HGG Advances

ARTICLE

Genetic regulation of circular RNA expression in human aortic smooth muscle cells and vascular traits

Redouane Aherrahrou,[1,3,*] Dillon Lue,[2,3] and Mete Civelek[1,2,4,*]

Summary

Circular RNAs (circRNAs) are a class of non-coding RNAs that have cell-type-specific expression and are relevant in cardiovascular disease. Aortic smooth muscle cells (SMCs) play a crucial role in cardiovascular disease. In this study, we employed a systems genetics approach to identify SMC circRNA transcripts and their relevance in cardiovascular traits across the genome. We quantified circRNA expression across 151 quiescent and proliferative human aortic SMCs from donors of various genetic ancestries. We identified 1,589 expressed circRNAs. Between quiescent and proliferative SMCs, we identified 173 differentially expressed circRNAs. To characterize the genetic regulation of circRNA expression, we associated the genotypes of 6.3 million single nucleotide polymorphisms (SNPs) with circRNA abundance and found 96 circRNAs that were associated with genetic loci. Three SNPs were associated with circRNA expression in proliferative SMCs but not quiescent SMCs. We identified six SNPs that had distinct association directions with circRNA isoforms from the same gene. Lastly, to identify the relevance of circRNAs in cardiovascular disease, we overlapped genetic loci associated with circRNA expression with vascular disease-related genome-wide association studies loci. We identified 14 blood pressure, one myocardial infarction, and three coronary artery disease loci, which were associated with a circRNA transcript but not an mRNA transcript. Overall, our results provide insight into the genetic basis of vascular disease traits mediated by circRNA expression.

Introduction

Smooth muscle cells (SMCs) are the major cellular component of the blood vessel wall. In physiological conditions, SMCs exist primarily in a contractile phenotype that allows them to maintain vascular homeostasis through vasodilation or contraction and proper vessel wall elasticity by secreting extracellular proteins.[1] During pathophysiological conditions, SMCs differentiate from a contractile (quiescent) to a synthetic (proliferative) phenotype that reduces their ability to contract, alters the composition of secreted extracellular proteins, and increases their capacity for proliferation, migration, and calcification.[1] This phenotypic switching results in vascular remodeling and therefore plays a role in hypertension, aortic aneurysms, coronary artery disease (CAD), strokes, and heart attacks.[2]

circRNAs are an abundant, tissue-specific class of non-coding RNAs formed through a covalent linkage between the 5′ and 3′ end of an RNA molecule. circRNAs have numerous functions, including acting as microRNA and RNA binding protein sponges, modulating the expression of linear transcripts, and being translated themselves.[3–6] Because of their lack of ends, circRNAs are protected from exonucleases and have a longer lifespan than their linear counterparts.[7] *CircANRIL*,[8] *circACTA2*,[9] *circLRP6*,[10] *circSFMBT2*,[11] and *circMAP3K5*[12] are involved in SMC migration, proliferation, and phenotypic modulation. Further, 9p21, which is the most significantly associated genomic locus with CAD, regulates the expression of *circANRIL* in atherosclerotic plaques.[8] However, a comprehensive genome-wide analysis linking genetic variation to SMC circRNA expression is missing.

Genome-wide association studies (GWAS) have linked genetic loci with many cardiovascular traits including CAD, stroke, and high blood pressure.[13] The intermediate molecular effector of individual loci is largely unknown. The majority of the significantly associated loci map to non-coding regions of the genome, pointing to their effect on regulating gene expression. Large-scale studies such as the Genotype Tissue Expression (GTEx) project[14] map GWAS loci to mRNA expression levels in what is known as an expression quantitative trait loci (eQTL) analysis. The integration of GWAS and eQTL studies enables the prediction of causal genes affecting complex phenotypic traits.

In a recent study, we had found that 37% of CAD GWAS loci are associated with mRNA gene expression levels in aortic SMCs.[15] To uncover additional vascular-related GWAS loci mediated by SMCs, we extended the traditional eQTL study to identify variants associated with circular RNA (circRNA) expression in aortic SMCs in what we term a circQTL (Figure 1). We cultured SMCs from 151 donors in quiescent and proliferative culture conditions and performed RNA sequencing. We identified 1,589 circRNAs,

[1]Center for Public Health Genomics, University of Virginia, Old Med School 3836, PO Box 800717, Charlottesville, VA 22908-0717, USA; [2]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA
[3]These authors contributed equally
[4]Lead contact
*Correspondence: ra2qy@virginia.edu (R.A.), mete@virginia.edu (M.C.)
https://doi.org/10.1016/j.xhgg.2022.100164.

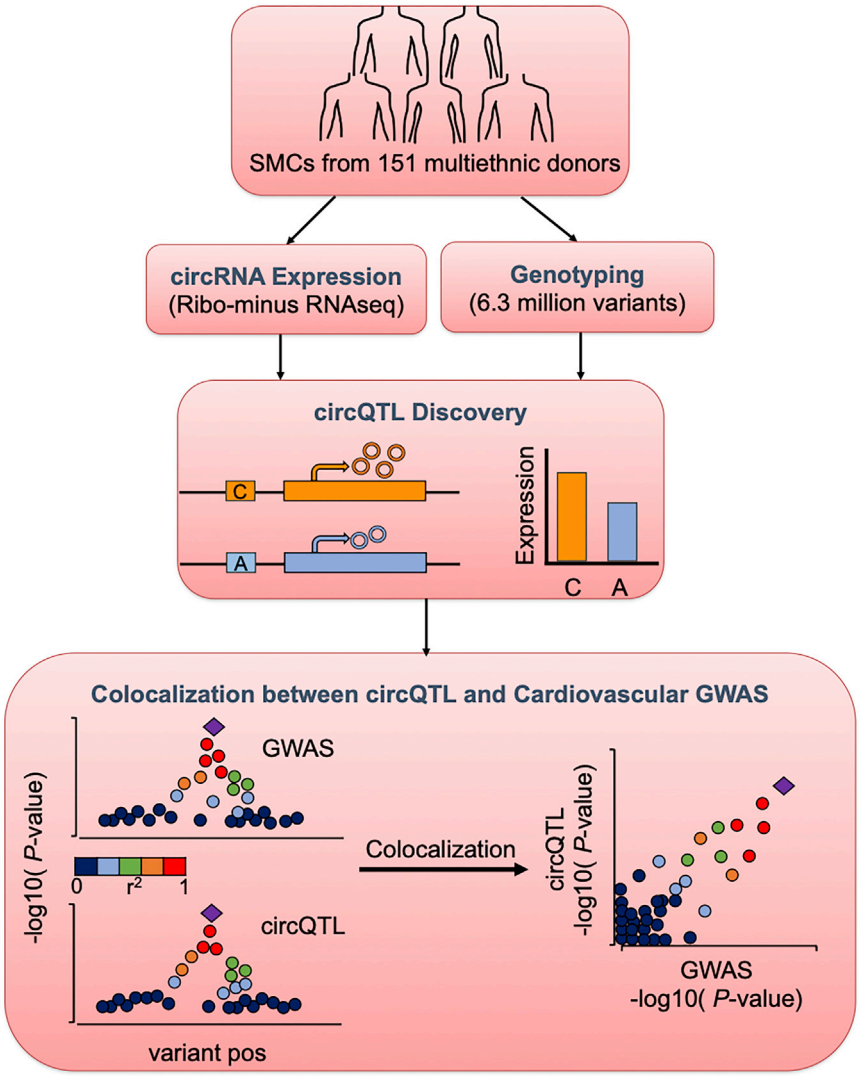

Figure 1. Study design and overview of analysis

We quantified circRNA expression using RNA sequencing of quiescent and proliferative human aortic smooth muscle cells (SMCs) from 151 donors of various genetic ancestries. We associated circRNA expression with ~6.3 million variants to perform circRNA quantitative trait loci (circQTL) mapping. The "C" and "A" represent alleles within non-coding regions of the genome. Subsequently, we performed colocalization between circQTLs and vascular GWAS traits.

many of which were SMC phenotype specific. From these 1,589 circRNAs, we identified 139 genetically regulated circRNAs, three of which were phenotype specific. Using colocalization analyses, we identified 14 blood pressure, one myocardial infarction, and three CAD GWAS loci that had association with circRNA expression but not mRNA expression. This study reveals that circRNA expression, independently from mRNA expression, mediates some of the genetic risk for vascular traits.

Methods

University of Virginia Institutional Review Board determined that the studies performed in this manuscript do not meet the definition of human subjects research since the aortic pieces were obtained from cadavers during heart transplant operations.

Cell culture, genotyping, and RNA sequencing

As previously described, we collected SMCs from 151 heart transplant donors of various genetic ancestries. We cultured each of the 151 SMC samples in media with fetal bovine serum to represent proliferative SMCs and media without fetal bovine serum to represent quiescent SMCs.[15] We performed RNA-seq of the ribosomal RNA-depleted total RNA. We prepared sequencing libraries with the Illumina TruSeq Stranded mRNA Library Prep Kit. Psomogen sequencing facility sequenced each sample to ~100 million read depth with 150-bp paired-end reads. We had genotype information for each donor for ~6.3 million SNPs with minor allele frequency >5%.[16]

circRNA detection

We trimmed paired-end RNA-seq reads with low average Phred scores (<20) using Trim Galore and performed circRNA mapping and quantification with BWA and circRNA Identifier 2 (CIRI2) in accordance with CIRI's recommended pipeline with all default parameters except for –0, which outputs all backsplicing sites without filtering.[17,18] We used the hg38 genome and GENCODE v32 annotations. CIRI counted circular junction counts, which is the number of reads overlapping a backsplicing site. We used circular junction counts as the quantitative measure of circRNA expression. CIRI also calculated the circular junction ratio, which is the ratio of circRNA expression to linear transcript expression as given by

$$Circular\ ratio = \frac{C}{C+L},$$

where C represents the number of circular counts, and L represents the number of reads supporting a linear transcript at the junction.

circRNA preprocessing

We considered a circRNA expressed if the junction count was ≥ 3 and the circular ratio was ≥ 0.05 in at least 20% of the samples of a particular SMCs phenotype. We used a threshold of counts ≥ 3 to reduce the number of circRNA false positives, and we used a threshold of ≥ 0.05 to reduce the possibility of mRNAs being miscounted as circRNAs. For downstream differential circRNA expression and circRNA quantitative trait locus mapping, we tested only expressed circRNAs. We classified circRNAs as exonic, intronic, and intergenic based on GENCODE v32 annotations. CIRI classified a circRNA as intronic or intergenic if at least one splice site

was "intergenic" or "intronic" respectively; otherwise, CIRI classified a circRNA as exonic. CIRI determined a circRNA's parent gene if the backsplicing site overlapped an intron or exon of an annotated gene. Intergenic circRNAs were not assigned a parent gene.

Differential circRNA expression

We performed differential circRNA expression between quiescent and proliferative SMCs using edgeR.[19,20] In order to compare expression values across different samples, we applied trimmed mean of M values (TMM) normalization.[21] Within edgeR, we used sex and the first four genotype principal components (PCs) as covariates. We considered a circRNA differentially expressed if the q-value false discovery rate was less than 0.05.[22]

Circular RNA quantitative trait loci mapping

To correct for library depth, we converted circRNA junction counts to counts per million (CPM), whereby a particular circRNA junction count was divided by the total junction counts in millions for that sample. We next normalized the CPM using TMM[21] followed by inverse normalization. We then tested the association of circRNA expression and ~6.3 million variants. We term each significant association a circRNA quantitative trait locus (circQTL). As covariates, we included technical confounders estimated through probabilistic estimation of expression residuals (PEER),[23] four genotype PCs, and sex. From 2, 4, 6, 8, and 10 PEER covariates, we chose the number of PEER covariates that maximized the number of circQTLs. We performed circQTL detection using tensorQTL, which performs linear regression between genotypes and circRNA expression with a permutation scheme that accounts for linkage disequilibrium by approximating the distribution of p values within a locus and subsequently accounts for genome-wide testing using q-value correction.[15] We considered a circQTL significant if the q-value false discovery rate was less than 0.05. We identified secondary circQTLs by rerunning permutation analysis but conditioning on the lead SNP of the locus.[24] We tested variants within 500 kb of either end of the backsplicing site. We term eSNP as the lead or most significant circQTL SNP in the locus.

Phenotype-specific circQTLs

To determine phenotype-specific circQTLs, we tested if the eSNP for every significant circQTL had different effect sizes between proliferative and quiescent SMCs using a Z score formulated as follows:

$$z = \frac{\beta_{quiescent} - \beta_{proliferative}}{\sqrt{\sigma^2_{quiescent} + \sigma^2_{proliferative}}},$$

where β is the effect size of the eSNP, and σ is the variance of the effect size.[25] We determined a phenotype-specific circQTL if the Bonferroni-corrected p value of the Z-test was <0.05.

For each circRNA that passed this initial test, we performed an interaction test where we modeled circRNA expression as linear function of genotype, covariates, SMC phenotype, and an an interaction term for genotype and SMC phenotype.

circRNA isoform-specific circQTLs

We define circRNA isoforms when distinct backsplicing sites arise from the same parent gene. To identify if any eSNPs were associated with the expression of more than one circRNA isoform at a time, we tested for differences in effect size of an eSNP on two circRNAs deriving from the same parent gene. We performed analyses separately in each SMC phenotype. For every parent gene that had two significant circQTLs on different circRNA isoforms, we performed two statistical tests. First, we tested if the effect size of the first eSNP on the first circRNA isoform had opposite effect compared with the effect size of the first eSNP on the second circRNA isoform. Similarly, we tested if the effect of the second eSNP had opposite effect size on the expression of the first circRNA isoform than the second circRNA isoform. For each test we used the Z score formulated as follows:

$$z = \frac{\beta_1 - \beta_2}{\sqrt{\sigma^2_1 + \sigma^2_2}},$$

where β_1 and σ_1 are the effect size and variance of the primary circQTL, and β_2 and σ_2 are the effect size and variance of the eSNP associated with another circRNA from the same parent gene. To account for the two tests performed per parent gene, we applied an initial Bonferroni correction by multiplying the p value of the Z-test by 2. Lastly, to account for multiple testing hypothesis across the genome, we applied a second Bonferroni correction to all parent genes tested.

To determine if there was a difference in directionality of effect sizes between an eSNP and two circRNAs, we took 90% confidence intervals of the effect size (β) of the eSNP on a circRNA. We classified an effect as positive if the confidence interval did not include 0 and the Z score was positive. Likewise, we classified an effect as negative if the confidence interval did not include 0 and the Z score was negative. Using the effect size annotations for directionality, we identified circRNA isoform-specific direction if the effect sizes differed in sign.

Colocalization

To test whether a genetic locus was associated with circRNA expression and mRNA expression, we performed colocalization of circQTL loci with eQTL loci using COLOC.[26] Using eQTL summary statistics that we generated previously,[15] we performed colocalization for each significant circQTL and considered a locus colocalized if the posterior probability of colocalization (PPH4) was greater than 0.5.

To identify if circRNAs contribute to the genetic risk for vascular related diseases, we performed colocalization of circQTL loci with vascular disease-related GWAS loci using COLOC[26] and eCAVIAR.[27] We downloaded summary statistics for CAD, myocardial infarction, stroke, aortic aneurysm, and blood pressure GWAS catalog or UK Biobank (UKBB) (Table S1).[13,28] If available, we downloaded GWAS data from both sources. We performed colocalization analysis separately for each GWAS. For each circQTL that passed q-value <0.05, variants within 500 kb of the eSNP were considered for colocalization. For each GWAS, we used a 1,000 Genomes project population-matched linkage disequilibrium. For eCAVIAR, we set the maximum number of causal variants to 2 and considered variants colocalized if the colocalization posterior probability (CLPP) was greater than 0.01. For COLOC, which uses a single causal variant assumption, variants were considered colocalized if the posterior probability of colocalization (PPH4) was greater than 0.5. As a final filtering step, a region was considered colocalized if at least one colocalized variant passed 10^{-4} in the GWAS.

Gene ontology enrichment analysis

We performed gene ontology enrichment analysis using Fischer's exact test as implemented in topGO.[29] We used a background of

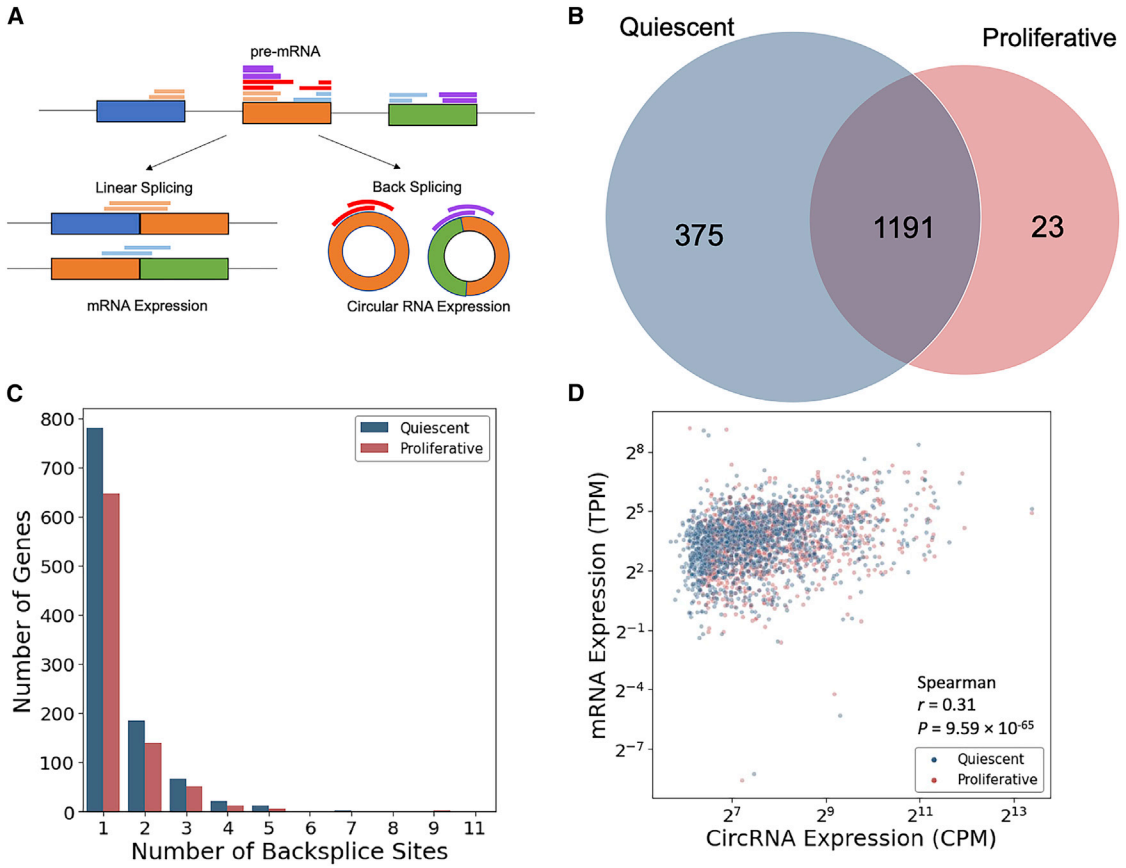

Figure 2. Circular RNA profile in human aortic SMCs

(A) mRNAs are made through linear splicing where a splice acceptor appears downstream of the splice donor. In contrast, circRNAs are made through backsplicing where the acceptor site appears upstream of the donor site creating a circular molecule. circRNAs are detected by mapping RNA-seq reads (rectangles above exon) that overlap a backsplice site. The red and purple reads support two different circRNA transcripts respectively from the same gene.

(B) Venn diagram comparing the number of expressed circRNAs in quiescent versus proliferative SMCs. circRNAs were considered expressed if the junction counts ≥ 3 and the ratio of mRNA reads to backsplice reads was ≥ 0.05 in at least 20% of samples.

(C) Histogram of number of circRNA transcripts (represented by the number of backsplice sites) per gene.

(D) Spearman correlation (*r*) and its significance (*P*) between circRNA expression and mRNA expression of the corresponding parent gene.

SMC genes that had greater than six read counts and 0.1 transcripts per million in at least 20% of samples in either SMC phenotype. We considered gene ontology terms significant if the term had nominal p value <0.01.

PCR validation of circRNA

We performed PCR validation of circRNAs exhibiting phenotype-specific regulation and circRNAs from colocalized circQTLs with vascular GWAS traits.[30] We cultured SMCs from seven randomly chosen donors in proliferative and quiescent conditions and collected total RNA using the miRNeasy Kit (QIAGEN), following the manufacturer's instructions. Because circRNAs expression varies widely across individuals, we pooled the RNA from the donors for each condition. For each SMC phenotype, we digested 2 μg of total RNA for 30 min at 37°C with 1 μL of RNase R enzyme (Lucigen), followed by RNA isolation with the PureLink RNA isolation kit.[30] We used total RNA without RNase R enzyme as control. We used primers for the linear transcript of GAPDH and ACTA2 as a control to test the effect of RNAse R. Next, we performed reverse transcription using Maxima reverse transcriptase (Thermo Fisher Scientific) with random hexamers (Thermo Fisher Scientific). Using divergent primers targeting the backsplicing site (Table S2) with NEBNext PCR master mix (New England BioLabs), we performed PCR amplification as previously described.[30] The final products were analyzed using 2% agarose gel.

Results

Landscape of circular RNA

We performed RNA sequencing on aortic SMCs derived from 151 transplant donors (118 male and 33 female). Previously, we had clustered the donor genotypes with 1000 Genomes Project reference population samples and identified 6, 12, 64, and 69 of the individuals with East Asian, African, Admixed American, and European ancestry.[16] After quantification and quality control, we obtained 145 and 139 samples cultured in the presence (proliferative) or absence (quiescent) of fetal bovine serum, respectively.[15] We used CIRI[18] to *de novo* identify and quantify circRNA

Table 1. Number of circRNAs and circQTLs identified

		circRNA expression	
Phenotype	parent gene type	number of circRNA tested	number of circRNA with circQTL
Quiescent	protein coding	1,490 (95.1%)	86 (90.0%)
	lncRNA	25 (1.6%)	3 (3.1%)
	pseudogene	12 (0.8%)	0 (0.0%)
	multiple parent gene	6 (0.4%)	2 (0.21%)
	no parent gene	33 (2.1%)	5 (5.2%)
	total	1,566	96
Proliferative	protein coding	1,151 (94.8%)	81 (84.4%)
	lncRNA	19 (1.6%)	4 (4.2%)
	pseudogene	10 (0.8%)	4 (4.2%)
	multiple parent gene	5 (0.4%)	2 (2.1%)
	no parent gene	29 (2.4%)	5 (5.2%)
	total	1,214	96

backsplicing sites (Figure 2A). Although we assumed that each backsplicing site represents a unique circRNA transcript, a backsplicing site could represent multiple circRNA isoforms with differences in internal splicing. We identified 1,214 and 1,566 expressed circRNAs (junction counts ≥ 3 and the circular ratio ≥ 0.05 in at least 20% of the samples) in the quiescent and proliferative SMCs, respectively. From the two phenotypes combined, we identified 1,589 unique circRNAs (Figure 2B). A particular circRNA has a corresponding parent gene if the circRNA's backsplicing site overlaps an annotated exon or intron of a gene. Out of the expressed circRNA transcripts, approximately 95% had a protein-coding parent gene, and only 2.1% had no parent gene (Table 1). Most circRNAs were exonic (92.5%) followed by intronic (5.3%) and intergenic (2.2%) (Figure S1).

Additionally, we found that >70% of parent genes had a single backsplicing site (Figure 2C). In our RNA-seq data, we identified 18,964 expressed mRNAs, 1,083 of which had a circRNAs, suggesting that 6% of SMC genes could be circularized. There was a statistically significant but low Spearman correlation ($r = 0.31$, $p = 9.59 \times 10^{-65}$) between circRNA expression and its parent gene expression (Figure 2D).

Differential circRNA expression

Of the 1,589 expressed circRNAs, 173 were differentially expressed between the quiescent and proliferative phenotypes (q-value < 0.05), with 62 circRNAs having a higher expression in quiescent SMCs and 111 circRNAs having a higher expression in proliferative SMCs. (Figure 3A and Table S3). Using topGO, we tested the enrichment in gene ontology terms for the 139 parent genes of differentially expressed circRNAs against a background of SMC genes.[29] The parent genes of the differentially expressed circRNAs were enriched in processes related to SMC contraction and migration including regulation of sodium ion transmembrane transport, cell polarity, and substrate adhesion-dependent cell spreading (Figure 3B). To further compare mRNA and circRNA expression, we tested if the parent genes of differentially expressed circRNAs were also differentially expressed. For each significant circRNA with a single parent gene, we performed a t test for differences in transcript per million mRNA gene expression values across quiescent and proliferative SMCs. We found that 87% (143/164) of parent genes from expressed circRNA were also differentially expressed (q-value < 0.05) (Figure 3C). Furthermore, we also found a high Spearman correlation ($r = 0.66$, $p = 5.8 \times 10^{-22}$) between circRNA fold change and parent gene mRNA fold change (Figure 3C). 41/164 (25%) of differentially expressed circRNAs showed opposite fold change directions to that of their parent gene (Figure 3C).

Quantitative trait loci mapping

We next tested the association between circRNA expression and single nucleotide polymorphisms (SNPs) within 500 kb away from the backsplicing site in each SMC phenotype separately using tensorQTL.[24] Eight and four PEER covariates maximized the number of circQTLs identified.[24] circQTL summary statistics are available at the URL indicated in the data and code availability section. Of the 1,589 expressed circRNAs tested, we identified 96 circQTLs in both SMC phenotypes separately and 53 circQTLs that were common to both SMC phenotypes (q-value < 0.05) (Figure 4A and Table S4). For each circQTL, we termed eSNP as the most significant or lead SNP in the locus. After conditioning on the eSNP of each circQTL, we identified one secondary circQTL in quiescent SMCs and two in proliferative SMCs. 86% of the circRNAs with circQTL had a parent protein-coding gene (Table 1). Using topGO, we tested the enrichment in gene ontology terms for the 139 circQTL parent genes against a background of SMC genes.[29] We identified enrichment in various pathways associated

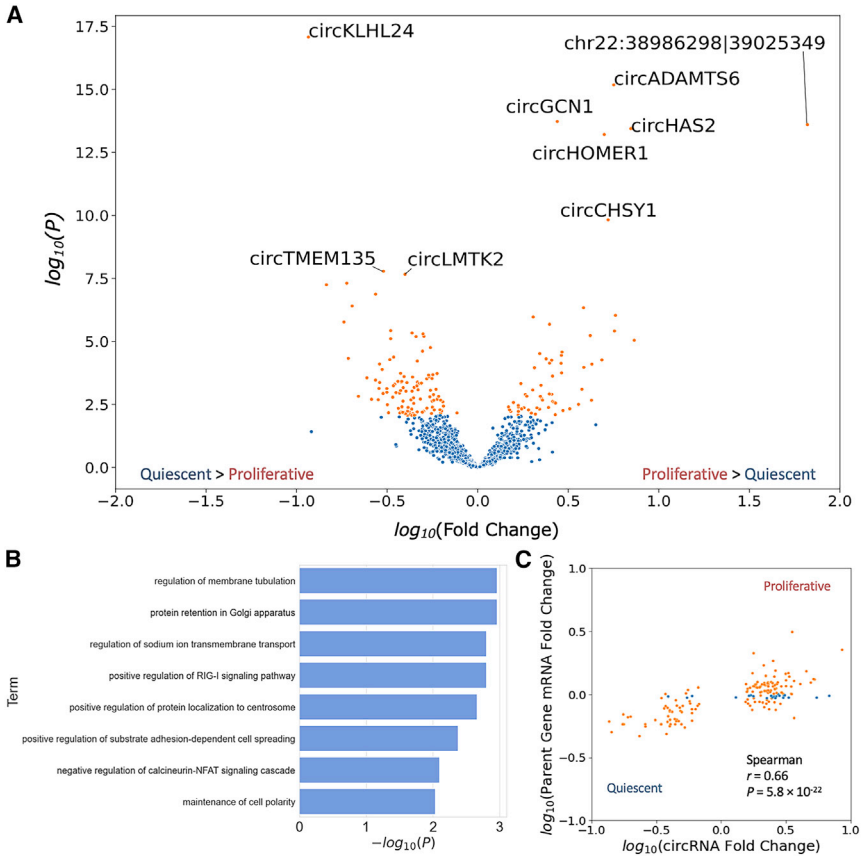

Figure 3. circRNA differential expression between quiescent and proliferative SMCs

(A) Volcano plot representing differentially expressed circRNAs. circRNAs enriched in quiescent SMCs are further to the left, and circRNAs enriched in proliferative SMCs are further to the right. The orange points represent the 173 circRNAs that passed significance q-value <0.05. The top 10 most significant circRNAs are labeled.

(B) Gene ontology enrichment analysis of the parent genes of differentially expressed circRNAs. All terms with p < 0.01 are shown.

(C) Spearman correlation (r) and its significance (P) of expression level fold changes of circRNA and its parent gene. Positive fold changes represent higher expression in proliferative SMCs, while negative fold changes represent higher expression in quiescent SMCs. The orange points represent the 143 parent genes that had statistically significant fold changes after a t test with q-value correction (q-value <0.05). The blue points represent non-significant fold changes.

with SMC biology including Rho protein signal transduction, sodium ion transmembrane transporter activity, and cytoplasmic microtubule organization (Figure 4B).

Next, we tested whether the effect size of circQTL SNPs on circRNA expression had a similar effect size on mRNA expression of its parent gene. We calculated the correlation of the eSNP effect size on circRNA expression and the eSNP effect size on parent gene expression. We determined low but significant Spearman correlation (r = 0.37, p = 3.57×10^{-7}) (Figure 4C). Although simple, this analysis did not consider the many SNPs in high linkage disequilibrium within a genetic locus. To formally test whether a genetic locus is associated with circRNA expression but not mRNA expression of the parent gene, we performed colocalization for each significant circQTL locus with the eQTL locus in matched SMC phenotypes using COLOC.[26] Across the 179 circQTL loci with parent genes in quiescent or proliferative SMCs, we identified 26 loci that were associated with mRNA expression of their parent gene (PPH4 > 0.5) (Table S5). For each SMC phenotype, we highlighted an example of a colocalized and not colocalized locus (Figure S2). To annotate circQTL SNPs, we determined the distribution of eSNP distance from the backsplicing site and found that 82% of eSNPs were within 100 kb of the circRNA backsplicing site (Figure 4D). Additionally, we found that 53% eSNPs were within introns and 14.1% were in non-coding transcripts (Figure 4E).

Phenotype-specific circQTLs

We previously showed that phenotypic state of SMCs affects the genetic regulation of mRNA expression[15]; therefore, we tested for circQTLs that were specific to either proliferative or quiescent SMCs. For each eSNP, we correlated the effect size of the eSNP on circRNA expression in the proliferative and quiescent SMCs. We found low Pearson correlation ($r = 0.36$, p = 7.01×10^{-48}) indicating phenotype-specific genetic regulation (Figure 5A). The low correlation could be a result of our low sample size and the fact that effect sizes tend to be small. Having identified low correlation, we formally tested differences in eSNP effect size between SMC phenotypes taking into consideration the uncertainty of our effect size estimates. Using a statistical test for differences in effect sizes between phenotypes, we identified three circQTLs (*circSTRN3*, *circCHD9*, and *circRIPK1*) that had a significant association with a SNP in one phenotype but not the other phenotype (Bonferroni-corrected p value < 0.05) (Figures 5B–5D). Statistical test results for each circRNA are in Table S6. Next we verified the phenotype-specific genetic regulation in *circSTRN3*, *circCHD9*, and *circRIPK1* using a linear model with an interaction term for genotype and phenotype, and the interaction term p value was significant for all three circRNAs (p value = 1.3×10^{-5}, 5.3×10^{-5}, 1.4×10^{-5}, respectively). Using RT-PCR, we showed the expression of all three circRNAs with phenotype-specific genetic regulation (Figure S3).

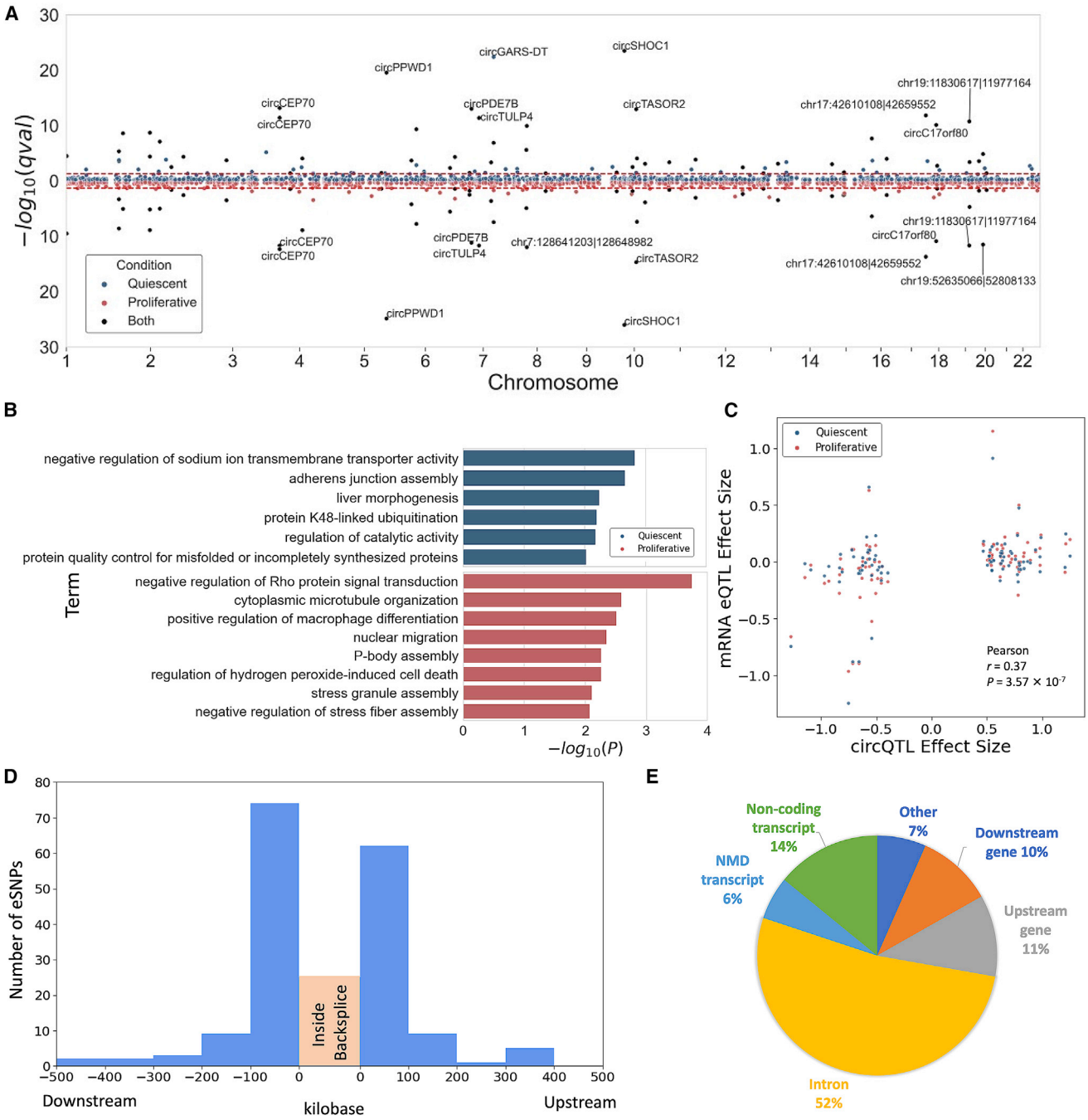

Figure 4. Genetic regulation of circRNA expression

(A) Miami plot showing the most significant association between a circRNA and a variant within a locus. The top half represents quiescent circQTLs, while the bottom half represents proliferative circQTLs. The red line indicates the significance level threshold at q-value = 0.05. All circRNAs with q-value $<10^{-10}$ are labeled.

(B) Gene ontology enrichment analysis of the parent genes of circQTLs in proliferative and quiescent SMCs. All terms with $p < 0.01$ are shown.

(C) Pearson correlation (r) and its significance (P) between SNP effect size on circRNA expression and SNP effect size on parent gene mRNA expression.

(D) Histogram of distances of the eSNPs from the backsplice site.

(E) Annotated effects of eSNPs using Variant Effect Predictor.

circRNA isoform-specific circQTLs

We detected circRNA isoforms (circRNAs that differed in backsplicing sites but were derived from the same parent gene) arising from 30% of the parent genes (Figure 2C). Across both SMC phenotypes, we identified 34 parent genes that had two significant circRNA isoforms with QTLs. We identified no parent gene with more than two circRNA isoforms with QTLs. Therefore, we tested if an

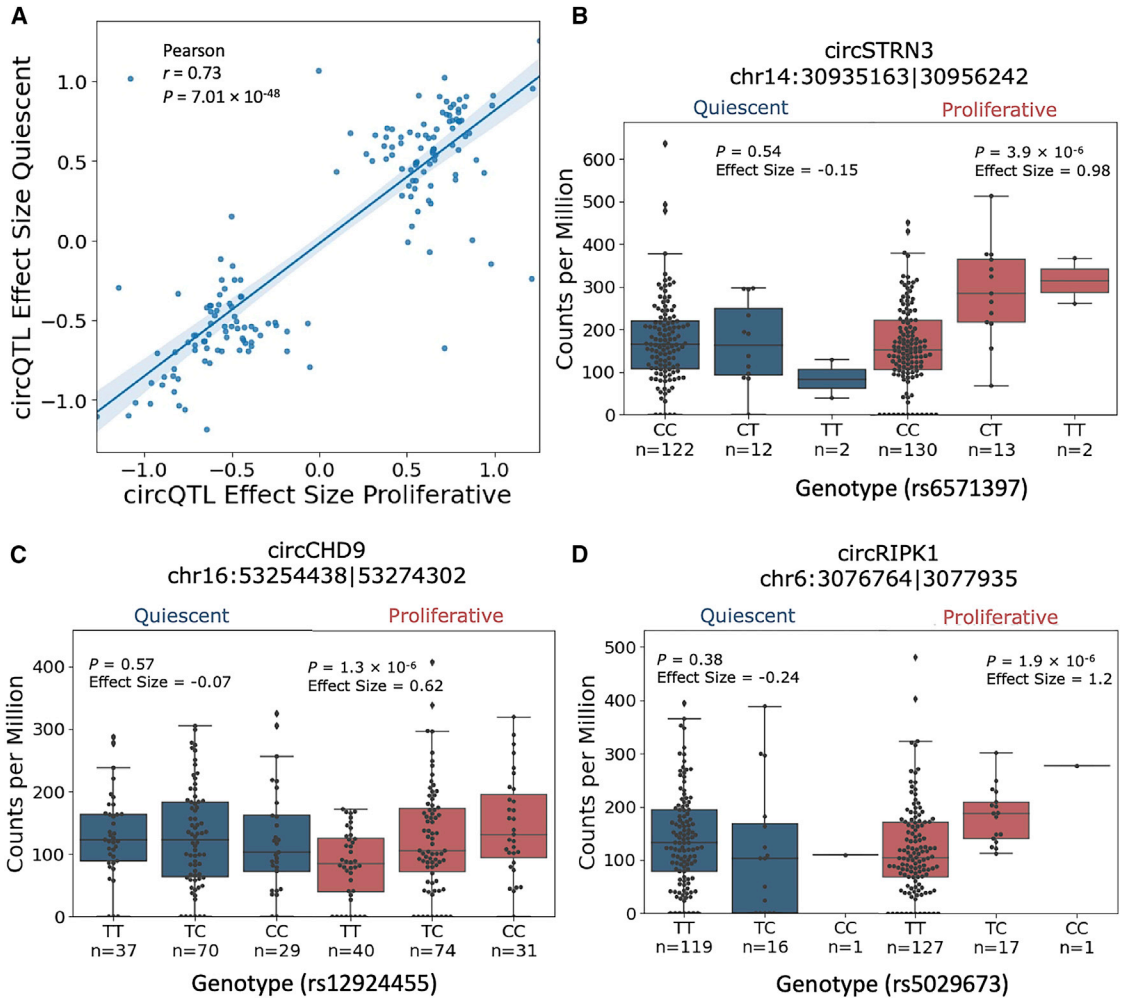

Figure 5. Phenotype-specific circQTL in SMCs

(A) Pearson correlation (*r*) and its significance (*P*) between circQTL effect sizes in the proliferative and quiescent SMCs.

(B–D) Phenotype-specific circQTLs that had a significant association in one phenotype but not the other. P-values (*P*) were calculated using tensorQTL, which performs linear regression between the indicated SNP genotype and circRNA abundance with a permutation scheme.

eSNP had opposite effect size on the expression of more than one circRNA isoform within a parent gene. In quiescent SMCs, we identified four eSNPs with circRNA isoform-specific direction; in proliferative SMCs, we identified two eSNPs with circRNA isoform-specific direction (Bonferroni-corrected p value < 0.05) (Table S7). An example of an eSNP with circRNA isoform-specific direction is the SNP rs16875285. The C allele of this SNP is associated with lower expression of one circRNA isoform but higher expression of another circRNA isoform arising from the long non-coding RNA *GARS-DT* (Figure 6A). We obtained a similar result for circRNA isoforms in *TASOR2* (Figure 6B).

circQTL overlap with vascular disease loci

To uncover circRNAs that may be playing a role in the pathogenesis of diseases that involve SMCs, we tested if vascular disease-related GWAS loci were associated with circRNA expression using colocalization. A locus was considered colocalized if either PPH4 > 0.5 using COLOC[26] or CLPP >0.01 using eCAVIAR.[27] We tested for colocalization with CAD, myocardial infarction, stroke, aortic aneurysm, and blood pressure GWAS loci. Across quiescent and proliferative SMCs, we identified 16 circQTLs that colocalized with CAD, myocardial infarction, and blood pressure GWAS loci with lead SNP p value $<10^{-4}$ (Figures 7A and S4, Tables S8 and S9). Using RT-PCR, we showed SMC expression of 14/16 circRNAs that had colocalization with at least one trait (Figure S3). Using the more common threshold of p value $<5 \times 10^{-8}$, we identified six circQTLs (*circZKSCAN1*, *chr17:42610108|42659552*, *circLARP4*, *circEPHB4*, *circSHKBP1*, *circSLC4A7*) that colocalized with GWAS loci. We highlight the colocalization of *circSLC4A7* circQTL with high blood pressure GWAS locus (Figure 7B). As previously reported, *SLC4A7* is a Na^+, HCO_3^- cotransporter that inhibits NO-mediated vasorelaxation, smooth muscle Ca^{2+} sensitivity, and higher blood pressure in mice.[31] We

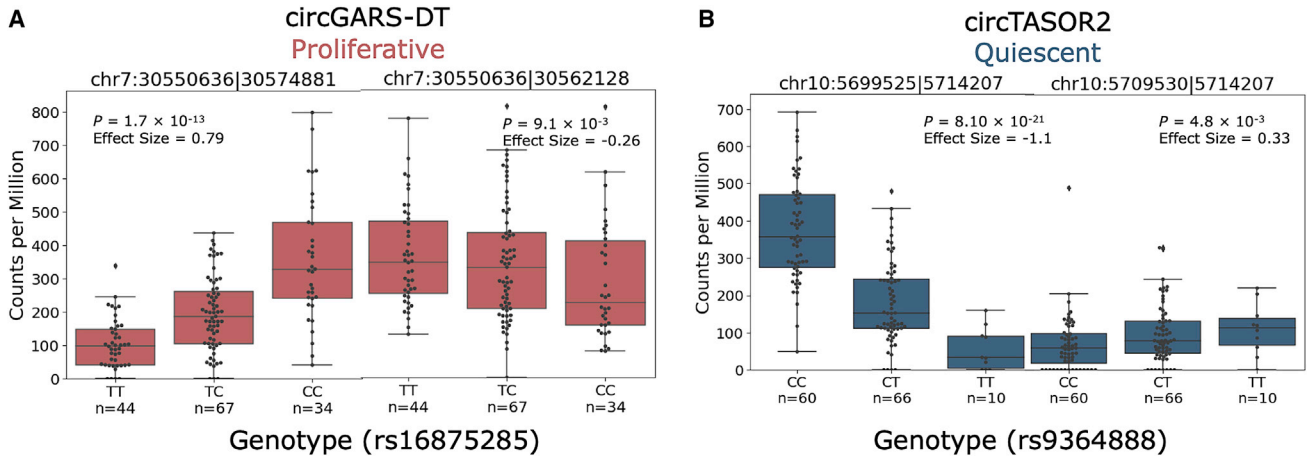

Figure 6. Genetic regulation of circRNA isoform abundance

(A and B) Two examples where a genetic variant decreases the expression of one circRNA isoform but increases the expression of another circRNA of the same parent gene. P-values (*P*) were calculated using tensorQTL, which performs linear regression between the indicated SNP genotype and circRNA abundance with a permutation scheme.

demonstrated that the *circSLC4A7* eSNP shows significant association with circular but not mRNA expression (Figure 7C). For the other 15 circQTLs colocalized with vascular traits, we also found that the eSNP was not associated with parent mRNA expression, suggesting that circRNAs and not mRNA expression are the causal mediators for these vascular trait loci (Figure S4). As previously described, we identified 26 circQTL loci associated with mRNA expression of their parent gene (Table S5). None of these loci were colocalized with a vascular trait locus. This further supports circRNAs-specific mediation of vascular trait loci.

Discussion

Recent studies have identified roles for circRNAs in many complex diseases, including cardiovascular disease.[32] It is often difficult to systematically measure and associate circRNA expression with a disease phenotype in a case-control setting. To overcome this, we used a systems genetics approach that utilizes naturally occurring randomly assigned genetic variants as causal mechanisms for changes in circRNA expression and disease phenotype. Previous GWASs have associated hundreds of loci with vascular-related disease traits where SMCs likely play a key role by differentiating from a contractile (quiescent) to a synthetic (proliferative) phenotype.[13,33] To annotate these loci, we identified and characterized the *cis*-acting genetic variation on circRNA expression in SMCs cultured with and without fetal bovine serum to artificially mimic the contractile and synthetic SMCs, respectively.

Using a unique source of 151 aortic SMCs from donors of various genetic ancestries, we identified 1,589 circRNAs of which 173 were differentially expressed between quiescent and proliferative SMCs. Although previous studies have performed differential expression of circRNAs between platelet-derived growth factor BB-treated and control vascular SMCs,[11,12,34,35] our study tested differences in circRNA expression across SMCs cultured in the presence or absence of fetal bovine serum. While we did not treat our RNA with linear RNA degradation (RNase R) enzyme like these other studies, we included many more samples (150 donors in two conditions). Our dataset provides unique insight into highly expressed circRNAs present across many individuals.

We identified 96 SMC circRNAs associated with genetic loci, which we term a circQTL. Other studies have performed circQTL analyses in lymphoblastoid cell lines[36] and dorsolateral prefrontal cortex samples.[37] One previous study had identified that *circANRIL* had a circQTL in atherosclerotic plaques.[8] Due to stringent expression threshold cutoffs, *circANRIL* was removed from our analysis. This implies that the *circANRIL* QTL detected in atherosclerotic plaques could be due to *circANRIL* expression in *in vivo* SMCs, a SMC phenotypic state that was not represented in our cultures, or cells other than SMCs.

A unified source for genome-wide circRNA annotations has yet to be established. To derive functional insight for our circRNAs, we assumed that a circRNA has a related function to its parent gene and performed gene ontology enrichment using annotation of the parent genes. Parent genes of the differentially expressed circRNAs were enriched in processes related to blood pressure and migration: regulation of sodium ion transmembrane transport, cell polarity, and substrate adhesion-dependent cell spreading. The circQTL parent genes were also enriched for processes related to blood pressure and migration: Rho protein signal transduction, sodium ion transmembrane transporter activity, and cytoplasmic microtubule organization. Overall, these results suggest that circRNAs play an important role in SMC processes and vascular disease traits. Previous studies have also demonstrated that circRNAs are relevant to SMC biology. For example, they showed that *circACTA2*[9] and *circSFMBT2*[11] may reduce SMC contraction. Although we

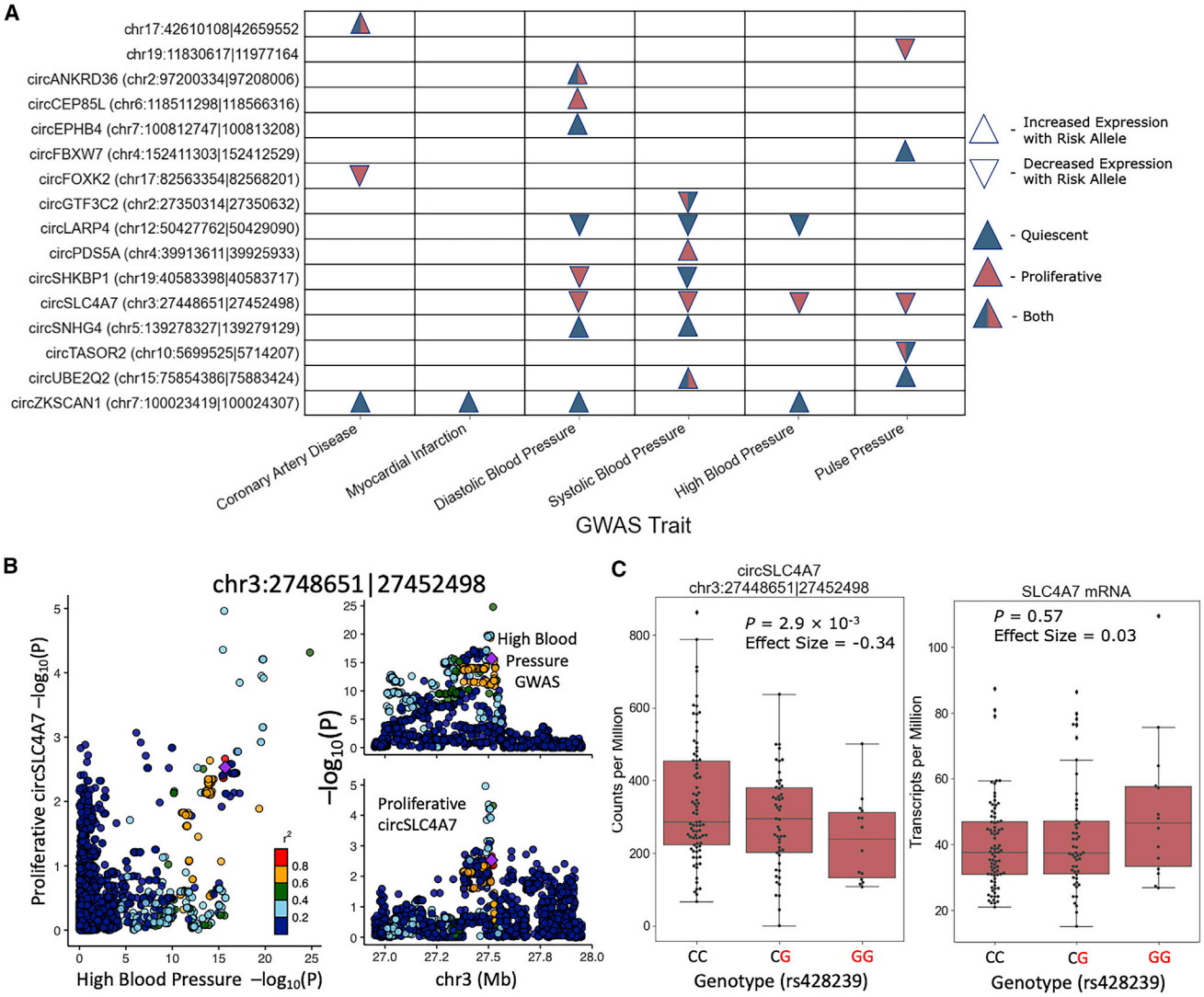

Figure 7. Cardiovascular GWAS loci associated with circRNA expression

(A) Heatmap showing circQTL colocalized with cardiovascular GWAS loci across proliferative and quiescent SMCs using eCAVIAR and COLOC.

(B) circSLC4A7 circQTL signal colocalizes with the high blood pressure GWAS. The colocalized SNP, rs428239, is highlighted with a purple diamond.

(C) The risk allele, G, of rs428239 is associated with lower expression of circSLC4A7 but is not associated with mRNA SLC4A7 expression. P-value (*P*) was calculated using tensorQTL, which performs linear regression between the indicated SNP genotype and circRNA abundance with a permutation scheme.

detected *circACTA2* and *circSFMBT2*, these circRNAs were removed after stringent expression threshold cutoffs. Previous studies demonstrated that overexpression of *circMAP3K5*[12] and *circLRP6*[10] inhibited the proliferation of SMCs and could play a role in vascular pathogenesis. In our study, we did not find *circMAP3K5* and *circLRP6* differentially expressed between quiescent and proliferative cells.

To further characterize genetic architecture of circRNA expression in SMCs, we identified phenotype-specific genetic regulation of *circSTRN3*, *circCHD9*, and *circRIPK1* expression where the effect size of eSNPs associated with circRNA expression differed between proliferative and quiescent cells. This is analogous to our previous studies that identified condition-specific effects of genetic regulation on mRNA expression in SMCs.[15] Although further studies will be needed, phenotype-specific genetic regulation of circRNA could be explained by transcription factor presence in one condition but not the other. Previous studies have identified transcription factor binding of DNA motifs that increases expression of circRNAs but not the host gene.[38] In addition, we identified six SNPs that had distinct association directions with circRNA isoforms from the same gene. This could be explained by previous studies that found that competition of RNA pairing of inverted sequences surrounding backsplicing sites could affect circRNA splicing selection.[39]

To uncover the relationships between circRNAs and their parent genes, we compared circRNA and mRNA SMC expression profiles. Surprisingly, there has been little consensus

over the relationship between the abundance of a circRNA and the mRNA of its parent gene. Studies have suggested a strong negative correlation,[39] no correlation,[40] or strong positive correlation.[38,41] Our study supports a low positive correlation. To further investigate this discrepancy, we showed that there was a strong correlation in fold change of circRNA expression between quiescent and proliferative SMCs and the fold change in parent gene expression. This supports the hypothesis that in response to stimuli, circRNA and its parent gene are regulated together; however, we also observed that 41/164 (25%) of differentially expressed circRNAs showed opposite fold change directions than their parent gene. These differences in correlation suggest distinct mechanisms of circRNA formation across genes. Although still under study, distinct mechanisms of circRNA biogenesis have been identified.[42] circRNAs formed under mechanisms involving competition between linear splicing and backsplicing of exons would support a negative correlation, while circRNAs formed as a byproduct from excised introns or exons would support a positive correlation.[42]

To further explore the relationship between circRNAs and their parent genes, we compared the genetic regulation of circRNA with the genetic regulation of mRNA. We found that a circQTL locus is usually not associated with mRNA expression of the parent gene. Two previous circQTL studies showed that the effect direction of a SNP on a circRNA expression is mostly concordant with the effect size of the same SNP on the parent gene expression.[36,37] Overall, these results suggest that circRNAs and mRNAs have distinct methods of expression regulation. Previous studies have suggested that SNPs associated with circRNAs are within inverted sequences around a backsplicing site or canonical splicing sites that play crucial roles in the biogenesis of circRNAs.[37]

Lastly, we utilized a colocalization approach to annotate the disease relevance of SMC circRNAs. Our previous study[15] identified 58 and 127 CAD loci that were mediated by changes in gene expression and alternative splicing, respectively. In this study, we demonstrated that circRNAs also contribute to the genetic architecture of complex traits by colocalizing circQTL loci with vascular-related GWAS loci. We identified 14 circQTL loci—including *circZKSCAN1*—that overlap with blood pressure GWAS loci. In addition, we identified that the *circZKSCAN1* locus was associated with myocardial infarction and CAD. A previous study showed that silencing of *circZKSCAN1* promoted proliferation, migration, and invasion of hepatocellular carcinoma cell lines.[41] We also identified that *circFOXK2* and *chr17:42610108|42659552* were associated with CAD. *CircFOXK2* was previously shown to promote growth and metastasis of pancreatic ductal adenocarcinoma.[42] The *chr17:42610108|42659552* backsplicing site is in an intergenic region between *TUBG* and *RETREG3*.

Utilizing colocalization approaches, our study identified putative circRNAs relevant for vascular-related diseases. Our results may include false positives due to small sample size, limitations in colocalization approaches,[43] and our lenient p value/posterior probability cutoffs. While culturing SMCs in artificial conditions may reduce physiological relevance, our culturing approach controls for confounders involving heterogeneous cell types/states across donors. Future in depth studies of these circRNAs would be needed to identify the molecular and cellular mechanisms. A relevant first step would be to identify the exact sequence of the circRNAs and not just the backsplicing site.

Overall, our results suggested that SMC circRNAs can play a role in the genetic risk for higher blood pressure, myocardial infarction, and CAD. Notably, for all circQTLs colocalized with vascular-related GWAS loci, we identified genetic regulation specific to circRNA expression but not mRNA expression of its parent gene. This suggests that circRNAs explain an additional part of the genetic architecture of disease that is missed when looking at just mRNA gene expression.

Data and code availability

The datasets generated during this study are available at GEO with the accession number GSE193817. circQTL summary statistics for both SMC phenotypes are available at https://virginia.box.com/s/6731ysx3mgfsxhnrejchyd908q2scu9x. We are not able to share genotype data for our donors. Code can be obtained from the corresponding author.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100164.

Acknowledgments

This work was supported by an American Heart Association Postdoctoral Fellowship 18POST33990046 (to R.A.) and Transformational Project Award 19TPA34910021 (to M.C.).

Declaration of interests

The authors declare that they have no competing interests.

Received: June 24, 2022
Accepted: November 23, 2022

References

1. Basatemur, G.L., Jørgensen, H.F., Clarke, M.C.H., Bennett, M.R., and Mallat, Z. (2019). Vascular smooth muscle cells in atherosclerosis. Nat. Rev. Cardiol. *16*, 727–744.
2. Zhuge, Y., Zhang, J., Qian, F., Wen, Z., Niu, C., Xu, K., Ji, H., Rong, X., Chu, M., and Jia, C. (2020). Role of smooth muscle cells in Cardiovascular Disease. Int. J. Biol. Sci. *16*, 2741–2751.
3. Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.
4. Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 is a circular RNA that can Be translated and functions in myogenesis. Mol. Cell *66*, 22–37.e9.

5. Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol. Cell *66*, 9–21.e7.
6. Militello, G., Weirick, T., John, D., Döring, C., Dimmeler, S., and Uchida, S. (2017). Screening and validation of lncRNAs and circRNAs as miRNA sponges. Briefings Bioinf. *18*, 780–788.
7. Enuka, Y., Lauriola, M., Feldman, M.E., Sas-Chen, A., Ulitsky, I., and Yarden, Y. (2016). Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. Nucleic Acids Res. *44*, 1370–1383.
8. Holdt, L.M., Stahringer, A., Sass, K., Pichler, G., Kulak, N.A., Wilfert, W., et al. (2016). Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. Nat. Commun. *7*, 1–14.
9. Sun, Y., Yang, Z., Zheng, B., Zhang, X.-H., Zhang, M.-L., Zhao, X.-S., Zhao, H.-Y., Suzuki, T., and Wen, J.-K. (2017). A novel regulatory mechanism of smooth muscle α-actin expression by NRG-1/circACTA2/miR-548f-5p Axis. Circ. Res. *121*, 628–635.
10. Hall, I.F., Climent, M., Quintavalle, M., Farina, F.M., Schorn, T., Zani, S., Carullo, P., Kunderfranco, P., Civilini, E., Condorelli, G., and Elia, L. (2019). Circ_Lrp6, a circular RNA enriched in vascular smooth muscle cells, acts as a sponge regulating miRNA-145 function. Circ. Res. *124*, 498–510.
11. Luo, Y., and Huang, C. (2021). CircSFMBT2 facilitates vascular smooth muscle cell proliferation by targeting miR-331-3p/HDAC5. Life Sci. *264*, 118691.
12. Zeng, Z., Xia, L., Fan, S., Zheng, J., Qin, J., Fan, X., Liu, Y., Tao, J., Liu, Y., Li, K., et al. (2021). Circular RNA CircMAP3K5 acts as a MicroRNA-22-3p sponge to promote resolution of intimal hyperplasia via TET2-mediated smooth muscle cell differentiation. Circulation *143*, 354–371.
13. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.
14. Consortium, T.G. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.
15. Aherrahrou, R., Lue, D., Perry, R.N., Aberra, Y.T., Khan, M.D., Soh, J.Y., Ord, T., Singha, P., Gilani, H., Benavente, E.D., et al. (2022). Genetic regulation of human aortic smooth muscle cell gene expression and splicing predict causal coronary artery disease genes. Preprint at bioRxiv. 2022.01.24.477536.
16. Aherrahrou, R., Guo, L., Nagraj, V.P., Aguhob, A., Hinkle, J., Chen, L., Yuhl Soh, J., Lue, D., Alencar, G.F., Boltjes, A., et al. (2020). Genetic regulation of atherosclerosis-relevant phenotypes in human vascular smooth muscle cells. Circ. Res. *127*, 1552–1565.
17. Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol. *16*, 4.
18. Gao, Y., Zhang, J., and Zhao, F. (2018). Circular RNA identification based on multiple seed matching. Briefings Bioinf. *19*, 803–810.
19. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
20. McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.
21. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.
22. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA *100*, 9440–9445.
23. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507.
24. Taylor-Weiner, A., Aguet, F., Haradhvala, N.J., Gosai, S., Anand, S., Kim, J., Ardlie, K., Van Allen, E.M., and Getz, G. (2019). Scaling computational genomics to millions of individuals with GPUs. Genome Biol. *20*, 228.
25. Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., et al. (2017). Giangreco N., et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat. Commun. *8*, 1–10.
26. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
27. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL signals detects target genes. Am. J. Hum. Genet. *99*, 1245–1260.
28. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012.
29. Alexa, A., and Rahnenfuhrer, J. Gene Set Enrichment Analysis with topGO. 26.
30. Das, A., Das, D., and Panda, A.C. (2022). Validation of circular RNAs by PCR. In PCR Primer Design, C. Basu, ed. (Springer US)), pp. 103–114.
31. Boedtkjer, E., Praetorius, J., Matchkov, V.V., Stankevicius, E., Mogensen, S., Füchtbauer, A.C., Simonsen, U., Füchtbauer, E.M., and Aalkjaer, C. (2011). Disruption of $Na+$, HCO_3^- cotransporter NBCn1 (slc4a7) inhibits NO-mediated vasorelaxation, smooth muscle Ca^{2+} sensitivity, and hypertension development in mice. Circulation *124*, 1819–1829.
32. Altesha, M.-A., Ni, T., Khan, A., Liu, K., and Zheng, X. (2019). Circular RNA in cardiovascular disease. J. Cell. Physiol. *234*, 5588–5600.
33. van der Harst, P., and Verweij, N. (2018). Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ. Res. *122*, 433–443.
34. Tian, J., Fu, Y., Li, Q., Xu, Y., Xi, X., Zheng, Y., Yu, L., Wang, Z., Yu, B., and Tian, J. (2020). Differential expression and bioinformatics analysis of CircRNA in PDGF-BB-induced vascular smooth muscle cells. Front. Genet. *11*, 530.
35. Chen, W., Lin, J., Li, B., Cao, S., Li, H., Zhao, J., Liu, K., Li, Y., Li, Y., and Sun, S. (2020). Screening and functional prediction of differentially expressed circRNAs in proliferative human aortic smooth muscle cells. J. Cell Mol. Med. *24*, 4762–4772.
36. Ahmed, I., Karedath, T., Al-Dasim, F.M., and Malek, J.A. (2019). Identification of human genetic variants controlling circular RNA expression. RNA *25*, 1765–1778.

37. Liu, Z., Ran, Y., Tao, C., Li, S., Chen, J., and Yang, E. (2019). Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. Genome Biol. *20*, 99.
38. Meng, J., Chen, S., Han, J.-X., Qian, B., Wang, X.-R., Zhong, W.-L., Qin, Y., Zhang, H., Gao, W.-F., Lei, Y.-Y., et al. (2018). Twist1 regulates vimentin through Cul2 circular RNA to promote EMT in hepatocellular carcinoma. Cancer Res. *78*, 4150–4162.
39. Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. Cell *159*, 134–147.
40. Chen, L.-L. (2016). The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol. *17*, 205–211.
41. Yao, Z., Luo, J., Hu, K., Lin, J., Huang, H., Wang, Q., Zhang, P., Xiong, Z., He, C., Huang, Z., et al. (2017). ZKSCAN1 gene and its related circular RNA (circZKSCAN1) both inhibit hepatocellular carcinoma cell growth, migration, and invasion but through different signaling pathways. Mol. Oncol. *11*, 422–437.
42. Wong, C.H., Lou, U.K., Li, Y., Chan, S.L., Tong, J.H., To, K.-F., and Chen, Y. (2020). CircFOXK2 promotes growth and metastasis of pancreatic ductal adenocarcinoma by complexing with RNA-binding proteins and sponging MiR-942. Cancer Res. *80*, 2138–2149.
43. Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H.K., and Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. Am. J. Hum. Genet. *108*, 25–35.