recombination hot spots (Wall and Pritchard 2003).

Furthermore, a distinction is made as to whether the SNP alters the resulting protein sequence. SNPs located in the exons of a gene may result in altered codon information due to base swapping. If the triplet still codes for the same amino acid and thus there is no altered protein sequence, the SNPs are defined as synonymous SNPs (Brieuc and Naish 2011). Non-synonymous SNPs lead to amino acid substitution and thus altered protein function, which in turn may have physiological effects. SNPs in regulatory regions can affect the concentration of the corresponding gene product, and SNPs in splice sites (i.e., a non-protein-coding region) can lead to differentially spliced mRNAs, thereby increasing the risk for a disease-specific phenotype (Valentonyte et al. 2005).

1.2.2 Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) are used to identify common genetic variants (SNPs) associated with a particular disease. For this purpose, DNA of patients who are suffering from the disease and exhibiting the corresponding phenotype and DNA of a control group is analyzed for genetic markers (SNPs) representing the entire variability of the human genome by determining the respective allele of a dinucleotide. If a genotype (SNP allele) occurs significantly more frequently in the group of patients than in the controls, this SNP is associated with the disease. A GWAS meta-analysis combines the results of multiple studies for analysis. However, only representative SNPs (tagging SNPs) of a haplotype block are tested for causing the association (Visscher et al. 2012) which means that all other SNPs in the same haplotype block and in strong LD to the tagging SNP can be also associated with the disease (Gabriel et al. 2002). The identification of the causal variant poses a challenge because the most significant associated variant, called sentinel variant or GWAS lead SNP, most often is not identical with the functional variant(s) that caused the disease-association. Consequently, genetic associations need to be leveraged to biological meaning to identify causal variants.

1.2.3 Expression quantitative trait loci (eQTL)

Expression quantitative trait loci (eQTL) analysis aims to identify genetic variants that affect variation in gene expression of one or more genes (Nica and Dermitzakis 2013). Whereas GWASs identify SNPs associated with a particular disease, eQTL analysis assigns SNPs that are associated with gene expression, making it a useful tool to identify candidate susceptibility genes. A strong candidate susceptibility gene represents a gene for which reported eQTL effects of a risk variant are given in a relevant tissue or cell type (Lawrenson et al. 2015). Usually, associated SNPs are located in non-coding regions (Rockman and Kruglyak 2006) in regulatory elements such as enhancers (**Figure 2**) or promoters, indicating that these variants influence the expression of closely by (cis) or remote genes (trans) (Bryois et al. 2014). **Figure 2** shows how a SNP can influence gene expression in an enhancer. A SNP changes the activator binding site for an activator protein which results in altered gene expression of gene A.

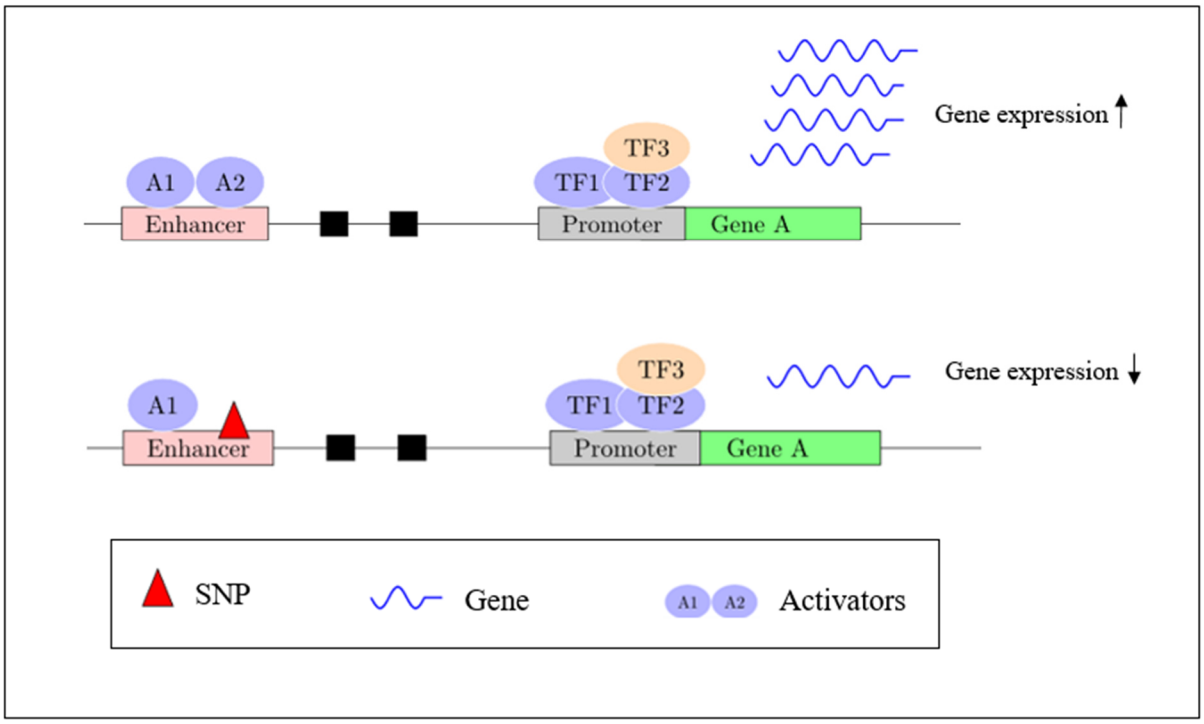


Figure 2: Schematic of regulation for specific gene expression by non-coding DNA elements. The upper panel shows that activator proteins bind to an enhancer element and activates gene expression of a distant gene A. The lower panel shows altered gene expression caused by a SNP (red triangle) that changes the activator binding site (modified after (Acharya et al. 2016)).

1.3 Elements of genetic regulation and predictive features for the identification of putative causal variants

Most of the identified disease-associated variants are located in non-protein-coding regions, suggesting that these non-coding variants are likely to be involved in gene regulation by affecting gene expression through effects on splicing, mRNA stability and transcription (Gallagher and Chen-Plotkin 2018). The identification of causal variants that affect gene expression is a major challenge because of many co-inherited variants in strong LD with the GWAS lead SNP, comprising a haplotype block.

Several studies have shown that causal SNPs are mostly located within features relating to the histone modifications H3K27Ac, H3K4me1 and H3K4me3, transcription factor binding sites (TFBSs) and chromatin accessibility (DNase I hypersensitive sites (DHSs)) (Kreimer et al. 2017; Kwasnieski et al. 2014). These DHSs are functionally related to transcriptional activity, since TFBS are mostly located in DHSs (Maurano et al. 2012). Genetic variants (SNPs) located in these regions showed differences in TF binding between individuals (Kasowski et al. 2010). In conclusion, mapping genetic variants to DHS is an accurate method to identify functional SNPs (Meuleman et al. 2020).

Transcriptional regulation is mostly mediated by binding of transcription factors (TFs) to TFBSs that are specific regions on the DNA. The transcription of a target gene is mostly controlled by different combinations of TFs acting together and only rarely does a single TF binding to DNA regulate the transcription of a target gene by its own (Narlikar and Ovcharenko 2009). TFBSs are located within enhancer regions or near the transcriptional start site (TSS) in (core promoters) or close by (proximal promoters) promoters of genomic DNA and can either repress or activate gene expression of the target gene. Enhancers are regulatory regions located within an intron or up- or downstream of the target gene (Blackwood and Kadonaga 1998). One theory for enhancer activity is the looping theory that states that once TFs bind the enhancer, the TFs (activators) are brought closer to the promoter by chromatin looping (**Figure 3**).

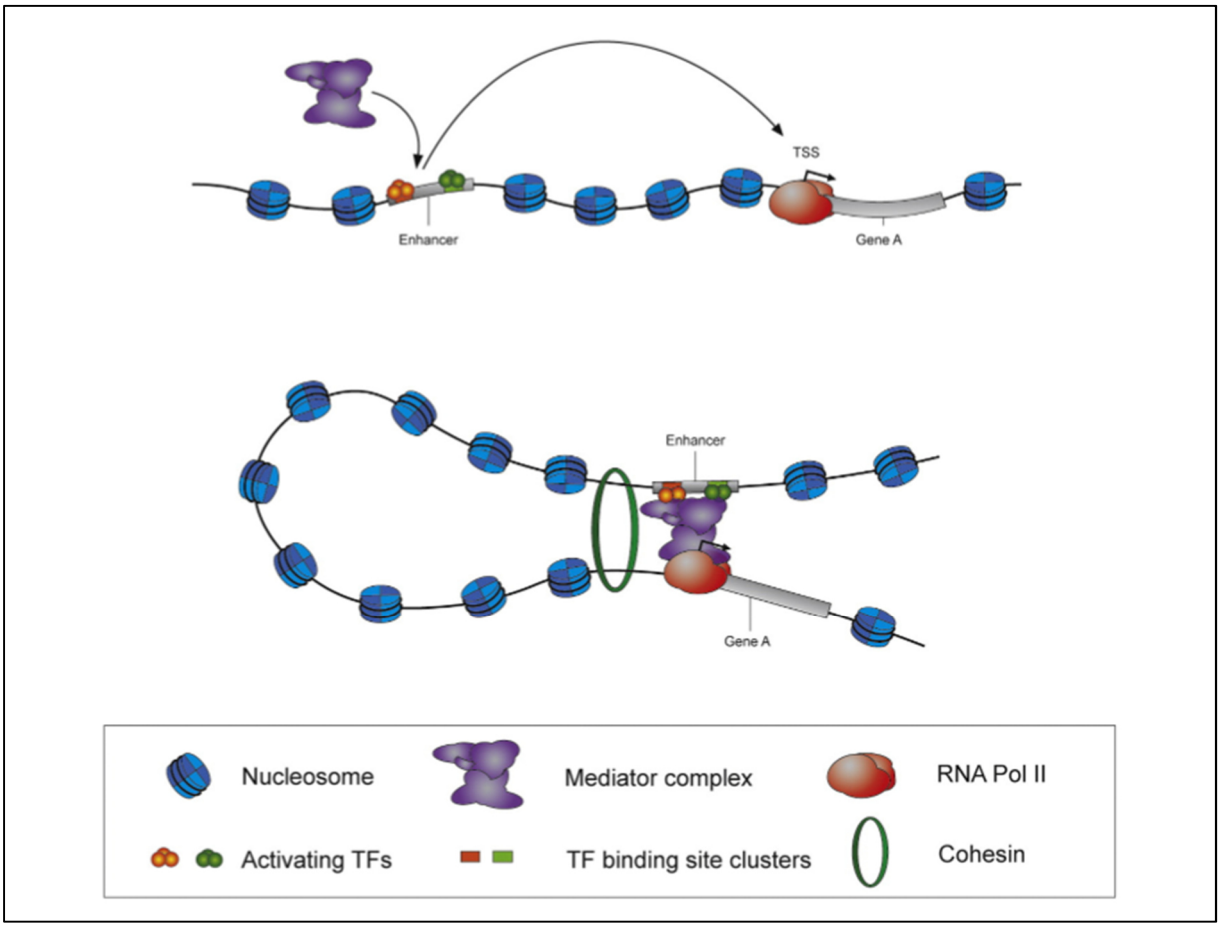


Figure 3: A schematic for enhancer regulation of gene expression. Binding of TF to DNA sequences located within enhancers stimulates chromatin looping. Cohesin stabilizes the enhancer-promoter interaction (taken from (Marand et al. 2017)).

Putative TFBSs are analyzed by scanning a specific DNA sequence of interest against position weight matrices (PWMs) of different TF (Jayaram et al. 2016). SNPs that are in strong LD to the GWAS lead SNP and are located in a TFBS are more likely to play a biological role than SNPs in the same associated region for which there is no evidence of overlap with a TFBS or other regulatory elements (Schaub et al. 2012).

ENCODE, the Encyclopedia of DNA Elements, is a project to identify all functional elements in the human genome such as TFBS, histone modifications and DHS, making it useful for the functional interpretation of disease associated SNPs (Consortium 2012; Schaub et al. 2012). However, because these predictive features do not provide functional or quantitative evidence of the regulatory activity, downstream experiments are required.

1.3.1 Non-coding sequences in the human genome

As described above, most of the identified disease-associated variants are located in non-coding regions of the human genome. Sequencing of the human genome showed that it contains about 20,000 - 25,000 protein-coding genes. This accounts for less than 2% of the total genomic sequence (International Human Genome Sequencing 2004). The nematode *Caenorhabditis elegans* has a similar number of protein-coding genes as humans (Wilusz et al. 2009). The organisms differ only in the number of non-protein-coding intronic and intergenic sequences in the genome, which increases with the complexity of the organisms. This fact suggests a specific function of the transcribed but not translated regions, which still account for over 90% of the human genome (Consortium et al. 2007). There are many different types of non-coding DNA sequences including non-coding RNA genes like long non-coding RNAs (lncRNA) and microRNAs (miRNAs), regulatory elements like promoters and enhancers, introns, pseudogenes and transposons (Palazzo and Gregory 2014). The Encyclopedia of DNA Elements (ENCODE), mentioned above, identified that 95% of the human genome is located within 8 kilobases (kb) of a protein-DNA interaction and that at least 80% of the human genome have at least one biochemical function in at least one cell type (ENCODE-Project-Consortium 2012).

Long non-coding RNAs (lncRNAs) belong to the non-coding sequences in the human genome. They have a length of more than 200 nucleotides (nt) and are not protein-coding (Kapranov et al. 2007). Since their discovery, it has become clear that lncRNAs regulate and influence fundamental biological processes such as histone modifications, gene activation or gene

repression at the transcriptional and posttranscriptional level (Gong and Maquat 2011; Wilusz et al. 2009) and are modifying chromatin states by recruiting epigenetic factors (Rinn and Chang 2012). LncRNAs were also shown to be involved in the pathogenesis of many diseases (Wapinski and Chang 2011) such as PD (Schaefer et al. 2011).

1.3.1.2 Identification of miRNAs that are upregulated in blood and gingiva of PD patients

MicroRNAs (miRNAs) are small non-coding RNAs, ranging in length from 18-25 nt, and primarily negatively regulate post-transcriptional gene expression (Lu and Rothenberg 2018). MicroRNAs recognize their targets mainly with a small region in the 5'-end (seed region) that interact with the complementary sequences in the 3' untranslated regions (3'UTR) of the target mRNA (Nilsen 2007). Through mRNA degradation or, more rarely, by translational repression, gene expression of the target gene is suppressed. The miRNAs function in association with Argonaute proteins which are guided together with associated factors by the miRNAs to the target mRNA. As a consequence, translation is inhibited or the mRNA is degraded (Dueck et al. 2012). MicroRNAs have been demonstrated to play a major role in development processes including proliferation, metabolism and apoptosis (Mattick and Makunin 2005). Giving evidence of their importance, miRNAs are predicted to control around 50% of all protein-coding genes in the human genome (Krol et al. 2010). A single mRNA can be modulated by various miRNAs and an individual miRNA can downregulate multiple mRNA targets (Lim et al. 2005). In recent years, dysregulation of miRNAs have been implicated in many human diseases (Paul et al. 2018) such as PD (Lee et al. 2011; Xie et al. 2011). Identification of differentially expressed miRNAs and their associated target genes in cases of disease compared to controls gives the opportunity to elucidate the molecular mechanisms involved in disease pathogenesis.

Differential miRNA expression in healthy and inflamed oral tissues was reported by various studies generated by array-based profiles for oral mucosa (Amaral et al. 2019; Lee et al. 2011; Stoecklin-Wasmer et al. 2012; Xie et al. 2011), saliva (Fujimori et al. 2019; Nisha et al. 2019), gingival crevicular fluid (Saito et al. 2017), and peripheral blood (Keller et al. 2011). However, interpretation of these results is difficult because differentially expressed miRNAs from different studies showed little overlap due to variable isolation and quantification methods, as well as environmental and genetic background of the respective donors. In addition, there are

differences in the analyzed tissues, tissue collection sites or collection methods, which influence the distribution of cell types that also contribute differently to the respective miRNA profiles.

To identify such miRNAs that are differentially expressed between PD cases and healthy controls, miRNAs were quantified using a miRNA expression array to identify characteristic blood-borne miRNA patterns of the complex disease PD in previous work of our research group. As some miRNAs are also secreted into venous blood (Zhu et al. 2017), the venous blood of 18 untreated PD cases with an age range from 20-55 of Dutch and German background and 70 German controls with a mean age of 44 years was collected (Mueller et al. 2021). The study population has been described in detail in a previous publication (Keller et al. 2011). In this previous work, total miRNAs were isolated using the miRNeasy Mini kit and were quantified on the Geniom Biochip array "miRNA Homo sapiens" and analyzed using the microarray-based screening approach described in (Keller et al. 2009a) (Keller et al. 2009b).

1.4 *SIGLEC5* as a risk gene of periodontitis

In 2017, a GWAS on periodontal disease identified rs4284742, located at Sialic acid binding IG like lectin 5 (*SIGLEC5*), as a risk variant of PD (Munz et al. 2017). This GWAS included samples from the severe, early-onset form AgP and controls from Germany and the Netherlands (N cases = 896, N controls = 7104), samples of the more moderate phenotype CP and controls from Germany (N cases = 993, N controls = 1419) and samples from AgP cases and controls from Turkey (N cases = 223, N controls = 564) (**Figure 4**).

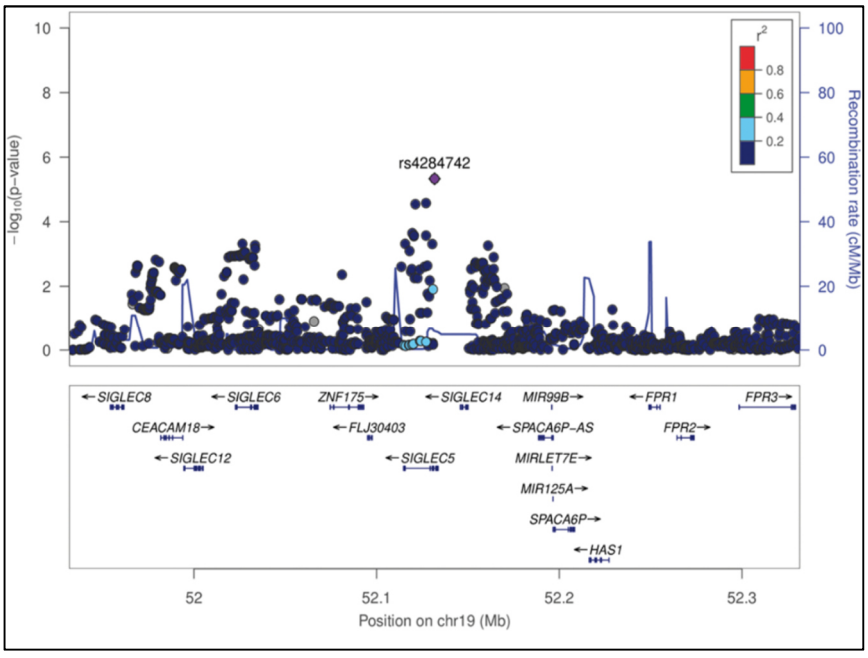


Figure 4: Regional association plot of rs4284742 (taken from (Munz et al. 2017))

The association at *SIGLEC5* with PD had a significance level of $p = 1.34 \times 10^{-8}$ (odds ratio (OR) = 1.34), whereby the p-value was calculated without the Turkish controls because of poor genotype quality for rs4284742. **Figure 4** is a regional association plot of rs4284742 and shows that there is no variant in strong LD ($r^2 > 0.8$) to rs4284742.

In addition, another GWAS meta-analysis on PD with AgP and CP cases (N cases = 5095, N controls = 9908) of North-West Europeans identified also an association located at *SIGLEC5* (rs11084095) with a significance level of $p = 5.09 \times 10^{-8}$ (OR = 1.17) (**Figure 5**) (Munz et al. 2019).

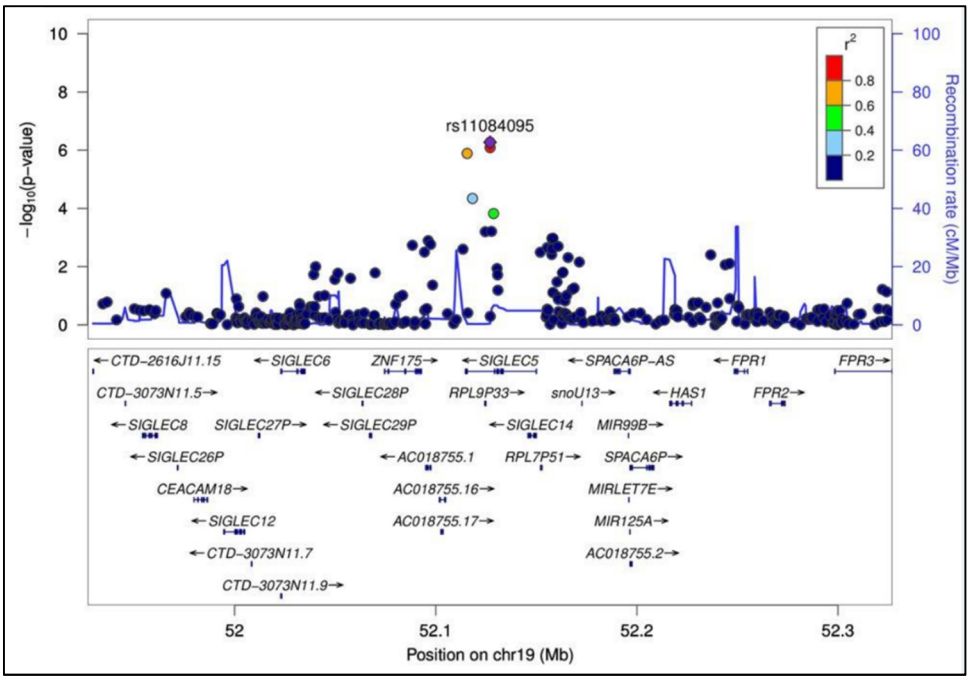


Figure 5: Regional association plot of rs11084095 (taken from (Munz et al. 2017)).

Interestingly, also the largest GWAS meta-analysis for PD performed to date, which includes PD cases from seven studies (N cases = 17.353, N controls = 28.210) from both the UK Biobank, which mainly contains self-reported PD status, and the Gene-Lifestyle Interactions in Dental Endpoints (GLIDE) consortium, also identified an association with the intronic SNP rs12461706 located to the gene *SIGLEC5* with a genome wide significance threshold of $p = 3.9 \times 10^{-9}$ (OR = 1.05) (Shungin et al. 2019).

1.4.1 Sialic acid binding IG like lectin 5 (*SIGLEC5*)

Sialic acid binding IG like lectin 5 or also known as CD170 is a member of the subfamily CD33-related siglecs. *SIGLEC5* is mainly expressed in various myeloid cells of the innate immune system that regulate inflammation mediated by PAMPs and danger-associated molecular patterns (DAMPs) (Pillai et al. 2012) and is expressed in B-cells (Crocker and Varki 2001). It is an inhibitory transmembrane receptor that poses an N-terminal domain for binding sialic acids and sialic acid-containing glycan ligands and a domain, located in the N-terminal domain, that is a conserved arginine residue for binding carbohydrates (von Gunten and Bochner 2008). The sialic acids are expressed by certain pathogens and bacteria on their surfaces (Khatua et al. 2013). The cytoplasmic tyrosine residues of Siglec receptors (one or more depending on the Siglec receptor) are located within inhibitory signaling motifs named immunoreceptor tyrosine-based inhibitory motifs (ITIMs) (Zhuravleva et al. 2008). Binding of sialyated ligands with *SIGLEC5* leads to recruitment of the tyrosine phosphatases, SHP-1 and SHP-2, to its two intracellular ITIMs. This initiates an inhibitory intracellular signaling cascade (Nordstrom et al. 2011).

SIGLEC5 shares almost complete sequence identity in two terminal immunoglobulin domains with the activating receptor *SIGLEC14* that interacts with the transmembrane immune signaling adaptor (DAP12) (Cao and Crocker 2011). It is assumed that *SIGLEC14* and *SIGLEC5* are paired inhibitory and activating receptors (Angata et al. 2006).

The role of *SIGLEC5* in the etiology of periodontal diseases is currently unknown. However, host immune responses are modulated through binding of sialic acids to Siglecs as described above.

1.5 Gene-gene interaction studies

Thousands of GWASs identified many genetic risk variants for complex diseases and traits (Visscher et al. 2017). However, the effect size of these identified genetic variants is generally small (Visscher et al. 2017) and explains only a small proportion of the disease heritability, leading to the "missing heritability" problem (Manolio et al. 2009). While there are few explanations for this problem, genetic interaction between genes at different loci is one potential (Boyle et al. 2017; Eichler et al. 2010; Wray et al. 2018; Zuk et al. 2012). This genetic interaction, also called gene-gene interaction, is the effect of one gene related to a disease modified by another one or more other genes that are also associated with the disease (polygenic

diseases). This implies that polygenic diseases are caused by interacting effects of the genetic variants (Boyle et al. 2017; Wray et al. 2018). Identifying those gene-gene interactions can potentially help to identify genes involved in the pathogenesis of a complex disease like PD, as well as gene functions and pathways (Chattopadhyay and Lu 2019).

In previous work of our research group, a gene-gene interaction study (currently unpublished) with the identified risk variants of PD in the GWAS (Munz et al. 2017) and the GWAS meta-analysis study (Munz et al. 2019) was performed to identify potential genetic risk variants that interact with *SIGLEC5*. The most significant gene-gene interaction with p_{cases} = 0.0057 and $p_{controls}$ = 0.6478 was identified for rs11084095 at *SIGLEC5* and rs1122900 at *CTD-2353F22.1* (**Table 1**).

Table 1: Most significant gene-gene interaction with identified SNPs associated with PD from German and Dutch cases with $p < 0.05$ and controls with $p > 0.05$.

Gene 1	SNP 1	SNP 2	Gene 2	p-value cases	p-value controls
SIGLEC5	rs11084095	rs1122900	*CTD-2353F22.1*	0.0057	0.6487

1.6 *CTD-2353F22.1* as a risk gene of periodontitis

The GWAS on periodontal disease that identified rs4284742, located at *SIGLEC5*, also identified rs6887423 and rs1122900, located at the long non-coding RNA *CTD-2353F22.1*, as risk variants of PD (Munz et al. 2017). This GWAS included samples from the severe, early-onset form AgP and controls from Germany and the Netherlands (N cases = 896, N controls = 7104) and samples of the more moderate phenotype CP and controls from Germany (N cases = 993, N controls = 1419). SNP rs6887423 showed a stronger association with $p = 1.42 \times 10^{-6}$ (OR = 1.29) compare to variant rs1122900 with $p = 1.77 \times 10^{-5}$ (OR = 1.26). Including samples from AgP cases and controls from Turkey (N cases = 223, N controls = 564), rs112200 showed a stronger association with $p = 8.00 \times 10^{-7}$ (OR = 1.27), whereby there are no data for the pooled analysis for rs6887423. Interestingly, the gene-gene interaction study mentioned above performed in previous work by our research group (currently unpublished) identified the strongest gene-gene interaction between rs1122900 at *CTD-2353F22.1* and rs11084095 at *SIGLEC5*.

1.6.1 Expression and transcripts of CTD-2353F22.1

CTD-2353F22.1 is an uncharacterized antisense lncRNA. Antisense lncRNAs have been shown to regulate the expression of their target gene in *cis* or in *trans* (Pelechano and Steinmetz 2013; Tufarelli et al. 2003; Zhang et al. 2014b). *CTD-2353F22.1* is also known as AC008957.1 and is primarily expressed in EBV (Epstein-Barr virus)-transformed lymphocytes (Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)). There are currently three known transcripts of *CTD-2353F22.1*, *CTD-2353F22.1*-001, *CTD-2353F22.1*-002 and *CTD-2353F22.1*-003 (Ensembl Genome Browser).

1.7 Aims of this thesis

The aims of the current thesis were the identification of the putative causal variants of the GWAS associated haplotype blocks at *SIGLEC5* and at *CTD-2353F22.1*, and validation of *SIGLEC5* and *CTD-2353F22.1* as the target genes. In addition, this thesis aimed to identify a microRNA that is significantly upregulated in blood and cells of the oral mucosa upon exposure to the periodontitis main risk factors oral inflammation and tobacco smoke and to subsequently identify and validate its target gene.

More specifically, the objectives of this thesis were:

1. Identification and characterization of periodontitis associated-functional regulator(s) at *SIGLEC5* and *CTD-2353F22.1*
2. Identification of the putative causal variant(s) of the GWAS associated haplotype blocks
3. Validation of *SIGLEC5* and *CTD-2353F22.1* as the target genes of the associations
4. Identification and characterization of the genes and gene networks that respond to increased expression of *CTD-2353F22.1*
6. Identification of a microRNA that is significantly upregulated in blood and cells of the oral mucosa upon exposure to the periodontitis main risk factors oral inflammation and tobacco smoke
7. Identification and validation of the target gene of the microRNA

2 Materials

2.1 Chemicals, solutions and cell culture media

Table 2: Chemicals, solutions and cell culture media.

Chemicals, solutions and cell culture media	Manufacturer
Acrylamide-solution (30%)	SERVA Electrophoresis
Agar	AppliChem
Agarose	SERVA Electrophoresis
Ampicillin	Gibco by life technologies
Ammonium Persulfate (APS)	Amresco
β-mercaptoethanol	Carl Roth
Bacto Tryptone	BD Bacto
Bacto Yeast Extract	BD Bacto
Bovine serum albumin (BSA)	SERVA Electrophoresis
DermaLife K, fully supplemented	Cell Systems
Dispase II	Sigma Aldrich
Dulbecco's Modified Eagle Medium (DMEM) supplemented with L-Glutamine, 3.7 g/L $NaHCO_3$ und 1 g/L D-Glucose	PAN Biotech
Deoxynucleotides (dNTPs) Mix (100 mM)	Thermo Fisher Scientific
Dulbecco's phosphate-buffered saline (DPBS)	Gibco by life technologies
Earle's MEM supplemented with 0.85 g/L $NaHCO_3$, w/o L-Glutamine	PAN Biotech
Ethanol (70%, dehydrated)	Carl Roth
Ethanol (99.9%)	Merck
Ethidium bromide (EtBr)	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Fetal Bovine Serum (FBS)	Gibco by life technologies
GeneRuler 1 kb DNA Marker	Thermo Fisher Scientific
Glycerol	Carl Roth
GlycoBlue	Thermo Fisher Scientific
Histopaque 1077	Sigma Aldrich
Isopropanol	Sigma Aldrich
jetPEI-DNA transfection reagent	Polyplus-transfection
L-Glutamine	Biochrom
Lipofectamine RNAiMax	Thermo Fisher Scientific
Loading Dye 6x	Thermo Fisher Scientific
non-essential amino acids (NEA) (100 x)	PAN Biotech
Opti-Mem Reduced Serum Medium	Life Technologies
Oligo dTs	Thermo Fisher Scientific
O'Range Ruler 50 bp Ladder	Thermo Fisher Scientific
Penicillin/Streptomycin	Biochrom

RPMI-1640 (2mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 4500 mg/L glucose, 1500 mg/L sodium bicarbonate)	ATCC
SOC Medium	Invitrogen
Sodium acetate 3M	Sigma Aldrich
Sodium chloride 150 mM	Polyplus-transfection
Tetramethylethylendiamine (TEMED)	Sigma-Aldrich
Tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich
Trizol Reagent	Ambion
Trypan blue	Biochrom
Trypsin/EDTA	PAN Biotech
Ultra Pure Water	Qiagen

2.2 Enzymes and buffers

Table 3: Enzymes and buffers.

Enzymes and buffers	**Manufacturer**
AflII, 20,000 U/mL	New England Biolabs
Alkaline Phosphatase, Calf Intestinal (CIP), 10,000 U/mL	New England Biolabs
BbsI-HF, 20,000 U/mL	New England Biolabs
Cut Smart Buffer	New England Biolabs
DNase I recombinant, 10 U/µL	Roche
HF Buffer	Thermo Fisher Scientific
KpnI-HF, 20,000 U/mL	New England Biolabs
MultiScribe Reverse Transkriptase, 50 U/µL	Invitrogen
Phusion High-Fidelity Polymerase 2 U/µL	Thermo Fischer Scientific
Q5 Hot Start High-Fidelity Master Mix	New England Biolabs
Reaction Buffer 10 x	Biozym
RNaseOut Recombinant Ribonuclease Inhibitor, 40 U/µL	Invitrogen
RT-Buffer 10 x	Invitrogen
Taq DNA Polymerase, 5 U/µL	Biozym
T4 Ligase	New England Biolabs
T4 Ligase Buffer	New England Biolabs
T4 Polynucleotide Kinase	New England Biolabs
XbaI, 20,000 U/mL	New England Biolabs
XhoI, 20,000 U/mL	New England Biolabs

2.3 Reaction systems (Kits)

Table 4: Reaction systems (Kits).

Kit	Manufacturer
DC Protein Assay	Bio Rad
Dual Luciferase Reporter Assay	Promega
Gelshift Chemiluminescent EMSA Kit	Active Motif
Gibson Assembly Cloning Kit	New England Biolabs
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems
Neon Transfection System 10 µL Kit	Invitrogen
Nuclear Protein extraction Kit	Active Motif
Q5 Site-Directed Mutagenesis Kit	NEB
Qiagen Plasmid Maxi Kit	Qiagen
Qiagen Plasmid Midi Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
RNeasy Mini Kit	Qiagen
SYBR Select Master Mix	Thermo Fisher Scientific

2.4 Consumables

Table 5: Consumables.

Consumables	Manufacturer
1.5 mL and 2 mL microcentrifuge tubes	Eppendorf
6-Well plate (sterile)	Techno Plastic Products (TPP)
Biosphere Filtertips (10 µL, 20 µL, 100 µL, 200 µL, 1000 µL)	Sarstedt
Cell culture flasks (T75-T300)	TPP
Cell scraper (16 cm, 25 cm)	Sarstedt
Falcon tubes (15 mL, 50 mL)	Falcon
Hemacytometer (Neubauer cell counting chamber)	Brand
Microseal'B ' PCR plate sealing foil	Bio-Rad
Multiplate 96-well white PCR plates	Bio-Rad
Nunclon Delta-Treated 96-well plate	Thermo Fisher Scientific
Nylon membrane, positively charged	Sigma Aldrich
PCR strips and reaction tubes	Carl Roth
Petri dishes (plastic)	Sarstedt
Precision Wipes	Kimtech Science
Scalpel, sterile	Braun
Serological pipettes (5 mL, 10 mL, 25 mL)	Sarstedt
Syringe	Braun
Whatmann Paper	Bio-Rad

X-ray film	Thermo Fisher Scientific

2.5 Devices

Table 6: Devices.

Devices	**Manufacturer**
Battery-operated pipette controller	Brand
Benchtop centrifuge	Thermo Fisher Scientifc/Heraeus
CFX Connect Real-Time PCR Detection System	Bio-Rad
Chemiluminescence Imager ChemoStar Touch	Intas
Confocal microscope LSM 700	Zeiss
Electrophoresis Chamber for agarose gels	Biometra
Electrophoresis chamber	Bio-Rad
Freezer -80°C MDF-U72V	SANYO
Freezer -20 °C	Liebherr
Incubator for cell culture	Heraeus Instruments
Light microscope	Leitz
Liquid nitrogen tank Arpege 40	Air Liquide
Microwave	Bosch
Mini Trans-Blot Cell	Bio-Rad
Multifuge X1R Centrifuge	Thermo Fisher Scientific
Multiskan GO Spectrophotometer	Thermo Fisher Scientific
Neubauer counting chamber	Brand
Orion II Microplate Luminometer	Berthold
PCR FlexCycler	Analytik Jena
pH-Meter 766 Calimatic	Knick
Shaking incubator for bacterial culture	VWR
Standard Power Pack P26 Power Supplies	Biometra
Sterile bench	Thermo Fisher Scientific
Thermomixer	Biometra
UVLink 1000 Crosslinker	Analytik Jena
UV transilluminator (E-BOX VX5)	Vilber Lourmat
UV transparent gel trays	Biometra
Vortex Genie 2	Scientific Industries
Waterbath	VWR
Waterbath for cell culture	Julabo MWB

2.6 Software and databases

Table 7: Software and databases.

Software and databases	Manufacturer and/or Website
ApE- A plasmid Editor	Copyright M. Wayne Davis
CFX Manager Software 3.1 for qRT-PCR cycler	Bio-Rad
Chemiluminescence Imager software	Intas
Clone Manager 9, Professional, version 9.2	Sci Ed Software LLC.
CRISPR-ERA, version 1.2	http://crispr-era.stanford.edu/ (Liu et al. 2015)
ENCODE	https://www.encodeproject.org/
Ensembl genome browser 104	https://www.ensembl.org/index.htmL
GraphPad Prism 6, version 6.01	GraphPad
ImageJ, 1.48v	https://imagej.nih.gov/ij/index.htmL (Rueden et al. 2017)
JASPAR	http://jaspar.genereg.net/(Sandelin et al. 2004)
LDproxy Tool - LDlink	https://ldlink.nci.nih.gov/?tab=ldproxy (Machiela and Chanock 2015)
Multiskan GO Spectrophotometer Software	Thermo Fischer Scientific
NEB Tm Calculator, version 1.13.0	https://tmcalculator.neb.com/#!/main
NEBioCalculator, version 1.13.1	https://nebiocalculator.neb.com/#!/ligation
Partek Genomics Suite Software	Partek
Primer3web, version 4.1.0	https://primer3.ut.ee/
PWMtools	Swiss Institute of Bioinformatics https://ccg.epfl.ch/pwmtools/pwmbrowse.htmL
QTlizer	http://genehopper.de/qtlizer (Munz et al. 2020)
SNPInspector	Genomatix
TargetScanHuman 7.1	http://www.targetscan.org/vert_71/(Agarwal et al. 2015)
Transfac professional 2018.2	geneXplain (Wingender 2008)
UCSC Genome Browser	http://genome.ucsc.edu (Lee et al. 2020)

2.7 Antibodies

Table 8: Antibodies.

Antibody	Manufacturer
AHR (200 µg/mL)	Santa Cruz Biotechnology
BACH2 (533 µg/mL)	Proteintech
ERG-1/2/3 (200 µg/mL)	Santa Cruz Biotechnology
GATA1 (200 µg/mL)	Santa Cruz Biotechnology
MAFB (400 µg/mL)	Santa Cruz Biotechnology
PRDM14 (288 µg/mL)	LSBio

2.8 Plasmids

Table 9: Plasmids.

Plasmids	Manufacturer
pGL4.24	Promega
pRL-SV40	Promega
dCas9-VP64_GFP	Addgene (# 61422)
MS2-P65-HSF1_GFP	Addgene (# 61423)
sgRNA (MS2) cloning backbone	Addgene (# 61424)
SP-dCas9-VPR	Addgene (#63798)
gRNA_Cloning Vector	Addgene (#41824)

2.9 Oligonucleotides

Table 10: Oligonucleotides of EMSA probes. Oligonucleotide were ordered with and without Biotin 3' modification.

Designation	Forward (5'-3')	Reverse (5'-3')
rs4284742-(G)	CTGGTTCTTTCCACAGTCAC(C)A AGGACCACTCCATGCCCCTC	GAGGGGCATGGAGTGGTCCTT(G)GTGACTGTGGAAAGAACCAG
rs4284742-(A)	CTGGTTCTTTCCACAGTCAC(T)A AGGACCACTCCATGCCCCTC	GAGGGGCATGGAGTGGTCCTT(A)GTGACTGTGGAAAGAACCAG
rs11084095-(A)	CAATCTTTAGGTGATGCTAAA(A) GAAAGCTCGTGTGTGTTAGT	ACTAACACACACGAGCTTTC(T) TTTAGCATCACCTAAAGATTG
rs11084095-(G)	CAATCTTTAGGTGATGCTAAA(G) GAAAGCTCGTGTGTGTTAGT	ACTAACACACACGAGCTTTC(C) TTTAGCATCACCTAAAGATTG

rs34984145-(A)	TCCCAACTACTCGGAGGGCTG(A)GGCAGGAGAATGGCGTGAACC	GGTTCACGCCATTCTCCTGCC(T)CAGCCCTCCGAGTAGTTGGGA
rs34984145-(T)	TCCCAACTACTCGGAGGGCTG(T)GGCAGGAGAATGGCGTGAACC	GGTTCACGCCATTCTCCTGCC(A)CAGCCCTCCGAGTAGTTGGGA
rs1122900-(A)	TTAATTTCAATCGGTTTCTCT(A)ACAGATCTTACGACTAAAAAA	TTTTTTAGTCGTAAGATCTGT(T)AGAGAAACCGATTGAAATTAA
rs1122900-(C)	TTAATTTCAATCGGTTTCTCT(C)ACAGATCTTACGACTAAAAAA	TTTTTTAGTCGTAAGATCTGT(G)AGAGAAACCGATTGAAATTAA
rs6887423-(C)	TGGACCACCTACCTTCTCCA(C)GCAGCTGGACTCAGCAAGCT	AGCTTGCTGAGTCCAGCTGC(G)TGGAGAAGGTAGGTGGTCCA
rs6887423-(T)	TGGACCACCTACCTTCTCCA(T)GCAGCTGGACTCAGCAAGCT	AGCTTGCTGAGTCCAGCTGC(A)TGGAGAAGGTAGGTGGTCCA
rs56038114-(T)	AGCCTGGGCAACAGAGTGAGA(T)ACTGTTTCAAAAAAAAAAAAA	TTTTTTTTTTTTTGAAACAGT(A)TCTCACTCTGTTGCCCAGGCT
rs56038114-(C)	AGCCTGGGCAACAGAGTGAGA(C)ACTGTTTCAAAAAAAAAAAAA	TTTTTTTTTTTTTGAAACAGT(G)TCTCACTCTGTTGCCCAGGCT

Table 11: Oligonucleotides for cloning reporter gene assay plasmids. Forward-oligonucleotides for cloning genomic DNA sequences including rs4284742, rs11084095, rs34984145, 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. Forward- and reverse oligonucleotides for cloning 3' UTR of *UHMK1* include *XbaI* restriction sites.

Designation	**Forward (5'-3')**	**Reverse (5'-3')**
rs4284742	ggccGGTACCCTCAAAGCAGTGAACAGACTTT	ggccCTCGAGATGCAGGAGTGGAAGGGTG
rs11084095	ggccGGTACCTGAGTTGTTTCCATTTGAGCCG	ggccCTCGAGAAGATGCCCATTCACATGCC
rs34984145	ggccGGTACCGTGAAACCCCGTCTCTACTAAA	cgaCTCGAGGCGATCTCCTCTCACTGCAA
rs1122900	ggccGGTACCTGTGGGGAAAGTTTTGTGCA	cgaCTCGAGTGCCTTCACAAAAGTTAGGGTG
rs6887423	ggccGGTACCACGACTCCAGCAGAATTGAT	cgaCTCGAGGATCACGTGTTAGGCTGTGG
rs56038114	ggccGGTACCAGGAGGAAGAGGTTTCAGTGA	cgaCTCGAGGATGCCTGACTTGCCATGTA
3'UTR-*UHMK1*	AATTGGCCTCTAGATGCCTGGAAATAGCCTTGGT	AATTGGCCTCTAGATCTGCTGGAAAAGATAGGGC

Table 12: Oligonucleotides for SNP exchange (Q5) in reporter gene assay plasmids.

Designation	Forward (5'-3')	Reverse (5'-3')
rs4284742-Q5	CCACAGTCACtAAGGACCACT	AAAGAACCAGACCACAGG
rs11084095-Q5	ACGAGCTTTCtTTTAGCATCAC	GTGTGTTAGTACGGTGAG
rs34984145-Q5	CGGAGGGCTGtGGCAGGAGAA	AGTAGTTGGGACTACAGGCG
rs1122900-Q5	CGGTTTCTCTaACAGATCTTAC	ATTGAAATTAAGAATTTTACCCC
rs6887423-Q5	ACCTCTCCAtGCAGCTGGAC	AGGTGGTCCAGGAGAGGA
rs56038114-Q5	CAGAGTGAGAcACTGTTTCAAAA AAAAAAAAAATCTCATTTACG	TTGCCCAGGCTGGAGTGC

Table 13: Oligonucleotides for cloning sgRNAs for CRISPR activation using VPR-system.

Designation	Forward (5'-3')	Reverse (5'-3')
gRNA-VPR-CTD-2353F22.1-1	TTTCTTGGCTTTATATATCTTG TGGAAAGGACGAAACACCGA TGAGTACGAACATGAAGG	GACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAACCCTTCA TGTTCGTACTCATC
gRNA-VPR-CTD-2353F22.1-6	TTTCTTGGCTTTATATATCTTG TGGAAAGGACGAAACACCGT GGGTTGTCTGTCTTGTCT	GACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAACAGACAA GACAGACAACCCAC
gRNA-VPR-CTD-2353F22.1-9	TTTCTTGGCTTTATATATCTTG TGGAAAGGACGAAACACCGG ATCTCCTTCTTGAGGGAG	GACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAACCTCCCT CAAGAAGGAGATCC
gRNA-VPR-scrambled	TTTCTTGGCTTTATATATCTTG TGGAAAGGACGAAACACCGC ACTACCAGAGCTAACTCA	GACTAGCCTTATTTTAACTTGCT ATTTCTAGCTCTAAAACTGAGTT AGCTCTGGTAGTGC

Table 14: Oligonucleotides for cloning sgRNAs for CRISPR activation using SAM-system.

Designation	Forward (5'-3')	Reverse (5'-3')
gRNA-rs4284742-1	CACCGTTATTTCCTGTCACAT TAG	AAACCTAATGTGACAGGAAAT AAC
gRNA-rs4284742-2	CACCGCTCCTCCCCTGGCCTA TGC	AAACGCATAGGCCAGGGGAGG AGC
gRNA-rs4284742-3	CACCGACCACAGGAGGGATG GGAGA	AAACTCTCCCATCCCTCCTGTG GTC
gRNA-rs11084095-1	CACCGCATCGCTGGGCATGTG AAT	AAACATTCACATGCCCAGCGAT GC
gRNA-rs11084095-2	CACCGTGTGCATCAATATGAG CTT	AAACAAGCTCATATTGATGCAC AC
gRNA-rs11084095-3	CACCGGCATAGATGAGAGTA TAAA	AAACTTTATACTCTCATCTATG CC
gRNA-rs11084095-4	CACCGCTGACATCCCAGAGG AGAT	AAACATCTCCTCTGGGATGTCA GC
gRNA-rs11084095-5	CACCGCTTGCTGAATCAAAGC AGC	AAACGCTGCTTTGATTCAGCAA GC
gRNA-rs11084095-6	CACCGTTCTGTGTTGCCAAAT GGT	AAACACCATTTGGCAACACAG AAC

gRNA-rs11084095-7	CACCGGGCAGTTTGTGACTATGGA	AAACTCCATAGTCACAAACTGCCC
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	CACCGGATCTCCTTCTTGAGGGAG	AAACCCTCCCTCAAGAAGGAGATCC
gRNA-scrambled	CACCGCACTACCAGAGCTAACTCA	AAACTGAGTTAGCTCTGGTAGTGC

Table 15: Oligonucleotides used for qRT-PCR.

Designation	**Forward (5'-3')**	**Reverse (5'-3')**
GAPDH-RT	GCATCTTCTTTTGCGTCG	TGTAAACCATGTAGTTGAGGT
SIGLEC5-RT	AAGGTCAACTCCAGCTCAGC	TTCGATCTCCCTTGCAGCAG
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
PTK9-RT	AGCTCAACTATGTGCAGT TGGAAA	ACGAGCTGAATCCTTGGGAA
Luciferase-RT	ACGTGCAAAAGAAGCTACCG	GGCAAATGGGAAGTCACGAA
pGL4.24 backbone	TTCAACCCAGTCAGCTCCTT	CAAGAACTCTGTAGCACCGC

2.10 Mediums and Buffer

Medium

YT medium	16 g Tryptone
	10 g Yeast extract
	5 g NaCl
	(15 g Agar)
	adjust volume to 1 L
	pH 7.0

Annealing Buffer

Oligonucleotide Annealing Buffer	10 mM Tris
	1 mM EDTA
	50 mM NaCl
	pH 8.0

Gel electrophoresis buffer

50 x TAE buffer	242 g Tris (base)
	57.1 mL Glacial Acetic Acid
	100 mL 0.5 M EDTA
	adjust volume to 1 L
	pH 8.3

10x TBE buffer	108 g Tris (base)
	55 g boric acid
	40 mL 0.5 M EDTA
	adjust volume to 1 L
	pH 8.3

Elution buffer

TE-buffer	10 mM TRIS-HCl
	0.1 mM EDTA
	pH 8.0

3 Methods

3.1 Identification of putative causal variant(s) underlying the associations at *SIGLEC5* and *CTD-2353F22.1*

3.1.1 Selection of putative causal variant(s)

The identification of putative causal variant(s) was done by determining all SNPs in strong LD to the GWAS lead-SNPs (Munz et al. 2017; Shungin et al. 2019) in the GBR [British in England and Scotland] (Genomes Project et al. 2010) population and North-Western European populations CEU [Utah Residents with Northern and Western European Ancestry] of the International Genome Sample Resource (IGSR), using the online tool LDproxy (Machiela and Chanock 2015). Highly predictive features for DNA sequences with a regulatory function on gene expression were predicted by ENCODE data (ENCODE-Project-Consortium 2012) of experimentally confirmed TF binding in several cell types, chromatin accessibility, DNase-hypersensitivity and H3K27Ac and H3K4Me1 histone modifications. Associated SNPs were analyzed whether they were located in those predictive features.

Using Transfac professional (geneXplain) (Wingender 2008), SNPInspector (Genomatix) (Quandt et al. 1995) and the open-access database Jaspar (Sandelin et al. 2004), sequences including the SNPs were aligned with matrix descriptions for transcription factor binding sites (TFBS) to analyze if transcription factors (TF) bind allele-specific at the sequences of the SNPs. To confirm the TF binding motif, PWMTools (Ambrosini et al. 2018), a web tool for Position Weight Matrix model generation and evaluation, was used.

3.1.2 Expression quantitative trait loci analysis

The software tool QTLizer (Munz et al. 2020) was used to annotate eQTL effects of the associated SNPs. The results and eQTLs of the GWAS lead-SNPs rs4284742, rs11084095, rs1122900 and rs6887423 and are given in chapter 8.1 (**Table 23 – 26**).

3.1.3 Isolation of Peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMCs) were isolated from 10 mL whole blood, collected in sodium heparin treated tubes. Whole blood was transferred into a 50 mL falcon and diluted with an equal volume of 1 x Dulbecco‘s phosphate-buffered saline (DPBS) and mixed

thoroughly. Histopaque-1077 (15 mL) were placed into another falcon and the diluted blood was carefully layered over it. The tube was centrifuged for 30 minutes (min) at 400 x g with very low acceleration and natural deceleration. After centrifugation, the PBMC layer was carefully transferred into a new falcon and washed with 3 volumes of DPBS. The tube was centrifuged at 350 x g for 10 min. The PBMC pellet was washed twice with 50 mL DPBS and cells were counted using hemocytometer. Subsequently, the nuclear extract of PBMCs was isolated for EMSAs as described in chapter 3.1.5.

3.1.4 Cultivation of Raji cells

The suspension B lymphocyte Raji cell line was cultivated at 37°C in a T75 flask in an upright position with 5% CO_2. Raji cells were grown in RPMI-1640 Medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin (pen/strep) until they reached approximately 2 million cells/mL confluence. For passaging the cells, Raji cells were resuspended in their culture medium and centrifuged at 200 x g for 10 min. The medium was aspirated and the cells were resuspended in appropriate volume of medium to reach a cell density of 600.000 cells/mL.

3.1.5 Nuclear protein extraction isolation

Nuclear protein extract was isolated from PBMCs and B lymphocyte cells (Raji cells) using the Nuclear Extract Kit (Active Motif).

Suspension cells (9 x 10^6) were centrifuged for 5 min at 200 x g and 4 °C, supernatant was discarded and cells were washed with 3 mL ice-cold PBS/phosphatase inhibitors. After another centrifugation step, cell pellet was resuspended in 500 µL hypotonic buffer and incubated for 15 min on ice. Detergent (25 µL) was added and vortexed for 10 seconds (sec) at the highest setting. Hypotonic buffer and detergent were used to isolate cytoplasmic proteins by bursting the cell wall and to keep the nuclear membrane intact. Suspension was centrifuged for 30 sec at 14000 x g and 4 °C. The supernatant including cytoplasmic protein fraction was transferred to a microcentrifuge tube and stored at -80 °C and pellet was resuspended in 50 µL Complete lysis Buffer. This buffer bursts the nuclear membrane. Suspension was vortexed at the highest setting for 10 sec and was incubated for 30 min at 150 rpm and 4°C on a rocking platform. Finally, suspension was vortexed for 30 sec, centrifuged for 10 min at 14000 x g and 4°C and

supernatant was transferred to a microcentrifuge tube. Protein concentration was measured, protein was aliquoted and stored at -80 °C for further use in EMSA (chapter 3.1.7).

3.1.6 Quantitative protein concentration measurement with Protein DC Assay

The DC (Detergent Compatible) Protein Assay is a colorimetric assay for the quantitative determination of protein concentration. The absorbance of the protein solution was measured at a wavelength of 540 nm, 650 nm or 750 nm and a standard curve was used to determine the concentration of protein present in the solution.

For the standard curve, bovine serum albumin (BSA) with a concentration of 1 mg/mL, 0.8 mg/mL, 0.6 mg/mL, 0.4 mg/mL, 0.2 mg/mL and 0 mg/mL was used. The protein solution (extracted from PBMCs and Raji cells) was diluted 1:10 and 1:5 before it was used for the assay. An amount of 5 µl of each standard and sample were transferred in a microtiter plate and then 25 µl of solution A' was added. Solution B (200 µl) was added and incubated for 15 min at RT. Solutions A' and B were supplied by the manufacturer and are not further described in the manufacturer's instructions. Finally, the concentration of proteins was measured using the Multiskan Go spectrophotometer. The spectrophotometer automatically generates a standard curve and uses its slope to calculate the concentration of each protein sample.

3.1.7 Electrophoretic Mobility Shift Assay

The Electrophoretic Mobility Shift Assay (EMSA) is used to detect protein-DNA interactions even with low protein concentration within the extract (Carey et al. 2012). It is based on the principle that protein-DNA complexes migrate through a native polyacrylamide gel more slowly than free DNA, with the mobility of the complex determined by the size, shape, charge and multimeric state of the protein. In this thesis it was used to determine allele-specific binding of the oligonucleotides, which included the SNP (rs4284742, rs11084095, rs34984145, rs1122900, rs6887423 and rs56038114) and 21 bp of the up- and downstream sequence, to nuclear proteins, extracted from Raji cells or PBMCs. The reaction mixture containing small biotin-labeled DNA fragments (including the SNP) with the cell extract was electrophoresed through a native polyacrylamide gel, which separates the free biotin-labeled probe DNA from the molecules bound by proteins. Afterwards, the samples were transferred to a nylon membrane, crosslinked with a UV cross-linker and finally the free and bound DNA molecules

were detected using streptavidin conjugated to horseradish peroxidase (HRP) by chemiluminescence. The EMSAs were performed using the Gelshift Chemiluminescent EMSA Kit.

3.1.7.1 Annealing of Oligonucleotides

All Biotin 3' end-labeled and unlabeled oligonucleotides had a total length of 43 bp including each SNP allele. Sequences are given in **Table 10**. Complementary oligonucleotides were mixed in a ratio 1:1 in a microcentrifuge tube and diluted with Oligonucleotide Annealing buffer (chapter 2.10) to obtain a final concentration of 1 pmol/µL.

The components were mixed and incubated for 5 min at 95°C in a heating block. The heating block was turned off and the samples were further incubated in the heating block until the annealed oligonucleotides have reached RT. Annealed oligonucleotides were stored at 4°C until further use.

3.1.7.2 Protein-DNA binding reactions and electrophoresis

For each allele of a SNP, five binding reactions were loaded. The first reaction was a control reaction in which no protein extract was added, so no protein-DNA shift can occur. In the second reaction, protein extract was added, so the target protein binds to the biotin-labeled oligonucleotide and a shift causes as compared to reaction 1. Unlabeled oligonucleotides were added to the binding reaction of protein extract and biotin-labeled DNA in reaction 3, so no shift can occur because the excess of unlabeled DNA competes for binding to the target protein. This reaction verifies the specific DNA-protein interaction in reaction 2. Reaction 4 is the same as reaction 2, but additionally specific antibody was added to verify the predicted TFBS and the allele-specific binding of the predicted protein. Unlabeled oligonucleotide and specific antibody were added to the binding reaction of protein extract and biotin-labeled DNA in reaction 5, to verify the specific antibody-shift in reaction 4. Specific antibodies are given in **Table 8**. Biotin-labeled oligonucleotides were diluted 1:100 before use in binding reaction to get a final concentration of 10 fmol/µL. Poly d(I-C) was added to reactions to reduce non-specific binding.

A 5% native polyacrylamide gel was prepared as follows:

4.2 mL	UPW
600 µL	5 x TBE
1 mL	30% acrylamide solution
60 µL	50% Glycerol
10 µL	TEMED
40 µL	10% APS

The native polyacrylamide gel was pre-run in 0.5 x TBE for 45-60 min at 100 V.

Table 16: Composition of the five reactions in EMSA (as mentioned above).

Reagent	**Final Concentration**	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**
UPW	-	12 µL	10 µL	6 µL	9 µL	5 µL
10 x Binding Buffer	1 x	2 µL	2 µL	2 µL	2 µL	2 µL
1 µg/µL Poly d(I-C)	50 ng/µL	1 µL	1 µL	1 µL	-	-
Unlabeled DNA	4 pmol	-	-	4 µL	-	4 µL
Nuclear extract	10 µg	-	5 µL	5 µL	5 µL	5 µL
Specific antibody	dependent on antibody concentration	-	-	-	2 µL	2 µL
Biotin-labeled DNA	20 fmol	2 µL	2 µL	2 µL	2 µL	2 µL
Total volume	-	20 µL	20 µL	20 µL	20 µL	20 µL

Reactions were incubated for 20 min at room temperature (RT). After the pre-run, 5 µL of 5 x loading buffer were added and 20 µL of the reaction were loaded onto the 5% native polyacrylamide gel and run in 0.5 x TBE at 100 V for 1 hour.

3.1.7.3 Transfer of binding reactions to Nylon Membrane

The previously prepared native polyacrylamide gel (chapter 3.1.7.2) was placed in the "Wet Transfer" cell with a nylon membrane between two blot papers and two sponges, after the bromophenol blue dye has migrated approximately 4/5 down the length of the gel. The membrane as well as the blot papers and the sponges were previously placed in 0.5 x TBE for

10 min. A glass stirring rod was used to ensure no air bubbles between the membrane and the gel. The transfer was performed at 380 mA for 60 min. Once the transfer was completed, the membrane was placed on a dry paper towel and was cross-linked at 120 mJ/cm^2 for 60 sec. Irradiation with ultraviolet light of the protein-DNA complex forms covalent bonds between the oligonucleotides and proteins (Chodosh 2001).

To detect the allele-specific protein-DNA binding by chemiluminescent, the nylon membrane was incubated with 20 mL blocking buffer for 15 min at RT. Afterwards the blocking buffer was replaced with streptavidin conjugate diluted in blocking buffer (1:300) and the membrane was incubated for another 15 min. The membrane was washed five times for 5 min in wash buffer and then it was incubated for another 5 min in substrate equilibration buffer. All incubation steps were performed on an orbital shaker.

Chemiluminescent reagent and reaction buffer were mixed in a ratio of 1:1 to prepare a chemiluminescent working solution. This solution was added onto the nylon membrane until the whole membrane was covered and was incubated for 5 min. The excess chemiluminescent working solution was removed and the membrane was wrapped into a plastic wrap. Biotin-labeled oligonucleotides were detected using X-ray films or CHEMOSTAR Touch ECL & Fluorescence Imager (Intas).

The intensity of the allele-specific shifted bands was analyzed by calculating the absolute value area using ImageJ.

3.1.8 Luciferase based reporter gene assay

Luciferase based reporter gene assays are used to study intracellular signaling, transcription factors and regulation of gene expression in various cellular contexts. To identify and characterize functional variants, luciferase reporter gene assay is a well suited method (Nair and Baier 2018). In this thesis, the SNPs and the up- and downstream sequences were cloned into the pGL4.24 vector to prove whether these SNPs have a regulatory function in gene expression and to quantify the allele-specific strength of the associated regulatory elements and the effect directions. The sequence of interest was cloned upstream of a minimal promoter and the luciferase reporter gene *luc2P*. In the present work, the Dual-Luciferase Reporter Assay System was used. In this system, two individual reporter enzymes, the Firefly luciferase that was cloned downstream of the sequence including the SNP, and the *Renilla* luciferase, as an internal control were measured from a single sample. Plasmid pRL-SV40 was used as a control

plasmid that expresses the *Renilla* luciferase. Both luciferases have dissimilar enzyme structures and substrate requirements which makes it possible to discriminate between their bioluminescent reactions and do not require post-translational processing for enzymatic activity (Matthews et al. 1977; Wood et al. 1984).

After transfection of the reporter gene expression plasmid (pGL4.24) into cells, the reporter gene (luciferase) is controlled by the same signals as the DNA sequence cloned into the plasmid. The luciferase activity can be measured directly and functions as a reporter of the transcriptional status of the sequence of interest (Bauer 2017). Depending on the strength of the signal, the cloned sequence can act as a repressor, an activator or has no effect on gene expression.

3.1.8.1 Cloning of Luciferase Reporter Plasmids

The genomic regions at rs4284742, rs11084095, rs34984145, rs1122900, rs6887423 and rs56038114 were cloned with an up- and downstream sequence up to 75 bp in an expression vector (pGL4.24) upstream of the luciferase reporter gene. For each SNP, both alleles were cloned into the pGL4.24 plasmid. The exchange of the alleles was performed with the Q5-mutagenesis kit as described in chapter 3.1.8.2.

3.1.8.1.1 Polymerase chain reaction

Polymerase chain reaction (PCR) is a method for amplification of DNA sequences. To amplify the sequences described above for cloning into the reporter gene plasmid pGL4.24 the PCR reaction was as follows:

2.5 µL 10 x Taq Reaction Buffer
0.5 µL 100 mM dNTPs
1 µL fwd-Primer (10 µM)
1 µL rev-Primer (10 µM)
1 µL Template DNA (genomic DNA; 100-200 ng)
0.25 µL Taq DNA polymerase
18.75 µL UPW

---

25 µL total volume

The reaction conditions of PCR are given in Table 17.

Table 17: Reaction conditions of PCR.

Number of Steps	Temperature	Time	cycle Step	
1	95 °C	5 min		
2	95 °C	1 min	Denaturation	
3	55-65 °C	15 sec	Annealing	
4	72°C	15 sec/kb	Elongation	to step 2 (40 cycles)
5	72°C	5 min		

3.1.8.1.2 Restriction digest

The PCR products were cloned into the pGL4.24 plasmid using the *Xho*I and *Kpn*I restriction enzymes. The reaction mixture was as follows:

2 µL DNA (~ 2 µg PCR product/pGL4.24)
1 µL *Xho*I
1 µL *Kpn*I
5 µL 10 x CutSmart Buffer
41 µL UPW

---

50 µL total volume

The reaction was incubated at 37°C for 2-3 hours. The product was then loaded onto a 1% agarose gel as described in chapter 3.1.8.1.3 and the DNA of the restricted plasmid was isolated from the agarose gel as specified in chapter 3.1.8.1.4.

3.1.8.1.3 Gel electrophoresis

Gel electrophoresis is used to examine the size of DNA fragments generated during PCR or restriction digest. By applying an electrical voltage, the DNA fragments are separated in an agarose gel according to their size. 1% - 2.5% agarose gels were used. To prepare the 1% agarose gel, 2.5 g of agarose powder were dissolved in 250 mL of 1xTAE buffer by heating

and 5 µL of ethidium bromide were added. The gel pockets of the polymerized gel were loaded with 50 µL of DNA previously prepared for gel electrophoresis with a dye- and glycerol-containing loading dye in a 1:5 ratio. The marker GeneRuler 1 kb Plus DNA Ladder or 50 bp O'Range Ruler (5 - 7 µL) were pipetted into a pocket for later identification of the size of DNA bands in the gel. A voltage of 90 - 120 V for 20 -120 min was applied.

3.1.8.1.4 Purification of DNA fragments from agarose gel

The DNA fragments were purified from agarose gel using the QIAquick Gel Extraction Kit. For this purpose, the appropriate DNA bands were first cut out of the agarose gel using a scalpel and weighed. For each 100 mg gel, 300 µL of binding buffer was added to the DNA fragments and the mixture was incubated at 50°C for 10 min. After the agarose gel slice had been completely dissolved, the entire content was pipetted onto a column and centrifuged at 14000 x g for 1 min. During this step, the DNA was adsorbed onto the silica membrane of the column due to the high salt concentration of the previously added buffer, while primers, enzymes, and other contaminants were discarded as supernatant. The DNA was washed with 750 µL of wash buffer to remove excess salts, and finally eluted with 30 µL of UPW.

3.1.8.1.5 Ligation

The formation of a phosphodiester bonds between two single strands of DNA (or a linearized plasmid DNA) is referred to as DNA ligation. In this process, a phosphodiester bond is formed by a ligase that catalyzes the connection between the 3'- hydroxyl end of the PCR products and the 5'- phosphate end of the restricted plasmid DNA.

The ligation reactions for cloning the reporter gene plasmids were as follows:

x µL pGL4.24 +*Xho*I +*Kpn*I (50 ng)
x µL PCR product +*Xho*I +*Kpn*I (6 ng)
2 µL T4 ligation buffer
1 µL T4 ligase
15 µL UPW

---

20 µL total volume

After incubation of the ligation reaction at RT for 2 hours, the transformation in *E. coli* cells was performed as reported in chapter 3.1.8.1.6.

3.1.8.1.6 Transformation

The uptake of DNA by competent bacterial cells is understood as transformation. 100 µL of self-prepared competent Top10 *E. coli* cells were slowly thawed on ice for 10 min for transformation and incubated with 10 µL of ligation mixtures for 20 min on ice. Cells were heated for 50 sec at 42 °C (called heat shock) and kept on ice for additional 2 min. During the heat shock, plasmid DNA is introduced into the cells. Next, 400 µL SOC medium was added and the cell mixture was incubated for 1 h at 37 °C and 250 rpm. Finally, the cell mixture was plated onto a YT-agar plate (mixed with 100 mg/mL ampicillin (amp)) and incubated overnight at 37 °C. Individual colonies were picked and grown overnight in 8 mL YT-medium complemented with amp (100 mg/mL). Plasmid DNA was isolated as reported in chapter 3.1.8.3.

3.1.8.2 Q5 Site-Directed Mutagenesis Cloning

The alleles in the cloned pGL424+SNP plasmids were exchanged using the Q5 Site-Directed Mutagenesis Kit (NEB) that utilizes a Hot Start High-Fidelity DNA polymerase and mutagenic primers to create substitutions. Primer sequences are given in **Table 12**.

PCR reaction was as follows:

12.5 µL Q5 Hot Start High-Fidelity 2x Master Mix
1.25 µL fw-Primer (10 µM)
1.25 µL rev-Primer (10 µM)
1 µL Template DNA (pGL4.24+SNP; 25-50 ng)
9 µL UPW

---

25 µL total volume

The reagents were mixed and transferred to a thermocycler. After PCR, the product was incubated for 30 min with a KLD enzyme mix that includes a kinase for phosphorylation, a ligase for ligation and DpnI for degradation the template DNA. Subsequently, the transformation of the KLD mix was performed as described in chapter 3.1.8.1.6.

3.1.8.3 Plasmid DNA preparation using the example of a Midi

Plasmid DNA preparation is the isolation of plasmid DNA from bacterial cells. For a midi preparation, a 25 mL overnight culture was centrifuged at 6000 x g for 15 min and DNA was isolated from the resulting pellet using the QIAGEN Plasmid Midi Kit. First, the bacterial cell pellet was resuspended in 4 mL of resuspension buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 ug/mL RNase A), 4 mL of lysis buffer (200 mM NaOH, 1% SDS) were added and the mixture was incubated for 5 min at RT. The lysis buffer contains NaOH and SDS, which help to denature plasmid DNA (NaOH) and form SDS-protein complexes. After addition of 4 mL of pre-chilled neutralization buffer, the mixture was incubated on ice for 15 min and then centrifuged at 12000 x g for 60 min. Potassium acetate and acetic acid in the buffer provide neutralization of the preparation, renaturation of plasmid DNA, and precipitation of SDS-protein complexes. After preparation of the column with 4 mL of activation buffer, the DNA solution, remaining cell components, and buffers were pipetted onto the column. The DNA was bound to the column while the buffer residues and remaining components passed through the column and were discarded as supernatant. By adding 2 x 10 mL of alcohol-containing wash buffer (1.0 M NaCl, 50 mM MOPS, pH 7.0, 15% isopropanol), the DNA was washed and the last buffer and salt residues were removed from the column. DNA was then eluted by adding 5 mL of elution buffer (1.25 M NaCl, 50 mM Tris-Cl, pH 8.5, 15% isopropanol) and collected in a 15 mL Falcon. DNA was precipitated by adding 3.5 mL isopropanol followed by centrifugation for 40 min at 4°C and 12000 x g. The DNA pellet was washed with 2 mL of 70% ethanol and centrifuged again for 20 min at 12000 x g and 4°C. Finally, the pellet was resuspended in 50 µL TE buffer and the concentration was determined.

DNA and RNA concentrations were measured using the Multiskan Go spectrophotometer. For this purpose, 2 µl standard (UPW or TE-buffer) and sample solution were pipetted onto the µDrop plate. For each measurement, 2 blanks were used. The absorbance of DNA and RNA solutions was measured at 260 nm and 280 nm.

3.1.8.5 Sanger Sequencing

To verify the correct sequence of PCR products and cloning of gRNAs into the sgRNA (MS2) cloning backbone vector, sequencing was performed at LGC Genomics. For plasmids, 100 ng/µL in a volume of 20 µL and for PCR products with a size of 500 bp-1000 bp, 20 ng/ µL in a volume of 15 µL were submitted to the company. Sequencing analyses were performed using Clone Manager 9 software.

3.1.9 Transfection of Raji cells

Raji cells were cultivated as described in chapter 3.1.4. One day prior to transfection, Raji cells were split to 600.000 cells per mL medium. The day of the experiment Raji cells were resuspended in their culture medium and centrifuged at 200 x g for 10 min. The medium was aspirated and the cells were resuspended in DPBS and counted using hemocytometer. 500,000 cells per replicate were transfected using the Neon Transfection System. Therefor an appropriate number of Raji cells were centrifuged at 200 x g for 10 min, DPBS was aspirated and Raji cells were resuspended in suitable volume of Resuspension Buffer R. Per replicate, 3 µg DNA was transferred to a sterile 1.5 mL microcentrifuge tube and the cell suspension was added and gently mixed. This reaction was aspirated into Neon pipette and transfected using electroporation. Electroporation is a transfection method in which the permeability of the cell membrane is increased through external electrical fields. The optimized electroporation parameters for Raji cells were 1600 V, 5 ms and 6 pulses. After electroporation, the cells were transferred into a prepared 24-well plate containing prewarmed RPMI-1640 Medium supplemented with 10% FBS but without antibiotics. Depending on the experiment, transfected cells were harvested 24-48 hours after transfection.

3.1.10 Cultivation of HeLa cells and HEK293 cells

HeLa cells and HEK293 cells were cultivated at 37°C in a T75 flask in an incubator with 5% CO_2. Passaging of the cells was performed at an approximate confluence of 90%. For this purpose, the medium was removed from the confluent grown cells and cells were washed twice with DPBS. The DPBS was then removed and the cells were incubated at 37°C for 5 min with 2 mL Trypsin/EDTA. The trypsinization of HeLa cells was stopped by adding 10 mL culture medium (Earle's MEM (+ 10% FBS, + 1% NEA, + 2 mM L-Glutamine, + 1% pen/strep)) and trypsinization of HEK293 cells was stopped by adding 10 mL culture medium (DMEM with

4.5 g/L glucose (+ 10% FBS, + 1% pen/strep)). Cells were transferred to a 50 mL falcon and centrifuged for 5 min at 300 x g. The supernatant was removed and the cell pellet resuspended in fresh medium. Depending on the confluence of cells, between 0.5 and 2 mL of cell suspension was transferred to a new T75 cell culture flask containing 12 mL culture medium, and cells were incubated in the incubator until the cells again reached 90% confluence.

3.1.10.1 Transfection of HeLa cells and HEK293 cells

One day prior to transfection, cells were seeded at 100.000 cells per well in 6-well tissue culture plates. Cells were always transfected in three independent biological replicates. The day of the experiment, the medium of cells was changed before transfection. Cells were transfected using jetPEI, a linear polyethlenenimine derivative, which belongs to the cationic polymers transfection reagents. Transfection reactions for one 6-well were as follows:

1. Reaction:

2 µL (2.7 µg pGL4.24+SNP and 0.3 µg pRL-SV40) DNA
100 µL 150 mM NaCl

2. Reaction:

6 µL jetPEI
100 µL 150 mM NaCl

The jetPEI solution was added to the DNA solution at once. The solution was vortexed immediately, spinned down briefly and incubated for 30 min at RT. Afterwards, the DNA-jetPEI complex was added dropwise to the cells. 24 hours after transfection, the cells were washed twice with PBS. Cell disruption and total RNA extraction were carried out as described in chapter 3.1.13.

3.1.11 Dual-Luciferase Reporter Assay System

In the Dual-Luciferase Reporter Assay System (Promega) system, two individual reporter enzymes, the Firefly luciferase that was cloned downstream of the sequence including the SNP, and the *Renilla* luciferase, as an internal control to normalize the data and therefrom to minimize

experimental variability, were measured from a single sample. HeLa cells were cultivated and transfected as reported in chapter 3.1.10 and chapter 3.1.11. 24 hours after transfection, cells were washed with PBS and incubated for 15 min with 1 x Passive lysis buffer (PLB) on a rocking platform. Cell suspension was transferred into a microcentrifuge tube and 5 µL were pipetted into a 96-well plate. The 96-well plate was put in Orion Microplate Luminometer (Berthold, Germany) which was used according to the manufacturer's instructions. The luminometer was programmed to perform a 2-second premeasurement delay, followed by a 10-second measurement period for each reporter assay. 25 µL of luciferase assay reagent II (LAR II) was added to each sample. Afterwards, the luminescent signal of the Firefly luciferase reporter was measured. With adding 25 µL of Stop&Glo reagent, this reaction was quenched and the *Renilla* luciferase reaction was initiated and measured.

3.1.11.1 Data Analysis

The activities of the reporter gene assay experiments were quantified as relative light units that were normalized as the ratio of Firefly luciferase activity to *Renilla* luciferase activity. Relative luciferase activity was calculated as the ratio of the average (Firefly/*Renilla*) from the sample (SNP + up- and downstream sequence) to the average (Firefly/*Renilla*) from the empty plasmid pGL4.24 (as control). Data was analyzed using two-tailed Student's t-test.

3.1.12 Cloning of gRNAs for CRISPRa

3.1.12.1 CRISPR/Cas

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-associated (Cas) protein systems form an adaptive immune system in many bacteria and most archaea and provide a robust tool for genome engineering (Nair and Baier 2018). In nearly 40% genomes of sequenced bacteria and around 90% genomes of sequenced archaea, CRISPR/Cas was observed (Nair and Baier 2018) as an adaptive immune system against viruses and plasmids by using a guide CRISPR RNA (crRNA) for directed target sequence recognition and protein (Cas)-mediated DNA cleavage (Jinek et al. 2012). Depending on the structures and sequences of Cas proteins, CRISPR/Cas systems are classified into type I, type II and type III systems (Makarova et al. 2011). In this thesis, the type II system was used. Type II systems use a single Cas protein, Cas9, that requires crRNA and trans-activating crRNA (tracrRNA) to cleave

nuclease domains. Cas9 contains two nuclease domains, the RuvC-like nuclease domain that cleaves the non-complementary strand and the HNH nuclease domain required for cleaving the DNA strand complementary to the guide RNA sequence (Nair and Baier 2018). In addition, a protospacer adjacent motif (PAM) sequence downstream of the target sequence is required for CRISPR/Cas9 mediated cleavage (**Figure 6**). The 3 nt PAM sequence 5'-NGG is recognized by the most used Cas9 protein (SpCas9) which comes from *Streptococcus pyogenes*.

The characteristic of Cas9-cRNA complex for directed target sequence recognition and protein (Cas)-mediated DNA cleavage makes it a universal tool for genome engineering with RNA-guided endonucleases.

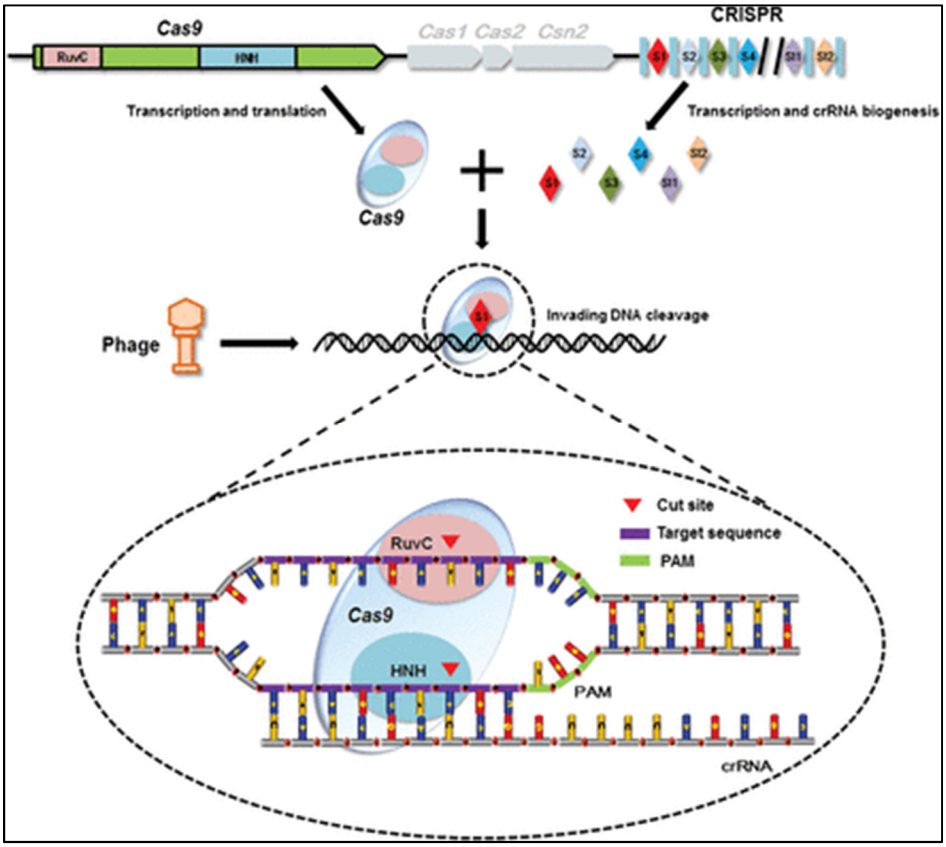


Figure 6: Illustration of CRISPR/Cas9 mediated cleavage. Cas9 protein is guided by mature crRNA to the target site of invading plasmid or phage DNA. The HNH nuclease domain and the RuvC-like nuclease domain cut the complementary and the non-complementary strand of DNA to the matching crRNA (taken from(Zhang et al. 2014a)).

3.1.12.2 CRISPR activation

In this thesis, CRISPR activation (CRISPRa) was used. CRISPRa uses a mutant form of Cas9, dead Cas9 (dCas9), which still retains the ability to bind to the target DNA, but has no nuclease activity anymore, as HNH nuclease domain and RuvC-like nuclease domain are both inactivated (D10A and H840A) (Sander and Joung 2014). The protein dCas9 is recruited by

single guide RNA (sgRNA) to specific target DNA sequence (complementary to the sgRNA). dCas9 fused to a transcriptional activator (VP64) have been shown to activate expression of endogenous genes in human cells (Maeder et al. 2013; Perez-Pinera et al. 2013). The efficiency of the VP64-p65-Rta (VPR) and the Synergistic Activation Mediator (SAM) system were compared, in which first the promoter of *CTD-2353F22.1* was activated.

3.1.12.3 CRISPR activation using the SAM system

The Synergistic Activation Mediator (SAM) system, which was used in this thesis among others, is an efficient dCas9-based activator system. Additionally, to the dCas9-VP64, a modified gRNA containing MS2 RNA aptamers at the tetraloop and stem-loop 2 and MS2-p65-HSF1 activator helper complex are required for stable and robust transcriptional activation with SAM (Konermann et al. 2015).

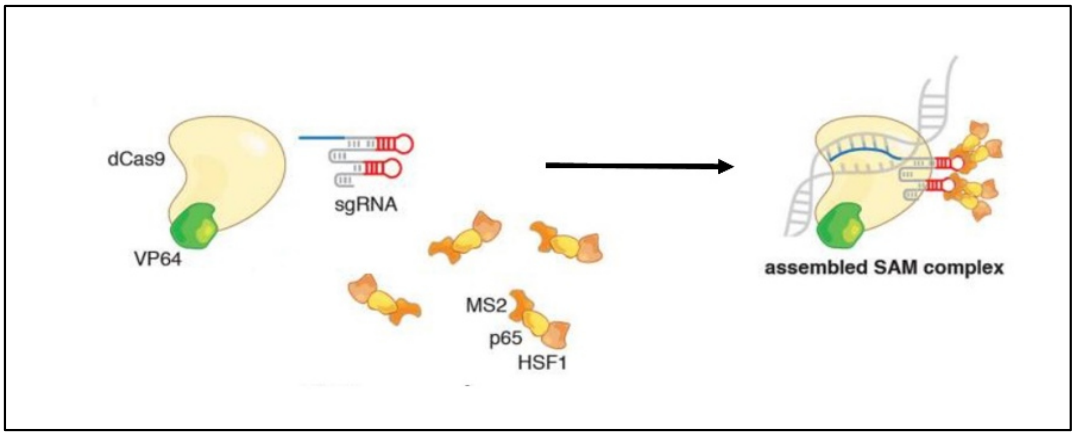


Figure 7: Schematic of the SAM system. For SAM activation additional RNA-binding helper activators are required (modified after (Konermann et al. 2015))

In this thesis, CRISPRa with the SAM system was used, first, to activate the ncRNA *CTD-2353F22.1* in the genomic context for further analysis and second, to validate, if the associated regulatory elements (SNPs) have *cis* or *trans* regulatory effects on the expression of the predicted genes (*SIGLEC5*, *CTD-2353F22.1*).

3.1.12.3.1 Design of sgRNAs

Single guide RNAs (sgRNAs) were designed with the CRISPR-ERA tool (Liu et al. 2015) to identify optimized sgRNA sequences. All sgRNAs for gene activation of *CTD-2353F22.1* were designed within -400 bp distance to the transcription start site (TSS) and next to the PAM sequence NGG. The sgRNAs that were used to validate, if the associated regulatory elements (SNPs) have regulatory effects on the expression of the predicted genes (*SIGLEC5*, *CTD-2353F22.1*) were designed as close as possible to the associated SNP and also next to the PAM sequence NGG. Using BLAST (basic local alignment search tool) to find regions of similarity between different sequences, sgRNAs were tested for binding unique in the human genome. A scrambled sequence was cloned into the gRNA backbone as negative control. All sgRNAs were synthesized by metabion international AG (Planegg/Steinkirchen) with flanking overhangs. Sequences can be found in **Table 14**.

3.1.12.3.2 Annealing of sgRNAs

For cloning sgRNAs, Golden Gate assembly based on Type IIS restriction enzymes, was used. Type IIS restriction enzymes create four base flanking overhangs by cleaving outside of their recognition site.

The annealing reaction of each sgRNA was as follows:

1 µL forward (fwd) oligo
1 µL reverse (rev) oligo
2 µL 10 x T4 ligation buffer
0.5 µL T4 Polynucleotide Kinase (PNK)
15.5 µL UPW

The components were mixed and incubated for 30 min at 37°C followed by incubation for 5 min at 95°C in a heating block. The heating block was turned off and the samples were further incubated in the heating block until the annealed oligonucleotides have reached RT.

3.1.12.3.3 Restriction digest

The annealed oligonucleotides of sgRNAs were cloned into the sgRNA(MS2) cloning backbone (Addgene Plasmid No. 61424) using the *BbsI-HF* restriction enzyme.

The reaction mixture was as follows:

2 µL sgRNA(MS2) cloning backbone plasmid DNA (2 µg)
2 µL *BbsI-HF*
5 µL 10 x CutSmart Buffer
41 µL UPW

---

50 µL total volume

The reaction was incubated at 37°C for 2-3 hours. The product was then loaded onto a 1% agarose gel as described in chapter 3.1.8.1.3 and the DNA of the restricted plasmid was isolated from the agarose gel as specified in chapter 3.1.8.1.4.

3.1.12.3.4 Ligation

The ligation was performed as described in 3.1.8.1.5. To prevent self-ligation of the restricted vector plasmid sgRNA(MS2) cloning backbone, a dephosphorylation reaction was carried out with 2 µL alkaline phosphatase (CIP) for 45 min at 37 °C.

Subsequently, the ligation reaction was as follows:

1 µL sgRNA(MS2) cloning backbone plasmid DNA + BbsI (50 ng)
1 µL annealed sgRNAs
2 µL T4 ligation buffer
1 µL T4 ligase
15 µL UPW

---

20 µL total volume

After incubation of the ligation reaction at RT for 2 hours, it was used for transformation in *E. coli* cells and plasmid DNA was isolated as reported in chapter 3.1.8.3.

3.1.12.4 CRISPR activation using the VPR system

The VP64-p65-Rta (VPR) system is a tripartite activator fused to dCas9. This system uses the transcriptional activator VP64 fused to the dCas9, as already shown for the SAM system, as well as the transcriptional activator p65 and Rta. Again, a gRNA is required to guide the dCas9 complex to the target sequence (Chavez et al. 2015). For CRISPR activation with the VPR system, Addgene plasmid #63798 was used, which includes the tripartite activator complex. The gRNA was cloned into the gRNA cloning vector (Addgene #41824).

3.1.12.4.1 Polymerase chain reaction for Gibson Assembly

Guide RNAs for the CRISPRa were cloned into the gRNA cloning vector using Gibson assembly. 19 bp of the target sequence were ordered with 41 bp overhang (exact sequences are given in **Table 13**) whereby the target sequence was complementary to each other in the fw- and rev-Primers. PCR with Phusion polymerase was performed to anneal and extend the oligonucleotides. The PCR reaction was as follows:

4 µL HF Buffer
0.4 µL 10 mM dNTPs
5 µL fwd-Primer (10 µM)
5 µL rev-Primer (10 µM)
0.2 µL Phusion DNA polymerase
5.4 µL UPW

---

20 µL total volume

PCR conditions were the same as reported in chapter 3.1.8.1.1.

3.1.12.4.2 Restriction digest

The annealed and amplified oligonucleotides were cloned into the gRNA cloning vector (Addgene Plasmid No. 41824) using the AflII restriction enzyme and the Gibson assembly cloning Kit (NEB).

The reaction mixture for the restriction digest was as follows:

2 µL gRNA cloning vector (2 µg)
2 µL AfIII
5 µL 10 x 3.1 Buffer (NEB)
41 µL UPW

---

50 µL total volume

The reaction was incubated at 37°C for 2-3 hours. The product was loaded onto a 1% agarose gel as described in chapter 3.1.8.1.4 and the DNA of the restricted plasmid was isolated from the agarose gel as specified in chapter 3.1.8.1.4.

3.1.12.4.3 Gibson Assembly

The Gibson Master Mix includes 5' exonuclease, DNA polymerase and DNA ligase. The 5'-3' exonuclease activity creates single stranded 3' overhangs. These complementary sequences then anneal and creating the double-stranded DNA of interest. DNA polymerase extends 3' ends and filling in the gaps and the DNA ligase seals the remaining nicks.

The reaction was as follows:

1 µL linearized plasmid (50-100 ng)
2 µL annealed oligonucleotides (5-fold molar excess)
10 µL Gibson Assembly Master Mix
7 µL UPW

---

20 µL total volume

The linearized plasmid and the annealed oligonucleotides were incubated with the Gibson assembly Master Mix (2 x) in a thermocycler at 50 °C and 60 min. Afterwards, the reaction was transformed into *E. coli* cells and plasmid DNA was isolated as reported in chapter 3.1.8.1.4.

3.1.12.5 Transfection of HeLa cells with CRISPR plasmids

HeLa cells were cultivated and transfected as described in chapter 3.1.10 and chapter 3.1.10.1. HeLa cells were seeded at 90.000 cells per well in a 6-well tissue culture plate for CRISPRa of the associated regulatory elements and the promoter of *CTD-2353F22.1*. For CRISPRa using the SAM system, Addgene Plasmid #61422, which expresses the dCas9-VP64 activator, Addgene Plasmid #61423, which expresses the MS2-p65-HSF1 activator helper complex and Addgene Plasmid #61424 which includes the gRNA and MS2 loops, were transfected. Of each plasmid, 1 µg was used to transfect one well of a 6-well plate. 24 hours after transfection, culture medium of transfected cells was changed and cells were harvested 40-48 hours after transfection as specified in chapter 3.1.13.

For CRISPRa using the VPR system, Addgene Plasmid #63798, which expresses the tripartite activator complex and Addgene Plasmid #41824 which includes the gRNA, were transfected. Of each plasmid, 1.5 µg was used to transfect one well of a 6-well plate. 24 hours after transfection, culture medium of transfected cells was changed and cells were harvested 40-48 hours after transfection as reported in chapter 3.1.13.

3.1.13 Isolation of Total RNA from cells

Total RNA from HeLa cells was isolated 24-48 hours after transfection. The medium of the 6-well cell culture plates was aspirated and cells were washed twice with 1 x DPBS. After the last washing step, DPBS was removed and 350 µL lysis buffer mixed with 3.5 µL ß-mercaptoethanol was added to each well for sufficient lysis. ß-mercaptoethanol was added to inhibit RNases. Cells were detached from the bottom of the plates using a rubber policeman, and the entire cell lysate was transferred to a QIAShredder spin column and centrifuged for 3 min at 14000 x g. The QIAShredder spin column consists of a biopolymer-shredding system that filtered out insoluble components and cell debris and homogenized the lysate completely. The flow-through, containing the RNA, was frozen at -80°C until further use.

For RNA extraction, the RNeasy Mini Kit was used. The sample was thawed on ice and an equal volume of 70% ethanol was added. The ethanol increases the binding affinity of the RNA to the RNeasy Mini spin column that consists of a silica membrane to which the sample was transferred. During the subsequent centrifugation step for 30 sec at 8000 x g, the RNA bound to the silica membrane, while other cellular components still present in the solution were discarded as flow-through. After washing with 350 µL of washing buffer, DNase-I-digestion

was performed on the column (70 µL RDD buffer + 10 µL DNase I) for 15 min at RT. DNase-I-digestion is performed to prevent contamination of RNA with genomic DNA. After washing several times to remove the last salt impurity from the column, RNA was eluted in 50 µL of ultrapure water (UPW) and frozen at -80°C until further use.

3.1.14 cDNA synthesis

cDNA synthesis describes the synthesis from an RNA template to complementary DNA (cDNA) by reverse transcription. For cDNA synthesis, oligo dT primer were used, that bind complementary to the poly-A tail of eukaryotic mRNA. 500 ng RNA were heated with 1 µL dNTPs and 1 µL Oligo dTs at 65°C for 5 min to prevent secondary structures of the RNA. Afterwards, the reaction was immediately incubated on ice and 6 µL UPW, 2 µL 10 x reverse transcriptase buffer, 1 µL RNase inhibitor and 1 µL Multiscribe reverse transcriptase were added. Finally, the reaction was incubated at 37°C for 2 hours and reheated for 5 min at 85°C. The cDNA was stored at -20°C until further use.

3.1.15 Quantitative Real-Time PCR

Quantitative real-time polymerase chain reaction (qRT-PCR) is a method for amplifying cDNA and is thus used for quantification of nucleic acid and gene expression. In this thesis, SYBR green I was used as fluorescent dye. Once bound, it intercalates in the double helix of cDNA and emits light at a wavelength of 520 nm. Thus, as the cDNA amplicons increase, the fluorescence signal also increases proportionally from cycle to cycle. Excitation of the fluorophores can be performed with halogen, laser or light-emitting diode (LED) light sources. The CFX Connect Real-Time PCR Detection System, used for this thesis, uses three light-emitting diodes (LEDs) with an excitation wavelength of 450-480 nm (excitation wavelength of SYBR Green: 497 nm). The SYBR Green Master Mix contains the appropriate SYBR Green I dye, dNTP’s, Ampli Taq Gold polymerase and the appropriate buffer for optimal reaction conditions.

The reaction mixture was as follows:

1 µL cDNA
5 µL SYBR Green Master Mix
0.02 µL forward Primer (100 µM)
0.02 µL reverse Primer (100 µM)
3.96 µL UPW

---

10 µL total volume

RT-PCR proceeds similar to conventional PCR (chapter 3.1.8.1.1), but a melting curve is recorded at the end of a run to check for the formation of non-specific products.

3.1.16 qRT-PCR Primer design

PCR primers for the qRT-PCR were designed with the software primer3 (Rozen and Skaletsky 2000). The amplicon length was chosen to be approximately between 70 bp and 150 bp and the annealing temperature of all primers was 60°C. If possible, primers were designed to span an exon-exon junction, with one primer spanning the exon-intron boarder. If this was not possible, primers were chosen to be located on different exons with a long intron in between. Especially the exon-spanning primers reduce the amplification of genomic DNA and thus false positive amplicons.

3.1.17 RNA sequencing

RNA sequencing (RNAseq) is a method to measure expression of transcriptome in a biological sample. RNA of transfected HeLa cells with the activated promoter of *CTD-2353F22.1* and the scrambled gRNA (each in triplicates) were extracted as described above. Of the transfected HeLa cells, 1000 ng total RNA (of each sample) 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). Quality control, differential gene expression, gene set enrichment and the hypergeometric test was performed as described in (Bao et al. 2021). Using Benjamini-Hochberg correction, the p-values of the differently expressed genes were corrected for multiple testing and the corrected p-values are given as p_{adj} values (false discovery rate [FDR]). RNAseq was performed and analyzed as described above at the Berlin Institute of Health Core Facility Genomics.

3.2 Identification and characterization of miRNA miR-374b-5p

3.2.1 Selection criteria of miRNAs

MiRNAs that showed significantly higher expression in the blood of PD patients compared to healthy controls are given in **Table 21**. To choose miRNAs for the identification of their target genes, miRNAs were selected for subsequent analysis with the following criteria (Mueller et al. 2021):

(I) Significant different expression in blood of PD patients compared to healthy controls with a p-value of $p < 5 \times 10^{-06}$

(II) Up- or downregulated with a FC ≥ 2 in venous blood of PD cases compared to healthy controls

(III) Published differentially upregulation (>2 fold) in at least two independent array-based miRNA expression profiling studies of cells that were exposed to inflammation or tobacco smoke metabolites (PD risk factors) or of human gingival tissues. The following PubMed search terms were used to identify appropriate studies in the literature: periodontitis, oral inflammation, miRNA, periodontal cells, microarray, oral mucosa, smoke and gingiva.

3.2.2 Isolation of human epithelial and gingival fibroblast cells

Human gingival fibroblasts and epithelial cells were isolated from gingival tissue. The tissue was washed for 30 sec in 80% EtOH, rinsed with Dulbecco‘s phosphate-buffered saline (DPBS) and incubated overnight at 4°C in dispase II (5 mg/mL) dissolved in DMEM medium (+ 1% pen/strep). Dispase II is a neutral protease that can be used for separating the gingival tissue in epithelial and fibroblast cells by cleaving type IV collagen and fibronectin. The next day, the gingival tissue was separated using forceps. The epithelial tissue was shaken in 5 mL Trypsin/EDTA at 37°C at 236 rpm for a maximum of 30 min. Afterwards, 5 mL DMEM medium (+ 1% pen/strep, + 5% FBS) was added and the solution containing the dissolved epithelial tissue was mixed with a 5-mL pipette for around eighty times. To remove clumps or debris and to isolate primary epithelial cells, the solution was added to a 70 µm cell strainer located in a 50-mL falcon and was processed using a cell strainer pestle. The cell strainer was washed with 20 mL DMEM medium (+ 1% pen/strep, + 5% FBS) and the falcon with the cells was centrifuged for 5 min at 300 x g. The supernatant was aspirated and the cell pellet was

resuspended in 5 mL DermaLife K, fully supplemented medium and transferred into a T25 cell culture flask coated with collagen A (1 mg/mL).

Connective tissue was cut with scalpels into small pieces and transferred into a T25 cell culture flask with 1 mL DMEM medium (+ 1% pen/strep, + 10% FBS, + 1% non-essential amino acids (NEA), 2 mM L-glutamine) to allow fibroblast cells to grow out. After around two weeks the fibroblasts were splitted into a T75 cell culture flask and were cultivated until they were used for transfection with miRNAs (Passage 3-4).

3.2.3 Cultivation of human primary gingival fibroblast cells

Primary gingival fibroblasts (pGFs) were cultivated at 37°C in a T75 flask in an incubator with 5% CO_2. Passaging of cells was performed under a sterile workbench once the cells reached approximately 90% confluence. For this purpose, the medium was removed from the confluent grown fibroblasts and the cells were washed twice with DPBS. DPBS was then removed and the human gingival fibroblast cells were incubated at 37°C for 5 min with 2 mL Trypsin/EDTA. The trypsinization was stopped by adding 10 mL culture medium (DMEM medium (+ 1% pen/strep, + 10% FBS, +1% non-essential amino acids (NEA), 2 mM L-glutamine)). Depending on the confluence of the cells, between 1 and 3 mL of cell suspension was transferred to a new T75 cell culture flask and incubated in the incubator until the cells again reached 90% confluence. For transfection, the pGFs (passage 3-4) were seeded in 6-well tissue culture plates (1.8×10^5 cells per well) one day prior to transfection.

3.2.4 Transfection of human gingival fibroblasts

Three independent pGFs from unrelated healthy donors were transfected in three technical replicates (N = 9) with three different mirVana miRNA mimics. For normalizing between samples, the negative control mirVana miRNA Mimic Negative Control #1 was used. As a positive control, mirVana miRNA mimic miR-1, that specifically down-regulates the expression of the gene Protein Tyrosine Kinase 9 (*PTK9*), was transfected into pGFs and to mimic miRNA hsa-miR-374b-5p, mirVana miRNA Mimic miR-374b-5p was used (Mueller et al. 2021). The day of the experiment, medium of 6-well tissue culture plates was changed before transfection. The miRNAs were transfected using Lipofectamine RNAiMAX reagent.

RNAiMAX is a cationic-lipid transfection reagent which uses lipofection to get genetic material (DNA, miRNA) into the cells. Transfection reactions for one 6-well were as follows:

1. Reaction:	9 µL Lipofectamine RNAiMAX reagent
	150 µL Opti-Mem Medium
2. Reaction:	3 µL mirVana miRNA mimic (30 pmol)
	150 µL Opti-Mem Medium

Both reactions were mixed by adding the second reaction to the first one and were incubated for 5 min at RT. Afterwards the miRNA-lipid complex was added dropwise to the cells. 24 hours after transfection, the cells were washed twice with PBS. Cell disruption and total RNA extraction was carried out as described in chapter 3.1.13. qRT-PCR (chapter 3.1.15) was performed to prove the downregulation of *PTK9* by the positive control.

3.2.5 Genome-wide gene expression analysis

RNA isolated from transfected pGF cells with mirVana miRNA mimic miR-374b-5p and mirVana miRNA Mimic Negative Control #1 was hybridized in biological triplicates (one sample of each donor) with the Human Clariom D Expression Array (Affymetrix) according to the manufacturer's instructions. Clariom D Expression Arrays were used to generate expression profiles and to identify different expressed genes and transcripts. The array covered > 540,000 transcripts of all known coding and non-coding splice variants of the transcribed human genome. Genome-wide gene expression analysis was performed by the sequencing platform at the Institute of Clinical Molecular Biology (IKMB) at the Kiel University. Partek Genomics Suite software (Partek) was used to analyze the expression data.

3.2.6 Prediction of microRNA targets

The online tool TargetScanHuman (version 7.1) (Agarwal et al. 2015) was used to predict biological targets of miRNAs. Therefore, the 3'UTR of mRNA sequences were analysed for the presence of conserved 8mer and 7mer sites that matched the seed region of miRNA hsa-miR-374b-5p.

3.2.7 Cloning of the 3'UTR UHMK1 into the pGL4.24 plasmid

In this thesis, 2055 bp of the 3' UTR of UHMK1, including two binding sites that match to the seed region of miRNA hsa-miR-374b-5p, were cloned in the expression vector pGL4.24 upstream of the luciferase reporter gene. To amplify the 3' UTR of UHMK1, the PCR reaction was done as described in chapter 3.1.8.1.1.

3.2.7.1 Restriction digestion and ligation

The PCR product was cloned into the pGL4.24 plasmid using the *Xba*I restriction enzyme.

The reaction mixture was as follows:

10 µL DNA (~ 2 µg PCR product/pGL4.24)
1 µL *Xba*I
5 µL 10 x CutSmart Buffer
34 µL UPW

---

50 µL total volume

The reaction was incubated at 37°C for 2 hours. The products were then loaded onto a 1% agarose gel and the DNA of the restricted plasmid and PCR product was isolated from the agarose gel as reported in chapter 3.1.8.1.4.

The ligation reactions for cloning the reporter gene plasmid were as follows:

1 µL pGL4.24 +*Xba*I (50 ng)
3 µL PCR product +*Xba*I (78 ng)
2 µL T4 ligation buffer
1 µL T4 ligase
13 µL UPW

---

20 µL total volume

After incubation of the ligation reaction at RT for 2 hours, the transformation in *E. coli* cells was performed as reported in chapter 3.1.8.1.6.

3.2.8 Transfection of HeLa cells with pGL4.24+3'UTR-*UHMK1*

24 hours after transfection of the HeLa cells with the reporter gene plasmid pGL4.24 + 3'UTR-UHMK1 (N = 9) using jetPEI as reported in chapter 3.1.10, cells were transfected with mirVana miRNA mimics in triplicates as described in chapter 3.2.4 for another 16 hours. Total RNA of cells was extracted as specified in 3.1.13.

3.2.9 DNase-I-digestion

At the end of RNA extraction, an additional DNase-I-digestion was performed to completely ensure no genomic DNA contamination in the RNA. For maximum activity, DNase-I requires divalent cations (Mg^{2+} and/or Ca^{2+}) and an optimal pH of 7.8. These optimal conditions were achieved by the addition of the 10 x reaction buffer (200 mM Tris-HCl, 500 mM KCl, 10 mM MnCl2, pH 8.3). Addition of a RNase inhibitor, an acidic protein of approximately 52 kDa that forms a 1:1 complex with RNase A, B and C, inhibited the RNase and protected the RNA from degradation.

The reaction was as follows:

30 µL	RNA (2-3 µg)
5 µL	10 x reaction buffer
1 µL	DNase-I
1 µL	RNaseOut Recombinant Ribonuclease Inhibitor
13 µL	UPW
50 µL	total volume

This reaction was incubated for 45 min at 37°C.

Subsequently, the RNA was precipitated with isopropanol and purified. For this purpose, 1 µL GlycoBlue, 5 µL sodium acetate (3M, pH 5.2) and 50 µL isopropanol were added to the

reaction, mixed and then incubated on ice for 15 min to precipitate the RNA. After the reaction was centrifuged for 10 min at 4 °C and 14000 x g, the supernatant was removed and the pellet was washed with 1 mL 70% EtOH. The sample was centrifuged for 5 min at 4 °C and 14000 x g and the alcohol was removed. The pellet was air dried and resuspended in 20 µL UPW.

DNase-I digest was performed until no more genomic DNA contamination in the RNA could be amplified by PCR (chapter 3.1.8.1.1).

Transcript levels of the Luciferase gene and GAPDH were quantified by qRT-PCR (chapter 3.1.15).

3.2.10 Luciferase-based reporter assay of 3'UTR of *UHMK1* and miRNA miR-374b-5p

To validate the effect of the miR-374b-5p on the protein level, HeLa cells were transfected as described in 3.1.10 and 24 hours after miRNA transfection, the effect was measured with the Dual luciferase reporter gene assay as reported in chapter 3.1.11.

3.2.11 Statistical analysis

All qRT-PCR experiments and luciferase reporter gene assays were performed with three independent biological replicates. p-values were calculated using two-tailed Student's-t-test. Graphs of these experiments were generated using GraphPad Prism software. Stars in the diagrams show significance level.

4 Results

4.1 Identification of putative causal variant(s) underlying the associations at *SIGLEC5* and *CTD-2353F22.1*

In **Figure 8** the workflow of the study is shown.

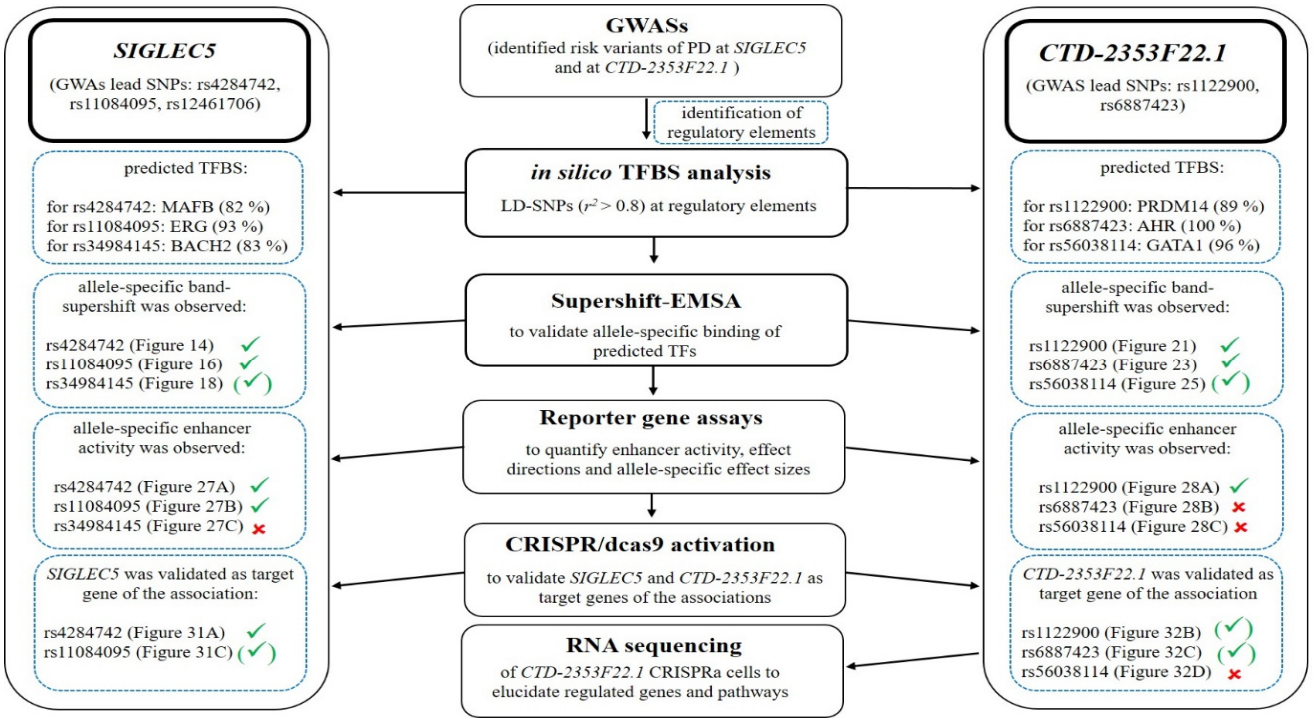


Figure 8: Workflow of the study. GWASs identified two associated regions at *SIGLEC5* and two associated regions at *CTD-2353F22.1* to increase the risk for periodontitis. GWAS lead SNPs and all SNPs in strong LD ($r^2 > 0.8$) were mapped to DNA elements with predictive features of regulatory functions and were screened for TFBS. For rs4284742, rs11084095, rs34984145, rs1122900, rs6887423 and rs56038114, given TFBSs were predicted. Matrix similarities are given in brackets. EMSA showed that all predicted TFs bind allele-specific to the corresponding SNPs. Ticks are given in brackets for rs34984145 and rs56038114 because TFs bind contrary to the alleles that were predicted. Allele-specific enhancer activity was validated for rs4284742, rs11084095 and rs1122900 using luciferase reporter gene assay. Using CRISPR-dCas9, the enhancer at rs4284742 showed strong activation of *SIGLEC5* expression, validating this gene as the target gene of the association. The enhancers at rs11084095 and rs1122900 demonstrated moderately activated *SIGLEC5* and *CTD-2353F22.1* expression. RNA sequencing of *CTD-2353F22.1* CRISPRa cells indicated significant upregulation of the gene sets “angiogenesis” and “TNFalpha signaling via NFκB”.

4.1.1 Assignment of putative causal associated variants at *SIGLEC5* by LD and predictive regulatory features

The two periodontitis associated haplotype blocks at *SIGLEC5* encompassed few co-segregating SNPs (**Figure 9 A-B**). GWAS lead SNP rs4284742 (Munz et al. 2017) had no other variant in strong LD ($r^2 > 0.8$) (**Table 18**). The associated haplotype block that was tagged by the lead SNPs of the GWAS meta-analyses, rs11084095 (Munz et al. 2019) and rs12461706 (Shungin et al. 2019), comprised a total of five SNPs in strong LD. These included the two GWAS lead SNPs and three additional tagging SNPs (**Table 18**).

Table 18: SNPs in LD ($r^2 > 0.8$) with the GWAS lead SNPs rs4284742, rs12461706 and rs11084095.

SNP ID	Location (GRCh37/hg19)	common allele	minor allele	MAF	*r2* with rs12461706	GWAS lead SNP
rs4284742	chr19:52131733	G	A	0.25	0.19	(Munz et al. 2017)
rs12461706	chr19:52121235	A	T	0.40	1.00	(Shungin et al. 2019)
rs11084095	chr19:52127030	G	A	0.40	1.00	(Munz et al. 2019)
rs11880807	chr19:52120522	A	G	0.30	0.81	
rs34984145	chr19:52120410	A	T	0.43	0.81	
rs4801882	chr19:52127053	G	A	0.44	0.81	

MAF = minor allele frequency, indicated for CEU (MAFs for human populations with other ethnic background were similar to North-West Europeans). The lead SNPs are highlighted in bold letters.

Of the six SNPs that tagged both haplotype blocks, only rs4284742 mapped to regulatory DNA elements as determined by DNAse I hypersensitivity (DHS) and transcription factor binding site (TFBS) experimentally confirmed by ChIP-Seq. rs11084095, rs34984145 and rs11880807 located to a region that showed H3K4Me1 methylation in the B cell line GM12878, a methylation mark that is enriched at active and primed cell type specific enhancers. Both haplotype blocks were separated from one another by an insulator element (**Figure 10**).

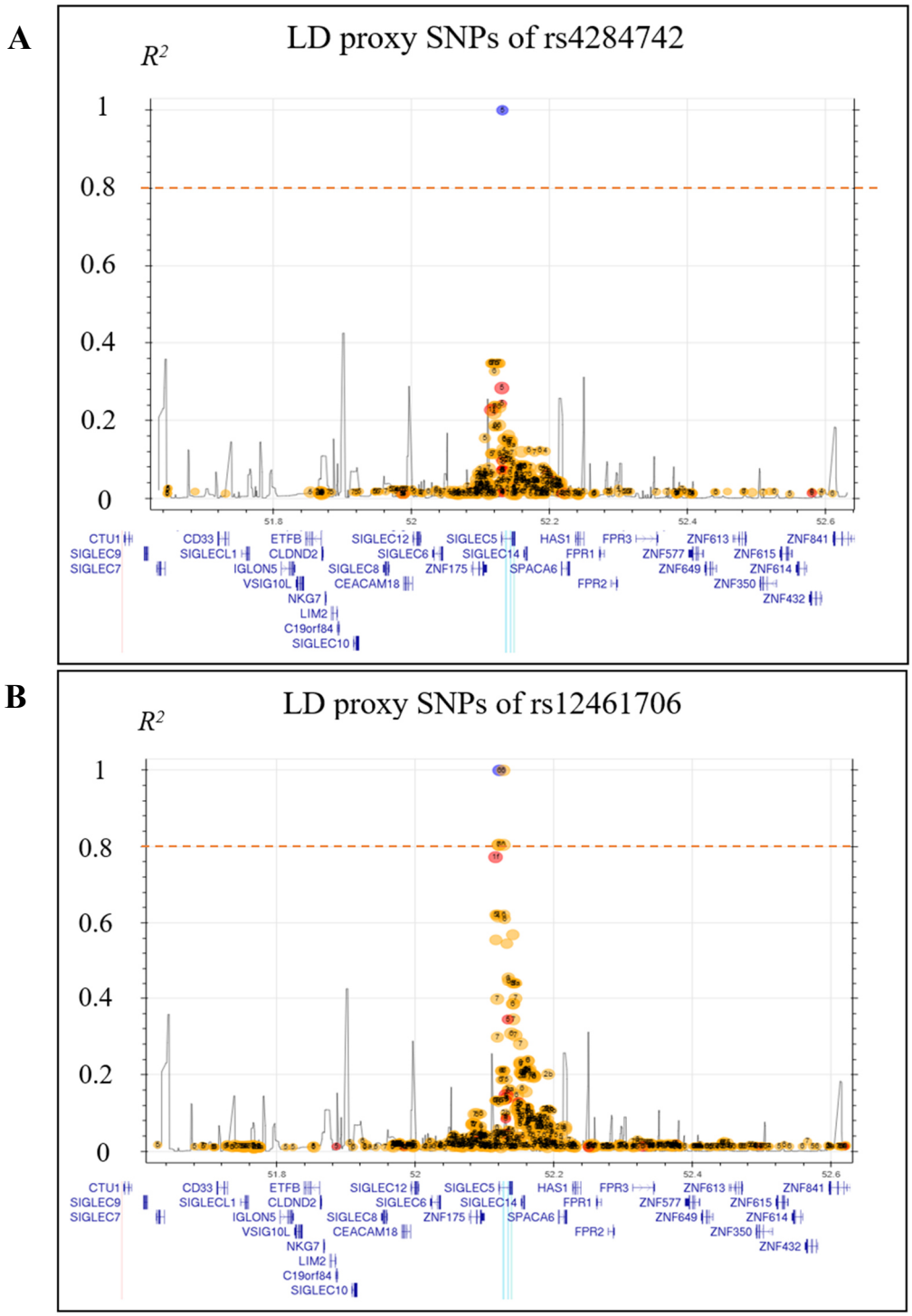


Figure 9: Linkage and chromosomal positions of the SNPs that showed genome-wide significant associations at *SIGLEC5* with increased periodontitis susceptibility. (A) rs4284742 is not in strong LD with other SNPs ($r^2 > 0.8$). **(B)** rs12461706 shows strong LD with four SNPs ($r^2 > 0.8$). The plots were generated with rs12461706 as index SNP, genotype data of the 1000genomes populations CEU and GBR in a 500,000 base pairs window (LDproxy Tool). The dashed horizontal line indicates $r^2 = 0.8$. The SNPs were aligned to their chromosomal positions (x-axis). The blue vertical line shows the position of the LD SNPs at $r^2 > 0.8$ (rs3829655 missed the LD threshold with $r^2 = 0.77$. This SNP is a benign exonic synonymous SNP and was not included in the TFBS analysis.

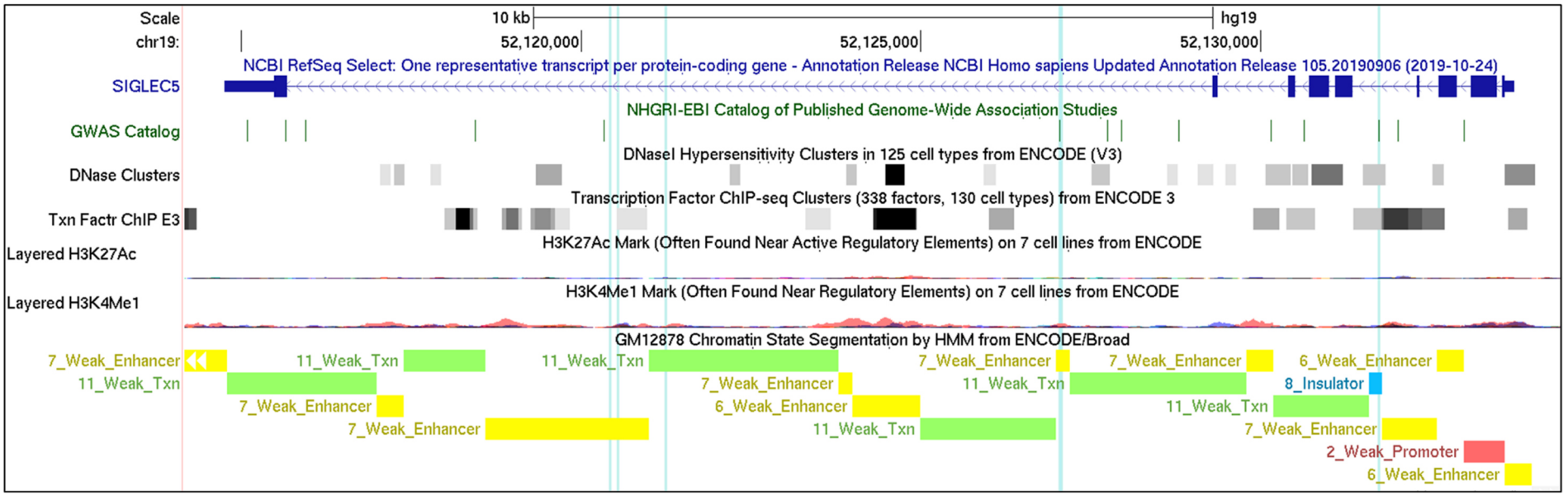


Figure 10: Chromosomal positions of the SNPs that showed genome-wide significant associations at *SIGLEC5* with increased periodontitis susceptibility. rs4284742 locates at chromatin elements that correlate with regulatory functions of gene expression (open chromatin as determined by DNAse I hypersensitivity (DHS) and transcription factor binding sites (TFBS) experimentally confirmed by ChIP-Seq), determined in various cell types (data from ENCODE). Chromatin segmentation in B lymphocyte cells (GM12878) indicated enhancer elements at rs34984145, rs11880807 and rs11084095 (The GWAS lead SNPs rs12461706, rs11084095, rs4284742 and their LD-SNPs ($r^2 > 0.8$) are highlighted with blue vertical lines. SNP order from left to right: rs34984145, rs11880807, rs12461706, rs11084095, rs4801882 and rs4284742. Because rs11084095 and rs4801882 have a distance of 24 bp, they are indicated as one line). Figure was generated with UCSC Genome Browser.

4.1.2 Assignment of putative causal associated variants at *CTD-2353F22.1* by LD and predictive regulatory features

The periodontitis associated haplotype block at *CTD-2353F22.1* tagged by the GWAS lead-SNPs encompassed eight co-segregating SNPs (**Figure 11A-B**). The associated haplotype block that was tagged by the GWAS lead SNP rs1122900 (Munz et al. 2017) comprised a total of four SNPs in strong LD ($r^2 > 0.8$). These included the sentinel variant and three additional tagging SNPs (**Table 19**). The associated haplotype block that was tagged by the GWAS lead SNP rs6887423 (Munz et al. 2017) comprised also a total of four SNPs in strong LD. These included the sentinel variant and three additional tagging SNPs (**Table 19**). Global allele frequencies in **Table 19** are taken from the 1000 Genomes Project Phase 3 (Ensembl genome browser (Yates et al. 2020)).

Table 19: SNPs in LD ($r^2 > 0.8$) with the GWAS lead SNPs rs1122900 and rs6887423.

SNP ID	Location (GRCh37/hg19)	common allele	minor allele	MAF (global)	*r2* with rs1122900	GWAS lead SNP
rs1122900	chr5:36689181	A	C	0.36	1.00	(Munz et al. 2017)
rs56038114	chr5:36683801	T	C	0.36	0.96	
rs56039629	chr5:36683903	C	T	0.36	0.89	
rs6862950	chr5:36681820	T	C	0.41	0.85	
rs6887423	chr5:36696501	T	C	0.40	0.67	(Munz et al. 2017)
rs17585785	chr5:36695331	A	G	0.43	0.67	
rs56162483	chr5:36710830	G	C	0.33	0.63	
rs56066032	chr5:36695601	G	A	0.39	0.53	

Of the eight SNPs that tagged both haplotype blocks, rs1122900, rs56038114, rs6887423 and rs56039629 mapped to regulatory DNA elements as determined by DNAse I hypersensitivity (DHS) and transcription factor binding site (TFBS) experimentally confirmed by ChIP-Seq. rs1122900, rs56038114, rs6887423 and rs56039629 are also located to a region that showed H3K4Me1 methylation and H3K27Ac acetylation which is a reliable predictive feature of active gene regulation and active enhancers (Creyghton et al. 2010). rs1122900 and rs6887423 are located in insulator elements in the B cell line GM12878 (**Figure 12**).

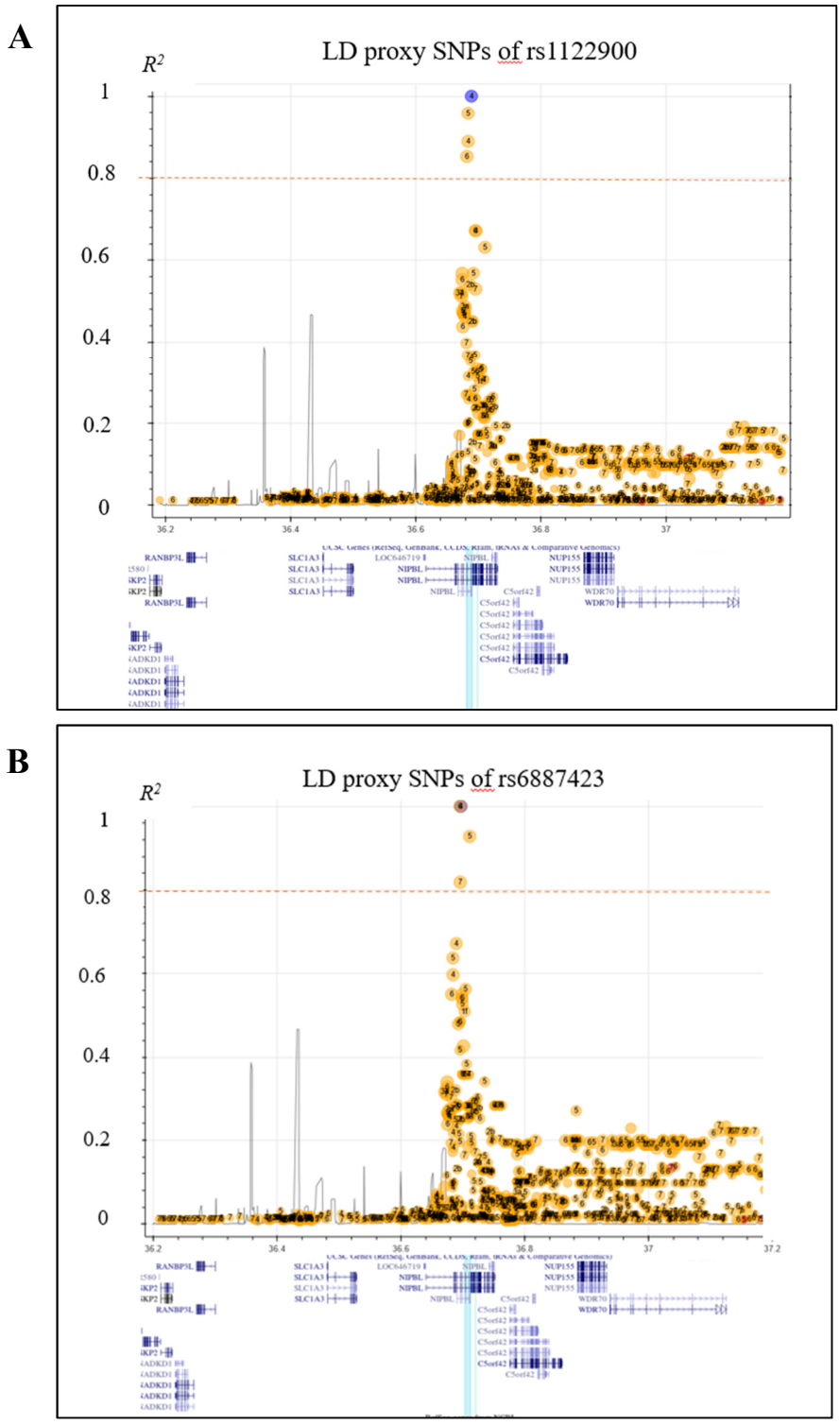


Figure 11: Linkage and chromosomal positions of the SNPs that showed genome-wide significant associations at *CTD-2353F22.1* with increased periodontitis susceptibility. **(A)** rs1122900 shows strong LD with three SNPs ($r^2 >$ 0.8). **(B)** rs6887423 shows strong LD with three SNPs ($r^2 > 0.8$). The plots were generated with rs1122900 as index SNP, genotype data of the 1000genomes populations CEU and GBR in a 500,000 base pairs window (LDproxy Tool). The dashed horizontal line indicates $r^2 = 0.8$. The SNPs were aligned to their chromosomal positions (x-axis). The blue vertical line shows the position of the LD SNPs at $r^2 > 0.8$.

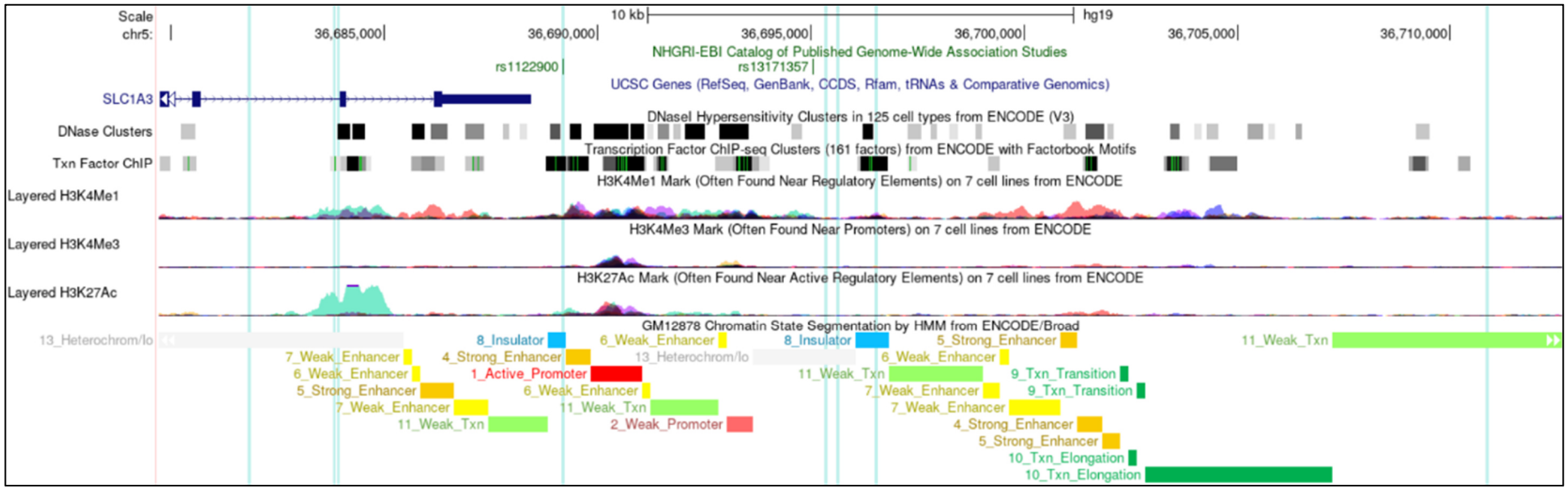


Figure 12: Chromosomal positions of the SNPs that showed genome-wide significant associations at *CTD-2353F22.1* with increased periodontitis susceptibility.

rs1122900, rs56038114, rs6887423 and rs56039629 are located at chromatin elements that correlate with regulatory functions of gene expression (open chromatin as determined by DNAse I hypersensitivity (DHS) and transcription factor binding sites (TFBS) experimentally confirmed by ChIP-Seq), determined in various cell types (data from ENCODE). The GWAS lead SNPs rs1122900, rs6887423 and their LD-SNPs ($r^2 > 0.8$) are highlighted with blue vertical lines. SNP order from left to right: rs6862950, rs56038114, rs56039629, rs1122900, rs17585785, rs56066032, rs6887423 and rs56162483. Figure was generated using UCSC Genome Browser.

4.1.3 The TFs MAFB and ERG show allele-specific binding at rs11084095 and rs4284742

It was tested whether the nucleotide variants of the six SNPs changed predicted TFBS. For the common allele of rs4284742, a TFBS for the TF MAF bZIP transcription factor B (MAFB) with a matrix similarity of 82% was predicted (**Figure 13A**), for the common allele of rs11084095 a TFBS for the ETS transcription factor ERG with a matrix similarity of 93% was predicted (**Figure 13B**), and for the common allele of rs34984145, a TFBS for BACH2 with a matrix similarity of 83% was predicted with SNPInspector (**Figure 13C**). PWMTools confirmed the high similarity with those TF binding motifs at these SNPs. For the non-effect alleles at all other SNPs, no TFBS was predicted (chapter 8.1 (**Table 27**)). The three SNPs rs4284742, rs11084095 and rs34984145 were selected for subsequent validation of TF binding *in vitro*.

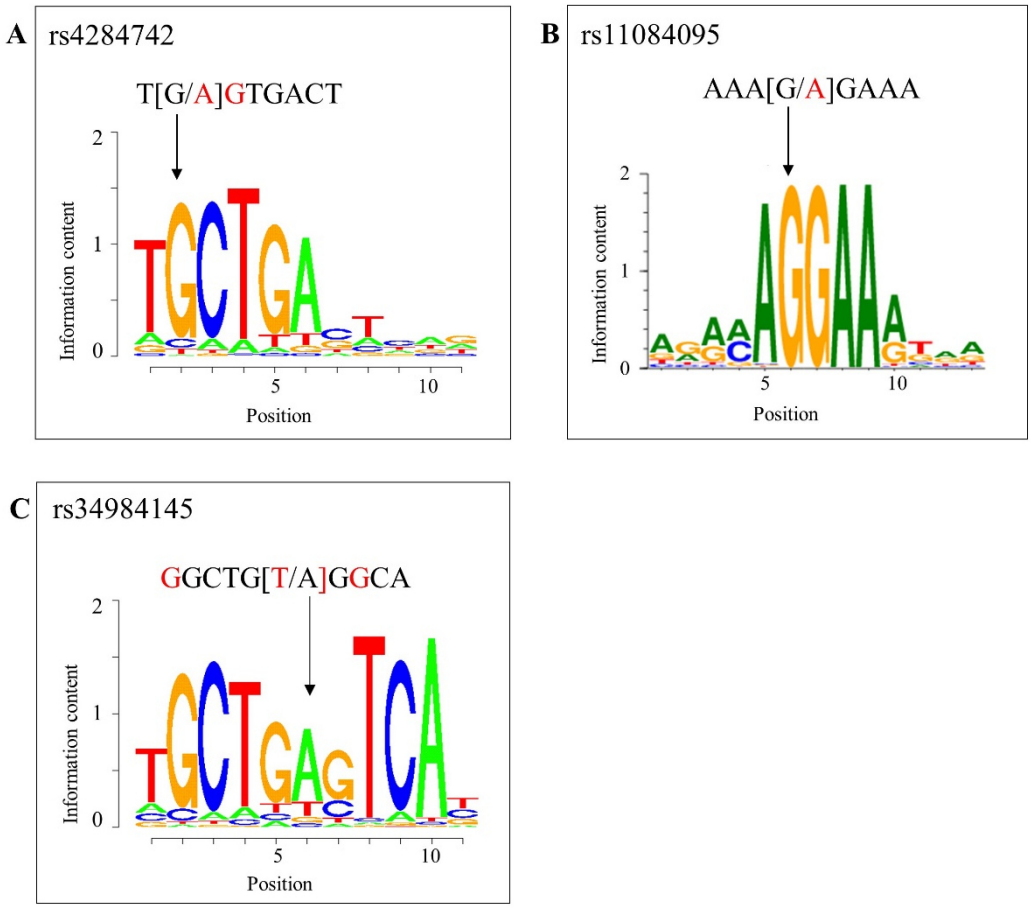


Figure 13: PWMs of TFBS for ERG, MAFB and BACH2. (A) The MAFB TF binding motif shows a matrix similarity of 82% with the DNA sequence at the common G-allele of rs4284742. **(B)** The TF binding motif of ERG has a matrix similarity of 93% with the DNA sequence at the common G-allele of rs11084095. **(C)** The BACH2 TF binding motif shows a matrix similarity of 83% with the DNA sequence at the common A-allele of rs34984145. The SNP alleles are shown in brackets. Letters that do not match to the PWM are given in red.

To prove protein binding at the three selected SNPs and to provide evidence for the specificity of binding of the predicted TFs, EMSAs with MAFB, ERG and BACH2 specific antibodies with allele-specific DNA probes were performed. *SIGLEC5* is mainly expressed in various lymphocytes of the innate immune system and in B cells (Crocker and Varki 2001). MAFB is mainly expressed in whole blood (including B lymphocytes), which is why the EMSA for rs4284742 was performed with protein extract from Raji cells. An allele-specific band-supershift was observed with the TF-specific antibody MAFB and allele-specific oligonucleotides. Probes with the non-effect allele (A-allele) abrogated antibody binding compared to probes with the common and effect G-allele (**Figure 14**).

Allele-specific shifted bands intensities indicated that the background of the rare A-allele of rs4284742 reduced MAFB binding to 69% compared to the common G-allele (**Figure 14 and 15**).

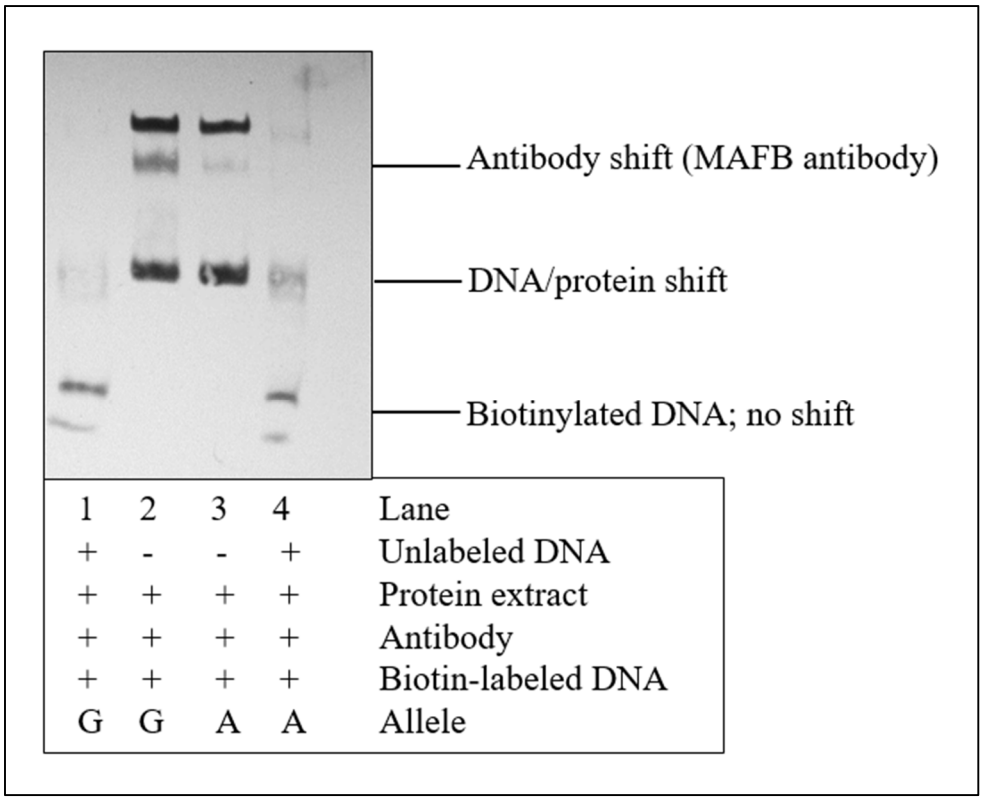


Figure 14: MAFB-antibody EMSA performed with rs4284742 allele-specific oligonucleotide probes and nuclear protein extract from Raji cells. Binding of MAFB-antibody to allele-specific probes is shown in lane 2 and 3. Probes with the non-effect allele (A-allele) abrogated antibody binding compared to probes with the common and effect G-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

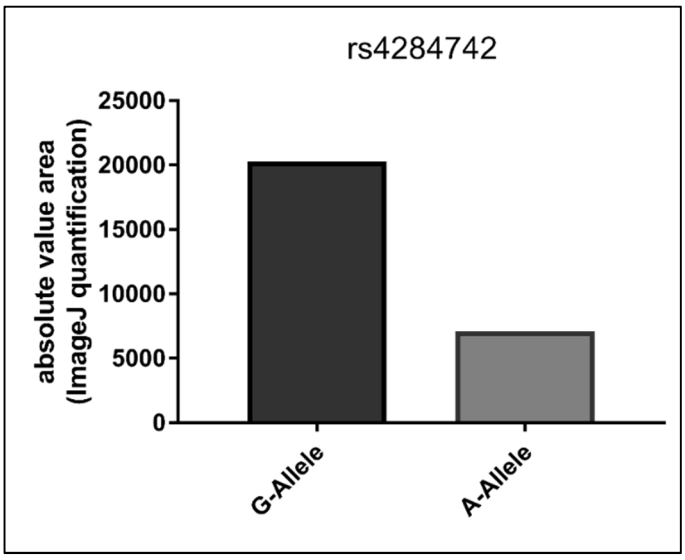


Figure 15: Absolute value area of the antibody specific bands of MAFB binding at rs4284742. In the background of the rare A-allele of rs4284742, MAFB binding to the allele-specific oligonucleotides was 69% reduced compared to the common G-allele.

To prove protein binding at rs11084095 and to give evidence for the specificity of the binding of the predicted TF ERG which is weakly expressed in PBMCs, the EMSA for rs11084095 was performed with protein extract from these cells (**Figure 16**). An allele-specific band-supershift was observed with the TF-specific antibody ERG and allele-specific oligonucleotides. The probes with the effect allele (A-allele) abrogated antibody binding compared to the probe with the common G-allele. The intensity of the allele-specific shifted bands was analyzed by calculating the absolute value area using ImageJ (**Figure 17**). In the background of the rare A-allele of rs11084095, ERG binding to the rare A-allele specific oligonucleotides was 99% reduced compared to the common G-allele.

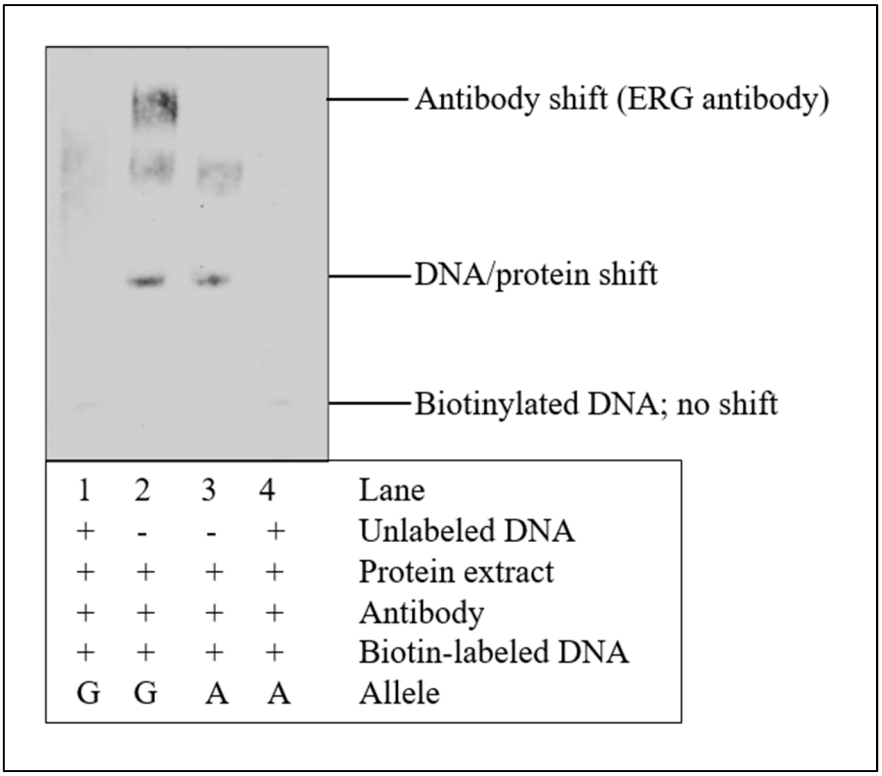


Figure 16: ERG-antibody EMSA performed with rs11084095 allele-specific oligonucleotide probes and nuclear protein extract from PBMCs. Binding of ERG-antibody to allele-specific probes is shown in lane 2 and 3. Probes with the effect allele abrogated antibody binding compared to probes with the common G-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

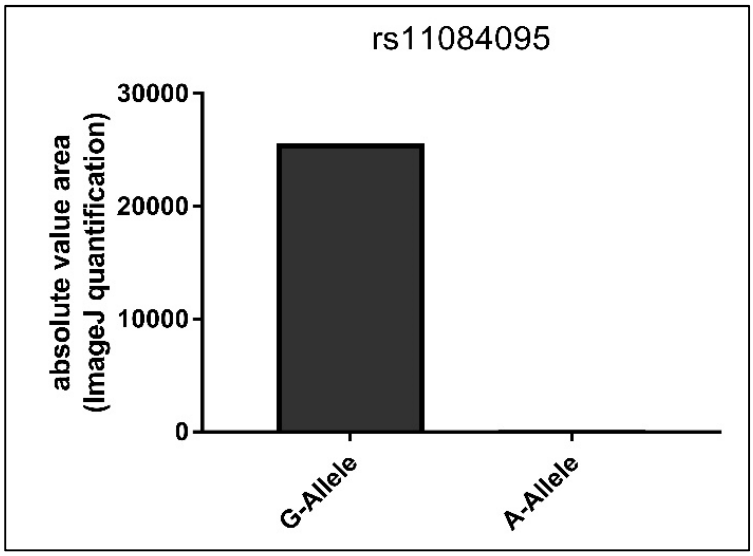


Figure 17: Absolute value area of the antibody specific bands of ERG binding at rs11084095. In the background of the rare A-allele of rs11084095, ERG binding to the rare A-allele specific oligonucleotides was 99% reduced compared to the common G-allele.

To prove protein binding at rs34984145 and to give evidence for the specificity of binding of the predicted TF BACH2 which is expressed in B cells, an EMSA using BACH2 specific antibody with allele-specific DNA probes and protein extract from Raji cells was performed. An allele-specific band-supershift was observed with the TF-specific antibody BACH2 and allele-specific oligonucleotides. However, contrary to what was predicted in PWM of BACH2, a weaker band was observed for probes with the common and non-effect A-allele compared to probes with the rare and effect T-allele (**Figure 18**). At rs34984145, BACH2 binding showed a reduction of TF binding by 59% at the common A-allele compared to the rare T-allele (**Figure 18 and 19**).

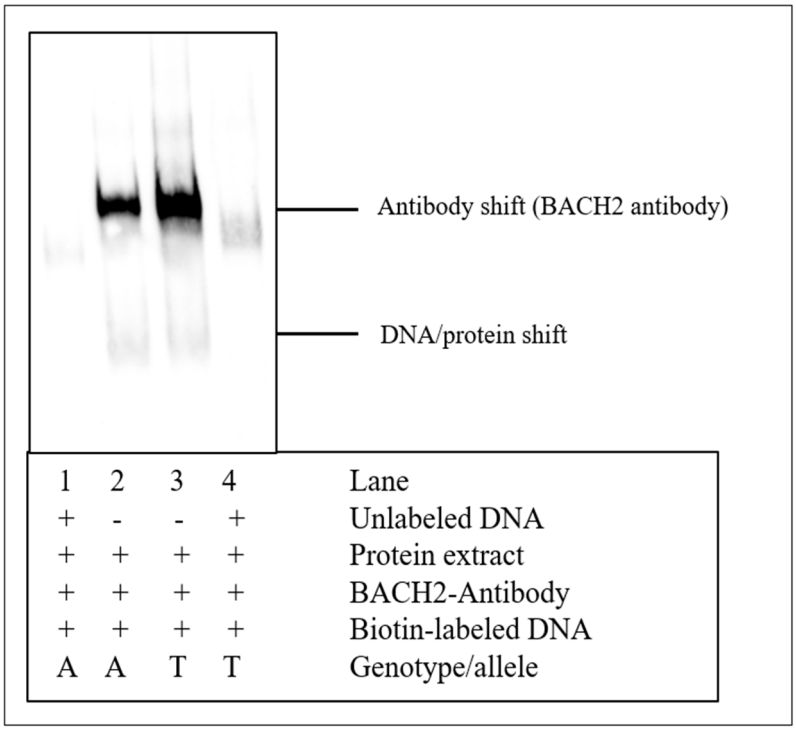


Figure 18: BACH2-antibody EMSA performed with rs34984145 allele-specific oligonucleotide probes and nuclear protein extract from Raji cells. Binding of BACH2-antibodies to allele-specific probes is shown in lane 2 and 3. Probes with the non-effect allele abrogated antibody binding compared to probes with the common T-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

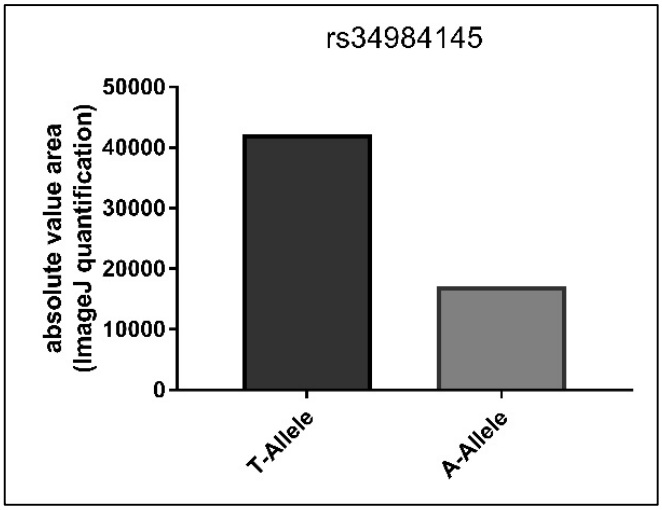


Figure 19: Absolute value area of the antibody specific bands of BACH2 binding at rs34984145. BACH2 binding to the allele-specific oligonucleotides of rs34984145 was 59% reduced for the common A-allele compared to the rare T-allele.

4.1.4 The TFs PRDM14 and AHR show allele-specific binding at rs1122900 and rs6887423

It was tested whether the nucleotide variants of the eight SNPs changed predicted TFBS. For the common allele of rs1122900 a TFBS for the TF PR Domain Zinc Finger Protein 14 (PRDM14) with a matrix similarity of 89% was predicted (**Figure 20A),** for the common allele of rs6887423, a TFBS for the TF Aryl Hydrocarbon Receptor (AHR) with a matrix similarity of 100% was predicted (**Figure 20B**), and for the common allele of rs56038114, a TFBS for GATA binding protein 1 (GATA1) with a matrix similarity of 96% was predicted (**Figure 20C**). PWMTools confirmed the high similarity to the TF binding motifs at these SNPs. For the non-effect alleles at all other SNPs, no TFBS was predicted (**Appendix** (**Table x**). The three SNPs rs1122900, rs6887423 and rs56038114 were used for subsequent validation of TF binding *in vitro*.

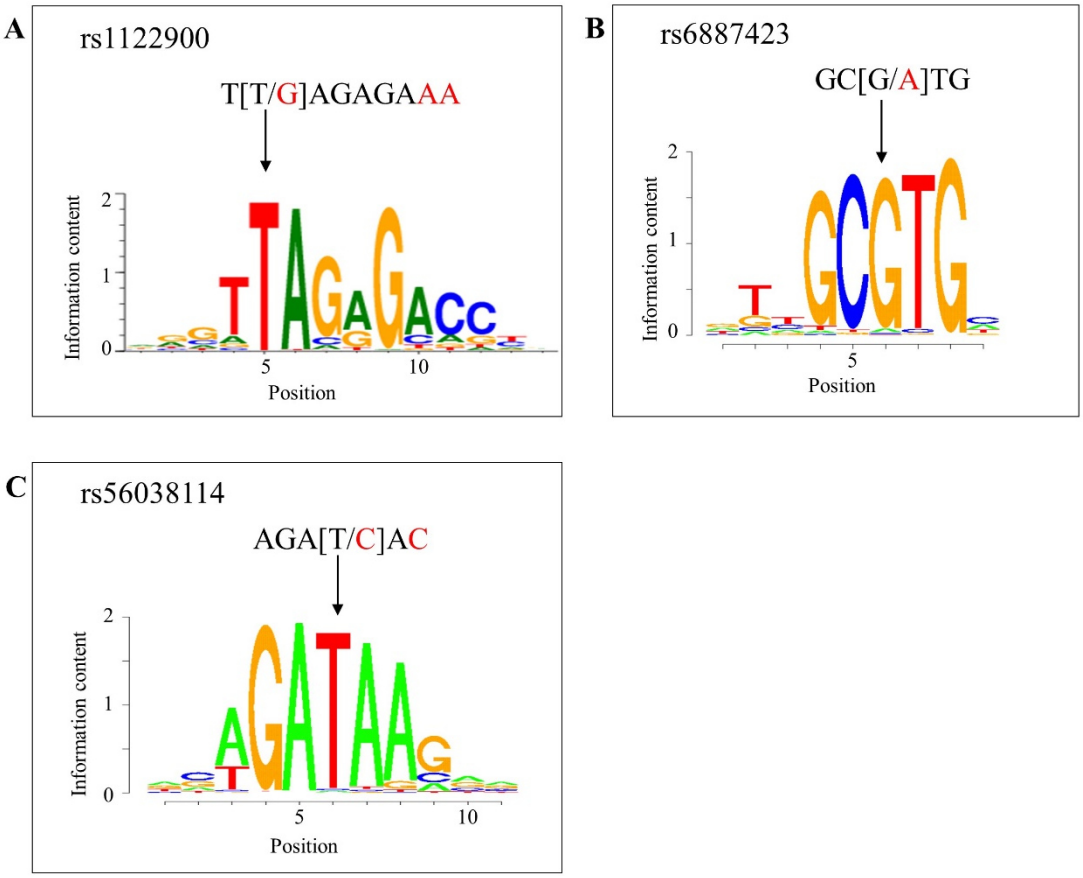


Figure 20: PWMs of TFBS for PRDM14, AHR and GATA1. (A) The TF binding motif of PRDM14 has a matrix similarity of 89% with the DNA sequence at the common G-allele of rs1122900. (B) The AHR TF binding motif shows a matrix similarity of 100% with the DNA sequence at the common G-allele of rs6887423. (C) 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.

To prove protein binding at the three selected SNPs and to give evidence for the specificity of binding of the predicted TFs, an EMSA was performed using PRDM14, AHR, and GATA1 specific antibodies with allele-specific DNA probes. *CTD-2353F22.1* is mainly expressed in B cells. PRDM14 is weakly expressed in B cells, which is why the EMSA for rs1122900 was performed with protein extract from Raji cells. An allele-specific band-supershift was observed with the TF-specific antibody PRDM14 and allele-specific oligonucleotides (**Figure 21**). Probes with the effect allele (C-allele) abrogated antibody binding compared to probes with the common A-allele.

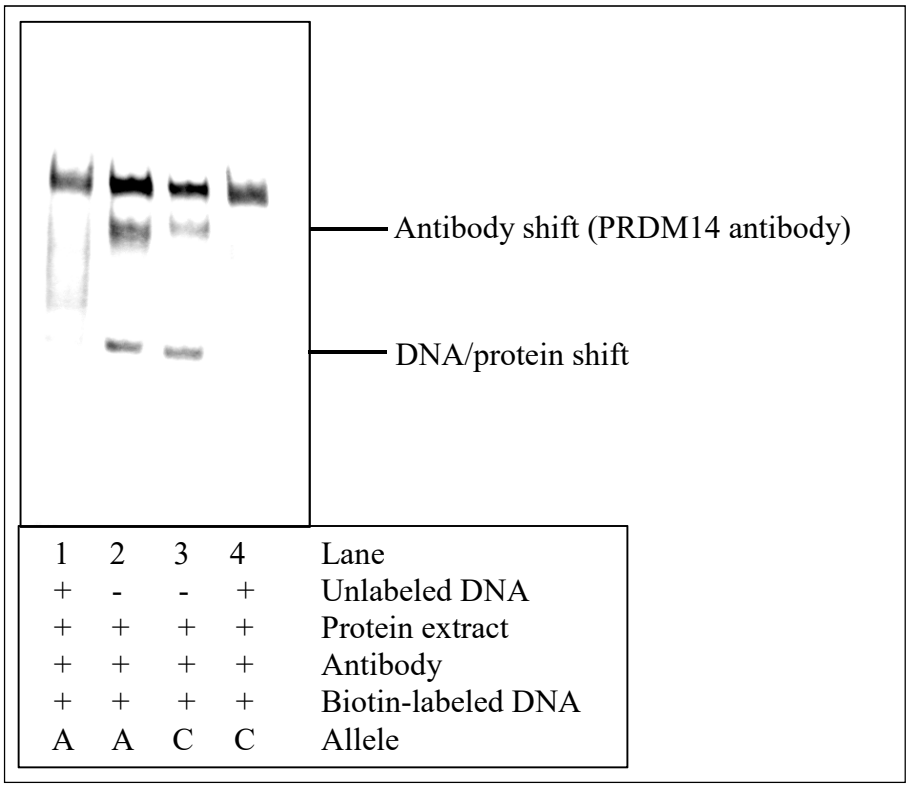


Figure 21: PRDM14-antibody EMSA performed with rs1122900 allele-specific oligonucleotide probes and nuclear protein extract from Raji cells. Binding of PRDM14-antibody to allele-specific probes is shown in lane 2 and 3. Probes with the effect allele abrogated antibody binding compared to probes with the common A-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

The intensity of the allele-specific shifted bands was analyzed by calculating the absolute value area using ImageJ (**Figure 22**). The background of the rare C-allele of rs1122900 reduced PRDM14 binding to 76% of binding in the background of the common A-allele (**Figure 21 and Figure 22**).

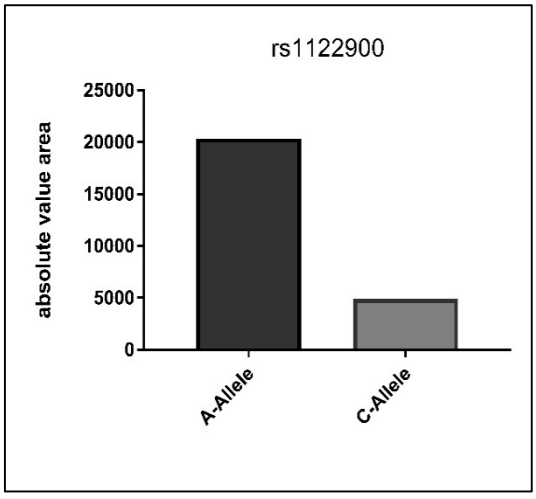


Figure 22: Absolute value area of the antibody specific bands of PRDM14 binding at rs1122900. In the background of the rare C-allele of rs1122900, PRDM14 binding to the rare C-allele specific oligonucleotides was 76% reduced compared to the common A-allele.

AHR is, like *CTD-2353F22.1*, weakly expressed in B cells, which is why the EMSA for rs6887423 was performed with protein extract from Raji cells. An allele-specific band-supershift was observed with the TF-specific antibody AHR and allele-specific oligonucleotides. Probes with the effect allele (T-allele) abrogated antibody binding compared to probes with the common C-allele (**Figure 23**).

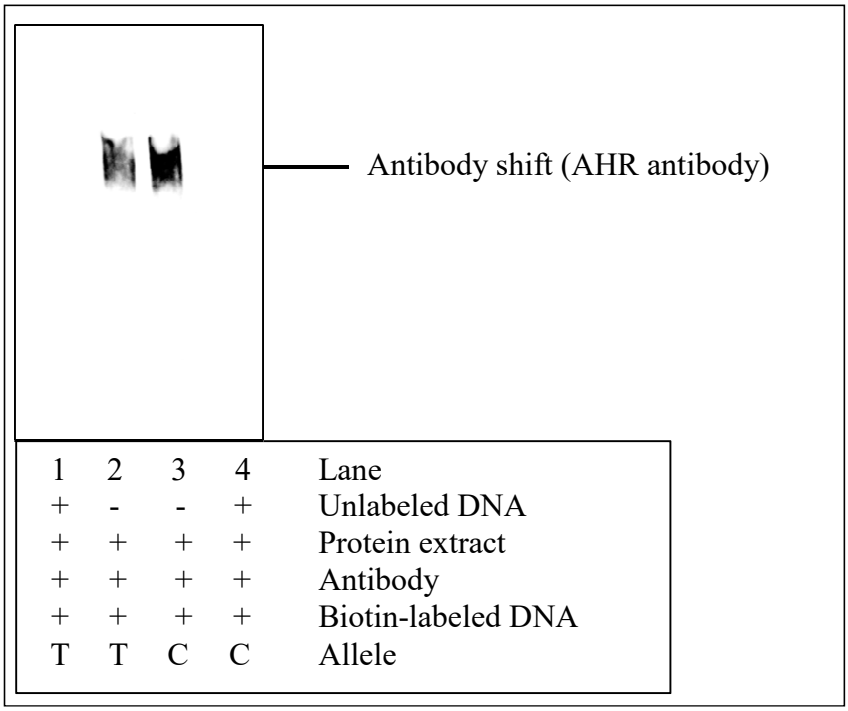


Figure 23: AHR-antibody EMSA performed with rs46887423 allele-specific oligonucleotide probes and nuclear protein extract from Raji cells. Binding of AHR-antibody to allele-specific probes is shown in lane 2 and 3. Probes with the effect allele (T-allele) abrogated antibody binding compared to probes with the common C-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

The absolute value area using ImageJ shows that the background of the rare T-allele of rs6887423 reduced AHR binding to 32% of binding in the background of the common C-allele (**Figure 23 and Figure 24**).

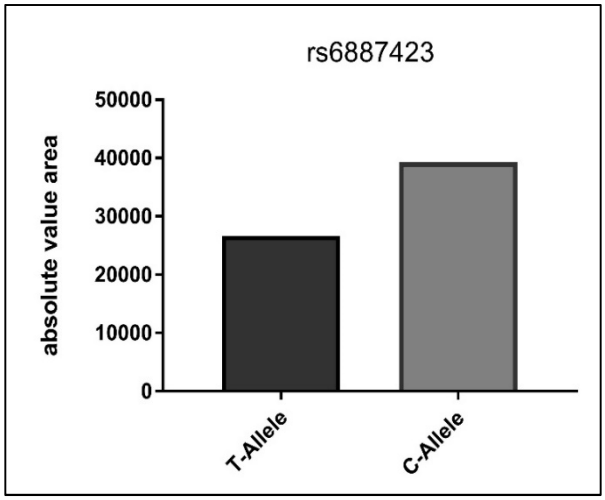


Figure 24: Absolute value area of the antibody specific bands of AHR binding at rs6887423. In the background of the rare T-allele of rs6887423, AHR binding to the rare T-allele specific oligonucleotides was 32% reduced compared to the common C-allele.

GATA1 is weakly expressed in PBMCs, which is why the EMSA for rs56038114 was performed with protein extract from PBMCs extracted as described in chapter 3.1.5. An allele-specific band-supershift was observed with the TF-specific antibody AHR and allele-specific oligonucleotides (**Figure 25**). However, contrary to what was predicted in PWM of GATA1, a weakening of binding was observed for probes with the common and non-effect T-allele compared to probes with the rare and effect C-allele (**Figure 25**). At rs56038114, GATA1 binding showed a reduction of TF binding by 73% at the common T-allele compared to the rare C-allele (**Figure 25 and Figure 26**).

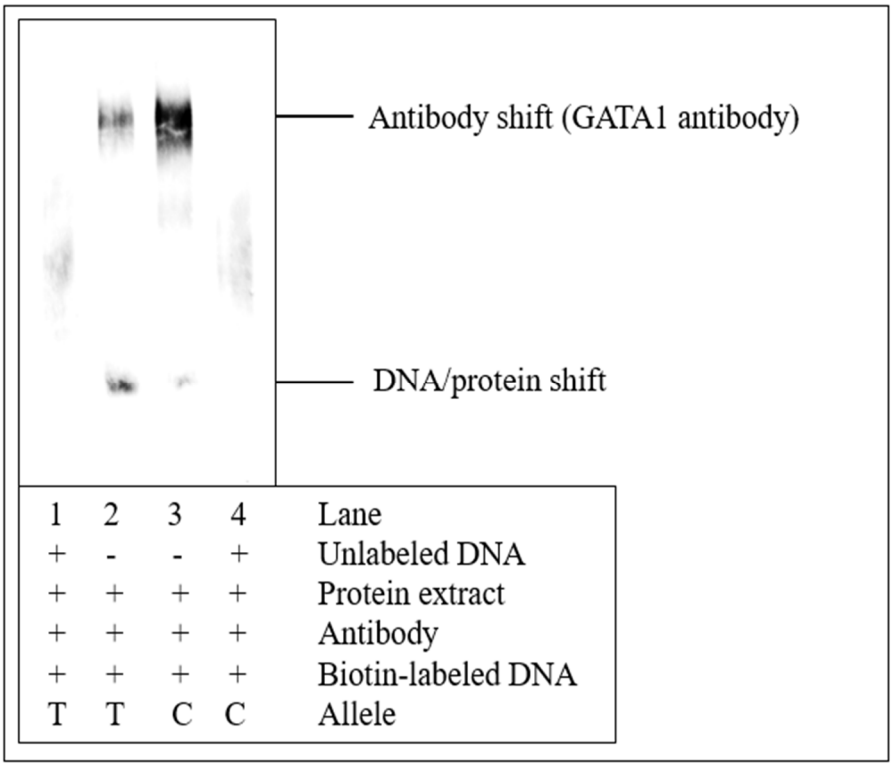


Figure 25: GATA1-antibody EMSA performed with rs56038114 allele-specific oligonucleotide probes and nuclear protein extract from PBMCs. Binding of GATA1-antibody to allele-specific probes is shown in lane 2 and 3. Probes with the effect allele (T-allele) abrogated antibody binding compared to probes with the common C-allele. Unlabeled DNA was added to verify that the band-shift was antibody specific in lane 1 and 4.

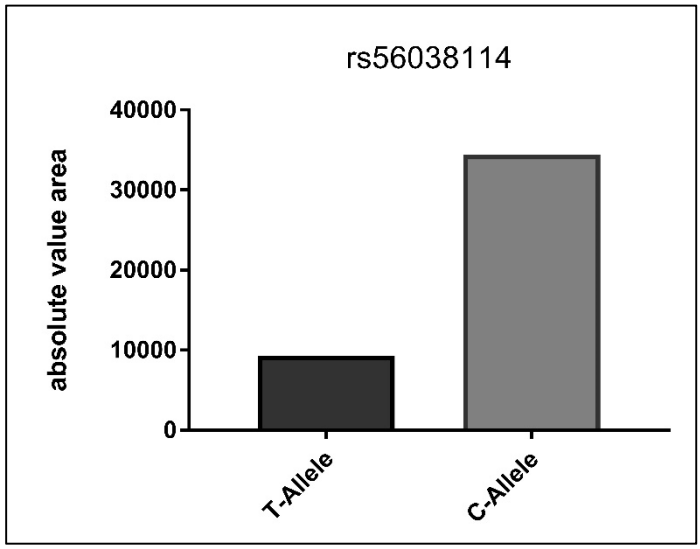


Figure 26: Absolute value area of the antibody specific bands of GATA1 binding at rs56038114. In the background of the common T-allele of rs56038114, GATA1 binding to the common T-allele specific oligonucleotides was 73% reduced compared to the rare C-allele.

4.1.5 Validation of effect direction and allele-specific effect sizes of rs11084095, rs4284742 and rs34984145 at *SIGLEC5* and rs1122900, rs6887423 and rs56038114 at *CTD-2353F22.1*

Regulatory DNA elements can either repress or activate transcription. To measure the activity of the regulatory elements at rs4284742, rs11084095 and rs34984145 at *SIGLEC5*, and at rs1122900, rs6887423 and rs56038114 at *CTD-2353F22.1*, and to discriminate their effect directions and allele-specific effect sizes, luciferase reporter gene assays were employed.

4.1.5.1 rs4284742, rs11084095 and rs34984145 at *SIGLEC5* are located in DNA regions with regulatory activity on gene expression

To measure the effect directions and allele-specific effect sizes of the regulatory elements at rs11084095, rs4284742 and rs34984145 at *SIGLEC5*, and at rs1122900, rs6887423 and rs56038114 at *CTD-2353F22.1*, luciferase reporter gene assays with the reporter gene assay plasmids were performed. Because of higher transfection efficiency of HeLa cells compared to PBMCs and Raji cells, reporter gene assays were performed in this cell type. Details of cloning of the reporter gene plasmids, agarose gel pictures and sequence alignments are given in chapter 8.2.

The reporter gene of rs4284742 with the common G-allele revealed a 13-fold upregulation of luciferase activity ($p = 0.008$) compared to the reporter gene without this regulatory element (**Figure 27A**). The reporter gene with the minor A-allele showed an upregulation of 4-fold ($p = 0.002$). The difference between both alleles was significant with $p = 0.010$. The luciferase activity of rs11084095 showed a significant increase in the background of the common G-allele of rs11084095 by 9.9-fold ($p = 0.015$) compared to the empty plasmid. The rare A-allele did not result in increase of luciferase activity. The difference between both alleles was significant with $p = 0.013$ (**Figure 27B).** The luciferase gene activity of rs34984145 showed weak but significant upregulation for both alleles (T-allele: 4.2-fold ($p = 0.0001$); A-allele: 2.3-fold ($p = 0.002$)) in HeLa cells (**Figure 27C**).

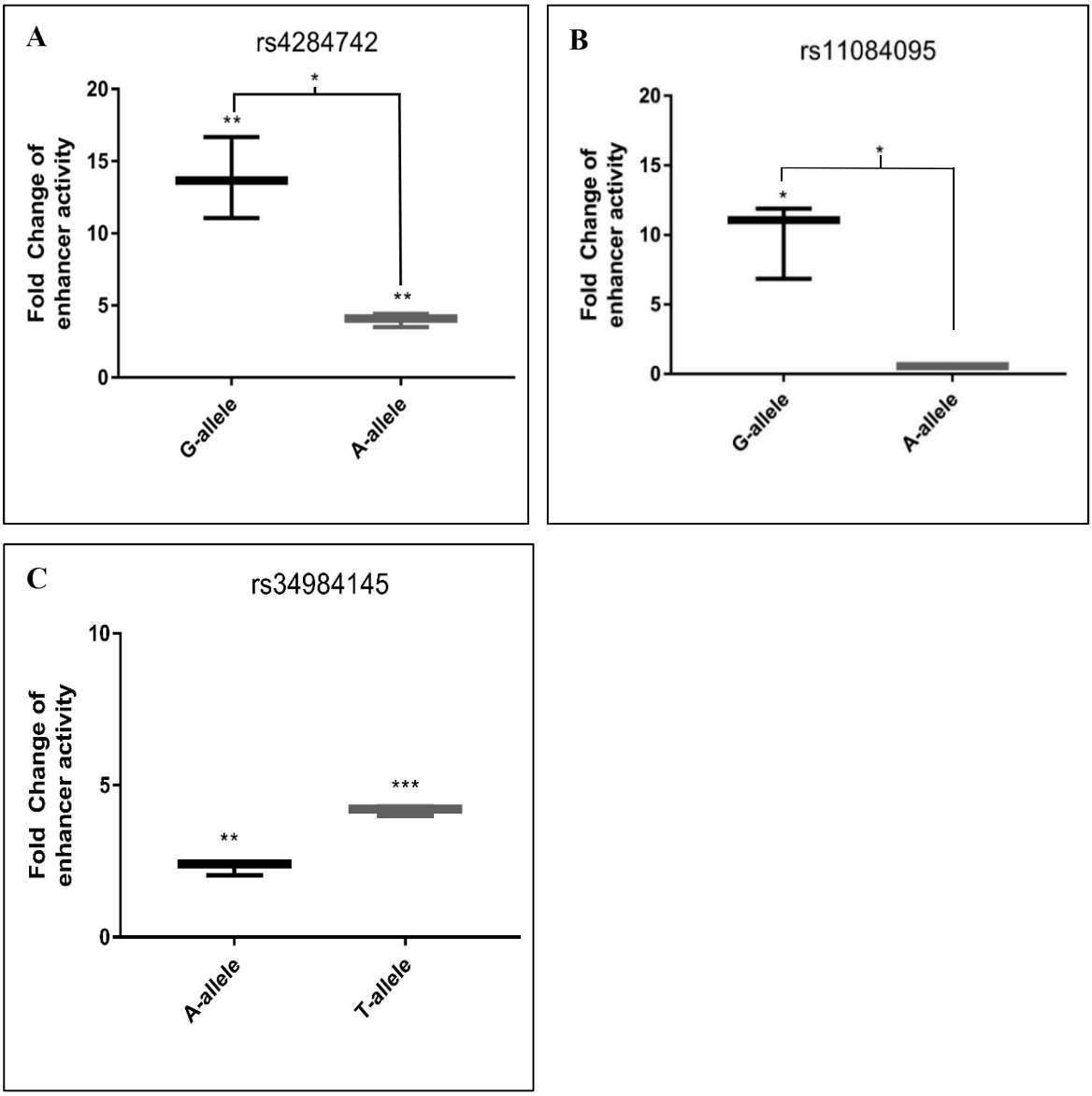


Figure 27: rs4284742, rs11084095 and rs34984145 at *SIGLEC5* are located in DNA regions with regulatory activity on gene expression. (A) rs4284742 has allele-specific effects on Luciferase activity. The 150 bp DNA sequence of rs4284742 showed allele-specific enhancer activity in HELA cells. The common G-allele increased luciferase activity 13-fold ($p = 0.008$) compared to the rare A-allele, that increased the activity 4-fold ($p = 0.002$). The differences between both alleles was significant with $p = 0.010$. **(B)** rs11084095 has allele-specific effects on Luciferase activity. The 65 bp DNA sequence up-and downstream of rs11084095 showed allele-specific enhancer activity in HELA cells. In the background of the common rs11084095 G-allele luciferase activity was 9.9 fold increased ($p = 0.015$) compared to the rare A-allele, which showed no upregulation. The difference between both alleles was significant with $p = 0.013$. **(C)** rs34984145 has no allele-specific effects on Luciferase activity. The 140 bp DNA sequence of rs34984145 showed significant enhancer activity in HELA cells. The rare T-allele showed stronger upregulation of 4.2-fold ($p = 0.0001$) compared to the common A-allele, which showed an upregulation of 2.3-fold ($p = 0.002$). The difference between both alleles was not significant. $p^* < 0.05$. $p^{**} < 0.005$. $p^{***} < 0.0005$.

4.1.5.2 rs1122900 at *CTD-2353F22.1* is located in a DNA region with regulatory activity on gene expression

The luciferase activity of rs1122900 showed a significant increase in the background of the common A-allele of rs1122900 by 5.5-fold (p = 0.0003) compared to the empty plasmid. The rare C-allele showed no increase in luciferase activity. The difference between both alleles was significant with p = 0.0005 (**Figure 28A).** This indicated that the regulatory elements at rs1122900 act as enhancer and showed strong differences of the regulatory effects on enhancer activity between the different alleles of rs1122900.

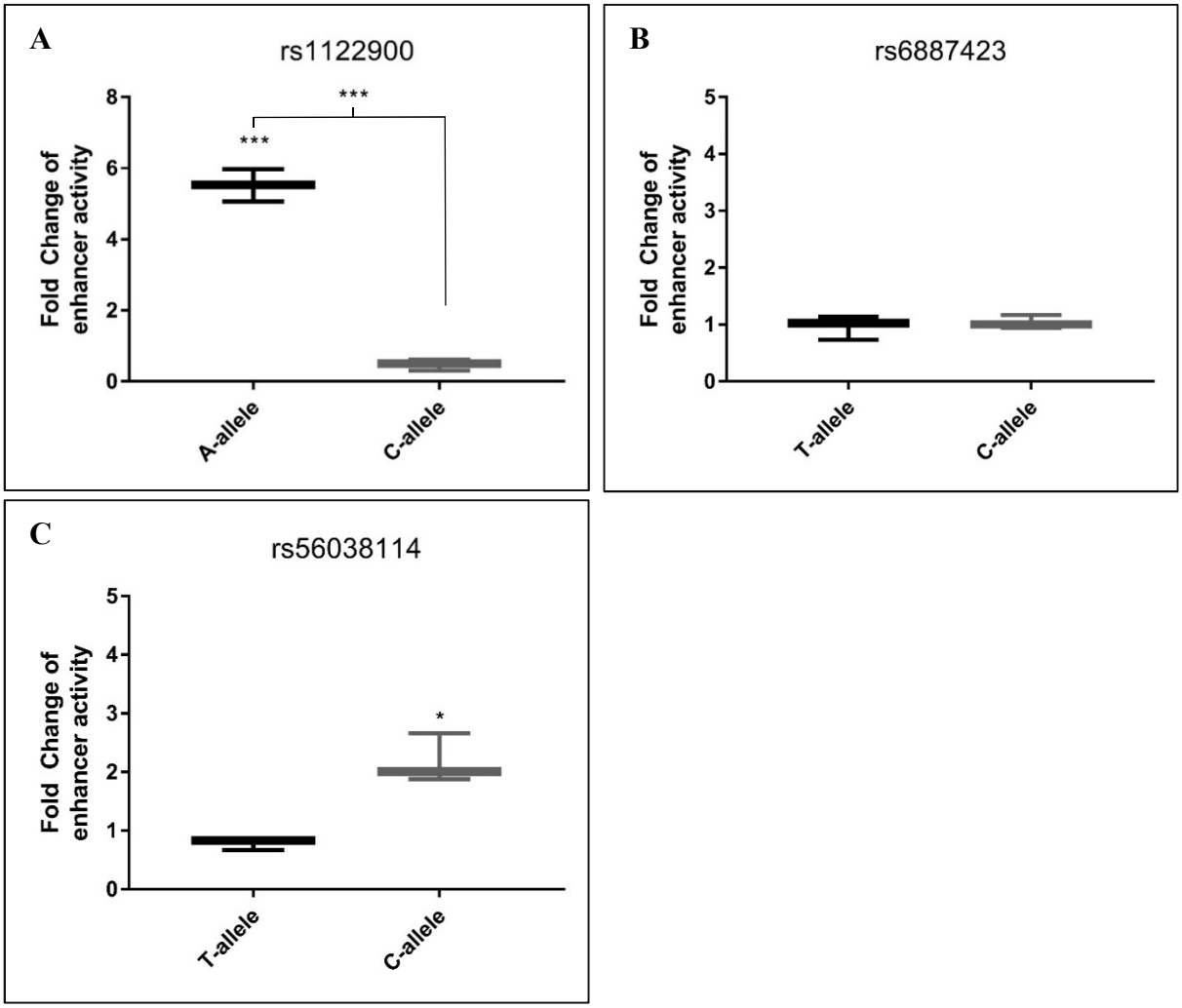


Figure 28: Results of Dual Luciferase Reporter gene assays. (A) rs1122900 has allele-specific effects on Luciferase activity. The 73 bp DNA sequence up-and downstream of rs1122900 showed allele-specific enhancer activity in HELA cells. In the background of the common A-allele luciferase activity was 5.5-fold increased (p = 0.0003) compared to the rare C-allele, which showed no upregulation. The difference between both alleles was significant with p = 0.0005. **(B)** rs6887423 has no allele-specific effects on Luciferase activity. The 143 bp DNA sequence of rs6887423 showed no regulatory effect in HELA cells. **(C)** rs56038114 has no allele-specific effects on Luciferase activity. The T-allele of the 147 bp DNA sequence of rs56038113 showed no regulatory effect in HeLa cells. The rare C-allele showed weak (2.2-fold) significant upregulation of the reporter gene with p = 0.01. p* < 0.05. p** < 0.005. p*** ≤ 0.0005.

The reporter gene of rs6887423 shows no regulatory effect, neither up- nor downregulation of the reporter gene. This indicates that rs6887423 is not located in a regulatory element (**Figure 28B**). The luciferase gene activity of rs56038114 showed weak but significant upregulation for the rare C-allele (2.2-fold ($p = 0.01$) in HeLa cells (**Figure 28C**). However, there is no allele-specific effect on the reporter gene, indicating that rs56038114 is not a putative causal variant.

4.1.6 Comparison of the CRISPR activation systems SAM and VPR using the example of *CTD-2353F22.1*

CRISPRactivaiton was used to identify the target genes of the associations. To ensure efficient upregulation of the regulatory elements, two different CRISPRactivation systems, SAM and VPR, were compared with each other. For this purpose, the promoter of *CTD-2353F22.1* was attempted to be upregulated with both systems. Details of cloning of the gRNAs for CRISPRa with the VPR and the SAM system and agarose gel pictures are given in chapter 8.3 and chapter 8.4.

4.1.6.1 CRISPR-dCas9 activation of the promoter of *CTD-2353F22.1* using VPR and SAM system

For CRISPR-dCas9 activation using the VPR-system, cloned gRNAs were transfected together with the SP-dCas9-VPR plasmid in three technical replicates into HEK293 cells (**Figure 29A**). For CRISPR-dCas9 activation using the SAM-system, cloned gRNAs were transfected together with the dCas9-VP64-GFP and the MS2-P65-HS1-GFP plasmids in three technical replicates into HEK293 cells (**Figure 29B**). Both systems showed significant upregulation of *CTD-2353F22.1* (transcript *CTD-2353F22.1*-002 and transcript *CTD-2353F22.1*-003) with all three different gRNAs. For the VPR system, CTD-1 gRNA showed an upregulation of 3-fold ($p = 0.0003$), CTD-6 gRNA showed an upregulation of 11-fold ($p < 0.0001$) and CTD-9 gRNA showed an upregulation of 4.7-fold ($p < 0.0001$). For the SAM-system, CTD-1 gRNA showed an upregulation of 4-fold ($p = 0.0003$), CTD-6 gRNA showed an upregulation of 23-fold ($p < 0.0001$) and CTD-9 gRNA showed an upregulation of 2.5-fold ($p < 0.0001$).

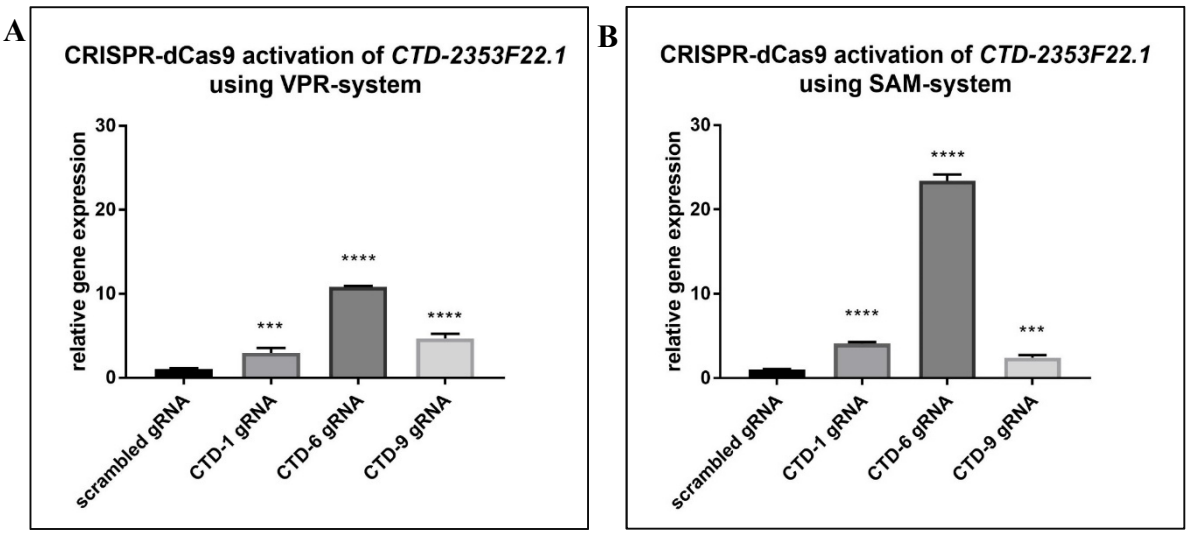


Figure 29: CRISPR-dCas9 activation of the promoter of *CTD-2353F22.1* using VPR and SAM system. (A) The VPR-system showed significant upregulation of *CTD-2353F22.1* with all three different gRNAs. CTD-1 gRNA showed an upregulation of 3-fold ($p = 0.0003$), CTD-6 gRNA showed an upregulation of 11-fold ($p < 0.0001$) and CTD-9 gRNA showed an upregulation of 4.7-fold ($p < 0.0001$). **(B) The SAM-system showed also significant upregulation of *CTD-2353F22.1* with all three different gRNAs.** CTD-1 gRNA showed an upregulation of 4-fold ($p = 0.0003$), CTD-6 gRNA showed an upregulation of 23-fold ($p < 0.0001$) and CTD-9 gRNA showed an upregulation of 2.5-fold ($p < 0.0001$). $p^* < 0.05$. $p^{**} < 0.005$. $p^{***} \leq 0.0005$. $p^{****} \leq 0.0001$.

The results uncovered a stronger upregulation of *CTD-2353F22.1* by CRISPR-dCas9 SAM activation system than by the VPR-system (except for gRNA CTD-9). Therefore, CRISPR/dCas9 gene activation (CRISPRa) using the SAM-system was employed to validate if the associated regulatory elements had *cis* regulatory effects on *SIGLEC5* or *CTD-2353F22.1* expression.

4.1.7 Validation of *cis*-regulation of rs11084095, rs4284742 and rs34984145 of *SIGLEC5* and of *cis*-regulation of rs1122900, rs6887423 and rs56038114 of *CTD-2353F22.1*

4.1.7.1 CRISPR-dCas9 activation of rs4284742 and rs11084095 shows *cis*-regulation of *SIGLEC5*

For rs11084095, an eQTL effect on the expression of *SIGLEC5* was reported in monocytes with $p = 6.4 \times 10^{-23}$ (Zeller et al. 2010) and for rs4284742 an eQTL effect on the expression of *SIGLEC5* was reported in whole blood with 7.7×10^{-14} (Blood eQTL browser (Westra et al. 2013)) (chapter 8.1 (**Table 23-24)**), suggesting that the disease-associated elements have *cis*-regulatory effects on the expression of *SIGLEC5*.

To validate the regulatory potential of the associated haplotype blocks tagged by rs4284742 and rs11084095 on the expression of *SIGLEC5*, CRISPRa was performed in HeLa cells. Raji cells were not used because this cell type showed poor survival after transfection with the CRISPRa plasmids.

Guide RNAs for CRISPRa using the SAM system were cloned as described in chapter 3.1.12.3 and successful cloning was proved by Sanger sequencing. Co-transfection of HeLa cells with the SAM system and three plasmids encoding different gRNAs for rs4284742 (20-300 bp distance from the SNP (**Figure 30**)) strongly increased the expression of *SIGLEC5* compared to a co-transfected unspecific control sgRNA (FC = 380, **Figure 31A**). This indictes that the DNA element at rs4284742 has a cis-regulatory effect on *SIGLEC5* and implies that *SIGLEC5* is the target gene of the disease-associated genetic variant.

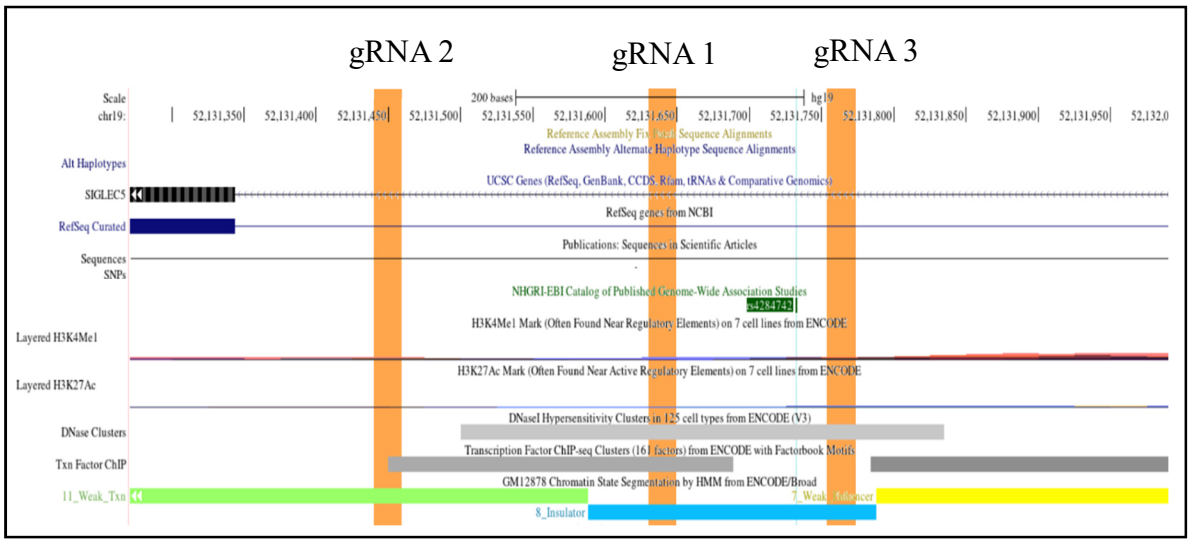


Figure 30: Position of gRNAs to test the genomic region at rs4284742 for its potential to activate *SIGLEC5* expression. CRISPRa of the genomic region at rs4284742 induced *SIGLEC5* expression compared to the scrambled gRNA. gRNA 1 (chr19:52131631-52131649), gRNA 2 (chr19:52131441-52131459), and gRNA3 (chr19:52131754-52131773) have 20-300 bp distance from rs4284742.

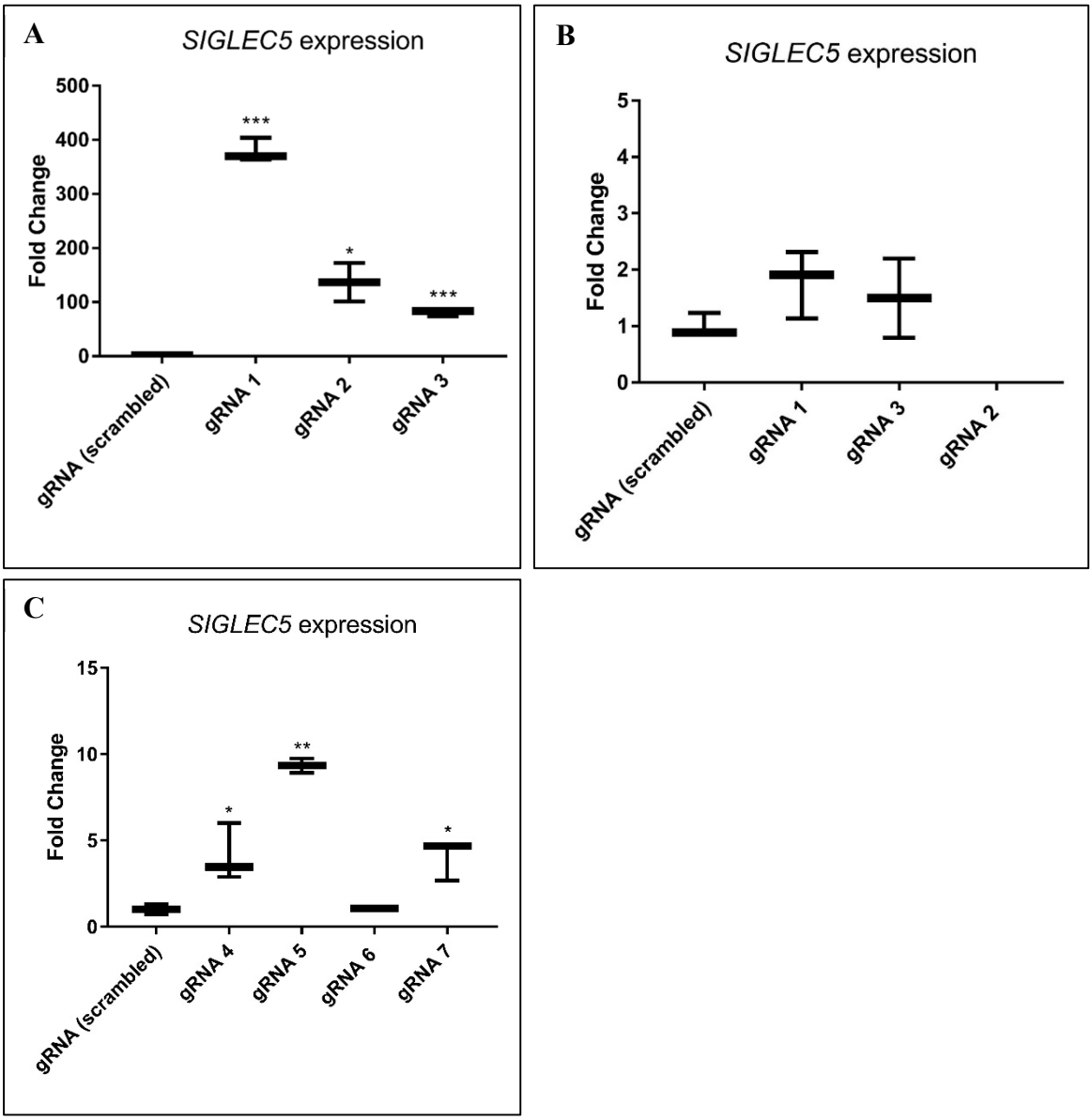


Figure 31: CRISPRa of genomic regions at rs4284742 and at rs11084905 (A) The haplotype block at rs4284742 has a cis-regulatory effect on *SIGLEC5* expression. CRISPRa of the genomic region at rs4284742 induced *SIGLEC5* expression compared to the scrambled gRNA. gRNA-1, gRNA-2 and gRNA-3 induced *SIGLEC5* expression with 380-fold upregulation ($p = 0.00008$), 137-fold ($p = 0.03$) and 80-fold ($p = 0.0009$), respectively. **(B)** CRISPRa of the genomic region at rs11084095 did not significantly induce *SIGLEC5* expression. **(C)** CRISPRa of the genomic region at rs11084095 with gRNAs in distance of 2100 - 2600 bp to rs11084095 shows weak effect on *SIGLEC5* expression (gRNA-4 (4.1-fold; $p = 0.04$); gRNA-5 (9.3-fold; $p = 0.002$); gRNA-7 (4-fold; $p = 0.02$)). gRNA-6 showed no upregulation of *SIGLEC5*. $p^* < 0.05$. $p^{**} < 0.005$. $p^{***} \leq 0.0005$.

Co-transfection of HeLa cells with the SAM system and three plasmids encoding different gRNAs for rs11084095 (25-120 bp distance from the SNP) showed no significant upregulation of *SIGLEC5* (**Figure 31B**), whereby CRISPRa of a region close to rs11084095 (2100 – 2600 bp distance from the SNP) showed up to 9-fold induced *SIGLEC5* expression (**Figure 31C**). Position of gRNAs are shown in **Figure 30** and chapter 8.5 (**Figure 52-53**).

As the alleles of rs34984145 showed no different effects on luciferase activity and no blood-bourne eQTLs of rs34984145 on the expression of *SIGLEC5* were reported, it was assumed that rs34984145 is not a causal variant of the association and accordingly, no CRISPRa was performed for this SNP.

4.1.7.2 CRISPR-dCas9 activation of rs6887423 shows *cis*-regulation of *CTD-2353F22.1*

All reported eQTL effects of rs1122900, rs6887423 and rs56038114 showed an effect on the expression of *CTD-2353F22.1*. For rs1122900, an eQTL effect on the expression of *CTD-2353F22.1* was reported in whole blood with $p = 6.1 \times 10^{-15}$ (Consortium 2013) and in EBV-transformed lymphocytes with 4.3×10^{-9} (chapter 8.1 (**Table 25-26**)). The strongest reported eQTL effects of rs6887423 and rs56038114 is on the expression of *CTD-2353F22.1* in whole blood with $p = 2.6 \times 10^{-26}$ and with $p = 3.5 \times 10^{-13}$ (Consortium 2013).

To validate the regulatory potential of the associated haplotype blocks tagged by rs1122900, rs6887423 and rs56038114 on the expression of *CTD-2353F22.1*, CRISPRa was performed for rs1122900, rs6887423 and rs56038114 in HeLa cells. Positions of gRNAs are given in chapter 8.5 (**Figure 54 - 57**). CRISPRa of rs1122900 with gRNAs in 50 – 180 bp distance to rs1122900 had no effect on *CTD-2353F22.1* expression (**Figure 32A**). One gRNA that was located in distance of 1500 bp to rs1122900 showed very weak effect on *CTD-2353F22.1* expression (2.7-fold; $p = 0.003$) (**Figure 32B**). This could be an indicator that the DNA element at rs1122900 has a *cis*-regulatory effect on *CTD-2353F22.1* and implies that *CTD-2353F22.1* might be the target gene of the disease-associated genetic variant. CRISPRa of rs6887423 with gRNA-1 in 80 bp distance to rs6887423 displayed a weak effect on the expression of *CTD-2353F22.1* with FC = 6.5 and $p = 0.003$. All other gRNAs showed no effect (**Figure 32C**). This indicates that *CTD-2353F22.1* might be the target gene of the disease-associated genetic variant. CRISPRa of rs56038114 with gRNAs in 60 - 360 bp distance to rs56038114 had no effect on *CTD-2353F22.1* expression (**Figure 32D**).

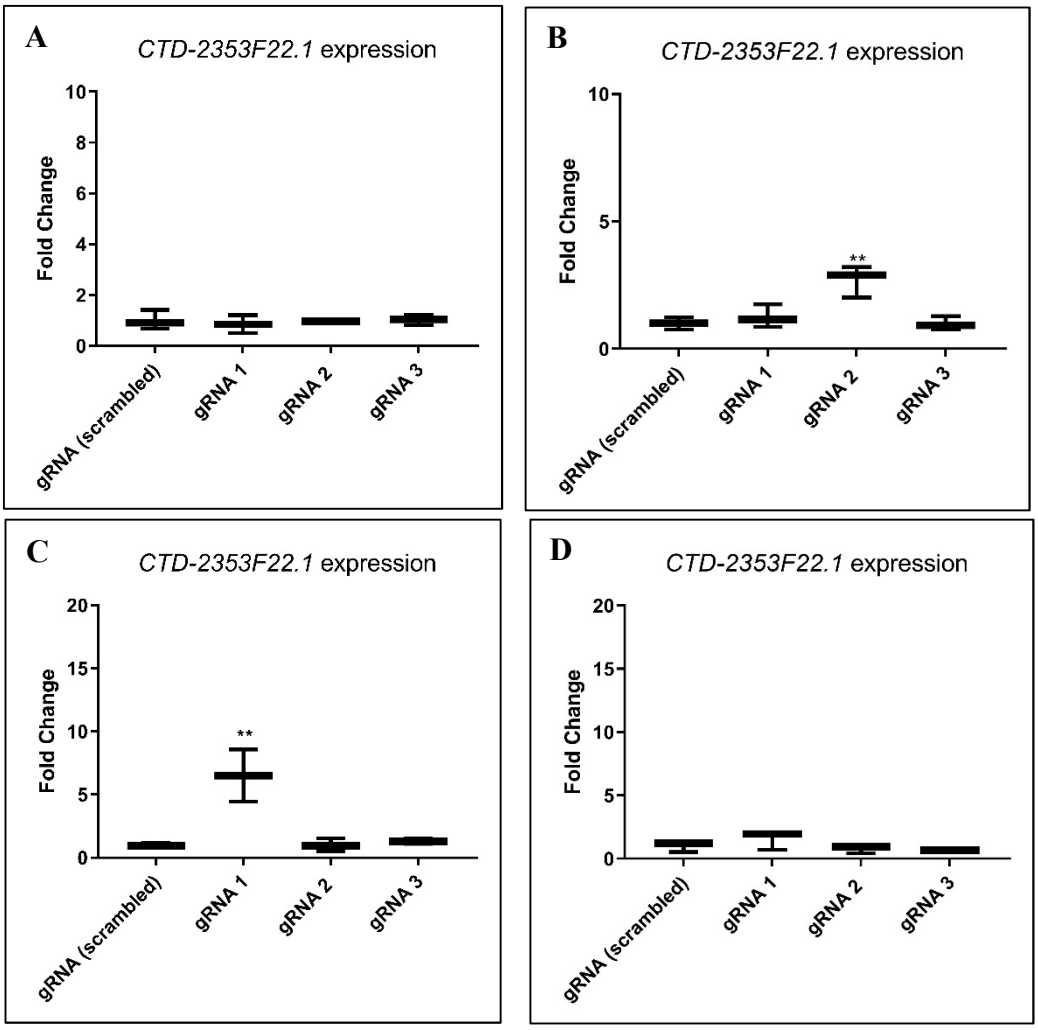


Figure 32: CRISPRa of genomic regions at rs1122900, rs6887423 and rs56038114. (A) CRISPRa of the genomic region at rs1122900 shows no effect on *CTD-2353F22.1* expression. **(B)** CRISPRa of the genomic region at rs1122900 with gRNAs in distance of 2100-2600 bp to rs1122900 shows very weak effect on *CTD-2353F22.1* expression (2.7-fold; p = 0.003). **(C)** CRISPRa of the genomic region at rs6887423 with gRNA 1 induced *CTD-2353F22.1* expression weakly (6.5-fold; p = 0.003) compared to the scrambled gRNA. The two other gRNAs showed no effect on the expression of *CTD-2353F22.1*. **(D)** CRISPRa of the genomic region at rs56038114 showed no effect on *CTD-2353F22.1* expression. p* < 0.05. p** < 0.005.

4.1.8 CRISPR-dCas9 activation of *CTD-2353F22.1*

To identify the functional context of *CTD-2353F22.1*, CRISPR/dCAS9 activation of *CTD-2353F22.1* in HeLa cells was performed followed by RNA-sequencing.

4.1.8.1 CRISPR-dCas9 activation of all three different transcripts of *CTD-2353F22.1*

HeLa cells were transfected in three technical replicates with gRNA CTD-6, l, as it showed the best activation of *CTD-2353F22.1* and scrambled gRNA as control in HEK293 cells (**Figure 33**). After 44 hours, cells were harvested and Total RNA was extracted. To prove the induction of *CTD-2353F22.1* expression, qRT-PCR was performed using primers for each of the three transcripts of *CTD-2353F22.1*. PCR products were loaded onto a 2% agarose gel. The transcript *CTD-2353F22.1-001* showed the smallest, but significant upregulation (FC = 5.1; p = 0.01) (**Figure 33A-B**). *CTD-2353F22.1-002* showed increased expression with a FC = 34 and with p = 0.002 (**Figure 33C-D**). The strongest upregulation after CRISPR/dCas9 activation presented the transcript *CTD-2353F22.1-003* (FC = 37; p = 0.0004). (**Figure 33E-F**).

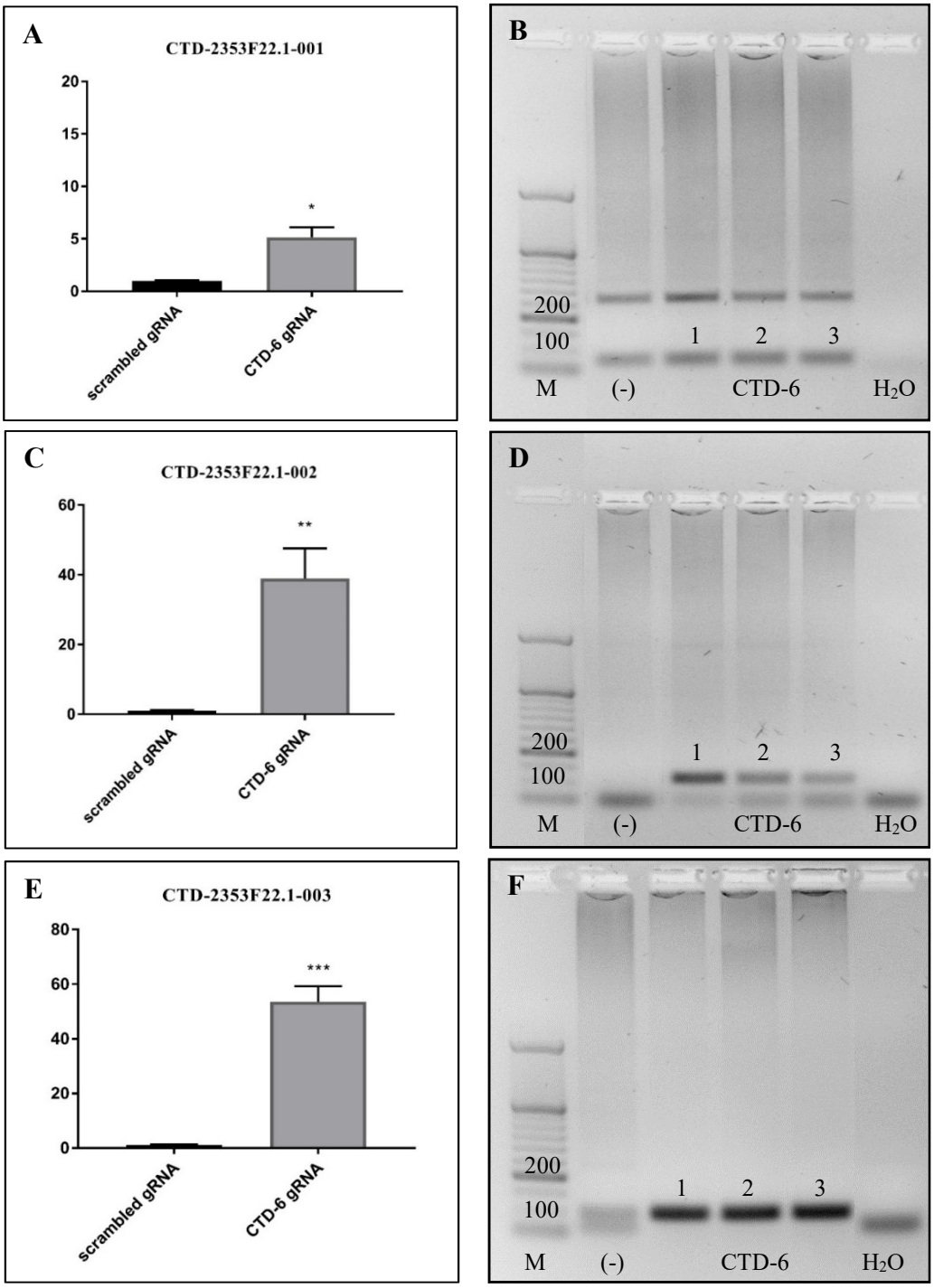


Figure 33: CRISPR-dCas9 activation of all three different transcripts of *CTD-2353F22.1*. (A) *CTD-2353F22.1-001* showed the smallest, but significant upregulation (FC = 5.1; p = 0.01) **(B)** Agarose gel picture of qRT-PCR products for the transcript *CTD-2353F22.1-001*. The first line shows the product of the CRISPRa with the scrambled gRNA, lane 2-4 show the three independent biological replicates of CRISPRa with gRNA CTD-6 of *CTD-2353F22.1-001* and the last lane shows the water control. **(C)** Expression of *CTD-2353F22.1-002* was significantly increased (FC = 34; p = 0.002) **(D)** Agarose gel picture of qRT-PCR products for the transcript *CTD-2353F22.1-002*. The first line shows the product of the CRISPRa with the scrambled gRNA, lane 2-4 show the three independent biological replicates of CRISPRa with gRNA CTD-6 of *CTD-2353F22.1-002* and the last lane shows the water control. **(E)** Expression of *CTD-2353F22.1-003* was significantly upregulated (FC = 37; p = 0.0004) **(F)** Agarose gel picture of qRT-PCR products for the transcript *CTD-2353F22.1-003*. The first line shows the product of the CRISPRa with the scrambled gRNA, lane 2-4 show the three independent biological replicates of CRISPRa with gRNA CTD-6 of *CTD-2353F22.1-003* and the last lane shows the water control. p* < 0.05. p** < 0.005. p*** ≤ 0.0005.

4.1.8.2 Overexpression of *CTD-2353F22.1* upregulates the gene sets "angiogenesis" and "TNF alpha signaling via NFKB"

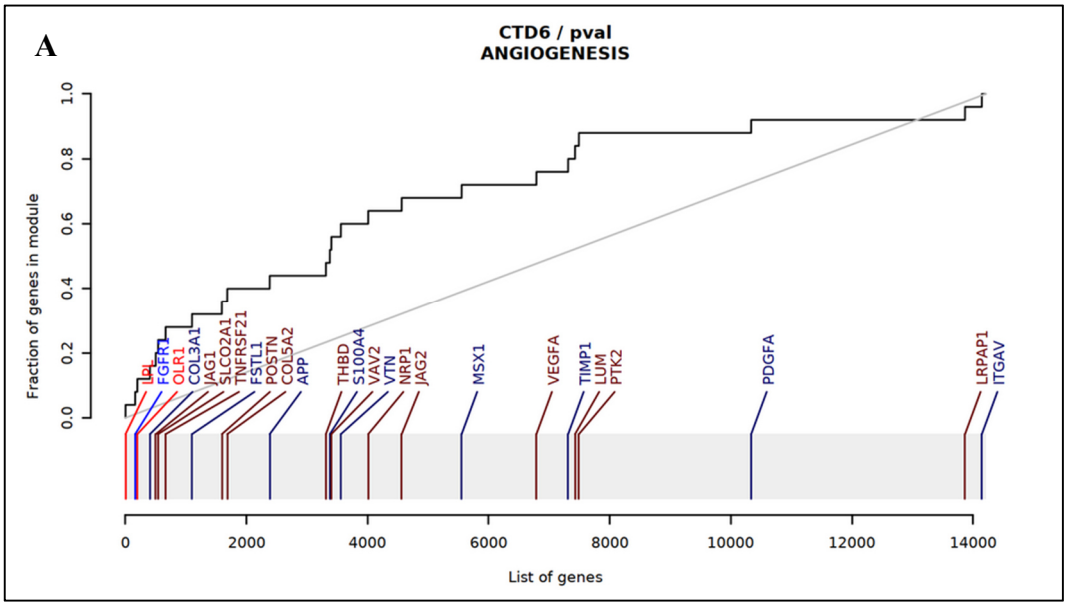
To identify *CTD-2353F22.1* responding genes and pathways, and to determine its function in the cell, RNA-sequencing of three replicates of *CTD-2353F22.1* CRISPRa cells compared to three replicates of scrambled sgRNA as control was performed. *CTD-235322.1* was the most upregulated gene with a FC (log_2) = 5.6 ($p_{adj} < 7.5 \times 10^{-35}$). The most significant downregulated gene is the Growth Arrest and DNA Damage Inducible Alpha (GADD45A) with $p_{adj} < 4.9 \times 10^{-11}$ and the second most significant downregulated gene is Thrombospondin 1, which has been shown to play a role in platelet aggregation and angiogenesis (**Table 20**).

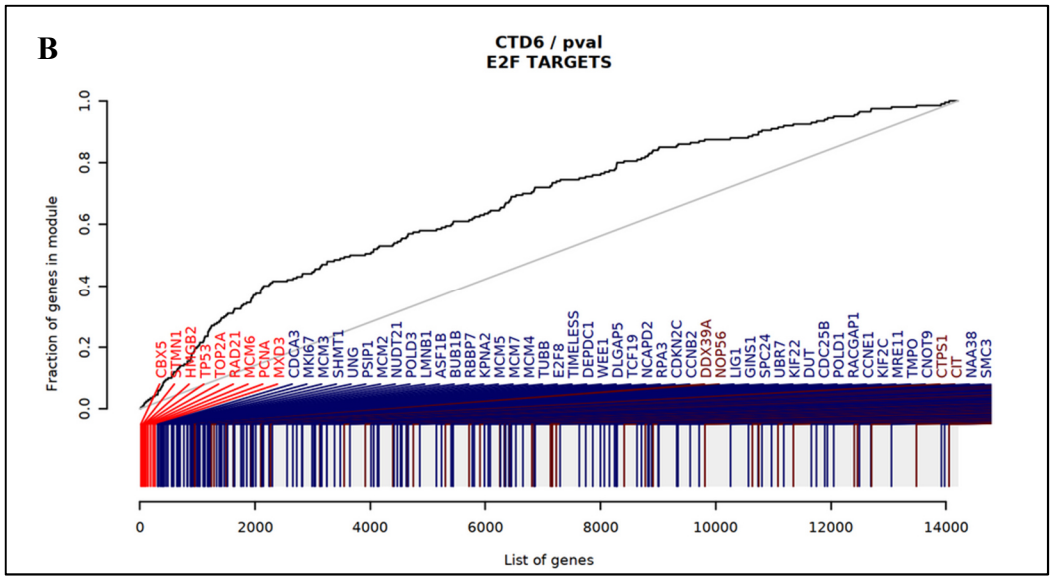
Table 20: Most significant up-and downregulated genes ($p_{adj} < 5 \times 10^{-6}$) after CRISPRa of *CTD-2353F22.1* in HeLa cells.

Gene	Description	Fold change (log_2)	*p* value	p_{adj} value
UPREGULATED GENES				
CTD-2353F22.1	uncharacterized lncRNA	5.61	6.4E-39	7.5E-35
LPL	Lipoprotein Lipase	0.34	1.1E-14	4.9E-11
CPS1	Carbamoyl-Phosphate Synthase 1	0.31	7.9E-14	2.3E-10
RNF182	Ring Finger Protein 182	0.29	6.2E-10	8.1E-07
ASS1	Argininosucciante Synthase 1	0.38	7.7E-10	9.0E-07
ATP9A	ATPase Phospholipid Transporting 9A	0.36	4.5E-09	4.4E-06
CBX5	Chromobox 5	0.29	4.9E-09	4.4E-06
DOWNREGULATED GENES				
GADD45A	Growth Arrest And DNA Damage Inducible Alpha	-0.55	1.25E-14	4.89E-11
THBS1	Thromospondin 1	-0.35	4.98E-13	1.17E-09
SERPINE1	Serpin Family E Member 1	-0.46	2.45E-12	4.78E-09
F3	Coagulation Factor III, Tissue Factor	-0.50	1.65E-10	2.76E-07
FOSL1	FOS Like 1, AP-1 Transcription Factor Subunit	-0.64	2.82E-10	4.12E-07
NBPF19	Neuroblastoma Breakpoint Family Member 19	-0.63	1.22E-09	1.30E-06

Gene set enrichment analysis with the Hallmark gene sets using a 2nd generation algorithm, showed the highest effect sizes of the *CTD-2353F22.1* CRISPRa cells compared to scrambled

sgRNA for the gene sets “angiogenesis” with an area under curve (AUC) = 0.71 and adjusted p-value $p_{adj} = 8.2 \times 10^{-5}$, “E2F Targets” with an AUC = 0.67 ($p_{adj} = 1.5 \times 10^{-13}$), “MYC Targets V2” with an AUC = 0.65 ($p_{adj} = 1.1 \times 10^{-4}$) and “TNF alpha signaling via NFKB” with an AUC = 0.63 ($p_{adj} = 1.0 \times 10^{-17}$) (**Figure 34A-D**).





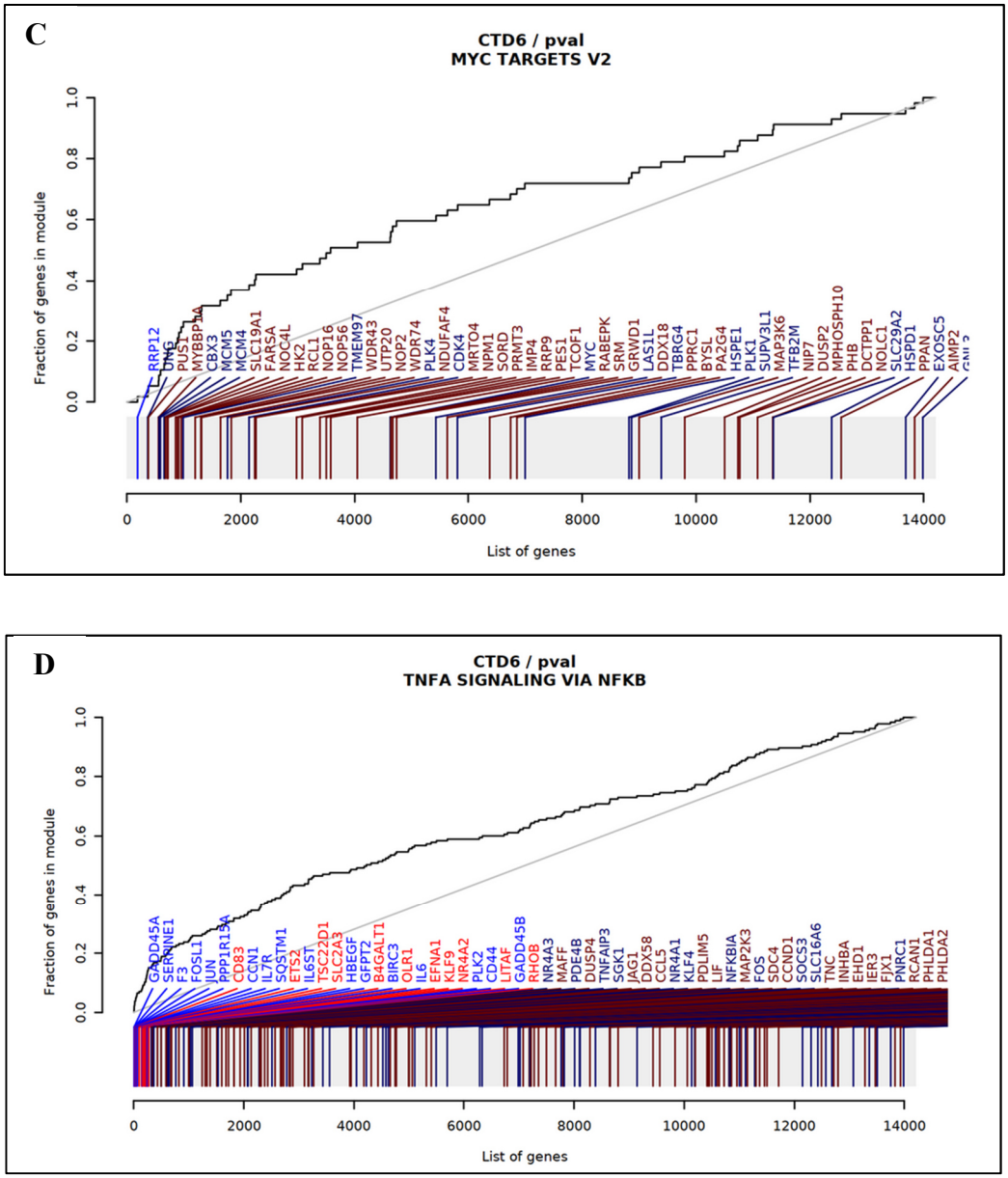


Figure 34: Gene set enrichment analysis of CRISPRa induced *CTD-2353F22.1* expression in HeLa cells. (A,B, C, D) Evidence plots (receiver-operator characteristic curves) for the top 4 gene sets. Each panel corresponds to one gene set. The grey rug plot underneath each curve corresponds to genes sorted by p-value, with the genes belonging to the corresponding gene sets highlighted in red (upregulated genes) or blue (downregulated genes). Bright red or bright blue indicates that the genes are significantly regulated. The area under the curve (AUC) corresponds to the effect size of the enrichment, with 0.5 being no enrichment and 1.0 being maximal possible enrichment.

In addition, Gene Ontology (GO) enrichment analysis with hypergeometric test showed enrichment of gene clusters related to the biological processes: response to wounding, blood vessel morphogenesis, angiogenesis, inflammatory response, wound healing, extracellular structure and matrix organization and response to bacterium. Dot plot results in **Figure 35**.

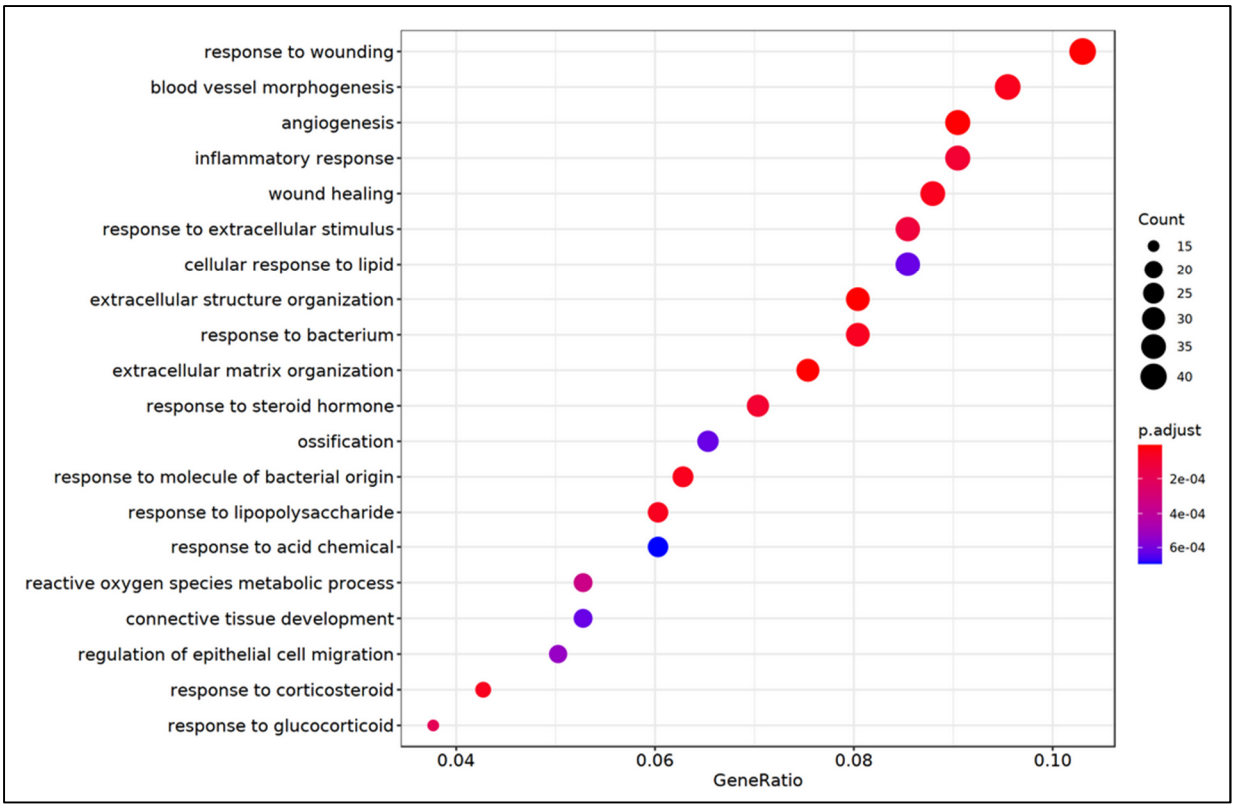


Figure 35: Dot plot of cluster profiler results of GO enrichment analysis with hypergeometric test. Dots represent term enrichment with color coding: red indicates high enrichment, blue indicates low enrichment. The sizes of the dots represent the counts of each row (GO category).

4.2 Identification and characterization of miRNA miR-374b-5p

In **Figure 36** the workflow of the miRNA study is shown.

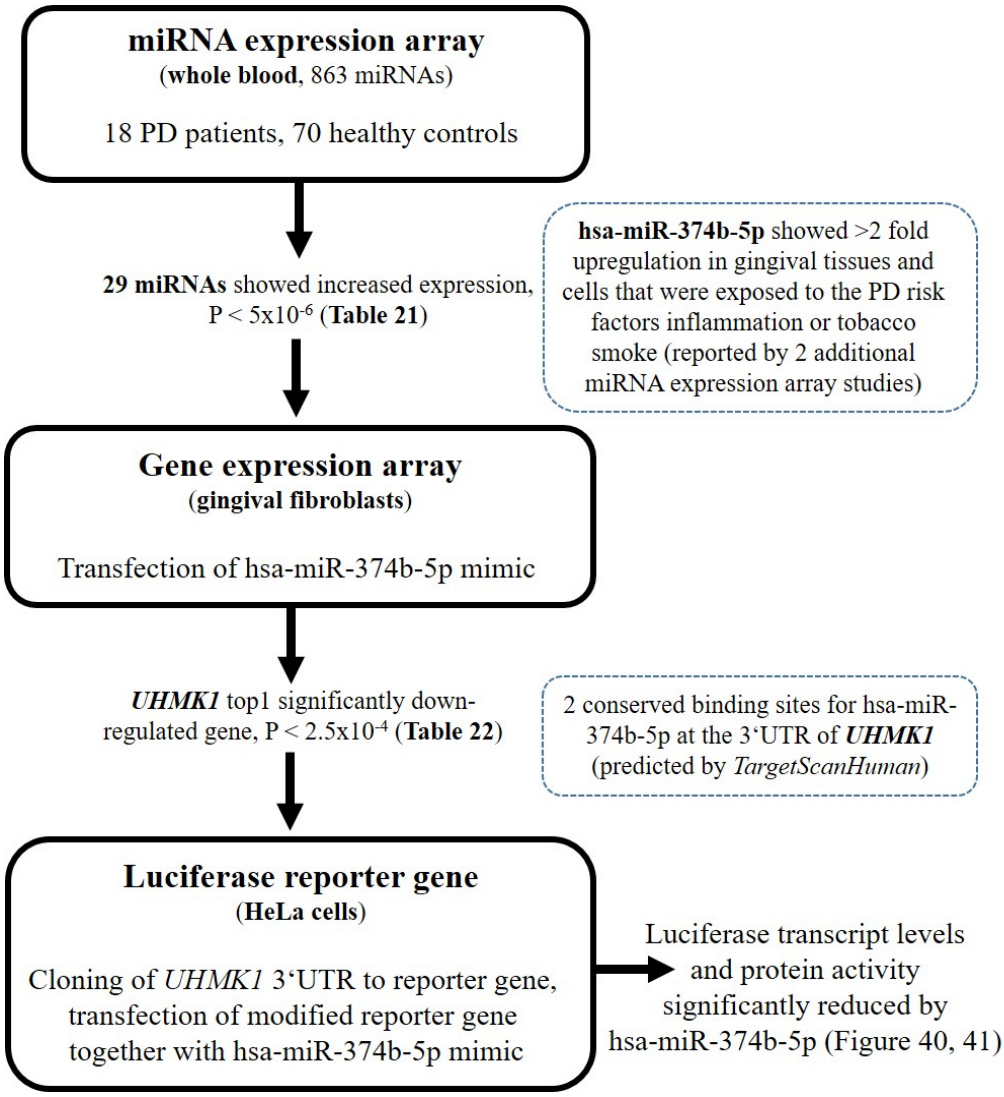


Figure 36: Workflow of the miRNA study. 29 miRNAs showed significant upregulation in whole blood of 18 PD patients compared to 70 healthy controls (Geniom Biochip). Of these, only hsa-miR-374b-5p showed increased expression in more than one additional study that used various miRNA arrays to quantify miRNA expression in gingival tissues and cells that were exposed to inflammation or nicotine as risk factors of PD. To identify the target gene of hsa-miR-374b-5p regulation, a hsa-miR-374b-5p mimic was transfected into primary gingival fibroblasts (3 biological replicates) and changes in expression of protein coding genes were monitored (Clariom D expression array). 24 hours after transfection, *UHMK1* was identified as the top 1 most significantly downregulated gene. 2 conserved hsa-miR-374b-5p binding sites were predicted at the 3'UTR of *UHMK1* (TargetScanHuman). To proof the molecular mechanism of miR-374b-5p regulation of *UHMK1*, the *UHMK1*-3'UTR sequence that included the predicted miR-374b-5p binding sites was cloned to the reporter gene Luciferase. After co-transfection of a hsa-miR-374b-5p mimic and the modified reporter gene into HeLa cells, Luciferase transcript levels and protein activity were significantly reduced. These results give evidence that hsa-miR-374-5p binds to the 3'UTR of *UHMK1* to suppress *UHMK1* gene expression at the post-transcriptional level indicating that the miRNA-mediated regulation is exerted through mRNA degradation but not through translational repression (taken from (Mueller et al. 2021)).

In preliminary works, total miRNAs of venous blood were isolated from 18 untreated PD cases of Dutch and German background and 70 German controls, quantified on the Geniom Biochip array "miRNA Homo sapiens" and analyzed using the microarray-based screening approach described in (Keller et al. 2009a) (Keller et al. 2009b).

4.2.1 Identification of miRNAs that are upregulated in blood and gingiva of PD patients

Out of 863 miRNAs that were quantified with the Geniom Biochip array, 29 miRNAs were significantly higher expressed with $p < 5x10^{-6}$ in blood from PD patients compared to healthy controls, 23 of which showed >2 fold upregulation (**Table 21).** All different expressed miRNAs between PD cases and healthy controls are given in chapter 8.6 (**Table 29**).

Table 21: List of upregulated genes with $P < 5 \times 10^{-6}$ in whole blood from PD cases compared to healthy controls. MiRNA hsa-miR-374b (marked in bold) solely showed increased expression with a FC >2 in blood of PD patients, in addition to ≥ 1 miRNA expression study of gingival cells and tissues (taken from (Mueller et al. 2021)).

miRNA (miRBAv20)	**p-value**	**Fold change**
hsa-miR-744-5p	1.30E-12	2.76
hsa-miR-1228-3p	2.53E-11	5.72
hsa-miR-758-3p	1.56E-10	28.70
hsa-let-7d-5p	2.30E-10	4.02
hsa-miR-567	2.54E-09	20.22
hsa-miR-1281	2.54E-09	7.02
hsa-miR-361-5p	4.12E-09	2.75
hsa-miR-1182	2.45E-08	24.77
hsa-miR-1181	2.45E-08	1.87
hsa-miR-765	3.73E-08	17.85
hsa-miR-548h-5p	1.28E-07	11.03
hsa-miR-146b-5p	1.28E-07	1.86
hsa-miR-199b-5p	1.69E-07	19.90
hsa-miR-384	1.69E-07	3.13
hsa-miR-9-3p	2.26E-07	3.97
hsa-miR-421	2.52E-07	3.92
hsa-miR-653-5p	3.52E-07	34.38
hsa-miR-637	3.92E-07	15.49

hsa-miR-187-5p	3.92E-07	2.46
hsa-miR-125a-3p	4.40E-07	16.25
hsa-let-7b-3p	6.62E-07	20.38
hsa-miR-522-3p	9.56E-07	7.21
hsa-miR-490-3p	1.33E-06	0.66
hsa-miR-1256	1.33E-06	0.56
hsa-miR-551b-5p	1.67E-06	3.10
hsa-miR-634	1.67E-06	1.66
hsa-miR-374b-5p	**2.06E-06**	**2.16**
hsa-let-7f-2-3p	2.31E-06	11.01
hsa-miR-191-5p	3.87E-06	1.45

MiRNA hsa-miR-374b-5p was >2 fold upregulated in blood of PD cases ($P = 2.1x10^{-6}$) and was additionally reported by two other miRNA array studies to be > 2 fold upregulated in biopsies of the oral masticatory mucosa and cells from PD cases compared to healthy controls. Owing to these results, this miRNA was chosen to identify the target gene of its regulatory function (Mueller et al. 2021).

4.2.2 Genome-wide expression profiling of the effects of hsa-miR-374b-5p on primary human gingival fibroblasts

To identify the target gene of miRNA hsa-miR-374b-5p, mirVana miRNA mimic Negative Control #1, mirVana miRNA mimic miR-1 as positive control and hsa-miR-374b-5p mimic were transfected in three technical replicates in three independent primary gingival fibroblast cells (pGF) from three independent donors. 24 hours after transfection, total RNA was extracted and successful transfection was confirmed by qRT-PCR. The positive control mirVana miRNA mimic miR-1 significantly downregulated the gene Protein Tyrosine Kinase 9 (*PTK9*) in all three donors (FB1: p = 0.008; FB2: p = 0.004, FB3: p = 0.003) (Mueller et al. 2021) (**Figure 37**), indicating that the experiment was successful.

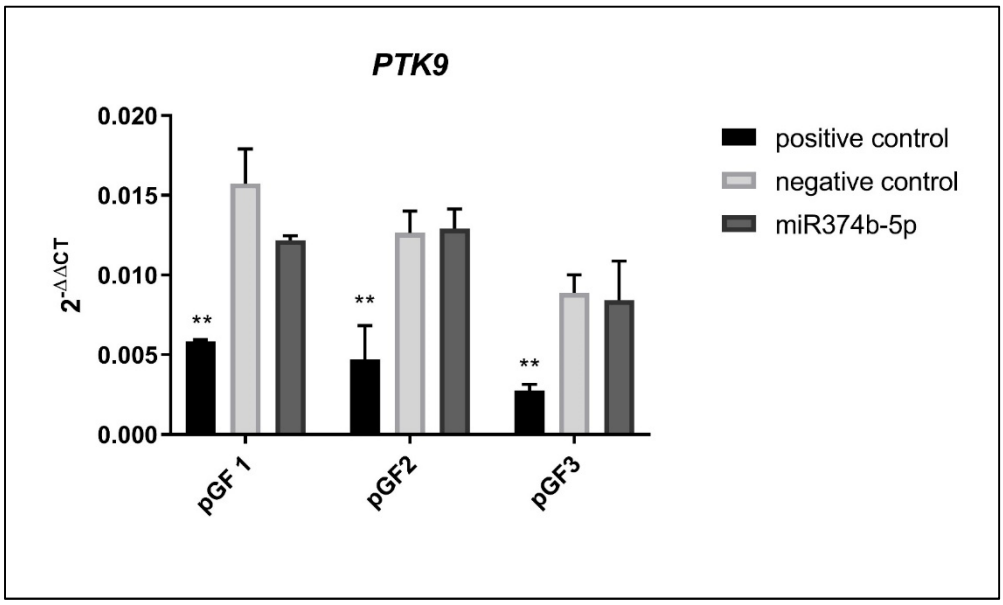


Figure 37: *PTK9* expression after 24 hours miR-1 transfection in pGF cells. The positive control downregulates PTK9 in all three donors (FB1: p = 0.008; FB2: p = 0.004, FB3: p = 0.003).

RNA expression of transfected cells of one technical replicate of hsa-miR-374b-5p mimic and miRNA mimic negative control #1 from each donor were quantified using the Clariom D Expression Array. A total of 21,448 individual transcripts were detected in pGFs and the expression of pGFs that were transfected with hsa-miR-374b-5p mimic was compared to pGFs transfected with miRNA mimic Negative Control #1.

Since biological variation between the transfected cell cultures and technical variation between individual miRNA mimic transfections could result in false positive findings, a pooled analysis of the expression array data was performed using Partek Genomic Suite Software. On the transcript level, miRNAs generally suppress gene expression through mRNA degradation, which is why the focus of the analyzed data was on downregulated genes. Genes that were downregulated in the pooled analysis with FC < -1.5 and p < 0.005 after miR-374b-5p mimic transfection in pGF cells are shown in **Table 22**. The smallest p-value in the pooled analysis with $p = 2.5x10^{-4}$ and the third largest fold change downregulation, with FC = -1.77 showed the gene *UHMK1* (U2AF homology motif (UHM) kinase 1) (Mueller et al. 2021).

All genes that were downregulated are given in chapter 8.6 (**Table 30**). Genes that showed upregulation after transfection of miRNA hsa-miR-374b-5p mimic are listed in chapter 8.6 (**Table 31**).

Table 22: Genes with < -1.5-fold change downregulation at p < 0.005 after miR-374b-5p mimic transfection in pGF cells for 24 hours (taken from (Mueller et al. 2021)).

Chr.	Gene	P-value	Fold change	GWAS catalog entries
1	*UHMK1*	0.00025	-1.767	Bone mineral density, Takayasu arteritis, emotional problems
16	*HSBP1*	0.00031	-1.530	chronic obstructive pulmonary disease, scoliosis
10	*CACUL1*	0.00040	-1.505	Height, obesity
14	*STYX*	0.00122	-2.017	-
19	*ZNF30*	0.00196	-1.563	-
12	*TAS2R19*	0.00245	-1.609	Bitter taste perception
11	*MS4A7*	0.00265	-1.798	myeloid leukemia, taurine metabolite levels
1	*TMED5*	0.00319	-1.894	HDL levels, educational attainment, myeloid leukemia
6	*SLC25A27*	0.00397	-1.629	FEV/FEC ratio, QRS complex, thyroiditis, hyperthyroidism
17	*METTL16*	0.00424	-1.567	HDL levels, mean corpuscular hemoglobin, insomnia
1	*CSRP1*	0.00424	-1.624	-
15	*RAB8B*	0.00462	-1.670	mean corpuscular hemoglobin, schizophrenia, social communication impairment

4.2.3 Validation of the effect of hsa-miR-374b-5p on *UHMK1* expression

To validate the effect of hsa-miR-374b-5p on *UHMK1* expression on the molecular biological level, the 3' UTR of *UHMK1* was cloned into the reporter gene plasmid pGL4.24 and was transfected together with the hsa-miR-374b-5p mimic, mirVana miRNA Mimic Negative Control #1 and mirVana miRNA mimic miR-1 as positive control into HeLa cells. The effect of hsa-miR-374b-5p on *UHMK1* expression was validated using qRT-PCR and Dual luciferase reporter gene assay.

Biological targets of miRNAs were predicted using the online tool TargetScanHuman (version 7.1). The 3'UTR of mRNA sequences were searched for the presence of conserved 8mer and 7mer sites that matched the seed region of miRNA hsa-miR-374b-5p. In the 3'UTR of UHMK1, two conserved 8mer sites, located 961 and 2,859 nucleotides upstream of the end of the 3'UTR, matched the seed region of hsa-miR-374b-5p (Mueller et al. 2021).

4.2.3.1 Cloning of the *UHMK1* 3'UTR into the reporter plasmid pGL4.24

A 2,055 nucleotide sequence of the 3'UTR that included the two 8mer sites were cloned to the poly-adenylation site of the luciferase gene in the reporter vector pGL4.24. The 2,055 nucleotide sequence of the 3'UTR of *UHMK1* were amplified using PCR and the PCR product and the pGL4.24 plasmid were restricted with *XbaI* and loaded onto a 1% agarose gel (**Figure 38**).

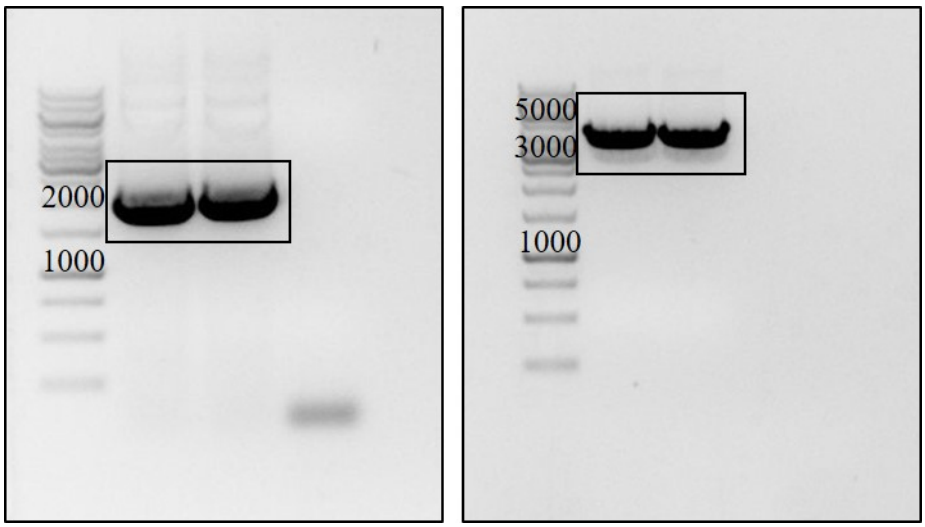


Figure 38: Picture of the 1% agarose gel with restricted PCR-product of 3' UTR of *UHMK1* (left) and restricted plasmid pGL4.24 (right). Electrophoresis was performed for 45-70 min at 100 V. The labeled fragments were isolated from the gel as described in chapter 3.1.8.1.3.

The expected size for the PCR product of the 3' UTR of *UHMK1* of 2,055 bp, and the expected size for the linearized pGL4.24 plasmid of 4411 bp were obtained. The fragments were excised and purified from the agarose gel, and then ligated. Successful ligation was tested by restriction digestion and finally confirmed by Sanger sequencing.

4.2.3.2 Validation of the effect of hsa-miR-374b-5p on *UHMK1* expression using qRT-PCR

To suppress gene expression through mRNA degradation, miRNAs usually interact with the 3' UTRs of the target mRNAs. To give evidence that hsa-miR-374b-5p suppresses the target mRNA on the transcript level, the expression of the reporter gene after miRNA transfection was quantified using qRT-PCR. Complementary DNA (cDNA) was tested for genomic DNA contamination using PCR. DNase I digest was performed until no genomic DNA was detectable (**Figure 39**).

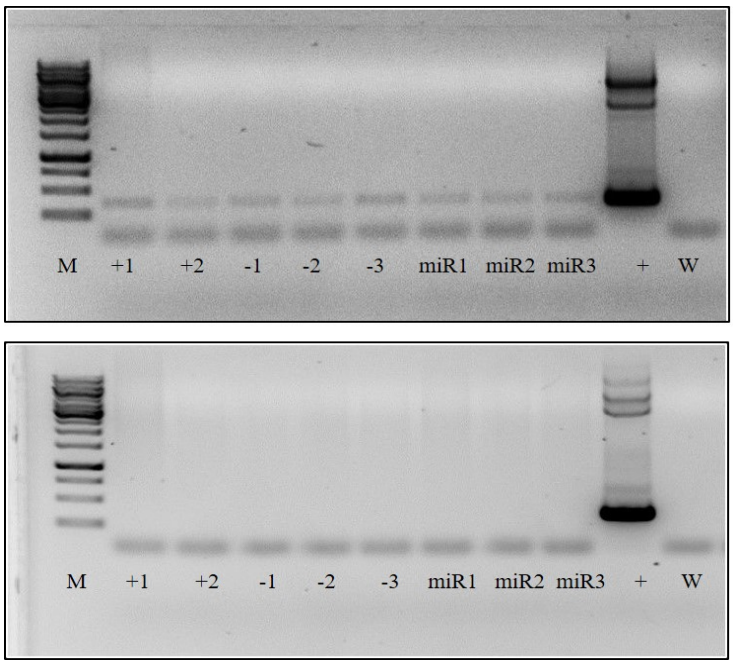


Figure 39: Agarose gel pictures of genomic DNA contamination of cDNA. (A) Agarose gel pictures of genomic DNA contamination of cDNA after DNAseI digest on column. PCR product of genomic DNA at 300 bp was obtained. **(B)** Agarose gel picture genomic DNA contamination of cDNA after more DNAseI digests. No PCR product was obtained, except from the positive control.

Once no genomic DNA was detectable in the cDNA, a qRT-PCR with *GAPDH* primers and Luciferase backbone primers was performed. 16 hours after miRNA transfection, luciferase mRNA expression showed -1.5 FC downregulation in HeLa cells with p = 0.02 (**Figure 40**).

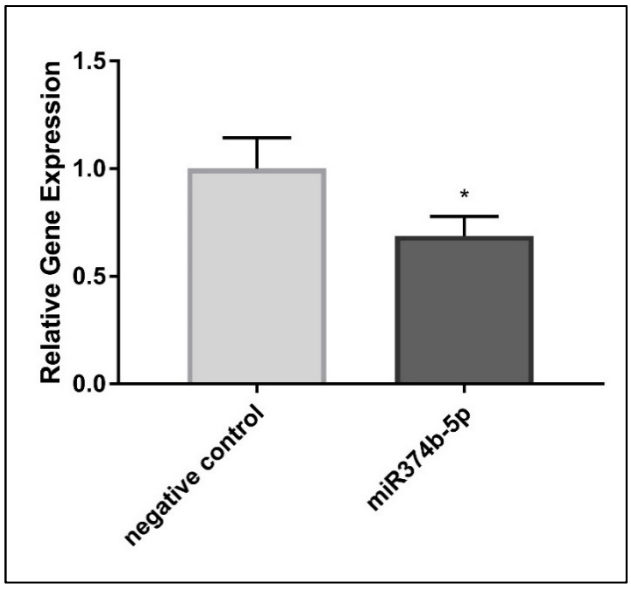


Figure 40: miR-374b-5p downregulates Luciferase-*UHMK1*-3'UTR reporter gene expression. Significantly reduced activity of the Luciferase-*UHMK1*-3'UTR reporter gene was observed on the transcript level (p = 0.02; FC = -1.5) 16 hours after miRNA transfection (taken from (Mueller et al. 2021))

4.2.3.3 Validation of the effect of hsa-miR-374b-5p on *UHMK1* expression using Dual luciferase reporter gene assay

To demonstrate the effect of hsa-miR-374b-5p on *UHMK1* expression on the protein level, luciferase activity of the reporter gene was measured 24 hours after miRNA transfection in HeLa cells. Luciferase activity was downregulated with -1.3-fold change and p = 0.013 **(Figure 41)**.

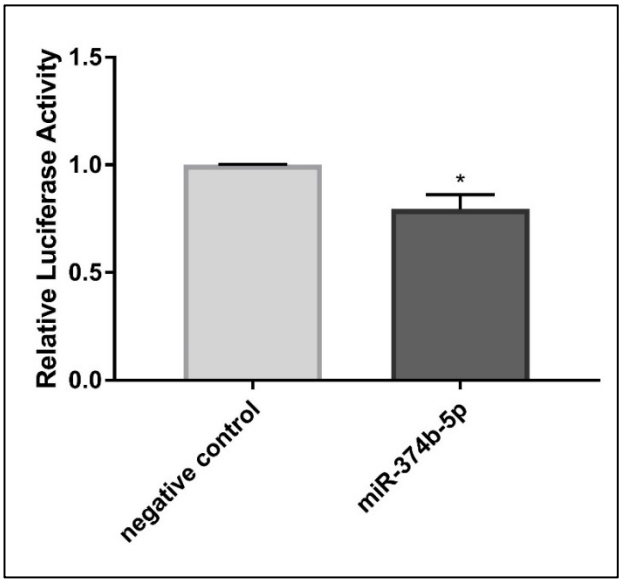


Figure 41: miR-374b-5p downregulates Luciferase-*UHMK1*-3'UTR reporter gene expression. Significantly reduced activity of the Luciferase-*UHMK1*-3'UTR reporter gene was observed on the transcript level (p = 0.013; FC = -1.3) following 24 h after miR-374b-5p transfection (taken from (Mueller et al. 2021))

5 Discussion

5.1 Identification of putative causal variant(s) underlying the associations at *SIGLEC5* and *CTD-2353F22.1*

Numerous SNPs statistically associated with risk for complex human diseases and traits were identified by GWASs. To elucidate the specific molecular mechanisms that predispose to increased disease susceptibility, genetic associations need to be leveraged to biological meaning. This poses a challenge because the most significant associated variant(s), the GWAS lead SNP(s) is most often not identical with the functional variant(s) that caused the association. This is explained by the fact that numerous SNPs are in strong linkage disequilibrium (LD) with the GWAS lead SNPs, comprising associated haplotype block(s). Consequently, the sentinel variant does not have to cause the associations, and empirical validation is required to determine which variants, the GWAS lead SNP(s) or the linked variants are functional and causing the associations. Furthermore, the associated SNPs usually fall within non-protein-coding sequences like introns or intergenic regions. These may either influence the expression of the most closely located gene (*cis*) or remote genes (*trans*), thus an experimental approach to determine the target gene of the association is necessary. To explain the role of the effect alleles in a disease, the molecular mechanism that the causal variant(s) impairs needs characterization. Answering these problems can help to improve the understanding of the underlying disease biology, to point to a regulatory genetic pathway and hopefully leading to new treatment options.

This thesis aimed to identify the putative causal variants of the GWAS associated haplotype blocks at *SIGLEC5* and at *CTD-2353F22.1*. Genetic variants at *SIGLEC5*, located on different haplotype blocks, were independently found in two large GWAS' to be associated with aggressive and chronic periodontitis, indicating broad relevance of this locus for disease etiology. To identify rs4284742, rs11084095 and rs1122900 as putative causal variants, all SNPs that were in strong LD to the GWAS lead SNPs were analyzed for their location at a predicted TFBS. The two complementary TF databases allowed prediction of binding sites of mostly all human TFs currently known. TFBS for the TFs MAFB, ERG and BACH2 at the positions of rs4284742, rs11084095 and rs34984145 for the associated haplotype blocks at *SIGLEC5* (**Figure 12**) and TFBS for the TFs PRDM14, AHR and GATA1 at the positions of rs1122900, rs6887423 and rs56038114 for the associated haplotype blocks at *CTD-2353F22.1* (**Figure 19**) were predicted. For none of the others LD-SNPs to the sentinel variants, a TFBS was predicted, indicating that these variants may not putative causal ones. Moreover, different algorithms and experimental approaches result in different sequences and PWMs, with the

consequence that the PWMs are limited in predictive accuracy (Weirauch et al. 2013). To overcome these limitations complementary TF databases were used to identify putative TFBS. However, it is still possible that a TFBS was missed because of currently unknown binding sites or because of the used algorithms of the TF databases.

5.1.2 rs4284742 and rs11084095 as putative causal variants of the association at *SIGLEC5*

The results of this thesis indicated that rs4284742 is a putative causal variant of the associations with PD. The effect allele of rs4284742, which is the common allele, provided strong MAFB binding affinity compared to the non-effect allele in the Supershift-EMSA (**Figure 13**). Allele-specific enhancer activity was also shown using reporter gene assays (**Figure 26A**). The common G-allele showed stronger enhancer activity than the rare A-allele. These experiments indicate that rs4284742 is a putative causal variant of the association with PD. Additionally, it was shown that CRISPR-dCas9 activation of the genomic sequences at rs4284742 activated *SIGLEC5* expression (**Figure 30A**). These results suggest that MAFB regulation is linked with activation of *SIGLEC5* expression and increases the risk for early-onset periodontitis. MAFB belongs to the subfamily of the large Maf transcription factors (Santos-Gallego 2016) and it is expressed by monocytes and is required for differentiation to macrophages (Gemelli et al. 2006; Kelly et al. 2000). Furthermore, MAFB negatively regulates osteoclast generation via inhibition of the transcription factor NFATc1 and the osteoclast-associated receptor (OSCAR) (Kim et al. 2007), implying a functional context of the genetic association with osteoclast differentiation. Moreover, MAFB was also shown to promote sprouting angiogenesis (Jeong et al. 2017). During wound healing of tissue injuries, angiogenic capillary sprouts invade the wound clot and organize into a microvascular network throughout the granulation tissue (reviewed in (Tonnesen et al. 2000). A challenge for healing of aseptic tissue injuries is discrimination from infections to those tissue injuries. This is achieved by the innate immune system through danger- or pathogens-associated molecular patterns, DAMPs and PAMPs, which are recognized through pattern recognition receptors, such as Toll-like receptors (TLR) and/or Nod-like receptors (NLR). Sialoside-based pattern recognition was shown to selectively suppress the immune response to DAMPs, which suggests a mechanism by which aseptic tissue injury and infection are distinguished (Chen et al. 2009). Normally it is a well controlled mechanism to avoid autoimmune destruction. However, many viruses and pathogenic bacteria express sialidase as virulence factors ((Crennell et al. 1993), reviewed in (Crennell et al. 1993; Drzeniek 1972)). It is discussed that microbial-expressed sialidase might have the potential to abrogate

the SIGLEC-mediated inhibitory effects of DAMPs on the innate immune system during healing of injured tissues. As a result of this, DAMPs and PAMPs would become indistinguishable, which could provide an explanation for massive inflammation and tissue destruction (Liu et al. 2009), which is also typically seen in severe periodontitis phenotypes with a rapid rate of progression (AgP). This may also explain why, in specific situations, *SIGLEC5* expression may increase the risk for alveolar bone loss.

Notably, the maximal position weight matrix (PWM) score of MAFB was only 82%. The incomplete similarity score could be due to the fact that the specificity of protein-DNA binding does not solely depend on the DNA sequence, but also on the 3D structure of DNA and TF protein macromolecules (Rohs et al. 2010). This result in variation of functional binding motifs and accordingly, the predicative accuracy of a PWM should be always interpreted with caution (Weirauch et al. 2013). However, the rare allele further reduced the matrix similarity to 69%, which was corresponding with reduced MAFB binding in the EMSA and reduced reporter gene activity.

Notably, the strongest eQTL of rs4284742 affected *SIGLEC5* expression *(e.g.* $p = 6.4x10^{-23}$ in peripheral blood) (Zeller et al. 2010). Validation of reported eQTLs is necessary, because the assignment of a regulatory effect of an associated region on a specific gene by an eQTL alone is error prone. This is due to the fact, that eQTLs are statistical observations (Zeng et al. 2017) and enhancers do not necessarily influence expression of the nearest gene but often map large distances away from their actual targets. Thus, appointing the target gene of an association requires direct evidence form molecular biological experiments. To validate the regulatory potential of the associated haplotype blocks, CRISPR-mediated activation was performed in HeLa cells. CRISPR-mediated activation of enhancer regions is a tool for efficient analysis of enhancer functions and to analyze if the enhancer region regulates transcription of a target gene (Ginley-Hidinger et al. 2019; Li et al. 2020b). Comparison of two different activation systems, SAM and VPR, which has been shown to be the most potent ones, featured that the promoter of *CTD-2353F22.1* was stronger upregulated with the CRISPR/dCas9 SAM activation system (**Figure 28A-B**). This system has been shown previously to result in a more potent activation than the VPR system in some contexts (Chavez et al. 2016). This could be because of the additional activator helper complex. Regarding the more efficient upregulation with the SAM activation system, all CRISPRa, also for the genomic region at rs4284742, were performed with this method. CRISPRa validated that the enhancer at rs4284742 strongly regulates *SIGLEC5* expression, indicating that *SIGLEC5* is the target gene of the association with PD. *SIGLEC5* is an inhibitory receptor that regulate inflammation mediated by PAMPs and DAMPs (Pillai et al.

2012). *SIGLEC5* have been shown to recognize and bind β-protein of group B streptococcus, resulting in recruitment of SHP and modulation of host immune response (Ali et al. 2014; Carlin et al. 2009). Interestingly, *SIGLEC5* also has been shown to bind to meningococcal sialyated lipopolysaccharide (LPS) of *Neisseria meningitides* and that this binding leads to enhance bacterial uptake (Jones et al. 2003). Thus, it could be that some inhibitory SIGLECS like *SIGLEC5*, are used by pathogens for cellular entry (Pillai et al. 2012), indicating an important role of *SIGLEC5* in the pathogenesis of PD.

Another TFBS was predicted at the position of rs11084095 for the TF ERG with 93% PWM matrix similarity. The reduction of ERG binding in the background of the rare effect allele and the allele-specific enhancer activity of the genomic region at rs11084095 was confirmed using Supershift-EMSA (**Figure 15**) and reporter gene assays (**Figure 26B**), indicating that rs11084095 is a putative causal variant for the association of PD. ERG is an important TF for endothelial homeostasis, a system control state that encompasses acute responses to injury to support repair of damaged endothelium (Heiss et al. 2015). The vascular structure is affected by changes in endothelial homeostasis (*i.e.,* endothelial injury) by interacting with extracellular matrix turnover and influence endothelial membrane function and adhesiveness to proteins of the coagulation cascade and platelets, membrane permeability and integrity (Gulino-Debrac 2013). ERG is mainly expressed in endothelial cells but also in leukocytes (Uhlen et al. 2015). Interestingly, in endothelial cells (ECs) inhibition of ERG expression was shown to increase expression of the intracellular adhesion molecule (ICAM-1), which is a pro-inflammatory transmembrane glycoprotein, and to induce NF-κB (nuclear factor kappa-light-chain enhancer of activated B cells) gene expression (Dryden et al. 2012). Activation of NF-κB and the NF-κB pathway induces the transcription of multiple genes involved in inflammation including chemokines and cytokines (Arabaci et al. 2010). This may explain why loss of ERG expression result in inflammatory diseases like periodontitis. In addition, activation of NF-κB is associated with many systemic diseases such as atherosclerosis and coronary artery disease (Brand et al. 1996), which are epidemiologically linked with PD (Munz et al. 2018). Additionally, multiple studies have shown that ERG plays an important role in endothelial apoptosis and angiogenesis, the growth of blood vessels from existing ones (Birdsey et al. 2008). The vascular endothelial growth factor (VEGF) plays an important role in neo-vascularization (Dvorak et al. 1995). One study proved that VEGF concentration is increased in the gingival crevicular fluid (GCF) of PD cases (R et al. 2014). Another result of angiogenesis in CP cases is the increase in vessel profiles in periodontal pocket walls (Chapple et al. 2000).

It is reported by the GWAS catalog (Welter et al. 2014) that genetic variations of ERG are associated with numerous blood cell traits and blood pressure, as well as bone mineral density and osteoarthritis. Notably, it was recently shown that soluble *SIGLEC5* appears to be efficient in blocking leukocyte-rolling over P-selectin (platelet selectin) and E-selectin (endothelial selectin) (Pepin et al. 2016). Furthermore, it was shown in human endothelial cells that ERG associates with enhancers of von Willebrand factor (vWF) which is an adhesive and multimeric glycoprotein (Kalna et al. 2019) and that vWF may act as a ligand for *SIGLEC5* (Pegon et al. 2012), suggesting a possible link between endothelial homeostasis and *SIGLEC5* function.

Using CRISPRa of the genomic region at rs11084095 with up to 200 bp distance from the SNP, *SIGLEC5* could not be validated as target gene of the association. However, CRISPRa of a region close by rs11084095 (with up to 2.6 kb distance) showed weak upregulation of *SIGLEC5* (**Figure 30C**) indicating that *SIGLEC5* might be the target gene of the association. Notably, the strongest eQTL of rs11084095 affected *SIGLEC5* expression *(e.g.* $p = 6.4x10^{-23}$ in monocytes) (Zeller et al. 2010) but no other genes. Reasons for low upregulation could be technical problems that impeded the function of the CRISPRa system at this haplotype block, *e.g.* poor targeting or binding of the gRNAs at the selected PAM sequences. Future studies could use Chromatin Conformation Capture (3C) to prove the *cis*-regulatory effects of rs11084095 on *SIGLEC5* expression indicated by the eQTL studies (Consortium 2013; Zeller et al. 2010). Chromatin Conformation Capture (3C) is a technique that can demonstrate the interaction of two genomic loci. For 3C, cells are crosslinked with formaldehyde to fixate chromatin followed by restriction digest. Chromatin fragments are ligated and the DNA is purified. The interaction between two genomic regions of interest can be detected by PCR or RT-PCR (Sati and Cavalli 2017).

For the common allele of rs34984145, a TFBS for BACH2 was predicted with a PWM matrix similarity of 83%. The antibody specific EMSA gave evidence for BACH2 binding at the sequence of this SNP (**Figure 17**) and the luciferase reporter gene showed weak enhancer activity (**Figure 26C**). However, unlike rs4284742 and rs11084095, the difference of the reporter gene activity between the alleles of rs34984145 was low (4.2-fold vs. 2.3-fold), which does not suggest meaningful biological relevance but rather implies natural biological variation between the experiments. This corresponds to the low information content of the SNP allele in the BACH2 consensus motif which means that BACH2 binds at this DNA sequence, however, the different alleles do not significantly change the binding affinities, implying that rs34984145 is not a causal variant.

In conclusion, rs4284742 and rs11084095 are functional and putative causal variants that reduce enhancer activity and binding affinity of the TFs MAFB and ERG. CRISPRa of *SIGLEC5* expression proved *SIGLEC5* as the target gene of the associations. Based on the functions of MAFB, ERG and *SIGLEC5*, a functional context of the associations with impaired regulation of endothelial homeostasis and healing of aseptic tissue injuries can be hypothesized.

5.1.3 rs1122900 as putative causal variant of the association with *CTD-2353F22.1*

For the common allele of rs1122900 a TFBS for the TF PR Domain Zinc Finger Protein 14 (PRDM14) with a matrix similarity of 89% was predicted. The reduction of PRDM14 binding in the background of the rare effect allele and the allele-specific enhancer activity of the genomic region at rs1122900 was confirmed using Supershift-EMSA (**Figure 20**) and reporter gene assay (**Figure 27A**), indicating that rs1122900 is a putative causal variant for the association of PD. In addition, CRISPR-dCas9 activation of the genomic sequences at rs1122900 showed very weak effects on *CTD-2353F22.1* expression (**Figure 31B**), suggestting that *CTD-2353F22* might be the target gene of the association. PRDM14 is a crucial epigenetic regulator in primordial germ cells (Sybirna et al. 2020). However, in recent years PRDM14 has been shown to play a key role in many types of human cancer (Tracey and Justice 2019) especially in lymphoblastic leukemia (Dettman et al. 2011) and lymphoblastic lymphoma (Dettman and Justice 2008). Acute lymphoblastic leukemia (ALL) is a type of cancer that is characterized by infiltration of the bone marrow and blood with immature lymphoblasts. The bone marrow is the soft inner part of the bones, where new blood cells are produces, it contains many different cell types like osteoclasts, osteoblasts and endothelial cells and lymphoblasts (Anderson et al. 2020). However, PRDM14-induced T-cell acute lymphoblastic leukemia (T-ALL) was exhibited to be driven by NOTCH1 in mice by binding of PRDM14 to the NOTCH1 locus (Carofino et al. 2016). Downregulation of NOTCH1 was shown to contribute to the severity of bone loss in AgP (Mijailovic et al. 2020). This could explain why decrease of PRDM14 may increase the risk for PD.

Using CRISPRa of the genomic region at rs1122900 spanning up to 200 bp distance from the SNP, *CTD-2353F22.1* could not be validated as target gene of the association. CRISPRa of more remote of rs1122900 (with up to 1.8 kb distance) showed very weak upregulation of *CTD-2353F22.1* indicating that *CTD-2353F22.1* might be the target gene of the association. However, all reported eQTLs of rs1122900 indicated a regulatory effect on *CTD-2353F22.1* expression, with $p < 9 \times 10^{-4}$ which indicates that *CTD-2353F22.1* might the target gene of the

association. The strongest eQTL of rs1122900 on *CTD-2353F22.1* expression was reported in whole blood with $p = 6.1 \times 10^{-15}$ (Consortium 2013). Reasons for low upregulation by CRISPRa were discussed above.

For the common allele of rs6887423 a TFBS for the TF aryl hydrocarbon receptor (AHR) with a matrix similarity of 100% was predicted. The antibody specific EMSA gave evidence for AHR binding at the sequence of this SNP (**Figure 22**). However, reporter gene assay showed no effect (**Figure 27B**), indicating that rs6887423 does not have any regulatory effect. This corresponds with the low information content of the Supershift-EMSA which only showed a reduction of 32% of AHR antibody by binding to the rare T-allele (**Figure 23**). This shows that the different alleles do not significantly change the binding affinities, implying that rs6887423 is not a causal variant. CRISPR-mediated activation of the genomic region at rs6887423 showed weak upregulation of *CTD-2353F22.1* (**Figure 31C**) indicating that the region has a *cis*-regulatory effect on the expression of *CTD-2353F22.1*, but not rs6887423 itself. This confirms the predicted eQTLs for rs6887423, which were reported for *CTD-2353F22.1* and not for other genes.

Another predicted TFBS was at the position of rs56038114 for the TF GATA1 with 96% PWM matrix similarity. The antibody specific EMSA gave evidence for GATA1 binding at the sequence of this SNP although GATA1 binds stronger to the rare allele than to the common allele (**Figure 24**). The luciferase reporter gene exhibited only very weak enhancer activity for the rare C-allele (**Figure 27C**) which indicates the same effect direction. However, the difference of the reporter gene activity between the alleles of rs56038114 was very small (2.2-fold vs. 1-fold), which does not suggest meaningful biological relevance but rather implies natural biological variation between the experiments. Likewise, CRISPRa did not exhibit an effect on the expression of *CTD-2353F22.1* (**Figure 31D**), implying that rs56038114 is not a causal variant.

In conclusion, rs1122900 is a functional and putative causal variants that reduce enhancer activity and binding affinity of the TF PRDM14 and proved *CTD-2353F22.1* as the target gene of the association.

One limitation of this thesis was that the binding of the TFs to the corresponding alleles was only shown *in vitro*. Although the binding of the TFs was shown with the nuclear protein extract of the cells of interest, future studies could use Chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-Seq) to determine whether the protein binds to the specific chromatin sequence in the native context *in vivo* (Wiehle and Breiling 2016). To this end, cells

of interest are crosslinked to covalently stabilize and fixate the protein-DNA complexes. Because protein-DNA interactions occur primarily in the nucleus, cell membranes are dissolved to solubilize protein-DNA complexes. To analyze the protein-DNA interactions the extracted genomic DNA has to be sheared enzymatically or by sonication. ChIP-validated antibodies are then used to immunoprecipitate and isolate the specific protein-DNA interactions of interest. After washing and purifying, the protein-DNA complex is eluted and can be analyzed using DNA specific primers (Gade and Kalvakolanu 2012). Using ChIP sequencing, protein-DNA interactions can be analyzed in a genome-wide scale.

Another possible limitation of this thesis was that the analysis was confined to SNPs that indicated strong linkage according to r^2>0.8. This LD measure was used because the r^2 coefficient of correlation takes account of allele frequency. Strong LD indicated by D' but not by r^2 includes alleles that are inherited with the particular GWAS lead SNP but are not carried by the majority of cases due to their frequency. Such alleles are not suggestive as causal variants because they do not explain the association for most cases. However, this does not exclude that rare susceptibly variants at *SIGLEC5* exist but such variants have no disease relevant role in the general population and their effects would manifest in individual cases.

Another limitation was that reporter gene assays and CRISPRa were performed in HeLa cells instead of blood cells. Since HeLa cells can be transfected more efficiently than PBMCs or B lymphocytes, which improves the detection of differences in reporter gene expression, the reporter gene experiments were performed in HeLa cells. Transfection of the immortalized Raji cell line also showed bad transfection efficiency. Prior to transfection of HeLa cells it was verified that the genes of interest were expressed in HeLa cells to ensure a validation of the regulatory effect. After transfection with the CRISPRa plasmids, Raji cells showed poor survival, probably because of DNA toxicity (Kim et al. 2014), whereas HeLa cells showed high survival after transfection of the CRISPRa system. Even if an enhancer was not currently active in the assayed cells it was recently demonstrated that recruitment of a strong transcriptional activator to an enhancer can be sufficient to drive target gene expression (Simeonov et al. 2017). However, HeLa cells are of different developmental origin than blood or Raji cells and it is possible that their chromatin state to *CTD-2353F22.1* upregulation differs from blood cells.

5.1.4 Overexpression of *CTD-2353F22.1* upregulates the gene sets "angiogenesis" and "TNF alpha signaling via NFKB"

Because *CTD-2353F22.1* seems to be the target gene of the association with PD one aim was to functional characterize *CTD-2353F22.1*, to identify responding genes and pathways, and to determine its function in the cell. Therefore, RNA-sequencing with *CTD-2353F22.1* CRISPRa HeLa cells was performed.

RNA-sequencing with *CTD-2353F22.1* CRISPRa HeLa cells showed that the most significant upregulated gene after *CTD-2353F22.1* was lipoprotein lipase (LPL) with an adjusted p-value of $p_{adj} = 4.9 \times 10^{-11}$ (**Table 20**). LPL was also shown to play an important role in the gene set "angiogenesis" (**Figure 33A**). LPL is an important enzyme for the triglyceride hydrolysis of lipoproteins and is mainly produced in adipose and skeletal muscle tissue whereat small amounts are also synthesized in endothelial cells. LPL was shown to significantly suppresses TNF alpha- induced gene expression as a result of its inhibition of NF-κB activity (Kota et al. 2005). As described above, activation of NF-κB induces the transcription of multiple genes involved in inflammation including chemokines and cytokines (Arabaci et al. 2010). Activation of NF-κB is associated with many systemic diseases such as atherosclerosis and coronary artery disease (Brand et al. 1996) that are linked to PD (Munz et al. 2018).

Serpine proteinase inhibitor, also known as endothelial plasminogen activator inhibitor (SERPINE1) was the third most significant downregulated gene after CRISPRa of *CTD-2353F22.1* with an adjusted p-value of $p_{adj} = 4.78 \times 10^{-9}$ (**Table 20**). SERPINE1 is a member of the superfamily of serpins, and inhibits the tissue-type and the urinary-type plasminogen activator, the two plasminogen activators able to activate plasminogen. The plasminogen activator inhibitor plays an important role in coagulation and fibrosis (Kruithof 2008). Hyperactivation of fibroblasts and excessive accumulation of extracellular matrix (ECM) protein in the wound area is a result of deregulation of wound healing and leads to manifestation of fibrosis (Ghosh and Vaughan 2012). SERPINE1 is also a key regulator in wound healing. Tissue homeostasis and wound healing is regulated by SERPINE1 by inhibiting plasmin-mediated matrix metalloproteinases (MMPs) (Flevaris and Vaughan 2017). Deficient levels of SERPINE1 cause healing anomalies like impaired wound healing, excessive bleeding or thrombosis (Simone et al. 2014). Gingival wound healing plays an important role in the pathogenesis of PD. To prevent the invasion of bacteria into tissues and to avoid developing of a chronic disease, fast wound healing is of critical importance (Smith et al. 2015). Downregulation of SERPINE1 may result in a bacterial contamination and delayed wound healing, resulting in chronic manifestation of PD. The dot plot in **Figure 34** shows enrichment

of gene clusters related to the biological processes after CRISPR-mediated activation of *CTD-2353F22.1*. This dot plot also shows significant downregulation of the biological processes "response to wounding", "blood vessel morphogenesis2, "angiogenesis", "inflammatory response", "wound healing", extracellular structure organization" and "response to bacteria". All these biological processes are critical in the pathogenesis of PD as described above.

Results of RNA-seq also showed the highest effect size for the gene set "TNF alpha signaling via NFKB" with an AUC = 0.63 (p_{adj} = 1.0 x 10^{-17}) (**Figure 33D**). Tumor necrosis factor α (TNFα) plays a crucial role in PD by mediating periodontal tissue destruction through stimulating bone resorption. It was shown to contribute to periodontal damage by its direct effect on osteoclastogenesis and by amplification of inflammatory immune reactions (Di Benedetto et al. 2013). This indicates a functional context of *CTD-2353F22.1* that has a described relevance in periodontal disease.

One limitation of this thesis was, as described above, that CRISPRa was performed in HeLa cells because of poor survival of Raji cells after transfection with CRISPRa plasmids. However, the observed effects give evidence that HeLa cells are an appropriate cell model for these experiments.

RNA-seq data and the identified TFs binding at rs4284742, rs11084095 and rs1122900 indicate that the associations both with *SIGLEC5* and *CTD-2353F22.1* have a functional context with impaired regulation of endothelial homeostasis, angiogenesis and healing of aseptic tissue injuries. The results also give evidence that *SIGLEC5* and *CTD-2353F22.1* have a role in the same functional context.

5.2 Identification and characterization of miRNA miR-374b-5p

The identification of differentially expressed miRNAs and the associated target genes of their regulation directly has the potential to elucidate the molecular mechanisms involved in the pathogenesis of diseases such as PD. In preliminary works, venous blood of PD patients was analyzed for differently expressed miRNAs compared to healthy controls. It is possible that individual cases of the samples developed periodontitis for different reasons. In a few cases periodontitis was diagnosed in young adults (<30 years of age) which indicates aggressive periodontitis whereas other cases developed periodontitis > 50 years of age (chronic periodontitis). This implies a general role of miRNA-374b-5p in the etiology of periodontitis and not only in a specific PD disease (Mueller et al. 2021).

Out of 863 miRNAs, 29 miRNAs showed significantly higher expression ($P < 5x10^{-6}$) in blood from PD cases compared to healthy controls, 23 of which showed >2-fold upregulation (**Table 21**). MiRNA hsa-miR-374b-5p was the only miRNA which was >2-fold upregulated in blood of PD cases ($P = 2.1x10^{-6}$) and additionally reported by two other miRNA array studies to be > 2-fold upregulated in PD cases or periodontal cells and was chosen for the identification of its target gene. Three different cultures of pGF were transfected with hsa-miR-374b-5p mimic and in the mRNA gene expression profiling experiment, *UHMK1* was observed as the most significant downregulated gene (**Table 22**). For analysis of the gene expression profiling experiment, all replicates were combined and the effects of the single experiments were added collectively, indicating robustness of the finding that *UHMK1* is downregulated by hsa-miR-374b-5p (Mueller et al. 2021). U2AF homology motif (UHM) kinase 1 (*UHMK1*) is a serine/threonine protein kinase that promotes cell cycle progression through G1 by phosphorylation of the cyclin-dependent kinase inhibitor 1B (provided by RefSeq (Wei et al. 2019). The GWAS catalog reported that *UHMK1* is genomewide associated with bone mineral density (Choi et al. 2016). Additionally, it was shown in functional experiments that *UHMK1* expression increased osteoblast and decreased osteoclast differentiation while *UHMK1* knockdown decreased osteoblast and increased osteoclast differentiation (Choi et al. 2016). This suggests a role of *UHMK1* in the regulation of bone homeostasis. Additionally, *UHMK1* does also promote cell cycle progression (Wei et al. 2019). Successful resolution of inflammation requires transition from the inflammatory to the proliferative phase, an important step in the healing process of oral inflammation.

However, the observation that *UHMK1* was downregulated after miRNA hsa-miR-374b-5p transfection in gingival fibroblasts was not significant after correction for genome-wide testing and could be a false positive finding and did not provide direct biological evidence for an inhibitory effect of hsa-miR-374b-5p. Thus, it was necessary to validate on the molecular level that hsa-miR-374b-5p downregulates the gene *UHMK1* and that it is a target gene of the analyzed miRNA. Therefore, the 3'UTR sequence stretches of *UHMK1* that included the bioinformatically predicted conserved binding sites for miR-374b-5p was cloned to a reporter gene plasmid. The 3'UTR of *UHMK1* is comprised of 7.200 kb. This exceeds the average length of human 3'UTR of 800 nucleotides ~10-fold (Mignone et al. 2002). 3'UTRs were likely shaped under selection to eliminate or acquire miRNA target sites (Stark et al. 2005). Correspondingly, the average 3'UTR length differs considerably in correlation with tissue specificity of gene expression. By way of example, genes that are expressed ubiquitously and involved in basic cellular processes such as ribosomal protein genes have 6-fold shorter 3'UTRs

than tissue specific genes. The length of the 3'UTR of *UHMK1* indicates stringent regulation, which might reflect potential damage from misregulation. 95% of genes with conserved target sites have only one site for a single miRNA, but most identified genetic targets contain multiple sites for the specific miRNA (Abrahante et al. 2003; Brennecke et al. 2005). In the current thesis, two conserved 8mer binding sites for hsa-miR-374b-5p in the 3'UTR of *UHMK1* were identified. This supports the specific regulatory role of hsa-miR-374b-5p for *UHMK1* expression adding to the results of *in vitro* experiments. For *in vitro* experiments, HeLa cells were used as they are an accepted cell model that allows efficient transfection. PD relevant cells were not used because mRNA degradation by miRNA binding at the 3'UTR is a general mechanism which is independent of the cell type and the genomic context. That hsa-miR-374b-5p binds to the 3'UTR of *UHMK1* to suppress *UHMK1* gene expression at the post-transcriptional level was shown by measuring both luciferase protein activity (**Figure 41**) and luciferase transcript levels (**Figure 40**). Although, the observed effect size was small, it was consistent with other studies that quantified luciferase activity to validate the regulatory effect of miRNAs (Khandelwal et al. 2019; Li et al. 2020a). This indicates that suppression mediated by miRNA is exerted at the mRNA level, probably by mRNA degradation, and not on the protein level through translational repression (Ha and Kim 2014).

A biological pathway analysis was not performed. The main interest of this study was the identification of a specific target gene of miR-374b-5p regulation. Furthermore, it is questionable if the small *in vitro* effect of the overexpressed miRNA would result in significant changes of an entire biological pathway. Nonetheless, in a natural *in vivo* situation, that is subject to multiple physiological factors that often act additive and over time, the effects of miRNA regulation may also be observed on the level of biological pathways.

In conclusion, this thesis identified miRNA hsa-miR-374b-5p to be upregulated in blood and gingiva of periodontitis patients and identified miRNA hsa-miR-374b-5p as a transcriptional repressor of *UHMK1*.

6 Conclusion

This thesis showed that rs4284742 and rs11084095 are functional and putative causal variants that reduce enhancer activity and binding affinity of the TFs MAFB and ERG. CRISPRa of genomic regions at rs4284742 and rs11084095 proved *SIGLEC5* as the target gene of the associations. Based on the functions of MAFB, ERG and *SIGLEC5*, a functional context of the associations with impaired regulation of endothelial homeostasis and healing of aseptic tissue injuries can be hypothesized. Additionally, rs1122900 was identified to be a functional putative causal variant of the association with PD. rs1122900 reduces enhancer activity and binding affinity of the TF PRDM14. This thesis showed that *CTD-2353F22.1* is the target gene of the association. RNA-seq of *CTD-2353F22.1* CRISPRa cells indicate a functional context of the association with angiogenesis and impaired wound healing.

In addition, this thesis identified miRNA hsa-miR-374b-5p to be upregulated in blood and gingiva of periodontitis patients and identified miRNA hsa-miR-374b-5p as a transcriptional repressor of *UHMK1*. This gene has a role in the regulation of osteoblasts and osteoclasts homoeostasis. This thesis gave mechanistic evidence that overexpression of hsa-miR-374b-5p results in decreased transcript levels and protein activity of a reporter gene that expressed the *UHMK1* 3'UTR. By using gingival fibroblasts and HeLa cells, it was shown that the molecular mechanism of this effect is independent of a specific cell type. In summary, these experiments indicated that hsa-miR-374b-5p interacts with the 3' UTR of *UHMK1* to suppress gene expression through mRNA degradation.

By providing specific information on the genetic risk factors of periodontitis and by identifiying the possible pathways with which these factors interfere, investigation like the one presented here can contribute to eventually find optimized treatment options for periodontitis.

7 References

Abrahante JE, Daul AL, Li M, Volk ML, Tennessen JM, Miller EA, Rougvie AE. 2003. The caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microrna. Dev Cell. 4(5):625-637.

Acharya CR, McCarthy JM, Owzar K, Allen AS. 2016. Exploiting expression patterns across multiple tissues to map expression quantitative trait loci. BMC Bioinformatics. 17:257.

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microrna target sites in mammalian mrnas. Elife. 4.

Ali SR, Fong JJ, Carlin AF, Busch TD, Linden R, Angata T, Areschoug T, Parast M, Varki N, Murray J et al. 2014. Siglec-5 and siglec-14 are polymorphic paired receptors that modulate neutrophil and amnion signaling responses to group b streptococcus. The Journal of experimental medicine. 211(6):1231-1242.

Amaral SA, Pereira TSF, Brito JAR, Cortelli SC, Cortelli JR, Gomez RS, Costa FO, Miranda Cota LO. 2019. Comparison of mirna expression profiles in individuals with chronic or aggressive periodontitis. Oral Dis. 25(2):561-568.

Ambrosini G, Groux R, Bucher P. 2018. Pwmscan: A fast tool for scanning entire genomes with a position-specific weight matrix. Bioinformatics. 34(14):2483-2484.

Anderson D, Skut P, Hughes AM, Ferrari E, Tickner J, Xu J, Mullin BH, Tang D, Malinge S, Kees UR et al. 2020. The bone marrow microenvironment of pre-b acute lymphoblastic leukemia at single-cell resolution. Sci Rep. 10(1):19173.

Angata T, Hayakawa T, Yamanaka M, Varki A, Nakamura M. 2006. Discovery of siglec-14, a novel sialic acid receptor undergoing concerted evolution with siglec-5 in primates. FASEB J. 20(12):1964-1973.

Arabaci T, Cicek Y, Canakci V, Canakci CF, Ozgoz M, Albayrak M, Keles ON. 2010. Immunohistochemical and stereologic analysis of nf-kappab activation in chronic periodontitis. Eur J Dent. 4(4):454-461.

Armitage GC. 1999. Development of a classification system for periodontal diseases and conditions. Ann Periodontol. 4(1):1-6.

Bao X, Weiner J, 3rd, Meckes O, Dommisch H, Schaefer AS. 2021. Entamoeba gingivalis exerts severe pathogenic effects on the oral mucosa. J Dent Res.220345211004498.

Bao X, Wiehe R, Dommisch H, Schaefer AS. 2020. Entamoeba gingivalis causes oral inflammation and tissue destruction. J Dent Res. 99(5):561-567.

Bauer P. 2017. Luciferase reporter gene assays. In: Schwab M, editor. Encyclopedia of cancer. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 2543-2547.

Birdsey GM, Dryden NH, Amsellem V, Gebhardt F, Sahnan K, Haskard DO, Dejana E, Mason JC, Randi AM. 2008. Transcription factor erg regulates angiogenesis and endothelial apoptosis through ve-cadherin. Blood. 111(7):3498-3506.

Blackwood EM, Kadonaga JT. 1998. Going the distance: A current view of enhancer action. Science. 281(5373):60-63.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: From polygenic to omnigenic. Cell. 169(7):1177-1186.

Brand K, Page S, Rogler G, Bartsch A, Brandl R, Knuechel R, Page M, Kaltschmidt C, Baeuerle PA, Neumeier D. 1996. Activated transcription factor nuclear factor-kappa b is present in the atherosclerotic lesion. J Clin Invest. 97(7):1715-1722.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microrna-target recognition. PLoS Biol. 3(3):e85.

Brieuc MS, Naish KA. 2011. Detecting signatures of positive selection in partial sequences generated on a large scale: Pitfalls, procedures and resources. Mol Ecol Resour. 11 Suppl 1:172-183.

Brookes AJ. 1999. The essence of snps. Gene. 234(2):177-186.

Bryois J, Buil A, Evans DM, Kemp JP, Montgomery SB, Conrad DF, Ho KM, Ring S, Hurles M, Deloukas P et al. 2014. Cis and trans effects of human genomic variants on gene expression. PLoS Genet. 10(7):e1004461.

Cao H, Crocker PR. 2011. Evolution of cd33-related siglecs: Regulating host immune functions and escaping pathogen exploitation? Immunology. 132(1):18-26.

Carey MF, Peterson CL, Smale ST. 2012. Experimental strategies for the identification of DNA-binding proteins. Cold Spring Harb Protoc. 2012(1):18-33.

Carlin AF, Chang YC, Areschoug T, Lindahl G, Hurtado-Ziola N, King CC, Varki A, Nizet V. 2009. Group b streptococcus suppression of phagocyte functions by protein-mediated engagement of human siglec-5. The Journal of experimental medicine. 206(8):1691-1699.

Carofino BL, Ayanga B, Tracey LJ, Brooke-Bisschop T, Justice MJ. 2016. Prdm14 promotes rag-dependent notch1 driver mutations in mouse t-all. Biol Open. 5(5):645-653.

Chapple CC, Kumar RK, Hunter N. 2000. Vascular remodelling in chronic inflammatory periodontal disease. J Oral Pathol Med. 29(10):500-506.

Chattopadhyay A, Lu TP. 2019. Gene-gene interaction: The curse of dimensionality. Ann Transl Med. 7(24):813.

Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, E PRI, Lin S, Kiani S, Guzman CD, Wiegand DJ et al. 2015. Highly efficient cas9-mediated transcriptional programming. Nat Methods. 12(4):326-328.

Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJ, Buchthal J et al. 2016. Comparison of cas9 activators in multiple species. Nat Methods. 13(7):563-567.

Chen GY, Tang J, Zheng P, Liu Y. 2009. Cd24 and siglec-10 selectively repress tissue damage-induced immune responses. Science. 323(5922):1722-1725.

Chodosh LA. 2001. Uv crosslinking of proteins to nucleic acids. Curr Protoc Mol Biol. Chapter 12:Unit 12 15.

Choi HJ, Park H, Zhang L, Kim JH, Kim YA, Yang JY, Pei YF, Tian Q, Shen H, Hwang JY et al. 2016. Genome-wide association study in east asians suggests uhmk1 as a novel bone mineral density susceptibility gene. Bone. 91:113-121.

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

Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the encode pilot project. Nature. 447(7146):799-816.

Consortium GT. 2013. The genotype-tissue expression (gtex) project. Nat Genet. 45(6):580-585.

Corey LA, Nance WE, Hofstede P, Schenkein HA. 1993. Self-reported periodontal disease in a virginia twin population. J Periodontol. 64(12):1205-1208.

Crennell SJ, Garman EF, Laver WG, Vimr ER, Taylor GL. 1993. Crystal structure of a bacterial sialidase (from salmonella typhimurium lt2) shows the same fold as an influenza virus neuraminidase. Proc Natl Acad Sci U S A. 90(21):9852-9856.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA et al. 2010. Histone h3k27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 107(50):21931-21936.

Crocker PR, Varki A. 2001. Siglecs, sialic acids and innate immunity. Trends in immunology. 22(6):337-342.

Darveau RP, Tanner A, Page RC. 1997. The microbial challenge in periodontitis. Periodontol 2000. 14:12-32.

de Carvalho FM, Tinoco EM, Govil M, Marazita ML, Vieira AR. 2009. Aggressive periodontitis is likely influenced by a few small effect genes. J Clin Periodontol. 36(6):468-473.

Dettman EJ, Justice MJ. 2008. The zinc finger set domain gene prdm14 is overexpressed in lymphoblastic lymphomas with retroviral insertions at evi32. PLoS One. 3(11):e3823.

Dettman EJ, Simko SJ, Ayanga B, Carofino BL, Margolin JF, Morse HC, 3rd, Justice MJ. 2011. Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors. Oncogene. 30(25):2859-2873.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. J Bacteriol. 192(19):5002-5017.

Di Benedetto A, Gigante I, Colucci S, Grano M. 2013. Periodontal disease: Linking the primary inflammation to bone loss. Clinical & developmental immunology. 2013:503754.

Dryden NH, Sperone A, Martin-Almedina S, Hannah RL, Birdsey GM, Khan ST, Layhadi JA, Mason JC, Haskard DO, Gottgens B et al. 2012. The transcription factor erg controls endothelial cell quiescence by repressing activity of nuclear factor (nf)-kappab p65. J Biol Chem. 287(15):12331-12342.

Drzeniek R. 1972. Viral and bacterial neuraminidases. Curr Top Microbiol Immunol. 59:35-74.

Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G. 2012. Microrнas associated with the different human argonaute proteins. Nucleic Acids Res. 40(19):9850-9862.

Dvorak HF, Brown LF, Detmar M, Dvorak AM. 1995. Vascular permeability factor/vascular endothelial growth factor, microvascular hyperpermeability, and angiogenesis. Am J Pathol. 146(5):1029-1039.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 11(6):446-450.

Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ. 2012. Prevalence of periodontitis in adults in the united states: 2009 and 2010. J Dent Res. 91(10):914-920.

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

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet. 7(2):85-97.

Flevaris P, Vaughan D. 2017. The role of plasminogen activator inhibitor type-1 in fibrosis. Semin Thromb Hemost. 43(2):169-177.

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.

Fujimori K, Yoneda T, Tomofuji T, Ekuni D, Azuma T, Maruyama T, Mizuno H, Sugiura Y, Morita M. 2019. Detection of salivary mirnas reflecting chronic periodontitis: A pilot study. Molecules. 24(6).

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M et al. 2002. The structure of haplotype blocks in the human genome. Science. 296(5576):2225-2229.

Gade P, Kalvakolanu DV. 2012. Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. Methods Mol Biol. 809:85-104.

Gallagher MD, Chen-Plotkin AS. 2018. The post-gwas era: From association to function. Am J Hum Genet. 102(5):717-730.

Gemelli C, Montanari M, Tenedini E, Zanocco Marani T, Vignudelli T, Siena M, Zini R, Salati S, Tagliafico E, Manfredini R et al. 2006. Virally mediated mafb transduction induces the monocyte commitment of human cd34+ hematopoietic stem/progenitor cells. Cell Death Differ. 13(10):1686-1696.

Genco RJ, Borgnakke WS. 2013. Risk factors for periodontal disease. Periodontol 2000. 62(1):59-94.

Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. Nature. 467(7319):1061-1073.

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA et al. 2015. A global reference for human genetic variation. Nature. 526(7571):68-74.

Ghosh AK, Vaughan DE. 2012. Pai-1 in tissue fibrosis. J Cell Physiol. 227(2):493-507.

Ginley-Hidinger M, Carleton JB, Rodriguez AC, Berrett KC, Gertz J. 2019. Sufficiency analysis of estrogen responsive enhancers using synthetic activators. Life Sci Alliance. 2(5).

Gong C, Maquat LE. 2011. Lncrnas transactivate stau1-mediated mrna decay by duplexing with 3' utrs via alu elements. Nature. 470(7333):284-288.

Gulino-Debrac D. 2013. Mechanotransduction at the basis of endothelial barrier function. Tissue Barriers. 1(2):e24180.

Ha M, Kim VN. 2014. Regulation of microrna biogenesis. Nat Rev Mol Cell Biol. 15(8):509-524.

Hajishengallis G. 2014. Immunomicrobial pathogenesis of periodontitis: Keystones, pathobionts, and host response. Trends in immunology. 35(1):3-11.

Heiss C, Rodriguez-Mateos A, Kelm M. 2015. Central role of enos in the maintenance of endothelial homeostasis. Antioxid Redox Signal. 22(14):1230-1242.

International Human Genome Sequencing C. 2004. Finishing the euchromatic sequence of the human genome. Nature. 431(7011):931-945.

Jayaram N, Usvyat D, AC RM. 2016. Evaluating tools for transcription factor binding site prediction. BMC Bioinformatics. 17(1):547.

Jeong HW, Hernandez-Rodriguez B, Kim J, Kim KP, Enriquez-Gasca R, Yoon J, Adams S, Scholer HR, Vaquerizas JM, Adams RH. 2017. Transcriptional regulation of endothelial cell behavior during sprouting angiogenesis. Nat Commun. 8(1):726.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-rna-guided DNA endonuclease in adaptive bacterial immunity. Science. 337(6096):816-821.

Jones C, Virji M, Crocker PR. 2003. Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. Mol Microbiol. 49(5):1213-1225.

Kalna V, Yang Y, Peghaire CR, Frudd K, Hannah R, Shah AV, Osuna Almagro L, Boyle JJ, Gottgens B, Ferrer J et al. 2019. The transcription factor erg regulates super-enhancers associated with an endothelial-specific gene expression program. Circulation research. 124(9):1337-1349.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL et al. 2007. Rna maps reveal new rna classes and a possible function for pervasive transcription. Science. 316(5830):1484-1488.

Karki R, Pandya D, Elston RC, Ferlini C. 2015. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics. 8:37.

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE et al. 2010. Variation in transcription factor binding among humans. Science. 328(5975):232-235.

Kassebaum NJ, Bernabe E, Dahiya M, Bhandari B, Murray CJ, Marcenes W. 2014. Global burden of severe periodontitis in 1990-2010: A systematic review and meta-regression. J Dent Res. 93(11):1045-1053.

Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, Wendschlag A, Giese N, Tjaden C, Ott K et al. 2011. Toward the blood-borne mirnome of human diseases. Nat Methods. 8(10):841-843.

Keller A, Leidinger P, Borries A, Wendschlag A, Wucherpfennig F, Scheffler M, Huwer H, Lenhof HP, Meese E. 2009a. Mirnas in lung cancer - studying complex fingerprints in patient's blood cells by microarray experiments. BMC Cancer. 9:353.

Keller A, Leidinger P, Lange J, Borries A, Schroers H, Scheffler M, Lenhof HP, Ruprecht K, Meese E. 2009b. Multiple sclerosis: Microrna expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. PLoS One. 4(10):e7440.

Kelly LM, Englmeier U, Lafon I, Sieweke MH, Graf T. 2000. Mafb is an inducer of monocytic differentiation. EMBO J. 19(9):1987-1997.

Khandelwal N, Dey SK, Chakravarty S, Kumar A. 2019. Mir-30 family mirnas mediate the effect of chronic social defeat stress on hippocampal neurogenesis in mouse depression model. Front Mol Neurosci. 12:188.

Khatua B, Roy S, Mandal C. 2013. Sialic acids siglec interaction: A unique strategy to circumvent innate immune response by pathogens. Indian J Med Res. 138(5):648-662.

Kim K, Kim JH, Lee J, Jin HM, Kook H, Kim KK, Lee SY, Kim N. 2007. Mafb negatively regulates rankl-mediated osteoclast differentiation. Blood. 109(8):3253-3259.

Kim S, Kim D, Cho SW, Kim J, Kim JS. 2014. Highly efficient rna-guided genome editing in human cells via delivery of purified cas9 ribonucleoproteins. Genome Res. 24(6):1012-1019.

Kinane DF. 2001. Causation and pathogenesis of periodontal disease. Periodontol 2000. 25:8-20.

Kinane DF, Stathopoulou PG, Papapanou PN. 2017. Periodontal diseases. Nat Rev Dis Primers. 3:17039.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H et al. 2015. Genome-scale transcriptional activation by an engineered crispr-cas9 complex. Nature. 517(7536):583-588.

Kota RS, Ramana CV, Tenorio FA, Enelow RI, Rutledge JC. 2005. Differential effects of lipoprotein lipase on tumor necrosis factor-alpha and interferon-gamma-mediated gene expression in human endothelial cells. J Biol Chem. 280(35):31076-31084.

Kreimer A, Zeng H, Edwards MD, Guo Y, Tian K, Shin S, Welch R, Wainberg M, Mohan R, Sinnott-Armstrong NA et al. 2017. Predicting gene expression in massively parallel reporter assays: A comparative study. Hum Mutat.

Krol J, Loedige I, Filipowicz W. 2010. The widespread regulation of microrna biogenesis, function and decay. Nat Rev Genet. 11(9):597-610.

Kruithof EK. 2008. Regulation of plasminogen activator inhibitor type 1 gene expression by inflammatory mediators and statins. Thromb Haemost. 100(6):969-975.

Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. 2014. High-throughput functional testing of encode segmentation predictions. Genome Res. 24(10):1595-1602.

Lamont RJ, Koo H, Hajishengallis G. 2018. The oral microbiota: Dynamic communities and host interactions. Nat Rev Microbiol. 16(12):745-759.

Lang NP, Schatzle MA, Loe H. 2009. Gingivitis as a risk factor in periodontal disease. J Clin Periodontol. 36 Suppl 10:3-8.

Lawrenson K, Li Q, Kar S, Seo JH, Tyrer J, Spindler TJ, Lee J, Chen Y, Karst A, Drapkin R et al. 2015. Cis-eqtl analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. Nat Commun. 6:8234.

Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, Hinrichs AS, Lee BT, Nassar LR, Powell CC et al. 2020. Ucsc genome browser enters 20th year. Nucleic Acids Res. 48(D1):D756-D761.

Lee YH, Na HS, Jeong SY, Jeong SH, Park HR, Chung J. 2011. Comparison of inflammatory microrna expression in healthy and periodontitis tissues. Biocell. 35(2):43-49.
Li J, Zhang X, Tang J, Gong C. 2020a. Microrna-374b-5p functions as a tumor suppressor in non-small cell lung cancer by targeting foxp1 and predicts prognosis of cancer patients. Onco Targets Ther. 13:4229-4237.
Li K, Liu Y, Cao H, Zhang Y, Gu Z, Liu X, Yu A, Kaphle P, Dickerson KE, Ni M et al. 2020b. Interrogation of enhancer function by enhancer-targeting crispr epigenetic editing. Nat Commun. 11(1):485.
Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some micrornas downregulate large numbers of target mrnas. Nature. 433(7027):769-773.
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.
Liu Y, Chen GY, Zheng P. 2009. Cd24-siglec g/10 discriminates danger- from pathogen-associated molecular patterns. Trends in immunology. 30(12):557-561.
Lu TX, Rothenberg ME. 2018. Microrna. J Allergy Clin Immunol. 141(4):1202-1207.
Machiela MJ, Chanock SJ. 2015. Ldlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 31(21):3555-3557.
Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. Crispr rna-guided activation of endogenous human genes. Nat Methods. 10(10):977-979.
Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al. 2011. Evolution and classification of the crispr-cas systems. Nat Rev Microbiol. 9(6):467-477.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. Nature. 461(7265):747-753.
Marand AP, Zhang T, Zhu B, Jiang J. 2017. Towards genome-wide prediction and characterization of enhancers in plants. Biochim Biophys Acta Gene Regul Mech. 1860(1):131-139.
Marazita ML, Burmeister JA, Gunsolley JC, Koertge TE, Lake K, Schenkein HA. 1994. Evidence for autosomal dominant inheritance and race-specific heterogeneity in early-onset periodontitis. J Periodontol. 65(6):623-630.
Marcenes W, Kassebaum NJ, Bernabe E, Flaxman A, Naghavi M, Lopez A, Murray CJ. 2013. Global burden of oral conditions in 1990-2010: A systematic analysis. J Dent Res. 92(7):592-597.
Matthews JC, Hori K, Cormier MJ. 1977. Purification and properties of renilla reniformis luciferase. Biochemistry. 16(1):85-91.
Mattick JS, Makunin IV. 2005. Small regulatory rnas in mammals. Hum Mol Genet. 14 Spec No 1:R121-132.
Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science. 337(6099):1190-1195.
Meuleman W, Muratov A, Rynes E, Halow J, Lee K, Bates D, Diegel M, Dunn D, Neri F, Teodosiadis A et al. 2020. Index and biological spectrum of human dnase i hypersensitive sites. Nature. 584(7820):244-251.
Michalowicz BS. 1994. Genetic and heritable risk factors in periodontal disease. J Periodontol. 65(5 Suppl):479-488.

Michalowicz BS, Aeppli DP, Kuba RK, Bereuter JE, Conry JP, Segal NL, Bouchard TJ, Jr., Pihlstrom BL. 1991. A twin study of genetic variation in proportional radiographic alveolar bone height. J Dent Res. 70(11):1431-1435.

Michalowicz BS, Diehl SR, Gunsolley JC, Sparks BS, Brooks CN, Koertge TE, Califano JV, Burmeister JA, Schenkein HA. 2000. Evidence of a substantial genetic basis for risk of adult periodontitis. J Periodontol. 71(11):1699-1707.

Mignone F, Gissi C, Liuni S, Pesole G. 2002. Untranslated regions of mrnas. Genome Biol. 3(3):REVIEWS0004.

Mijailovic I, Nikolic N, Djinic A, Carkic J, Milinkovic I, Peric M, Jankovic S, Milasin J, Aleksic Z. 2020. The down-regulation of notch 1 signaling contributes to the severity of bone loss in aggressive periodontitis. J Periodontol. 91(4):554-561.

Mueller R, Bajric D, Keceli HG, Keller A, Dommisch H, Elsharawy A, Schaefer AS. 2021. Hsa-miR-374b-5p regulates expression of the gene u2af homology motif (uhm) kinase 1. J Periodontal Res.

Munz M, Richter GM, Loos BG, Jepsen S, Divaris K, Offenbacher S, Teumer A, Holtfreter B, Kocher T, Bruckmann C et al. 2019. Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. Eur J Hum Genet. 27(1):102-113.

Munz M, Richter GM, Loos BG, Jepsen S, Divaris K, Offenbacher S, Teumer A, Holtfreter B, Kocher T, Bruckmann C et al. 2018. Genome-wide association meta-analysis of coronary artery disease and periodontitis reveals a novel shared risk locus. Sci Rep. 8(1):13678.

Munz M, Willenborg C, Richter GM, Jockel-Schneider Y, Graetz C, Staufenbiel I, Wellmann J, Berger K, Krone B, Hoffmann P et al. 2017. A genome-wide association study identifies nucleotide variants at siglec5 and defa1a3 as risk loci for periodontitis. Hum Mol Genet.

Munz M, Wohlers I, Simon E, Reinberger T, Busch H, Schaefer AS, Erdmann J. 2020. Qtlizer: Comprehensive qtl annotation of gwas results. Sci Rep. 10(1):20417.

Nair AK, Baier LJ. 2018. Using luciferase reporter assays to identify functional variants at disease-associated loci. Methods Mol Biol. 1706:303-319.

Narlikar L, Ovcharenko I. 2009. Identifying regulatory elements in eukaryotic genomes. Brief Funct Genomic Proteomic. 8(4):215-230.

Nesse W, Abbas F, van der Ploeg I, Spijkervet FK, Dijkstra PU, Vissink A. 2008. Periodontal inflamed surface area: Quantifying inflammatory burden. J Clin Periodontol. 35(8):668-673.

Nica AC, Dermitzakis ET. 2013. Expression quantitative trait loci: Present and future. Philos Trans R Soc Lond B Biol Sci. 368(1620):20120362.

Nilsen TW. 2007. Mechanisms of microrna-mediated gene regulation in animal cells. Trends Genet. 23(5):243-249.

Nisha KJ, Janam P, Harshakumar K. 2019. Identification of a novel salivary biomarker mir-143-3p for periodontal diagnosis: A proof of concept study. J Periodontol. 90(10):1149-1159.

Nordstrom T, Movert E, Olin AI, Ali SR, Nizet V, Varki A, Areschoug T. 2011. Human siglec-5 inhibitory receptor and immunoglobulin a (iga) have separate binding sites in streptococcal beta protein. J Biol Chem. 286(39):33981-33991.

Page GP, George V, Go RC, Page PZ, Allison DB. 2003. "Are we there yet?": Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. Am J Hum Genet. 73(4):711-719.

Palazzo AF, Gregory TR. 2014. The case for junk DNA. PLoS Genet. 10(5):e1004351.

Paul P, Chakraborty A, Sarkar D, Langthasa M, Rahman M, Bari M, Singha RS, Malakar AK, Chakraborty S. 2018. Interplay between mirnas and human diseases. J Cell Physiol. 233(3):2007-2018.

Pegon JN, Kurdi M, Casari C, Odouard S, Denis CV, Christophe OD, Lenting PJ. 2012. Factor viii and von willebrand factor are ligands for the carbohydrate-receptor siglec-5. Haematologica. 97(12):1855-1863.

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

Pepin M, Mezouar S, Pegon J, Muczynski V, Adam F, Bianchini EP, Bazaa A, Proulle V, Rupin A, Paysant J et al. 2016. Soluble siglec-5 associates to psgl-1 and displays anti-inflammatory activity. Sci Rep. 6:37953.

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW et al. 2013. Rna-guided gene activation by crispr-cas9-based transcription factors. Nat Methods. 10(10):973-976.

Pillai S, Netravali IA, Cariappa A, Mattoo H. 2012. Siglecs and immune regulation. Annu Rev Immunol. 30:357-392.

Quandt K, Frech K, Karas H, Wingender E, Werner T. 1995. Matind and matinspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23(23):4878-4884.

R P, Sreedhara A, P I, Sarkar I, Kumar CS. 2014. Vascular endothelial growth factor levels in gingival crevicular fluid before and after periodontal therapy. J Clin Diagn Res. 8(11):ZC75-79.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding rnas. Annu Rev Biochem. 81:145-166.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. Nature. 405(6788):847-856.

Rockman MV, Kruglyak L. 2006. Genetics of global gene expression. Nat Rev Genet. 7(11):862-872.

Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. 2010. Origins of specificity in protein-DNA recognition. Annu Rev Biochem. 79:233-269.

Roshna T, Nandakumar K. 2012. Generalized aggressive periodontitis and its treatment options: Case reports and review of the literature. Case Rep Med. 2012:535321.

Rozen S, Skaletsky H. 2000. Primer3 on the www for general users and for biologist programmers. Methods Mol Biol. 132:365-386.

Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, Eliceiri KW. 2017. Imagej2: Imagej for the next generation of scientific image data. BMC Bioinformatics. 18(1):529.

Saito A, Horie M, Ejiri K, Aoki A, Katagiri S, Maekawa S, Suzuki S, Kong S, Yamauchi T, Yamaguchi Y et al. 2017. Microrna profiling in gingival crevicular fluid of periodontitis-a pilot study. FEBS Open Bio. 7(7):981-994.

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. 2004. Jaspar: An open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 32(Database issue):D91-94.

Sander JD, Joung JK. 2014. Crispr-cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 32(4):347-355.

Santos-Gallego CG. 2016. Mafb and the role of macrophage apoptosis in atherosclerosis: A time to kill, a time to heal. Atherosclerosis. 252:194-196.

Sati S, Cavalli G. 2017. Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma. 126(1):33-44.

Schaefer AS. 2018. Genetics of periodontitis: Discovery, biology, and clinical impact. Periodontol 2000. 78(1):162-173.

Schaefer AS, Richter GM, Dommisch H, Reinartz M, Nothnagel M, Noack B, Laine ML, Folwaczny M, Groessner-Schreiber B, Loos BG et al. 2011. Cdkn2bas is associated with periodontitis in different european populations and is activated by bacterial infection. J Med Genet. 48(1):38-47.

Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. 2012. Linking disease associations with regulatory information in the human genome. Genome Res. 22(9):1748-1759.

Shungin D, Haworth S, Divaris K, Agler CS, Kamatani Y, Keun Lee M, Grinde K, Hindy G, Alaraudanjoki V, Pesonen P et al. 2019. Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. Nat Commun. 10(1):2773.

Simone TM, Higgins CE, Czekay RP, Law BK, Higgins SP, Archambeault J, Kutz SM, Higgins PJ. 2014. Serpine1: A molecular switch in the proliferation-migration dichotomy in wound-"activated" keratinocytes. Adv Wound Care (New Rochelle). 3(3):281-290.

Slatkin M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 9(6):477-485.

Smith PC, Caceres M, Martinez C, Oyarzun A, Martinez J. 2015. Gingival wound healing: An essential response disturbed by aging? J Dent Res. 94(3):395-402.

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal microrna confer robustness to gene expression and have a significant impact on 3'utr evolution. Cell. 123(6):1133-1146.

Stoecklin-Wasmer C, Guarnieri P, Celenti R, Demmer RT, Kebschull M, Papapanou PN. 2012. Microrna and their target genes in gingival tissues. J Dent Res. 91(10):934-940.

Sybirna A, Tang WWC, Pierson Smela M, Dietmann S, Gruhn WH, Brosh R, Surani MA. 2020. A critical role of prdm14 in human primordial germ cell fate revealed by inducible degrons. Nat Commun. 11(1):1282.

Tonetti MS, Greenwell H, Kornman KS. 2018. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. J Periodontol. 89 Suppl 1:S159-S172.

Tonetti MS, Jepsen S, Jin L, Otomo-Corgel J. 2017. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. J Clin Periodontol. 44(5):456-462.

Tonnesen MG, Feng X, Clark RA. 2000. Angiogenesis in wound healing. J Investig Dermatol Symp Proc. 5(1):40-46.

Tracey LJ, Justice MJ. 2019. Off to a bad start: Cancer initiation by pluripotency regulator prdm14. Trends Genet. 35(7):489-500.

Trindade F, Oppenheim FG, Helmerhorst EJ, Amado F, Gomes PS, Vitorino R. 2014. Uncovering the molecular networks in periodontitis. Proteomics Clin Appl. 8(9-10):748-761.

Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR. 2003. Transcription of antisense rna leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet. 34(2):157-165.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A et al. 2015. Proteomics. Tissue-based map of the human proteome. Science. 347(6220):1260419.

Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, Franke A, Haesler R et al. 2005. Sarcoidosis is associated with a truncating splice site mutation in btnl2. Nat Genet. 37(4):357-364.

Verma D, Garg PK, Dubey AK. 2018. Insights into the human oral microbiome. Arch Microbiol. 200(4):525-540.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of gwas discovery. Am J Hum Genet. 90(1):7-24.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 2017. 10 years of gwas discovery: Biology, function, and translation. Am J Hum Genet. 101(1):5-22.
von Gunten S, Bochner BS. 2008. Basic and clinical immunology of siglecs. Ann N Y Acad Sci. 1143:61-82.
Wall JD, Pritchard JK. 2003. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 4(8):587-597.
Wapinski O, Chang HY. 2011. Long noncoding rnas and human disease. Trends Cell Biol. 21(6):354-361.
Wei T, Weiler SME, Toth M, Sticht C, Lutz T, Thomann S, De La Torre C, Straub B, Merker S, Ruppert T et al. 2019. Yap-dependent induction of uhmk1 supports nuclear enrichment of the oncogene mybl2 and proliferation in liver cancer cells. Oncogene. 38(27):5541-5550.
Weirauch MT, Cote A, Norel R, Annala M, Zhao Y, Riley TR, Saez-Rodriguez J, Cokelaer T, Vedenko A, Talukder S et al. 2013. Evaluation of methods for modeling transcription factor sequence specificity. Nat Biotechnol. 31(2):126-134.
Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L et al. 2014. The nhgri gwas catalog, a curated resource of snp-trait associations. Nucleic Acids Res. 42(Database issue):D1001-1006.
Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE et al. 2013. Systematic identification of trans eqtls as putative drivers of known disease associations. Nat Genet. 45(10):1238-1243.
Wiehle L, Breiling A. 2016. Chromatin immunoprecipitation. Methods Mol Biol. 1480:7-21.
Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding rnas: Functional surprises from the rna world. Genes Dev. 23(13):1494-1504.
Wingender E. 2008. The transfac project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform. 9(4):326-332.
Wood KV, de Wet JR, Dewji N, DeLuca M. 1984. Synthesis of active firefly luciferase by in vitro translation of rna obtained from adult lanterns. Biochem Biophys Res Commun. 124(2):592-596.
Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. 2018. Common disease is more complex than implied by the core gene omnigenic model. Cell. 173(7):1573-1580.
Xie YF, Shu R, Jiang SY, Liu DL, Zhang XL. 2011. Comparison of microrna profiles of human periodontal diseased and healthy gingival tissues. Int J Oral Sci. 3(3):125-134.
Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R et al. 2020. Ensembl 2020. Nucleic Acids Res. 48(D1):D682-D688.
Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H et al. 2010. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One. 5(5):e10693.
Zeng B, Lloyd-Jones LR, Holloway A, Marigorta UM, Metspalu A, Montgomery GW, Esko T, Brigham KL, Quyyumi AA, Idaghdour Y et al. 2017. Constraints on eqtl fine mapping in the presence of multisite local regulation of gene expression. G3 (Bethesda). 7(8):2533-2544.
Zhang F, Wen Y, Guo X. 2014a. Crispr/cas9 for genome editing: Progress, implications and challenges. Hum Mol Genet. 23(R1):R40-46.
Zhang X, Li H, Burnett JC, Rossi JJ. 2014b. The role of antisense long noncoding rna in small rna-triggered gene activation. RNA. 20(12):1916-1928.
Zhu JJ, Liu YF, Zhang YP, Zhao CR, Yao WJ, Li YS, Wang KC, Huang TS, Pang W, Wang XF et al. 2017. Vamp3 and snap23 mediate the disturbed flow-induced endothelial microrna secretion and smooth muscle hyperplasia. Proc Natl Acad Sci U S A. 114(31):8271-8276.

Zhuravleva MA, Trandem K, Sun PD. 2008. Structural implications of siglec-5-mediated sialoglycan recognition. J Mol Biol. 375(2):437-447.

Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A. 109(4):1193-1198.

8 Appendix

8.1 eQTL effects of rs4284742, rs11084095, rs1122900 and rs6887423

Table 23: eQTL effects of rs428472 annotated by the software tool QTLizer.

Index variant	LD-SNP (r^2>0.8)	Affected Gene	Tissue	p-value	beta	Effect Allele	Non-Effect allele	Source
rs4284742	-	SIGLEC5	Peripheral blood	7.7e-14	-	A	G	Blood eQTL Browser
rs4284742	-	SHANK1	Spleen	6.9e-7	-0.39	G	A	GTEx v8
rs4284742	-	RPL9P33	Whole blood	0.0000022	-0.18	G	A	GTEx v8
rs4284742	-	SIGLEC14	Pituitary	0.00002	-0.3	G	A	GTEx v8
rs4284742	-	SIGLEC5	Whole blood	0.000033	-0.11	G	A	GTEx v8
rs4284742	-	NDUFA3	Brain - Parietal lobe	0.0043	-	-	-	ScanDB
rs4284742	-	ZNF813	Brain - Cerebellum	0.0045	-	-	-	ScanDB
rs4284742	-	ZNF787	Brain - Parietal lobe	0.0064	-	-	-	ScanDB
rs4284742	-	SYT5	Brain - Parietal lobe	0.0087	-	-	-	ScanDB
rs4284742	-	CD33	Lymphoblastoid cell lines	0.12	-	-	-	seeQTL

Table 24: eQTL effects of rs11084095 and LD-SNPs with $p < 1 \times 10^{-6}$ annotated by the software tool QTLizer.

Index variant	LD-SNP (r^2>0.8)	Affected Gene	Tissue	p-value	beta	Effect Allele	Non-Effect allele	Source
rs11084095	-	SIGLEC5	Cells – Monocytes	6.4e-23	-	-	-	Zeller et al. (PlosONE 2010)
rs11084095	rs4801882	SIGLEC5	Cells – Monocytes	2.1e-16	-	-	-	Zeller et al. (PlosONE 2010)
rs11084095	-	SIGLEC5	Adipose - Subcutaneous	4.0e-16	0.4	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Adipose - Subcutaneous	4.0e-16	0.4	T	A	GTEx v8
rs11084095	-	SIGLEC5	Nerve - Tibial	2.2e-14	0.4	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Nerve - Tibial	2.2e-14	0.4	T	A	GTEx v8
rs11084095	-	SIGLEC5	Artery - Tibial	4.6e-14	0.38	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Artery - Tibial	6.2e-14	0.38	T	A	GTEx v8
rs11084095	rs4801882	SIGLEC5	Nerve - Tibial	1.7e-12	0.37	A	G	GTEx v8
rs11084095	-	SIGLEC5	Colon – Sigmoid	1.1e-11	0.48	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Colon – Sigmoid	1.1e-11	0.48	T	A	GTEx v8

rs11084095	rs4801882	SIGLEC5	Adipose - Subcutaneous	4.0e-11	0.32	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Cells - Macrophages	1.2e-10	-0.28	A	T	The Cardiogenics Project
rs11084095	-	SIGLEC5	Cells - Macrophages	1.4e-10	-0.28	G	A	The Cardiogenics Project
rs11084095	-	SIGLEC5	Esophagus - Muscularis	3.0e-10	0.37	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Esophagus - Muscularis	3.0e-10	0.37	T	A	GTEx v8
rs11084095	-	SIGLEC5	Adipose – Visceral (Omentum)	1.7e-9	0.23	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC5	Artery - Tibial	2.6e-9	0.29	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Adipose – Visceral (Omentum)	3.1e-9	0.23	T	A	GTEx v8
rs11084095	-	SIGLEC5	Artery - Aorta	1.2e-8	0.29	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Artery - Aorta	1.8e-8	0.29	T	A	GTEx v8
rs11084095	-	SIGLEC5	Colon - Transverse	3.5e-8	0.31	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Colon - Transverse	3.5e-8	0.31	T	A	GTEx v8
rs11084095	-	SIGLEC5	Stomach	3.8e-8	0.33	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Stomach	3.8e-8	0.33	T	A	GTEx v8
rs11084095	rs4801882	LLNLR-470E3.1	Whole blood	5.2e-8	-0.24	A	G	GTEx v8
rs11084095	-	LLNLR-470E3.1	Whole blood	8.0e-8	-0.25	A	G	GTEx v8
rs11084095	rs12461706	LLNLR-470E3.1	Whole blood	8.0e-8	-0.25	T	A	GTEx v8
rs11084095	rs4801882	SIGLEC5	Colon – Sigmoid	1.0e-7	0.37	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC5	Esophagus - Muscularis	1.1e-7	0.31	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Breast - Mammary tissue	1.2e-7	0.34	T	A	GTEx v8
rs11084095	-	SIGLEC5	Breast – Mammary tissue	1.3e-7	0.34	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC5	Breast - Mammary tissue	1.3e-7	0.33	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC5	Artery - Aorta	1.8e-7	0.25	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Skin - Sun exposed (Lower leg)	2.4e-7	0.25	T	A	GTEx v8
rs11084095	-	SIGLEC5	Skin - Sun exposed (Lower leg)	3.0e-7	0.25	A	G	GTEx v8
rs11084095	-	SIGLEC5	Adrenal gland	3.3e-7	0.45	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Adrenal gland	3.3e-7	0.45	T	A	GTEx v8
rs11084095	rs4801882	SIGLEC5	Cells - Macrophages	4.3e-7	-0.21	G	A	The Cardiogenics Project

rs11084095	rs4801882	SIGLEC5	Skin - Sun exposed (Lower leg)	5.7e-7	0.24	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC5	Adipose – Visceral (Omentum)	8.3e-7	0.18	A	G	GTEx v8
rs11084095	rs4801882	SIGLEC14	Whole blood	9.2e-7	-0.21	A	G	GTEx v8
rs11084095	-	SIGLEC5	Heart - Atrial appendage	9.7e-7	0.26	A	G	GTEx v8
rs11084095	rs12461706	SIGLEC5	Heart - Atrial appendage	9.7e-7	0.26	T	A	GTEx v8

Table 25: eQTL effects of rs1122900 and LD-SNPs with $p < 1 \times 10^{-5}$ annotated by the software tool QTLizer.

Index variant	LD-SNP	Affected Gene	Tissue	p-value	beta	Effect Allele	Non-Effect Allele	Source
rs1122900	-	CTD-2353F22.1	Whole blood	6.10E-15	0.42	C	A	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Whole blood	2.40E-14	0.4	T	C	GTEx v8
rs1122900	rs56038114	CTD-2353F22.1	Whole blood	3.50E-13	0.39	C	T	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Whole blood	4.50E-12	0.36	C	T	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Nerve - Tibial	3.50E-10	0.3	T	C	GTEx v8
rs1122900	-	CTD-2353F22.1	Nerve - Tibial	6.60E-10	0.31	C	A	GTEx v8
rs1122900	-	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	4.30E-09	0.67	C	A	GTEx v8
rs1122900	rs56038114	CTD-2353F22.1	Nerve - Tibial	1.80E-08	0.28	C	T	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Lung	6.10E-08	0.28	C	T	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	8.90E-08	0.6	T	C	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Nerve - Tibial	1.60E-07	0.25	C	T	GTEx v8
rs1122900	-	CTD-2353F22.1	Lung	2.30E-07	0.29	C	A	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Artery - Aorta	2.30E-07	0.3	T	C	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Lung	3.80E-07	0.27	T	C	GTEx v8
rs1122900	-	CTD-2353F22.2	Cells - EBV-transformed lymphocytes	3.90E-07	-0.54	C	A	GTEx v8

rs1122900	rs56038114	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	4.10E-07	0.57	C	T	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Artery - Aorta	4.30E-07	0.29	C	T	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Artery - Tibial	4.80E-07	0.25	T	C	GTEx v8
rs1122900	-	CTD-2353F22.1	Adipose - Subcutaneous	9.00E-07	0.22	C	A	GTEx v8
rs1122900	-	CTD-2353F22.1	Adipose - Visceral	9.10E-07	0.23	C	A	GTEx v8
rs1122900	-	CTD-2353F22.1	Artery - Tibial	9.80E-07	0.25	C	A	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Artery - Tibial	1.50E-06	0.24	C	T	GTEx v8
rs1122900	-	CTD-2353F22.1	Artery - Aorta	1.70E-06	0.29	C	A	GTEx v8
rs1122900	rs56039629	CTD-2353F22.2	Cells - EBV-transformed lymphocytes	1.80E-06	-0.5	T	C	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Adipose - Visceral	2.60E-06	0.21	T	C	GTEx v8
rs1122900	rs56038114	CTD-2353F22.1	Lung	3.00E-06	0.25	C	T	GTEx v8
rs1122900	rs56038114	CTD-2353F22.1	Artery - Aorta	3.40E-06	0.27	C	T	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Skin - Sun exposed (Lower leg)	4.30E-06	0.23	C	T	GTEx v8
rs1122900	-	CTD-2353F22.1	Skin - Sun exposed (Lower leg)	4.40E-06	0.25	C	A	GTEx v8
rs1122900	rs6862950	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	4.40E-06	0.52	C	T	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Adipose - Subcutaneous	5.00E-06	0.2	T	C	GTEx v8
rs1122900	rs56038114	CTD-2353F22.1	Adipose - Visceral	7.40E-06	0.2	C	T	GTEx v8
rs1122900	rs56039629	CTD-2353F22.1	Skin - Sun exposed (Lower leg)	9.20E-06	0.23	T	C	GTEx v8

Table 26: eQTL effects of rs6887423 and LD-SNPs with $p < 1 \times 10^{-8}$ annotated by the software tool QTLizer.

Index variant	LD-SNP	Affected Gene	Tissue	p-value	beta	Effect Allele	Non-Effect Allele	Source
rs6887423	rs56066032	CTD-2353F22.1	Whole blood	2.50E-28	0.57	A	G	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Whole blood	8.50E-27	0.55	G	A	GTEx v8
rs6887423	-	CTD-2353F22.1	Whole blood	2.60E-26	0.55	C	T	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Whole blood	1.30E-24	0.53	G	C	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Lung	2.20E-21	0.49	A	G	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Spleen	9.40E-21	0.79	A	G	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Lung	5.80E-17	0.44	G	C	GTEx v8
rs6887423	-	CTD-2353F22.1	Spleen	1.30E-16	0.74	C	T	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Spleen	3.40E-16	0.73	G	A	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Adipose - Subcutaneous	4.90E-16	0.34	A	G	GTEx v8
rs6887423	-	CTD-2353F22.1	Lung	5.50E-16	0.42	C	T	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Nerve - Tibial	5.50E-16	0.39	A	G	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Nerve - Tibial	6.80E-16	0.39	G	C	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Lung	3.70E-15	0.41	G	A	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Spleen	1.60E-14	0.64	G	C	GTEx v8
rs6887423	-	CTD-2353F22.1	Nerve - Tibial	2.10E-14	0.37	C	T	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Artery - Tibial	2.90E-14	0.37	A	G	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Artery - Tibial	1.20E-13	0.37	G	C	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Nerve - Tibial	1.40E-13	0.36	G	A	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Artery - Aorta	2.90E-13	0.4	A	G	GTEx v8
rs6887423	-	CTD-2353F22.1	Adipose - Subcutaneous	3.80E-13	0.31	C	T	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Adipose - Subcutaneous	4.40E-13	0.31	G	A	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Adipose - Subcutaneous	2.20E-12	0.3	G	C	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Artery - Tibial	4.50E-12	0.34	G	A	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	5.10E-12	0.7	A	G	GTEx v8

rs6887423	-	CTD-2353F22.1	Artery - Tibial	6.40E-12	0.34	C	T	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Adipose - Visceral	4.00E-11	0.3	A	G	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Adipose - Visceral	7.00E-11	0.29	G	C	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Artery - Aorta	7.20E-11	0.37	G	C	GTEx v8
rs6887423	-	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	1.90E-10	0.66	C	T	GTEx v8
rs6887423	rs56162483	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	1.90E-10	0.67	G	C	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Colon - Transverse	2.40E-10	0.4	A	G	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Cells - EBV-transformed lymphocytes	3.10E-10	0.66	G	A	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Esophagus - Gastroesophageal junction	3.20E-10	0.43	A	G	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Brain - Spinal cord	4.20E-10	0.71	A	G	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Artery - Aorta	5.90E-10	0.35	G	A	GTEx v8
rs6887423	-	CTD-2353F22.1	Artery - Aorta	8.90E-10	0.35	C	T	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Adipose - Visceral	2.10E-09	0.26	G	A	GTEx v8
rs6887423	-	CTD-2353F22.1	Adipose - Visceral	2.20E-09	0.26	C	T	GTEx v8
rs6887423	rs17585785	CTD-2353F22.1	Colon - Transverse	2.80E-09	0.38	G	A	GTEx v8
rs6887423	-	CTD-2353F22.1	Colon - Transverse	3.80E-09	0.38	C	T	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Small intestine - Terminal ileum	5.40E-09	0.37	A	G	GTEx v8
rs6887423	rs56066032	CTD-2353F22.1	Breast - Mammary tissue	6.10E-09	0.3	A	G	GTEx v8

Table 27: *SIGLEC5* lead SNPs rs4284742, rs11084095 and rs12461706 and SNPs in LD ($r^2 > 0.8$) and predicted TFBS.

SNP ID	Location (GRCh37/hg19)	Distance to next SNP (kb)	predicted TFBS (PWM similarity score)	*r2* with rs12461706
rs4284742	chr19:52131733	NA	MAFB (82%)	0.19
rs11084095	chr19:52127030	4.7	ERG (93%)	1.00
rs4801882	chr19:52127053	0	none	0.81
rs12461706	chr19:52121235	5.8	none	1.00
rs11880807	chr19:52120522	0.7	none	0.81

rs34984145	chr19:52120410	0.1	BACH2 (83%)	0.81

Table 28: *CTD-2353F22.1* lead SNPs rs1122900 and rs6887423 and SNPs in LD ($r^2 > 0.8$) and predicted TFBS.

SNP ID	Location (GRCh37/hg19)	Distance to next SNP (kb)	predicted TFBS (PWM similarity score)	*r2* with rs1122900
rs1122900	chr5:36689181	NA	PRDM14 (89%)	1.00
rs56039629	chr5:36683903	5.3	none	0.89
rs56038114	chr5:36683801	0.1	GATA1 (96%)	0.96
rs17585785	chr5:36695331	0.8	none	0.67
rs56066032	chr5:36695601	0.2	none	0.53
rs6887423	chr5:36696501	0.9	AHR (100%)	0.67
rs6862950	chr5:36681820	0.1	none	0.85
rs56162483	chr5:36710830	14.2	none	0.63

8.2 Cloning of reporter gene assay plasmids

Using PCR, the genomic sequences at the SNPs of interest were amplified from genomic DNA and the PCR-products were loaded onto a 1% agarose gel. The PCR-products and the pGL4.24 plasmid were restricted with *XhoI* and *KpnI* and again loaded onto a 1% agarose gel. The restricted PCR-product of rs4284742 and the restricted plasmid pGL4.24 are shown in **Figure 42**. The restricted PCR-products of rs11084095, rs34984145, rs1122900, rs6887423 and rs56038114 are shown in **Figure 43**.

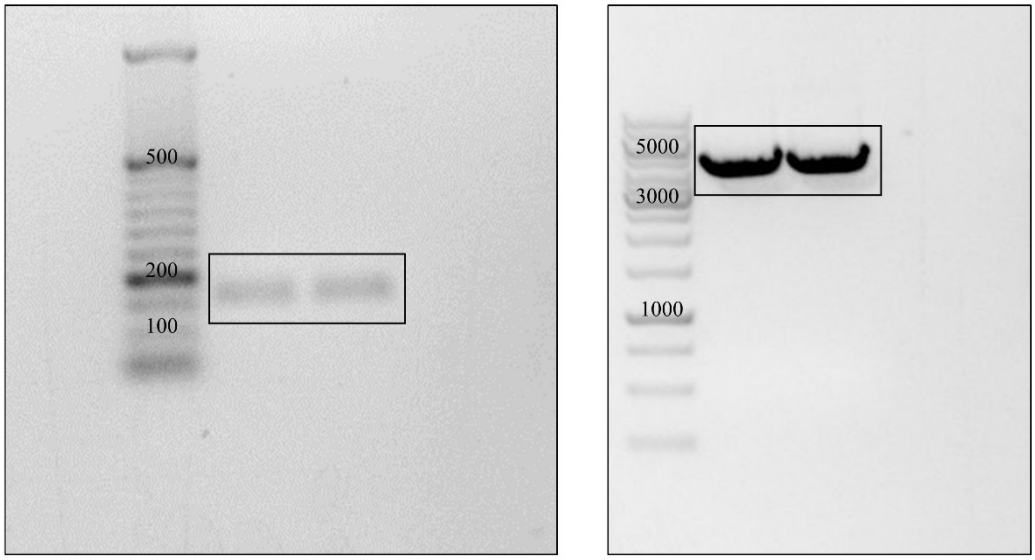


Figure 42: Picture of the 1% agarose gel with restricted PCR-product of rs4284742 (left) and the 2% agarose gel with restricted plasmid pGL4.24 (right). Electrophoresis was performed for 60 min at 100 V. The labeled fragments were isolated from the gel as described in 3.1.8.1.4.

The expected size for the PCR product of rs4284742 is about 150 bp, and the expected size for pGL4.24 is 4411 bp. The bands in the agarose gel establish that the required fragments were obtained. The labeled bands were excised and purified from the agarose gel, and then ligated. Successful ligation was tested by colony PCR and restriction digest and finally confirmed by Sanger sequencing. After successful cloning of the genomic sequence at rs4284742, the G-allele of rs4284742 was exchanged with the A-allele using Q5 site-directed mutagenesis kit as described in chapter 3.1.8.2 (**Figure 27**).

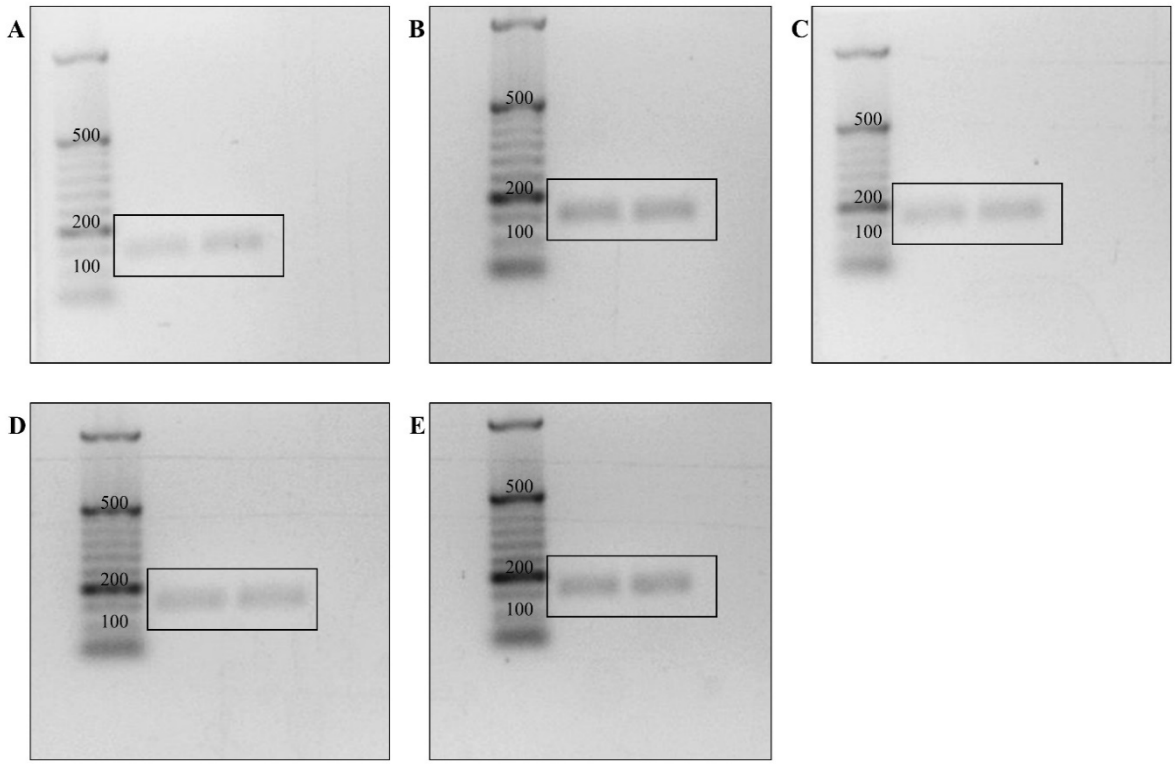


Figure 43: Pictures of 2% agarose gels with restricted PCR-products of the GWAS lead SNPs rs11084095, rs1122900 and rs6887423, and the two SNPs in strong LD rs34984145 and rs56038114. (A) rs11084095 **(B)** rs34984145 **(C)** rs1122900 **(D)** rs6887423 **(E)** rs56038114. Electrophoresis was performed for 60 min at 100 V. The labeled fragments were isolated from the gel as described in 3.1.8.1.4.

Figure 44: Sequence alignment of rs428472. The first line shows the reference sequence of rs4284742. The second line shows the amplified sequence of rs4284742 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs4284742. SNP rs4284742 is marked in green. Areas of matched sequences at same base pair positions are colored.

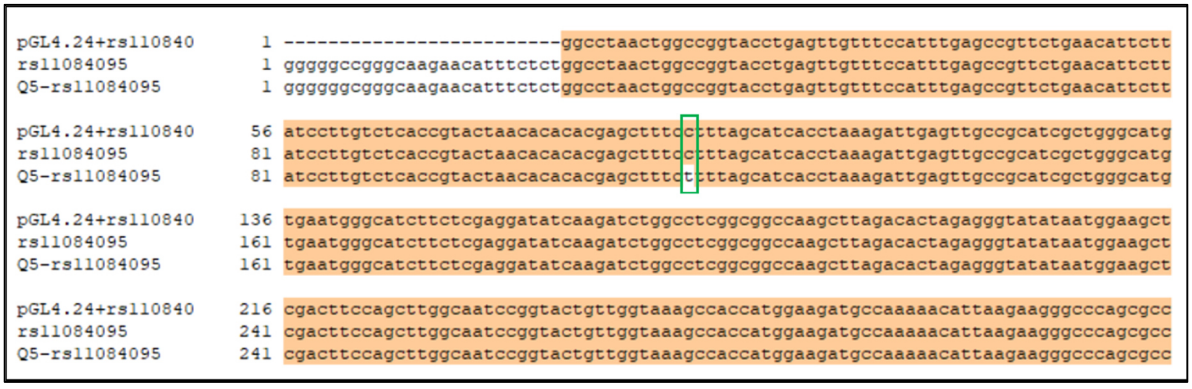


Figure 45: Sequence alignment of rs11084095. The first line shows the reference sequence of rs11084095. The second line shows the amplified sequence of rs11084095 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs11084095. SNP rs11084095 is marked in green. Areas of matched sequences at same base pair positions are colored.

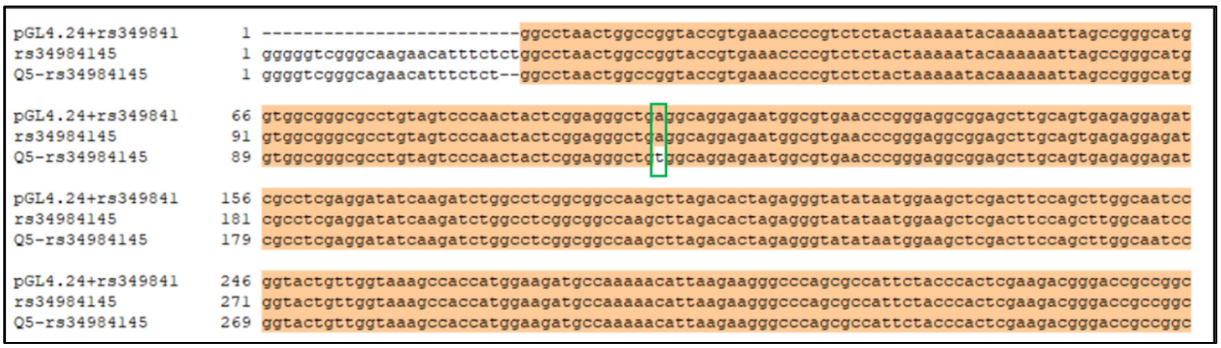


Figure 46: Sequence alignment of rs34984145. The first line shows the reference sequence of rs34984145. The second line shows the amplified sequence of rs34984145 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs34984145. SNP rs34984145 is marked in green. Areas of matched sequences at same base pair positions are colored.

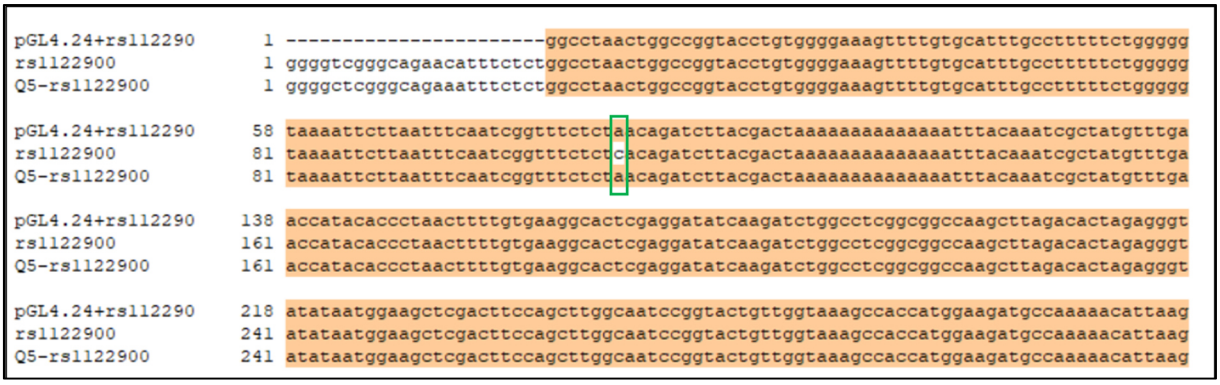


Figure 47: Sequence alignment of rs1122900. The first line shows the reference sequence of rs1122900. The second line shows the amplified sequence of rs1122900 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs1122900. SNP rs1122900 is marked in green. Areas of matched sequences at same base pair positions are colored.

Figure 48: Sequence alignment of rs6887423. The first line shows the reference sequence of rs6887423. The second line shows the amplified sequence of rs6887423 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs6887423. SNP rs6887423 is marked in green. Areas of matched sequences at same base pair positions are colored.

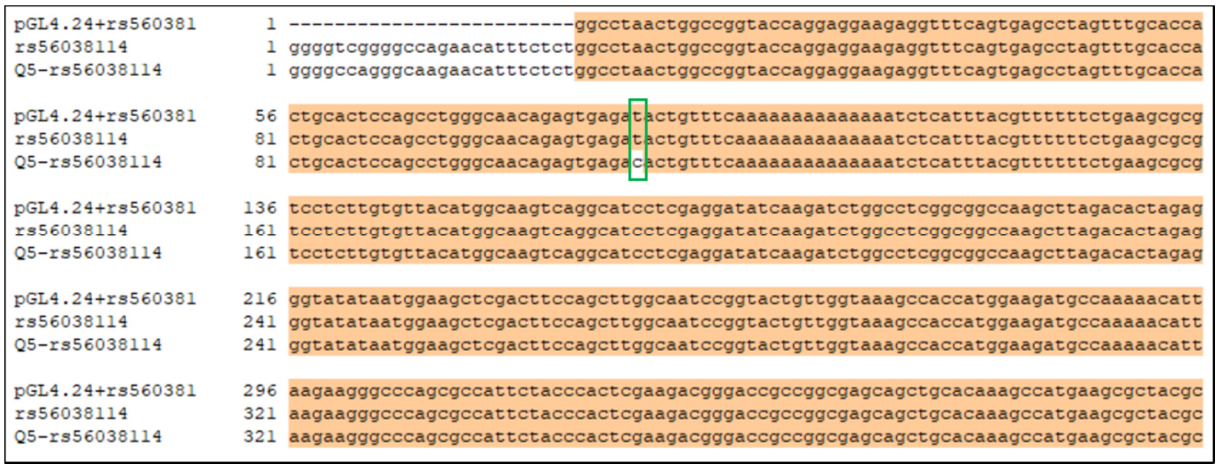


Figure 49: Sequence alignment of rs56038114. The first line shows the reference sequence of rs56038114. The second line shows the amplified sequence of rs56038114 from genomic DNA cloned in the pGL4.24 plasmid and the third line shows the sequence after Q5-cloning with the allele exchange of rs56038114. SNP rs56038114 is marked in green. Areas of matched sequences at same base pair positions are colored.

8.3 Cloning of gRNAs for VPR-activation system

For the CRISPRa with VPR-activation system, gRNAs were cloned into the gRNA cloning vector using Gibson assembly. Three different gRNAs were tested. PCR reactions with the primers as templates were performed as described in chapter 3.1.8.1.1 and PCR products were loaded onto a 1% agarose gel (**Figure 50**). The gRNA cloning vector was linearized with *AflII* and loaded onto a 1% gel (**Figure 50**). The expected size for the PCR products of the gRNAs of 100 bp, and the expected size for the linearized gRNA cloning vector of 3914 bp were obtained. The fragments were excised and purified from the agarose gel, and then ligated using Gibson assembly. Successful ligation was tested by restriction digest and finally confirmed by Sanger sequencing.

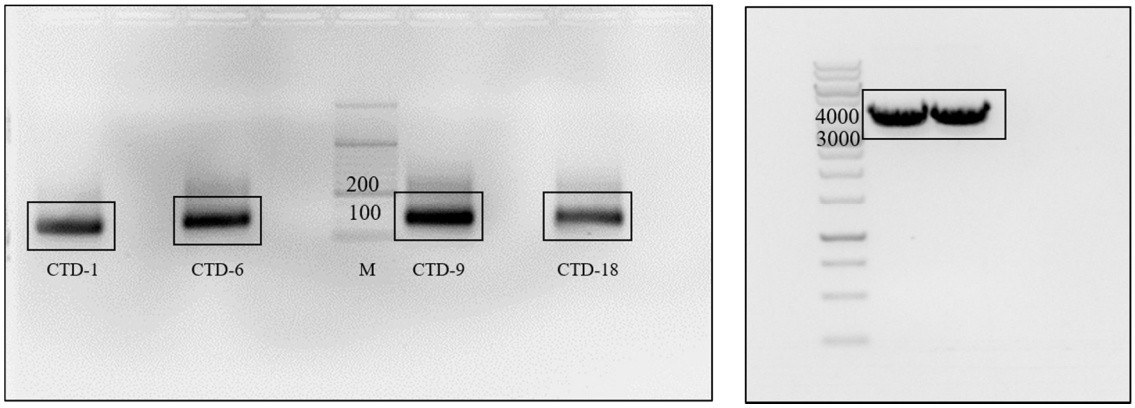


Figure 50: Picture of the 1% agarose gel with amplified PCR-products of gRNAs for upregulation of the promoter of *CTD-2353F22.1* (left) and restricted plasmid gRNA Cloning Vector (right). Electrophoresis was performed for 45 min at 100 V. The labeled fragments were isolated from the gel as described in chapter 3.1.8.1.4.

8.4 Cloning of gRNAs for SAM-activation system

For the CRISPRa with SAM-activation system, gRNAs were cloned into the sgRNA (MS2) cloning backbone. The same three different gRNAs that were used for the VPR-system were also cloned for the SAM system. Oligonucleotides were annealed as described in chapter 3.1.12.3.2 and the sgRNA (MS2) cloning backbone plasmid was linearized with *BbsI-HF* and loaded onto a 1% agarose gel (**Figure 51**).

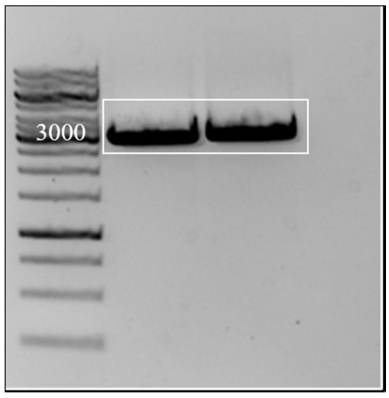


Figure 51: Picture of the 1% agarose gel with restricted plasmid gRNA Cloning Vector. Electrophoresis was performed for 100 min at 100 V. The labeled fragments were isolated from the gel as described in chapter 3.1.8.1.4.

The expected size for the linearized sgRNA (MS2) cloning backbone of around 3 kb was obtained. The fragment was excised and purified from the agarose gel and ligated with the annealed oligonucleotides. Successful ligation was confirmed by Sanger sequencing.

8.5 Positions of gRNAs for CRISPR-mediated activation

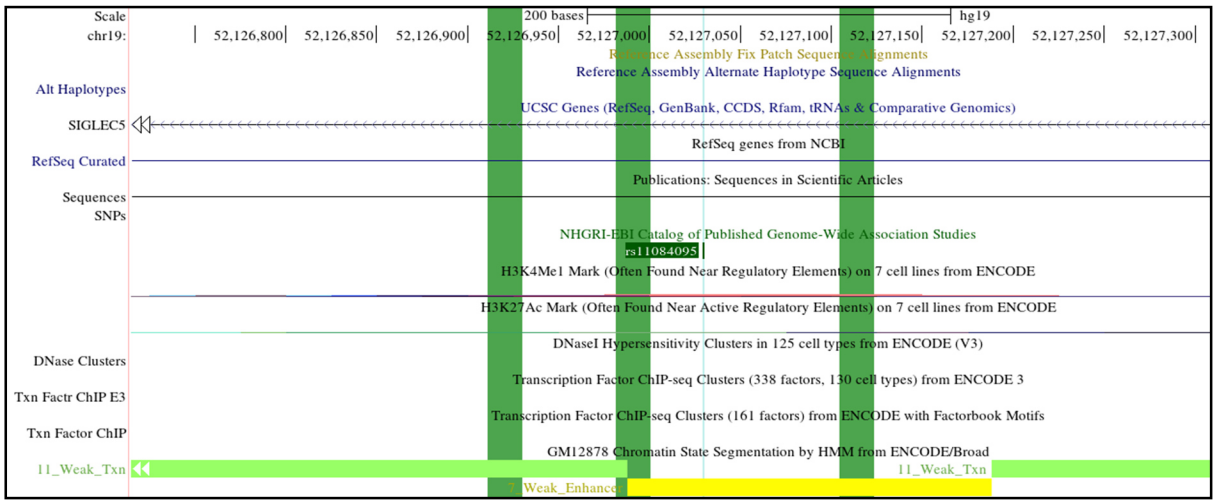


Figure 52: Position of gRNAs to test the genomic region at rs11084095 for its potential to activate *SIGLEC5* expression. CRISPRa of the genomic region at rs11084095 did not induce *SIGLEC5* expression compared to the scrambled gRNA. gRNA-1 (chr19:52126979-52127000), gRNA-2 (chr19:52127105-52127127) and gRNA-3 (chr19:52126912-52126930) have 50-120 bp distance from rs11084095.

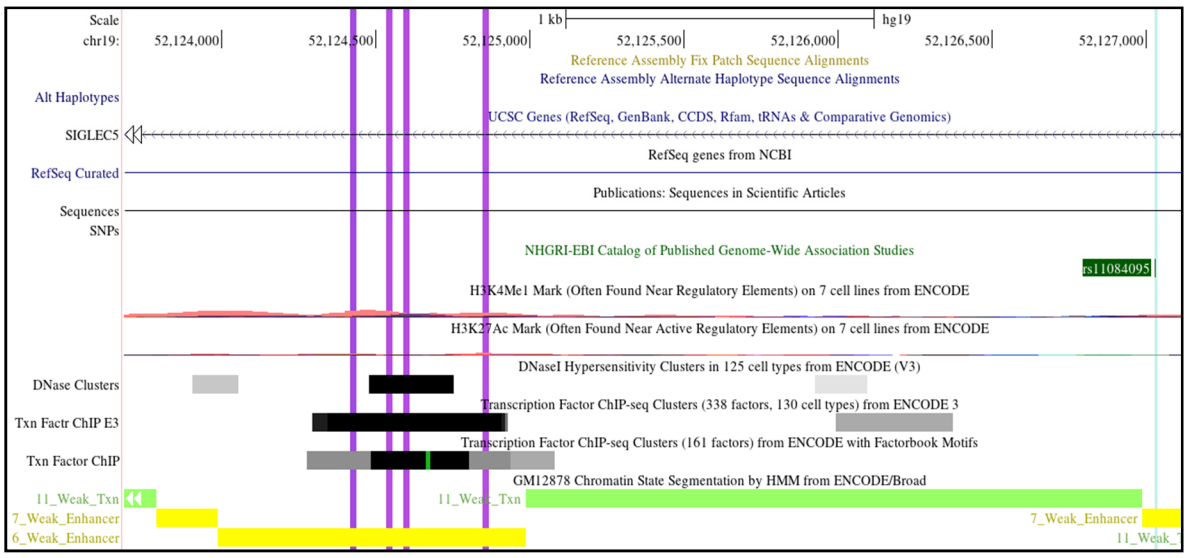


Figure 53: Position of gRNAs to test the genomic region close by rs11084095 for its potential to activate *SIGLEC5* expression. CRISPRa of the genomic region close by rs11084095 induced *SIGLEC5* expression weakly compared to the scrambled gRNA. gRNA-4 (chr19:52124417-52124435), gRNA-5 (chr19:52124848 - 52124866), gRNA-6 (chr19:52124534-52124552) and gRNA-7 (chr19:52124586 – 52124607) have 2100 - 2600 bp distance from rs11084095.

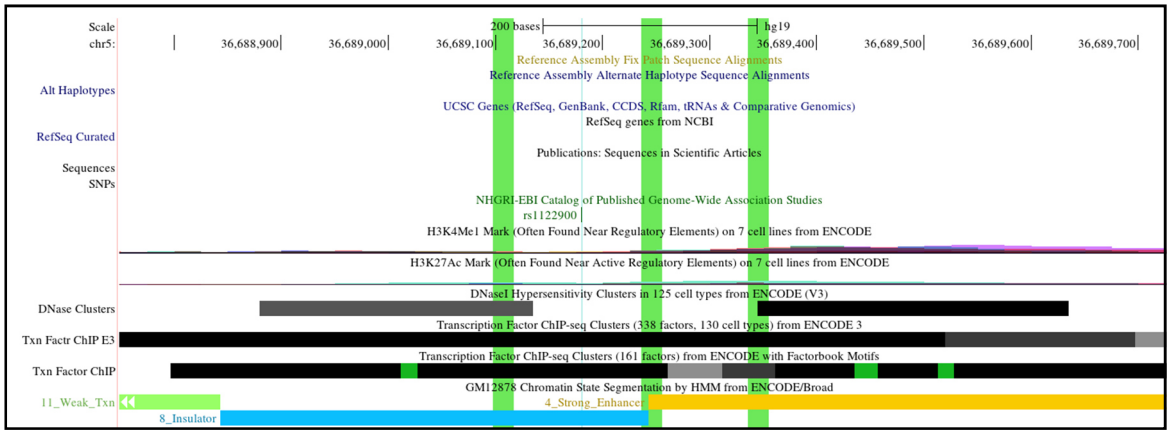


Figure 54: Position of gRNAs to test the genomic region at rs1122900 for its potential to activate *CTD-2353F22.1* expression. CRISPRa of the genomic region at rs1122900 did not induce *CTD-2353F22.1* expression compared to the scrambled gRNA. gRNA-1 (chr19:36689336-36689354), gRNA-2 (chr19:36689095-36689116) and gRNA-3 (chr19:36689237-36689255) have 50-180 bp distance from rs1122900.

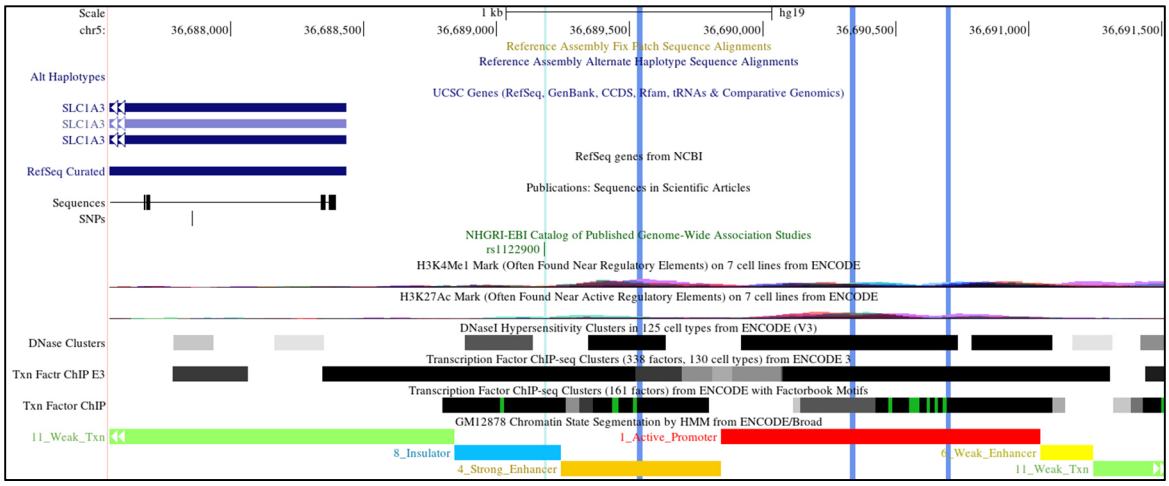


Figure 55: Position of gRNAs to test the genomic region close by rs1122900 for its potential to activate *CTD-2353F22.1* expression. CRISPRa of the genomic region close by rs1122900 induced *CTD-2353F22.1* expression very weakly compared to the scrambled gRNA. gRNA-1 (chr19: 36689528-36689546), gRNA-2 (chr19: 36690687-36690704) and gRNA-3 (chr19:36690328 – 36690346) have 390 - 1500 bp distance from rs1122900.

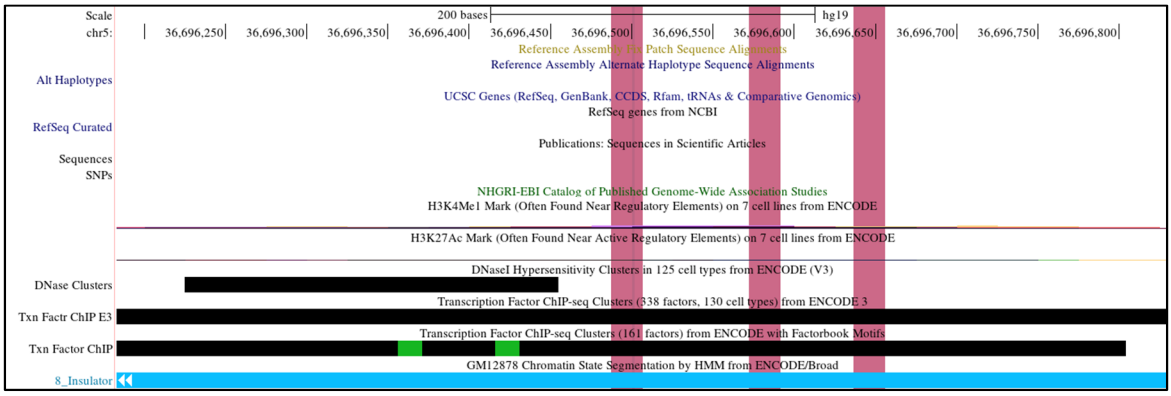


Figure 56: Position of gRNAs to test the genomic region at rs6887423 for its potential to activate *CTD-2353F22.1* expression. CRISPRa of the genomic region at rs6887423 did induce *CTD-2353F22.1* expression very weakly compared to the scrambled gRNA. gRNA-1 (chr19:36696573-36696591), gRNA-2 (chr19:36696637-36696655) and gRNA-3 (chr19:36696488-36696506) have 0-150 bp distance from rs6887423.

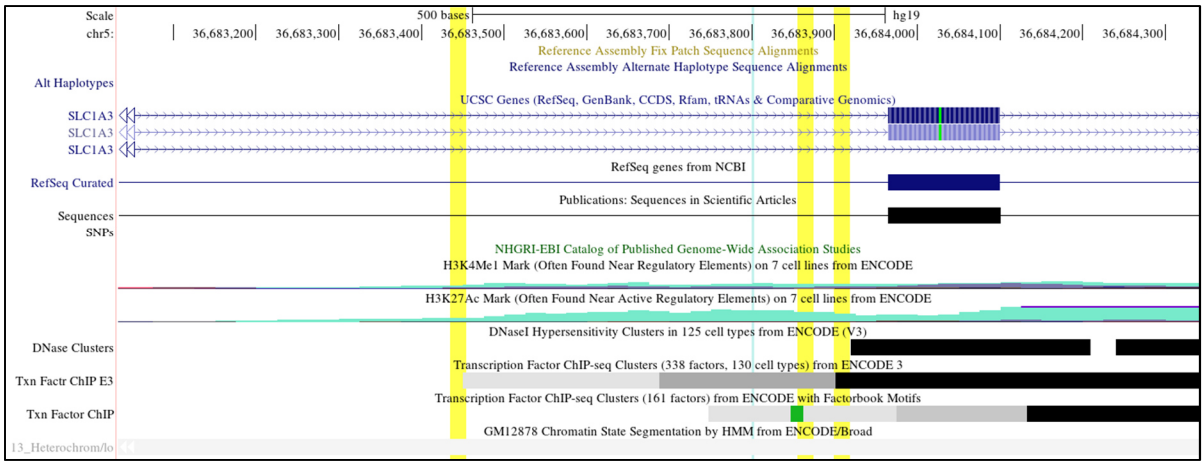


Figure 57: Position of gRNAs to test the genomic region at rs56038114 for its potential to activate *CTD-2353F22.1* expression. CRISPRa of the genomic region at rs56038114 did not induce *CTD-2353F22.1* expression compared to the scrambled gRNA. gRNA-1 (chr19:36683435-36683453), gRNA-2 (chr19:36683900-36683918) and gRNA-3 (chr19:36683855-36683837) have 60-360 bp distance from rs56038114.

8.6 Tables of miRNA study

Table 29: List of miRNAs in whole blood from periodontitis cases compared to healthy controls that were quantified with the Geniom Biochip array (Keller et al. 2011) ($p < 0.0005$) (taken from (Mueller et al. 2021)).

miRNA (miRBAv20)	p-value	Fold change
hsa-miR-744-5p	1.30E-12	2.76
hsa-miR-1228-3p	2.53E-11	5.72
hsa-miR-758-3p	1.56E-10	28.70
hsa-let-7d-5p	2.30E-10	4.02
hsa-miR-567	2.54E-09	20.22
hsa-miR-1281	2.54E-09	7.02
hsa-miR-361-5p	4.12E-09	2.75
hsa-miR-1182	2.45E-08	24.77
hsa-miR-1181	2.45E-08	1.87
hsa-miR-765	3.73E-08	17.85
hsa-miR-548h-5p	1.28E-07	11.03
hsa-miR-146b-5p	1.28E-07	1.86
hsa-miR-384	1.69E-07	3.13
hsa-miR-199b-5p	1.69E-07	19.90
hsa-miR-9-3p	2.26E-07	3.97
hsa-miR-421	2.52E-07	3.92
hsa-miR-653-5p	3.52E-07	34.38
hsa-miR-187-5p	3.92E-07	2.46
hsa-miR-637	3.92E-07	15.49
hsa-miR-125a-3p	4.40E-07	16.25
hsa-let-7b-3p	6.62E-07	20.38
hsa-miR-522-3p	9.56E-07	7.21
hsa-miR-1256	1.33E-06	0.56

hsa-miR-490-3p	1.33E-06	0.66
hsa-miR-551b-5p	1.67E-06	3.10
hsa-miR-634	1.67E-06	1.66
hsa-miR-374b-5p	2.06E-06	2.16
hsa-let-7f-2-3p	2.31E-06	11.01
hsa-miR-191-5p	3.87E-06	1.45
hsa-miR-101-5p	5.80E-06	2.95
hsa-miR-370-3p	5.89E-06	3.88
hsa-let-7a-3p	5.89E-06	16.22
hsa-miR-30c-2-3p	5.94E-06	10.38
hsa-miR-1293	6.10E-06	18.43
hsa-miR-92a-2-5p	6.46E-06	3.15
hsa-miR-499a-3p	6.61E-06	2.83
hsa-miR-770-5p	6.73E-06	0.62
hsa-miR-148a-5p	6.73E-06	18.09
hsa-miR-1307-3p	8.26E-06	2.54
hsa-miR-613	8.95E-06	1.40
hsa-miR-210-3p	1.05E-05	2.32
hsa-miR-1306-3p	1.48E-05	19.47
hsa-miR-10a-3p	1.52E-05	6.44
hsa-miR-548a-5p	1.60E-05	2.60
hsa-miR-518c-3p	1.60E-05	2.14
hsa-miR-574-5p	1.70E-05	1.93
hsa-miR-34c-3p	1.70E-05	1.52
hsa-miR-1229-3p	1.73E-05	1.58
hsa-miR-608	1.78E-05	26.02
hsa-miR-27a-3p	1.78E-05	2.50
hsa-miR-1271-5p	1.91E-05	2.41
hsa-miR-191-3p	2.07E-05	3.15
hsa-miR-450a-5p	2.59E-05	1.00
hsa-miR-10b-3p	2.61E-05	0.57
hsa-miR-936	2.74E-05	0.60
hsa-miR-671-5p	3.01E-05	0.64
hsa-miR-302b-3p	3.22E-05	1.52
hsa-miR-1294	3.23E-05	9.67
hsa-miR-382-5p	3.23E-05	4.50
hsa-miR-1178-3p	3.40E-05	22.40
hsa-miR-1237-3p	3.40E-05	3.14
hsa-miR-1252-5p	3.78E-05	1.00
hsa-miR-425-5p	3.78E-05	1.40
hsa-miR-548l	4.03E-05	1.00
hsa-miR-582-5p	4.60E-05	2.86
hsa-miR-520d-3p	4.62E-05	3.47
hsa-miR-1284	4.68E-05	33.02
hsa-miR-1231	4.68E-05	0.44

hsa-miR-509-3p	4.89E-05	1.00
hsa-miR-1245a	5.39E-05	0.49
hsa-miR-1470	5.55E-05	1.59
hsa-miR-890	5.65E-05	0.30
hsa-miR-122-5p	5.72E-05	1.76
hsa-miR-486-3p	5.72E-05	1.81
hsa-miR-18b-5p	5.97E-05	1.66
hsa-miR-654-3p	6.73E-05	36.28
hsa-miR-135b-5p	8.13E-05	1.00
hsa-miR-604	8.33E-05	0.59
hsa-miR-595	9.02E-05	2.26
hsa-miR-559	9.13E-05	0.62
hsa-miR-302a-3p	9.15E-05	1.00
hsa-miR-661	9.83E-05	27.35
hsa-miR-499a-5p	9.83E-05	0.20
hsa-let-7a-5p	1.14E-04	2.79
hsa-miR-601	1.21E-04	9.52
hsa-miR-578	1.33E-04	9.64
hsa-miR-432-5p	1.53E-04	1.00
hsa-miR-30b-3p	1.64E-04	3.11
hsa-miR-668-3p	1.64E-04	1.58
hsa-miR-26b-5p	1.70E-04	1.84
hsa-miR-154-5p	1.87E-04	1.00
hsa-miR-1206	1.96E-04	0.40
hsa-miR-615-3p	1.97E-04	27.05
hsa-miR-520f-3p	2.10E-04	1.00
hsa-miR-141-3p	2.23E-04	2.21
hsa-miR-572	2.23E-04	2.46
hsa-miR-623	2.24E-04	0.70
hsa-miR-548g-3p	2.26E-04	3.47
hsa-miR-25-5p	2.34E-04	0.57
hsa-miR-1244	2.38E-04	1.00
hsa-miR-1250-5p	2.50E-04	4.64
hsa-miR-187-3p	2.77E-04	6.20
hsa-miR-126-3p	2.85E-04	3.76
hsa-miR-196a-5p	3.06E-04	1.00
hsa-miR-520a-5p	3.21E-04	0.64
hsa-miR-1537-3p	3.31E-04	1.69
hsa-miR-1295a	3.34E-04	0.63
hsa-miR-132-5p	3.39E-04	1.00
hsa-miR-10b-5p	3.71E-04	0.56
hsa-miR-551a	3.73E-04	1.32
hsa-miR-1290	3.73E-04	1.00
hsa-miR-514a-3p	3.73E-04	0.43
hsa-miR-1277-3p	3.82E-04	0.56

hsa-miR-190a-5p	4.02E-04	1.00
hsa-miR-26b-3p	4.12E-04	5.54
hsa-miR-296-3p	4.21E-04	1.46
hsa-miR-363-5p	4.63E-04	0.57
hsa-miR-502-5p	4.75E-04	13.65
hsa-miR-193b-5p	4.85E-04	3.91
hsa-miR-1258	4.97E-04	1.00

Table 30: List of downregulated genes after miR-374b-5p mimic transfection in pGF cells for 24 hours ($p < 0.01$) (taken from (Mueller et al. 2021)).

Gene	p-value	Fold-Change
SGF29	0.00002	-1.16608
SNRK	0.00005	-1.42154
ASAP2	0.00007	-1.46182
UHMK1	0.00025	-1.76674
HSBP1	0.00031	-1.53028
CACUL1	0.00040	-1.50472
SOX5	0.00048	-1.38789
ABI3	0.00072	-1.32955
M1AP	0.00101	-1.35811
STYX	0.00122	-2.01684
DNAJB12	0.00195	-1.25404
ZNF30	0.00196	-1.56302
DCP2	0.00202	-1.3922
LIN7C	0.00238	-1.47123
TAS2R19	0.00245	-1.60884
C5orf24	0.00261	-1.36171
MS4A7	0.00265	-1.79752
C12orf54	0.00268	-1.3635
LAMTOR1	0.00269	-1.27399
TMED5	0.00319	-1.89425
SCN1B	0.00342	-1.40329
ZSWIM8	0.00361	-1.12218
SLC25A27	0.00397	-1.6291
ZNF404	0.00397	-1.3673
KCNB1	0.00408	-1.43155
METTL16	0.00424	-1.56672
CSRP1	0.00424	-1.62398
TEX101	0.00434	-1.47433
FRS2	0.00452	-1.44553
RAB8B	0.00462	-1.67045
PLB1	0.00503	-1.4728
LILRB5	0.00510	-1.45852
RMND5A	0.00521	-2.07187
PGBD1	0.00571	-1.35578
ABHD14A	0.00578	-1.25778
TTLL10	0.00610	-1.45378
PARP16	0.00612	-1.31418
MAFG	0.00631	-1.267
PKIB	0.00649	-1.31994

POMT2	0.00674	-1.27655
PRSS53	0.00691	-1.14143
TNRC6A	0.00698	-1.36914
EXOSC10	0.00709	-1.14406
DNAJC7	0.00710	-1.17521
TMEM245	0.00717	-1.59567
CHCHD4	0.00751	-1.29421
TTC1	0.00759	-1.1082
LRP6	0.00762	-1.3824
THAP8	0.00787	-1.23522
NTS	0.00800	-1.58473
KCNJ10	0.00834	-1.36135
THEM6	0.00850	-1.25718
COMMD8	0.00864	-1.48163
PRB2	0.00869	-1.25197
CAPN14	0.00872	-1.20963
SPAG7	0.00947	-1.29776
PPAT	0.00952	-1.55481
COQ7	0.00969	-1.26872
APEX2	0.00978	-1.29221
PSMC4	0.00983	-1.38408

Table 31: List of upregulated genes after miR-374b-5p mimic transfection in pGF cells for 24 hours ($p < 0.01$) (taken from (Mueller et al. 2021)).

Gene	**p-value**	**Fold-Change**
NGRN	0.00008	1.47908
TMEFF1	0.00033	1.85768
SRPR	0.00044	1.53706
ZHX1	0.00088	1.60994
ANKRD52	0.00093	1.49339
TTLL11	0.00149	1.22981
SLC39A9	0.00190	1.37483
H3F3AP4	0.00220	1.69793
RAB14	0.00229	1.24002
CCNC	0.00253	1.59468
H3F3A	0.00256	1.53788
ALS2CR12	0.00264	1.35933
PPP2R1B	0.00276	1.73566
MED27	0.00286	1.30194
SLC2A14	0.00336	2.00265
RSAD2	0.00373	1.2099
G3BP2	0.00374	1.46493
HOXB3	0.00406	1.25095
IL10RB	0.00450	1.38153
MBTPS1	0.00472	1.38964
SPZ1	0.00483	1.32809
RAB7A	0.00506	1.17801
FAM32A	0.00535	1.18994
LYSMD3	0.00543	1.66523

MOB3B	0.00551	1.49195
PRY2	0.00584	1.22484
SLC2A3	0.00586	2.09837
KLHL9	0.00614	1.71311
TOR4A	0.00654	1.18998
PPP3CB	0.00676	1.34937
STX3	0.00718	1.27401
RHO	0.00728	1.31922
PRPF38B	0.00749	1.27309
METTL8	0.00764	1.30281
ZCCHC6	0.00777	1.26866
CYB5D2	0.00781	1.42287
TLR6	0.00788	1.76508
TOMM20	0.00792	1.40739
BLMH	0.00795	1.27411
LCA5L	0.00804	1.41052
HNRNPCL4	0.00842	1.31855
KRT79	0.00852	1.27635
MFAP1	0.00924	1.26195
PCSK7	0.00959	1.30734
MSANTD3-TMEFF1	0.00970	1.79136
NFASC	0.00997	1.30148

9 Acknowledgments

I would like to take the opportunity to express my gratitude to all who have helped and encouraged me during my PhD time. The completion of this work could not have been possible without the participation, help and assistance of so many great people.

First of all, I would like to sincerely thank Prof. Dr. Arne Schäfer and Prof. Dr. Henrik Dommisch for being great supervisors during the last years and to give me the opportunity to do my PhD in such an interesting research field. I want to thank you for your guidance, support, supervision, for all of your help and for always believing in me. My special thanks also go to Prof. Dr. Sigmar Stricker, who kindly agreed to take over the second report.

I had the privilege of working in an amazing lab filled with great people that made my graduate studies enjoyable. I am grateful to work with all of you, Gesa, Avi, Bao Xin, Maren, Tugba and Nicole. Thanks for your help and support, for all the funny and interesting conversations and for making the last years to a great time for me.

Special thanks also to Ben Davies, who gave me the opportunity to work for 3 month in his amazing lab at the Oxford University. Him and all his group members teached me a lot and made this stay to a great time.

I also want to thank January who is working at at the Berlin Institute of Health Core Facility Genomics and who has done all the analysis of the RNA-seq data.

Special thanks also to the dentists and my friends Ana, Eduardo, Katrin and Kim for your support and for all the nice and funny coffee breaks.

I would especially like to thank my wonderful family. To my mom, Susanne, and my dad, Arno, for always helping and supporting me and for always being there for me. They have always believed in me and loved me unconditionally. To my siblings, Anna-Maria, Rebecca, Katharina, Lena and Thomas for all their support. Thank you for suffering and rejoicing with me and for the understanding at all times.

Lastly, I would like to thank my beautiful husband, Carsten, for his love, patience, encouragement, and support. He is my best friend, my favorite person, and my companion. Thank you for everything, for all the cooked food, for always being there for me and for always providing fun and wonderful distractions throughout my studies.

10 List of publications

List of my publications that were published until July 2021:

Mueller R et al. (2021) hsa-miR-374b-5p regulates expression of the gene U2AF homology motif (UHM) kinase 1. *J Periodontal Res*. PMID: 34160076

Chopra A, **Mueller R** et al. (2021) BACH1 Binding Links the Genetic Risk for Severe Periodontitis with ST8SIA1. *J Dent Res.* PMID: 34160287

Dommisch H, Stolte KN, Jager J, Vogel K, **Müller R** et al (2021) Characterization of an ester-based core-multishell (CMS) nanocarrier for the topical application at the oral mucosa. *Clin Oral Investig.* PMID: 33821321

Bao X, **Wiehe R**, Dommisch H, Schaefer AS (2020) *Entamoeba gingivalis* causes oral inflammation and tissue destruction. *J Dent Res.* PMID: 32023135

Freitag-Wolf S, Munz M, **Wiehe R** et al. (2019) Smoking modifies the genetic risk for early-onset periodontitis. *J Dent Res.* PMID: 31537151

Richter GM, Kruppa J, Munz M, **Wiehe R** et al (2019) A combined epigenome- and transcriptome-wide association study of the oral masticatory mucosa assigns CYP1B1 a central role for epithelial health in smokers. *Clin Epigenetics.* PMID: 31331382

Hubberten M, Bochenek G, Chen H, Häsler R, **Wiehe R** et al (2019) Linear isoforms of the long non-coding RNA CDKN2B-AS1 regulate c-myc-enhancer binding factor RBMS1. *Eur J Hum Genet.* PMID: 30108282

11. Declaration of independent work

I hereby declare that this thesis has been done by myself without any aids other than were cited. I confirm for any part of the work presented in this thesis has not previously submitted for a degree in university or institution.

Berlin, 19.07.2021_____

Ricarda Müller