CRISPR perturbations at many coronary artery disease loci impair vascular endothelial cell functions

Florian Wünnemann[1], Thierry Fotsing Tadjo[1], Melissa Beaudoin[1], Simon Lalonde[1], Ken Sin Lo[1], Guillaume Lettre[1,2]

[1] Montreal Heart Institute, Montréal, Québec, Canada

[2] Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada

Correspondence:

Guillaume Lettre

Montreal Heart Institute

5000 Belanger Street

Montreal, Quebec, Canada

H1T 1C8

guillaume.lettre@umontreal.ca

514-376-3330 ext. 2657

ABSTRACT

Genome-wide association studies have identified 161 genetic variants associated with coronary artery disease (CAD), but the causal genes and biological pathways remain unknown at most loci. Here, we used CRISPR knockout, inhibition and activation to target 1998 variants at 83 CAD loci to assess their effect on six vascular endothelial cell phenotypes (E-selectin, ICAM1, VCAM1, nitric oxide, reactive oxygen species, calcium signalling). We identified 42 significant variants located within 26 CAD loci. Detailed characterization of the RNA helicase *DHX38* and CRISPR activation at the *FURIN*/*FES*, *CCDC92*/*ZNF664* and *CNNM2* loci revealed a strong effect on vascular endothelial cell senescence.

INTRODUCTION

Coronary artery disease (CAD) remains the main cause of mortality in the world despite widely available drugs (*e.g.* statins) and the known benefits of simple prevention strategies (*e.g.* exercise). Part of the complexity to prevent and treat CAD resides in our incomplete understanding of atherosclerosis, the pathophysiological process largely responsible for CAD initiation and progression. Atherosclerosis is triggered by many environmental risk factors and other intrinsic stimuli, and results in the dysregulation of vascular wall homeostasis due to the accumulation of cholesterol-rich lipoproteins and a maladaptive inflammatory state[1,2].

Human genetics provide a framework to dissect the biological pathways and cellular networks implicated in atherosclerosis. Genome-wide association studies (GWAS) have already identified 161 loci associated with CAD[3,4]. However, the functional characterization of genes that modulate CAD risk at GWAS loci is labor-intensive. It is further complicated by the fact that most CAD variants are non-coding and are in linkage disequilibrium (LD) with a multitude of other DNA sequence variants.

Half of the CAD GWAS loci do not associate with traditional risk factors. We and others have hypothesized that some of the CAD variants, which are enriched in open chromatin regions found in human vascular endothelial cells, directly modulate endothelial cell functions[5–9]. The functional characterization of two CAD genes in endothelial cells, *PLPP3*[10] and *MIA3/AIDA*[9], has further supported this hypothesis. Vascular endothelial cells have critical roles in atherosclerosis[9,11,12]. Upon activation, they express adhesion molecules necessary for monocyte rolling and attachment (*e.g.* E-selectin, ICAM1, VCAM1) and weakening of their cell-cell junctions can facilitate

monocyte transmigration into the intima. Furthermore, dysfunctional endothelial cells adopt an atheroprone behaviour with changes in calcium (Ca^{2+}) signalling[13], decreased bioavailability of the vasodilator nitric oxide (NO) and increased production of reactive oxygen species (ROS).

The development of pooled CRISPR-based screens now allows perturbation experiments to test most sentinel and LD proxy variants associated with CAD for a role in human vascular endothelial cells[14]. Moreover, by using inhibition (KRAB) or activation (VP64) domains tethered to an inactivated Cas9 (dCas9), it is possible to mimic loss- or gain-of-function effects that might elude perturbations due to classic Cas9 insertion-deletions (indels)[15–17]. Here, we carried out comprehensive pooled CRISPR screens for six endothelial phenotypes relevant to atherosclerosis (presentation of adhesion proteins at the cell membrane (E-selectin, ICAM1 and VCAM1), production of NO and ROS, and intracellular Ca^{2+} concentration) using three different Cas9 perturbation modalities (double-strand break induction (Cas9), inhibition (dCas9-KRAB or CRISPRi) and activation (dCas9-VP64 or CRISPRa)). Our results identified 42 variants that modulate endothelial functions, including a subset that cause endothelial dysfunction by inducing vascular endothelial cell senescence.

RESULTS

FACS-based pooled CRISPR screens for endothelial functions

The design of our sgRNA library is summarized in **Figure 1A**. To target genomic regions associated with CAD, we collected 92 GWAS sentinel variants at 89 CAD-associated loci[5,6,18,19] and retrieved their proxy variants in strong LD (r^2 >0.8 in populations of European ancestry). Using this strategy, we derived a set of 2,893 variants (92 GWAS sentinel and 2,801 LD proxy variants) (**Figure 1A** and **Supplementary Table 1**). For each of these variants, we designed a maximum of five high-quality sgRNAs (**Figure 1B**). The mean distance between sgRNA potential cut-sites and the targeted variants was 22-bp (**Figure 1C**). After quality-control steps, we generated a list of 7,393 sgRNA that targeted 1,998 variants at 83 CAD loci (**Supplementary Table 2**). On average at each CAD locus, our sgRNA library covered 76±22% of the targeted variants (**Figure 1D** and **Supplementary Table 3**). Of the 83 tested loci, we could capture 100% of the targeted variants at 20 CAD loci and ≥80% of variants at 38 loci (**Supplementary Table 3**). The majority of the targeted variants were intronic (70.8%) or intergenic (10.2%) (**Figure 1E**)[20,21].

We utilized lentiviruses to deliver our pooled CRISPR libraries to immortalized human aortic endothelial cells (teloHAEC) that stably express one of three Cas9 variants (Cas9, CRISPRi, CRISPRa) (**Figure 1F**). We treated Cas9 and CRISPRi (but not CRISPRa) infected cells with TNFα in order to find genes that can block (Cas9, CRISPRi) or induce (CRISPRa) a pro-inflammatory response. Then, we labelled cells with fluorescent antibodies against E-selectin, VCAM1, or ICAM1, or with fluorescent dyes for signalling molecules (ROS, NO, Ca^{2+}), and sorted cell populations by flow cytometry (FACS) to collect the bottom and top 10% cells based on fluorescence

intensity (**Figure 1F** and **Supplementary Figures 1-3**). We amplified and sequenced the sgRNAs from the FACS cell fractions to identify sgRNAs that have a significant effect on endothelial functions. Quality-control analyses of sorted cell fractions showed a good representation of sgRNA diversity (mean Gini index=0.076±0.01) and a good read coverage per sgRNA (mean number of aligned reads per sgRNA=1995±2981) (**Supplementary Figure 4**). Analysis of the 10% most variable sgRNAs across all assays revealed clustering of samples along the Cas9 modalities (**Figure 2A**).

Effects of CRISPR knockout, inhibition and activation in teloHAEC

To assess Cas9 efficiency in our experiments, we included in the library 330 sgRNAs against the coding sequence of genes essential for cell viability. For Cas9 and CRISPRi, we found a strong depletion of sgRNAs targeting essential genes among the sequenced FACS cell fractions (Kolmogorov-Smirnov (KS) test $P<2.2\times10^{-16}$ and $P=1.6\times10^{-13}$, respectively) (**Figure 2B**). We also noted a minor but significant shift toward depletion in the sgRNA count distribution of essential genes for the CRISPRa experiments (KS test $P=3.7\times10^{-6}$), potentially due to steric hindrance effects by the dead Cas9 moiety near the transcriptional start site of these genes or the toxic impact of gene over-expression (**Figure 2B**).

As an additional quality-control step in our experiment, we designed sgRNAs against the coding and promoter sequences of *SELE*, *ICAM1* and *VCAM1*, which encode the three adhesion proteins measured in our FACS assays (**Figure 1F**). We observed significant depletion of sgRNAs targeting coding exons and promoter regions of these genes in the top vs. bottom 10% FACS fractions with Cas9 or CRISPRi (**Figures 2C-D**, **Supplementary Figure 5**). In the CRISPRa experiments, the same sgRNAs were

enriched in FACS fractions with high E-selectin, ICAM1 or VCAM1 levels (**Figure 2E**, **Supplementary Figure 5**). The three other endothelial phenotypes measured in our experiments - NO and ROS production, and Ca^{2+} signalling - are physiological readouts that are not the product of a single gene. In the absence of confirmed positive control genes that we could target to validate our system, we carefully calibrated the flow cytometry assays for these readouts using appropriate agonists/inducers (**Methods**). Our screens are sufficiently sensitive to detect sgRNAs targeting CAD loci that have strong effects on these hallmarks of endothelial dysfunction.

Our pooled CRISPR screens identified 51 significant variant-endothelial phenotype results (false discovery rate (FDR) ≤10%) involving 42 different variants located within 26 CAD loci (**Figure 3A** and **Supplementary Table 4**). We found significant results for almost all combinations of Cas9 modality and FACS phenotypes, and most of these results were specific to a single combination (**Figure 3A**). This highlights the importance to test several cellular phenotypes and Cas9 modalities to carry out comprehensive perturbation screens in order to characterize GWAS loci. For 15 CAD loci where we could target all LD proxies with sgRNAs (**Figure 1D** and **Supplementary Table 3**), we detected no significant signals in our CRISPR assays, suggesting that genes within these genomic regions modulate CAD risk through different functions or cell types. When compared with genomic loci with no significant results, CAD loci with at least one significant variant in our CRISPR screens were not better captured by designed sgRNAs (median coverage 73% vs. 78% of LD proxies, Wilcoxon's test P=0.70) but had significantly more LD proxies (median 41 vs. 9 variants, Wilcoxon's test P=1.1x10^{-4}).

Several of the CAD loci identified by GWAS have been implicated in blood lipid metabolism (*e.g. LDLR*, *APOE*, *PCSK9*). Because genetic variation within these loci are likely to influence risk through an effect on lipid levels, we did not anticipate identifying them in our endothelial cell functions CRISPR screens. Of the variants that mapped to 10 lipid loci included in our screens, all were negative across the different endothelial phenotypes tested except rs118039278 located in an intron of *LPA* (CRISPRa for ICAM1, FDR<0.001, **Figure 3A**). Although *LPA* is not expressed in teloHAEC, CRISPRa could induce its ectopic expression and the encoded Lp(a) lipoprotein has been shown to induce endothelial dysfunction[22].

Validation of a CAD-associated regulatory variant at the *FURIN*/*FES* locus

To validate our results, we selected eight SNPs at seven CAD loci and performed individual sgRNA infection and FACS experiments (**Supplementary Table 5**). For this validation step, we prioritized variants that were significant for >1 cellular phenotype and that had strong effect sizes in the CRISPR screens. For each experiment, we compared the distribution of the FACS-based cellular phenotype between control sgRNAs and the best sgRNA targeting each selected CAD variant (**Figure 3B**). Across three independent biological replicates, we could validate six of the eight selected SNPs (one-tailed *t*-test $P<0.05$, **Supplementary Table 5**).

To investigate the global transcriptional consequences of genome editing perturbations and elucidate the underlying impact of these loci on endothelial cell biology, we performed RNA-seq for five sgRNAs that passed primary validation. At the *FURIN/FES* locus, the top sgRNA (sgRNA_06939) maps to rs12906125, a variant in strong LD with the CAD sentinel variant rs2521501 (r^2=0.91). rs12906125 is located

in the *FES* promoter and overlaps an ATAC-seq peak as well as a H3K27ac-defined enhancer that physically interacts with the *FURIN* promoter (**Figure 4A**)[9]. The same SNP is an eQTL for *FES* in human primary aortic endothelial cells[23] and arterial tissues from GTEx. In the CRISPRa experiments, we found a significant up-regulation of both *FES* (log_2(fold-change (FC))=3.75, adjusted P=8.5x10^{-173}) and *FURIN* (log_2FC=0.78, adjusted P=1.5x10^{-10}) (**Figure 4B**).

Recently, a different CAD SNP at the same locus, rs17514846, was shown to be an eQTL for *FURIN* in vascular endothelial cells[24]. This variant was not tested in our CRISPR screens because of weaker LD with the sentinel variant rs2521501 (r^2=0.47). *FURIN*, which encodes a proprotein convertase, represents a strong candidate CAD causal gene at this locus: its specific knockdown in human endothelial cells reduced atheroprone characteristics such as monocyte-endothelial adhesion and transmigration[24]. Consistent with these previous *FURIN*-related results, gene-set enrichment analysis (GSEA) of the *FURIN*/*FES* RNA-seq data highlighted genes implicated in inflammatory responses (interferon, TNFα/TGFβ) and cell cycle regulation (p53, apoptosis) (**Figure 4C** and **Supplementary Table 6**). Although our results are consistent with *FURIN* representing a strong CAD candidate gene, we cannot rule out a role for *FES* given the strong CRISPRa effect (**Figure 4B**). *FES*, which encodes a tyrosine protein kinase that can control cell growth, differentiation and adhesion, has not been implicated in vascular endothelial cell biology.

Loss of *DHX38* function induces vascular endothelial cell senescence

Two of the validated sgRNAs target synonymous variants in *DHX38* (rs2074626, rs2240243) and mediate Cas9 nuclease effects on E-selectin (**Figures 3A-B**) and

198 VCAM1 (as validated by subsequent analyses, **Supplementary Table 5**). We
199 confirmed the E-selectin result using Cas9 ribonucleoprotein complexes
200 (**Supplementary Figure 6**). In our screens, we also tested but found no significant
201 effects for two *DHX38* missense variants (rs1050361, rs1050362). However, in
202 contrast to the *DHX38* synonymous variants, these missense variants are located in
203 early exons (exons 2-3, with transcripts potentially escaping nonsense-mediated
204 mRNA decay) that are expressed at low levels. *DHX38* encodes a RNA helicase
205 involved in splicing (**Figure 5A**).
206
207 In the RNA-seq experiments with a sgRNA targeting *DHX38* (seven days post-
208 infection, TNFα treatment), the gene was not down-regulated and we found few reads
209 with Cas9-mediated indels (<2%), yet a strong gene expression signature suggesting
210 an effect on cell proliferation with the modulation of genes involved in the p53, G2/M
211 checkpoint and E2F target genes pathways (**Figure 5B**). Together, these results
212 suggest a complex scenario: two sgRNAs targeting Cas9 nuclease to *DHX38* exons
213 produce robust cell adhesion phenotypes as well as a cell cycle-related gene
214 expression signature without apparently introducing many indels nor impacting *DHX38*
215 expression levels. To reconcile these observations, we hypothesized that endothelial
216 cells with *DHX38* detrimental indels have a growth disadvantage and induce a
217 response in surrounding cells without *DHX38* indels through paracrine signalling. To
218 test this model, we replaced the antibiotic resistance marker by a fluorescence protein
219 (CRIMSON) in the sgRNA vector so that we can sort and characterize at different
220 timepoints teloHAEC stably expressing Cas9 that have or not received a *DHX38*
221 sgRNA (**Figure 5C**). While the fraction of CRIMSON^{+} cells is similar for safe harbor
222 and *DHX38* sgRNAs two- and four-days post-infection, it is significantly lower after

seven days (**Figure 5D**). This observation is aligned with our pooled CRISPR screens results since cells were also tested for endothelial dysfunction phenotypes seven days post-infection. Although we did not capture many *DHX38* indels in the RNA-seq experiment, we could detect a high frequency of indels (15-40%) in CRIMSON^{+} cells already two days post-infection (**Figure 5E** and **Supplementary Table 7**). Importantly, we also measured a down-regulation of *DHX38* expression levels in CRIMSON^{+} cells, confirming that *DHX38* is likely the gene at this CAD locus that mediates the endothelial phenotypes (**Figure 5F**).

Our hypothesis would be consistent with vascular endothelial cell senescence and the senescence-associated secretory phenotype (SASP)[25,26]. In CRIMSON^{+} cells with *DHX38* sgRNA, we measured an up-regulation of *CDKN1A* (encoding the CDK2 inhibitor p21$^{WAF1/Cip1}$) and detected a higher number of cells with β-galactosidase activity when compared to CRIMSON^{-} cells or CRIMSON^{+} cells with a safe harbor sgRNA (**Figures 5F-G**). These characteristics are hallmarks of cell senescence. Activation of the senescence program is specific to *DHX38* and not a general response to DNA damage induced by this particular sgRNA as four different sgRNAs targeting two different *DHX38* exons impaired endothelial functions in the CRISPR screens. Collectively, our observations point towards the induction of senescence in teloHAEC with dysregulated *DHX38* expression and help explain the results from our RNA-seq experiment. In the cell population profiled by RNA-seq, we found few *DHX38* indels and did not detect *DHX38* down-regulation because cells with *DHX38* detrimental indels underwent senescence-mediated growth arrest. However, these *DHX38*-edited senescent cells secreted pro-inflammatory molecules as part of the SASP which, in

247 combination with the TNFα added to the cell medium, activated a transcriptional
248 response in teloHAEC without *DHX38* edits.
249
250 **A transcriptional signature of senescence triggered by CRISPRa perturbations**
251 **at three CAD loci**
252 To assess the role of vascular endothelial cell senescence in explaining our CRISPR
253 screens results, we re-analyzed our RNA-seq data using senescence and SASP
254 curated gene sets[27]. This analysis did not yield significant enrichments for sgRNAs
255 that target *DHX38* or *MAT2A* (**Figure 6A**). Despite our results that implicate *DHX38* in
256 senescence (**Figure 5**, CRIMSON experiments without TNFα), we did not expect to
257 find a gene expression signature of senescence in the RNA-seq experiment (with
258 TNFα) because too few cells had *DHX38* indels (see above). Additionally, the SASP
259 was masked by the presence of exogenous pro-inflammatory TNFα, which we added
260 to the cell medium to validate the E-selectin phenotype (**Figure 3B**). For *MAT2A*,
261 targeting Cas9 to the synonymous variant rs1078004 increased ROS production in
262 TNFα-treated teloHAEC (**Figure 3C**). *MAT2A* encodes a methionine
263 adenosyltransferase that is responsible for the biosynthesis of S-adenosylmethionine,
264 a precursor of the potent antioxidant glutathione[28]. The analysis of the *MAT2A* RNA-
265 seq experiment (with TNFα) revealed a profile very similar to *DHX38*: we found few
266 *MAT2A* indels, the gene was not differentially expressed, and we did not detect a
267 signature of senescence (**Figure 6A**). Thus, it is possible that *MAT2A* edits decrease
268 *MAT2A* expression, increased oxidative stress and trigger senescence[25,26], but that
269 we would only observe this response by enriching for cells with *MAT2A* sgRNA in
270 medium without exogenous TNFα, as we did for *DHX38* sgRNA (**Figure 5**).
271

Analysis of the RNA-seq data for the three CRISPRa experiments showed a strong senescence signature (**Figure 6A**). While it has recently been reported that CRISPRa can lead to non-specific transcriptional effects such as the up-regulation of *IL6*[29], we used safe harbor sgRNAs to control for such effects and *IL6* was not differentially expressed in our experiments. Additional controls suggested specificity of our CRISPRa results (**Supplementary Figure 7**). Beside the *FURIN*/*FES* locus described above, the two other CRISPRa experiments targeted intronic variants in *ZNF664* (rs12311848) and *CNNM2* (rs78260931). Across these three CAD loci, the CRISPRa experiments shared 734 differentially expressed genes and highlighted biological pathways that are relevant to senescence, such as the down-regulation of cell cycle-related genes (**Supplementary Tables 6** and **8**).

Targeting CRISPRa at rs12311848 did not increase the expression of *ZNF664* but the expression of *CCDC92*, a gene located 82 kb upstream ($log_2FC=0.74$, adjusted $P=9.2x10^{-5}$, **Figure 6B**). The sentinel CAD variant identified by GWAS at this locus is rs11057401, a missense variant in *CCDC92*. We targeted four sgRNAs at rs11057401 but did not detect significant effects in the CRISPR screens. This result suggests that CRISPRa gain-of-function experiments are necessary to detect the impact of *CCDC92* on endothelial dysfunction and senescence. *CCDC92* has been found to be over-expressed in senescent cells[30]; accordingly, we detected increased *CCDC92* expression in the *FURIN*/*FES* (sgRNA_06939, $log_2FC=0.6$, adjusted $P=3.3x10^{-5}$) and *CNNM2* (sgRNA_25344, $log_2FC=1.9$, adjusted $P=1.8x10^{-42}$) CRISPRa experiments. *CCDC92* over-expression is specific to these sgRNAs as a CRISPRa experiment with a "silent" sgRNA targeting the *LPL* locus did not induce its expression (**Supplementary Figure 7B**). Finally, although we measured a robust transcriptional

297 signature of senescence when targeting CRISPRa at the *CNNM2*-rs78260931 locus
298 (**Figure 6A**), we found no evidence of differential expression for nearby genes (in *cis*,
299 the closest differentially expressed gene was *NFKB2* located 568 kb away
300 (log_2FC=0.33, adjusted *P*=0.008)). We also manually inspected the sequence reads
301 that mapped to the *CNNM2* region but did not find un-annotated genes that were
302 differentially expressed. Thus, based on our results, we cannot prioritize a candidate
303 causal gene at this CAD locus, but emphasize that the effect is specific to this region
304 and not a sgRNA-specifc artifact because three of the four sgRNAs that we targeted
305 at rs78260931 gave consistent results in the CRISPRa-ICAM1 screen.

DISCUSSION

As for most complex human diseases, many GWAS loci associated with CAD do not include obvious candidate causal genes nor implicate known pathophysiological mechanisms. To elucidate their mechanisms and gain insights into atherosclerosis, we carried out multiple CRISPR screens to test if CAD variants impact vascular endothelial functions. By combining six different endothelial cell readouts and three Cas9 modalities, we identified 42 variants at 26 CAD loci (**Figure 3A**). This list is depleted of variants that modulate CAD risk through an effect on lipid metabolism and enriched for loci of unknown functions (**Supplementary Table 3**). We found variants near *ARHGEF26* and *ADAMTS7*, genes previously implicated in leukocyte transendothelial migration[5] and endothelial cell angiogenesis[31], respectively. We also retrieved rs17163363, an intronic variant in *MIA3* that controls the expression of *AIDA* in endothelial cells[9].

There were also variants and genes that we expected to find but did not recover. For instance, we did not identify rs17114036, a likely functional variant that controls the expression of *PLPP3* in endothelial cells, although this negative result may be because the underlying enhancer requires hemodynamic stress to be active[10]. Furthermore, our screens did not yield variants at CAD loci that include *PECAM1* (adhesion protein CD31) and *NOS3* (endothelial NO synthase), two genes with important roles in endothelial cells. As for *PLPP3*, it might be that we did not activate endothelial cells with the right stimulus to detect the functional impact of these variants/genes in our assays. It is also possible that some loci will require the precise engineering of alleles (*e.g.* using base editing) to detect a cellular phenotype, or that the phenotypic effect

of a variant at the cellular level is too low to distinguish a true signal from the experimental noise inherent to any large-scale omics approach.

We designed our sgRNA library using a variant-focused approach. Our rationale was that the identification of causal variants by CRISPR perturbations would lead us to the causal genes and biological pathways. This strategy worked at a few CAD loci (*e.g. MIA3*/*AIDA*[9]). At the *FURIN*/*FES* locus, we found a candidate regulatory variant that was also prioritized using orthogonal methodologies (**Figure 4**)[23]. However, it is also likely that some of the findings from our CRISPR screens result from loss- or gain-of-function effects on causal genes independently of the causal variants. For instance, we identified and validated sgRNAs near synonymous variants in *DHX38* and *MAT2A* using the Cas9 nuclease. While synonymous variants can have phenotypic consequences, it is more likely that these variants are in LD with the causal variants but were captured in our screens because they targeted loss-of-function indels to the *DHX38* and *MAT2A* coding sequences. Similarly, ectopic activation or inhibition of gene expression by CRISPRa and CRISPRi can highlight potential nearby causal genes (*e.g. CCDC92*) even if the sgRNAs do not directly overlap causal variants.

Vascular endothelial cell senescence emerged as an important response in our CRISPR perturbation screens. Senescent endothelial cells are characterized by growth arrest, but also a pro-inflammatory and atheroprone phenotype that involves increased production of adhesion molecules and ROS, and reduced NO bioavailability[32]. For *DHX38* with Cas9, we had to design a TNFα-free CRIMSON-based strategy to confirm senescence (**Figure 5**). Conversely, the RNA-seq results for the CRISPRa experiments (without TNFα) at *FURIN*/*FES*, *CCDC92*/*ZNF664* and

CNNM2 were less ambiguous, with a clear gene expression signature indicating senescence (**Figure 6A**). Our controls do not support the hypothesis that this difference is due to non-specific effects of Cas9 or CRISPRa (**Supplementary Figure 7**). Instead, the presence of TNFα, which can induce premature endothelial cell senescence[33], has led to a stronger growth arrest phenotype in the presence of Cas9-mediated indels in *DHX38* (and maybe also *MAT2A*).

While *FURIN* has been shown to influence endothelial functions[34], most other genes (*FES*, *DHX38*, *MAT2A*, and *CCDC92*) have been less studied. Supportive of their role *in vivo*, we re-analyzed single-cell RNA-seq data from human atherosclerotic right coronary arteries and confirm that these genes are expressed in human vascular endothelial cells (**Figure 6C**). Focusing specifically on *CCDC92*, the same CAD variant is also associated with waist-to-hip ratio, and knockdown of *CCDC92* expression in immortalized mouse OP9-K stromal cells impaired adipogenesis[35]. This result prompted the authors to propose a role for *CCDC92* in adipocyte differentiation and insulin secretion, although the authors did not exclude senescence as a possible mechanism. In a different study, *CCDC92* over-expression reduced Ebola viral infections in HEK293 and endothelial HUVEC by blocking the virus without killing the host cells[36]. It was not investigated, but senescence has been proposed as an antiviral strategy[37]. Given that *CCDC92* was up-regulated both in *cis* and *trans* in our CRISPRa experiments, it suggests that it is a key gene located in a CAD GWAS locus that can impact normal endothelial functions.

Endothelial cell senescence is both a physiological and pathological process[38]. In health, it signals the system for vascular endothelium repair. Senescence also

380 increases with age and in response to traditional CAD risk factors. When it overcomes
381 the regeneration capacity of the system or upon stress, senescence causes
382 endothelial dysfunction and can lead to vascular diseases. Senescent cells
383 accumulate at the sites of atherosclerosis in human blood vessels[39,40] and their
384 selective elimination using transgenic strategies or drugs (senolytics) delays
385 atherogenesis progression in mice[41]. Our data suggest that a subset of variants
386 associated with CAD in humans affect key endothelial functions, potentially by
387 inducing premature senescence. This observation links a large body of literature that
388 has implicated senescence in atherosclerosis with a genetic program that modulates
389 endothelial functions. As clinical trials to test the efficacy of senolytics on vascular
390 diseases are now in discussion[42], it will be important to explore whether specific CAD
391 variants or polygenic scores are predictive of their clinical response.
392

ONLINE METHODS

Design of the sgRNA library

We retrieved 92 sentinel genetic variants associated with coronary artery disease (CAD) at genome-wide significant levels (P-value ≤$5x10^{-8}$) from four GWAS meta-analyses available at the time of the design of this experiment[5,6,18,19]. For the design of the sgRNA library, we included all sentinel variants as well as variants in strong LD (r^2 >0.8 in the 1000 Genomes Project European-ancestry populations). For each variant - sentinel and LD proxy - we identified all possible sgRNA in a 100-bp window centered on the variant itself. We prioritized sgRNA with the highest predicted quality using the CRISPR OffTarget Tool (version 2.0.3)[43] with a Targeting_guide_score ≥ 20 and the "matches with 0 mismatches" = 1 and "matches with 1 mismatch" = 0 settings. We discarded sgRNA that overlapped heterozygous variants, indels and/or multi-allelic variants in the teloHAEC genome (build hg19). We selected sgRNA targeting essential genes from a previously published study[44]. For potential positive control genes (*SELE*, *SELP*, *ICAM1*, *VCAM1*, *PECAM1*, *NOS3*, *VWF*, *SOD2*, *SOD3*, *GPX3*, *CAT*, *ITPR1*, *ITPR2*, *ITPR3*, *ATP2A2*, *ATP2A3*, *PLN, CAV1*, and *TRPV4*), we selected sgRNA from the Human GeCKOv2 CRISPR knockout pooled library[45]. We also selected sgRNA that targeted the promoter (300-bp window before the transcriptional start site) of positive control genes for the CRISPRa (dCas9-VP64) experiments. For all selected loci (variants, coding sequences, gene promoters), we retained the five top scoring sgRNA for the library design. Finally, we added two sgRNA for the *SELE* locus (SELE_g1, SELE_g2) that we frequently use to validate TNFα stimulation. This resulted in a final library of 8051 sgRNA (**Supplementary Table 2**).

418 The sgRNA were synthesized in duplicates by Agilent Technologies (Cat-#: G7555B)
419 to accommodate the specific requirements of the Cas9/dCas9-KRAB and dCas9-
420 VP64 (specific MS2 tracrRNA) experiments. We amplified each specific pool of
421 oligonucleotides as previously described[16], with the following small modifications: we
422 performed two PCR using NebNext High fidelity Master mix (Cat-#: M0541L). The first
423 PCR was used to amplify each pool separately using 2.5ng of pooled oligonucleotides
424 and 500nM of each primer (for the Cas9/dCas9-KRAB library, we used
425 U6_subpool_fwd and Guide_CM_barcode1_rev; for the dCas9-VP64 library, we used
426 U6_subpool_fwd and Guide_MS2_Barcode2_rev). Cycling conditions for PCR1 were
427 98℃ for 30 sec, then 15 cycles of 98℃ for 10 sec; 55℃ for 10 sec; 72℃ for 15 sec and
428 a final step of 72℃ for 2 min and a 10℃ hold. We performed the second PCR to add
429 homologous sequences, using the U6_screen_fwd and Tracr_rev oligonucleotides for
430 the Cas9/dCas9-KRAB library, and the U6_screen_fwd and Tracr_MS2_rev
431 oligonucleotides for the dCas9-VP64 library, in both cases using ⅕ of PCR1 as
432 template. Cycling conditions for PCR2 were 98℃ for 30 sec, then 10 cycles of 98℃ for
433 10 sec; 55℃ for 10 sec; 72℃ for 15 sec and a final cycle of 72℃ for 2 min and 10℃
434 hold. See table **Supplementary Table 9** for primer details.
435
436 After gel extraction and PCR purification, we performed Gibson assembly in both
437 respective vectors (pHKO9-Neo and lentisgRNA(MS2)-zeo backbone addgene
438 61427). For pHKO9-Neo, we replaced the Crimson fluorescent gene in the pHKO9-
439 Crimson-CM vector (gift from Dan Bauer's lab) by a neomycin resistance (NeoR)
440 sequence. Briefly, we amplified the NeoR gene from our pCas9-Neo vector[46] using
441 BsiWI-Neo-Fwd and MluI-Neo_rev primer (**Supplementary Table 9**). After digestion
442 by BsiWI and MluI, we cloned the segment in pHKO9-Crimson_CM, which had been

443 digested with BsiWI and MluI. We amplified each library using ten independent maxi-
444 preparations (Macherey-Nagel cat# 740424). To control the quality of both libraries,
445 we sequenced them on an Illumina HiSeq4000 instrument and calculated the Gini
446 index, which summarizes read distribution across sgRNA in a given pool. For a good-
447 quality sgRNA library, the expected Gini index is ≤0.2, and we obtained Gini indexes
448 of 0.050 and 0.052 for the Cas9/dCas9-KRAB and dCas9-VP64 library, respectively.
449
450 **Engineering of teloHAEC cell lines to stably express Cas9 variants**
451 TeloHAEC are immortalized human aortic endothelial cells obtained by over-
452 expressing telomerase (ATCC CRL-4052). These cells have a normal female
453 karyotype (46;XX) and exhibit many of the properties and functions of human vascular
454 endothelial cells[9]. We generated our teloHAEC cells models expressing either Cas9,
455 dCas9-KRAB or dCas9-VP64 + MPHv2 using Addgene vectors #52962, #46911,
456 #61425 and #89308, and lentiviral infection as previously described[46].
457
458 **Pooled CRISPR screen experiments**
459 We produced four batches of lentiviruses for each sgRNA library pool (Cas9/dCas9-
460 KRAB, dCas9-VP64). We infected each teloHAEC cell line (Cas9, dCas9-KRAB,
461 dCas9-VP64) at a multiplicity of infection of 0.3 using each batch of viruses separately.
462 Following viral infection, we selected cells using zeocin (teloHAEC-dCas9-VP64) or
463 G418 (teloHAEC-Cas9/-dCas9-KRAB) for five (teloHAEC-dCas9-VP64) or seven
464 days (teloHAEC-Cas9/-dCas9-KRAB) in vascular cell basal medium (ATCC PCS-100-
465 030) to remove any cells that did not incorporate a vector. After selection, we
466 stimulated cells expressing Cas9 or dCas9-KRAB using TNFα (10ng/µl) for four hours
467 to induce a pro-inflammatory response; we did not stimulate cells expressing dCas9-

VP64, reasoning that the VP64 transcriptional domain should activate gene expression. Following TNFα stimulation, we immunostained cells (around 50M cells) with antibodies linked to phycoerythrin for adhesion molecules (E-selectin (BD BIOSCIENCES Cat-#: 551145), VCAM-1 (Cat-#: 12-1069-42), ICAM-1 (Cat-#: 12-0549-42)) or we incubated with fluorescent dye-based reagents for endothelial signaling markers: (nitric oxide (NO) (DAF-FM Diacetate, Cat-#: D23844)), reactive oxygen species (ROS) (CM-H2DCFDA, Cat-#: C6827), calcium signaling (Fura Red, Cat-#: F3021)). We calibrated the FACS assays with positive control treatments to make sure that we could robustly detect changes in the measured phenotypes. Antibodies and fluorescent dye-based reagents were titrated to use optimal concentrations. We also quantified how teloHAEC were responding to ionomycin for calcium signaling, to sodium nitroprusside for NO and to TNFα for reactive oxygen species. For adhesion molecules, we utilized sgRNA targeting coding exons and promoter regions of *SELE*, *ICAM1* and *VCAM1* as positive controls. Unless otherwise stated, we purchased all antibodies and dyes from ThermoFisher Scientific. Subsequently, we sorted stained cells by flow cytometry on a BD FACSARIA FUSION flow cytometer to collect the top and bottom 10% of fluorescently labeled cells. FACS traces were generated with FlowJo (BD Biosciences). We extracted genomic DNA from both top and bottom 10% cell fractions separately (around 5M cells in each fraction) using the QIAGEN DNeasy Blood and Tissue kit (Cat No. 69504) according to manufacturer's instructions.

Amplification and sequencing of pooled CRISPR experiments

We amplified sgRNA sequences from genomic DNA via PCR, followed by a cleanup step using the QIAGEN QIAquick PCR purification kit (Cat-#: 28104) according to the

493 manufacturer's instructions. We used the primer sequences and PCR settings as
494 previously described in ref. [16]. We created sequencing libraries using Illumina TruSeq
495 adapters according to the manufacturer's protocols. We sequenced the libraries on an
496 Illumina Hiseq4000 instrument at the McGill University and Genome Quebec
497 Innovation Centre (MUGQIC). Generally 6 samples were multiplexed per sequencing
498 lane for a target read coverage of ~500 reads per sgRNA per sample
499 (**Supplementary Figure 4B**).
500
501 **Computational analysis of pooled CRISPR screen data**
502 We processed raw sequencing data from the BCL to the FASTQ format using
503 bcl2fastq at MUGQIC. Raw FASTQ reads were quality-controlled using FastQC
504 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC[47]. We
505 performed downstream analysis of sgRNA sequencing data using MAGECK
506 (v.0.5.9)[48]. We quantified sgRNA sequences using MAGECK count against the list of
507 sgRNA sequences in the library (**Supplementary Table 3**), allowing for no
508 mismatches in the sgRNA sequence. We then tested the difference in sgRNA counts
509 between cell fractions or conditions using MAGECK maximum likelihood estimation
510 (*mle*) method with median normalization[49].
511
512 **UMAP representation of sample-level count data**
513 We normalized raw sgRNA counts using variance stabilizing transformation (vst) in
514 DESeq2[50]. To account for baseline differences between plasmid preparations, we
515 further normalized samples to their respective vector library by dividing the vst
516 normalized sgRNA count by the vst normalized count of the Cas9/dCas9-KRAB or
517 dCas9/VP64 library, respectively. We calculated principal components using the top

10% most variable sgRNA (805 sgRNA) across all cell sorted samples based on normalized counts. Next, we used the loadings from the first three principal components in Uniform Manifold Approximation and Projection (UMAP)[51] to create a two-dimensional embedding of the normalized sgRNA count data. Each dot in the UMAP plot represents one sequenced sample (top or bottom 10% of stained cells).

Analysis of sgRNAs targeting essential genes

To test for potential effects of sgRNA on endothelial cell death and proliferation, we compared sgRNA counts of all samples across the same cellular model (Cas9, dCas9-KRAB, dCas9-VP64) against the respective baseline vector library sgRNA count using MAGECK *mle*. We used sgRNAs targeting essential genes in the teloHAEC Cas9 cellular model as positive controls.

Single sgRNA validation

We individually cloned each sgRNA for validation as previously described[52]. We produced lentiviruses, infected cells, performed antibiotic selection, and stained cells as for the pooled CRISPR screen. We analyzed cells using flow cytometry (BD FACSCelesta (BD Biosciences, San Jose, CA, USA) equipped with a 20 mW blue laser (488 nm), a 40 mW red laser (640 nm), and a 50 mW violet laser (405 nm). For each experiment, we measured the mean fluorescent intensity (MFI) obtained for sgRNA of interest and compared it with the MFI for control sgRNA (safe-harbor and/or scrambled sgRNA). We performed each experiment at least three times. For statistical analyses, we used Student's *t*-test and determined that a sgRNA had a significant effect on the measured phenotype when a one-tailed P-value ≤0.05.

For *DHX38* Crimson experiments, we individually cloned each sgRNA in pHKO9-Crimson-CM vector (gift from Dan's Bauer lab). We produced lentiviruses, infected cells and performed flow cytometry on a BD FACSARIA FUSION flow cytometer at day 2, day 4 and day 7 post-infection. We analyzed the percentage of Crimson positive cells and we sorted Crimson positive and negative cells to extract RNA in each fraction. We extracted total RNA using RNeasy Plus Mini Kit (Qiagen cat #: 74136). We measured RNA integrity and concentration using Agilent RNA 6000 Nano II assays (Agilent Technologies) on an Agilent 2100 Bioanalyzer and Take3 on Cytation V (Biotek). We reverse transcribed 750ng of total RNA using random primers and 1 U of the MultiScribe Reverse Transcriptase (Applied Biosystems) in a 20 µL reaction volume at 100 mM dNTPS and 20 U of RNase inhibitor with these three steps: 10 min at 25 °C, 120 min at 37 °C and 5 min at 85 °C. We followed the MIQE guidelines to assess quality and reproducibility of our qPCR results [53]. We performed qPCR in triplicates for all samples using: 1.25 µL of cDNA (1/50 dilution), 5 µL of Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies) and 3.75 µL of primer pair mix at 0.8 µM on a CFX384 from Biorad. We used the following thermal profile: 10 min at 95 °C, and 40 cycles of 30 s at 95 °C, 30 s at 55 °C and 45 s at 72 °C. We carried out melting curve analyses after the amplification process to ensure the specificity of the amplified products. We also simultaneously performed qPCR reactions with no template controls for each gene to test the absence of non-specific products. Cq values were determined with the CFX Manager 3.1 (Bio-Rad) software and expression levels were normalized on the expression levels of the house-keeping genes TATA-box binding protein (*TBP*), hypoxanthine-guanine phosphoribosyltransferase (*HPRT*), and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) using the ΔΔCt method. The primer sequences are in **Supplementary Table 8**.

568

569 **PCR for determination of CRISPR-Cas9-induced indels**

570 We isolated gDNA using QuickExtract DNA Extraction Solution (Epicentre, QE0905)
571 from 1x10^5 cells. We used 100 or 200 ng of gDNA as a template for PCR reaction with
572 the corresponding primers (see **Supplementary Table 9**). gDNA from parental
573 teloHAEC cells was used as control. Obtained PCR products were analysed by
574 electrophoresis on a 1% agarose gel prior to Sanger sequencing. We used TIDE
575 (tracking of indels deconvolution) software for analysis[54].

576

577 **Assays for assessment of cell senescence**

578 Using the same experimental design (DHX38-Crimson), we performed beta-
579 galactosidase staining using the CellEvent Senescence Green Flow Cytometry Assay
580 Kit from Invitrogen at day 2, day 4 and day 7 following the manufacturer's protocol.
581 Briefly, we trypsinized and we fixed the cells with 2% paraformaldehyde solution during
582 10 minutes at room temperature, washed them in 1%BSA/PBS and incubated for 1h30
583 in 1/500 working solution. After incubation, we washed the cells with 1%BSA/PBS and
584 analyzed them by flow cytometry. We measured β-galactosidase fluorescence signal
585 in positive and negative Crimson cells independently. As positive control, non-infected
586 cells were treated with 20 µM of Etoposide (Sigma, E1383-25) for 2, 4 and 7 days.

587

588 **Transcriptome data analysis**

589 For RNA-seq analysis, we extracted RNA using RNeasy plus mini kit from Qiagen (cat
590 #: 74136). RNA-seq experiments were carried out by the Centre d'Expertise et de
591 Services Genome Quebec using rRNA-depleted TruSeq stranded (HMR) libraries
592 (Illumina) on an Illumina Hiseq 4000 instrument (paired-ends, 100-bp reads) and by

The Center for applied Genomics (Toronto) using rRNA-deletion library prep on an Illumina NovaSeq-SP flow cell. We quality-controlled raw fastq files with FastQC and multiQC[47]. We used kallisto (v. 0.46.0) to quantify transcript abundances[55] against ENSEMBL reference transcripts (release 94) followed by tximport to calculate gene-level counts[56]. We utilized regularized log-transformation (rlog) in DESeq2[50] as input for principal component analysis (PCA). DESeq2[50] was further used to identify differentially-expressed genes between teloHAEC cell models (Cas9, dCas9-VP64) infected by lentiviruses with safe-harbor sgRNA or sgRNA identified in the pooled CRISPR screens. We excluded genes expressed with less than 10 reads across all samples from the analysis. We performed shrinkage for effect size estimates using apeglm using the lfcShrink method[57]. Genes differentially expressed with a Benjamini-Hochberg adjusted p-value ≤ 0.05 were considered significant. Gene set enrichment analysis was performed using the R package fgsea using 100,000 permutations against the Hallmark gene sets from msigdbr (https://igordot.github.io/msigdbr/)[58,59]. We quantified short indels in the RNA-seq data of *DHX38* (sgRNA_10966) and *MAT2A* (sgRNA_02249) using the tools transIndel and Genesis-Indel, which are specifically designed to identify indels in the unmapped read fraction of samples[60,61].

Analysis of scRNA-seq data from human coronary arteries

Single-cell gene expression matrix from human right atherosclerotic coronary arteries (three male and one female donors), was downloaded from NCBI GEO (GSE131780, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131780). The data was re-analyzed using the Seurat package in R with a standard single-cell clustering pipeline. Gene expression data was normalized using the SCTransform function from Seurat (v.3.2.3), regressing out the percentage of mitochondrial gene expression. Principal

618 components analysis was performed, followed by dimensional reduction with Uniform
619 Manifold Approximation and Projection (UMAP) using the first 20 principal components
620 as input. Gene expression was visualized on the first two UMAP dimensions using the
621 kernel density function (plot_density) from the Nebulosa package (v.0.99.92)[62] for
622 endothelial cell marker and candidate genes.
623
624 **Statistics and data analysis**
625 Unless noted otherwise, we performed all data and statistical analyses in R (v.3.6.0)
626 using Rstudio. We ran our analyses on a high performance computing cluster (Beluga)
627 from Calcul Quebec/Compute Canada. For MAGECK variant-level analyses,
628 permutation-based FDR of ≤10% were considered significant. For RNA-seq analysis,
629 genes with a Benjamini-Hochberg adjusted P-value in DESeq2 ≤0.05 were considered
630 significant[50].
631

Funding statements

Florian Wünnemann is supported by a postdoctoral fellowship from the Fonds de Recherche Santé - Québec (FRQS). This work was funded by the Canadian Institutes of Health Research (MOP #136979), the Heart and Stroke Foundation of Canada (Grant #G-18-0021604), the Canada Research Chair Program, the Fondation Joseph C. Edwards and the Montreal Heart Institute Foundation.

Conflict of Interest

The authors declare no conflict of interests.

Acknowledgements

This research was enabled in part by support provided by Calcul Quebec (https://www.calculquebec.ca/en/) and Compute Canada (www.computecanada.ca). We thank John D. Rioux and Catherine Martel for flow cytometry and FACS analysis. We thank Éric Thorin, Mike Sapieha, Dan Bauer and Luca Pinello for providing comments on an earlier version of this manuscript. Finally, we thank Génome Québec and the Center for Applied Genomics (TCAG) in Toronto for performing next-generation DNA sequencing for this project.

Data availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[63] and are accessible through GEO Series accession number GSE165925 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165925).

REFERENCES

1. Hopkins, P. N. Molecular biology of atherosclerosis. *Physiol. Rev.* **93**, 1317–1542 (2013).
2. Libby, P., Ridker, P. M. & Hansson, G. K. Progress and challenges in translating the biology of atherosclerosis. *Nature* **473**, 317–325 (2011).
3. van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **122**, 433–443 (2018).
4. Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease. *Cell* **177**, 132–145 (2019).
5. Klarin, D. *et al.* Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat. Genet.* **49**, 1392–1397 (2017).
6. Howson, J. M. M. *et al.* Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* **49**, 1113–1119 (2017).
7. Codina-Fauteux, V.-A., Beaudoin, M., Lalonde, S., Lo, K. S. & Lettre, G. PHACTR1 splicing isoforms and eQTLs in atherosclerosis-relevant human cells. *BMC Med. Genet.* **19**, 97 (2018).
8. Wang Xiao & Musunuru Kiran. Confirmation of Causal rs9349379-PHACTR1 Expression Quantitative Trait Locus in Human-Induced Pluripotent Stem Cell Endothelial Cells. *Circulation: Genomic and Precision Medicine* **11**, e002327 (2018).
9. Lalonde, S. *et al.* Integrative analysis of vascular endothelial cell genomic features identifies AIDA as a coronary artery disease candidate gene. *Genome Biol.* **20**, 133 (2019).

10. Krause, M. D. *et al.* Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11349–E11358 (2018).
11. Cahill, P. A. & Redmond, E. M. Vascular endothelium - Gatekeeper of vessel health. *Atherosclerosis* **248**, 97–109 (2016).
12. Godo, S. & Shimokawa, H. Endothelial Functions. *Arterioscler. Thromb. Vasc. Biol.* **37**, e108–e114 (2017).
13. Sandoval, R. *et al.* Ca(2+) signalling and PKCalpha activate increased endothelial permeability by disassembly of VE-cadherin junctions. *J. Physiol.* **533**, 433–445 (2001).
14. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
15. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
16. Joung, J. *et al.* Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
17. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
18. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
19. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
20. Maurano, M. T. *et al.* Systematic localization of common disease-associated

706 variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
707 21. Vierstra, J. *et al.* Global reference mapping of human transcription factor
708 footprints. *Nature* **583**, 729–736 (2020).
709 22. Ioka, T. *et al.* Association between plasma lipoprotein(a) and endothelial
710 dysfunction in normocholesterolemic and non-diabetic patients with
711 angiographically normal coronary arteries. *Circ. J.* **66**, 267–271 (2002).
712 23. Stolze, L. K. *et al.* Systems Genetics in Human Endothelial Cells Identifies Non-
713 coding Variants Modifying Enhancers, Expression, and Complex Disease Traits.
714 *Am. J. Hum. Genet.* (2020) doi:10.1016/j.ajhg.2020.04.008.
715 24. Yang, X. *et al.* FURIN Expression in Vascular Endothelial Cells Is Modulated by
716 a Coronary Artery Disease–Associated Genetic Variant and Influences Monocyte
717 Transendothelial Migration. *Journal of the American Heart Association* vol. 9
718 (2020).
719 25. Acosta, J. C. *et al.* A complex secretory program orchestrated by the
720 inflammasome controls paracrine senescence. *Nat. Cell Biol.* **15**, 978–990
721 (2013).
722 26. Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–
723 827 (2019).
724 27. Binet, F. *et al.* Neutrophil extracellular traps target senescent vasculature for
725 tissue remodeling in retinopathy. *Science* **369**, (2020).
726 28. Tchantchou, F., Graves, M., Falcone, D. & Shea, T. B. S-adenosylmethionine
727 mediates glutathione efficacy by increasing glutathione S-transferase activity:
728 implications for S-adenosyl methionine as a neuroprotective dietary supplement.
729 *J. Alzheimers. Dis.* **14**, 323–328 (2008).
730 29. Soubeyrand, S., Lau, P., Peters, V. & McPherson, R. Off-target effects of

CRISPRa on interleukin-6 expression. *PLoS One* **14**, e0224113 (2019).

30. Limbad, C. *et al.* Astrocyte senescence promotes glutamate toxicity in cortical neurons. *PLoS One* **15**, e0227887 (2020).
31. Pu, X. *et al.* Effect of a coronary-heart-disease-associated variant of ADAMTS7 on endothelial cell angiogenesis. *Atherosclerosis* **296**, 11–17 (2020).
32. Gimbrone, M. A., Jr & García-Cardeña, G. Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. *Circ. Res.* **118**, 620–636 (2016).
33. Khan, S. Y. *et al.* Premature senescence of endothelial cells upon chronic exposure to TNFα can be prevented by N-acetyl cysteine and plumericin. *Sci. Rep.* **7**, 39501 (2017).
34. Kim, W. *et al.* Loss of endothelial furin leads to cardiac malformation and early postnatal death. *Mol. Cell. Biol.* **32**, 3382–3391 (2012).
35. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
36. Kuroda, M. *et al.* Identification of interferon-stimulated genes that attenuate Ebola virus infection. *Nat. Commun.* **11**, 2953 (2020).
37. Baz-Martínez, M. *et al.* Cell senescence is an antiviral defense mechanism. *Sci. Rep.* **6**, 37007 (2016).
38. Goligorsky, M. S., Chen, J. & Patschan, S. Stress-induced premature senescence of endothelial cells: a perilous state between recovery and point of no return. *Curr. Opin. Hematol.* **16**, 215–219 (2009).
39. Vasile, E., Tomita, Y., Brown, L. F., Kocher, O. & Dvorak, H. F. Differential expression of thymosin beta-10 by early passage and senescent vascular endothelium is modulated by VPF/VEGF: evidence for senescent endothelial cells

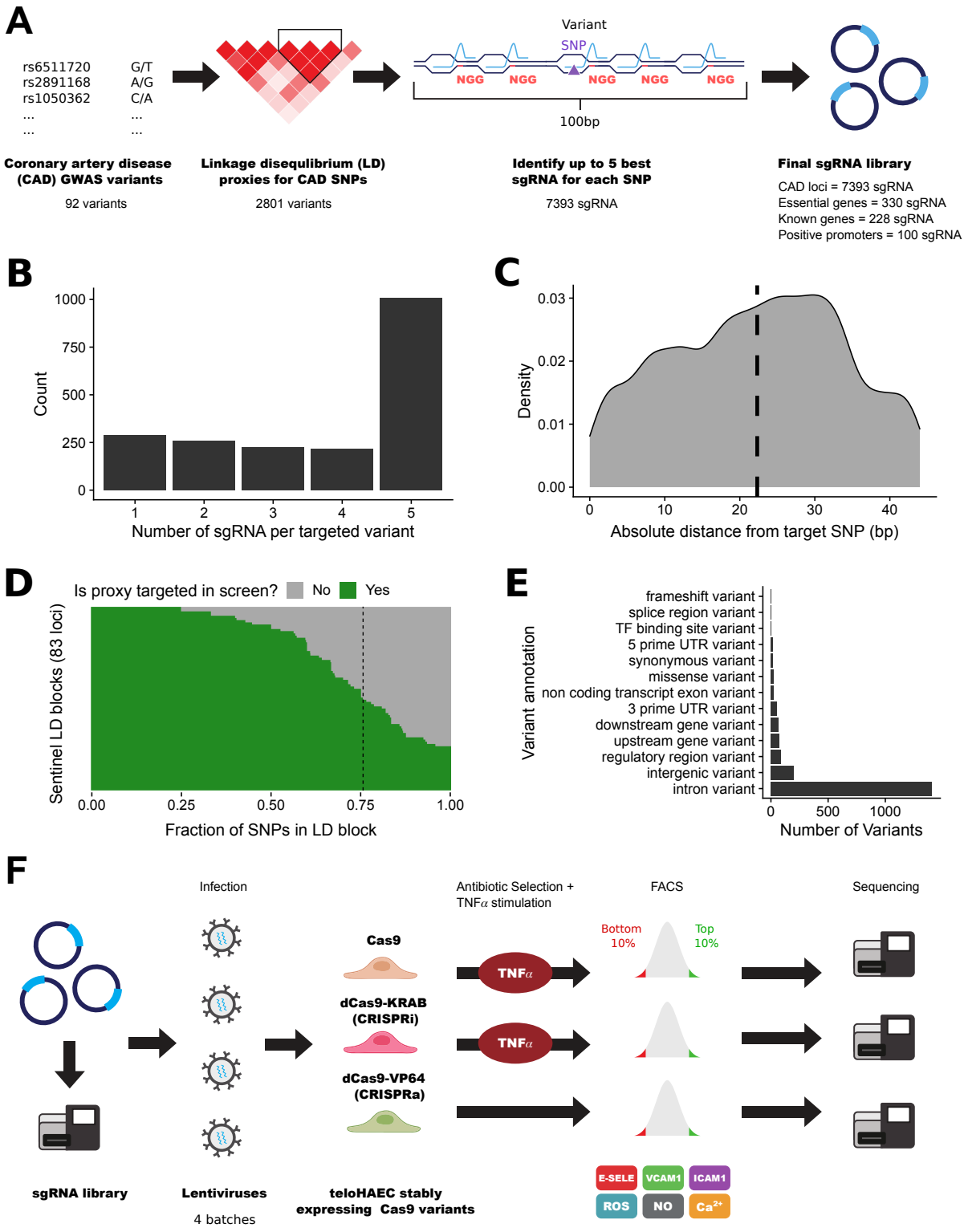
in vivo at sites of atherosclerosis. *FASEB J.* **15**, 458–466 (2001).

40. Minamino Tohru *et al.* Endothelial Cell Senescence in Human Atherosclerosis. *Circulation* **105**, 1541–1544 (2002).
41. Childs, B. G. *et al.* Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science* **354**, 472–477 (2016).
42. Regnault, V., Challande, P., Pinet, F., Li, Z. & Lacolley, P. Cell senescence: basic mechanisms and the need for computational networks in vascular ageing. *Cardiovasc. Res.* (2020) doi:10.1093/cvr/cvaa318.
43. Lessard, S. *et al.* Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11257–E11266 (2017).
44. Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733 (2014).
45. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
46. Lalonde, S. *et al.* Frameshift indels introduced by genome editing can lead to in-frame exon skipping. *PLoS One* **12**, e0178700 (2017).
47. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
48. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
49. Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* **16**, 281 (2015).

781 50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
782 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
783 51. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and
784 Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).
785 52. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human
786 cells. *Science* **343**, 84–87 (2014).
787 53. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of
788 quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
789 54. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative
790 assessment of genome editing by sequence trace decomposition. *Nucleic Acids*
791 *Res.* **42**, e168 (2014).
792 55. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic
793 RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
794 56. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq:
795 transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521
796 (2015).
797 57. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence
798 count data: removing the noise and preserving large differences. *Bioinformatics*
799 **35**, 2084–2092 (2019).
800 58. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene
801 set collection. *Cell Syst* **1**, 417–425 (2015).
802 59. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for
803 comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
804 60. Yang, R., Van Etten, J. L. & Dehm, S. M. Indel detection from DNA and RNA
805 sequencing data with transIndel. *BMC Genomics* **19**, 270 (2018).

806 61. Hasan, M. S., Wu, X. & Zhang, L. Uncovering missed indels by leveraging
807 unmapped reads. *Sci. Rep.* **9**, 11093 (2019).
808 62. Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single cell gene
809 expression signals by kernel density estimation. Bioinformatics (2021).
810 63. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene
811 expression and hybridization array data repository. Nucleic Acids Res. 30, 207–
812 210 (2002).
813 64. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from
814 thousands of CRISPR perturbations. Nat. Genet. 51, 1664–1669 (2019).
815 65. Wirka, R. C. et al. Atheroprotective roles of smooth muscle cell phenotypic
816 modulation and the TCF21 disease gene as revealed by single-cell analysis. Nat.
817 Med. 25, 1280–1289 (2019).

818 **FIGURES**



819

820 **Figure 1.** Pooled CRISPR screens to identify CAD variants and genes that modulate

821 vascular endothelial functions.

(**A**) From 92 loci associated with coronary artery disease (CAD) risk by genome-wide association studies (GWAS), we identified 2893 sentinel and linkage disequilibrium proxy variants for testing. For each of these variants, we attempted to design a maximum of five high-quality guide RNAs (sgRNAs) within a 100-bp window. In the design of the library, we also included sgRNAs that target genes essential for cell viability, as well as sgRNAs that target the coding sequence and promoter of genes that control endothelial cell functions (known genes, positive controls). (**B**) Number of sgRNAs per targeted variant that passed stringent quality-control filters. In total, we designed 7393 sgRNAs against 1998 CAD-associated variants (mean and median number of sgRNA per variant are 3.7 and 5, respectively). (**C**) Distribution of the absolute distance of the sgRNA cut-site relative to the targeted variant in base pairs (the vertical dashed line indicates mean sgRNA distance). (**D**) Fraction of variants at each locus that are successfully targeted by our pooled CRISPR screens. Each row represents one of the CAD loci that we tested. In green is the fraction of variants - including sentinel and LD proxies - for which we designed high-quality sgRNAs and obtained results for the endothelial function phenotypes. On average, 76% of variants at any given CAD locus are captured in the screens (vertical dashed line). (**E**) Most severe annotation for the 1998 CAD variants targeted by the lentiviral sgRNA libraries using ENSEMBL's Variant Effect Predictor (VEP) module. (**F**) As a control step, we sequenced the plasmid library to ensure even representation of sgRNAs in the pool. Then, we produced four independent batches of lentiviruses which we used to infect teloHAEC cells that stably express Cas9, dCas9-KRAB (CRISPRi) or dCas9-VP64 (CRISPRa). Following antibiotic selection and TNFα treatment (for Cas9 and CRISPRi), we stained teloHAEC for cell surface markers (E-selectin, ICAM-1, VCAM-1) or intracellular signaling molecules (reactive oxygen species (ROS), nitric oxide

847 (NO), calcium (Ca^{2+})). By flow cytometry, we sorted cells from the bottom and top 10

848 percentiles of the marker distributions, and sequenced sgRNAs found in each fraction.

849

850

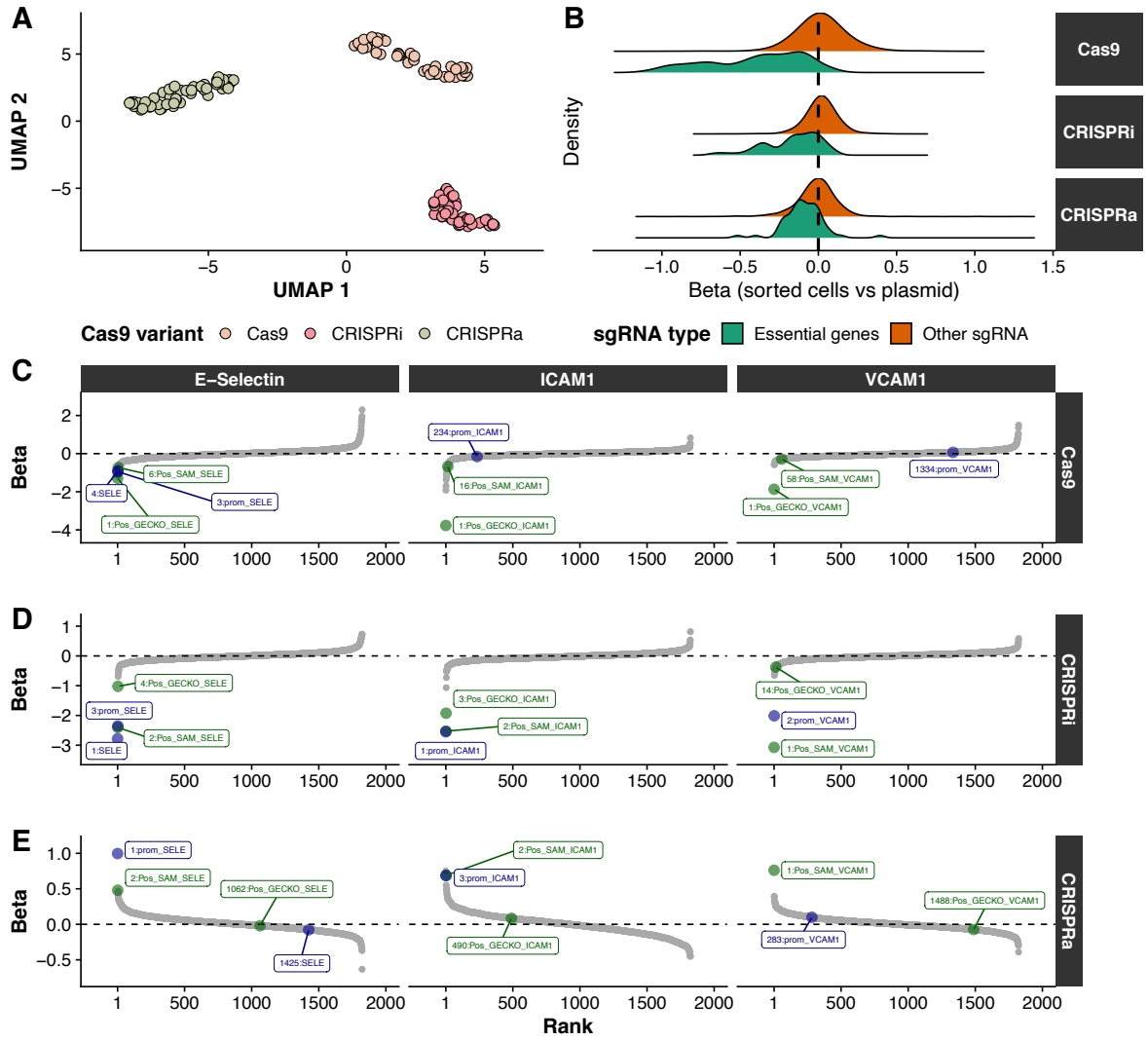


Figure 2. Quality-controls of the pooled CRISPR screens for vascular endothelial cell phenotypes.

(**A**) Two-dimensional uniform manifold approximation and projection (UMAP) representation of 148 fluorescence-activated cell sorting (FACS) samples based on the normalized read counts of the top 10% most variable sgRNAs across all samples. (**B**) Density distributions of effect sizes (Beta, *x*-axis) across all Cas9 variants for essential genes and the rest of the sgRNA library. Positive betas indicate that sgRNA are enriched in the cell fractions when compared to the input library, while negative betas indicate a depletion of sgRNA across all samples. We observed a depletion of sgRNA targeting essential genes with all three Cas9 variants. (**C-E**) Rank of all control

862 sgRNAs and targeted CAD variants in the (**C**) Cas9, (**D**) CRISPRi and (**E**) CRISPRa
863 screens for three adhesion proteins: E-selectin (left), VCAM1 (middle) and ICAM1
864 (right). For each panel, the *y*-axis corresponds to the effect sizes (Beta, comparing top
865 *vs* bottom FACS 10% fractions). For the Cas9 and CRISPRi experiments, we found
866 an enrichment of sgRNAs targeting the coding and promoter sequences of genes
867 encoding adhesion proteins in the bottom 10% cell fractions (negative Betas). In
868 contrast, sgRNAs targeting the promoter of these genes were enriched in the top 10%
869 cell fractions in the CRISPRa experiments. In green and blue, we highlight sgRNAs
870 targeting coding exons and promoters, respectively. The number in front of the name
871 of each control sgRNA indicates its rank in the corresponding analysis.
872

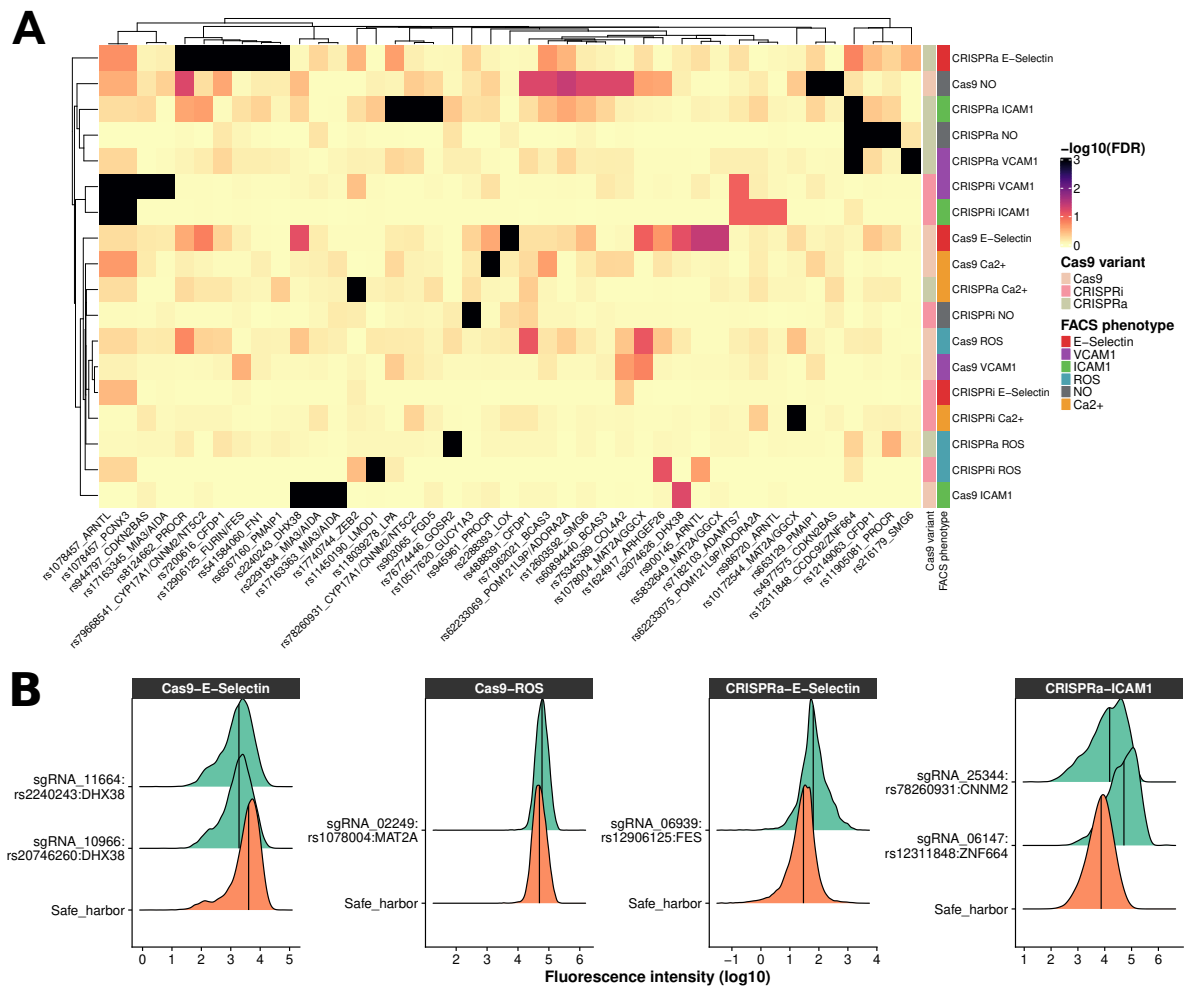


Figure 3. Discovery and validation of CRISPR perturbations that induce atheroprone vascular endothelial cell phenotypes.

(**A**) Heatmap of CAD-associated variants that are significant (false discovery rate (FDR) ≤10%) for at least one of six endothelial phenotypes tested in the teloHAEC pooled CRISPR screens. Each row corresponds to a combination of Cas9 variant and cellular readout, and each column corresponds to a CAD variant. For each variant, we added the name of a nearby gene to simplify locus identification, although we do not imply that these genes are causal. Dendrograms of rows and columns represent hierarchical clustering based on euclidean distance. The FDR is capped at 0.1%. (**B**) Validation by flow cytometry of six hits from the pooled CRISPR screens. For each validation, we used the top sgRNA from the pooled CRISPR screens to target the

variant/locus with the corresponding Cas9 variant. We compared the distribution of the fluorescence intensity of the cellular markers (*x*-axis) between the sgRNA identified in the screens and a safe harbor negative control sgRNA. We assessed statistical significance using the Kolmogorov-Smirnov (KS) test, all validations shown are significant (KS P-value $<2.2x10^{-16}$). Validations were performed in at least three independent experiments for each sgRNA (**Supplementary Table 5**).

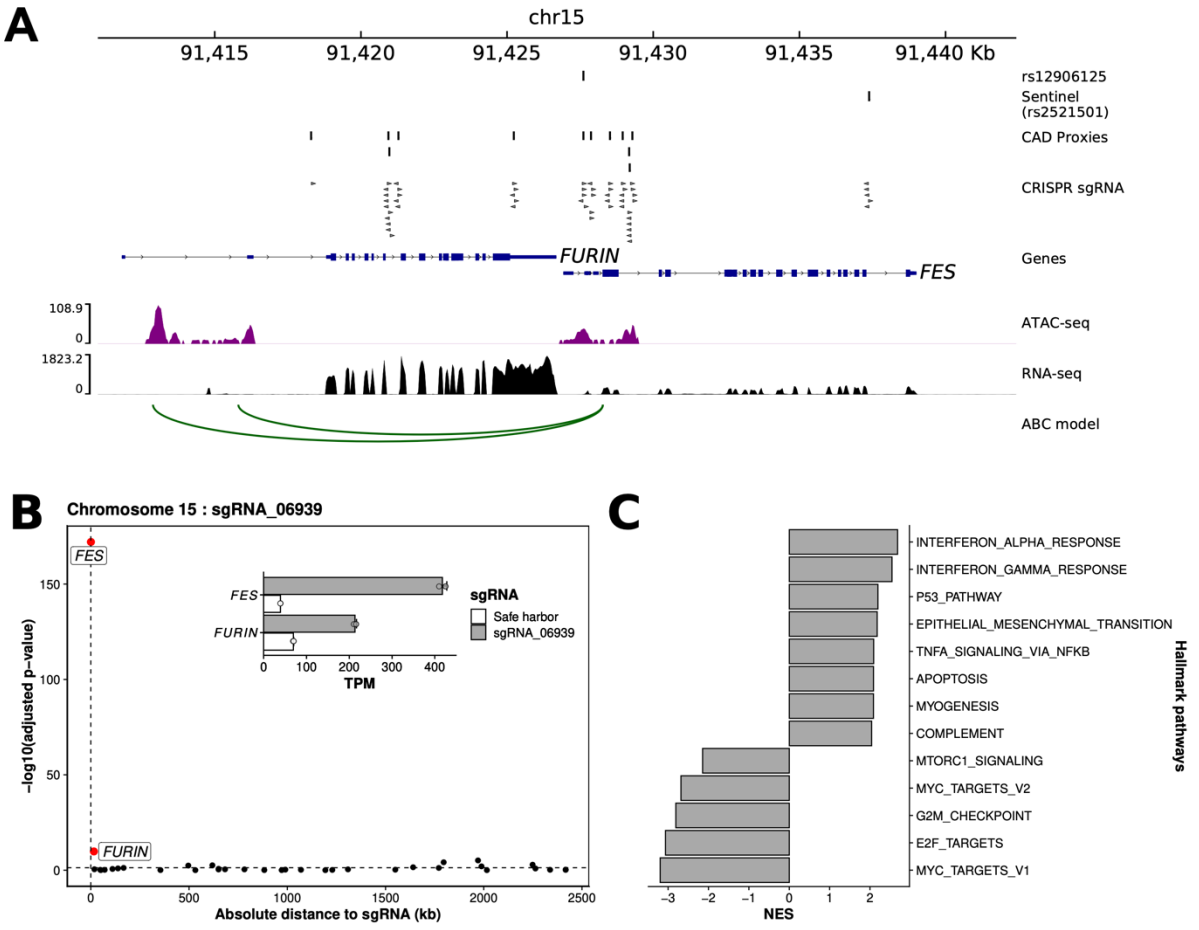


Figure 4. Characterization of a CAD-associated regulatory variant located within an enhancer at the *FURIN*/*FES* locus.

(**A**) CRISPRa perturbations highlighted rs12906125 as a potential regulatory variant for *FURIN* and *FES*. The variant overlaps an ATAC-seq peak in the promoter of *FES* and a H3K27ac-defined enhancer that physically interacts with the *FURIN* promoter through chromosomal loops predicted by the ABC model applied to teloHAEC Hi-C data[9,64]. (**B**) Within a 2.5-Mb window, *FES* and *FURIN* are the top two differentially expressed genes when targeting rs12906125 by CRISPRa in teloHAEC. The inset plot shows the induction of both *FES* and *FURIN* expression with sgRNA_06939 when compared to the control safe harbor sgRNA. (**C**) Enrichment results of differentially expressed genes for the MSigDB (H) hallmark gene-sets, comparing RNA-seq data for CRISPRa experiments with sgRNA_06939 targeting variant rs12906125 to a

906 control safe harbor sgRNA. Only pathways with a Benjamini-Hochberg-corrected P-
907 value <0.05 and a normalized enrichment score (NES) <-2 or >2 are highlighted. All
908 results from the pathway analyses are presented in **Supplementary Table 6**.
909

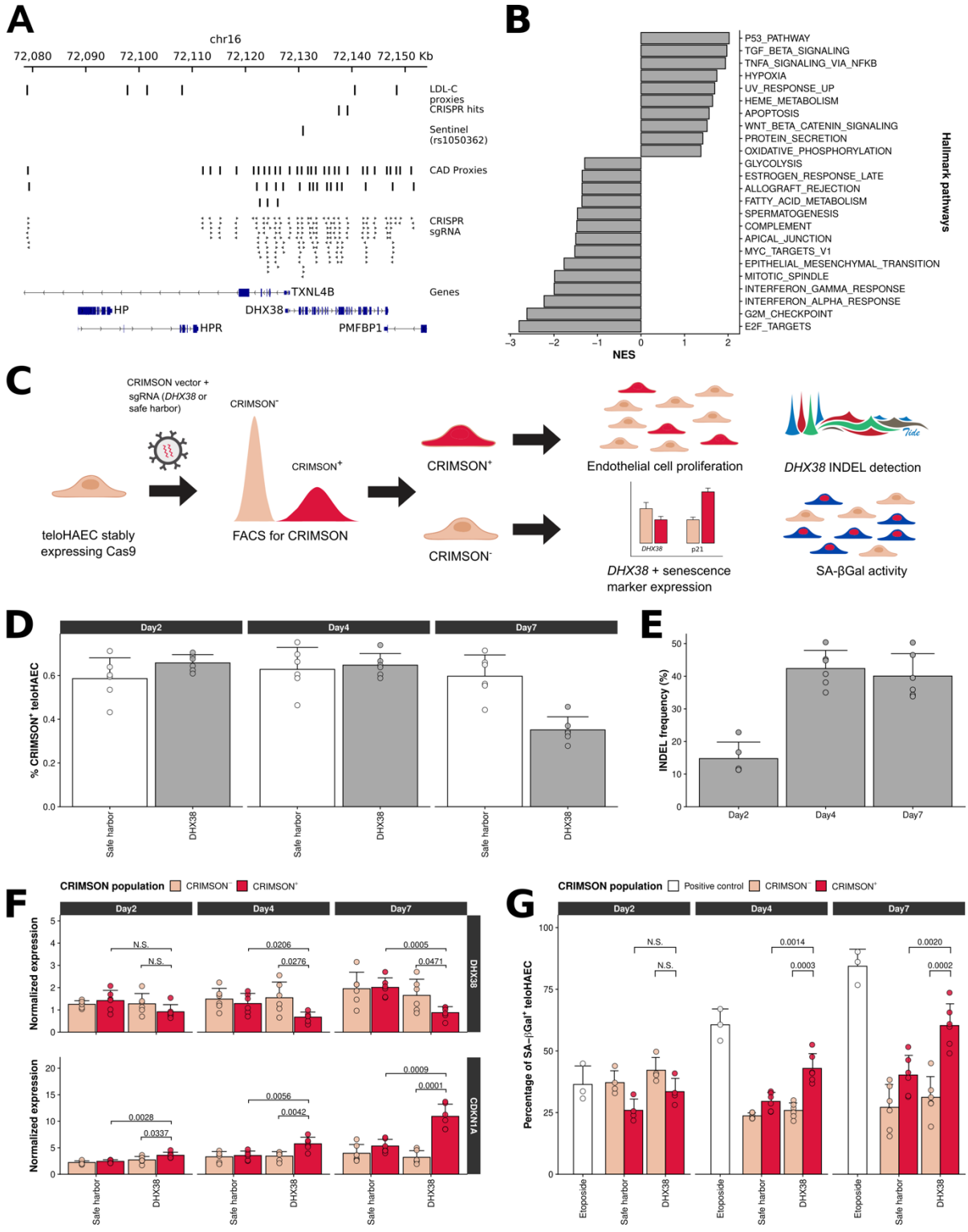
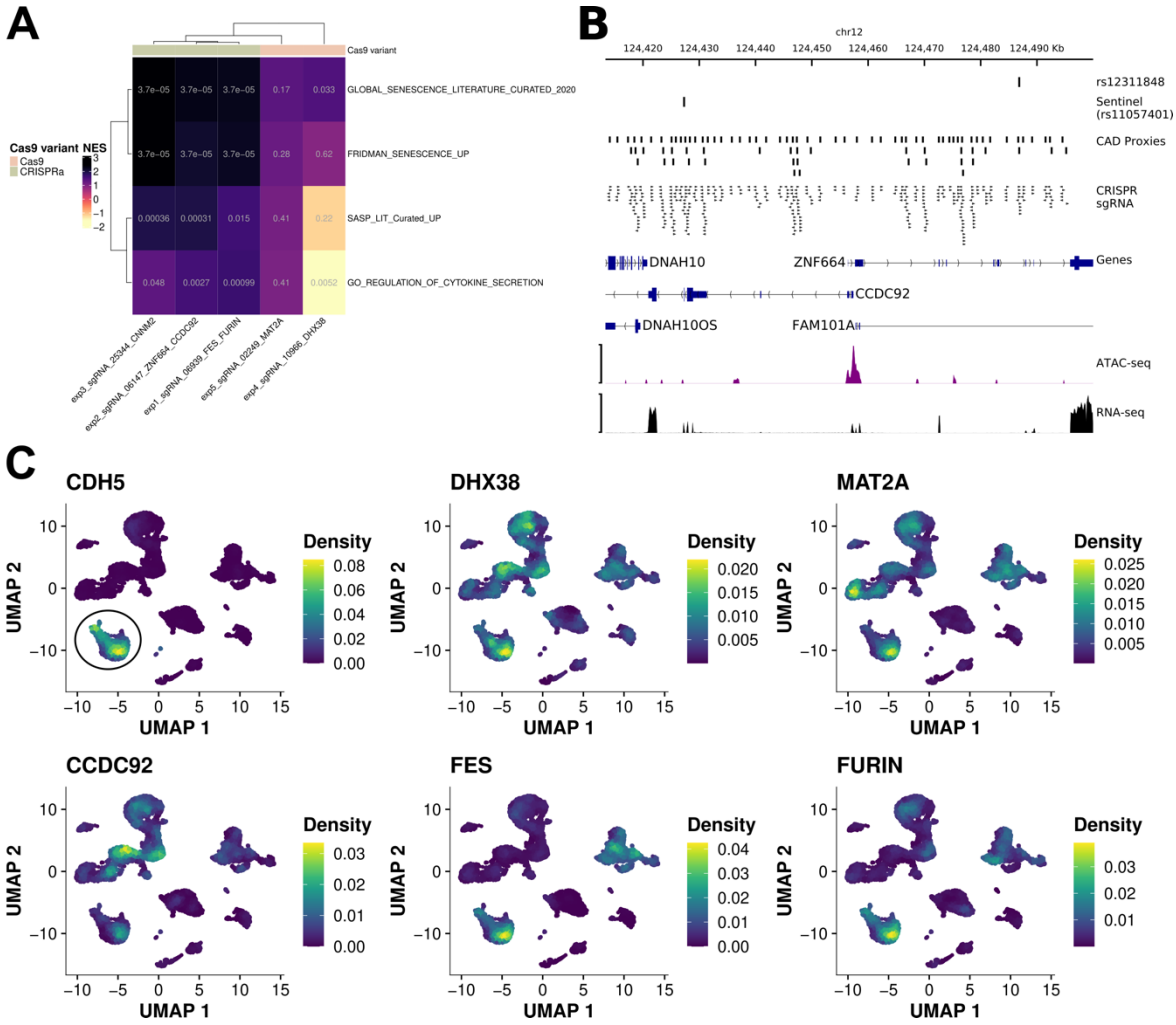


Figure 5. Disruption of *DHX38* induces vascular endothelial cell senescence. (**A**) Perturbations with the Cas9 nuclease highlighted two synonymous variants (rs2074626, rs2240243) in the *DHX38* gene for several endothelial phenotypes. *DHX38* is located downstream of the *HP* and *HPR* genes, which have previously been

915 associated with LDL-C levels. However, the CAD and LDL-C GWAS signals are
916 distinct. (**B**) Gene-set enrichment analysis results for differentially expressed genes
917 identified by RNA-seq in teloHAEC between a sgRNA targeting a *DHX38* coding exon
918 and a safe harbor negative control sgRNA. Only pathways with a Benjamini-Hochberg-
919 corrected P-value <0.05 and normalized enrichment scores (NES) <-1 or >1 are
920 shown. (**C**) Experimental design for the characterization of *DHX38* using the
921 fluorescent marker CRIMSON in place of an antibiotic resistance gene. We did all
922 experiments in teloHAEC that stably express Cas9. We monitored the impact of a
923 *DHX38* sgRNA on cell proliferation, indel induction, gene expression and senescence-
924 associated β-galactosidase (SA-βGal) activity. (**D**) Comparison of endothelial cell
925 proliferation between teloHAEC with a *DHX38* sgRNA or a safe harbor negative
926 control sgRNA. The differences in the number of CRIMSON^{+} cells were not significant
927 two or four days post-infection. However, there were 27% less CRIMSON^{+} cells with
928 *DHX38* sgRNA relative to the safe harbor control at seven days post-infection
929 (Student's *t*-test *P*-value = 7.3×10^{-8}). Results are mean ± standard deviation for 6
930 replicates for safe harbor and three replicates for two DHX38 targeting sgRNA. (**E**)
931 Quantification of *DHX38* indels by tracking of indel by decomposition (TIDE) analysis.
932 As expected, we found no indels in the CRIMSON^{-} cells (**Supplementary Table 7**).
933 However, in CRIMSON^{+} cells that received a *DHX38* sgRNA, we found indels with an
934 average frequency of 15%, 42% and 40% at day 2, 4 and 7, respectively. Results are
935 mean ± standard deviation for 6 replicates for safe harbor and three replicates for two
936 DHX38 targeting sgRNA. (**F**) Expression levels of *DHX38* and *CDKN1A* in CRIMSON^{-}
937 and CRIMSON^{+} teloHAEC that have received a sgRNA that targets *DHX38* or a safe
938 harbor region (negative control). There were no significant differences in *DHX38*
939 expression levels at day 2. However, at day 4 and 7, *DHX38* was significantly down-

940 regulated and *CDKN1A* was significantly up-regulated in CRIMSON^{+} cells that
941 received the *DHX38* sgRNA. N.S., not significant. We provide Student's *t*-test P-values
942 when $P<0.05$. Bars are mean normalized expression and error bars represent one
943 standard deviation. (**G**) Quantification of senescent teloHAEC by flow cytometry using
944 senescence-associated β-galactosidase (SA-βGal) staining. At day 4 and 7 post-
945 infection, there were significantly more senescent cells in the CRIMSON^{+} *DHX38*
946 sgRNA experiment than in the CRIMSON^{-} cells or in the CRIMSON^{+} cells that received
947 the safe harbor sgRNA. We used the DNA damaging agent etoposide as a positive
948 control to induce senescence. N.S., not significant. We provide Student's *t*-test P-
949 values when $P<0.05$. Results are mean percentage SA-βGal^{+} teloHAEC and error
950 bars represent one standard deviation.
951



952

953 **Figure 6.** CRISPRa at the *FURIN*/*FES*, *CCDC92*/*ZNF664* and *CNNM2* loci activates
954 a gene expression signature of senescence. (**A**) Gene-set enrichment analyses of the
955 five RNA-seq experiments for four curated gene sets that capture senescence or the
956 SASP [27]. The three CRISPRa experiments are highly significant for the senescence
957 pathways. In each cell of the heatmap, the color and number indicate, respectively,
958 the normalized enrichment score (NES) and the Benjamini-Hochberg adjusted P-
959 value. (**B**) Locus view for the CAD locus with nearby genes *ZNF664* and *CCDC92*.
960 We provide the position of the sentinel CAD variant (rs11057401) and the functional
961 variant identified in the pooled CRISPR screen (rs12311848). The LD proxies and
962 sgRNAs tested are also shown. ATAC-seq and RNA-seq data in resting teloHAEC are
963 from ref. [9]. *DNAH10*, *DNAH10OS*, *FAM101A*, and *ZNF664* were not differentially

964 expressed in the CRISPRa experiment. (**C**) Uniform manifold approximation projection
965 (UMAP) for 11,756 cells from human right coronary arteries analyzed by single-cell
966 RNA-sequencing[65]. We color-coded cells based on the level of expression of
967 candidate causal CAD genes identified and characterized in this study. We used the
968 expression of the endothelial cell marker gene *CDH5* (encoding VE-Cadherin) to
969 identify endothelial cells (circle in top left panel). All five candidate genes are
970 expressed in human vascular endothelial cells from coronary arteries.

971