The atherosclerosis risk variant rs2107595 mediates allele-specific transcriptional regulation of *HDAC9* via E2F3 and Rb1

Cover title: Functional noncoding variant at HDAC9

Authors:

Matthias Prestel, PhD*, Caroline Prell-Schicker, PhD*, Tom Webb, PhD, Rainer Malik, PhD, Barbara Lindner, Natalie Ziesch, Monika Rex-Haffner, B.Sc., Simone Röh, Dipl., Thanatip Viturawong, PhD, Manuel Lehm, MD, Michal Mokry, MD, PhD, Hester den Ruijter, PhD, Saskia Haitjema, MD, Yaw Asare, PhD, Flavia Söllner, M.A., M.Sc., Maryam Ghaderi Najafabadi, M.Sc., Rédouane Aherrahrou, PhD, Mete Civelek, PhD, Nilesh J. Samani, MD, Matthias Mann, PhD, Christof Haffner, PhD, Martin Dichgans, MD

Affiliation:

From the Institute for Stroke and Dementia Research, Klinikum der Universität München, 81377 Munich, Germany (M.P., C.P.S., R.M., B.L., N.Z., M.L., Y.A., F. S., C.H., M.D.);

Department of Cardiovascular Sciences, University of Leicester and National Institute for Research Leicester Biomedical Research Centre, Leicester, United Kingdom, Glenfield Hospital, Leicester, LE3 9QP, UK Health (T.W., M.G.N., N.J.S.);

Department of Translational Research in Psychiatry, Max-Planck-Institute for Psychiatry, 80804 Munich, Germany (M.R.H., S.R.);

Department of Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, 82152 Martinsried, Germany (T.V., M.L., M.M.);

Abteilung für Diagnostische und Interventionelle Neuroradiologie, Klinikum rechts der Isar, 81675 Munich, Germany (M.L.) ;

Department of Pediatrics, University Medical Center Utrecht, 3584 CX Utrecht, Netherlands (M.Mo.)

Laboratory of Experimental Cardiology, University Medical Center Utrecht, 3584 CX Utrecht, Netherlands (H.d.R., S.H.);

Department of Physiological Chemistry, Biomedical Center Munich, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany (F.S.);

Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22904-4741 (R.A., M.C.)

Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany (M.D.);

* These authors contributed equally to this article.

Address of correspondence:

Martin Dichgans, MD

Institute for Stroke and Dementia Research

Feodor-Lynen-Straße 17

81377 Munich

Germany

martin.dichgans@med.uni-muenchen.de

phone:+49(0)894400-46019

twitter: @ISD_Research

Total word count: 6030

Figures: 6

Subjects codes:

Gene Expression and Regulation, Ischemic Stroke, Atherosclerosis, Mechanisms

ABSTRACT

Background and Purpose: Genome-wide association studies have identified the *histone deacetylase 9* (*HDAC9*) gene region as a major risk locus for atherosclerotic stroke and coronary artery disease in humans. Previous results suggest a role of altered *HDAC9* expression levels as the underlying disease mechanism. rs2107595, the lead single nucleotide polymorphism for stroke and coronary artery disease resides in noncoding DNA and colocalizes with histone modification marks suggestive of enhancer elements.

Methods: To determine the mechanisms by which genetic variation at rs2107595 regulates *HDAC9* expression and thus vascular risk we employed targeted resequencing, proteome-wide search for allele-specific nuclear binding partners, chromatin immunoprecipitation, genome-editing, reporter assays, circularized chromosome conformation capture (4C), and gain- and loss-of-function experiments in cultured human cell lines and primary immune cells.

Results: Targeted resequencing of the HDAC9 locus in patients with atherosclerotic stroke and controls supported candidacy of rs2107595 as the causative SNP. A proteomic search for nuclear binding partners revealed preferential binding of the E2F3/TFDP1/Rb1 complex to the rs2107595 common allele, consistent with the disruption of an E2F3 consensus site by the risk allele. Gain- and loss-of-function studies showed a regulatory effect of E2F/Rb proteins on *HDAC9* expression. Compared to the common allele the rs2107595 risk allele exhibited higher transcriptional capacity in luciferase assays and was associated with higher *HDAC9* mRNA levels in primary macrophages and genome-edited Jurkat cells. 4C revealed a genomic interaction of the rs2107595 region with the *HDAC9* promoter, which was stronger for the common allele as was the *in vivo* interaction with E2F3 and Rb1 determined by chromatin immunoprecipitation. Gain-of-function experiments in isogenic Jurkat cells demonstrated a key role of E2F3 in mediating rs2107595-dependent transcriptional regulation of *HDAC9*.

Conclusions: Collectively, our findings imply allele-specific transcriptional regulation of *HDAC9* via E2F3 and Rb1 as a major mechanism mediating vascular risk at rs2107595.

KEY WORDS

Gene regulation, cell cycle, transcription factors, large artery stroke, HDAC9

INTRODUCTION

Stroke is the leading cause of permanent disability and the second most common cause of death worldwide.[1] Genome-wide association studies (GWAS) have mapped more than 35 genomic loci for stroke most residing in noncoding DNA.[2] However, at many loci the causal variant, gene, and mechanism remain undetermined[3] thus impeding the identification of novel pathways and possible targets for intervention. The histone deacetylase 9 (*HDAC9*) gene region on chromosome 7p21.1 represents the strongest risk locus for atherosclerotic stroke (large artery stroke)[2] and has further been established as a major risk locus for myocardial infarction, coronary artery disease,[4] and peripheral artery disease,[5] thus implying a broader involvement in atherosclerosis and a major impact on human health.

rs2107595, the lead single nucleotide polymorphism (SNP) in recent GWAS for stroke[2, 6] and coronary artery disease[4] resides in noncoding DNA 3' to the *HDAC9* gene. rs2107595 colocalizes with DNase I hypersensitive sites (DHS) and histone modification marks H3K27ac and H3K4me1 (ENCODE,[7] genome build hg19) indicating a possible involvement in gene regulatory mechanisms.[8]

We and others recently provided evidence for a central role of HDAC9 expression levels in atherogenesis and stroke: first, *Hdac9* deficiency attenuates atherogenesis in mouse models of atherosclerosis.[9, 10] Second, HDAC9 expression levels were found to be elevated in human atherosclerotic plaques.[11] Third, gene expression studies in peripheral blood mononuclear cells (PBMCs) revealed an association between the rs2107595 risk allele and elevated levels of HDAC9 mRNA expression with a gene dosage effect.[10] The same variant further associates with both carotid intima media thickness and the presence of atherosclerotic plaques in the common carotid artery.[11, 12] Collectively, these findings point to the possibility that the rs2107595 region mediates disease risk by influencing HDAC9 expression levels.

In the current study, we aimed to elucidate the molecular mechanisms linking genetic variation in the rs2107595 region to HDAC9 expression. For this we employed targeted resequencing of the HDAC9 locus, proteome-wide search for allele-specific nuclear binding partners, chromatin immunoprecipitation (ChIP), genome-editing, reporter assays, circularized chromosome conformation capture (4C), and gain- and loss-of-function experiments in cultured human cell lines and primary vascular and immune cells. We provide evidence for a regulatory effect of rs2107595 on HDAC9 expression involving a direct physical interaction between the rs2107595 region and the *HDAC9* promoter. We further demonstrate a role of the E2F3 and Rb1 proteins in mediating allele-specific effects of rs2107595 on HDAC9 transcription.

METHODS

All data and supporting materials have been provided with the published article. Tables and a detailed description of the methodology used for Targeted Resequencing, Proteome-Wide Analysis of SNPs (PWAS), Chromatin immunoprecipitation (ChIP), Cell culture and Transfection, RNA isolation and cDNA synthesis, Protein isolation and Immunoblotting, Gene expression analysis, Cell cycle synchronization, the Isolation and Culture of Human primary Aortic Smooth Muscle cells (HAoSMCs) and human blood-derived macrophages (MΦ), Dual luciferase reporter assay, Generation of genome-edited Jurkat cell lines**,** Circular Chromosome Conformation Capture and Cell proliferation assays is provided in the Supplementary Methods (http://stroke.ahajournals.org.)

Experiments in primary human cells were approved by the local institutional review board (project #17-693). Primary human blood-derived MΦ were obtained from healthy volunteers. Primary HAoSMC were obtained from Dr. Civelek (University of Virginia).

Statistical analysis

The Shapiro-Wilks Test was utilized to determine the distribution of data sets. Normally distributed data were statistically analysed with the parametric T-Test, else a Wilcoxon Rank-Sum Test or

Wilcoxon Signed-Rank Test were applied. Data are represented as mean values and standard error of the mean unless specified otherwise. Significance is depicted as follows; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. HDAC9 regional plots (**Figure 1A**) were constructed using locuszoom. The upper panel uses data from the large artery stroke analysis of the MEGASTROKE collaboration.[2]

RESULTS

Targeted resequencing of the HDAC9 region supports candidacy of rs2107595 as the causal variant for large artery stroke

rs2107595 gave the strongest signal in previous GWAS for atherosclerotic phenotypes,[2, 6] (**Figure 1A, upper panel**) and had a >95% posterior probability of being the only causal SNP at this locus in the most recent stroke GWAS.[2] To further examine the candidacy of rs2107595 as the causal variant at this locus while also capturing rare variants, low-frequency variants, and haplotypes, we performed targeted resequencing of the *HDAC9* gene region including the nearby *TWIST1* and *FERD3L* genes in 176 patients with large artery stroke and 176 stroke-free controls (**Figure 1A**, **middle panel**; **Figure I**). Genotypes for rs2107595 showed 99.8% agreement with previously obtained microarray and TaqMan genotyping data demonstrating the reliability of our sequencing approach. Overall, we identified 9,428 variants (8,496 SNPs, 932 insertions/deletions) and 169 haplotype blocks but no rare or low-frequency variants in the rs2107595 haplotype block. Following correction for multiple testing, none of the variants or haplotypes significantly associated with large artery stroke thus arguing against variants with large effect sizes in this region. Next, we used variant-collapsing methods (SKAT and SKAT-O) to analyse the conserved 2.5-kb sequence block around rs2107595, the intergenic region between *HDAC9* and *TWIST1*, and the *HDAC9*, *TWIST1*, and *FERD3L* genes (**Figure 1A**, **lower panel**). SKAT-O analyses revealed a significant association (p=0.017) for the conserved sequence block encompassing rs2107595, while all other equally-sized sequence blocks showed higher p-values. Of note, all proxy SNPs (r^2 with rs2107595 >0.8) localize outside the

conserved sequence block. Collectively, these findings support rs2107595 as the causative variant at this locus. Hence, we focused on this variant in our functional analyses.

The rs2107595 risk variant interferes with E2F3 binding

The rs2107595 region shows enrichment for marks of regulatory chromatin (DHS, H3K27ac, H3K4me1, H3K9me3) in various cell types listed in HaploReg,[13] Roadmap Epigenomics,[14] and ENCODE[7] (**Table I and II**, **Figure IIA**) suggesting a potential involvement of rs2107595 in transcriptional regulation. To identify transcription factors with allele-specific binding at rs2107595 and hence a possible role in transcriptional regulation, we performed proteome-wide analysis of SNPs (PWAS). This approach is based on the interaction of synthetic oligonucleotides with metabolically labelled nuclear factors that are subsequently identified by mass spectrometry.[15] 41-bp-SNP-centered oligonucleotides differing only at rs2107595 (**Table III**) were incubated either with light or heavy isotope labelled nuclear factors from HeLa cells. A comparison of the heavy/light ratios of all binding proteins revealed six factors surpassing the predefined FDR of 0.01: NFATC2, a member of the nuclear factor of activated T-cells,[16] L3MBTL3, a putative polycomb group protein functioning as transcriptional regulator in large protein complexes,[17] SAMD1, a protein with a potential role in immobilizing low density lipoprotein (LDL) in the arterial wall,[18] and all constituents of the E2F3/TFDP1/Rb1 complex (**Figure 1B**).

E2F3 and TFDP1 represent transcription factors of the E2F and DP1 families known to complex with Rb proteins.[19] The observed enrichment of E2F3 at the common allele is supported by the prediction of an E2F3 consensus site[20] within the common allele sequence which is disrupted by the risk allele (**Figure 1C**). To validate allele-specific binding of E2F3 we further incubated biotinylated synthetic oligonucleotides with nuclear extracts from HeLa cells and purified the assembled allele-specific nucleoprotein complexes by DNA pull-down. Subsequent immunoblotting revealed enriched binding of E2F3 to the common allele (**Figure IIB**). Finally, we performed ChIP experiments in HeLa cells, which are homozygous for the rs2107595 common allele and thus suited to explore E2F3 binding *in vivo*. ChIP revealed a significant occupancy of E2F3 at rs2107595 (**Figure 1D**). Given these results

and the known role of E2F and Rb proteins in transcriptional regulation[21, 22] we considered these proteins to be strong candidates for regulating HDAC9 expression.

E2F3 and Rb1 regulate HDAC9 expression

To determine the effect of E2F and Rb proteins on endogenous HDAC9 expression we next conducted gain- and loss-of-function (**Table IV**) experiments in HeLa cells. Overexpression of E2F3a resulted in a 6-fold increase in HDAC9 mRNA levels compared to empty vector control. In contrast, overexpression of Rb1 led to a reduction in HDAC9 expression (**Figure 2A and III A, B**). siRNA-mediated knockdown of E2F3, E2F4, or both resulted in a significant decrease of HDAC9 mRNA compared to non-targeting control (**Figure 2B and IIIC, D**). In contrast, knockdown of Rb proteins caused a significant increase in HDAC9 expression (**Figure 2C and III E, F**).

E2F and Rb act as transcriptional regulators of cell cycle genes. At the G_1/S boundary repressive Rb proteins become phosphorylated by cyclin-dependent kinases and dissociate from E2F proteins, which then activates the expression of target genes.[21, 22] Hence, we analysed cell cycle-dependent variations in HDAC9 expression. Synchronization of HeLa cells by hydroxyurea-(HU)-induced cell cycle arrest at the G_1/S boundary led to a significant increase in HDAC9 mRNA expression compared to untreated cells (**Figure 2D**). Following release of the cell cycle arrest HDAC9 mRNA expression further increased during progression through S phase and declined upon reaching G2, thus paralleling the activity of E2F proteins across the cell cycle.[23]

The rs2107595 risk variant is associated with elevated HDAC9 transcription

To examine the association between rs2107595 and HDAC9 gene expression in cells relevant to atherosclerosis we first examined primary human macrophages (MΦ) and human aortic smooth muscle cells (HAoSMCs) with defined carrier status at rs2107595. Proinflammatory MΦ were isolated from PBMCs obtained from healthy donors (GG genotype: n=7; GA: n=7; AA: n=5, matched for age and gender) and differentiated *in vitro* (**Figure IV A**). Upon stimulation with TNFα and IFNγ, MΦ homozygous for the risk allele showed significantly elevated HDAC9 expression levels

compared to MΦ homozygous for the common allele (**Figure 3A**). Gene expression analysis in cultured HAoSMC (GG genotype: n=9; AA: n=6) revealed no allele-specific differences in HDAC9 expression before and after 4 or 8 h of TNFα stimulation (**Figure 3B**). Also, there was no allele-specific effect on TWIST1 expression in HAoSMCs and MΦ (**Figure IV B and** results not shown).

To examine the effects of rs2107595 on transcriptional regulation, we further performed luciferase reporter assays in T-lymphoid Jurkat cells, THP-1 monocytes and PMA-induced THP-1 MΦ, HAoEC, and HAoSMC. 41-bp-SNP centered fragments containing either the rs2107595 common or risk variant were cloned into a firefly luciferase reporter vector (**Figure 4A**) and tested for a *cis*-regulatory function by measuring luciferase activity after transient transfection. Transcriptional activity was significantly higher for the risk allele compared to the common allele both in Jurkat cells and PMA-induced THP-1 MΦ[24] (**Figure 4A**) whereas we found no allele-specific differences in HAoEC, HAoSMCs, and THP-1 monocytes, (**Figure V A-C**).

Next, we specifically genome-edited rs2107595 in Jurkat cells using recombinant adeno-associated virus[25] (rAAV) resulting in isogenic cells differing solely at rs2107595. Jurkat cells were chosen because of (1) their immunological origin, (2) the presence of open chromatin marks both in the rs2107595 region (**Figure II**) and *HDAC9* promoter, and (3) their diploidy and heterozygosity for rs2107595[7, 14, 26] allowing a one-step editing procedure in either direction (**Figure 4B**). Cells homozygous for the risk allele exhibited 2-fold higher mRNA levels of HDAC9 compared to cells carrying the common allele (**Figure 4C**). Heterozygous cells displayed intermediate mRNA levels compatible with a gene dosage effect. TWIST1 and FERDL3 expression levels were below detection limit in these cells (data not shown). Collectively, these results show that rs2107595 regulates HDAC9 transcription in an allele-specific manner. We further examined allele-specific effects of rs2107595 on HDAC9 transcription across the cell cycle. Following synchronization at the G_1/S-boundary, HDAC9 levels were significantly elevated in risk allele cells compared to common allele cells (**Figure 4D**) in accordance with the results obtained in unsynchronized cells (**Figure 4C**). This difference was sustained for 6 h following release of the HU block. Because of the allele-specific effects on cell cycle associated HDAC9 expression we analysed the effect of rs2107595 on cell

proliferation in genome-edited Jurkat cells. Pulse-chase labeling with the thymidine analogue EdU and detection by flow cytometry revealed no allele-specific differences for rs2107595 (**Figure VI A and B**).

The rs2107595 region physically interacts with a HDAC9 promoter

Given the observed effect of rs2107595 on HDAC9 transcription we next tested for physical interactions of the rs2107595 region with the HDAC9 promoter by circularized chromosome conformation capture (4C) in isogenic Jurkat cells. Based on Jurkat cell-specific open chromatin structure (DHS) and promoter information (H3K4me3)[7] we selected the promoter viewpoint at nt ~18,330,000. Mapping the 4C-seq signals to the HDAC9 gene region revealed a significant interaction between rs2107595 and the promoter region in common allele ("GG" in **Figure 5**) but not in risk allele cells ("AA") indicating allele-specific differences in chromatin organisation. Analyses for an alternative HDAC9 promoter lacking detectable chromatin marks in Jurkat cells showed lower significance for allele-specific interactions at both viewpoints (**Figure VII**). These results provide further mechanistic evidence for a role of the rs2107595 region in regulating HDAC9 transcription.

E2F3 mediates allele-specific effects of rs2107595 on HDAC9 transcription.

To determine whether the binding of E2F3 and Rb1 at rs2107595 observed in HeLa cells occurs in a truly allele-specific manner *in vivo* we next performed ChIP experiments in genome-edited isogenic Jurkat cells. Since E2F3 and Rb1 control cell cycle progression at the G1/S boundary,[23] we arrested these cells with HU. Upon synchronization, we found a significantly enriched occupancy of E2F3 and Rb1 proteins at the common allele compared to the risk allele (**Figure 6A and B**), which was not present in unsynchronized cells (**Figure VIII A and B**) possibly reflecting cell cycle-dependent binding of E2F3 and Rb1 to the common allele.

Finally, to examine whether the allele-specific effects on HDAC9 transcription at rs2107595 are mediated by allele-specific binding of E2F3 and Rb1, we tested the influence of exogenous E2F3a and Rb1 expression in isogenic Jurkat cells. Compared to empty vector control, overexpression of

E2F3a but not Rb1 resulted in a significant increase of the ratio between HDAC9 expression in cells homozygous for the common allele vs cells homozygous for the risk allele (**Figure 6C and VIIIC**). Collectively, these results suggest allele-specific interactions between rs2107595 and the HDAC9 promotor and show a mediating effect of E2F3 on HDAC9 expression via rs2107595 (**Figure 6D**).

DISCUSSION

We present a mechanism by which a noncoding variant at the large artery stroke and coronary artery disease risk locus 7p21.1 regulates HDAC9 transcription. We show that rs2107595, the likely causal variant at this locus, has allele-specific transcriptional capacity and associates with elevated HDAC9 expression in cell types relevant to atherosclerosis. We further identify a physical interaction of the rs2107595 region with the HDAC9 promoter, demonstrate preferential binding of the E2F3/TFDP1/Rb1 cell-cycle complex to the common allele, and show that E2F3 mediates HDAC9 transcription in an allele-specific manner. This novel mechanism for transcriptional regulation of HDAC9 by E2F3/Rb1 complexes provides a plausible mechanistic link between genetic variation at rs2107595 and disease risk.

Several lines of evidence point to rs2107595 as the causal variant mediating vascular risk: rs2107595 was the lead SNP in GWAS for stroke[2, 6] and coronary artery disease,[4] it was the only variant contained in the 95% credible SNP set in the MEGASTROKE,[2] and here using targeted sequencing and SKAT-O analyses we find no variants with large effect sizes in the HDAC9 region. A transcriptional effect of rs2107595 on HDAC9 expression is demonstrated by our results in genome-edited T-lymphoid Jurkat cells and in primary proinflammatory MΦ, and is further substantiated by the 4C results, which showed a physical interaction between the rs2107595 region and the HDAC9 promoter. The directionality of the transcriptional effect was consistent with results from luciferase assays in Jurkat cells and THP-1 MΦ. It was further consistent with the effects on HDAC9 transcription reported previously for PBMCs[10] in that the risk allele was associated with higher

HDAC9 expression levels. Of note, however, our earlier observations in PBMCs did not allow attributing allele-specific effects to a specific genetic variant. As such, the current findings represent a major advance.

Our results suggest that the effects of rs2107595 on HDAC9 expression might be cell-type dependent. While the rs2107595 risk allele was associated with higher HDAC9 expression levels in proinflammatory human MΦ and genome-edited T-lymphoid Jurkat cells, we found no indication for an allele-specific effect in cultured HAoSMC. Similarly, luciferase assays showed a higher transcriptional activity with the risk allele in Jurkat cells and proinflammatory THP-1 MΦ but not in undifferentiated THP-1 monocytes, HAoSMCs and HAoECs. Genome-edited Jurkat cells showed a constitutive allele-specific effect on HDAC9 expression but an inflammatory stimulus was required to uncover a risk allele-dependent increase in HDAC9 expression in THP-1 cells and proinflammatory human MΦ. This might be due to a cell-type-specific chromatin conformation affecting the accessibility of transcription factors and thus gene expression.[14] Despite a cell cycle-dependent HDAC9 expression and the proposed role of HDAC9 in cell proliferation and cancer,[27-30] we found no allele-specific effect on cell proliferation in isogenic Jurkat cells. However, this might relate to Jurkat cells lacking functional p53,[31] which is transcriptionally regulated by HDAC9.[30] Additional work is needed to determine a possible role of rs2106595-mediated control of HDAC9 in cell proliferation.

An allele-specific interaction between rs2107595 and E2F3/Rb1 complexes is supported by four independent lines of evidence: our proteome-wide analysis of allele-specific binding partners, DNA pull-down experiments in combination with immunoblotting, ChIP, and the presence of a consensus-binding site for E2F3 at the common allele. Importantly, the directionality was consistent across all approaches in that the risk allele disrupted binding to E2F3. There is evidence for a role of Rb in atherosclerosis.[32] MΦ specific deletion of Rb has previously been shown to enhance atherosclerosis in ApoE deficient mice. Aside from their crucial function in cell cycle regulation, Rb and E2F proteins cooperatively regulate transcriptional programs for development, metabolism and cell differentiation.[21] For instance, both proteins are required for proper myeloid cell development[33, 34] and control migration and senescence of vascular SMC in human atherosclerotic lesions.[35, 36] E2F and Rb further induce foam cell formation through the mTor/SREBP-2 pathway upon inflammatory stress.[37]

Hence, loss of Rb/E2F binding at the rs2107595 risk variant and the associated increase in HDAC9 expression might provide a proinflammatory environment promoting atheroprogression. HDAC9 is known to act as a proinflammatory factor,[9, 10, 38-40] and a recent gene expression study in LAS patients found the rs2107595 risk allele to be associated with enhanced IL-6 signalling in peripheral blood.[41] However, HDAC9 also mediates cholesterol efflux in mouse MΦ.[9] Thus, there may be a synergistic effect of Rb/E2F-mediated HDAC9 expression on both cholesterol metabolism and inflammation in mediating atherosclerosis risk.[39, 41, 42] Our proteome-wide experiment identified differential interactors aside from E2F3 and Rb proteins and we cannot exclude a role of these factors in mediating allele-specific effects.[17, 43] Yet, the binding of three proteins belonging to the same complex (E2F3, TFDP1, and Rb1) together with our functional results strongly support a major role of E2F3/Rb1 in mediating the effects of rs2107595 on HDAC9 expression.

HDAC9 has emerged as a potential target for drug development. For one, there is evidence from different mouse models of atherosclerosis that lowering HDAC9 expression may attenuate atherogenesis.[9, 10] Second, rs2107595 has been associated with early stages of atherogenesis,[11, 12] which makes HDAC9 an attractive target for early intervention. Third, recent drug discovery programs have resulted in the development of selective class IIa HDAC inhibitors with reasonable specificity and inhibitory activity against HDAC9.[44] Interest in HDAC9 further emerges from the observation that the HDAC9 locus is implicated in three major manifestations of atherosclerosis: stroke, coronary artery disease, and peripheral artery disease. More work is needed to better understand the mechanisms linking genetic variation in the rs2107595 region to atherosclerosis and stroke.

ACKNOWLEDGMENTS

We thank Joseph R. Nevins and Alexander Brehm for providing reagents, Horizon Discovery, Cambridge, UK for support in generating genome-edited cell lines, Noortje A. M. van den Dungen for technical assistance and the Utrecht Sequencing Facility for performing sequencing of the 4C libraries.

SOURCES OF FUNDING

This work was supported by grants from the Deutsche Forschungsgemeinschaft (CRC 1123 and Munich Cluster for Systems Neurology), Bundesministerium für Bildung und Forschung (e:AtheroSysMed), FP7/2007-2103 European Union project CVgenes@target (Health-F2-2013-601456), Leducq Foundation CADgenomics programme, European Union Horizon2020 projects SVDs@target (No.66688) and CoSTREAM (No.667375), and the Vascular Dementia Research Foundation. Tom Webb and Nilesh J. Samani are supported by the British Heart Foundation. Nilesh J. Samani is a National Health Research senior investigator. Matthias Mann was funded by the Max Planck Society and Rédouane Aherrahrou by the American Heart Association. Michal Mokry is supported by the Netherlands Organisation for Scientific Research, the Netherlands Organisation for Health Research and Development, Dutch Heart Foundation and Leducq Foundation.

DISCLOSURES

None.

REFERENCES

1. Collaborators GBDLRoS, Feigin VL, Nguyen G, Cercy K, Johnson CO, Alam T, et al. Global, Regional, and Country-Specific Lifetime Risks of Stroke, 1990 and 2016. *N Engl J Med*. 2018;379:2429-2437. doi: 10.1056/NEJMoa1804492.
2. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524-537. doi: 10.1038/s41588-018-0058-3.
3. Dichgans M, Pulit SL, Rosand J. Stroke genetics: discovery, biology, and clinical applications. *Lancet Neurol*. 2019;18:587-599. doi: 10.1016/S1474-4422(19)30043-2.
4. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49:1385-1391. doi: 10.1038/ng.3913.
5. Matsukura M, Ozaki K, Takahashi A, Onouchi Y, Morizono T, Komai H, et al. Genome-Wide Association Study of Peripheral Arterial Disease in a Japanese Population. *PLoS One*. 2015;10:e0139262. doi: 10.1371/journal.pone.0139262.
6. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, et al. Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology*. 2016;86:1217-1226. doi: 10.1212/WNL.0000000000002528.
7. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46:D794-D801. doi: 10.1093/nar/gkx1081.
8. Meng H, Bartholomew B. Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. *J Biol Chem*. 2018;293:13786-13794. doi: 10.1074/jbc.R117.813485.
9. Cao Q, Rong S, Repa JJ, St Clair R, Parks JS, Mishra N. Histone deacetylase 9 represses cholesterol efflux and alternatively activated macrophages in atherosclerosis development. *Arterioscler Thromb Vasc Biol*. 2014;34:1871-1879. doi: 10.1161/ATVBAHA.114.303393.
10. Azghandi S, Prell C, van der Laan SW, Schneider M, Malik R, Berer K, et al. Deficiency of the stroke relevant HDAC9 gene attenuates atherosclerosis in accord with allele-specific effects at 7p21.1. *Stroke*. 2015;46:197-202. doi: 10.1161/STROKEAHA.114.007213.
11. Markus HS, Makela KM, Bevan S, Raitoharju E, Oksala N, Bis JC, et al. Evidence HDAC9 genetic variant associated with ischemic stroke increases risk via promoting carotid atherosclerosis. *Stroke*. 2013;44:1220-1225. doi: 10.1161/STROKEAHA.111.000217.
12. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun*. 2018;9:5141. doi: 10.1038/s41467-018-07340-5.
13. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016;44:D877-881. doi: 10.1093/nar/gkv1340.
14. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317-330. doi: 10.1038/nature14248.
15. Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet*. 2012;8:e1002982. doi: 10.1371/journal.pgen.1002982.
16. Modiano JF, Johnson LD, Bellgrau D. Negative regulators in homeostasis of naive peripheral T cells. *Immunol Res*. 2008;41:137-153. doi: 10.1007/s12026-008-8017-1.
17. Meier K, Brehm A. Chromatin regulation: how complex does it get? *Epigenetics*. 2014;9:1485-1495. doi: 10.4161/15592294.2014.971580.
18. Lees AM, Deconinck AE, Campbell BD, Lees RS. Atherin: a newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis*. 2005;182:219-230. doi: 10.1016/j.atherosclerosis.2005.01.041.
19. Girling R, Partridge JF, Bandara LR, Burden N, Totty NF, Hsuan JJ, et al. A new component of the transcription factor DRTF1/E2F. *Nature*. 1993;362:83-87. doi: 10.1038/362083a0.

20. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res*. 2018;46:D252-D259. doi: 10.1093/nar/gkx1106.
21. Korenjak M, Brehm A. E2F-Rb complexes regulating transcription of genes important for differentiation and development. *Curr Opin Genet Dev*. 2005;15:520-527. doi: 10.1016/j.gde.2005.07.001.
22. Blais A, Dynlacht BD. E2F-associated chromatin modifiers and cell cycle control. *Curr Opin Cell Biol*. 2007;19:658-662. doi: 10.1016/j.ceb.2007.10.003.
23. Stevaux O, Dyson NJ. A revised picture of the E2F transcriptional network and RB function. *Curr Opin Cell Biol*. 2002;14:684-691. doi: 10.1016/S0955-0674(02)00388-5.
24. Lund ME, To J, O'Brien BA, Donnelly S. The choice of phorbol 12-myristate 13-acetate differentiation protocol influences the response of THP-1 macrophages to a pro-inflammatory stimulus. *J Immunol Methods*. 2016;430:64-70. doi: 10.1016/j.jim.2016.01.012.
25. Jones PD, Kaiser MA, Ghaderi Najafabadi M, McVey DG, Beveridge AJ, Schofield CL, et al. The Coronary Artery Disease-associated Coding Variant in Zinc Finger C3HC-type Containing 1 (ZC3HC1) Affects Cell Cycle Regulation. *J Biol Chem*. 2016;291:16318-16327. doi: 10.1074/jbc.M116.734020.
26. Schneider U, Schwenk HU, Bornkamm G. Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer*. 1977;19:621-626. doi: 10.1002/ijc.2910190505.
27. Yang R, Wu Y, Wang M, Sun Z, Zou J, Zhang Y, et al. HDAC9 promotes glioblastoma growth via TAZ-mediated EGFR pathway activation. *Oncotarget*. 2015;6:7644-7656. doi: 10.18632/oncotarget.3223.
28. Yuan Z, Peng L, Radhakrishnan R, Seto E. Histone deacetylase 9 (HDAC9) regulates the functions of the ATDC (TRIM29) protein. *J Biol Chem*. 2010;285:39329-39338. doi: 10.1074/jbc.M110.179333.
29. Zhang Y, Wu D, Xia F, Xian H, Zhu X, Cui H, et al. Downregulation of HDAC9 inhibits cell proliferation and tumor formation by inducing cell cycle arrest in retinoblastoma. *Biochem Biophys Res Commun*. 2016;473:600-606. doi: 10.1016/j.bbrc.2016.03.129.
30. Zhao YX, Wang YS, Cai QQ, Wang JQ, Yao WT. Up-regulation of HDAC9 promotes cell proliferation through suppressing p53 transcription in osteosarcoma. *Int J Clin Exp Med*. 2015;8:11818-11823.
31. Gioia L, Siddique A, Head SR, Salomon DR, Su AI. A genome-wide survey of mutations in the Jurkat cell line. *BMC Genomics*. 2018;19:334. doi: 10.1186/s12864-018-4718-6.
32. Boesten LS, Zadelaar AS, van Nieuwkoop A, Hu L, Jonkers J, van de Water B, et al. Macrophage retinoblastoma deficiency leads to enhanced atherosclerosis development in ApoE-deficient mice. *FASEB J*. 2006;20:953-955. doi: 10.1096/fj.05-4530fje.
33. Bergh G, Ehinger M, Olsson I, Jacobsen SE, Gullberg U. Involvement of the retinoblastoma protein in monocytic and neutrophilic lineage commitment of human bone marrow progenitor cells. *Blood*. 1999;94:1971-1978.
34. Trikha P, Sharma N, Opavsky R, Reyes A, Pena C, Ostrowski MC, et al. E2f1-3 are critical for myeloid development. *J Biol Chem*. 2011;286:4783-4795. doi: 10.1074/jbc.M110.182733.
35. Andres V. Control of vascular cell proliferation and migration by cyclin-dependent kinase signalling: new perspectives and therapeutic potential. *Cardiovasc Res*. 2004;63:11-21. doi: 10.1016/j.cardiores.2004.02.009.
36. Bennett MR, Macdonald K, Chan SW, Boyle JJ, Weissberg PL. Cooperative interactions between RB and p53 regulate cell proliferation, cell senescence, and apoptosis in human vascular smooth muscle cells from atherosclerotic plaques. *Circ Res*. 1998;82:704-712. doi: 10.1161/01.RES.82.6.704.
37. Ma KL, Liu J, Wang CX, Ni J, Zhang Y, Wu Y, et al. Activation of mTOR modulates SREBP-2 to induce foam cell formation through increased retinoblastoma protein phosphorylation. *Cardiovasc Res*. 2013;100:450-460. doi: 10.1093/cvr/cvt203.
38. Beier UH, Wang L, Han R, Akimova T, Liu Y, Hancock WW. Histone deacetylases 6 and 9 and sirtuin-1 control Foxp3+ regulatory T cell function through shared and isoform-specific mechanisms. *Sci Signal*. 2012;5:ra45. doi: 10.1126/scisignal.2002873.

39. Han X, Han X, Wang Z, Shen J, Dong Q. HDAC9 regulates ox-LDL-induced endothelial cell apoptosis by participating in inflammatory reactions. *Front Biosci (Landmark Ed)*. 2016;21:907-917. doi: 10.2741/4428.
40. Lu S, Li H, Li K, Fan XD. HDAC9 promotes brain ischemic injury by provoking IkappaBalpha/NF-kappaB and MAPKs signaling pathways. *Biochem Biophys Res Commun*. 2018;503:1322-1329. doi: 10.1016/j.bbrc.2018.07.043.
41. Shroff N, Ander BP, Zhan X, Stamova B, Liu D, Hull H, et al. HDAC9 Polymorphism Alters Blood Gene Expression in Patients with Large Vessel Atherosclerotic Stroke. *Transl Stroke Res*. 2019;10:19-25. doi: 10.1007/s12975-018-0619-x.
42. Blaschke F, Leppanen O, Takata Y, Caglayan E, Liu J, Fishbein MC, et al. Liver X receptor agonists suppress vascular smooth muscle cell proliferation and inhibit neointima formation in balloon-injured rat carotid arteries. *Circ Res*. 2004;95:e110-123. doi: 10.1161/01.RES.0000150368.56660.4f.
43. Baksh S, Widlund HR, Frazer-Abel AA, Du J, Fosmire S, Fisher DE, et al. NFATc2-mediated repression of cyclin-dependent kinase 4 expression. *Mol Cell*. 2002;10:1071-1081. doi.
44. Lobera M, Madauss KP, Pohlhaus DT, Wright QG, Trocha M, Schmidt DR, et al. Selective class IIa histone deacetylase inhibition via a nonchelating zinc-binding group. *Nat Chem Biol*. 2013;9:319-325. doi: 10.1038/nchembio.1223.

Figure Legends

Figure 1: The rs2107595 risk variant interferes with E2F3 binding. (**A**) Top: regional association plot of the HDAC9 gene region (18123000-19188000, GRCh37/hg19) showing association signals around rs2107595 for large artery stroke in the MEGASTROKE dataset.[2] Middle: association plot of the same region showing variants identified by targeted resequencing. Bottom: -$\log_{10}$ p-values for the conserved sequence element around rs2107595, the intergenic region between HDAC9 and TWIST1, and the HDAC9, TWIST1, and FERD3L genes, calculated by variant-collapsing methods (SKAT and SKAT-O). The conserved 2.5-kb sequence block around rs2107595 (position marked by the dashed line) significantly associated with large artery stroke (p=0.017). (**B**) Identification of allele-specific binding partners of rs2107595 using PWAS. E2F3, Rb1, TFDP1, SAMD1 and L3MBTL3 preferentially interacted with the common allele (G) whereas NFATC2 preferentially bound to the risk allele (A). (**C**) Position Weight Matrix[20] for the consensus site of the human E2F3 protein aligned to the genomic sequence surrounding rs2107595. (**D**) ChIP experiments showing *in vivo* binding of E2F3 to the rs2107595 region in HeLa cells. Fold change **(**FC). (n=7-8, mean±SD. Wilcoxon Signed-Rank Test).

Figure 2: E2F3 and Rb1 regulate HDAC9 expression. **(A-C)** FC in HDAC9 mRNA expression assessed in HeLa cells after (**A**) overexpression of E2F3a and Rb1, (**B**) siRNA mediated knockdown of E2F3 and E2F4 and (**C**) siRNA mediated knockdown of Rb1, Rbl1 and Rbl2. n=7. (**D**) Cell cycle analysis by flow cytometry and propidium iodide staining in HeLa cells following cell cycle arrest at the G1/S boundary by hydroxurea (HU). HDAC9 expression is increased at the G_1/S boundary and during S phase. (n=6-7. FC mean±SEM. Wilcoxon Signed-Rank Test).

Figure 3: The rs2107595 risk variant is associated with elevated HDAC9 transcription in human primary MΦ. (**A**) Human blood-derived monocytes were differentiated *in vitro* to proinflammatory MΦ. Upon TNFα and IFNγ stimulation MΦ homozygous for the risk allele displayed significantly higher HDAC9 expression levels compared to common allele carriers (GG: n=5; GA: n=5; AA: n=7).

(**B**) Cultured post-mortem-derived HAoSMC showed no significant expression differences in unstimulated or TNFα-stimulated (4h or 8h) HAoSMCs. (GG: n=9; AA: n=6).

Figure 4: The rs2107595 risk variant is associated with elevated HDAC9 transcription in reporter assays and genome-edited Jurkat cells

(**A**) Risk allele associated with a significant increase in luciferase activity compared to the common allele in T-lymphoid Jurkat cells and PMA-induced THP-1 MΦ. (**B**) Sanger sequencing of genome-edited Jurkat cells containing either the (*) common allele (G) or risk allele (A). (**C**) rs2107595 risk allele-dependent increase in HDAC9 mRNA expression in genome-edited Jurkat cells. n=16 or 24, mean±SD. T-test. (**D**) Comparative expression analysis during cell cycle progression in isogenic Jurkat cells carrying either the common (G) or risk allele (A). HDAC9 expression levels increased during the first 8 h after HU removal. Risk allele carrying cells showed a significantly increased expression of HDAC9 until 6 h. (mean±SD. T-test).

Figure 5: The rs2107595 region physically interacts with a HDAC9 promoter. (**A**) Domain plot of the 4C-seq results obtained in isogenic Jurkat cells homozygous for the common (G) or risk allele (A). Shown are the significance levels of the 4C-seq signal coverage with viewpoints in the HDAC9 promoter (top) and rs2107595 region (bottom). For both viewpoints results for individual alleles are depicted in the upper panels with difference plots depicted below. Region of interactions (arrows) are defined by an enrichment of covered fragends within a running window of *1 to 50* fragends. Grey boxes represent the location of the 4C viewpoints. DHS and H3K4me3 histone marks are displayed at the top.

Figure 6: E2F3 mediates allele-specific effects of rs2107595 on HDAC9 transcription.

(**A and B**) Comparative ChIP experiments in isogenic Jurkat cells homozygous for the common (G) or risk allele (A). G1/S boundary arrested cells showed an enriched E2F3 (**A**) and Rb1 (**B**) occupancy in common vs risk allele cells at rs2107595. (n=6, mean±SEM, Wilcoxon Rank-Sum Test). (**C**)

553 Overexpression of E2F3a resulted in a significant increase of the ratio between HDAC9 expression in
554 cells homozygous for the rs2107595 common allele (A) vs cells homozygous for the risk allele (G).
555 (n=8-10, mean±SD, T-test). (**D**) Proposed model for the regulatory effect of rs2107595 on HDAC9
556 expression by allele-specific binding of the E2F3/Rb1/TFDP1 complex. In the presence of the
557 common allele (G) the E2F3/Rb1/TFDP1 complex is recruited to the rs2107595 region and mediates a
558 repressive effect on HDAC9 transcription. The risk allele (A) disrupts binding of the
559 E2F3/Rb1/TFDP1 complex thus resulting in elevated HDAC9 expression.