Chapter title: Finding and verifying enhancers for endothelial-expressed genes

Alice Neal[1,2]*, Helena Rodriguez-Caro[1] and Sarah De Val[1,2]*

*co-corresponding authors

[1] Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3PT, United Kingdom.
[2] Ludwig Institute for Cancer Research Ltd, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7DQ, United Kingdom.

*Corresponding authors:
Sarah De Val
Department of Physiology, Anatomy & Genetics
University of Oxford
Ludwig Institute for Cancer Research
Old Road Campus Research Building
Roosevelt Drive
Headington
Oxford OX3 7DQ
01865 617435
sarah.deval@dpag.ox.ac.uk

Alice Neal
Department of Physiology, Anatomy & Genetics
University of Oxford
Ludwig Institute for Cancer Research
Old Road Campus Research Building
Roosevelt Drive
Headington
Oxford OX3 7DQ
01865 617435
Alice.neal@dpag.ox.ac.uk

Chapter title: Finding and verifying enhancers for endothelial-expressed genes

Summary/Abstract: Identification and analysis of enhancers for endothelial-expressed genes can provide crucial information regarding their upstream transcriptional regulators. However, enhancer identification can be challenging, particularly for people with limited access or experience of bioinformatics, and transgenic analysis of enhancer activity patterns can be prohibitively expensive. Here we describe how to use publicly available datasets displayed on the UCSC Genome Browser to identify putative endothelial enhancers for mammalian genes. Further, we detail how to utilize mosaic Tol2-mediated transgenesis in zebrafish to verify whether a putative enhancer is capable of directing endothelial-specific patterns of gene expression.

1. Introduction

Eukaryotic genes typically contain two types of transcriptional regulatory DNA elements: core promoters proximal to the transcriptional start site, and more distal regulatory elements including enhancers, insulators and silencers [1]. Regulation of complex spatio-temporal patterns of gene transcription during mammalian development is often accomplished by enhancers, although those located close to core promoters are sometimes referred to as proximal promoters. A single gene can be regulated by multiple different enhancers, each responsible for a particular aspect of expression [2].

Enhancers are densely clustered aggregations of transcription factor binding motifs able to influence transcription independent of location, distance or orientation [3]. Identification and analysis of enhancers can prove a direct link between a transcription factor which binds the enhancer and expression of a particular gene. Additionally, detailed analysis of the sequences within a specific enhancer can help identify a cohort of transcription factors required for particular patterns of gene expression (e.g. analysis of an arterial-specific enhancer can reveal factors which collectively regulate arterial differentiation). Analysis of enhancers containing disease associated SNPs (which are most frequently located in non-coding DNA) can also provide crucial insight into both normal development and disease-causal pathways.

Enhancers are typically found relatively close to their target gene promoter, with a recent genome wide analysis of over 600 enhancer-core promoter pairs finding an average distance of 24 kb between enhancer and promoter [4]. However, enhancers can be located upstream, downstream, within introns or even within exons of both target or neighbouring genes. Additionally, although apparently less common, enhancers can influence reporter gene expression from much greater distances, and there are many examples of enhancers located around 1MB from their target gene (e.g. the Shh limb enhancer [5] and the Coup-TFII endothelial enhancer [6]). Consequently, identifying the enhancer(s) regulating your gene of interest in endothelial cells can be a challenging task. However, some hallmarks of enhancers can assist in this endeavour. Active enhancers are usually in more open regions of the chromatin, which can be easily identified by hypersensitivity to DNase digest (DNase-seq or DNase HS), by identifying regions that fail to cross-link to nucleosomes (FAIRE-seq) or by identifying regions with increased accessibility to transposases (ATAC-seq) [7]. Enhancers are also associated with increased levels of the specific histone modifications H3K4Me1 and H3K27Ac, and an absence of the histone modification H3K4Me3 (which instead is associated

with promoters). Additionally, enhancers are often bound by EP300 and cell-type specific transcription factors. For example, nearly all known endothelial enhancers are bound by members of the ETS transcription factor family [3,8]. Direct enhancer-promoter binding can also be detected by analysis of 3D proximity (e.g. chromatin conformation capture techniques) [9], and enhancer-transcribed RNA (eRNAs) can be detected through RNA-seq [10]. However, none of these characteristics either together or alone appear to be able to reliably identify functional enhancers every time, nor can they indicate the expression pattern regulated by the enhancer. Consequently, identification of putative enhancers ideally needs to be combined with some level of *in vivo* validation to determine both the spatial (e.g. does the enhancer drive ubiquitous or tissue-specific expression) and temporal (e.g. is the enhancer active during a particular developmental window?) characteristics of each putative enhancer.

In this protocol paper, we describe simple ways to identify putative endothelial gene enhancers for those with limited knowledge or access to bioinformatic assistance, using publicly available data displayed on the UCSC Genome Browser. Further, we detail how to verify that these enhancers are able to direct endothelial gene expression *in vivo* using Tol2-mediated zebrafish transgenesis, which combines cost effectiveness with speed and relative ease of access.

2. Materials

Non-Standard Equipment:

Programmable Microinjector and micromanipulation system mounted on an inverted microscope (e.g Eppendorff Femtojet^{R} 4X and Adapter Leica$^{®}$ 1)

Capillary glass 1.0 OD x 0.58ID X 150 Lmm (Harvard apparatus 30-0020/Clark Electromedical Instruments GC100F-15)

A micropipette puller (e.g Flaming Brown P-1000 micropipette puller)

96 well plate for confocal imaging (e.g. Thermo Scientific Nunc MicroWell 96-well Optical-Bottom Plates)

Confocal microscope

Access to Sanger sequencing (e.g Eurofinsgenomics custom DNA sequencing services)

2 Fine tipped forceps (e.g Jewelers forceps, Dumont No.5)

Materials

E1b-GFP-Tol2 plasmid (Addgene plasmid #37845)

Expression plasmid for production of Tol2 transposase (e.g. pCS2FA CO Tol2 TPase Addgene #133032)

pCR^{TM}8/GW/TOPO/TA Cloning Kit with One Shot^{TM} TOP10 E.Coli (ThermoFisher Scientific cat# K250020)

Gateway^{TM} LR Clonase^{TM} II Enzyme mix (ThermoFisher Scientific cat# 11791020)

Mini prep kit (e.g. QIAprep Spin Miniprep kit Cat. No. 27104)

Midi prep kit (e.g. QIAGEN plasmid plus midi kit Cat. No.: 12941)

Various restriction enzymes (New England BioLabs)

Antibiotics: Spectinomycin and Ampicillin

mMessage mMachine^{TM} SP6 Transciption Kit (ThermoFisher Scientific Cat. No. AM1344)

RNA purification columns (e.g. Monarch RNA purification Columns NEB T2010)

RNAse free H_2o

Non-Standard Solutions

60X Embryo Medium (2L of 60x solution)

NaCl – 34.8g (5mM)
KCl – 1.6g (0.17mM)
5.8 g $CaCl_2{\cdot}2H_2O$ (0.33mM)
9.78 g $MgCl_2{\cdot}6H_2O$ (0.33mM
Dissolve in H_2O to a final volume of 2L
Adjust pH to 7.2 with NaOH. Autoclave and store at room temperature.

1X Embryo Medium (1L)

Dilute 16.5ml 60X Embryo medium to 1L
To prevent fungal growth add 100µl of 1% methylene blue

Tricaine Stock Solution

400mg Ethyl 3-aminobenzoate methanesulfonate (MS-222/Tricaine) (Sigma-Aldrich)
98 ml ddH_2O
2 ml 1 M Tris (pH 9)
Adjust pH to ~7.
Store at -20

Tricaine Working solution

5ml tricaine stock solution
45ml embryo media

3. Methods

3.1 Identifying putative endothelial enhancers *in silico* using publicly available datasets displayed on UCSC Genome Browser

3.1.1 Accessing data readily available on UCSC Genome Browser to identify human endothelial enhancers

While there are a myriad of ways of acquiring and interrogating information about the enhancer-associated characteristics for your gene loci of interest, one of the most accessible ways is to use standard tracks and data from the ENCODE project [11] available on the UCSC Genome Browser. Currently, access to required data tracks is best using the Human Feb. 2009 (GRCh37/hg19) assembly. This approach will find human putative enhancer regions, although orthologous regions for other species (e.g. mouse) can easily be inferred from this data where sequence conservation exists. In addition to the protocol here, we have also provided a public session, see Note 1.

1. At https://genome.ucsc.edu/index.html, click on "Genomes" in the toolbar, select "Human" and build "Feb. 2009 (GRCh37/hg19)" from the drop-down menu. Alternative European (https://genome-euro.ucsc.edu/index.html) and Asian https://genome-asia.ucsc.edu/index.html) mirror sites are available.

2. Type gene of interest or co-ordinates of the region of interest into the central search bar. Because enhancers can be identified away from the gene of interest, begin your enhancer search by focusing on 20-80 kb at either side of the transcriptional start site (TSS). This can later be widened if necessary. Our example in Figure 1 uses a 70 kb region around *EPHB4* (chr7:100,385,001-100,455,000). See Note 2 for ways to configure your view of the browser.

3. After selecting "default order", and "hide all" (Fig. 1A) in the display configuration tab, use the track display controls to access required information:
   a. Gene(s) of interest: Under "Gene and Gene Predictions" go to "UCSC Genes" and switch from "hide" to "pack" (Fig. 1B). You will now see all protein-coding and non-coding RNA genes, including splice variants and predictions. For a simpler display, click on "UCSC Genes" to access the configuration page and de-select "non-coding genes" and "splice variants", although this will remove potential e-RNAs from view.
   b. Enhancer-associated histone modifications: Under "Regulation" go to "ENCODE Regulation" (Fig. 1C), switch to "show" and click through to access configuration page. From there, select "Layered H3K4Me1", "Layered H3K4Me3" and "Layered H3K27Ac" subtracks and switch to "full" display mode. Click through "Layered H3K4Me1" to access settings page, where you can specifically select HUVEC and untick all others. Repeat for H3K4Me3 and H3K27Ac subtracks. Light blue enrichment peaks indicate enhancer-associated H3K4Me1 and H3K27Ac and promoter-associated H3K4Me3 histone marks in HUVECS for your region of interest (Fig. 1D). See Note 3.
   c. Enhancer-associated open chromatin regions: Under "Regulation" go to "ENC DNase/FAIRE" (Fig. 1E), switch to "show" then click the link to access configuration page. Select "UW DNAseI DGF" subtrack (which provides DNase I Digital Genomic Footprinting from ENCODE/University of Washington and includes a selection of endothelial cell lines) and switch to "dense". Leave all other subtrack options as "hide", click the link to UW DNaseI DGF settings page and select cell lines of interest. See Notes 4-5. Ensure maximum display mode is set to "dense" and return to the main genome browser. The DNase HS footprint will be displayed as a greyscale heatmap (see Fig. 1F).
   d. Sequence conservation: Under "Comparative Genomics", set "Conservation" to "pack", and click through to track settings (Fig. 1G). It is up to you to decide which species are

important for your research. We find that mouse, tenrec, opossum, chicken, x.tropicalis and zebrafish are good choices for comparative genomics. Select the species of interest, set Multiz Alignments to "pack" and leave the other options as "hide". Conservation will be displayed on browser by relative intensity or scarcity of black vertical lines, each indicating a conserved nucleotide (Fig. 1H).

e. ChIP-seq data for transcription factor binding patterns: While the UCSC Genome Browser main page includes publicly available ChIP-seq data tracks, there are currently few directly relevant to endothelial enhancer identification. However, you can browse ChIP-seq data via "ENCODE Regulation" through the "TF ChIP" subtrack, or by searching for specific transcription factors or cell lines using the "track search" feature in the display configuration buttons in the main browser page.

3.1.2 Using datasets accessed via ChIP-Atlas, the Sequence Read Archive (SRA) or via the Gene Expression Omnibus (GEO)

While the standardised tracks on UCSC can be very informative, there are increasing amounts of public ChIP-seq data available that can also assist in the identification of enhancers. In particular, all investigated endothelial enhancers contain binding motifs for the ETS family of transcription factors, and in virtually all cases these are essential for the activity of the enhancer [8]. Consequently, publicly available ChIP-seq datasets detailing peaks of different ETS factors can be used alongside histone marks and open chromatin patterns to aid the identification of functional enhancers. Examples of ETS factor ChIP-seq datasets in the public domain include ETS1 [12], ERG (e.g. from [13,14]), FL1 (e.g. from [15]) and ETV2 (e.g. from [16]). Additionally, there are a number of ChIP-seq datasets for EP300 (a transcriptional coactivator bound to most tissue specific enhancers) binding in endothelial cells (e.g. [13] which compares HUVEC with HUAEC, [17] which uses human pulmonary arterial ECs, and [18] in Tie2 Cre+ mouse endothelial and blood cells). Additional endothelial H3K4Me1 and H3K27Ac ChIP-seq datasets are also available for a variety of human and mouse cell lines/tissues cultured under different conditions.

These datasets can be accessed via the accession numbers provided within the published papers (usually through GEO https://www.ncbi.nlm.nih.gov/geo/). However, we find the simplest way to find the required ChIP-seq is often via ChIP-Atlas https://chip-atlas.org/. It is important to note that the quality of the ChIP data available on both ChIP-Atlas and GEO is highly variable, therefore datasets must be selected and interpreted with caution after considering the methods sections of the relevant papers.

1. To find ChIP-seq datasets using ChIP-Atlas, select "Dataset Search" from the top toolbar and enter transcription factor of interest and/or cell type of interest into the search window. Click on the SRX ID of each entry to access details about each dataset (e.g. culture conditions, cell line). Once the appropriate ChIP-seq sample is selected, you can visualise the data directly on IGV (Integrative Genomics Viewer) as a BigWig file (to visualise a histogram) or a peak-called Bed file (to see the location of statistically significant binding sites). Alternatively, you can note the SRXID number and genome build for use as a custom track on UCSC browser, or note the GEO ID number to access and download the data directly from GEO.

2. To access ChIP-seq datasets using GEO, enter the accession number provided within the paper into the search window. BigWig, BED and other files can then be downloaded directly from the results page. Alternatively, click through the SRA number (often provided in the "Relations" section on the GEO results page) to access the SRXID number for each run. Note that data downloaded from GEO is often of higher quality than raw reads accessed via SRX.

3. To visualise ChIP-seq data on IGV, first open an IGV browser window (this can be downloaded from https://software.broadinstitute.org/software/igv/). You can now directly look at data previously downloaded from GEO or ChIP-Atlas by going to "file" and selecting "download file". Alternatively, return to your page of interest in ChIP-Atlas, select "Visualize" and chose either BigWig or Peak Call at the required statistical significance. The relevant track will automatically open on your IGV browser.

4. To visualise ChIP-seq data on UCSC Browser alongside your histone and DNase HS data, return to your UCSC browser and click on "My Data" in the top toolbar (Fig. 1I), and select "Custom Tracks". Ideally, you want to visualize both the BigWig data (to see the pattern of binding as a histogram) and the BED files (which will show the statistically significant peak calls). You can upload a file previously downloaded from GEO or elsewhere, or paste a URL from ChIP-Atlas (see Notes 6-8).

5. Once custom tracks are loaded on UCSC Browser, you can alter the way the ChIP-seq information is displayed by right-clicking within the track and selecting "configure". "Dense" displays data as a heatmap and "full" displays data as a histogram (see Note 9 and Fig. 1J-K). Always check that the data looks similar to that shown in the original paper, as mislabelling of GEO deposited data can easily occur accidently. It is important to remember that local peak heights are not comparable across different ChIP-seq data tracks, as they depend heavily on experimental design and analysis pipelines. ENCODE data is the exception to this rule as it includes rigorous controls and has been processed using standardised pipelines that ensure data quality and replicability [11].

3.1.3 *Selecting your putative human endothelial enhancers*

1. Use the information now displayed on the Genome Browser (and/or IGV) to identify putative endothelial enhancers for your gene of interest. Putative endothelial enhancers should be defined as regions enriched in endothelial-specific H3K4Me1 and H3K27Ac marks alongside EC-specific patterns of DNase hypersensitivity, and will likely be bound by ETS factors and EP300 in endothelial cells. In Fig. 1, the validated Ephb4-2 venous enhancer is marked in red, while a second putative enhancer is marked in grey (Fig. 1L).

2. Enhancers can be located either within a gene (usually in introns but occasionally in exons) or either 5' upstream or 3' downstream. Additionally, endothelial genes are often regulated by more than one endothelial enhancer. Consequently, it is advised to test all putative enhancers within the gene loci. In cases where the neighbouring genes are not expressed in endothelial cells (this can be detected by lack of endothelial H3K4Me3 and/or absence of DNAse HS in endothelial cells lines), all putative enhancer regions within at least 100 kb either side of the TSS should be tested, and further if obvious putative enhancers

are observed. We have found that enhancers are often more distal from the TSS when the gene has no neighbouring genes, or contains large introns.

3. Once identified, you need to obtain the full DNA sequence of your human endothelial enhancer. While enhancers have no well-defined size limits, validation in zebrafish works optimally with putative enhancers between 100-900 bp. To obtain the sequence of your putative enhancer in UCSC browser, centre your region of interest using DNase HS region, zoom in until you are displaying the desired size of DNA, select "View" from the toolbar at the top of the page (Fig. 1I), "DNA" in the dropdown menu, and click "get DNA" button at the bottom of the window.

3.3.4 Finding endothelial enhancers for other species

A considerable number of publicly available ChIP-seq datasets are now available for enhancer-associated traits in many species, particularly mouse and zebrafish.

1. The UCSC Browser Mouse July 2007 (NCBI37/mm9) Assembly includes a "LICR Histone" track (in the "Expression and Regulation" section) with H3K4Me1, H3K4Me3 and H4K27Ac marks for a number of mouse tissues including embryonic and adult heart.

2. Other public ChIP-seq datasets that can be accessed via GEO or ChIP-Atlas, including ETS factor and EP300 binding in mouse endothelial cells (e.g. [16,18]), TF binding and histone marks in embryonic mouse cardiac cells (e.g. [19]), and ATAC-seq in endothelial vs non endothelial cells in zebrafish [20].

3. If your human enhancer region is conserved between species, the orthologous enhancer sequence from an alternative species can be obtained from the Multiz Alignment track in the human UCSC genome browser. Begin at the most conserved 100-200 bp region of your human enhancer and click directly on the alignment display (Fig. 1H) to go to the "Multiz Alignments of 100 Vertebrate" track page. Select a region of high conservation, find the required species and select the blue "B" at the far left-hand side to go to the the orthologous region on the UCSC genome browser. Alternatively, clicking "D" on the Multiz Alignment will take you directly to the "Get DNA" page, but you will have to adjust the genome coordinates to get the entire putative enhancer sequence.

3.2 Validating enhancer activity in zebrafish

While the exact definition of a functional enhancer can vary, enhancers are generally considered validated *in vivo* if they are shown capable of affecting gene expression levels. This is usually done either by cloning the putative enhancer upstream of a reporter gene and generating transgenic animals, or through germline deletion or mutation of the enhancer followed by analysis of expression levels of the assumed target gene [21]. In the case of endothelial enhancers, an *in vivo* validated endothelial-expressed enhancer would need to either drive reporter gene expression (e.g. GFP) in endothelial cells, or show reduced endothelial gene expression after germline deletion/mutation. Here, we detail how to use Tol2-driven mosaic transgenic zebrafish embryos [22,23], to validate enhancers. We

and many others have used this to successfully validate many endothelial enhancers, both using mammalian and teleost enhancer sequences [20,24].

3.2.1 Gateway Cloning to Create an Enhancer:E1b:Reporter Tol2 Plasmid.

First, you need to use pCR™8/GW/TOPO to create an entry vector plasmid containing your putative enhancer region:

1. Perform PCR with forward and reverse primers designed to amplify the putative enhancer region.

2. Using 10µl of PCR product, check the amplified region is the expected size using agarose gel electrophoresis and a standard DNA 1kb ladder.

3. Using the pCR™8/GW/TOPO TA Cloning Kit, combine
   2µl of PCR product (your putative enhancer)
   1µl of pCR™8/GW/TOPO plasmid
   1µl of salt solution (from kit)
   2.5µl of ddH_2O
   Incubate at room temperature for 10 minutes.

4. Transfer 2µl of pCR™8-enhancer/GW/TOPO mixture to a thawed vial of TOP10-competent E.Coli. Mix well by gently flicking the bottom of the tube. Incubate on ice for 10 minutes.

5. For transfection by heat shock, place E.coli in a 37.5°c water bath for 2 minutes, immediately on ice for 1 minute, then add 200µl of L.B broth to E.Coli and incubate in a 37.5°C shaking incubator for 30 minutes.

6. Plate transfected E.coli by spreading the entire 250µl volume onto a pre warmed agar plate containing 50ug/ml spectinomycin.

7. Incubate plates overnight at 37°C or until there are clearly visible but distinct colonies.

8. Using a pipette tip, pick 6 colonies and transfer each to an individual 15ml plastic canonical tube containing 2.5ml LB broth with 50µg/ml spectinomycin, 1 colony per tube. Grow overnight in a 37°c shaking incubator.

9. Perform DNA miniprep on bacteria from all overnight cultures to isolate plasmid DNA.

10. Assess putative enhancer integration into pCR™8/GW/TOPO plasmid by restriction enzyme digestion. See Note 10.

11. Select 2 pCR™8-enhancer/GW/TOPO plasmids containing the correct enhancer insert and Sanger sequence to check for any errors that may be introduced by PCR amplification.

Now that your putative enhancer sequence is in the pCR™8/GW/TOPO Gateway compatible entry vector, you can create an Tol2-enhancer:E1b:GFP-Tol2 expression plasmid. E1b is a minimal promoter commonly used to validate enhancer activity in zebrafish transgenesis, as it is does not drive notable expression of GFP alone. An E1b:GFP-Tol2 plasmid can be obtained from Addgene (see Materials). However, it should be noted that some researchers have reported better enhancer activity when using the endogenous promoter, particularly when looking at teleost enhancers [13].

The Tol2-enhancer:E1b:GFP-Tol2 expression plasmid is created using LR clonase:

15. Combine
    1µl of 37ng/µl pCR™8:enhancer/GW/TOPO mid
    1 µl of 37.5ng/µl pE1B:GFP-Tol2 expression plasmid
    0.5µl Gateway™ LR Clonase™ II Enzyme mix
    Spin briefly
    Incubate at room temperature for 1 hour.

12. To stop LR Clonase™ activity, add 0.5µl of proteinase K (from Invitrogen LR clonase kit) and incubate at 55°C for 15 minutes.

13. Transfect plasmid into TOP10 competent E.Coli by repeating above steps 4 to 8 replacing spectinomycin with ampicillin throughout.

14. Perform a DNA miniprep from bacteria from all overnight cultures.

15. Assess integration of putative enhancer region into E1b:GFP-Tol2 expression plasmid by restriction digest. The LR reaction is directional, so insert will be in the same orientation as in the pCR™8-enhancer/GW/TOPO plasmid.

16. Select an overnight culture that contains the desired plasmid and midiprep for a greater concentration of plasmid.

17. Sequence midiprep of your putative enhancer:E1b:GFP-Tol2 plasmid.

18. Store at -20°C.

3.2.2 Making Tol2 mRNA

1. Linearize your Tol2 expression plasmid by restriction enzyme digestion and column purify linearized plasmid.

2. Using the mMessage mMachine™ SP6 Transciption Kit combine:
   10µl 2 x NTP/CAP
   2µl 10 x reaction buffer
   2µl enzyme mix
   1µg linear expression plasmid

RNase-free H_2O to 20µl
Mix, spin and incubate for 2 hours at 37°C.

3. Add 1µl of TURBO DNase. Incubate for 15minutes at 37°C.

4. Column purify RNA and measure concentration on spectrophotometer.

5. Check mRNA is not denatured by running ~100ng mRNA on a 1% agarose gel in 0.5 x TBE buffer. See Note 11

6. Dilute to 1µg/µl in RNAse free H_2O. 1µg/µl aliquots are enough for the injection of 2 constructs.

7. Store aliquots at -80°C.

3.2.3 Creating Enhancer:E1b:GFP-Tol2 Transgenic Zebrafish Embryos

You are now ready to inject fertilized zebrafish eggs with a combination of your putative enhancer:E1b:GFP-Tol2 plasmid and Tol2 mRNA. These embryos will become F0 transgenic zebrafish (sometimes referred to as transient transgenic zebrafish). Remember that these embryos will be mosaic, as Tol2-induced transgenesis is not uniform.

This protocol requires long taper microinjection pipettes, normally made in-house by pulling capillary glass in a micropipette puller. It is important that this step is optimised for the specific capillaries/capillary puller/microinjector set up being used, as a good needle is crucial for injection success. Needles with a very thin shaft will break easily and be difficult to use, whilst needles with a shaft that is too thick will damage the chorion as they enter. See Note 12.

1. Thaw 1µg/µl aliquot of Tol2 mRNA and add 9µl of RNAse free H_2O to create a 100ng/µl solution.

2. Combine Enhancer:E1b:GFP-Tol2 plasmid with Tol2 mRNA for a 10µl solution of 60ng/µl plasmid 50ng/µl Tol2mRNA (e.g 5µl of 120ng/µl plasmid and 5µl of 100ng/µl mRNA).

3. Load the mixture into a fine tipped microinjection needle, break the tip of the needle with fine-tipped forceps and mount on micromanipulation system.

4. Using a calibration micrometre slide, adjust injection pressure and time on microinjector controls to deliver an injection volume of 0.5nl.

5. Collect zebrafish eggs immediately after fertilisation and arrange eggs in a single line along the long edge of a microscope slide in a petri dish orientated such that the single cell of the embryo is pointing away from the slide.

6. Inject plasmid mRNA mixture directly into the single cell. The needle should pierce the chorion and directly enter the cell but should not go through the yolk.

7. Once the line has been injected collect embryos with a plastic Pasteur pipette and transfer to a petri dish. Keep the embryos in embryo medium at 27°C until analysis.

3.2.4. Analysing zebrafish embryos

You can now assess putative enhancer activity by looking for GFP expression in the injected zebrafish embryos as they develop. We usually restrict our analysis to 22-96 hours post fertilisation (hpf), although later timepoints will be required if you are interested in the lymphatic vasculature. In our experience, validated mammalian enhancers (e.g. ones that also drive expression in mouse transgenic embryos) usually drive GFP expression in between 10-70% of injected zebrafish. Remember these F0 *tg(enhancer:E1b:GFP)* zebrafish will have only mosaic integration of the transgene, so GFP activity will not be in every endothelial cell. Figure 2 shows an example of an F0 *tg(Ephb4-2:E1b:GFP)* zebrafish, but note that sometimes GFP expression may be restricted to a small number of endothelial cells.

1. If imaging prior to 48 hpf, embryos require manually dechorionating. Keeping zebrafish embryos in E3 medium, use fine tipped forceps and a suitable microscope to tear a small hole in the chorion and release the embryo. Care must be taken not to apply excessive force as this can easily damage the embryo.

2. Anaesthetise fish by immersing in tricaine working solution and use a plastic pasteur pipette place 1 fish in each well of a 96 well plate suitable for confocal microscopy. Ensure fish remain submerged in tricaine throughout.

3. Image fish on a confocal microscope. Multiple Z-stacks will be required to assess GFP reporter gene expression in the zebrafish vasculature.

4. When imaging is complete, remove fish from 96 well plates back into E3 medium and keep at 27°C. Fish will recover quickly and can be re-anaesthetised and re-imaged as often as required.

4. Notes

1. We have created a UCSC public session with the basic tracks for endothelial enhancer identification. This session contains the UCSC genes, histone modifications, DNA accessibility, and comparative genomics tracks described here as well as relevant Chip-seq custom tracks of HUVEC cells extracted from Chip-Atlas. To access this public session, click on "My Data" in the toolbar, then click "Public Sessions", and search for it writing the session name (AngiogenesisProtocol_2021_DeVal) or the author (HRC) in the search space on the right. This public session is available at the main UCSC site and the European mirror site.

2. The configuration of the USCS browser can be altered (e.g. to increase font size, width of image or label location) by selecting “view” from the top tool bar, then selecting “configure browser”.

3. It is often very helpful to compare the histone marks in endothelial cells with those of other cells. To do this, go into the track settings page for each histone mark (e.g. Layered H3K4Me1 Track Settings, accessed from “ENCODE regulation”), and select any other cell line. The marks will be displayed in the colour indicated next to the cell line name.

4. The endothelial cell lines on offer include HUVEC (human umbilical vein endothelial cells), HMVEC-dBl-Ad (human dermal derived adult blood microvascular endothelial cells), HMVEC-dBL-Neo (human dermal derived neonatal blood microvascular endothelial cells), HMVEC-dLy-Neo (human dermal-derived neonatal lymphatic microvascular endothelial cells), HMVEC-LBl (human lung-derived blood microvascular endothelial cells) and HMVEC-LLy (human lung-derived lymphatic microvascular endothelial cells).

5. It is advisable to also select some non-endothelial cell lines for comparative purposes.

6. To add a BigWig file from ChIP-Atlas into UCSC as a custom track:

   track type=bigWig name=track name.bw description=track description
   bigDataUrl=http://dbarchive.biosciencedbc.jp/kyushu-u/hg19 or hg38 /eachData/bw/SRXID.bw

7. To add a BigBed file from ChIP-Atlas into UCSC as a custom track (red text is for you to change as needed, green text can be adjusted for required q):

   track type=bigBed name=track name_q05 description=track description
   bigDataUrl=http://dbarchive.biosciencedbc.jp/kyushu-u/hg19 or hg38/eachData/bb05/SRXID.05.bb

   * q refers to the MACS2 Q-value threshold for statistical significance and can displayed as 05, 10, or 20. As a reference: if the MACS2 threshold of significance is set to 50, Chip-Atlas shows a q value < 1E-05.

8. An example URL to use in UCSC browser as a custom track, using part of a dataset from [13], showing binding of the ETS factor ERG in HUVECs:

   track type=bigWig name=ERG_vein2.bw description=LawsonSRX5527659
   bigDataUrl=http://dbarchive.biosciencedbc.jp/kyushu-u/hg19/eachData/bw/SRX5527659.bw
   track type=bigBed name=ERG_vein2_q05 description=Lawson_qE05_SRX5527659
   bigDataUrl=http://dbarchive.biosciencedbc.jp/kyushu-u/hg19/eachData/bb05/SRX5527659.05.bb

9. Within the “configure” window for each custom track there are also options for scaling of the y axis. We often find it useful to autoscale the axis in order to view locally enriched binding, but it is important to note that the apparent size of any local peak is relevant only to what is currently within the browser window and therefore these ‘sizes’ are not comparable across regions.

10. As pCR™8/GW/TOPO/TA cloning is not directional it is helpful to choose enzymes that will give distinct band sizes for each integration option, that is; no insert, enhancer inserted in 5'-3' orientation, enhancer inserted in 3'- 5' orientation. Typically we select the 5'-3' orientation for enhancers located in the 5' upstream region, and the 3' - 5' orientation for enhancers located 3' of the TSS of the relevant gene.

11. Specific RNA gel protocols are available, but in our experience standard agarose gels are sufficient to visualise Tol2 mRNA at a high concentration. DNA/RNA dye must be added directly to mRNA

12. As a **guide only**, Flaming/Brown micropipette puller model P-1000 with capillary glass 1.0 OD x 0.58 ID X 150 L (mm) uses the following programme:
    a. Heat 500
    b. Pull 110
    c. Velocity 75
    d. Time 90
    e. Pressure 500
    f. This will pull a needle of ~0.5um that is ~150 – 200µm long.

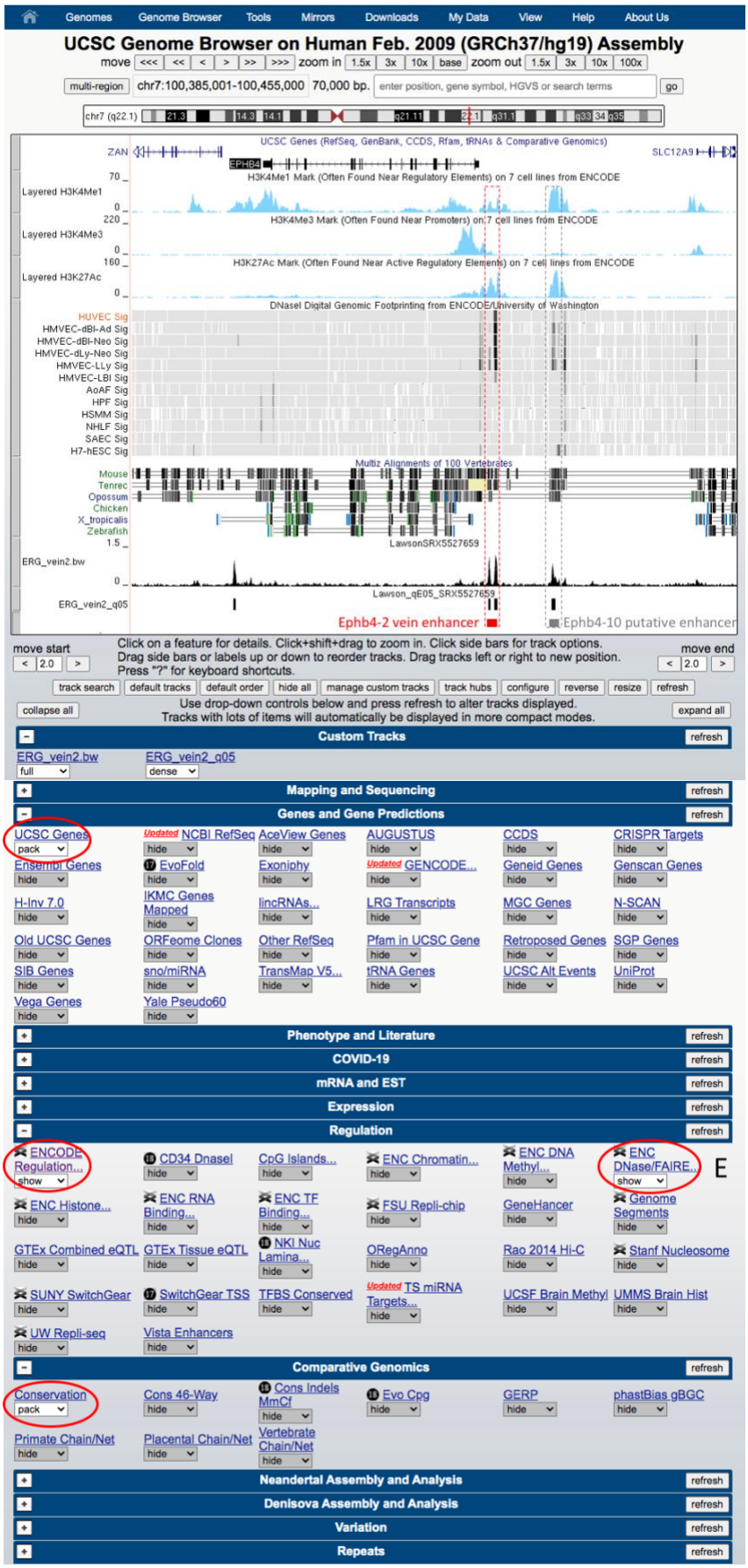

Figure 1. Screenshot of the USCS browser showing relevant portions discussed in text. Ephb4-2 and Ephb4-10 enhancers are described in detail in [6]. ERG ChIP-seq data comes from [13].

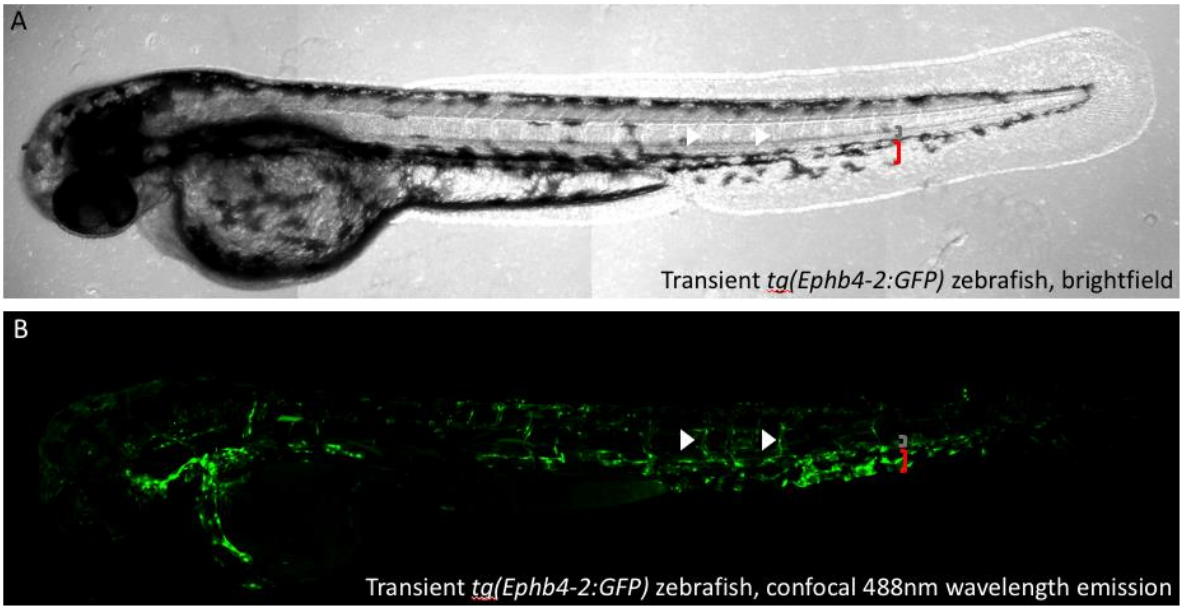

Figure 2. (**A**) brightfield and **(B)** GFP (488nm wavelength emission) images of an anaesthetised transgenic zebrafish embryo at 48 hours post fertilisation (hpf). This embryo was injected at the 1 cell stage with an E1b-GFP-Tol2 expression plasmid containing the Ephb4-2 enhancer [6] along with Tol2 mRNA, making an F0 *tg(Ephb4-2:E1b:GFP)* transgenic zebrafish. The GFP signal demonstrates that the Ephb4-2 enhancer can drive gene expression in the venous endothelium. Red brackets indicate the caudal vein expressing GFP, grey brackets indicate the aorta which is GFP negative, white arrow heads indicate GFP expression in venous intersegmental vessels.

5. References

1. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7,** 29–59 (2006).
2. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461,** 199–205 (2009).
3. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15,** 272–286 (2014).
4. Gasperini, M., Hill, A. J., McFaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., Jackson, D., Leith, A., Schreiber, J., Noble, W. S., Trapnell, C., Ahituv, N. & Shendure, J. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176,** 377–390.e19 (2019).
5. Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. & de Graaff, E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12,** 1725–1735 (2003).
6. Neal, A., Nornes, S., Payne, S., Wallace, M. D., Fritzsche, M., Louphrasitthiphol, P., Wilkinson, R. N., Chouliaras, K. M., Liu, K., Plant, K., Sholapurkar, R., Ratnayaka, I., Herzog, W., Bond, G., Chico, T., Bou-Gharios, G. & De Val, S. Venous identity requires BMP signalling through ALK3. *Nat Commun* **10,** 453 (2019).
7. Ma, S. & Zhang, Y. Profiling chromatin regulatory landscape: insights into the development of ChIP-seq and ATAC-seq. 1–13 (2020). doi:10.1186/s43556-020-00009-w
8. De Val, S. & Black, B. L. Transcriptional Control of Endothelial Cell Development. *Dev Cell* **16,** 180–195 (2009).
9. Sewitz, S. A., Fahmi, Z. & Lipkow, K. Higher order assembly: folding the chromosome. *Curr Opin Struct Biol* **42,** 162–168 (2017).
10. Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nature Structural & Molecular Biology* 1–8 (2020). doi:10.1038/s41594-020-0446-0
11. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
12. Chen, J., Fu, Y., Day, D. S., Sun, Y., Wang, S., Liang, X., Gu, F., Zhang, F., Stevens, S. M., Zhou, P., Li, K., Zhang, Y., Lin, R.-Z., Smith, L. E. H., Zhang, J., Sun, K., Melero-Martin, J. M., Han, Z., Park, P. J., Zhang, B. & Pu, W. T. VEGF amplifies transcription through ETS1 acetylation to enable angiogenesis. *Nat Commun* **8,** 383 (2017).
13. Sissaoui, S., Yu, J., Yan, A., Li, R., Yukselen, O., Kucukural, A., Zhu, L. J. & Lawson, N. D. Genomic Characterization of Endothelial Enhancers Reveals a Multifunctional Role for NR2F2 in Regulation of Arteriovenous Gene Expression. *Circ Res* **126,** 875–888 (2020).
14. Kalna, V., Yang, Y., Peghaire, C. R., Frudd, K., Hannah, R., Shah, A. V., Osuna Almagro, L., Boyle, J. J., Göttgens, B., Ferrer, J., Randi, A. M. & Birdsey, G. M. The Transcription Factor ERG Regulates Super-Enhancers Associated With an Endothelial-Specific Gene Expression Program. *Circ Res* **124,** 1337–1349 (2019).
15. Wang, S., Chen, J., Garcia, S. P., Liang, X., Zhang, F., Yan, P., Yu, H., Wei, W., Li, Z., Wang, J., Le, H., Han, Z., Luo, X., Day, D. S., Stevens, S. M., Zhang, Y., Park, P. J., Liu, Z.-J., Sun, K., Yuan, G.-C., Pu, W. T. & Zhang, B. A dynamic and integrated epigenetic

program at distal regions orchestrates transcriptional responses to VEGFA. *Genome Res* **29,** 193–207 (2019).
16. Liu, F., Li, D., Yu, Y. Y. L., Kang, I., Cha, M. J., Kim, J. Y., Park, C., Watson, D. K., Wang, T. & Choi, K. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO Rep* **16,** 654–669 (2015).
17. Miyagawa, K., Shi, M., Chen, P.-I., Hennigs, J. K., Zhao, Z., Wang, M., Li, C. G., Saito, T., Taylor, S., Sa, S., Cao, A., Wang, L., Snyder, M. P. & Rabinovitch, M. Smooth Muscle Contact Drives Endothelial Regeneration by BMPR2-Notch1-Mediated Metabolic and Epigenetic Changes. *Circ Res* **124,** 211–224 (2019).
18. Zhou, P., Gu, F., Zhang, L., Akerberg, B. N., Ma, Q., Li, K., He, A., Lin, Z., Stevens, S. M., Bin Zhou & Pu, W. T. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *eLife Sciences* **6,** e22039 (2017).
19. Akerberg, B. N., Gu, F., VanDusen, N. J., Zhang, X., Dong, R., Li, K., Zhang, B., Bin Zhou, Sethi, I., Ma, Q., Wasson, L., Wen, T., Liu, J., Dong, K., Conlon, F. L., Zhou, J., Yuan, G.-C., Zhou, P. & Pu, W. T. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun* 1–16 (2019). doi:10.1038/s41467-019-12812-3
20. Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L. J. & Lawson, N. D. Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq. *Cell Rep* **20,** 709–720 (2017).
21. Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K. E., Ni, M. & Xu, J. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun* **11,** 485–16 (2020).
22. Kawakami, K. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biology 2010 11:R7* **8 Suppl 1,** S7–10 (2007).
23. Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K., Kawakami, K. Transgenesis in zebrafish with the tol2 transposon system. *Methods Mol Biol*. **561**:41-6 (2009).
24. Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P. W., Sanchez-del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C. R., Patient, R., Bou-Gharios, G. & De Val, S. Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. *PNAS* **110,** 11893–11898 (2013).