



Stroke-associated intergenic variants modulate a human *FOXF2* transcriptional enhancer

Jae-Ryeon Ryu^{a,b}, Suchit Ahuja^{a,b}, Corey R. Arnold^{a,b}, Kyle G. Potts^{a,c,d}, Aniket Mishra^e, Qiong Yang^{f,g}, Muralidharan Sargurupremraj^{f,h,i}, Douglas J. Mahoney^{a,c,d}, Sudha Seshadri^{f,h,i}, Stéphanie Debette^{a,f,j}, and Sarah J. Childs^{a,b,1}

Edited by David Raible, University of Washington School of Medicine, Seattle, WA; received November 23, 2021; accepted July 28, 2022 by Editorial Board Member Mary-Claire King

SNPs associated with human stroke risk have been identified in the intergenic region between Forkhead family transcription factors *FOXF2* and *FOXQ1*, but we lack a mechanism for the association. FoxF2 is expressed in vascular mural pericytes and is important for maintaining pericyte number and stabilizing small vessels in zebrafish. The stroke-associated SNPs are located in a previously unknown transcriptional enhancer for *FOXF2*, functional in human cells and zebrafish. We identify critical enhancer regions for *FOXF2* gene expression, including binding sites occupied by transcription factors ETS1, RBPJ, and CTCF. rs74564934, a stroke-associated SNP adjacent to the ETS1 binding site, decreases enhancer function, as does mutation of RPBj sites. rs74564934 is significantly associated with the increased risk of any stroke, ischemic stroke, small vessel stroke, and elevated white matter hyperintensity burden in humans. *Foxf2* has a conserved function cross-species and is expressed in vascular mural pericytes of the vessel wall. Thus, stroke-associated SNPs modulate enhancer activity and expression of a regulator of vascular stabilization, *FOXF2*, thereby modulating stroke risk.

stroke | *FOXF2* | pericyte

Stroke is a leading cause of death and disability, caused by an interruption of blood flow to the brain that results in neural damage. Even though the cause of some rare inherited forms of stroke has been determined, genetic risk factors leading to most stroke types are unknown. Genetic susceptibility to stroke in the general population has been explored through genome-wide association studies (GWAS). Single-nucleotide polymorphisms (SNPs) that associate with higher stroke risk have been identified in individuals of multiple ancestries (1–4), but mechanistic insight as to how risk SNPs lead to stroke is lacking. GWAS-identified SNPs are often located in the noncoding genome, in regions without functional annotation. The lead SNP (the SNP having the strongest statistical association with the disease) is not necessarily the causal variant, and there are often many SNPs in a contiguous region of linkage disequilibrium (LD) surrounding the lead SNP. Which of these individual SNPs mediates risk, and how SNPs modulate disease at the DNA, protein, and cellular levels, is not obvious. One hypothesis is that stroke-associated SNPs may modify the transcription of genes essential for the stability of the vascular wall, either directly through affecting transcription factor binding sites, or indirectly, by affecting three-dimensional genome structure.

The vascular wall comprises an endothelial cell lining surrounded by vascular mural cells (pericytes and vascular smooth muscle cells [vSMCs]) that contribute extracellular matrix (ECM) to ensure support of the blood vessel. Endothelial and mural cells signal to one another to maintain quiescence. Pericytes support the smaller vessels of the brain, while vSMCs surround larger vessels. Both cells also modulate contractility of brain vessels. Critical pathways required for cerebral vascular stability include the platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF), Notch, transforming growth factor- β , Wnt, and angiopoietin pathways. In particular, mutations in the Notch receptor *NOTCH3*, basement membrane collagens *COL4A1* or *COL4A2*, serine peptidase *HTRA1*, transcription factor *FOXC1*, transcription and export protein *TREX*, and alpha galactosidase *GLA* lead to inherited cerebral small vessel disease (CSVD), a cause of 20% of ischemic strokes (5, 6). CSVD affects the microvessels of the brain that are stabilized by pericytes, and is particularly devastating as it is a major contributor to compromised blood–brain barrier, cognitive decline, dementia, and stroke.

While extrinsic signaling pathways are well studied in vascular stabilization, the roles of intrinsic transcription factors are less well characterized. Forkhead box transcription factors *FOXF2* and *FOXC1* are associated with vascular mural cell development (1, 7–10). *FOXC1* is required for basement membrane deposition and is associated with CSVD (8–11). *FOXC1* is dosage sensitive, as mutation, deletion, duplication, or SNPs

Significance

An increased risk of stroke is associated with the presence of high-risk genetic variants in the intergenic region between *FOXF2* and *FOXQ1*. We show that some variants are present in a short genomic region that acts as a transcriptional enhancer of *FOXF2*, a gene that stabilizes the blood vessel wall. We identify transcription factors that bind to the enhancer to drive activity. Activity of the enhancer is significantly decreased when high-risk versus low-risk variants are present. Thus, SNPs in a *FOXF2* enhancer modulate the fine levels of *FOXF2* gene transcription, which, in turn, stabilizes the vascular wall, preventing stroke.

Author contributions: J.-R.R., S.A., C.R.A., and S.J.C. designed research; J.-R.R., S.A., and C.R.A. performed research; J.-R.R., K.G.P., A.M., Q.Y., M.S., D.J.M., S.S., S.D., and S.J.C. contributed new reagents/analytic tools; J.-R.R., S.A., A.M., Q.Y., M.S., and S.J.C. analyzed data; and S.A. and S.J.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. D.R. is a guest editor invited by the Editorial Board.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: schilds@ucalgary.ca.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2121333119/-DCSupplemental>.

Published August 22, 2022.

within a single allele in humans lead to CSVD (8). *FOXQ1* is part of a conserved FOX cluster of genes, *FOXQ1*, *FOXF2*, and *FOXQ1*, linked within 300 kb. A small number of patients with a contiguous deletion of *FOXF2* and *FOXQ1* show an increase in MRI white matter hyperintensities, a marker of neuronal damage, over patients with *FOXQ1* alone deleted, suggesting that the loss of *FOXF2* could cause or enhance CSVD (1), but a mechanistic basis has not been determined. *Foxf2* is expressed in pericytes of the brain vasculature and developing vSMCs in mouse and fish, with some expression in mouse endothelial cells (1, 12). Loss of *Foxf2* in mice leads to hemorrhage, changes in pericyte number, and defects in blood–brain barrier formation (12). Loss of one or two zebrafish *foxf2* genes leads to decreased smooth muscle coverage of developing cerebral vessels (1).

Common variants leading to stroke in the general population have been identified in an intergenic region between *FOXF2* and *FOXQ1* (1, 3, 13). The lead SNP in the article by Chauhan et al., rs12204590 is associated with the risk of all types of stroke (1). A total of 38 additional SNPs with an $r^2 > 0.5$ with the lead SNP were identified spanning a 4.7-kb region. This genomic region has no known function, but spans a DNase hypersensitive region, with H3K27 acetylation peaks and is predicted to bind a large number of transcription factors, hallmarks of an enhancer. Genetic association of stroke with this genomic region was confirmed and expanded in additional cohorts (3, 13).

Here, we establish the function of stroke-associated SNPs adjacent to *FOXF2* using in vitro and in vivo experiments. We show that the SNPs are located within a strong transcriptional enhancer for human *FOXF2*. We identify three transcription factors critical for *FOXF2* activation in this region, ETS proto-oncogene1 (ETS1), recombination signal binding protein for immunoglobulin kappa J region (RBPJ) and CCCTC-binding factor (CTCF). We show that Notch signaling is functionally important for *FOXF2* expression in vitro and in vivo in zebrafish. We demonstrate a mechanistic link between a SNP in an ETS1 binding site that modulates enhancer activity, is associated with stroke, its clinical subtypes, and relevant MRI markers in humans.

Results

Foxf2 Promotes Pericyte Number and Vascular Stabilization.

The role of FoxF2 in zebrafish pericyte biology is not completely clear. We previously showed that mutation of *foxf2b*, one of the two human *FOXF2* orthologs, leads to decreased cerebral smooth muscle, but found no significant effects on vascular stabilization, potentially because of compensation from the *foxf2* paralog (1). To observe the full loss of function phenotypes, we created *foxf2a* mutants. *foxf2a^{ca71}* has a 11-bp deletion leading to a premature truncation after 78 amino acids, expected to produce no functional protein. We found a significant decrease in the number of brain pericytes in *foxf2a^{-/-};**foxf2b^{-/-}* double mutants versus *foxf2a^{+/+};**foxf2b^{+/+}* controls (average of 25 pericytes in the brains of controls that is reduced to 14 in double mutants, $P = 0.01$; *SI Appendix, Fig. S1*). Decreased pericyte number is strongly correlated with decreased vascular stability in mouse and zebrafish models (14). Correspondingly, we found that 6% of *foxf2a^{+/+};**foxf2b^{-/-}* mutants and 10% of *foxf2a^{+/-};**foxf2b^{-/-}* mutants have brain hemorrhage, and this is significantly increased to 36% in *foxf2a^{-/-};**foxf2b^{-/-}* double mutants ($P < 0.02$ and 0.03 respectively; *SI Appendix, Fig. S1*). These data suggest that Foxf2 controls vascular stabilization via pericytes in zebrafish.

Stroke-Associated Risk SNPs Are Located in a Transcriptional Enhancer. GWASs identify a strong region of risk linked with all stroke types on human chromosome 6, in an intergenic region of unknown function. A total of 38 variants in LD ($r^2 > 0.5$) with the lead variant rs12204590 were identified, many of which were also identified by a second study (1, 3) (*SI Appendix, Fig. S2*). The variants associated with “all stroke,” “ischemic stroke,” and “small vessel disease” are located between *FOXF2* and *FOXQ1* as shown by locus zoom plots (*SI Appendix, Fig. S3*); both genes are expressed by vascular mural cells (Fig. 1A). The regulatory potential of this genomic region has not been studied. Pericytes and vSMCs are differentiated cells that can only be cultured as primary cells, which are not ideal for enhancer studies. As a first step to understand *FOXF2* regulation, we used HEK293 cells that endogenously express high levels of *FOXF2* (*SI Appendix, Fig. S4*) as a model to determine (1) whether the GWAS-identified region could act as a transcriptional enhancer, (2) whether silencing or deletion of the enhancer affects *FOXF2* expression, (3) which transcription factors and pathways control expression from this region, and (4) whether variants affect expression of *FOXF2* from this enhancer.

We cloned construct A, a ~5-kb region including lead variants rs12204590 and rs4959130 and other risk SNPs as well as ~2-kb construct D, derived from expanded GWAS analysis (Fig. 1B and C (3)). We confirmed that HEK cells are homozygous for major (low-risk) SNP alleles in this region. Subconstructs 3.4-kb B and 1.6-kb C were cloned to narrow down critical regions of function within A, and 0.75-kb E and 1.3-kb F are derived from D. Locations of H3K27Ac peaks and candidate regulatory elements, H3K27 acetylation, DNase hypersensitivity, and transcription factor binding on ENCODE were used to choose subconstructs. Constructs were cloned upstream of a minimal promoter and the luciferase reporter. Transcriptional activation was determined by luciferase assay. Constructs A, B, and C elicit an average 5.5-, 4.8-, and 4.4-fold increase in luciferase activity over background, respectively (Fig. 1D). Furthermore, construct A activity is not significantly different in forward and reverse configurations. Construct D had only two-fold activity, while subconstruct E had the highest transcriptional enhancer activity of any tested (11.3-fold). In contrast, and construct F (construct D that omits E) had only threefold activity. We thus identify a transcriptional enhancer within the genomic region identified by GWAS and focus on construct E for further study as it had the highest transcriptional activation.

As a test of enhancer activity in vivo, we expressed constructs D through F in zebrafish. Human constructs were cloned upstream of a minimal promoter and green fluorescent protein (GFP) in a Tol2 transposon vector for expression. Interestingly, constructs D and E drive the expression of GFP in zebrafish perivascular cells (PVCs) in a high proportion of transgenic animals (71 and 73%, respectively), showing cross-species interpretation of the enhancer (Fig. 1F and *SI Appendix, Fig. S5*). There were no statistically significant differences in activity between construct D and E in vivo (chi-square test), although no perivascular expression was seen from construct F ($P < 0.001$), which is not surprising given that it lacks the enhancer elements. Thus, a human *FOXF2* enhancer functions as an enhancer in an evolutionary distant vertebrate, with expression in PVCs.

High-Risk SNPs Decrease Transcriptional Output of the Enhancer.

We predicted that the SNP haplotype may affect enhancer activity. Human DNA corresponding to constructs A, B, C, and E was cloned from the major (low-risk) and minor allele (high-risk)

Identification of Enhancer Regions Promoting Endogenous *FOXF2* Expression.

To identify the critical sequences for enhancer function, we used temporally controlled inducible CRISPR interference (iCRISPRi). Guide RNAs direct a catalytically dead Cas9 (dCas9) fused to the Krüppel-associated box (KRAB) domain to specific genomic sequences. dCas9-KRAB acts as a transcriptional repressor by inducing local, transient heterochromatin formation (15). We designed guides to target putative ETS1 and RBPJ sites within construct E, and an adjacent CTCF site (Fig. 2 *A* and *B*). Changes in *FOXF2* mRNA were detected by qRT-PCR. Compared to empty guide vector-transfected controls, iCRISPRi guides targeting the ETS1 sites significantly reduce *FOXF2* expression to 66% of control ($P < 0.001$, Fig. 2*D*), while guides targeting the CTCF site reduce expression to 62% of control ($P < 0.001$). However, guides to RBPJ-1 or RBPJ-2 do not change *FOXF2* expression ($P < 0.24$). As a positive control, guides directly targeting the *FOXF2* exon reduce the expression of endogenous *FOXF2* to 39% of wild type ($P < 0.0001$). We confirmed that iCRISPRi did not change the expression of the closely linked *FOXC1*, indicating that the enhancer specifically promotes endogenous *FOXF2* expression (Fig. 2*E*).

CTCF, RBPJ, and ETS1 Proteins Bind to the *FOXF2* Intergenic Enhancer.

ENCODE data from multiple cell types suggests that CTCF, ETS1, and RBPJ bind the enhancer (Fig. 3*A*). RBPJ sites are frequently located near CTCF sites (16), and ETS1 and RBPJ can physically bind each other, suggesting that physical interactions between these factors may be critical for enhancer

function (17–19). Interestingly, ENCODE data show RBPJ binding strongly over the canonical ETS1 binding sequence, but weakly to canonical RBPJ binding sites (RBPJ1 and -2) located 100 to 200 bp away from this site (Fig. 3*A*).

We used chromatin immunoprecipitation qPCR (ChIP-qPCR) with antibodies against CTCF, RBPJ, and ETS1. Negative controls for all of the experiments included glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and the human SAT2 repeat element. We included a positive control of a known RBPJ-binding site near HES1. Fig. 3 shows data for biological replicate 1, while replicates 2 and 3 are found in *SI Appendix, Fig. S8*. We show that ETS1 binds to the predicted ETS1 site (Fig. 3*B*), RBPJ binds to the RBPJ1 and -2 sites, and to the positive control HES1 (Fig. 3 *C*, *E*, and *F*), and CTCF binds to its predicted site in HEK cells (Fig. 3*D*). Thus, the predicted transcription factor binding sites in the enhancer are occupied by their cognate transcription factors.

As the transcription factor RBPJ is a transducer of Notch, we tested whether Notch activity modulates *FOXF2* expression in human cells using two γ -secretase inhibitors (Fig. 3*E*). DAPT 20 μ M or LY411575 reduced human *FOXF2* expression to 42% ($P = 0.007$) and 47% ($P = 0.0005$) of untreated cells, respectively, demonstrating that Notch activity positively regulates *FOXF2* expression (Fig. 3*F*). Zebrafish *foxf2b* is reduced ~50% by LY411575 and 40% by DAPT at 72 hpf ($P < 0.0028$ and $P < 0.010$, respectively) (Fig. 3*G*). Thus, Notch signaling positively regulates *FOXF2* in two species, which is consistent with occupied RBPJ sites in the *FOXF2* intergenic enhancer.

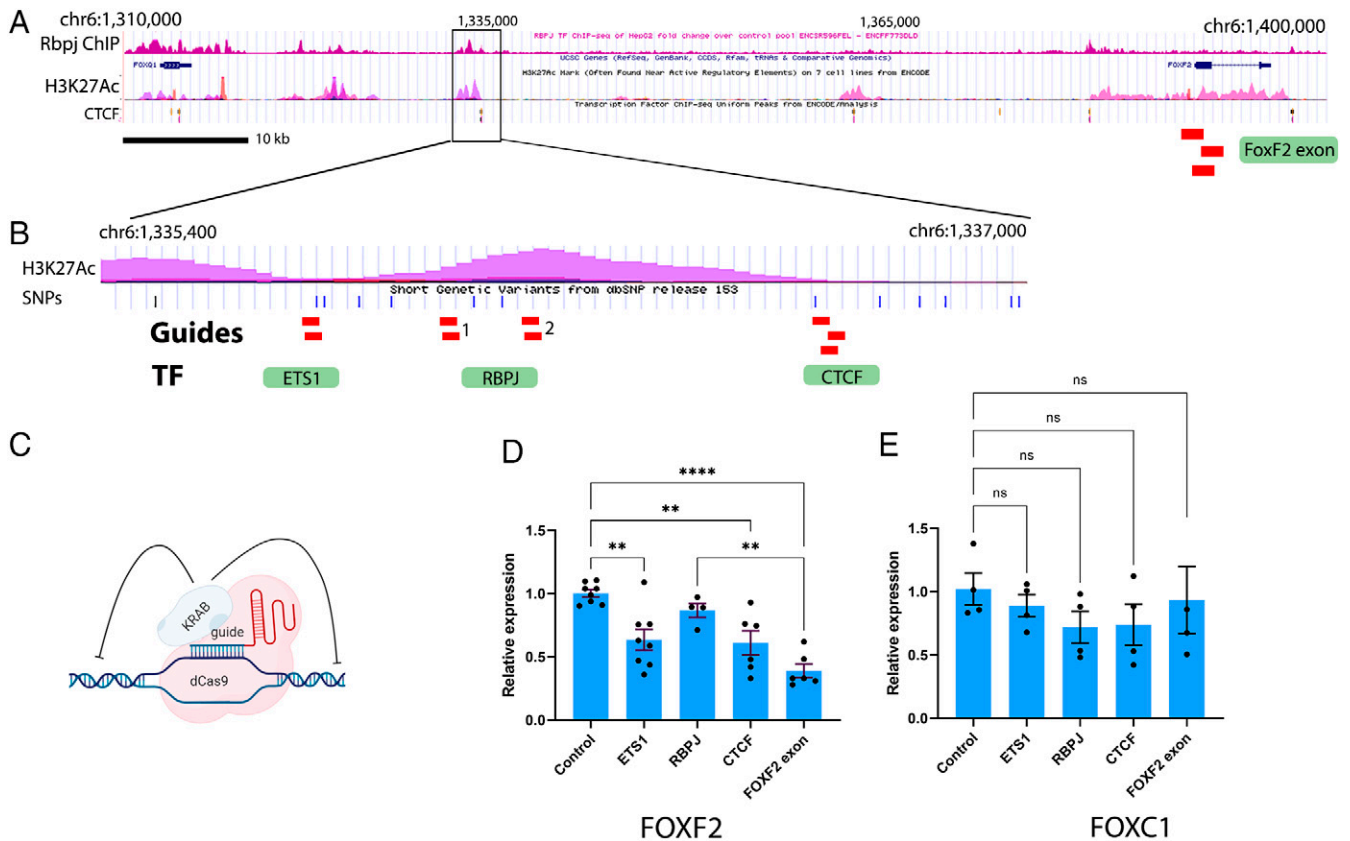


Fig. 2. iCRISPRi targeting of sites in the intergenic enhancer reduces *FoxF2* transcription. (A) Low-resolution map of the region containing the SNPs indicating ENCODE ChIP-seq peaks for the transcription factors ETS1, RBPJ, and CTCF. (B) Higher resolution view of the 750-bp enhancer (E) and locations of CRISPR guides (red) centered over putative sites of transcription factor binding. Note the 3 control sites in the *FoxF2* exon in (A). (C) Schematic of the iCRISPRi dead Cas9 (dCas9) fused to a KRAB transcriptional repressor and resulting repression of local gene expression. (D and E) Relative expression of (D) *FOXF2* and (E) *FOXC1* as determined by qPCR after iCRISPRi. (ns = not significant, $**P < 0.01$, $****P < 0.0001$ by 1-way ANOVA with Tukey's post hoc test; *FOXF2*: 3 technical replicates of 7 to 11 biological replicates; *FOXC1*: 3 technical replicates of 3 biologically independent replicates).

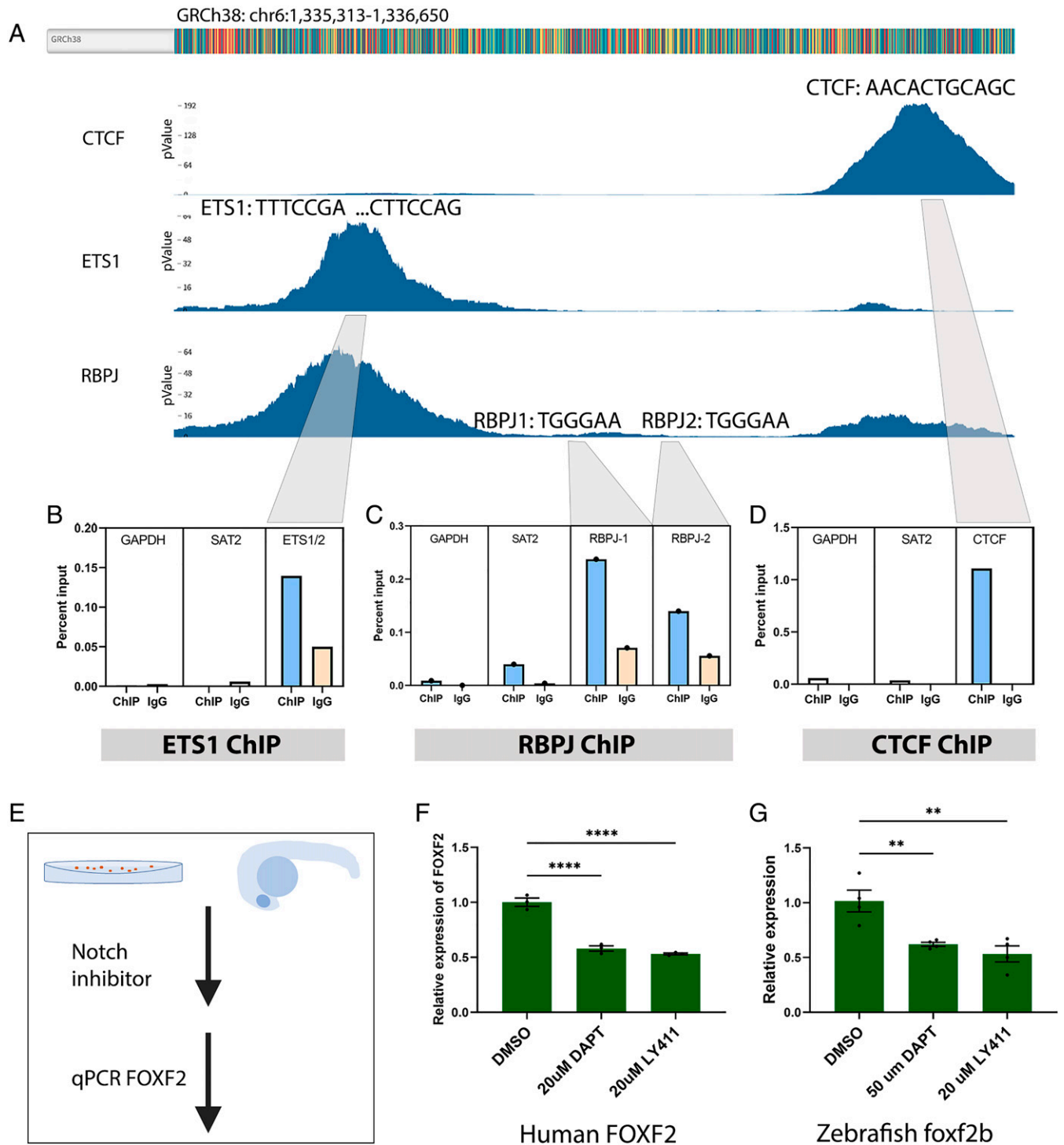


Fig. 3. RBPJ, CTCF, and ETS1 bind the FOXF2 enhancer. (A) Chromosome-level map and ChIP-seq peaks of CTCF, ETS1, and RBPJ from ENCODE marked with consensus transcription factor binding sequences under each peak. Note that the RBPJ peak that coincides with the ETS1 peak does not have RBPJ consensus binding sequences; however, there are adjacent RBPJ sites (1, 2) a short distance away. (B–D) Representative ChIP-qPCR analysis of binding of antibodies to (B) ETS1, (C) RBPJ, or (D) CTCF (blue bars) or immunoglobulin G control (orange bars) from a single experiment (replicates are in *SI Appendix, Fig. S4*). GAPDH and SAT2 are negative controls. Gray lines indicate site of PCR product. (E) Schematic of Notch inhibitor experiment. (F) Notch inhibitor treatment reduces expression of (F) human FOXF2 and (G) zebrafish *foxf2b* at 75 hpf using the indicated inhibitors and doses. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.001$ by 1-way ANOVA with Tukey's post hoc test; human cells, $n = 3$ biological replicates; zebrafish, $n = 4$ biological replicates.

Single SNPs Affect FOXF2 Enhancer Activity. To directly test an enhancer role for the region containing ETS1 and RBPJ1 sites, we used two CRISPR guides to delete a ~250-bp fragment in cells (Δ enhancer). As a control, we ablated FOXF2 transcription by deleting the promoter and exon1 coding sequence (Δ exon; Fig. 4C and *SI Appendix, Fig. S9*). Δ Enhancer cells express FOXF2 at 17% of control ($P < 0.015$; Fig. 4A), confirming

that this short region is an essential FOXF2 enhancer. Δ Exon cells have no detectable FOXF2 expression at all ($P < 0.0006$; Fig. 4B). As expected, FOXF2 is not changed by either deletion, while FOXQ1 is unaffected by Δ enhancer, but strongly reduced by Δ exon. This suggests that while the enhancer specifically regulates the transcription of FOXF2, transcription of FOXF2 regulates FOXQ1 likely through direct transcriptional activation via

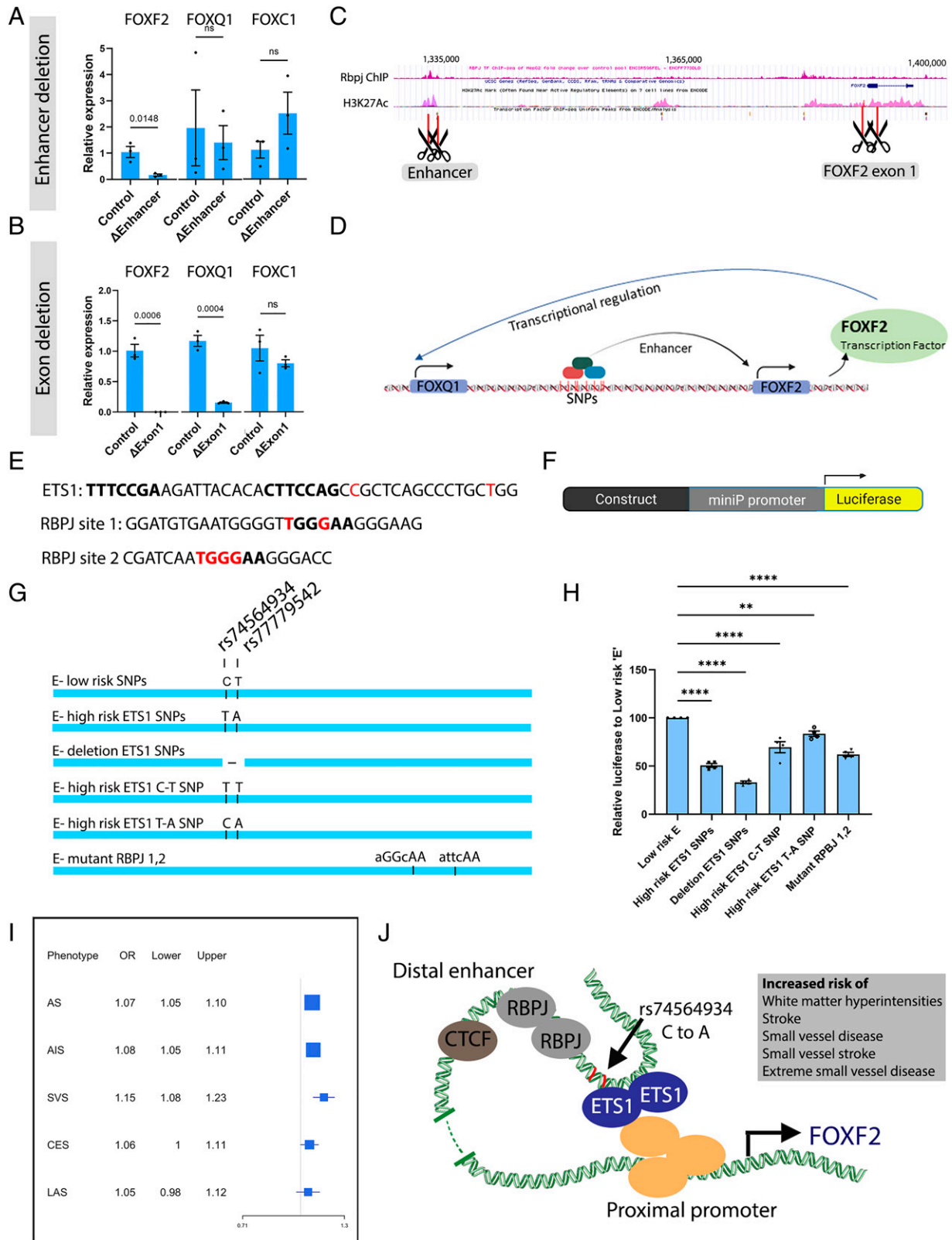


Fig. 4. The FOXF2 enhancer is critical for FOXF2 transcription; risk variants adjacent to ETS1 decrease enhancer activity. (A and B) Relative expression of FOXF2, FOXQ1, and FOXC1 after targeted excision of the FOXF2 (A) enhancer or (B) exon 1. P values are from a Student's *t* test for each gene pair (control and Δenhancer), 3 biological replicates. (C) Schematic of the locus, marking the CRISPR targets for enhancer and exon 1. (D) Model of regulatory relationships based on enhancer and exon 1 deletion. The enhancer regulates FOXF2, while FOXF2 transcription regulates FOXQ1 and LINC01394. (E) Sequences of mutated enhancer fragments. Boldface lettering indicates transcription factor binding sites, while red indicates SNPs. (F) Schematic of the luciferase construct used to test expression. (G) Maps of mutated constructs derived from construct E with the following modifications: low-risk SNPs (control), 2 high-risk SNPs, single high-risk C-T SNP, or single high-risk T-A SNP mutated, with high-risk SNPs both deleted, and a construct with 2 putative RBPJ binding sites mutated to sequences not predicted to bind RBPJ. (H) Luciferase activity reporting transcriptional activation of the constructs in B (***P* < 0.01, *****P* < 0.0001 by ANOVA and Dunnett's post hoc test, *n* = 4 biological replicates). (I) Forest plot of association of rs74564934 with stroke subtypes (AS, any stroke; AIS, all ischemic stroke; SVS, small vessel stroke; CES, cardioembolic stroke; LAS, large artery stroke). (J) Model of how ETS1, CTCF, and RBPJ bind the FOXF2 enhancer to activate FOXF2 expression in conjunction with the promoter. High-risk SNPs adjacent to the ETS1 site reduce the ability of the enhancer to promote FOXF2 expression.

binding of FOXF2 protein to the *FOXQ1* enhancer, as has been shown (20) (Fig. 4D). Thus, reporter assays, CRISPR inhibition, and CRISPR deletion data agree that this short intergenic region is a critical *FOXF2* enhancer.

We next asked whether GWAS variants modulate enhancer activity. The ETS1 site is proximate to a stroke-associated SNP. Single-cell RNA sequencing from human brain shows enrichment of *ETS1* expression in cells expressing *PDGFR β* , *FOXF2*, *NOTCH1*, *NOTCH3*, and *COL4A1*, which are canonical pericyte markers (*SI Appendix*, Fig. S10 (21)). We synthesized six permutations of construct E with mutations and deletions (Fig. 4E and G). rs74564934 (C/T) is located 2 bp from the canonical ETS1 site, while rs77779542 (T/A) is located 14 bp proximally. We designed mutant constructs with both major “low-risk” alleles (CT) and both minor “high-risk” alleles (TA). We created a single rs74564934 high-risk SNP in the low-risk background (TT) and a single high-risk rs77779542 in the low-risk background (CA). Furthermore, we created a 42-bp deletion of both the ETS1 sites and adjacent SNPs (–). We also created a construct in which RBPJ1 and -2 are mutated to sequences not predicted to bind RBPJ (mutants 1 to 2 (22)). These six construct E mutants were cloned upstream of luciferase, and expression measured relative to the low-risk construct (Fig. 4F). We show that all six mutated constructs had significantly decreased luciferase expression as compared to wild-type construct E (Fig. 4H). The most severe decrease in activity came after the deletion of a 42-bp segment containing low-risk SNPs near ETS1 (33% activity of control $P < 0.0001$). Mutating both SNPs from low risk to high risk (T-A) reduced expression to 51% of control. The mutation of rs74564934 reduces expression to 69% of wild-type ($P < 0.0001$). Mutation of the more proximal SNP rs77779542 reduces expression to 83% of wild-type ($P < 0.01$). Mutation of RBPJ1 and -2 together also decreased expression (62% of control, $P < 0.0001$). Our data strongly suggest that high-risk SNPs in the enhancer modulate transcriptional activity, with rs74564934 having the strongest effect on activity. RBPJ sites are also important for transcriptional enhancer activity.

rs74564943 Is Strongly Associated with Small Vessel Stroke and Its Endophenotype in Humans. rs74564934 is in moderate LD ($r^2 = 0.61$) with rs12204590, the lead SNP reported to be associated with “incident any stroke” in 1000 Genomes European ancestry samples (1), but was not previously followed up. We explored the association of rs7456934 with any stroke ($n = 40,585/406,111$), any ischemic stroke ($n = 34,217/404,630$), and clinical subtypes of ischemic stroke: cardioembolic stroke ($n = 7,193$), large artery stroke ($n = 4,373$), and small vessel stroke ($n = 5,386$), using European ancestry GWAS summary data from the MEGASTROKE Consortium (3). We observed that the less frequently occurring T allele (frequency 0.13 in Europeans) of rs7456943 is significantly associated with increased risk of any stroke, any ischemic stroke, and small vessel stroke (Fig. 4I and *SI Appendix*, Table S1). The association of rs7456943 was not significant for cardioembolic and large artery ischemic stroke, suggesting that the association with any stroke and ischemic stroke is primarily driven by small vessel stroke. We further explored the association of rs7456943 with MRI markers of CSVD using the data from the largest GWAS on white matter hyperintensities (WMH) burden ($n = 48,454$) (13) in European ancestry samples and subcortical brain infarct ($n = 2,021/17,223$), any brain infarct ($n = 3,726/17,223$) (23), and brain microbleeds ($n = 3,556/22,306$) (24) in transethnic samples. We observed that the T allele of rs7456943 is significantly associated with an increased burden of WMH in the

general population after correcting for multiple tests for 4 MRI markers of SVD ($P < 1.25E-2$), whereas the association was not significant for subcortical brain infarct, any brain infarct, and brain microbleeds. Thus, a SNP that decreases *FOXF2* enhancer transcriptional activity in vitro is significantly associated with markers of CSVD and stroke in humans.

Discussion

Despite the large number of GWASs identifying highly significant disease-associated loci, correlation between GWAS-identified SNPs and a biological function has been a major barrier in implementing the findings, particularly when variants are located in intergenic regions. Our data fill this gap for stroke-associated variants near *FOXF2*. We show that stroke-associated SNPs can decrease the activity of a previously unknown *FOXF2* transcriptional enhancer. Decreased expression of *Foxf2* affects vascular wall stability in mouse and fish models. Furthermore, rs7456943, the SNP adjacent to an ETS1 site in the enhancer, is associated with stroke in humans. Mechanistically, we propose that decreased expression of FOXF2 in the vascular wall over a lifetime in humans contributes to decreased vascular integrity and stroke.

Key to our approach is using cell- and animal-based methods to identify enhancer subregions by inhibiting, mutating, or deleting SNPs to demonstrate function. By dissecting a ~6-kb region encompassing risk SNPs guided by in vitro enhancer assays, putative candidate *cis*-regulatory elements (CREs) from ENCODE, and H3K27 acetylation sites, we narrowed the strongest enhancer to a 2-kb region D, and further to E, a 750-bp enhancer. E has the highest enhancer activity of all of the constructs in vitro, while D also has only modest enhancer activity in vitro. In vivo, constructs D and E are expressed in PVCs in developing zebrafish with roughly equal efficiency, showing that the sequence syntax is interpreted as a transcriptional enhancer across evolutionary distant species (25), even though there is no obvious sequence conservation to mouse or zebrafish. Why does construct D have only modest activity in vitro, but shows activity in PVCs in vivo? In vitro luciferase assays are quantitative, while expression assays in zebrafish are qualitative and are scored as presence or absence of expression in PVCs, but do not distinguish signal intensity (this is not possible to score in a transient mosaic expression assay). We dissected the strongest enhancer E activity down to 250 bp, then 42 bp, and identified critical SNPs that modulate its expression.

We find that ETS1 binds to canonical sites adjacent to risk SNPs. Two ETS1 sites are separated by 14 bp, suggesting ETS1 could bind as a homodimer. Mutation of these SNPs from low- to high-risk sites resulted in a significant decrease in enhancer activity, implicating ETS1 in *FOXF2* regulation. ETS1 family members (ETS1, ETV2, FLI1) have important roles in endothelial cell biology downstream of VEGF signaling (26, 27), but no known role in vascular mural cell biology or vascular stabilization. ETS1 also has no known association with stroke in humans. Thus, it was somewhat surprising that the critical SNP that we identified interfered with an ETS1 binding site. However, recent human brain single-cell sequencing analysis shows that ETS1 has enriched expression in pericytes similar to other pericyte markers (21).

In this light, a recently identified critical binding partner of ETS1 is the NOTCH intracellular domain (NICD). ETS1 facilitates NOTCH-dependent transcription in T-cell acute lymphoblastic leukemia (17) by binding to RBPJ and Mastermind-like downstream of Notch signaling to form a transcriptional complex. ETS1, NOTCH1, RBPJ, and H3K27Ac peaks show high co-occupancy in the human genome (17). The human *FOXF2*

enhancer has enriched H3K27Ac, completely overlapping RBPJ and ETS1 ChIP sequencing peaks. One model is that a higher-order ETS1-RBPJ-NICD complex may form, whether RBPJ binds adjacent sites in the enhancer or more remote sites is unknown (model Fig. 4J).

RBPJ can bind constantly at “static” sites in the genome or dynamically at “inducible” sites in regions of open chromatin where NICD cobinds (28). Dynamic NOTCH sites occur within enhancers also containing constitutive CTCF binding (16). CTCF is a DNA-binding protein with transcriptional activation, insulation, and structural activities (29). The human *FOXF2* enhancer we identify has a deeply conserved site that binds CTCF. Using iCRISPRi, we show that the CTCF site positively regulates *FOXF2* expression. CTCF and RBPJ cobind at NOTCH-responsive super-enhancers (16).

We show that Notch signaling is necessary for *FOXF2* enhancer activity and expression. Although RBPJ-targeted iCRISPRi did not show changes in transcriptional activity, mutation of the RBPJ sites decreased enhancer activity. We used two iCRISPRi guides for each of the two predicted RBPJ sites, but these may have not been enough to fully repress the enhancer, as iCRISPRi experiments often need multiple guides. The involvement of Notch signaling in the expression of *FOXF2* and *foxf2b* is supported by reduced *FOXF2* expression with the use of Notch inhibitors.

Mouse and zebrafish models demonstrate a role for FoxF2 in pericyte and smooth muscle differentiation, vascular stabilization, and in the intestine (1, 12, 30). Interestingly, zebrafish *foxf2* mutants have a decreased number of pericytes, while mouse *Foxf2* mutants have increased pericyte numbers (12). However, both gain- and loss-of-function models result in a disrupted vascular wall and vascular stability defects. Similarly, Notch1 and Notch3 activity is also critical for pericyte and smooth muscle differentiation (18, 31–35). Human autosomal dominant *NOTCH3* mutations cause cerebral arteriopathy with subcortical infarcts and leukoencephalopathy 1 (CADASIL1), a disease that leads to stroke and dementia (36). Recessive mutations in *NOTCH3* are associated with Sneddon syndrome, leading to pediatric stroke (37). Importantly, for our study, Notch3 is essential for the maturation of both vSMCs and pericytes (38–41). The downstream targets of Notch in vascular stabilization are unknown; here, we show that human *FOXF2* enhancer is a candidate. *FOXF2* is one of the few examples in which a gene outside of the canonical NOTCH targets (e.g., Hes1, Hey2) has been shown to be regulated by RBPJ. Notch has multiple roles in vascular mural cells. Both gain- and loss-of-function human mutations in *NOTCH3* lead to vascular stability phenotypes. Furthermore, loss of RBPJ in human aortic smooth muscle cells leads to an up-regulation of the vSMC program (18). As RBPJ binds DNA and represses gene expression in the absence of a NOTCH NICD signal, context-specific activation of NOTCH would be expected to transform bound RBPJ from repression to transcriptional activation. In both human cells and zebrafish embryos, we show that Notch inhibition results in decreased *FOXF2* expression, similar to blocking (iCRISPRi) or mutating RBPJ sites in luciferase assays.

Human SVD and resulting stroke can be caused by a number of genes that are not obviously linked in a direct pathway, although all of them affect the structural stability of the vascular wall. Mutations in *NOTCH3*, *COL4A1*, and *FOXC1* cause CSVD. A unifying convergent model for CSVD in humans has emerged, suggesting that causative genes are involved in the higher-level control of the matrisome (ECM proteins secreted around blood vessels) and vascular stability (42). The factors

modulating the expression of matrisome genes are incompletely known. Mouse and zebrafish FoxF2 mutants have compromised vascular integrity, and mouse FoxF2 mutants have decreased Collagen IV expression (12). FoxF2 is therefore a potential intermediate transcription factor in vascular stabilization pathways surrounding ECM deposition. Notch has not been reported to directly activate *COL4A1* gene expression, although the loss of Notch1 and -3 in mice and cells leads to decreased Collagen IV deposition (18, 35).

We show that the developmental loss of FoxF2 in zebrafish results in the reduction of vascular stabilization in development. In contrast, the GWASs that initiated the present work suggest that the risk SNPs in the *FOXF2* intergenic enhancer contribute to stroke in older populations. We identify rs74564934 as a critical SNP that potentially interferes with ETS1 binding in a *FOXF2* enhancer. We then reanalyzed population data that stratify risk for different stroke subtypes to understand whether rs74564934 modifies stroke risk for different stroke types (1, 3, 13, 23, 24). The presence of the minor allele of rs74564934 is significantly associated with any stroke, ischemic stroke, small vessel stroke, extreme small vessel disease, and white matter intensities, demonstrating that our in vitro experiments identify a disease-relevant SNP. The association of this SNP with small vessel disease, WMH, and small vessel stroke is consistent with a role for *FOXF2* in pericytes and vascular stabilization of small vessels, allowing further mechanistic studies on stroke susceptibility and possible treatments.

Mechanistically, patients with risk SNPs may undergo a lifetime with slightly reduced *FOXF2* expression, leading to slightly decreased ECM deposition. Mild defects would become more apparent with age. Additional enhancers and potentially additional SNPs may modulate *FOXF2* expression; however, our experiments suggest that we have identified a critical SNP for *FOXF2* gene expression as no other subconstructs in the region were able to modulate transcriptional activity as effectively. Taken together, the combination of GWASs with in vitro and in vivo models allowed us to dissect and identify the function of important genomic loci in human stroke.

Methods

The methods are described in detail in *SI Appendix, SI Text*, which includes sections covering the following: cell lines, zebrafish, expression of dCas9, CRISPR, RNA isolation and qPCR analysis, gateway cloning of intergenic enhancer constructs, luciferase reporter assay, ChIP, drug inhibition studies, statistics, and genetic association. The animal procedures used in this study were approved by the Institutional Animal Care and Use Committee at the University of Calgary.

Data, Materials, and Software Availability. All data and materials are available in the main text or the supplementary materials.

ACKNOWLEDGMENTS. We thank Jasper Greysson-Wong and Ordan Lehmann for insightful comments on the manuscript. Schematics were created in BioRender. Funding for the study is from the Cumming School of Medicine Postdoctoral Fellowship (to S.A.), Alberta Children’s Hospital Research Institute Postdoctoral Fellowship (to S.A.), Canadian Institutes for Health Research PJT-153023 (to S.J.C.), Canadian Institutes for Health Research PJT-795600 (to D.J.M.), Canadian Cancer Society Grant DF-18-5 (to D.J.M.), and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Grant AG052409 (to S.S. and S.D.).

Author affiliations: ^aAlberta Children’s Hospital Research Institute, University of Calgary, Calgary AB T2N 4N1, Canada; ^bDepartment of Biochemistry and Molecular Biology, University of Calgary, Calgary AB T2N 4N1, Canada; ^cDepartment of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Calgary AB T2N 4N1, Canada; ^dArnie Charbonneau Cancer Institute, University of Calgary, Calgary AB T2N

1. G. Chauhan *et al.*, Identification of additional risk loci for stroke and small vessel disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **15**, 695–707 (2016).
2. K. L. Keene *et al.*; COMPASS, SiGN, and METASTROKE Consortia, Genome-wide association study meta-analysis of stroke in 22 000 individuals of African descent identifies novel associations with stroke. *Stroke* **51**, 2454–2463 (2020).
3. R. Malik *et al.*; AFGen Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; INVENT Consortium; STARNET; BioBank Japan Cooperative Hospital Group; COMPASS Consortium; EPIC-CVD Consortium; EPIC-InterAct Consortium; International Stroke Genetics Consortium (ISGC); METASTROKE Consortium; Neurology Working Group of the CHARGE Consortium; NINDS Stroke Genetics Network (SiGN); UK Young Lacunar DNA Study; MEGASTROKE Consortium, Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
4. M. Traylor *et al.*; NINDS Stroke Genetics Network (SiGN); and International Stroke Genetics Consortium (ISGC), Subtype specificity of genetic loci associated with stroke in 16 664 cases and 32 792 controls. *Circ. Genom. Precis. Med.* **12**, e002338 (2019).
5. R. Y. Y. Tan *et al.*; NIH Rare Diseases Consortium, How common are single gene mutations as a cause for lacunar stroke? A targeted gene panel study. *Neurology* **93**, e2007–e2020 (2019).
6. A. Ter Telgte *et al.*, Cerebral small vessel disease: From a focal to a global perspective. *Nat. Rev. Neurol.* **14**, 387–398 (2018).
7. T. Kume, H. Jiang, J. M. Topczewska, B. L. Hogan, The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* **15**, 2470–2482 (2001).
8. C. R. French *et al.*, Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *J. Clin. Invest.* **124**, 4877–4881 (2014).
9. J. A. Siegenthaler *et al.*, Foxc1 is required by pericytes during fetal brain angiogenesis. *Biol. Open* **2**, 647–659 (2013).
10. T. R. Whitesell *et al.*, foxc1 is required for embryonic head vascular smooth muscle differentiation in zebrafish. *Dev. Biol.* **453**, 34–47 (2019).
11. J. M. Skarie, B. A. Link, FoxC1 is essential for vascular basement membrane integrity and hyaloid vessel morphogenesis. *Invest. Ophthalmol. Vis. Sci.* **50**, 5026–5034 (2009).
12. A. Reyahi *et al.*, Foxf2 is required for brain pericyte differentiation and development and maintenance of the blood-brain barrier. *Dev. Cell* **34**, 19–32 (2015).
13. M. Sargurupremraj *et al.*; International Network against Thrombosis (INVENT) Consortium; International Headache Genomics Consortium (IHGC), Cerebral small vessel disease genomics and its implications across the lifespan. *Nat. Commun.* **11**, 6285 (2020).
14. U. Lendahl, P. Nilsson, C. Betsholtz, Emerging links between cerebrovascular and neurodegenerative diseases—a special role for pericytes. *EMBO Rep.* **20**, e48070 (2019).
15. P. I. Thakore *et al.*, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143–1149 (2015).
16. H. Wang *et al.*, NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 705–710 (2014).
17. A. C. McCarter *et al.*, Combinatorial ETS1-dependent control of oncogenic NOTCH1 enhancers in T-cell leukemia. *Blood Cancer Discov.* **1**, 178–197 (2020).
18. J. M. Rozenberg, J. M. Taylor, C. P. Mack, RBPJ binds to consensus and methylated cis elements within phased nucleosomes and controls gene expression in human aortic smooth muscle cells in cooperation with SRF. *Nucleic Acids Res.* **46**, 8662 (2018).
19. C. P. Yu *et al.*, Discovering unknown human and mouse transcription factor binding sites and their characteristics from ChIP-seq data. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026754118 (2021).
20. L. J. Kang *et al.*, Reciprocal transrepression between FOXF2 and FOXQ1 controls basal-like breast cancer aggressiveness. *FASEB J.* **33**, 6564–6573 (2019).
21. L. Song *et al.*, STAB: A spatio-temporal cell atlas of the human brain. *Nucleic Acids Res.* **49**, D1029–D1037 (2021).
22. J. B. Miesfeld *et al.*, Rbpj direct regulation of Atoh7 transcription in the embryonic mouse retina. *Sci. Rep.* **8**, 10195 (2018).
23. G. Chauhan *et al.*; Stroke Genetics Network (SiGN); the International Stroke Genetics Consortium (ISGC); METASTROKE; Alzheimer's Disease Genetics Consortium (ADGC); and the Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, Genetic and lifestyle risk factors for MRI-defined brain infarcts in a population-based setting. *Neurology* **92**, e486–e503 (2019).
24. M. J. Knol *et al.*; Alzheimer's Disease Neuroimaging Initiative, Association of common genetic variants with brain microbleeds: A genome-wide association study. *Neurology* **95**, e3331–e3343 (2020).
25. D. Villar *et al.*, Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
26. S. De Val, B. L. Black, Transcriptional control of endothelial cell development. *Dev. Cell* **16**, 180–195 (2009).
27. A. Neal *et al.*, ETS factors are required but not sufficient for specific patterns of enhancer activity in different endothelial subtypes. *Dev. Biol.* **473**, 1–14 (2021).
28. D. Castel *et al.*, Dynamic binding of RBPJ is determined by Notch signaling status. *Genes Dev.* **27**, 1059–1071 (2013).
29. R. J. Lake, P. F. Tsai, I. Choi, K. J. Won, H. Y. Fan, RBPJ, the major transcriptional effector of Notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. *PLoS Genet.* **10**, e1004204 (2014).
30. C. Bolte *et al.*, Forkhead box F2 regulation of platelet-derived growth factor and myocardin/serum response factor signaling is essential for intestinal development. *J. Biol. Chem.* **290**, 7563–7575 (2015).
31. T. Grieskamp, C. Rudat, T. H. Lütke, J. Norden, A. Kispert, Notch signaling regulates smooth muscle differentiation of epicardium-derived cells. *Circ. Res.* **108**, 813–823 (2011).
32. F. A. High *et al.*, An essential role for Notch in neural crest during cardiovascular development and smooth muscle differentiation. *J. Clin. Invest.* **117**, 353–363 (2007).
33. K. Kurpinski *et al.*, Transforming growth factor-beta and notch signaling mediate stem cell differentiation into smooth muscle cells. *Stem Cells* **28**, 734–742 (2010).
34. T. L. Henshall *et al.*, Notch3 is necessary for blood vessel integrity in the central nervous system. *Arterioscler. Thromb. Vasc. Biol.* **35**, 409–420 (2015).
35. N. M. Kofler, H. Cuervo, M. K. Uh, A. Murtoniemi, J. Kitajewski, Combined deficiency of Notch1 and Notch3 causes pericyte dysfunction, models CADASIL, and results in arteriovenous malformations. *Sci. Rep.* **5**, 16449 (2015).
36. A. Joutel *et al.*, Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383**, 707–710 (1996).
37. E. K. Greisenegger *et al.*, A NOTCH3 homozygous nonsense mutation in familial Sneddon syndrome with pediatric stroke. *J. Neurol.* **268**, 810–816 (2021).
38. V. Domenga *et al.*, Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18**, 2730–2735 (2004).
39. C. Fouillade, M. Monet-Leprêtre, C. Baron-Menguy, A. Joutel, Notch signalling in smooth muscle cells during development and disease. *Cardiovasc. Res.* **95**, 138–146 (2012).
40. H. Liu, S. Kennard, B. Lilly, NOTCH3 expression is induced in mural cells through an autoregulatory loop that requires endothelial-expressed JAGGED1. *Circ. Res.* **104**, 466–475 (2009).
41. Y. Wang, L. Pan, C. B. Moens, B. Appel, Notch3 establishes brain vascular integrity by regulating pericyte number. *Development* **141**, 307–317 (2014).
42. A. Joutel, I. Haddad, J. Ratelade, M. T. Nelson, Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J. Cereb. Blood Flow Metab.* **36**, 143–157 (2016).