The human *FLT1* regulatory element directs vascular expression and modulates angiogenesis pathways *in vitro* and *in vivo*

Julian Stolper[1,2], Holly K. Voges[1], Michael See[1], Neda Rahmani Mehdiabadi[1,3], Gulrez Chahal[3], Mark Drvodelic[3,4], Michael Eichenlaub[3], Tanya Labonne[1], Benjamin G. Schultz[5], Alejandro Hidalgo[1], Lazaro Centanin[6], Jochen Wittbrodt[6], Enzo R. Porrello[1,7], David A. Elliott[1,2,3] and Mirana Ramialison,[1,3]**

[1] Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria, 3052, Australia.

[2] Department of Paediatrics, The Royal Children's Hospital, The University of Melbourne, Parkville, Victoria, 3052, Australia.

[3] Australian Regenerative Medicine Institute and Systems Biology Institute Australia, Monash University, Clayton, Victoria 3800, Australia.

[4] University of Melbourne, Parkville, Victoria, 3010, Australia

[5]Centre for Neuroscience of Speech, Department of Audiology and Speech Pathology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Carlton, Victoria, 3053, Australia

[6] Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany

[7] Department of Physiology, School of Biomedical Sciences, The University of Melbourne, Parkville, Victoria, 3010, Australia.

* Corresponding authors: Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria, 3052, Australia. Ph. +61 3 9936 6668. Fax. +61 3 9348 139. Email. david.elliott@mcri.edu.au; mirana.ramialison@mcri.edu.au

Julian Stolper: julian.stolper@mcri.edu.au

Holly K. Voges: holly.voges@mcri.edu.au

Michael See: michael.see@mcri.edu.au

Neda Rahmani Mehdiabadi: neda.rahmanimehdiaba@mcri.edu.au

Gulrez Chahal: gulrez.chahal@monash.edu

Mark Drvodelic: Mdrvodelic@student.unimelb.edu.au

Michael Eichenlaub: michael.eichenlaub@monash.edu

Tanya Labonne: tanya.labonne@mcri.edu.au

Benjamin G. Schultz: ben.schultz@unimelb.edu.au

Alejandro Hidalgo: alejandro.hidalgogon@mcri.edu.au

Lazaro Centanin: lazaro.centanin@cos.uni-heidelberg.de

Jochen Wittbrodt: jochen.wittbrodt@cos.uni-heidelberg.de

Enzo R. Porrello: Enzo.porrello@mcri.edu.au

David A. Elliott: david.elliott@mcri.edu.au

Mirana Ramialison: mirana.ramialison@mcri.edu.au

Abstract

There is growing evidence that mutations in non-coding *cis*-regulatory elements (CREs) disrupt proper development. However, little is known about human CREs that are crucial for cardiovascular development. To address this, we bioinformatically identified cardiovascular CREs based on the occupancy of the CRE by the homeodomain protein NKX2-5 and cardiac chromatin histone modifications. This search defined a highly conserved CRE within the *FLT1* locus termed *enFLT1*. We show that the human *enFLT1* is an enhancer capable of driving reporter transgene expression *in vivo* throughout the developing cardiovascular system of medaka. Deletion of the human *enFLT1* enhancer (*ΔenFLT1*) triggered molecular perturbations in extracellular matrix organisation and blood vessel morphogenesis *in vitro* in endothelial cells derived from human embryonic stem cells and vascular defects *in vivo* in medaka. These findings highlight the crucial role of the human *FLT1* enhancer and its function as a regulator and buffer of transcriptional regulation in cardiovascular development.

Keywords: enhanceropathies, VEGF signalling pathway, *FLT1* enhancer, VEGFR1, angiogenesis, blood vessel.

Introduction

Disruption of heart and major blood vessel formation during development results in congenital heart defects at birth and are a major factor underlying child mortality and morbidity (1). The development of the heart is strictly regulated by a tight network of genetic components (2) which when disrupted perturb normal heart development leading to disease. Genome wide association studies (GWAS) have been used to identify genes associated with CHD and begin to dissect the complex genetic architecture underlying heart development (3,4). Nevertheless, an established problem is definitively assigning pathogenicity to a given variant or single nucleotide polymorphism (SNP). Strikingly, the majority of SNPs associated with CHD are found in non-coding regions in the genome, highlighting the importance of *cis*-regulatory elements in developmental processes and disease (5,6). These regulatory elements (REs), such as enhancers, promote gene expression in a spatial temporal manner through the coordinated binding of specific transcription factors (TFs). However, since SNPs in such sequences can lead to disruptions in TF binding motifs and therefore have no impact on the protein sequence directly, it is challenging to pinpoint the how sequence alterations result in cardiac and blood vessel defects. Aberrations in regulatory elements can lead to the perturbation of a gene regulatory network which, in turn, causes genes to be over- or underexpressed, even for multiple targets at once. This ultimately results in enhanceropathies, a group of diseases caused by mutations in regulatory elements (7).

There is now mounting evidence that disrupted cardiovascular regulatory elements can impair heart development leading to disease (8–11). A prerequisite for functional studies to understand the effect of non-coding SNPs, is the accurate identification of the regulatory

regions in the human genome that are important for cardiac development and disease. Accessible datasets of the human genome, regulatory element associated chromatin marks (12,13), TF analyses (14) and chromatin capture experiments (15–17) provide valuable resources to define the human gene regulatory network in the heart. However, key challenges to identify non-coding elements relevant for disease remain. The search space is still large, for example, the current registry of human *cis*-regulatory elements in the Encode data set is comprised of 926,535 entries (18). Furthermore, it is crucial to have functional validation methods to determine both the sufficiency and necessity of a given human regulatory element for normal development (19).

In this study, we developed a bioinformatic pipeline to identify cardiac enhancers that are involved in development and disease with a particular focus on the highly conserved cardiac TF NKX2-5. NKX2-5 is essential for heart formation and homeostasis and is crucial for the development of heart muscle cells (14,20,21). Since mutations in NKX2-5 can lead to CHD (22), we reasoned that variants in NKX2-5 target enhancers may also impair cardiac development. By using datasets of histone modifications, evolutionary conservation and functional enrichment, we have identified a human enhancer of the *Fms Related Receptor Tyrosine Kinase 1* gene (*FLT1* also known as *Vascular growth factor receptor1* (*VEGFR1*)) termed *enFLT1*. Dysregulation of the *FLT1* genes leads to vascular abnormalities and cardiovascular phenotypes in fish, mice and humans (23–25). Despite extensive study of *FLT1* function, its regulatory network of enhancers has not been completely defined (26–28). How perturbations of these elements may alter the transcriptional regulation of *FLT1* has yet to be determined. Here we demonstrate that the human *enFLT1* enhancer was able to drive gene expression in cardiovascular tissues *in vivo* in medaka (*Oryzias latipes*). The deletion of the enhancer element

in endothelial cells derived from human embryonic stem cells revealed a molecular disruption which overlaps with *FLT1* gene loss-of-function. In addition, medaka *enFLT1* deletion resulted in impaired cardiovascular development *in vivo*. Thus, here were have defined a highly conserved *FLT1* enhancer and provide evidence that this enhancer plays an evolutionarily important role in the development of the cardiovascular system through the modulation of *FLT1* downstream pathways essential for blood vessel morphogenesis.

Results

Identification of *cis*-regulatory elements relevant for cardiac development and disease

In order to identify human *cis*-regulatory elements that are involved in cardiovascular development, we developed a bioinformatic pipeline to filter for sequences that were directly bound by NKX2-5, a TF essential for heart development (Fig 1a). We therefore made use of a previously generated dataset of NKX2-5 genomic targets identified in human pluripotent derived cardiomyocytes by chromatin immunoprecipitation sequencing (ChIP-seq) (14). From all of the ChIP-seq experiments, 20,879 regions were identified to be directly bound by NKX2-5. Since heart development is a highly conserved process, REs deeply embedded in such essential processes are under positive selective pressure compared to non-functional non-coding sequences to maintain correct activity (12). We therefore filtered these regions for high sequence conservation and obtained 62 sequences which were ultra-conserved across 100 vertebrate species, from fish to human. Furthermore, in order to filter for sequences that were shown to be active REs, we used publicly available datasets of histone modification marks as a measure to obtain active cardiac enhancers(13). We identified 38 regions that showed histone

marks for active enhancers H3K4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac). We further filtered these regions for those that could be associated with genes known to play a role in heart development. We identified 7 CREs associated with highly relevant genes expressed in the heart and deeply embedded in the genetic networks controlling heart development (Table S1). To further understand the mechanism of regulatory elements involved in cardiovascular development and to illustrate the evidence supporting our hypothesis, we set out to investigate an enhancer element located in the intron 10 of the gene *FLT1* (Fig. 1b).

The human enhancer of *FLT1* is able to drive cardiovascular gene expression *in vivo*

We set out to assess the *in vivo f*unction of the human enhancer sequence *enFLT1,* which has been validated to be bound by many TFs embedded in cardiovascular development (Fig. S1a). In order to determine whether the RE is able to drive GFP reporter gene expression in the heart, we cloned the enhancer in a modified ZED vector (29,30) to perform a transgenesis assay in medaka (Fig. S1b). Already established in the ENCODE project in 2012, medaka has been shown to be the ideal model to study human regulatory elements *in vivo* and is well suited for genetic engineering (12)**.** Of 121 injected embryos, 52 embryos (43%) showed consistent GFP expression in the cardiovascular system (Fig. S1c). All GFP positive fish were raised and subsequently crossed to wildtype fish to obtain a stable transgenic line (Fig. 1c) denoted *enFLT1:GFP*. Characterisation of the *enFLT1:GFP* line revealed consistent expression in cardiac and endothelial tissues such as intersegmental vessels, dorsal aorta (Fig. 1c'), the outflow tract (Fig. 1d), blood vessels in the myocardium, the endocardium and the heart valves (Fig. 1d').

153 This demonstrates that *enFLT1* consisting of 358 bases is sufficient to drive GFP expression in
154 endothelial and cardiac tissues *in vivo.*

155

156 Enhancer of *FLT1* is essential for pathways involved in blood vessel morphogenesis

157

158 To understand the effect of the *FLT1* enhancer on the gene regulatory network and its function
159 in humans, we deleted the RE via CRISPR/Cas9 mediated gene editing (Fig. S2a) in human
160 embryonic stem cells (hESC) (background line: H3, NKX2-5($^{eGFP/wt}$) (31)) to generate the
161 Δen*FLT1* line, which was subsequently differentiated into endothelial cells. In order to
162 understand the effects of the enhancer deletion on FLT1 and its related pathways, we also
163 generated a gene mutant cell line, Δex1*FLT1* to act as a positive control. To understand the
164 transcriptional consequences of deleting *enFLT1* (Fig. 2a) we performed RNA sequencing on
165 wildtype, *ΔenFLT1* and *Δex1FLT1* endothelial cells derived from hESC after 12 days of culturing.
166 Deletion of the enhancer had only a minor, not significant reduction on *FLT1* mRNA expression
167 (Fig. S2b) despite evidence from chromatin conformation capture data (32) that enFLT1
168 interacts with the FLT1 promoter in cardiac cells (Fig. S2c). The removal of exon 1 of the gene
169 however resulted in a truncated transcript (Fig. S2d). Nonetheless, when assessing the
170 expression of different *FLT1* isoforms in *ΔenFLT1*, we could not detect a significant decrease in
171 transmembrane bound *FLT1 (tFLT1)* expression (201) nor any difference in expression levels of
172 any of the soluble *FLT1 (sFLT1)* isoforms (204/207), suggesting a more redundant, subtle role
173 of the CRE in *FLT1* regulation. However, deletion of exon 1 of *FLT1* had a significant effect on
174 expression of both, transmembrane and soluble isoforms (Fig. 2b), highlighting *FLT1*
175 dysregulation. Furthermore, we identified 211 differentially expressed genes (DEGs) as a result
176 of deleting the regulatory element compared to cells with the intact wildtype *enFLT1*. The

majority of these DEGs (179) overlapped with the 2261 identified genes which were differentially expressed in the human *FLT1* gene mutant, suggesting an essential role of the enhancer in the FLT1 pathway (Fig. 2c). Furthermore, for the first time, we looked at gene ontology (GO) terms of the DEGs in a human *FLT1* gene mutant and found that the main pathways affected were blood vessel development, cell differentiation, cell division, heart development and extra cellular matrix organisation consistent with the molecular function of FLT1 (25,33). The genes shared by gene mutant and enhancer mutant human cell lines were involved in pathways such as organelle fission and extracellular structure organisation and other regulatory processes of extracellular matrix (ECM) composition. Pathways identified to be unique to differentially expressed genes in *ΔenFLT1* were also found to be involved in angiogenesis and blood vessel morphogenesis, suggesting a role of *enFLT1* in alternative pathways in blood vessel development distinct from the FLT1 pathway (Fig. 2d and S2e).

Deletion of *enFLT1* does not impair angiogenesis *in vitro*

In order to assess the potential of *ΔenFLT1* and *Δex1FLT1* endothelial cells to form tubes *in vitro*, we performed angiogenesis assays (34). Cells were labelled with anti-human CD31 and anti-human CD34 antibodies and imaged over the course of 48 hours (Fig. 3a). While wildtype (*NKX2-5*$^{eGFP/wt}$) and *ΔenFLT1* endothelial cells formed tubes within 10h and then stopped and coalesced, we observed a continuation of angiogenesis and proliferation in *Δex1FLT1* cells until 48h (Fig. 3b). To assess difference in tube formation potential across genotypes, we analysed skeleton length, area, junction count and branch count over the course of 48h (Fig. 3c and S3a). *ΔenFLT1* cells showed no significant difference compared to wildtype cells in any of the categories assessed. However, we did see a significant change in area and skeleton length

when comparing *Δex1FLT1* with the other cell lines (Fig. 3d), indeed suggesting a disruption of the *FLT1* pathway during angiogenesis consistent with published data on *FLT1* knockdown in HUVECS and *Flt1*$^{-/-}$ mutant in endothelial cells derived from murine embryonic stem cells (23). This data suggests the enhancer of *FLT1* is not essential for endothelial tube formation *in vitro* and that other regulatory elements act to buffer *FLT1* expression from the loss of *enFLT1.* Nevertheless, given the altered transcriptional profile observed in *ΔenFLT1* endothelial cells it remains possible that this enhancer is required for normal development.

Deletion of *enFLT1* perturbs cardiovascular development *in vivo*

To determine if loss of the conserved *FLT1* enhancer alters vasculogenesis *in vitro* we deleted the endogenous en*flt1* in medaka and assessed the effect on blood vessel development or morphology. In order to alter the regulatory sequence, which was located in intron 10 of the *tflt1* isoform and the 3'UTR of the *sflt1* isoform in medaka (Fig. 1b), we made use of the CRISPR/Cas9 System to introduce a 270 bp deletion (Fig. 3e). The excision of the endogenous RE fragment in medaka (*enflt1*) was verified by PCR (Fig. S3b) and injected embryos were analysed for defects in heart or blood vessel formation. To rule out an effect of the Cas9 enzyme or the process of microinjection itself on heart and blood vessel formation, controls were injected to target *oca2*, a gene important for pigment formation in the eye and body (35). Interestingly, embryos injected with the sgRNA targeting *enflt1* (Fig. 3f and g) resulted in a significant higher number of blood clots (14.68%) on the yolk in compared to control embryos (5.81%) (p value=0.0215, Mann-Whitney test, one-tailed) suggesting a crucial role of this sequence in angiogenesis and vascularisation *in vivo* in fish.

Discussion

FLT1 is an important regulator of blood vessel development, cell proliferation, migration, differentiation and cell survival (36). Loss of FLT1 or perturbation of *Flt1* regulation has been implicated in blood vessel defects in human, mouse, and zebrafish (23–25). Given that blood vessels are throughout the body, *FLT1* regulation is likely to be complex incorporating both tissue specific cues and endothelial cell type signals. Indeed, the precise combination of transcriptional regulators that bind at the regulatory elements of the *FLT1* locus are not well understood. However, a few studies have highlighted the importance of *FLT1 cis*-regulatory elements and suggested that disruptions in these non-coding regulators affect gene expression of *FLT1* and its isoforms and can ultimately lead to disease (26,28,37,38).

We aimed to functionally test regulatory elements which were linked to cardiovascular development and disease in humans and identified an enhancer in intron 10 of *FLT1*. This enhancer is well conserved across species suggesting it plays a critical role in modulating *FLT1* expression levels. For example, the endogenous zebrafish *enFlt1* enhancer in combination with the endogenous *flt1* promoter is able to drive transgenic reporter gene expression throughout the developing vasculature *in vivo* (27). Here we demonstrate that the human *enFLT1* was also able to drive stable reporter gene expression in combination with a synthetic minimal promoter (SCP1) (39). In human endothelial cells derived from *ΔenFLT1* hESCs levels of *FLT1* were only slightly reduced from wildtype levels. Previous studies have shown that the deletion of *cis*-regulatory elements, despite their capacity to drive gene expression, may only lead to subtle effects on transcription levels or phenotype (40). This occurs due to the fact that genomic regulatory domains often act additively to provide genetic and phenotypic robustness

during the development of an organism (9,40–43). Therefore, we hypothesize that other regulatory elements in the locus functionally compensate for *enFLT1* to maintain stable *FLT1* expression. Nevertheless, the transcriptional profile of endothelial cells is perturbed in the *enFLT1* knockout suggesting the enhancer may provide robustness during vessel morphogenesis. Supporting this idea is that the *in vivo* knockout of the endogenous conserved enhancer in medaka led to an increase of blood vessel disruptions on the yolk resulting in blood clot formation, hinting towards an important role of this sequence in vessel morphogenesis in fish (9,40–43).

Using an *in vitro* angiogenesis model we observed vessel collapse and a reduced rate of retraction of vessels in the human *FLT1* gene mutant cell line that was consistent with previous reports (23), however, this phenotype was not recapitulated in *ΔenFLT1* lines. These *in vitro* experiments might be not be the appropriate assay of enhancer function. An *in vivo* mammalian system could provide more insight into determining if the *FLT1* enhancer is integrated into transcriptional networks during organogenesis rather than angiogenesis. For example, the enhancer may have key regulatory roles during cardiac development given the strategy to identify the enhancer used sequences targeted by the cardiac transcription factor NKX2-5. Interestingly we do see an effect of the deletion of the RE in medaka *in vivo*. In fish, the conserved enhancer lies in the 3'UTR of the *sflt1* isoform, which might lead to premature degradation of the mRNA and, therefore, may have a stronger effect on vessel morphogenesis.

Genes differentially expressed in ΔenFLT1 are involved in blood vessel morphogenesis and cardiovascular development via the regulation of the extracellular matrix composition.

Supporting the notion that the *enFLT1* is a critical regulatory element we also observed 33 DEGs, which did not overlap with DEGs in *Δex1FLT1*, that require an intact enhancer. Among these genes, ANGPT1 is involved in angiogenesis and disruptions in the gene have been associated with angioedema (44,45). As an antagonist of angiogenesis, FLT1 has been predicted to be a functional partner of ANGPT1 (46). While the most straightforward explanation is that sub-optimal *FLT1* expression in ΔenFLT1 endothelial cells impairs these pathways, a more speculative hypothesis is that the enhancer acts in *trans* to regulate this subset of genes. However, in order to fully understand the interaction network of the enhancer with other genes, 3D interaction data of the enhancer is required to identify putative physical interaction partners.

In this study we have defined an evolutionarily conserved enhancer element that permits further dissection of the complex regulatory program controlling of *FLT1* expression. This study provides novel cellular and animal models to further study this enhancer element and provides a striking example of the robustness of the transcriptional network the enables the precise expression of *FLT1*. In conclusion, we show that perturbations in the sequence of a regulatory element of *FLT1* compromises the FLT1/VEGF signalling cascade impairing both transcriptional profiles and blood vessel formation. In this context, this work provides a framework for identifying other *FLT1* regulatory sequences that facilitate the complex interplay of spatial and temporal cues provided to the cells of the vasculature throughout the body.

Methods

Bioinformatic mining and enhancer prediction

Sequences bound by NKX2-5 were retrieved from Anderson et al (14), filtered for ultra-conserved sequences via MultiZ alignments (47) and intersected with histone modification marks datasets the Human Roadmap Epigenome Project (48). The GREAT tool (49) was used to link enhancer candidates to target genes based on proximity rules.

Fish maintenance and ethics

Fish lines were maintained under standard recirculating aquaculture conditions. Day-night cycles were set to 14 h of light and 10h of darkness. The whole fish facility is under the supervision of the local representative of the animal welfare agency and all experiments on Medaka (*Oryzias latipes*) were performed according to European Union animal welfare guidelines and national animal welfare standards in Germany (Tierschutzgesetz §11, Abs. 1, Nr.1, husbandry permit number AZ35-9185.64/BH Wittbrodt, line generation permit number AZ 35-9185.81/G-145-15). The wildtype strain Cab was used in this study.

Enhancer assay and generation of transgenic lines

The reporter line *enFLT1:GFP* was generated by injection of 5ng/µl donor DNA and 10 ng/µl Tol2 transposase mRNA into one cell stage medaka embryos. For transgenesis using the Meganuclease SceI, the injection mix consisted of 0.5x Yamamoto buffer, 0.5x I-SceI buffer, 0.3 U/µl I-SceI enzyme and 10-20n/µl of the donor constructs. The injected plasmid was a modified

zebrafish enhancer detection plasmid (29,30) containing a SCP1 promoter (39). The enhancer element was amplified using these primers: trans_enflt1_for: TTAGGGGGAGGGGAATGTGC; trans_enflt1_rev: CCTCCCTGCCATTGTACTTGG.

Imaging *in vivo*

Medaka hatchlings (sedated with tricaine when alive) were mounted in 1% low melting agarose on glass bottom MatTek dishes. High-resolution imaging was carried out using confocal laser scanning microscopes (Leica TCS SPE or Leica TCS SP8). Analysis and processing were performed using the ImageJ software.

Gene editing *in vivo*

Suitable sgRNAs with low predicted off targets were designed using the CRISPR/Cas9 target online predictor (CCTop) (50). Cas9 mRNA was transcribed from JDS246 by mMessage mMachine Sp6 Transcription Kit (Thermo Fisher) and sgRNA were cloned into DR274 (Addgene #42250) (51). DR274 was linearized using the restriction enzyme DraI and subsequently transcribed using MEGAscript T7 transcription Kit (Thermo Fisher). RNA purification was performed using the RNeasy Mini kit (Qiagen). One-cell medaka embryos were injected with 150 ng/µl Cas9 mRNA and 15 ng/µl per sgRNA used. Δenflt1_sgRNA1: CCAGACCCAACAGTGGACCC; Δenflt1_sgRNA2: GGGCTTGAGAGGTATGTGCT; Δoca2_sgRNA1: TTGCAGGAATCATTCTGTGT; Δoca2_sgRNA2: GATCCAAGTGGAGCAGACTG (35).

Gene editing *in vitro*

sgRNA oligos were cloned into px458 (Addgene #48138) and subsequently transfected via electroporation using 100µl Neon® Tips. Human embryonic stem cells were electroporated at 1050V for 30ms with 2 pulses. Cells transiently expressing GFP were single cell sorted, colonies were grown for 2 weeks and screened via PCR. ΔenFLT1_sgRNA1: TAAGGGCACAAGCCCTAGTA; ΔenFLT1_sgRNA2: ACCTGAAACAACTTAATTT; Δex1FLT1_sgRNA1: TAGTTGCAGCGGGCACGCTT; Δex1FLT1_sgRNA2: TTATAAATCGCCCCCGCCCT.

Cell culture and endothelial cell differentiation

HESCs (background line: H3, NKX2-5(^{eGFP}/^{wt})) were cultured on feeders and passaged as previously described (31). Endothelial cell differentiation was induced by using an adapted cardiomyocyte protocol (14). 2.2 x 10^6 cells were seeded per well on 6-well plates coated with Geltrex (Life Technologies). On day 0, the standard cardiac differentiation medium contained 12µM CHIR99021, 80ng/ml Activin A and 50µg/ml Ascorbic Acid. For day 3 and day 5, the standard media was supplemented with 5µM IWR-1, 50µg/ml Ascorbic Acid, 30ng/ml VEGF (PeproTech) and 50ng/ml SCF (PeproTech). At day 6 the cells were harvested and processed for flow cytometry sorting. The suspension was stained for the endothelial specific cell surface markers CD31 and CD34 to sort for endothelial progenitors. The antibodies anti-human CD31-APC (BioLegend) and anti-human CD34-PE/Cy7 (BioLegend) were used in a 1:400 ratio. Endothelial cells were maintained in endothelial growth medium, EGM2-2MV Bulletkit (Lonza).

Angiogenesis assay

Vessel formation assays or angiogenesis assays were performed as previously published (34). In brief, 40µl of Geltrex™ were added to each well of a 96 well glass bottom plates (Corning). The plates were kept on ice while pipetting, subsequently centrifuged and left to set in the incubator at 37˚C for 30min. Endothelial cells were harvested and resuspended in complete growth factor EGM-2V media supplemented with anti-human CD31-APC and anti-human CD34-PE/Cy7 antibodies (1:400). 15,000 cells were plated per well.

Imaging and Image analysis

Confocal images were acquired with Yokagawa CellVoyager CV8000 high-throughput discovery system under 37˚C and 5% CO_2. Maximum intensity projection (MIP) images were constructed from 15µm z slices (300µm total z distance), captured every 40 minutes for 48 hours. Images were analysed in CellPathfinder software based on CD34-APC intensity. Analysis algorithm was defined using skeleton function to generate the parameters vessel area, vessel length, branch count, and junction count. The Kruskal-Wallis test was used to evaluate differences in medians among three cell lines. If the Kruskal-Wallis test was significant, the Wilcoxon Rank Sum Test was used to assess the level of difference significance on a variable between two cell lines. The statistical analysis was performed in the R statistical programming language.

RNA sequencing and analysis

Three biological replicates with each three technical replicates (per cell line) of endothelial cells were harvested 12 days after differentiation initiation. RNA was extracted using the Direct-Zol RNA Miniprep Kit (Zymo research). Paired-end mRNA sequencing was performed at the

Victorian Clinical Genetics Services (VCGS) using the NovaSeq 6000 System (Illumina) with a 2 x 150 bp read length. The fastq files were processed using the RNAsik pipeline (52). The STAR aligner (53) was used to align reads to the GRCh38 Assembly. Aligned reads were assigned to features from the GRCh38 EnsEMBL Annotation (54) using the featureCounts program from RsubRead (55). Degust (56) was used to perform and visualise differential expression analysis. Firstly, the first dimension of unwanted variation were removed from counts using RUVr routine from the RUVSeq R package (57). Next, TMM normalisation and the quasi-likelihood test was performed using EdgeR's (58) standard workflow.

Salmon (59) was used to quantify transcript isoforms abundance, and the differential abundance of FLT1 was tested and visualised using the DRIMSeq R Package (60).

Declarations

Ethics approval and consent to participate. No ethics approval and consent required for this study.

Consent for publication. All authors provide consent for publication.

Competing interests. The authors declare no competing financial interests.

Funding. This work was supported by a 0 (1180905), the Royal Children's Hospital Foundation as well as the Stafford Fox Foundation. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

Authors' contributions.

MR designed the study with input from DE and ME. JS performed *in vivo* and *in vitro* experiments with input from TL, LC, JW, EP, DE and MR. JS and HV performed imaging and data analysis with input from AH. MR, ME, MD, MS, JS, NRM and BS performed bioinformatics and

414 statistical analysis. JS, MR and DE wrote the manuscript with input from all authors. All authors
415 reviewed and approved the manuscript.

416

417 **Availability of data and materials**. The RNA-seq dataset is accessible through NCBI Gene
418 Expression Omnibus, GEO160873.

419

420 **Acknowledgments.** We thank Jeannette Hallab, Karen Gross, Kathy Karavendzas, Francesca
421 Bolk, Markus Tondl, Ling Qian and Sebastian-Alexander Stamatis for support in laboratory work
422 and advice. Choon Boon (Evangelyn) Sim, Ali Seleit, James McNamara and Christine Wells for
423 scientific discussions and guidance. We are grateful to Jose Arturo Gutierrez-Triana for
424 providing the transgenesis vector. We would also like to thank the Victorian Clinical Genetics
425 Services (VCGS) for providing the RNA-seq service and the MCRI FACS facility for technical
426 assistance. Furthermore, we would like to thank all members of the Centanin, Elliott, Porrello,
427 Wittbrodt and Ramialison laboratory members for active discussions and feedback. The
428 Australian Regenerative Medicine Institute is supported by grants from the State Government
429 of Victoria and the Australian Government.

430

431 ## References

432

433 1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart
434 disease and stroke statistics—2020 update: A report from the American Heart
435 Association. Vol. 141, Circulation. Lippincott Williams and Wilkins; 2020. p. E139–596.
436 2. Edwards JJ, Gelb BD. Genetics of congenital heart disease. Curr Opin Cardiol. 2016

437 May;31(3):235–41.

438 3. Agopian AJ, Goldmuntz E, Hakonarson H, Sewda A, Taylor D, Mitchell LE. Genome-Wide
439 Association Studies and Meta-Analyses for Congenital Heart Defects. Circ Cardiovasc
440 Genet. 2017 Jun 1;10(3):e001449.

441 4. Wang Y, Wang J-G. Genome-Wide Association Studies of Hypertension and Several
442 Other Cardiovascular Diseases. Pulse. 2018;6(3–4):169–86.

443 5. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease.
444 Genome Med. 2014 Oct 28;6(10):85.

445 6. Chahal G, Tyagi S, Ramialison M. Navigating the non-coding genome in heart
446 development and Congenital Heart Disease. Vol. 107, Differentiation. Elsevier Ltd; 2019.
447 p. 11–23.

448 7. Smith E, Shilatifard A. Enhancer biology and enhanceropathies. Nat Struct Mol Biol. 2014
449 Mar 1;21(3):210–9.

450 8. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, et al. Large-scale
451 discovery of enhancers from human heart tissue. Nat Genet. 2012 Jan 4;44(1):89–93.

452 9. van den Boogaard M, van Weerd JH, Bawazeer AC, Hooijkaas IB, van de Werken HJG,
453 Tessadori F, et al. Identification and Characterization of a Transcribed Distal Enhancer
454 Involved in Cardiac Kcnh2 Regulation. Cell Rep. 2019 Sep 3;28(10):2704-2714.e5.

455 10. Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, et al.
456 Genome-wide compendium and functional assessment of in vivo heart enhancers. Nat
457 Commun. 2016 Oct 5;7:12923.

458 11. van Ouwerkerk AF, Bosada FM, van Duijvenboden K, Hill MC, Montefiori LE, Scholman
459 KT, et al. Identification of atrial fibrillation associated genes and functional non-coding
460 variants. Nat Commun. 2019 Dec 1;10(1).

12. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.

13. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317–29.

14. Anderson DJ, Kaplan DI, Bell KM, Koutsis K, Haynes JM, Mills RJ, et al. NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. Nat Commun. 2018;9(1):1–13.

15. Arvanitis M, Tampakakis E, Zhang Y, Wang W, Auton A, Agee M, et al. Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. Nat Commun. 2020 Dec 1;11(1):1122.

16. Montefiori LE, Sobreira DR, Sakabe NJ, Aneas I, Joslin AC, Hansen GT, et al. A promoter interaction map for cardiovascular disease genetics. Elife. 2018;7:1–35.

17. Bertero A, Fields PA, Ramani V, Bonora G, Yardimci GG, Reinecke H, et al. Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat Commun. 2019 Dec 1;10(1):1–19.

18. ENCODE Encyclopedia, Version 4: Genomic annotations – ENCODE [Internet]. [cited 2020 Oct 30]. Available from: https://www.encodeproject.org/data/annotations/

19. Catarino RR, Stark A. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. 2018;32(3–4):202–23.

20. Bouveret R, Waardenberg AJ, Schonrock N, Ramialison M, Doan T, de jong D, et al. NKX2-5 mutations causative for congenital heart disease retain functionality and are directed to hundreds of targets. Elife. 2015 Jul 6;4(JULY2015).

485 21. Lyons GE. Vertebrate heart development. Curr Opin Genet Dev. 1996 Aug 1;6(4):454–
486 60.
487 22. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, et al. Congenital
488 heart disease caused by mutations in the transcription factor NKX2-5. Science (80-).
489 1998 Jul 3;281(5373):108–11.
490 23. Nesmith JE, Chappell JC, Cluceru JG, Bautch VL. Blood vessel anastomosis is spatially
491 regulated by Flt1 during angiogenesis. Development. 2017;144(5).
492 24. Krueger J, Liu D, Scholz K, Zimmer A, Shi Y, Klein C, et al. Flt1 acts as a negative regulator
493 of tip cell formation and branching morphogenesis in the zebrafish embryo.
494 Development. 2011 May;138(10):2111–20.
495 25. Fong GH, Rossant J, Gertsenstein M, Breitman ML. Role of the Flt-1 receptor tyrosine
496 kinase in regulating the assembly of vascular endothelium. Nature. 1995 Jul
497 6;376(6535):66–70.
498 26. Kikas T, Inno R, Ratnik K, Rull K, Laan M. C-allele of rs4769613 Near FLT1 Represents a
499 High-Confidence Placental Risk Factor for Preeclampsia. Hypertens (Dallas, Tex 1979).
500 2020 Sep 1;76(3):884–91.
501 27. Bussmann J, Bos FL, Urasaki A, Kawakami K, Duckers HJ, Schulte-Merker S. Arteries
502 provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk.
503 Development. 2010 Aug 15;137(16):2653–7.
504 28. Menendez D, Krysiak O, Inga A, Krysiak B, Resnick MA, Schönfelder G. A SNP in the flt-1
505 promoter integrates the VEGF system into the p53 transcriptional network. Proc Natl
506 Acad Sci U S A. 2006 Jan 31;103(5):1406–11.
507 29. Bessa J, Tena JJ, De La Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A,
508 et al. Zebrafish Enhancer Detection (ZED) vector: A new tool to facilitate transgenesis

and the functional analysis of cis-regulatory regions in zebrafish. Dev Dyn. 2009;238(9):2409–17.

30. Gutierrez-Triana JA, Herget U, Castillo-Ramirez LA, Lutz M, Yeh C-M, De Marco RJ, et al. Manipulation of Interrenal Cell Function in Developing Zebrafish Using Genetically Targeted Ablation and an Optogenetic Tool. Endocrinology. 2015 Sep 1;156(9):3394–401.

31. Elliott DA, Braam SR, Koutsis K, Ng ES, Jenny R, Lagerqvist EL, et al. NKX2-5 eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes. Nat Methods. 2011 Dec;8(12):1037–43.

32. Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet. 2019 Sep 1;51(9):1380–8.

33. Chen TT, Luque A, Lee S, Anderson SM, Segura T, Iruela-Arispe ML. Anchorage of VEGF to the extracellular matrix conveys differential signaling responses to endothelial cells. J Cell Biol. 2010 Feb 22;188(4):595–609.

34. Arnaoutova I, Kleinman HK. In vitro angiogenesis: Endothelial cell tube formation on gelled basement membrane extract. Nat Protoc. 2010 Apr;5(4):628–35.

35. Lischik CQ, Adelmann L, Wittbrodt J. Enhanced in vivo-imaging in medaka by optimized anaesthesia, fluorescent protein selection and removal of pigmentation. Winkler C, editor. PLoS One. 2019 Mar 7;14(3):e0212956.

36. Simons M, Gordon E, Claesson-Welsh L. Mechanisms and regulation of endothelial VEGF receptor signalling. Vol. 17, Nature Reviews Molecular Cell Biology. Nature Publishing Group; 2016. p. 611–25.

37. Thomas CP, Raikwar NS, Kelley EA, Liu KZ. Alternate processing of Flt1 transcripts is

533 directed by conserved cis-elements within an intronic region of FLT1 that reciprocally
534 regulates splicing and polyadenylation. Nucleic Acids Res. 2010 Apr 10;38(15):5130–40.
535 38. Owen LA, Morrison MA, Ahn J, Woo SJ, Sato H, Robinson R, et al. FLT1 genetic variation
536 predisposes to neovascular AMD in ethnically diverse populations and alters systemic
537 FLT1 expression. Investig Ophthalmol Vis Sci. 2014 May 8;55(6):3543–54.
538 39. Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that
539 enhances gene expression. Nat Methods. 2006 Nov 23;3(11):917–22.
540 40. Cunningham TJ, Lancman JJ, Berenguer M, Dong PDS, Duester G. Genomic Knockout of
541 Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb
542 Development. Cell Rep. 2018;23(11):3146–51.
543 41. Osterwalder M, Barozzi I, Tissiéres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, et al.
544 Enhancer redundancy provides phenotypic robustness in mammalian development.
545 Nature. 2018;554(7691):239–43.
546 42. Letelier J, De La Calle-Mustienes E, Pieretti J, Naranjo S, Maeso I, Nakamura T, et al. A
547 conserved Shh cis-regulatory module highlights a common developmental origin of
548 unpaired and paired fins. Nat Genet. 2018;50(4):504–9.
549 43. Sarro R, Kocher AA, Emera D, Uebbing S, Dutrow E V., Weatherbee SD, et al. Disrupting
550 the three-dimensional regulatory topology of the Pitx1 locus results in overtly normal
551 development. Dev. 2018 Apr 1;145(7).
552 44. Bafunno V, Firinu D, D'Apolito M, Cordisco G, Loffredo S, Leccese A, et al. Mutation of
553 the angiopoietin-1 gene (ANGPT1) associates with a new type of hereditary
554 angioedema. J Allergy Clin Immunol. 2018 Mar 1;141(3):1009–17.
555 45. d'Apolito M, Santacroce R, Colia AL, Cordisco G, Maffione AB, Margaglione M.
556 Angiopoietin-1 haploinsufficiency affects the endothelial barrier and causes hereditary

557 angioedema. Clin Exp Allergy. 2019 May 1;49(5):626–35.

558 46. FLT1 protein (Sus scrofa) - STRING interaction network [Internet]. [cited 2020 Oct 31].
559 Available from: https://string-db.org/network/9823.ENSSSCP00000009946

560 47. Multiz Alignments Multiz Align Track Settings [Internet]. [cited 2020 Oct 31]. Available
561 from: http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=multiz100way

562 48. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative
563 analysis of 111 reference human epigenomes. Nature. 2015 Feb 18;518(7539):317–30.

564 49. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves
565 functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May
566 2;28(5):495–501.

567 50. Stemmer M, Thumberger T, del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: An Intuitive,
568 Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. Maas S, editor. PLoS One. 2015
569 Apr 24;10(4):e0124633.

570 51. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing
571 in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 2013 Mar 29;31(3):227–9.

572 52. MonashBioinformaticsPlatform/RNAsik-pipe: JOSS ready [Internet]. Zenodo; 2018.
573 Available from: https://zenodo.org/record/1403976#.X5kS3IgzaUl

574 53. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
575 universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.

576 54. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020.
577 Nucleic Acids Res. 2019 Oct 28;gkz966.

578 55. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better
579 for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019 May
580 1;47(8).

581 56. drpowell/degust 4.1.1 [Internet]. Zenodo; 2019. Available from:
582 https://zenodo.org/record/3501067#.X5kV8ogzaUk
583 57. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis
584 of control genes or samples. Nat Biotechnol. 2014 Oct 28;32(9):896–902.
585 58. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential
586 expression analysis of digital gene expression data. Bioinformatics. 2009 Nov
587 11;26(1):139–40.
588 59. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware
589 quantification of transcript expression using dual-phase inference. Nat Methods. 2017
590 Oct 28;14(4):417–9.
591 60. Nowicka M, Robinson MD. DRIMSeq: a Dirichlet-multinomial framework for multivariate
592 count outcomes in genomics. F1000Research. 2016 Oct 28;5.
593

Figures

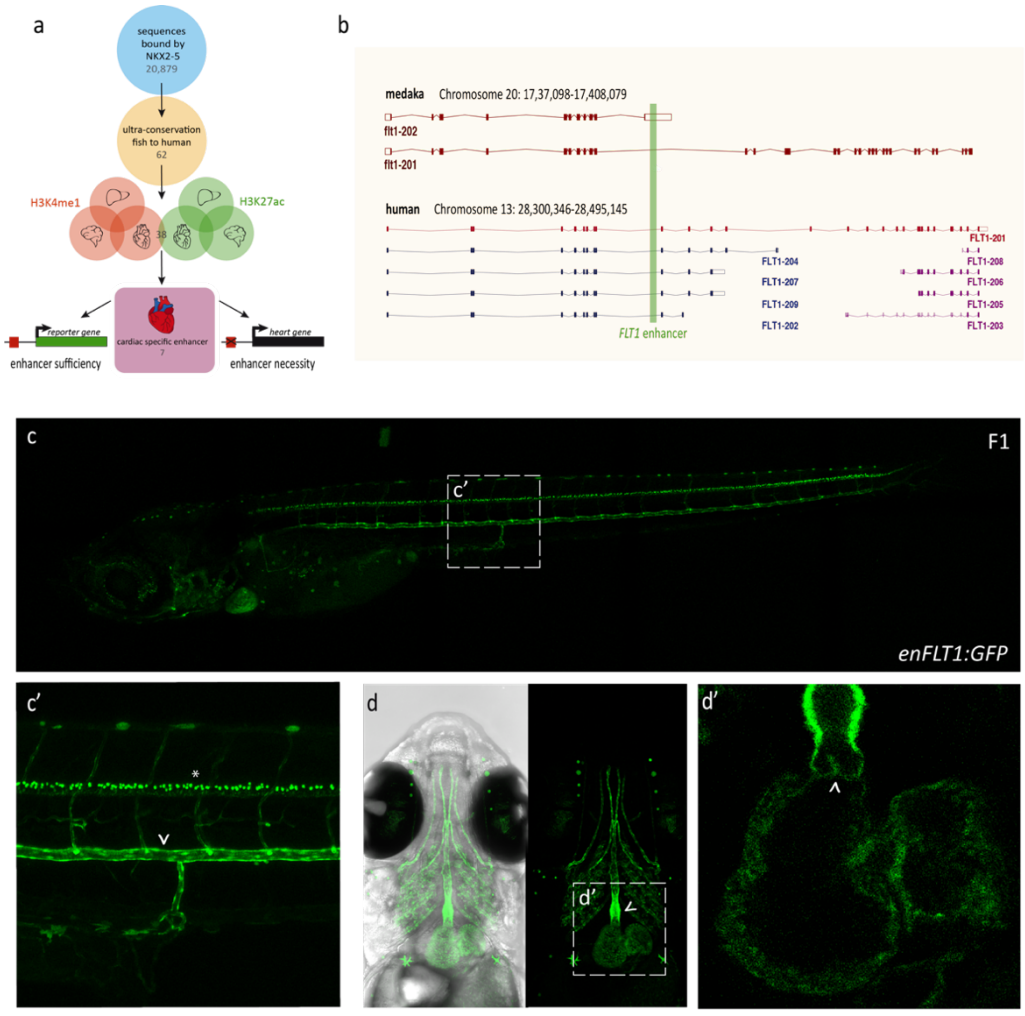


Figure 1: Identification and validation of cardiac specific *cis*-regulatory elements. a) Bioinformatic pipeline to identify cardiac regulatory elements (RE) based on NKX2-5 binding, sequence conservation and active histone marks (H3K4me1 and H3K27ac). Putative RE were tested for the ability to drive reporter gene expression and the necessity for endogenous *FLT1* expression. b) Overview of the *FLT1* locus and its isoforms in humans and medaka. The RE used for *in vivo* transgenesis is displayed in green. c) *enFLT1* drives stable reporter gene (green fluorescent protein (GFP)) expression in neurons (asterisk) and endothelium such as c') the dorsal aorta (arrowhead) d) the outflow tract (arrowhead) d') and valves (arrowhead) in the heart.

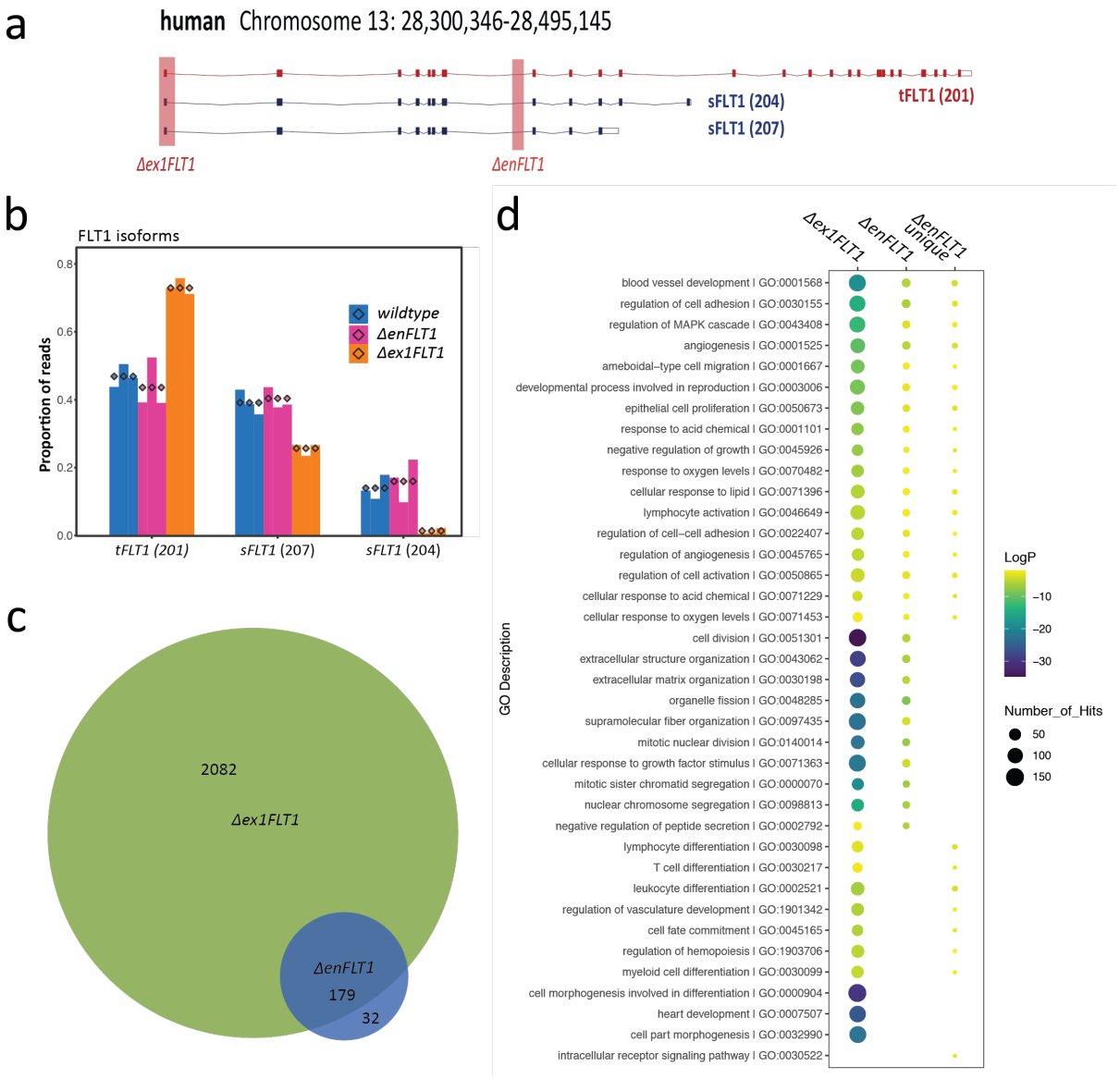


Figure. 2: *enFLT1* is implemented in FLT1 related and alternative pathways. a) Overview of the transmembrane bound FLT1 (tFLT1) and soluble FLT1 (sFLT1) isoforms. We generated a gene mutant (*Δex1FLT1*) and enhancer mutant (*ΔenFLT1*) cell line in human embryonic stem cells. b) Differentially expressed isoforms in *wildtype*, *ΔenFLT1* and *Δex1FLT1* cells. c) Venn diagram of differentially expressed genes (DEGs) (compared to wildtype) in *ΔenFLT1* and *Δex1FLT1*. The majority of DEGs in *ΔenFLT1* overlaps with genes in *Δex1FLT1*. d) Gene ontology of DEGs in Δex1FLT1, ΔenFLT1 and the 32 genes unique to enhancer mutant.

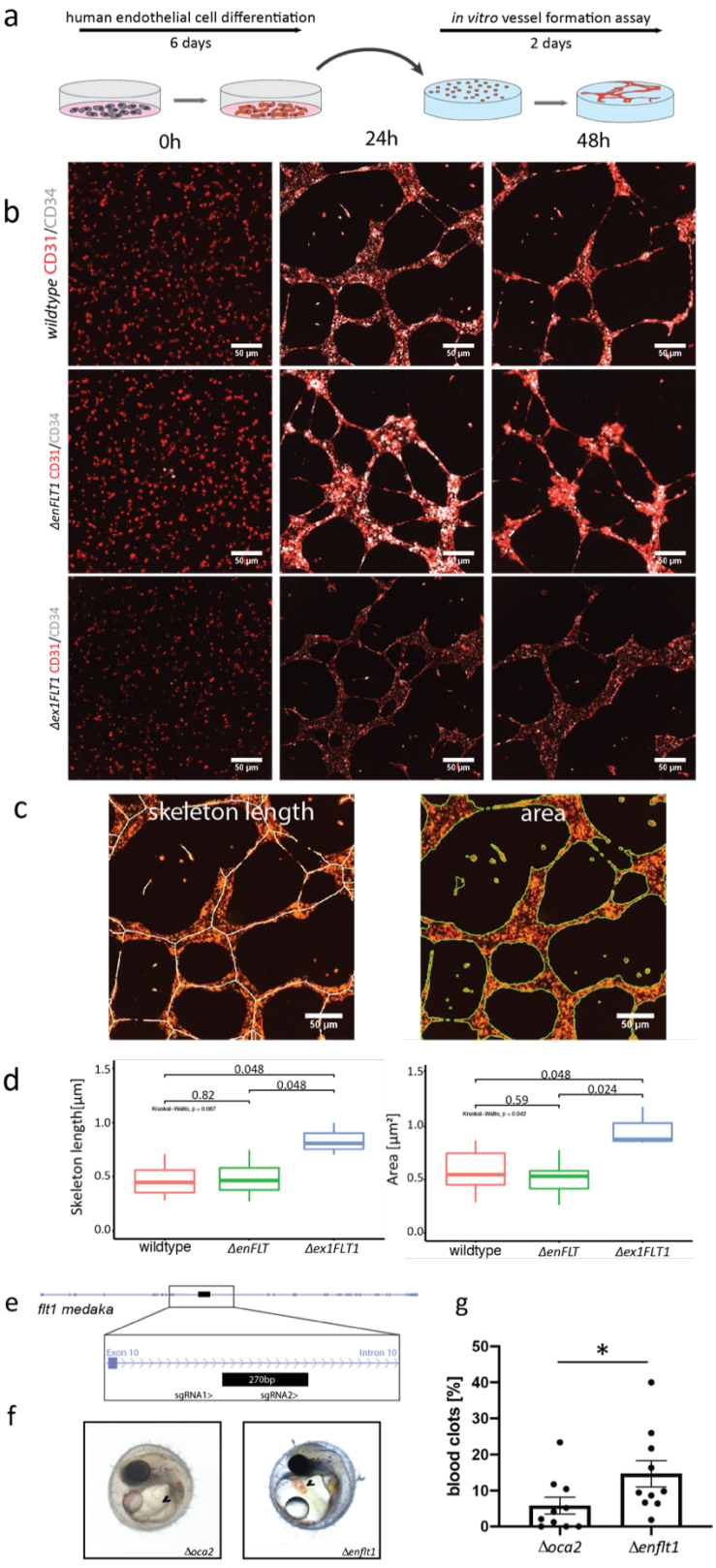


Figure 3: Functional analysis of *enFLT1* during angiogenesis. a) human embryonic stem cells were differentiated within 6 days and subsequently used for angiogenesis assays and imaged

for 48 hours. b) Angiogenesis assays of wildtype, *ΔenFLT1* and *Δex1FLT1* cell lines at 0h, 24h, and 48h. All cell lines stayed positive for CD31 (red) and CD34 (grey) for 48h. *Δex1FLT1* vessels maintained stronger connections and collapsed over time forming thicker vessels. c) CellPathfinder analysis of vessels based on skeleton length and area. d) Statistical analysis of normalised skeleton length and area covered over the course of the experiment. *Δex1FLT1* cells displayed longer connections and covered a significant larger area compared to wildtype and *ΔenFLT1* cells. e) *flt1* locus and endogenous *enflt1* in medaka. 270bp were targeted with 2 sgRNAs. f) Deletion of *enflt1* increased occurrence of blot clots (*Δenflt1*) compared to control injections targeting *oca2* (*Δoca2*) g) statistical analysis of blood clot occurrence. A significant difference (p= 0.0215) of blood clot numbers were observed upon deletion of *enflt1* (14.68%) compared to controls (5.81%).

Supplementary data

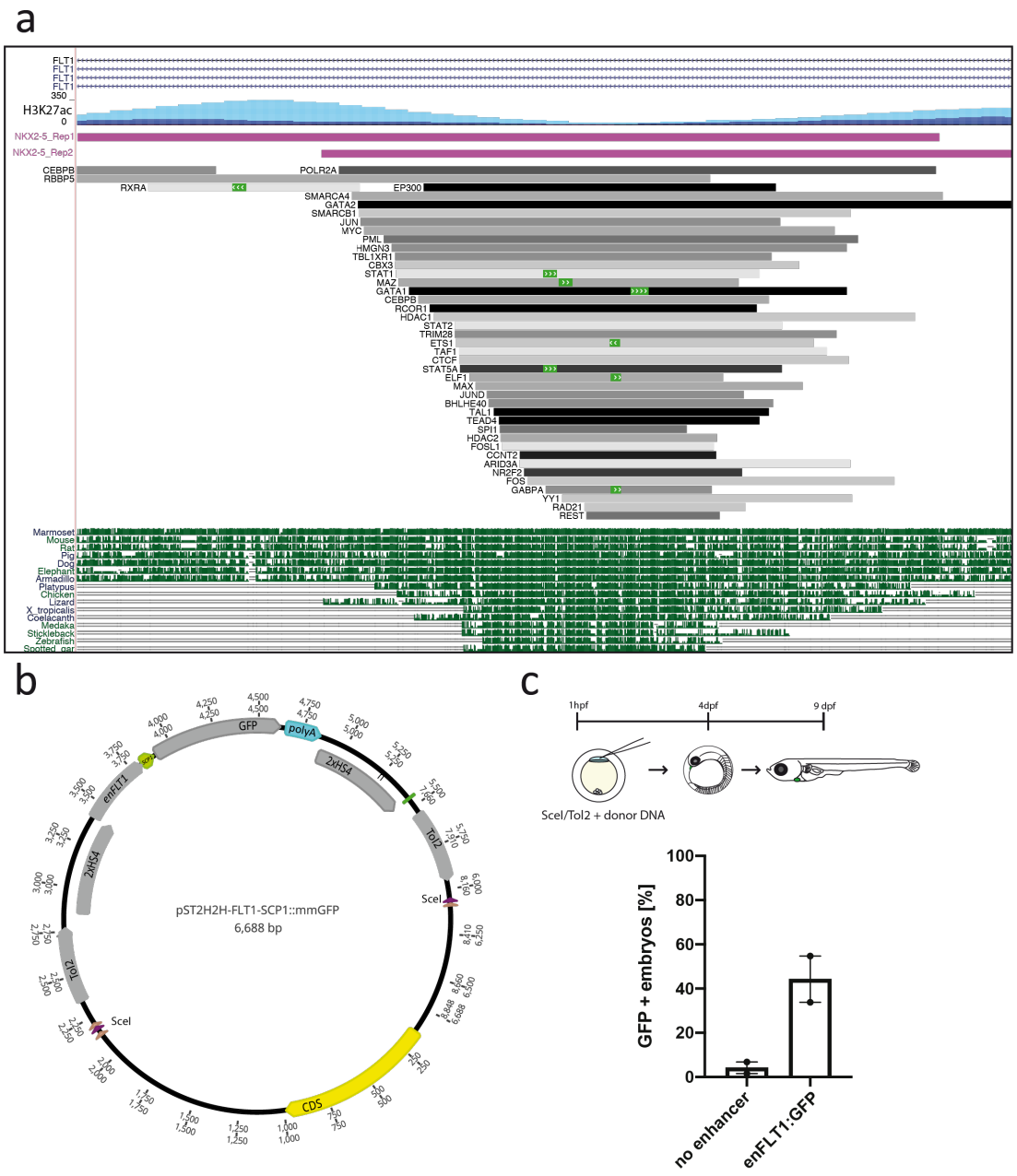


Figure S1: In vivo validation of *enFLT1*. a) Scheme of Transcription factor binding to *enFLT1*. Transcription factor binding overlaps with highly a conserved sequence across species and a valley between two H3K27ac activity marks suggesting open chromatin. b) Transgenesis assay plasmid containing *enFLT1*. Control vector had no RE sequence inserted. c) Plasmids were in injected in one cell stage medaka embryos. Occurrence of GFP expression was monitored until

636 hatching. Embryos injected with the plasmid containing *enFLT1* showed higher number of GFP

637 positive embryos compared to injection with a control vector without a putative RE.

638

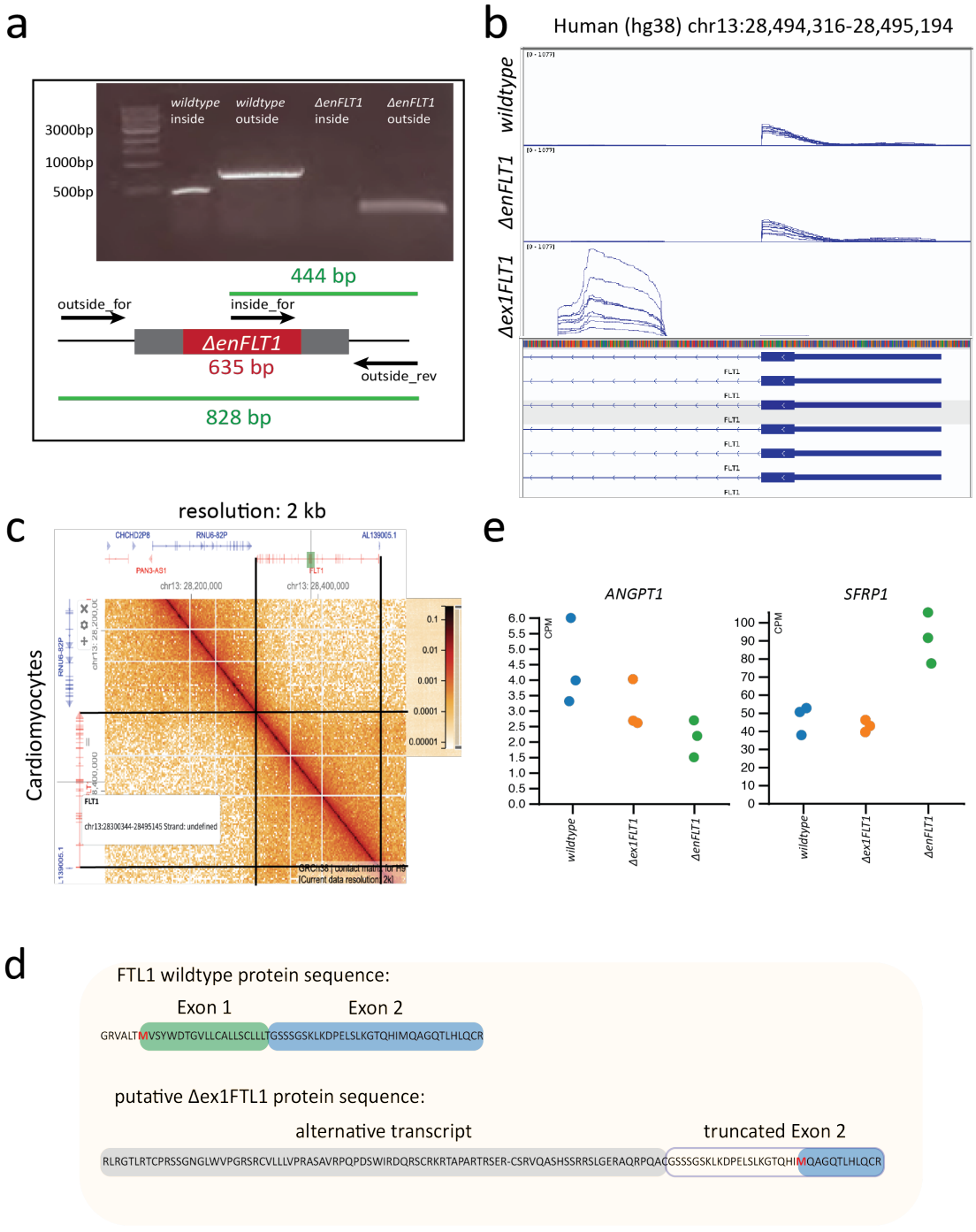
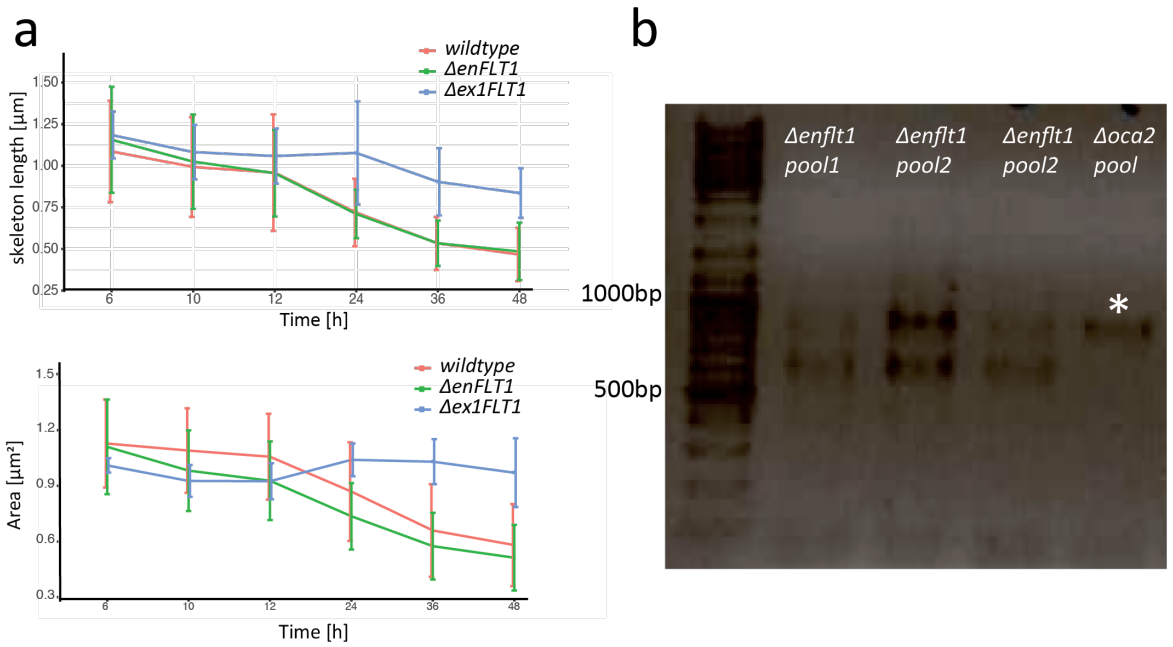


Figure S2: Generation of *ΔenFLT1* and *Δex1FLT1* human embryonic stem cells.

a) Screening PCR targeting the deleted RE in hESCs. 3 Primer PCR approach was used to amplify gene edited locus. Band in *ΔenFLT1* outside, was cut out and sent for sequencing. b) overview of sequencing reads of exon 1 of *FLT1* in wildtype, *ΔenFLT1* and *Δex1FLT1* endothelial cells. Expression of exon 1 has been disrupted in *Δex1FLT1* and transcription starts at an alternative start site. c) HiC interaction map of targets interacting with *enFLT1* in cardiomyocytes at 2kb

646 resolution. d) Predicted protein sequence of *FLT1* in wildtype and *Δex1FLT1*. Alternative Start
647 codon in Exon 2 might lead to a truncated protein in *Δex1FLT1*. e) ANGPT1 and SFRP1 were
648 identified from a list of 33 DEGs unique in *ΔenFLT1* involved in blood vessel morphogenesis.
649



650

651 **Figure S3: *in vitro* and *in vivo* analysis of *enFLT1* deletion.** a) Assessment of normalised skeleton

652 length and area of vessels in angiogenesis assays. Length and area are significantly different in

653 *Δex1FLT1* compared to cell lines at 48h. b) Screening PCR of gene edited embryos injected with

654 CRISPR/Cas9 targeting the *enflt1* or *oca2*. Asterisk indicate the wildtype band at 879bp.

655 Excision of enhancer led to a band with a size of 634 bp. Samples contained pooled gDNA from

656 several injected embryos.

657

658 Table S1: List of putative regulatory elements identified via bioinformatic mining.

RE #	Coordinates (chromosome #; start; end) human hg19 assembly			Location	Putative genes regulated by the RE (location of the RE with respect to the transcription start of the gene (in base pairs))
1	3	71573754	71574438	intronic	FOXP1 (-394,108)
2	5	88129361	88130230	intronic	MEF2C (+70,103)
3	8	106341006	106341556	intronic	FOG2 (+10,361)
4	13	28982241	28983259	intronic	FLT1 (+86,482)
5	15	96808076	96808846	intronic	COUP-TFII (-65,485)
6	16	54540547	54541137	intergenic	IRX5 (-423,932), IRX3 (-220,167)
7	21	38807473	38807875	intronic	DYRK1A (+15,072)

659