Published online 6 March 2018

Nucleic Acids Research, 2018, Vol. 46, No. 7 ***3517–3531***
doi: 10.1093/nar/gky166

A screen for deeply conserved non-coding GWAS SNPs uncovers a *MIR-9-2* functional mutation associated to retinal vasculature defects in human

Romain Madelaine[1,†], James H. Notwell[2,†], Gemini Skariah[1], Caroline Halluin[1], Charles C. Chen[2], Gill Bejerano[2,3,4,*] and Philippe Mourrain[1,5,*]

[1]Department of Psychiatry and Behavioral Sciences, Stanford Center for Sleep Sciences and Medicine, Stanford, CA 94305, USA, [2]Department of Computer Science, Stanford, CA 94305, USA, [3]Department of Developmental Biology, Stanford, CA 94305, USA, [4]Division of Medical Genetics, Department of Pediatrics, Stanford, CA 94305, USA and [5]INSERM 1024, Ecole Normale Supérieure Paris, 75005, France

Received August 10, 2017; Revised February 20, 2018; Editorial Decision February 21, 2018; Accepted March 01, 2018

ABSTRACT

Thousands of human disease-associated single nucleotide polymorphisms (SNPs) lie in the non-coding genome, but only a handful have been demonstrated to affect gene expression and human biology. We computationally identified risk-associated SNPs in deeply conserved non-exonic elements (CNEs) potentially contributing to 45 human diseases. We further demonstrated that human CNE1/rs17421627 associated with retinal vasculature defects showed transcriptional activity in the zebrafish retina, while introducing the risk-associated allele completely abolished CNE1 enhancer activity. Furthermore, deletion of CNE1 led to retinal vasculature defects and to a specific downregulation of *microRNA-9*, rather than *MEF2C* as predicted by the original genome-wide association studies. Consistent with these results, *miR-9* depletion affects retinal vasculature formation, demonstrating *MIR-9-2* as a critical gene underpinning the associated trait. Importantly, we validated that other CNEs act as transcriptional enhancers that can be disrupted by conserved non-coding SNPs. This study uncovers disease-associated non-coding mutations that are deeply conserved, providing a path for *in vivo* testing to reveal their *cis*-regulated genes and biological roles.

INTRODUCTION

Since the availability of the human genome sequence in 2000, widespread genome-wide association studies (GWAS) brought new hope for understanding the genetic basis of human diseases. The most common type of human genetic variation is a single nucleotide polymorphism (SNP), where two alternative bases occur at appreciable frequency in the human population. SNPs located in protein coding regions have been shown to change susceptibility to diseases by losing or changing protein function (1,2). An unexpected finding of the GWAS era was the prevalence of disease-associated SNPs located outside of protein coding transcripts (>90%), making the genetic contribution to these disorders much more difficult to understand (3–5). GWAS-SNPs located in non-exonic regions of the genome are thought to highlight disease-associated modifications to gene regulatory elements (6), but reported SNPs may not be the causal mutation and may gain their association from linkage with the actual causal mutation (7,8). Several elegant studies have uncovered the impact of non-coding SNPs. To our knowledge, the first reported example was a base substitution in a promoter element of the β-globin gene in patients suffering of β-thalassemia (9). Other studies showed that when present in the promoter region, SNPs can alter or even create transcription factor binding sites or affect promoter methylation and have substantial effects on both gene transcription and protein levels (10–12). When present in the 3′ UTR region, SNPs can affect mRNA stability by changing binding affinity of microRNA (13).

Deciphering how SNPs located in intronic or intergenic regions alter human disease susceptibility remains a significant challenge, due to our incomplete understanding of regulatory codes. Moreover, the genes they regulate, which could give us clues to their function, can be located tens or hundreds of kilobase-pairs away and mingled among other genes, making the discovery of the actual *cis*-regulated gene a tremendous challenge. To date, very few intronic or inter-

*To whom correspondence should be addressed. Tel: +1 650 724 3871; Fax: +1 650 723 9546; Email: mourrain@stanford.edu
Correspondence may also be addressed to Gill Bejerano. Tel: +1 650 723 7666; Fax: +1 650 725 2923; Email: bejerano@stanford.edu
†These authors contributed equally to the paper as first authors.

genic GWAS-SNPs have been functionally validated *in vivo* as likely causal mutations in human phenotypes. These rare studies (seven to the best of our knowledge) have demonstrated the role of non-coding variants in human disorders such as cancers, pigmentation, scoliosis and restless legs syndrome (RLS) (14–20). For example, rs12469063 (RLS) and rs2168101 (neuroblastoma) reduce the transcriptional activity of enhancers regulating *MEIS1* and *LMO1*, respectively (16,19), whereas rs339331 (prostate cancer) and rs11190870 (scoliosis) increase the activity of enhancers regulating the transcription of *RFX6* and *LBX1*, respectively (17,18). These reports highlight the importance of investigating the putative regulatory activity of non-coding SNPs identified by GWAS to determine their biological functions. The characterization of these regulatory SNPs not only revealed new enhancers, but also the genes and molecular and biological processes disrupted in human pathologies.

In this study, we developed an extremely sensitive method for detecting GWAS-SNPs that fall in evolutionarily conserved genomic regions which are likely functional *cis*-regulatory elements (conserved non-exonic DNA elements, CNEs) by focusing on the identification of non-coding SNPs deeply conserved across vertebrate genomes (21–24) that also preserve gene synteny (25–27). The intersection of conserved GWAS-SNPs with deep non-coding sequence conservation likely selects elements of functional relevance, and has the added benefit of providing an immediate path for *in vivo* testing in vertebrate animal models; we used zebrafish for the present study. This novel approach allowed us to identify a set of 45 non-coding SNPs located in deeply conserved CNEs, indicating a potential functional role. One of the non-coding regulatory mutations mentioned above, rs11190870, is located in CNE15 close to LBX1 which was previously demonstrated as responsible for scoliosis (17), suggesting the value of our set. We performed an in-depth investigation *in vivo* of our first SNP rs17421627/CNE1 pair associated with retinal vasculature defects in two related GWAS studies, and was predicted to regulate the neighboring gene *MEF2C* (28,29). Here, we showed that CNE1 acts as a specific enhancer in the retina regulating *microRNA-9-2,* but not *MEF2C,* expression. Importantly, we found that rs17421627 risk nucleotide abolishes CNE1 activity demonstrating its mutagenic capacity. In addition, deletion of CNE1 (ΔCNE1) in the zebrafish genome and downregulation of *miR-9* expression led to retinal vasculature defects. Thus, our evidence not only demonstrated that rs17421627 is a functional mutation, but we also unmasked the gene affected in the biological process, providing an entry point for future treatment strategies. Furthermore, for 11 of the human pathologies, such as artery calcification, nasopharyngeal carcinoma or macular degeneration, the associated CNE/SNP pair preserves synteny with a single gene across multiple species, significantly simplifying and improving the accuracy of predicting the *cis*-regulated gene contributing to the human phenotype. Finally, we confirmed the regulatory function of other CNE/SNPs pairs in our list, providing an avenue for understanding the biological processes underpinning other human medical conditions.

MATERIALS AND METHODS

Identification of zebrafish conserved non-coding GWAS SNPs

The NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/) was most recently downloaded on 4 November 2017 from the UCSC Genome Browser (30). Each SNP was padded with 100 bp on each side, creating a local sequence context to be used when querying the zebrafish genome. Those elements overlapping any (coding or non-coding) Ensembl exon (Ensembl release 75) (31), padded by 50 bp, were excluded to avoid all coding and splicing related events (while non-coding intronic sequences were not removed).

For each element, the corresponding alignment blocks from a varying number of species were extracted from the UCSC 46-way multiple alignment (30). Sequences that had been soft-masked in the UCSC alignment were hard-masked to prevent alignments between repetitive sequences. A profile hidden Markov model (HMM) was constructed from the alignment corresponding to each element using the hmmbuild utility from the HMMER package (version 3.1b1) (32). This model was used to query the zebrafish Zv9 assembly (danRer7) using nhmmer (version 3.1b1) with default parameters. All alignments were retained. The human and zebrafish matches were then realigned using LASTZ (version 1.02.00; –seed = match4 –hspthresh = 500 –gappedthresh = 500; HoxD55 scoring matrix; http://www.bx.psu.edu/~rsharris/lastz/). Only pairs scoring above 2000 were retained.

GWAS-SNPs embedded specifically in gene regulatory regions conserved to zebrafish are expected to lie next to the orthologous gene(s) in both human and zebrafish. This property is known as synteny. Ensembl gene transcripts and ortholog mappings (Ensembl release 75) (31) were used to identify syntenic alignments. The top-scoring zebrafish alignment for each human query was labeled as syntenic if there exists an Ensembl gene transcript within 500 kb of the hg19 SNP, as well as an Ensembl gene transcript corresponding to a human gene ortholog within 500kb of the zebrafish alignment.

This entire process was repeated for the chicken galGal4 assembly with a false discovery rate (FDR) of 0.

Transcription factor binding site prediction

We have curated an extensive library of monomer and dimer transcription factor binding motifs, each represented as a position weight matrix, from the UniPROBE (33), JASPAR (34) and TransFac (35) databases, secondary UniPROBE motifs, motifs from published ChIP-seq datasets and from other primary literature as described previously (36). Sequences with a MATCH (37) score of at least 0.8 to a position weight matrix (PWM) were overlapped with GWAS SNPs.

CRISPR/Cas-9 deletion of CNE1 in the zebrafish genome

To delete CNE1, we took advantage of the previously described genome editing protocol (38). We used two gRNA targeting the CNE1 locus (5′-CCCGGCGTCCCCCTT

CCT-3′ and 5′-AGGAAGGGGGACGCCGGG-3′). Co-injection of these gRNAs with the *Cas9* mRNA resulted in a deletion of 770 bp removing CNE1 most conserved core including SNP position. F1 clutches from F0 injected embryos were genotyped by polymerase chain reaction (PCR) to identify F0 fish carrying the CNE1 deletion in the germline. F1 adult fish from F0 positive carriers were genotyped individually by PCR (with primers upstream/forward: 5′-CAGACTCTACCTTTCCTGCAAC-3′ and downstream/reverse: 5′-CCTTACATTTGCATGCCTAAC-3′ of the CNE1 sequence) to identify F1 positive heterozygous carriers for the CNE1 deletion (validated by sequencing of the PCR product). F1 carriers were outcross with *kdrl:mcherry* transgenic fish. F2 heterozygous fish were identified by genotyping and intercross to generate F3 homozygous CNE1 mutants and siblings controls. F3 homozygous mutant larvae were identified by genotyping of the tail (the heads were individually keep to collect the retina) with a forward primer inside the CNE1 region 5′-GGACCAGGAAGCAGTAATGG-3′ (CNE1 was not amplified by PCR in the homozygous mutants).

Fish lines and developmental conditions

Embryos were raised and staged according to standard protocols (39) and in accordance with Stanford University animal care guidelines. The following transgenic lines were used: *Tg(kdrl:ras-mCherry)* (40) and *Tg(olig2:dsred2)* (41). Embryos were fixed overnight at 4°C in 4% paraformaldehyde/1×phosphate-buffered saline, after which they were dehydrated through an ethanol series and stored at −20°C until use.

Plasmid construction and transgenic line establishment

For the generation of *Tg(CNEx:egfp)* and *Tg(CNExSNP:egfp),* transgenic lines were made by PCR amplification of human genomic DNA. PCR products were directionally cloned into the XhoI and BglII sites of the pTol2-E1b:EGFP vector (42). Plasmids were injected into one-cell stage embryos with the Tol2 mRNA transposase (43). F1 clutches from F0 injected embryos were genotyped by PCR to identify F0 fish carrying the transgene in the germline. F1 adult fish from F0 positive carriers were genotyped individually by PCR to identify F1 positive carriers for the transgene. F2 embryos from F1 positive fish were analyzed to determine the activity of the enhancer. For each stable line, at least three independent integrations were identified by genotyping (PCR on the *egfp* sequence) and used to determine the transcriptional activity with the protective or the risk allele. For example, for the *Tg(CNE1:egfp),* four independent integrations with an identical EGFP expression pattern were identified. For the *Tg(CNE1SNP:egfp)*, *Tg(CNE8SNP:egfp)* and *Tg(CNE18SNP:egfp),* nine, five and four independent integrations with no detectable EGFP expression were isolated by genotyping and analyzed respectively. Of note, for *Tg(CNE18SNP:egfp)*, we identified five independent integrations. For one of these integrations, we observed EGFP expression in the spine at 24 hpf. At 48 hpf, this expression was gone and contrary to the three independent lines carrying the protective allele, but identically to the other four stable lines carrying the risk allele, the expression was not detected in muscle or retina. This expression was likely ectopic and due to the position of the integration in the zebrafish genome and does not reflect the intrinsic activity of the enhancer carrying the SNP mutation.

In situ hybridization and Immunostaining

In situ hybridizations (ISHs) were performed as previously described (44). *meis2a* and *sox6* ORFs were cloned in a pCS2+ vector using zebrafish cDNA and antisense DIG labeled probes were transcribed using the linearized pCS2+ plasmid containing the ORF. For *miR-9* ISH, the previously described miRCURY detection probe (LNA) hsa–miR-9 (Exiqon) was used (45). Probes for *pri-miR-9-5s* were amplified by PCR on genomic DNA to target the previously described regions in the *miR-9-5* pri-precursors (46). For pre-*miR-9-5* ISH, a miRCURY LNA probe (dre-mir-9-5; TATGAAGTGCAAAATACTC) was specifically designed by Exiqon. ISHs were revealed using either BCIP and NBT (Roche) or Fast Red (Roche) as substrates. Immunohistochemical stainings were performed as previously described (47), using either anti-GFP (1/1000, Torrey Pines Biolabs), anti-HuC/D (1/500, Molecular Probes), anti-glutamine synthetase (1/500, Millipore) or anti-DsRed (1/500, Clontech) as primary antibodies and Alexa 488 or Alexa 555-conjugated goat anti-rabbit IgG or goat anti-mouse IgG (1/1000) as secondary antibodies (Molecular Probes). FITC-dextran 2000 kDa (Sigma) was used to visualize the vasculature after injection in the cardinal vein.

Antisense morpholino injection

For morpholino knockdowns, we used the previously described miR-9 MO (TCATACAGCTAGATAACCAAAGA) and the control MO (CACCAAACCATATAGAAGTGATA) (45). Many studies previously reported and validated the efficiency and specificity of the miR-9 MO (45,48–51). Embryos were injected at the one-cell stage with 2 pmole of the miR-9 or control MO.

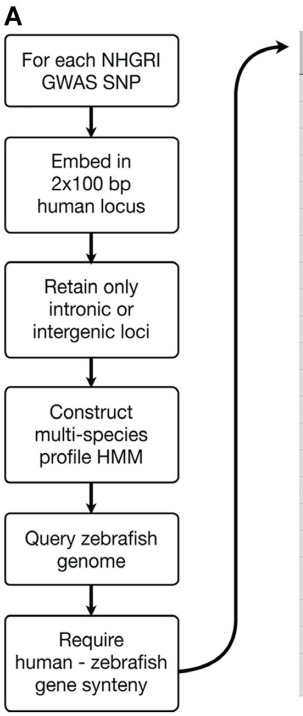
Image acquisition and blood vessels branching quantification

Confocal acquisitions were carried out using a Leica SP5 confocal microscope (Stanford cell science imaging facility). Images were prepared using Photoshop software (Adobe). For quantification of blood vessels number and branching in the retina, images were analyzed using ImageJ software with Skeletonize and AnalyzeSkeleton plugins. Statistical analyses associated with each figure are reported in the figure legends.

RESULTS

Computational identification of non-coding GWAS SNPs conserved in evolution

We began by extracting nearly 40 000 references SNP cluster ID (rsID) entries from the NHGRI-EBI GWAS catalog. We padded each SNP with 100 bp on either side, creating



Conserved non-coding element	Trait or disease	Variant rsID	Syntenic genes
CNE1	Retinal vascular caliber	rs17421627	MIR9-5, MEF2C, TMEM161B
CNE2	Hippocampal volume	rs17178006	MSRB3, LEMD3, WIF1, HMGA2
CNE3	Migraine	rs4759042	RDH16, SDR9C7, HSD17B6
CNE4	Migraine	rs4880487	ADARB2, WDR37, IDI1, IDI2
CNE5	Restless legs syndrome	rs12593813	SKOR1, C15orf61, IQCH, PIAS1, AAGAB
CNE6	Attention deficit hyperactivity disorder	rs2842643	TFEB, MDFI, FOXP4
CNE7	Response to amphetamines	rs564148	PHC2
CNE8	Response to antipsychotic treatment	rs1568679	MEIS2, C15orf41
CNE9	Response to statin therapy	rs6588480	GLIS1
CNE10	Obesity-related traits	rs12431307	SPRY2
CNE11	Coronary artery calcification	rs7246657	ZNF850
CNE12	Age-related macular degeneration	rs13095226	COL8A1
CNE13	Myopia (severe)	rs13382811	ZEB2
CNE14	Corneal structure	rs2307121	ADAMTS6, PPWD1, TRIM23, CWC27
CNE15	Scoliosis	rs11190870	LBX1, TLX1, FBXW4
CNE16	Ankylosing spondylitis	rs4333130	ANTXR2, PRDM8, FGF5, C4orf22
CNE17	Nasopharyngeal carcinoma	rs6774494	MECOM
CNE18	Capecitabine sensitivity	rs16932455	SOX6
CNE19	Immune reponse to smallpox (secreted TNF-alpha)	rs13414205	CAMKMT, PREPL, SLC3A1, PPM1B, LRPPRC
CNE20	Polycystic ovary syndrome	rs2272046	HMGA2, TMBIM4, MSRB3, IRAK3, HELB
CNE21	Freckling	rs2153271	BNC2, CNTLN
CNE22	Hair morphology	rs7349332	WNT10A, WNT6, CDK5R2, FEV, CRYBA2, MIR375

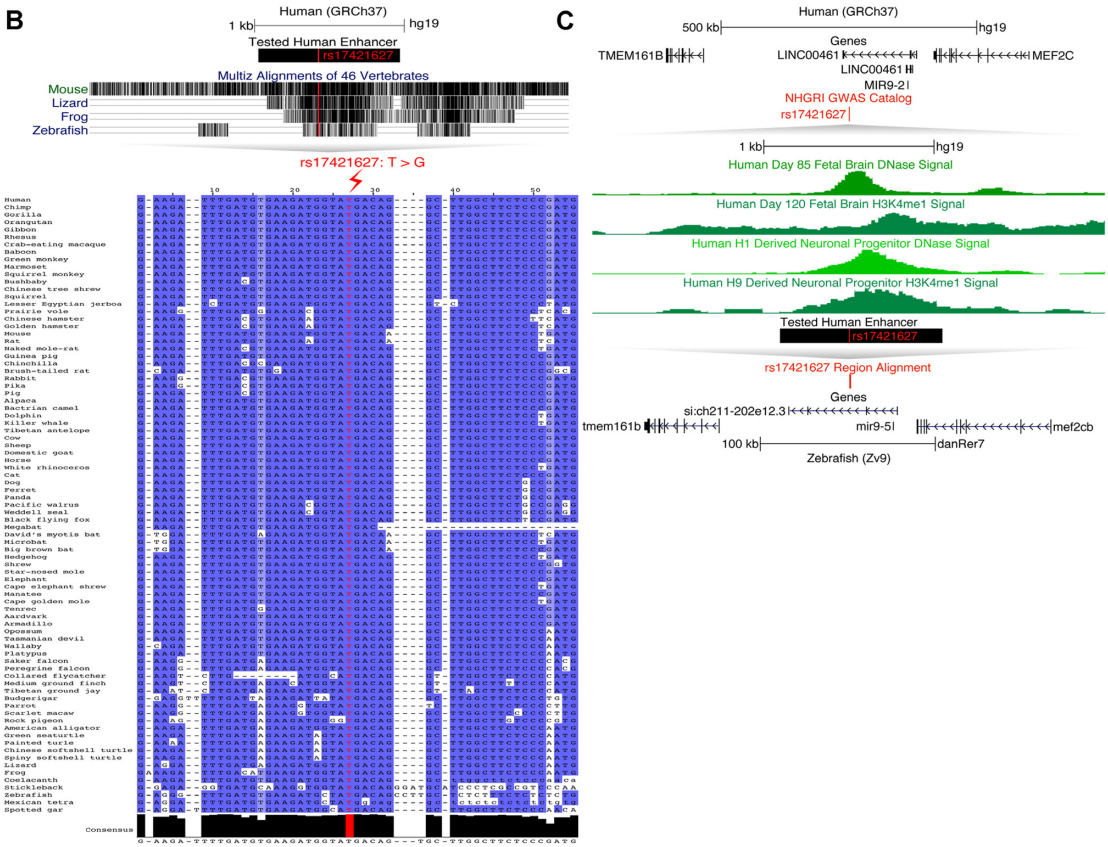


Figure 1. Identification of GWAS SNPs embedded in deeply conserved non-coding elements. (**A**) Methodology used to perform the screen and first list of regulatory GWAS SNPs conserved between human and zebrafish, and their conserved syntenic genes (see Supplementary Table S1 for the full list of 45 conserved CNE/SNPs). (**B**) The rs17421627 protective allele and its immediate sequence neighborhood have been conserved across dozens of species and hundreds of millions of years of evolution, including in human CNE1 (948 bp; hg19.chr5:87,847,186-87,848,133) and zebrafish CNE1 (860 bp; danRer7.chr5:49,927,812-49,928,671). (**C**) The human and zebrafish gene neighborhoods surrounding CNE1/rs17421627 contain three orthologous and conserved genes: *TMEM161B*, *MEF2C* and *MIR9-2*. The CNE1/SNP pair is marked by H3K4me1 and DNase signals in both human fetal brain and neuronal progenitors.

a local sequence context to be used when querying the zebrafish genome. We then removed all regions overlapping or in close proximity to any exon to avoid coding and splicing related mutations. For each human element, we extracted a multiple species alignment block aligned to the region, constructed a DNA profile HMM, and used nhmmer (32) to sensitively query the zebrafish genome. To specifically identify gene regulatory homologies we then applied a gene synteny filter, requiring that the human GWAS-SNP embedding region and its top zebrafish alignment are found near orthologous genes in human and zebrafish (Figure 1 and 'Materials and Methods' section). This resulted in a list of 45 human GWAS-SNPs embedded in non-coding elements conserved down to zebrafish next to orthologous genes (Figure 1A and Supplementary Table S1 for the full list). The list includes SNPs associated with traits and diseases ranging from psychiatric, metabolic to developmental disorders. To estimate the FDR, we used the same set of models to query the reversed but not complemented zebrafish genome, and this resulted in zero false positive matches. We overlapped our CNE/SNP pairs with chromatin states from the Roadmap Epigenomics Project (52) and found that 38 of our CNEs overlap enhancers from the core 15-state model integrating 5 epigenetic marks for 127 epigenomes (Supplementary Table S2), suggesting functional contexts for the majority of our CNEs. In addition, for 11 of the CNE/SNP pairs, we identified a single syntenic gene, suggesting the identity of the gene *cis*-regulated by the CNE containing the GWAS-SNP and potentially contributing to the human trait (Supplementary Table S1). Finally, all the CNEs identified in this analysis are strongly conserved during 450 million years of evolution since the teleost radiation, suggesting that they are functional *cis*-regulators (Figure 1B for an example).

Validation of deeply conserved CNEs as functional enhancers *in vivo*

To determine whether the conserved CNE/SNP pairs discovered by our screen act as *cis*-regulatory elements *in vivo*, we decided to establish stable CNE#:EGFP lines for a representative subset of our list. Based on our scientific interest and associated diseases, we picked 6 CNE/SNP pairs (#1, 8, 10, 15, 16, 18) and tested the transcriptional activity of these CNEs during early zebrafish development. We fused the human CNE regions to a basal promoter driving EGFP expression in at least three independent zebrafish lines per transgene to dismiss putative genome position effects (Figure 2A; integrations were identified by genotyping at least three independent integrations with identical expression patterns for each these CNEs, see 'Materials and Methods' section for details). We validated the transcriptional enhancer function for five of these conserved non-coding elements (CNE1/rs17421627; CNE8/rs1568679; CNE10/rs12431307; CNE15/rs11190870; CNE18/rs16932455; Figure 2B). The lack of activity observed for one of them (CNE16/rs4333130) could be attributed to the developmental stage we assayed, e.g. the associated disease arthritis affects elderly, or potentially to their identity as repressor or insulator elements. The CNEs with validated enhancer function drove restricted EGFP expression in a variety a cell types (neurons, glial cells, muscles and notochord), demonstrating their specific transcriptional activity (Figure 2 and Supplementary Figure S1). This result demonstrates the transcriptional role of these non-coding elements indicating that sequence conservation highlights functional conservation.

The human risk alleles abolish enhancer functions *in vivo*

We next tested the functionality of the human risk allele for three of our five positive CNE lines, CNE1/rs17421627, CNE8/rs1568679 and CNE18/rs16932455 by introducing the 1 bp change into the 948, 849 and 832 bp long human enhancers, respectively. Strikingly, in these three enhancers, the risk allele abolished the transcriptional activity completely in all independent integrations that we identified (Figure 2A–E). To confirm that the lack of expression was due to the risk mutation and not a genome integration effect, we used genotyping to identify at least 4 independent integrations for each CNE/SNP carrying the risk allele (see 'Materials and Methods' section for details). In addition, we performed EGFP immunostaining and ISH to compare CNE-dependent expression with the associated syntenic genes. We identified the likely *cis*-regulated gene for CNE1 (*MIR9-2*), CNE8 (*MEIS2*) and CNE18 (*SOX6*) (Supplementary Figure S1), as indicated by the coexpression of EGFP with the endogenous microRNA or mRNAs. Importantly, while the *P*-value for rs17421627 and rs1568679 are highly significant with the common standards used in GWAS ($P < 5 \times 10^{-8}$), this is not the case for rs16932455 ($P = 2 \times 10^{-6}$). Despite this fact, rs16932455 is a functional mutation abolishing most of the EGFP expression in the known domains of *sox6* expression (brain, retina and muscle), indicating that sequence conservation is a useful criterion for identifying functional non-coding mutations potentially contributing to human phenotypes. In addition to these three findings, a recent report independently supports our predictions that GWAS-SNPs embedded in conserved CNEs are regulatory mutations by demonstrating that rs11190870, associated with scoliosis, is a gain of function mutation in an enhancer (our CNE15 element) regulating *LBX1* activity (17). Overexpression of human *LBX1* or zebrafish *lbx* genes caused scoliosis (17). Together, these results validate *in vivo* our *in silico* predictions and demonstrate the approach effectively identifies functional CNE/SNP pairs likely contributing to human traits as suggested by GWAS.

CNE1 enhancer deletion leads to retinal vasculature defects *in vivo*

We showed above that our human CNE/SNP pair testing in a vertebrate model organism allows us (i) to reveal the specific enhancer activity of a CNE, (ii) to demonstrate the mutagenic effect of risk SNP on this activity and (iii) to unmask the *cis*-regulated gene likely underpinning the human disease. We next wanted to test whether we could phenocopy/model aspects of human defects caused by CNE dysregulation in zebrafish, and further investigated

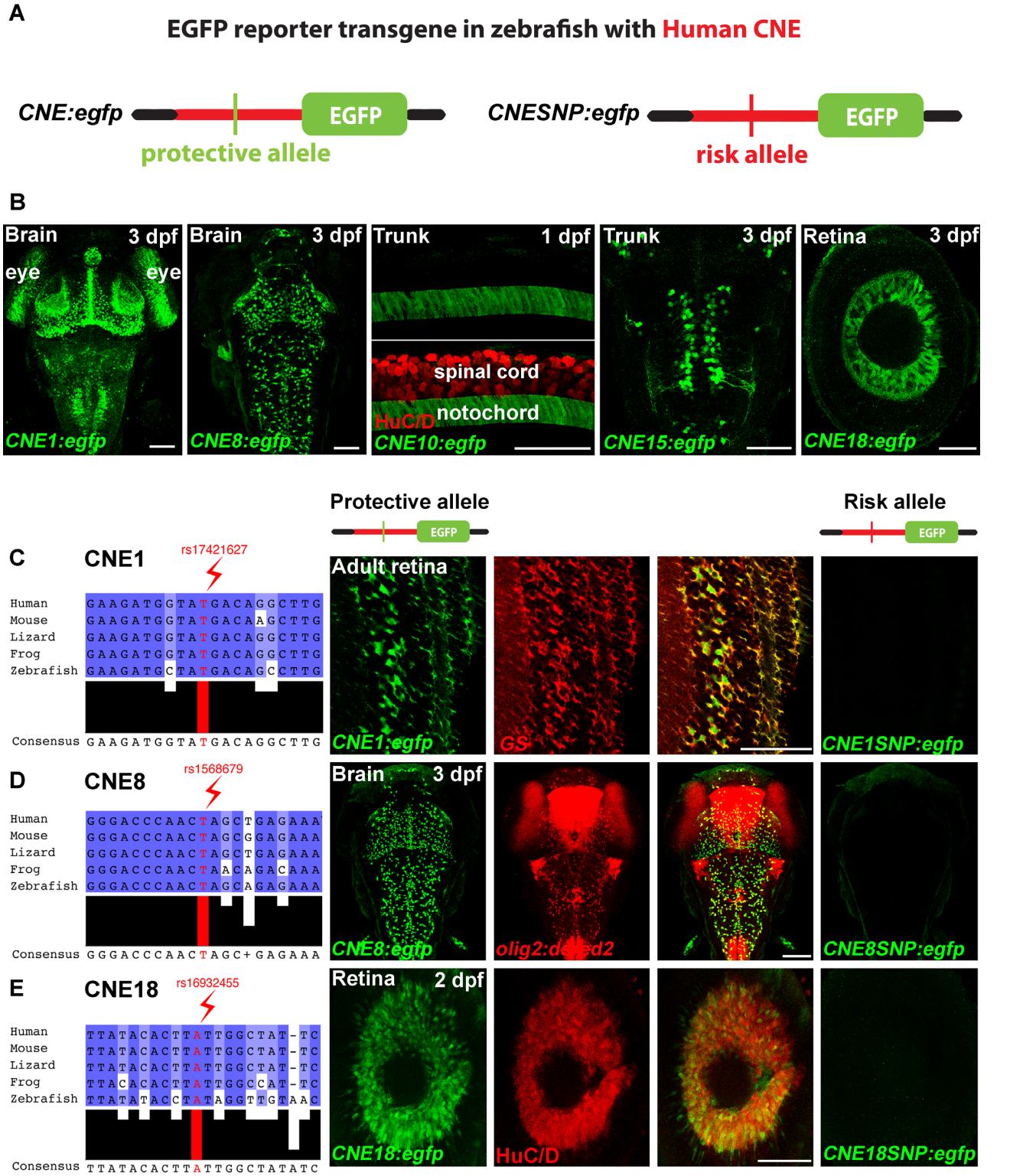


Figure 2. Functional validation of conserved CNEs enhancer activity and demonstration that the associated SNP can be a regulatory mutation. (**A**) Schematic of the transgenes carrying the protective and the risk alleles. (**B**) Confocal projections of EGFP immunolabeling in *Tg(CNE1:egfp)*, *Tg(CNE8:egfp)*, *Tg(CNE10:egfp)*, *Tg(CNE15:egfp)* and *Tg(CNE18:egfp)*, demonstrating the specific enhancer activity of these human CNEs in zebrafish. (**C**–**E**) The sequence surrounding rs17421627, rs1568679 and rs16932455 are conserved during evolution. (**C**) Confocal sections of EGFP immunolabeling in *Tg(CNE1:egfp)* containing the protective allele or *Tg(CNE1SNP:egfp)* containing the risk allele (rs17421627) in the adult retina immunolabeled with glutamine synthetase (GS). (D) Confocal projections of EGFP immunolabeling in *Tg(CNE8:egfp)* containing the protective allele or *Tg(CNE8SNP:egfp)* containing the risk allele (rs1568679) in the brain of *olig2:dsred2* transgenic larvae at 72 hpf. (E) Confocal projections of EGFP immunolabeling in *Tg(CNE18:egfp)* containing the protective allele or *Tg(CNE18SNP:egfp)* containing the risk allele (rs16932455) in retina immunolabeled with HuC/D at 48 hpf. Dorsal view of the brain with anterior up. Lateral view of the retina. Scale bars: 100 μm.

CNE1/rs17421627, associated to human retinal vasculature defects. To completely remove the transcriptional component due CNE1 enhancer activity and investigate its impact in retinal blood vessel formation, we deleted CNE1 in the zebrafish genome. Taking advantage of the CRISPR/Cas-9 system (38), we deleted most of the CNE1 enhancer (ΔCNE1: 770 bp, Figure 3A), including the deeply conserved region containing the human SNP position (Figure 1B). We next generated ΔCNE1 homozygous mutants and siblings to examine potential phenotypes in the eye. During the first days of development, ΔCNE1 mutants are indistinguishable from their wild-type siblings and display normal eye morphology (Figure 3B). However, when we used FITC-dextran as a tracer to check for vascular network integrity, micro-angiography of homozygous ΔCNE1 larvae revealed a disrupted blood vessel network and vascular defects in the retina (Figure 3C). We further characterized and quantified this phenotype using a *kdrl* (*vegfr2*) reporter line labeling the vasculature (40). Consistent with the human phenotype, we observed a reduction of ∼40% in the number of vessels and branching points and abnormal vascular caliber in homozygous ΔCNE1 mutants compared to their control siblings (Figure 3D and E). These results demonstrate that CNE1 is functionally associated with blood vessel development in the retina and support that SNP rs17421627 is a functional mutation affecting CNE1 activity and potentially involved in retinal vascular phenotype in humans.

CNE1 enhancer regulates *MIR-9-2/miR-9-5* expression

CNE1/rs17421627 occupies open chromatin in human fetal brain cells and is associated with brain expression in human by the GTEx project (53) (Figure 1C and Supplementary Table S2), suggesting that it is an active enhancer and regulates a syntenically-conserved neighboring gene in the central nervous system. Consistent with this observation, the human CNE1 (948 bp) used to establish *CNE1:egfp* transgenic zebrafish drove EGFP expression in the brain and retina (Figure 4B; Supplementary Figure S1b and c). The SNP is located ∼283 and ∼352 kb away from the closest neighboring protein-coding genes in the human genome, *TMEM161B* and *MEF2C* respectively (Figure 4A). In both GWAS studies that replicated rs17421627 (28,29), the authors hypothesized *MEF2C* to be the affected target gene because of its role in cardiovascular development (54). However, the *CNE1:egfp* expression profile matched neither the *tmem161b* nor *mef2cb* mRNA expression patterns revealed by whole mount ISH at the same stage (Figure 4C). Our own *in silico* predictions, which included both coding and non-coding gene synteny checks, suggested a third candidate target gene of CNE1, the *MIR-9-2* microRNA (Figure 1A). This microRNA was originally not well annotated, which may explain its absence in the original GWAS reports. *MIR-9-2* is located ∼147 kb away from CNE1 in the human genome and is in fact closer to CNE1 than both *MEF2C* and *TMEM161B* (Figure 4A). *miR-9-5*, the zebrafish ortholog of the human *MIR-9-2*, is also the closest conserved adjacent gene to CNE1 in the zebrafish genome (Figures 1C and 4A). We next performed ISH against *miR-9* in zebrafish larvae and adults, and we found that *miR-9* is expressed in the brain and retina in a pattern very similar to *CNE1:egfp* (Figure 4B–D and Supplementary Figure S1a–c). Importantly, *miR-9* colocalized with EGFP in the *CNE1:egfp*+ cells in the brain and retina (Figure 4D; Supplementary Figure S1b and c), supporting that *MIR-9-2/miR-9-5* could be a *cis*-regulatory target of CNE1. Consistent with a role in human retina vascular disease, *miR-9* expression has been reported in human retina extracts (55). As expected, we found that *miR-9* genes expression is also conserved in the inner nuclear layer of both the zebrafish and mouse retina (Figure 4E).

We next tested whether CNE1 is an actual *cis*-regulator of *MIR-9-2/miR-9-5* by examining *miR-9-5* expression in ΔCNE1 homozygous mutant larvae. In both human and zebrafish, *MIR-9-2/miR-9-5* is part of long non-coding precursor RNAs (pri-miRNAs) (56,57), *LINC00461* and *si:ch211-202e12.3* respectively (Figures 1C and 5A). In both species CNE1 is intronic to these pri-miRNA genes (Figures 1C and 5A), suggesting that CNE1 could act as an enhancer regulating these precursors. In zebrafish, it was recently shown that *miR-9-5* comes from the processing of the long host pri-miRNA (*si:ch211-202e12.3*) but also from an alternative intronic promoter located 1.1 kb upstream of the pre-*miR-9-5* sequence (46). Nepal *et al.* showed that transcripts from the host and intronic promoters have the same onset of expression (14 somite stage) and the same spatial distribution in the CNS, suggesting transcriptional co-regulation. Using similar riboprobes, we found a reduction in the expression of both *miR-9-5* precursors in the homozygous ΔCNE1 mutant compared to their control siblings (Figure 5B and C), suggesting that CNE1 acts as an enhancer shared by both the host and intronic promoters. We then confirmed that the downregulation of *miR-9-5* pri-precursors in ΔCNE1 led to a specific reduction of pre-*miR-9-5* by using a LNA probe that specifically distinguished pre-*miR-9-5* from the other mature miR-9 sequences (Figure 5D). Most importantly, we showed that at this stage, CNE1 likely acts as a specific enhancer of *MIR-9-2/miR-9-5,* as *tmem161b* and *mef2cb* expression patterns were unaffected in ΔCNE1 mutants (Figure 5E and F). Thus, these data suggest that a downregulation of *MIR-9-2* expression can contribute to human retinal vasculature defects associated to the rs17421627 mutation.

miR-9 depletion leads to retinal vasculature formation defects

To fully establish the functional role of *miR-9* in the retinal vascular phenotype, we first analyzed the expression of the pre-*miR-9-5* in the retina of homozygous ΔCNE1 mutant larvae. We found that *miR-9-5* expression was absent or strongly reduced in 88% of the homozygous mutant retinas ($n = 43$), while unaffected in their control siblings (Figure 6A). Then, to directly support the role of miR-9 in retinal angiogenesis *in vivo*, we used the previously described and extensively validated miR-9 morpholino (45,48–51) to deplete its expression in *kdrl:mCherry* fish revealing the entire eye vasculature (40). miR-9 depleted larvae displayed normal brain and eye morphology (Figure 6B), but the hyaloid vasculature in the retina was strongly reduced with abnormal vascular caliber (Figure 6C), similarly to the human phenotype. We further quantified this reduction in blood

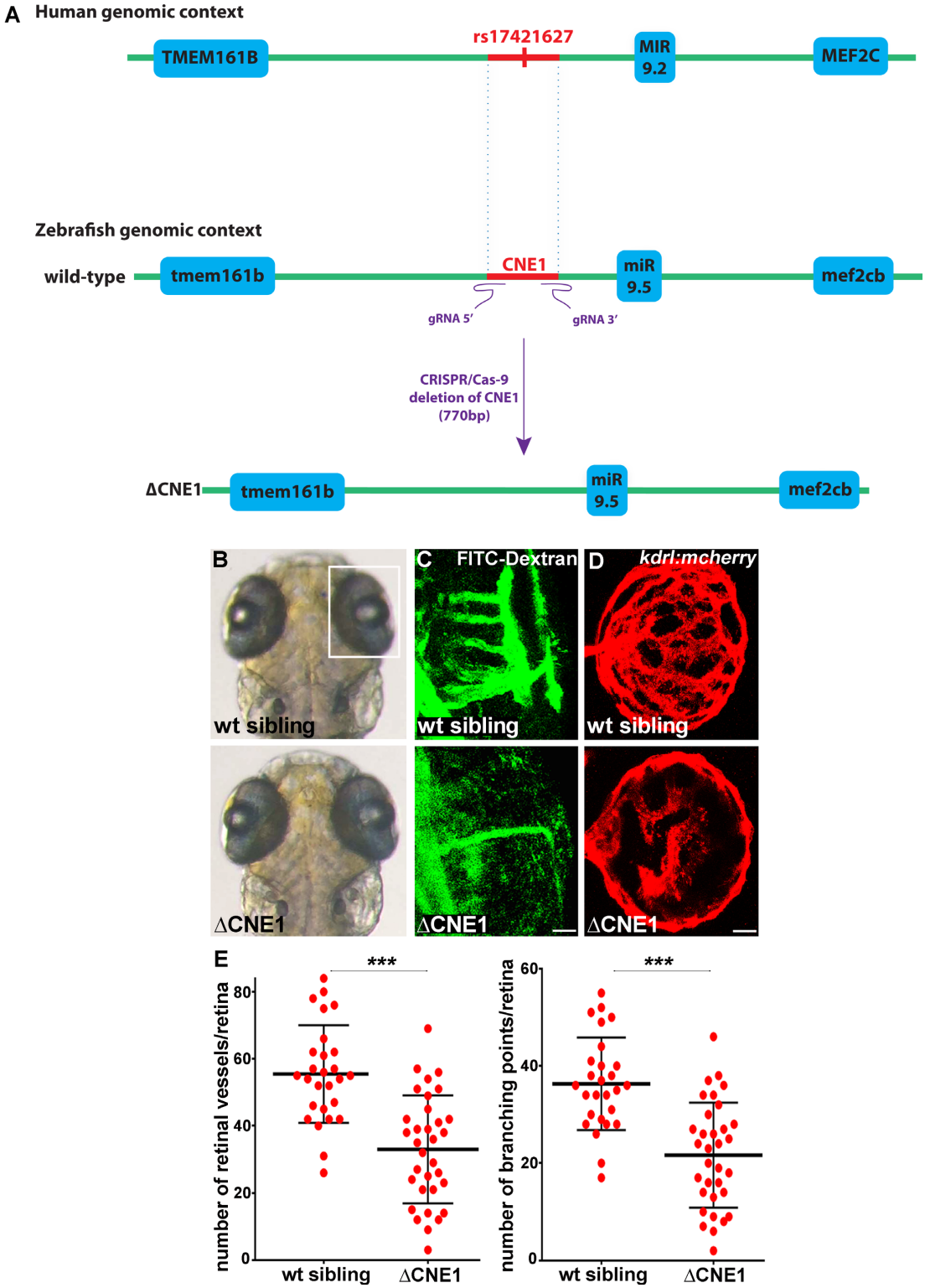


Figure 3. CNE1 regulates retinal vasculature formation *in vivo*. (**A**) Schematic of the human and zebrafish genomic DNA containing CNE1, which was deleted using the CRISPR/Cas-9 system with a pair of gRNAs (ΔCNE1 in the zebrafish genome). The deletion is 770 bp long, including the deeply conserved sequence of CNE1 (473 bp; danRer7. chr5:49,928,049–49,928,521) containing the SNP. (**B** and **C**) CNE1 homozygous mutants display normal brain and eye morphology (B), but the formation of the hyaloid vasculature in the retina is affected, as revealed by microangiography using FITC-dextran injection (C). The retinal vasculature shown in (C) are from larvae shown in (B). (**D**) Confocal projections of mCherry immunolabeling in *Tg(kdrl:mCherry)* retina at 72 hpf showing hyaloid vasculature formation in control and ΔCNE1 mutant larvae. (**E**) Quantification of the hyaloid vasculature network organization observed in control and homozygous ΔCNE1 mutant larvae at 72 hpf. A minimum of 28 retinas was analyzed for each context. Dorsal view of the brain with anterior up. Lateral view of the retina. Scale bars: 10 μm. Error bars represent s.d. $*P < 0.05$, $**P < 0.001$, $***P < 0.0005$, determined by *t*-test, two-tailed.

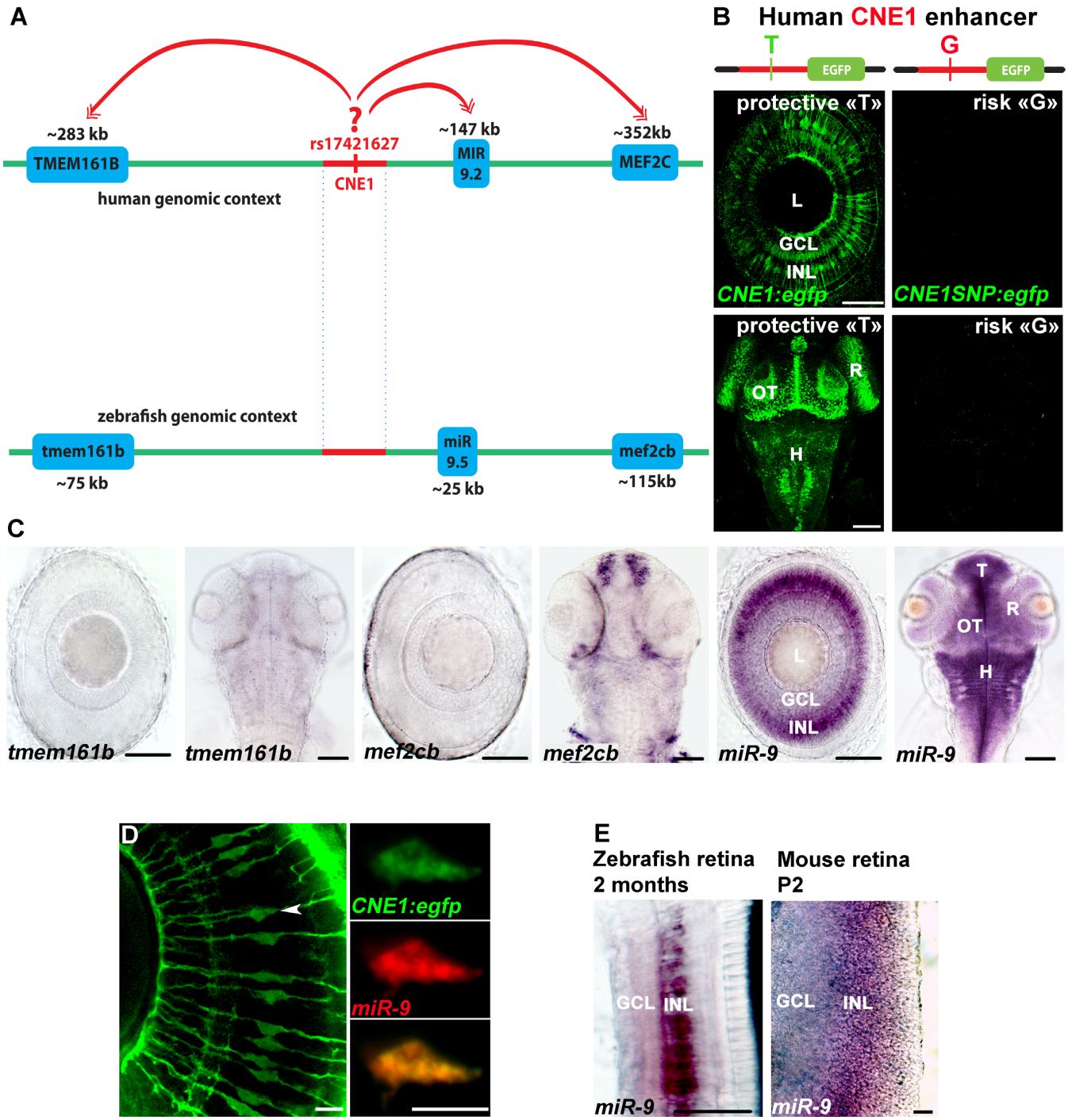


Figure 4. CNE1 enhancer activity is similar to *miR-9* expression, but not *mef2cb* and *tmem161b*. (**A**) Schematic of the human and zebrafish genomic DNA containing CNE1 where SNP rs17421627 is located. (**B**) Confocal projections of EGFP immunolabeling in *Tg(CNE1:egfp)* containing the protective T allele or *Tg(CNE1SNP:egfp)* containing the risk G allele (rs17421627) in larvae at 72 hpf. In *Tg(CNE1:egfp)*, EGFP expression is detected in cell bodies in the inner nuclear layer of the retina, telencephalon, optic tectum and hindbrain. The *Tg(CNE1:egfp)* expression pattern is reminiscent of *miR-9* expression. EGFP expression in the CNS is abolished in the presence of the SNP rs17421627 risk allele. (**C**) Whole-mount ISH against *tmem161b, mef2cb* and *miR-9* in larvae at 72 hpf. (**D**) Confocal section of double *in situ*/immunolabeling showing co-localization in the expression of endogenous *miR-9* and EGFP protein in *Tg(CNE1:egfp)* retina at 72 hpf. (**E**) Whole-mount ISH against miR-9 in zebrafish (2 months old) or mouse (P2 stage) showing similar expression in the inner nuclear layer of the retina. Retina (R), Ganglion cells layer (GCL), Inner nuclear layer (INL), Telencephalon (T), Optic Tectum (OT), Hindbrain (H). Dorsal view of the brain with anterior up. Lateral view of the retina. Scale bars: 100 μm.

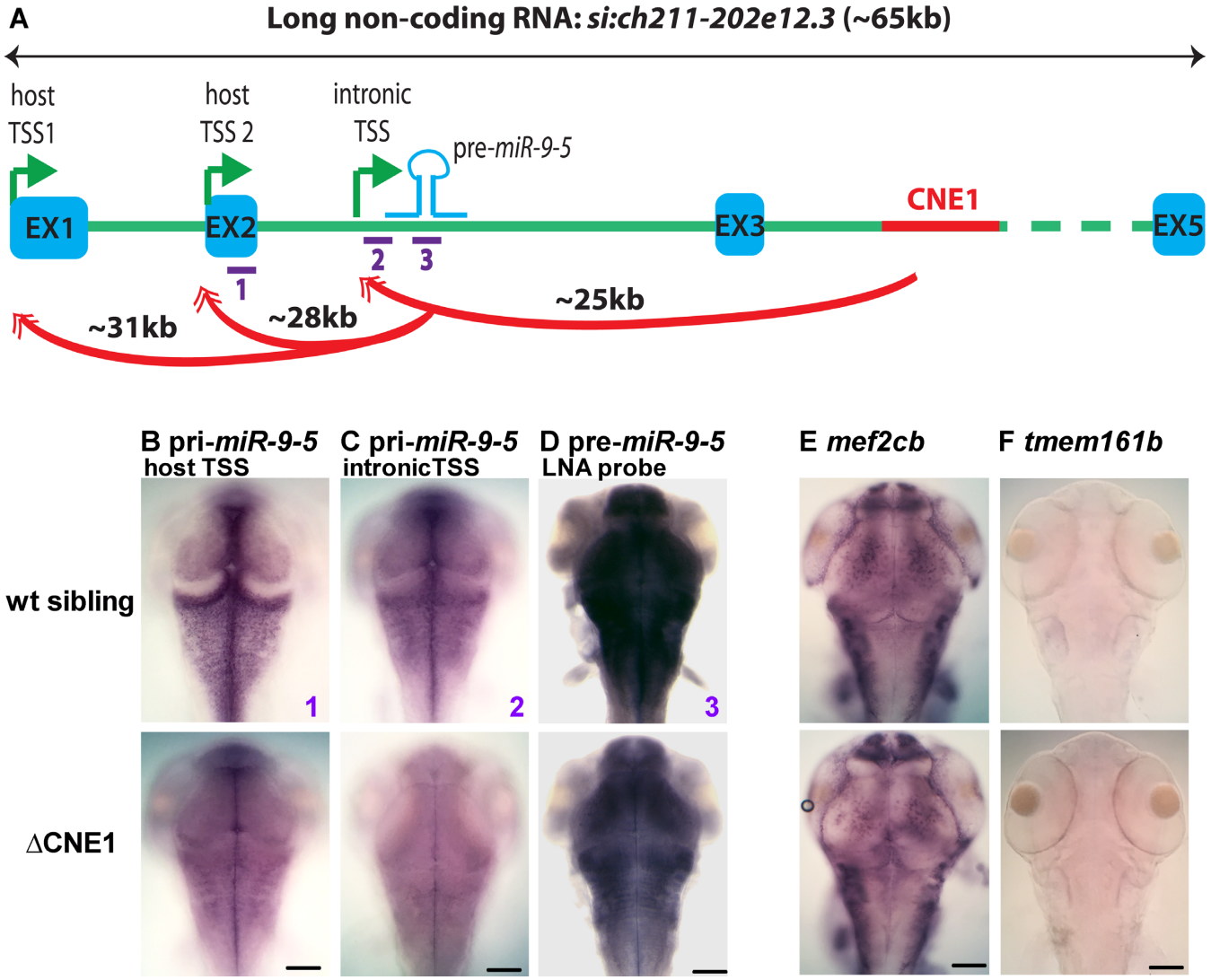


Figure 5. CNE1 regulates *MIR-9-2*/*miR-9-5* expression. (**A**) Schematic of the zebrafish genomic region containing pre-*miR-9-5* and CNE1. pre-*miR-9-5* and CNE1 are intronic in *si:ch211-202e12.3*, a long non-coding RNA. Three transcription starts (TSS, green arrows), host and intronic, are used for pri-*miR-9-5* transcription (46). (**B**–**F**) Whole-mount ISH against host and intronic pri-*miR-9-5s* (probe 1 and 2, respectively), pre-*miR-9-5* (probe 3), *tmem161b* or *mef2c* in control and ΔCNE1 larvae at 72 hpf. The expression of pri- and pre-*miR-9-5*, but not *tmem161b* and *mef2c*, is reduced in the brain of homozygous ΔCNE1 mutants. Dorsal view of the brain with anterior up. Scale bars: 100 μm.

vessel branching complexity and observed a decrease of ∼50% in the number of retinal vessels and branching points (Figure 6D), indicating that the vascular network organization is affected in the retina in absence of miR-9. This observation is consistent with a *cis*-regulation of *miR-9* by CNE1 and with the ΔCNE1 phenotype. Altogether these results indicate that *miR-9* expression is required for the normal development of the retinal vasculature and that its downregulation due to the abolition of CNE1 activity by the rs17421627 mutation is potentially a genetic factor contributed to retinal blood vessel defects in humans.

DISCUSSION

rs17421627 is a functional mutation likely regulating *MIR-9-2* and retinal vasculature formation in human

In this study, we computationally identified 45 non-coding CNE/SNP pairs deeply conserved across vertebrate evolution and associated with human traits by GWAS. This screen pinpoints high-confidence candidates for functional non-coding mutations and provides strong clues unmasking the *cis*-regulated, syntenic genes that may be involved in different human diseases. While the large evolutionary distance between human and zebrafish produces a small, although tractable, number of candidates for further experimentation, we expect the deep conservation to identify a high proportion of functional non-coding *cis*-regulatory sequences. By comparison, the evolutionary distance of 300 million years between human and chicken (versus 450 mil-

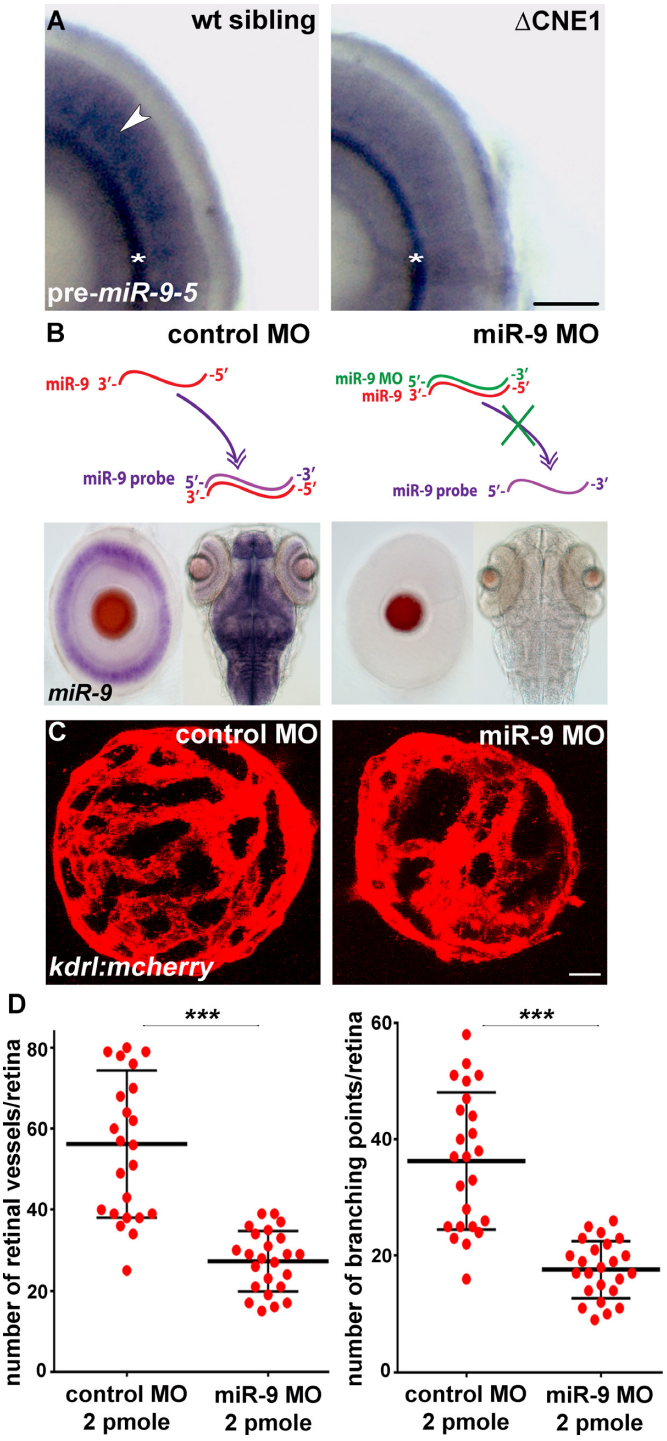


Figure 6. miR-9 controls retinal vasculature development. (**A**) Whole-mount ISH against pre-*miR-9-5* in control and ΔCNE1 mutant retina at 72 hpf. In 88% of the mutant retinas ($n = 43$), *miR-9-5* expression is reduced or absent, compared to the control larvae ($n = 34$). Arrowheads show *miR-9-5* expression in the inner nuclear layer of the retina. Asterisk shows background staining with the pre-*miR-9-5* LNA probe that is not observed with the LNA probe targeting the mature miR-9 sequence (see Figure 4C and panel (B)). (**B**) Schematic representation of the miR-9 MO binding to microRNA-9. With the control MO, miR-9 is freely available revealing the miR-9 expression pattern with the specific miR-9 LNA probe. In presence of the miR-9 MO, miR-9 is bound by the MO inhibiting the binding of the LNA probe. miR-9 morphant larvae show no obvious defects in the brain and eye morphogenesis compared to the control MO. (**C**) Confocal projections of mCherry immunolabeling in *Tg(kdrl:mCherry)* retina at 72 hpf showing hyaloid vasculature formation in control MO and miR-9 MO injected larvae. (**D**) Quantification of the hyaloid vasculature network organization observed in the control MO or the miR-9 MO at 72 hpf. A minimum of 24 retinas was analyzed for each context. Dorsal view of the brain with anterior up. Lateral view of the retina. Scale bars: 100 μm (A) or 10 μm (C). Error bars represent s.d. $*P < 0.05$, $**P < 0.001$, $***P < 0.0005$, determined by *t*-test, two-tailed.

lion years with zebrafish) led to the identification of 558 conserved CNE/SNPs pairs (Supplementary Table S3 and Methods). We experimentally validated the regulatory activity of some of these polymorphisms, and more importantly, provided evidence *in vivo* for the biological role of rs17421627 in the associated phenotype, retinal vasculature defects in human (28,29). Previously, the function of this SNP and the *cis*-regulated gene involved in this biological process were unknown. Here, we demonstrated that rs17421627 abolishes the enhancer activity of a deeply conserved non-coding element (CNE1) *cis*-regulating *MIR-9-2* in human. In addition, the deletion of CNE1 and the depletion of *miR-9* lead to a similar phenotype, suggesting that the downregulation of *MIR-9-2* expression is responsible of the retinal vascular defects observed in patient carriers of the rs17421627 risk allele. Our work illustrates how the combination of human and model organism genetics can reveal the biological function of non-coding GWAS-SNPs, which represent a major challenge for the future of biomedical research.

The activity of the CNE1 enhancer during development in zebrafish and human fetal brain, as indicated by epigenetic marks, suggest that developmental defects could be responsible for the retinal vascular phenotype observed in older carriers of the rs17421627 risk allele. Because changes in retinal vascular caliber have been linked with increased cardiovascular risk and are predictive of global vascular pathology (58,59), and because aging is a strong risk factor for developing vascular disorders, it is possible that early developmental conditions associated with rs17421627 increase the probability of developing a vascular pathology in the retina later in life, but also in the rest of the central nervous system, where CNE1 is also active.

Our study illustrates the necessity of working with an animal model system, in addition to cell culture, to reveal the biological effect of GWAS-SNPs at the whole organism level. Our experiments revealed numerous insights into the functional impact of these non-coding variants. First, a human CNE of only several hundred bases can drive a very precise expression pattern in space and time during development, demonstrating the specificity of the enhancer. Second, the mutagenic potential of risk polymorphisms can be directly tested in fish allowing us to identify functional mutations. Third, by comparing the enhancer reporter expression with the mRNA distribution of the neighboring genes and by deleting the CNE in the zebrafish genome, we identified the actual *cis*-regulated gene. Finally, our findings suggest that the vasculature defect studied in the GWAS is likely the result of *MIR-9-2* expression downregulation in neural cells and uncover a non-cell autonomous func-

tion for miR-9 during physiological angiogenesis *in vivo*. In contrast to its role during angiogenesis, miR-9 is a well-known regulator of neurogenesis (60). *In vivo*, miR-9 has been shown to control the proliferation and differentiation of neural stem cells and its inhibition transiently increases their proliferation, ultimately resulting in an increased number of late-differentiating neurons (49,50,61). Because *miR-9* is expressed in neural stem cells and regulates the neuronal differentiation (60,62), we hypothesize that it may control the development of the retinal vasculature by modulating neural stem cells fate and/or the formation of neurons, that are involved in blood vessels development (62–66).

Deeply conserved SNPs are functional variants underlying the human phenotype

The overwhelming majority of the GWAS-SNPs identified occupy non-genic portions of the genome that do not result in an obvious disruption in coding sequence, making them challenging to interpret. Furthermore, GWAS studies report tag SNPs that may be merely co-segregating markers with the causal mutation (7,8). The evolutionary distance between human and zebrafish can reveal functional non-coding SNPs of importance for gene *cis*-regulation (21). A hallmark of *cis*-regulatory sequence conservation is that it persists alongside the gene or genes it regulates, even in species as diverged as human and zebrafish (26), a property which we exploited in our computational screen. By identifying aligning non-coding regions adjacent to orthologous genes, we predicted the putative target genes whose expression may be modulated by the regulatory SNPs. In some cases, our predicted target genes (but not our CNEs) have already been implicated in the GWAS phenotype itself. For example, the predicted target gene of CNE12/rs13095226, associated with macular degeneration, is *COL8A1*, in which mutations are known to cause cornea abnormalities (67). Likewise, we link severe myopia CNE13/rs13382811 to *ZEB2*, in which mutations are known to deform lens morphology (68).

In addition to CNE1/rs17421627, we validated the specific regulatory function of GWAS-SNPs, embedded in other conserved CNEs. CNE18/rs16932455 drives EGFP expression in retinal ganglion cells of the retina and in muscles of the trunk, where *sox6* mRNA is also expressed and is critical during development (69). CNE10 activity is observed in the notochord, and *spry2*, the predicted *cis*-regulated gene, is known to be involved in the formation of this structure (70). Furthermore, rs1568679 is associated to antipsychotic response, and the EGFP driven by CNE8 co-localizes with oligodendrocytes in the zebrafish brain. Finally, a recent study demonstrated that rs11190870 (embedded in CNE15) regulates *LBX1* and is likely involved in the associated scoliosis disease (17). While a CNE/SNP pair can potentially regulate transcription at very long distances, it appears that the CNE/SNP pairs validated in this study control the expression of the nearest adjacent gene (CNE1/*MIR9-2,* CNE8/*MEIS2* CNE15/*LBX1* and CNE18/*SOX6*). While many human traits are likely multifactorial and are dependent of a combination of mutations affecting the activity of different regulatory elements, we observed that in the case of rs17421627 and rs11190870 regulating *MIR-9-2* and *LBX1* respectively, these non-coding mutations are likely one of the genetic components underpinning a biological process associated with the human disease.

Because polymorphisms identified as susceptibility loci in GWASs can affect the binding of particular transcription factors, we computationally predicted the transcription factors that bind to the conserved CNEs (Supplementary Table S4). As many of our CNEs are deeply conserved across vertebrates with a high-sequence identity, conserved binding site methods that rely on transcription factor binding sites being more conserved than the surrounding genomic neighborhood (71) offer little assistance. For this reason, we are limited to single sequence (human) binding site prediction. In the case of rs2168101, which is associated with sporadic neuroblastoma, our top prediction suggested a GATA transcription factor family binding site, and it was recently demonstrated that the rs2168101 reduces GATA3 binding to the LMO1 super-enhancer (19). In the case of CNE1/rs17421627, we predicted an homeodomain transcription factor binding site containing the polymorphism, but we were not able to experimentally validate the activity of TGIF1, PKNOX2, MEIS2 or SIX3 on this binding site by gain of function. This observation indicates either that the computational prediction is wrong for this specific CNE/SNP pair or that additional factors are required in combination with the homeodomain transcription factor to bind and/or activate CNE1. Further functional investigations are needed to reveal the mechanisms of these 45 CNE/SNP pairs relevant for human health.

While our approach cannot comprehensively address the problem of interpreting all non-genic GWAS-SNPs, it provides a tractable approach for functionally studying a subset of them in zebrafish. The 45 non-coding GWAS-SNPs revealed here are associated with a wide variety of human phenotypes and diseases, which range from developmental defects to psychiatric disorders. The value of experimentally testing each one of them to reveal the molecular mechanisms underlying the associated disease is immense. Studying these polymorphic sites in zebrafish or other amenable animal models provides an attractive first step in deciphering the biological mechanisms underpinning the human traits, allowing us to identify new molecular targets and treatment strategies in the future.

DATA AVAILABILITY

rsIDs were extracted from the NHGRI GWAS catalog (http://www.genome.gov/gwastudies/) and mapped to genomic loci in the GRCh37 human reference (hg19) using the dbSNP 137 database from the UCSC Genome Browser. Genome coordinates and statistics are listed in Supplementary Table S1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Drs Laure Bally-Cuif, Thomas S. Becker and Ben A. Barres for critical reading of the manuscript.

We thank Drs Jeffrey L. Goldberg, Steven Sloan, Louis Leung, Aaron Wenger, Geetu Tuteja, Harendra Guturu, Evan G. Cameron and Ana Meireles for insightful discussions and help. We are grateful to Drs Neil C. Chi, David Traver, Nathan Lawson and William S. Talbot for gift of reagents. We would also like to thank the Stanford CSIF Imaging platform.

Authors contributions: J.H.N., C.C.C. and G.B. designed the computational study, wrote the algorithms and analyzed the genomic data. R.M. and P.M. designed and analyzed all the *in vivo* experiments. R.M. performed all the *in vivo* experiments with technical support of G.S. and C.H. R.M., J.H.N., G.B. and P.M wrote the manuscript.

FUNDING

National Institute of Health [NS104950, MH099647 to P.M., HG005058 to J.H.N., C.C.C., G.B.]; BrightFocus Foundation (to P.M., R.M.); Simons Foundation and John Merck Fund (to P.M.); EMBO Long Term Fellowship [ALTF 413-2012 to R.M.]; National Science Foundation Fellowship [DGE-1147470 to J.H.N.]; Bio-X Stanford Interdisciplinary Graduate Fellowship (to J.H.N.); Packard Fellowship (to G.B.). Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Cargill,M., Altshuler,D., Ireland,J., Sklar,P., Ardlie,K., Patil,N., Shaw,N., Lane,C.R., Lim,E.P., Kalyanaraman,N. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, **22**, 231–238.
2. Lek,M., Karczewski,K.J., Minikel,E.V., Samocha,K.E., Banks,E., Fennell,T., O'Donnell-Luria,A.H., Ware,J.S., Hill,A.J., Cummings,B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
3. Khurana,E., Fu,Y., Chakravarty,D., Demichelis,F., Rubin,M.A. and Gerstein,M. (2016) Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.*, **17**, 93–108.
4. Ward,L.D. and Kellis,M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.*, **30**, 1095–1106.
5. Schaub,M.A., Boyle,A.P., Kundaje,A., Batzoglou,S. and Snyder,M. (2012) Linking disease associations with regulatory information in the human genome. *Genome Res.*, **22**, 1748–1759.
6. Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S. and Manolio,T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9362–9367.
7. Nicolae,D.L., Gamazon,E., Zhang,W., Duan,S., Dolan,M.E. and Cox,N.J. (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.*, **6**, e1000888.
8. Tuupanen,S., Turunen,M., Lehtonen,R., Hallikas,O., Vanharanta,S., Kivioja,T., Bjorklund,M., Wei,G., Yan,J., Niittymaki,I. *et al.* (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.*, **41**, 885–890.
9. Orkin,S.H., Antonarakis,S.E. and Kazazian,H.H. Jr (1984) Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. *J. Biol. Chem.*, **259**, 8679–8681.
10. Bond,G.L., Hu,W., Bond,E.E., Robins,H., Lutzker,S.G., Arva,N.C., Bargonetti,J., Bartel,F., Taubert,H., Wuerl,P. *et al.* (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell*, **119**, 591–602.
11. Horn,S., Figl,A., Rachakonda,P.S., Fischer,C., Sucker,A., Gast,A., Kadel,S., Moll,I., Nagore,E., Hemminki,K. *et al.* (2013) TERT promoter mutations in familial and sporadic melanoma. *Science*, **339**, 959–961.
12. Ziller,M.J., Gu,H., Muller,F., Donaghey,J., Tsai,L.T., Kohlbacher,O., De Jager,P.L., Rosen,E.D., Bennett,D.A., Bernstein,B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477–481.
13. Ryan,B.M., Robles,A.I., McClary,A.C., Haznadar,M., Bowman,E.D., Pine,S.R., Brown,D., Khan,M., Shiraishi,K., Kohno,T. *et al.* (2015) Identification of a functional SNP in the 3′UTR of CXCR2 that is associated with reduced risk of lung cancer. *Cancer Res.*, **75**, 566–575.
14. Guenther,C.A., Tasic,B., Luo,L., Bedell,M.A. and Kingsley,D.M. (2014) A molecular basis for classic blond hair color in Europeans. *Nat. Genet.*, **46**, 748–752.
15. Praetorius,C., Grill,C., Stacey,S.N., Metcalf,A.M., Gorkin,D.U., Robinson,K.C., Van Otterloo,E., Kim,R.S., Bergsteinsdottir,K., Ogmundsdottir,M.H. *et al.* (2013) A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell*, **155**, 1022–1033.
16. Spieler,D., Kaffe,M., Knauf,F., Bessa,J., Tena,J.J., Giesert,F., Schormair,B., Tilch,E., Lee,H., Horsch,M. *et al.* (2014) Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res.*, **24**, 592–603.
17. Guo,L., Yamashita,H., Kou,I., Takimoto,A., Meguro-Horike,M., Horike,S., Sakuma,T., Miura,S., Adachi,T., Yamamoto,T. *et al.* (2016) Functional investigation of a non-coding variant associated with adolescent idiopathic scoliosis in zebrafish: elevated expression of the ladybird homeobox gene causes body axis deformation. *PLoS Genet.*, **12**, e1005802.
18. Huang,Q., Whitington,T., Gao,P., Lindberg,J.F., Yang,Y., Sun,J., Vaisanen,M.R., Szulkin,R., Annala,M., Yan,J. *et al.* (2014) A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat. Genet.*, **46**, 126–135.
19. Oldridge,D.A., Wood,A.C., Weichert-Leahey,N., Crimmins,I., Sussman,R., Winter,C., McDaniel,L.D., Diamond,M., Hart,L.S., Zhu,S. *et al.* (2015) Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature*, **528**, 418–421.
20. Claussnitzer,M., Dankel,S.N., Kim,K.H., Quon,G., Meuleman,W., Haugen,C., Glunk,V., Sousa,I.S., Beaudry,J.L., Puviindran,V. *et al.* (2015) FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.*, **373**, 895–907.
21. Hiller,M., Agarwal,S., Notwell,J.H., Parikh,R., Guturu,H., Wenger,A.M. and Bejerano,G. (2013) Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: application to zebrafish. *Nucleic Acids Res.*, **41**, e151.
22. Navratilova,P., Fredman,D., Hawkins,T.A., Turner,K., Lenhard,B. and Becker,T.S. (2009) Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev. Biol.*, **327**, 526–540.
23. Shin,J.T., Priest,J.R., Ovcharenko,I., Ronco,A., Moore,R.K., Burns,C.G. and MacRae,C.A. (2005) Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. *Nucleic Acids Res.*, **33**, 5437–5445.
24. Taher,L., McGaughey,D.M., Maragh,S., Aneas,I., Bessling,S.L., Miller,W., Nobrega,M.A., McCallion,A.S. and Ovcharenko,I. (2011) Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res.*, **21**, 1139–1149.
25. Hufton,A.L., Mathia,S., Braun,H., Georgi,U., Lehrach,H., Vingron,M., Poustka,A.J. and Panopoulou,G. (2009) Deeply conserved chordate noncoding sequences preserve genome synteny but do not drive gene duplicate retention. *Genome Res.*, **19**, 2036–2051.
26. Kikuta,H., Laplante,M., Navratilova,P., Komisarczuk,A.Z., Engstrom,P.G., Fredman,D., Akalin,A., Caccamo,M., Sealy,I., Howe,K. *et al.* (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.*, **17**, 545–555.

27. Woolfe,A. and Elgar,G. (2008) Organization of conserved elements near key developmental regulators in vertebrate genomes. *Adv. Genet.*, **61**, 307–338.
28. Ikram,M.K., Sim,X., Jensen,R.A., Cotch,M.F., Hewitt,A.W., Ikram,M.A., Wang,J.J., Klein,R., Klein,B.E., Breteler,M.M. *et al.* (2010) Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet.*, **6**, e1001184.
29. Sim,X., Jensen,R.A., Ikram,M.K., Cotch,M.F., Li,X., MacGregor,S., Xie,J., Smith,A.V., Boerwinkle,E., Mitchell,P. *et al.* (2013) Genetic loci for retinal arteriolar microcirculation. *PLoS One*, **8**, e65804.
30. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.
31. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.
32. Wheeler,T.J. and Eddy,S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, **29**, 2487–2489.
33. Newburger,D.E. and Bulyk,M.L. (2009) UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.*, **37**, D77–D82.
34. Bryne,J.C., Valen,E., Tang,M.H., Marstrand,T., Winther,O., da Piedade,I., Krogh,A., Lenhard,B. and Sandelin,A. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.*, **36**, D102–D106.
35. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
36. Guturu,H., Doxey,A.C., Wenger,A.M. and Bejerano,G. (2013) Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **368**, 20130029.
37. Kel,A.E., Gossling,E., Reuter,I., Cheremushkin,E., Kel-Margoulis,O.V. and Wingender,E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–3579.
38. Varshney,G.K., Pei,W., LaFave,M.C., Idol,J., Xu,L., Gallardo,V., Carrington,B., Bishop,K., Jones,M., Li,M. *et al.* (2015) High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.*, **25**, 1030–1042.
39. Kimmel,C.B., Ballard,W.W., Kimmel,S.R., Ullmann,B. and Schilling,T.F. (1995) Stages of embryonic development of the zebrafish. *Dev. Dyn.*, **203**, 253–310.
40. Chi,N.C., Shaw,R.M., De Val,S., Kang,G., Jan,L.Y., Black,B.L. and Stainier,D.Y. (2008) Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. *Genes Dev.*, **22**, 734–739.
41. Kucenas,S., Takada,N., Park,H.C., Woodruff,E., Broadie,K. and Appel,B. (2008) CNS-derived glia ensheath peripheral nerves and mediate motor root development. *Nat. Neurosci.*, **11**, 143–151.
42. Birnbaum,R.Y., Everman,D.B., Murphy,K.K., Gurrieri,F., Schwartz,C.E. and Ahituv,N. (2012) Functional characterization of tissue-specific enhancers in the DLX5/6 locus. *Hum. Mol. Genet.*, **21**, 4930–4938.
43. Kwan,K.M., Fujimoto,E., Grabher,C., Mangum,B.D., Hardy,M.E., Campbell,D.S., Parant,J.M., Yost,H.J., Kanki,J.P. and Chien,C.B. (2007) The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.*, **236**, 3088–3099.
44. Oxtoby,E. and Jowett,T. (1993) Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. *Nucleic Acids Res.*, **21**, 1087–1095.
45. Leucht,C., Stigloher,C., Wizenmann,A., Klafke,R., Folchert,A. and Bally-Cuif,L. (2008) MicroRNA-9 directs late organizer activity of the midbrain-hindbrain boundary. *Nat. Neurosci.*, **11**, 641–648.
46. Nepal,C., Coolen,M., Hadzhiev,Y., Cussigh,D., Mydel,P., Steen,V.M., Carninci,P., Andersen,J.B., Bally-Cuif,L., Muller,F. *et al.* (2016) Transcriptional, post-transcriptional and chromatin-associated regulation of pri-miRNAs, pre-miRNAs and moRNAs. *Nucleic Acids Res.*, **44**, 3070–3081.
47. Masai,I., Heisenberg,C.P., Barth,K.A., Macdonald,R., Adamek,S. and Wilson,S.W. (1997) Floating head and masterblind regulate neuronal patterning in the roof of the forebrain. *Neuron*, **18**, 43–57.
48. Garaffo,G., Conte,D., Provero,P., Tomaiuolo,D., Luo,Z., Pinciroli,P., Peano,C., D'Atri,I., Gitton,Y., Etzion,T. *et al.* (2015) The Dlx5 and Foxg1 transcription factors, linked via miRNA-9 and -200, are required for the development of the olfactory and GnRH system. *Mol. Cell. Neurosci.*, **68**, 103–119.
49. Bonev,B., Pisco,A. and Papalopulu,N. (2011) MicroRNA-9 reveals regional diversity of neural progenitors along the anterior-posterior axis. *Dev. Cell*, **20**, 19–32.
50. Coolen,M., Thieffry,D., Drivenes,O., Becker,T.S. and Bally-Cuif,L. (2012) miR-9 controls the timing of neurogenesis through the direct inhibition of antagonistic factors. *Dev. Cell*, **22**, 1052–1064.
51. Katz,S., Cussigh,D., Urban,N., Blomfield,I., Guillemot,F., Bally-Cuif,L. and Coolen,M. (2016) A nuclear role for miR-9 and argonaute proteins in balancing quiescent and activated neural stem cell states. *Cell Rep.*, **17**, 1383–1398.
52. Roadmap Epigenomics,C., Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
53. Consortium,G.T., Laboratory,D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
54. Lin,Q., Schwarz,J., Bucana,C. and Olson,E.N. (1997) Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science*, **276**, 1404–1407.
55. Karali,M., Persico,M., Mutarelli,M., Carissimo,A., Pizzo,M., Singh Marwah,V., Ambrosio,C., Pinelli,M., Carrella,D., Ferrari,S. *et al.* (2016) High-resolution analysis of the human retina miRNome reveals isomiR variations and novel microRNAs. *Nucleic Acids Res.*, **44**, 1525–1540.
56. Rodriguez,A., Griffiths-Jones,S., Ashurst,J.L. and Bradley,A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.
57. Pauli,A., Valen,E., Lin,M.F., Garber,M., Vastenhouw,N.L., Levin,J.Z., Fan,L., Sandelin,A., Rinn,J.L., Regev,A. *et al.* (2012) Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.*, **22**, 577–591.
58. McGeechan,K., Liew,G., Macaskill,P., Irwig,L., Klein,R., Klein,B.E., Wang,J.J., Mitchell,P., Vingerling,J.R., de Jong,P.T. *et al.* (2009) Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. *Am. J. Epidemiol.*, **170**, 1323–1332.
59. Wang,J.J., Liew,G., Klein,R., Rochtchina,E., Knudtson,M.D., Klein,B.E., Wong,T.Y., Burlutsky,G. and Mitchell,P. (2007) Retinal vessel diameter and cardiovascular mortality: pooled data analysis from two older populations. *Eur. Heart J.*, **28**, 1984–1992.
60. Coolen,M., Katz,S. and Bally-Cuif,L. (2013) miR-9: a versatile regulator of neurogenesis. *Front. Cell. Neurosci.*, **7**, 220.
61. Shibata,M., Nakao,H., Kiyonari,H., Abe,T. and Aizawa,S. (2011) MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. *J. Neurosci.*, **31**, 3407–3422.
62. Madelaine,R., Sloan,S.A., Huber,N., Notwell,J.H., Leung,L.C., Skariah,G., Halluin,C., Pasca,S.P., Bejerano,G., Krasnow,M.A. *et al.* (2017) MicroRNA-9 couples brain neurogenesis and angiogenesis. *Cell Rep.*, **20**, 1533–1542.
63. Haigh,J.J., Morelli,P.I., Gerhardt,H., Haigh,K., Tsien,J., Damert,A., Miquerol,L., Muhlner,U., Klein,R., Ferrara,N. *et al.* (2003) Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.*, **262**, 225–241.
64. Mukouyama,Y.S., Shin,D., Britsch,S., Taniguchi,M. and Anderson,D.J. (2002) Sensory nerves determine the pattern of arterial differentiation and blood vessel branching in the skin. *Cell*, **109**, 693–705.
65. Ogunshola,O.O., Antic,A., Donoghue,M.J., Fan,S.Y., Kim,H., Stewart,W.B., Madri,J.A. and Ment,L.R. (2002) Paracrine and autocrine functions of neuronal vascular endothelial growth factor (VEGF) in the central nervous system. *J. Biol. Chem.*, **277**, 11410–11415.

66. Raab,S., Beck,H., Gaumann,A., Yuce,A., Gerber,H.P., Plate,K., Hammes,H.P., Ferrara,N. and Breier,G. (2004) Impaired brain angiogenesis and neuronal apoptosis induced by conditional homozygous inactivation of vascular endothelial growth factor. *Thromb. Haemost.*, **91**, 595–605.
67. Hopfer,U., Fukai,N., Hopfer,H., Wolf,G., Joyce,N., Li,E. and Olsen,B.R. (2005) Targeted disruption of Col8a1 and Col8a2 genes in mice leads to anterior segment abnormalities in the eye. *FASEB J.*, **19**, 1232–1244.
68. Manthey,A.L., Lachke,S.A., FitzGerald,P.G., Mason,R.W., Scheiblin,D.A., McDonald,J.H. and Duncan,M.K. (2014) Loss of Sip1 leads to migration defects and retention of ectodermal markers during lens development. *Mech. Dev.*, **131**, 86–110.
69. Jackson,H.E., Ono,Y., Wang,X., Elworthy,S., Cunliffe,V.T. and Ingham,P.W. (2015) The role of Sox6 in zebrafish muscle fiber type specification. *Skelet. Muscle*, **5**, 2.
70. Sivak,J.M., Petersen,L.F. and Amaya,E. (2005) FGF signal interpretation is directed by Sprouty and Spred proteins during mesoderm formation. *Dev. Cell*, **8**, 689–701.
71. Wenger,A.M., Clarke,S.L., Guturu,H., Chen,J., Schaar,B.T., McLean,C.Y. and Bejerano,G. (2013) PRISM offers a comprehensive genomic approach to transcription factor function prediction. *Genome Res.*, **23**, 889–904.