

UNIVERSITY OF NAPLES FEDERICO II
DOCTORATE IN
MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY
XXXV CYCLE



IDENTIFYING CRITICAL REGULATORY FACTORS
DURING CARDIOPHARYNGEAL MESODERM (CPM)
DIFFERENTIATION INTO ENDOTHELIAL CELLS (ECs)

Tutor

Prof. Antonio Baldini

Candidate

Ilaria Aurigemma

Co-tutor

Prof.ssa Elizabeth Anne Illingworth

2019-2023

UNIVERSITY OF NAPLES FEDERICO II

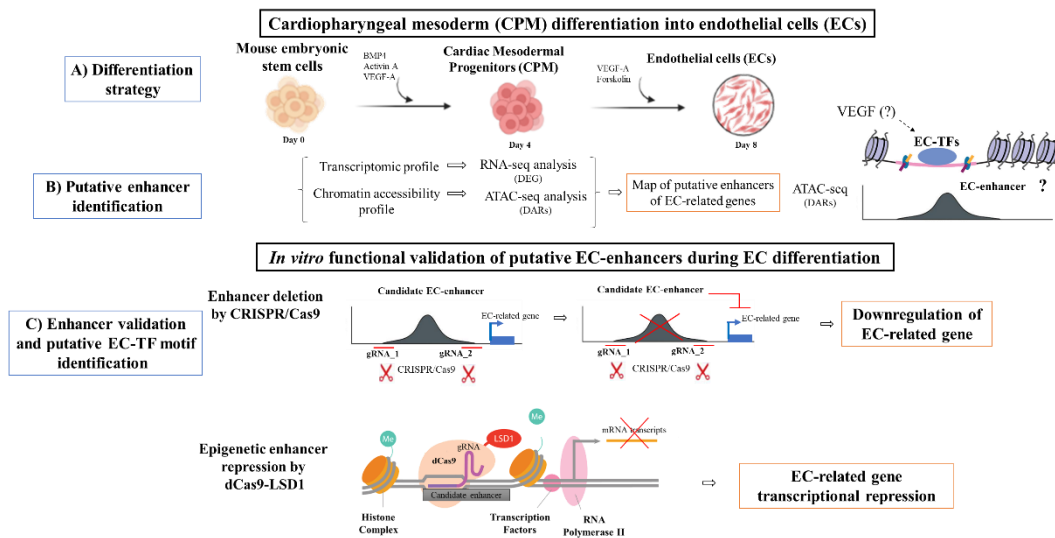
DOCTORATE IN
MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXV CYCLE



Ilaria Aurigemma

**IDENTIFYING CRITICAL REGULATORY FACTORS
DURING CARDIOPHARYNGEAL MESODERM (CPM)
DIFFERENTIATION INTO ENDOTHELIAL CELLS (ECs)**



2019-2023

TABLE OF CONTENTS

TABLE OF CONTENTS	2
ABSTRACT	6
ABBREVIATIONS	8
LIST OF FIGURES	10
LIST OF TABLES	27
INTRODUCTION	30
1. Cardiogenesis in mouse development: signaling pathways from mesoderm induction to cardiogenic mesoderm progenitors.....	30
2. Specification of the endothelial cell lineage (EC) during mouse development.....	35
3. Transcriptional control and regulation of endothelial cell and vascular development.....	38
3.1 <i>Tal1 (Sc11)</i>	38
3.2 <i>GATA2 (GATA binding protein 2)</i>	38
3.3 <i>ETS Transcription Factors in regulation of endothelial gene expression</i>	39
3.3.1 <i>Role of Etv2 in early vascular development</i>	40
3.4 <i>Forkhead Proteins are important regulators of endothelial transcription</i>	40
3.5 <i>HEY1/HEY2 (Hairy/enhancer-of-split related with YRPW motif protein1/2)</i>	41
3.6 <i>SOX F group transcription factors: SOX7, SOX17 and SOX18</i>	41
4. Signaling Pathways in Regulating Endothelial Transcription Factors and Vascular Development.....	43
4.1 <i>VEGF signaling</i>	43
4.2 <i>NOTCH signaling</i>	46
4.3 <i>Signaling via Bone Morphogenetic Proteins (BMP Signaling)</i>	47
4.4 <i>WNT Signaling</i>	48
5. In vitro model to induce multipotent cardiac progenitors to differentiate into endothelial cells (ECs), starting from mESC	49
5.1 <i>Undifferentiated ES cells</i>	49

5.1.1	Cell states and identities	50
5.2	Differentiation model of ES cells	52
5.3	Cardiac differentiation of mouse embryonic stem cell lines	54
5.4	Vascular endothelial cell differentiation from ES cells	57
6.	DNA-regulatory elements: Enhancers	61
6.1	Molecular assay to identify candidate enhancer regions	62
6.1.1	MNase-seq chromatin accessibility assay	63
6.1.2	DNase-seq chromatin accessibility assay	63
6.1.3	FAIRE-seq chromatin accessibility assay	64
6.1.4	ATAC-seq chromatin accessibility assay	64
6.2	Epigenetic features of enhancer	66
6.3	Functional validation enhancer assay	68
6.3.1	Traditional reporter assays for enhancer activity	69
6.3.2	Massively Parallel Reporter Assays (MPRA) to study enhancer function 70	
6.4	CRISPR/Cas9-based methods to functionally validate candidate enhancers	71
6.4.1	Genetic manipulation of enhancers	72
6.4.2	dCas9 system for gene regulation and epigenome control	73
6.4.3	CRISPR/Cas9-effector system mediated gene activation and repression	75
AIMS	82
MATERIALS AND METHODS	83
1.	Cell biological methods	83
1.1	Mouse embryonic stem cells (mESC)	83
1.2	In vitro differentiation protocol	83
1.3	CRISPR-Cas9-Mediated Targeting of mESCs	84
1.3.1	Tbx1 knockout mESC line by CRISPR-Cas9 technology	84
1.3.2	Pecam1 intron2-enhancer deletion by CRISPR-Cas9 technology	85
1.3.3	Notch1 intron15-enhancer deletion by CRISPR-Cas9 technology	85

1.3.4	<i>Generation of dCas9-LSD1 expressing mESC line by electroporation.....</i>	86
1.4	<i>mESC reverse transfection protocol</i>	87
1.5	<i>Tube formation assay of mESC -ECs plated on Matrigel.....</i>	88
1.6	<i>Fluorescence Activated Cell Sorting (FACS)</i>	88
1.6.1	<i>Evaluation of endothelial cells (EC) from differentiated mESC88</i>	
1.6.2	<i>gRNA transfection efficiency</i>	88
2.	<i>Molecular biological methods</i>	88
2.1	<i>Reverse transcription and PCR amplification (RT-PCR)</i>	88
2.2	<i>Quantitative real-time PCR (qRT-PCR)</i>	89
2.3	<i>Statistical Analysis</i>	89
2.4	<i>ATAC-seq</i>	89
2.5	<i>RNA-seq.....</i>	90
2.6	<i>Integrated analysis</i>	91
RESULTS.....		92
Chapter 1. Establishment of a model for differentiation of cardiopharyngeal mesoderm (CPM) into endothelial cells (ECs) starting from mouse embryonic stem cells (mESCs).		92
1.1	<i>Characterization of mESC-derived differentiated cells</i>	94
1.2	<i>mESC-derived endothelial cells form vascular network-like structures in vitro.....</i>	97
Chapter 2. Transcriptomic and chromatin accessibility profile in early phases of multipotent cardiac progenitors' differentiation into mature ECs. 98		
2.1	<i>Gene expression profile of multipotent cardiac progenitors at day2-day4 (RNA-seq analysis).....</i>	98
2.2	<i>Chromatin accessibility profile in early phases of multipotent cardiac progenitors' differentiation into mature ECs.</i>	106
2.3	<i>Intersection of DE genes and DARs regions between day2 and day4 of CPM-EC differentiation.....</i>	116
Chapter 3. Identification of putative endothelial regulatory elements (enhancers) regulated during CPM differentiation into endothelial cells (EC) by integration of RNA-seq and ATAC-seq data.		118
3.1	<i>Putative endothelial enhancers validation strategies</i>	124

Chapter 4. Generation of mutant engineered mouse embryonic stem cell lines (mESC) using the CRISPR/Cas9 system.	126
4.1 <i>Tbx1</i> knockout (<i>Tbx1</i> ^{-/-}) mESC lines.....	128
4.2 <i>Pecam1</i> -Δ <i>enh. intron2</i> mESC lines.....	131
4.3 <i>Notch1</i> -Δ <i>enh. intron15</i> mESC lines	134
4.4 <i>dCas9-LSD1</i> mESC lines	137
Chapter 5. Validation of putative endothelial enhancers and verification of their requirement during EC differentiation.....	141
5.1 <i>Notch1-intron15</i> enhancer validation: evaluation of gene expression consequences during ECs differentiation.	141
5.1.1 Assessment of <i>Notch1-intron15</i> putative enhancer using CRISPR/Cas9 deletion strategy.....	141
5.1.2 Assessment of <i>Notch1-intron15</i> putative enhancer using CRISPR/dCas9 – <i>LSD1</i> epigenetic repression strategy.....	150
5.2 <i>Pecam1-intron2</i> enhancer validation: evaluation of gene expression consequences during ECs differentiation, using CRISPR/Cas9 deletion strategy.....	155
5.3 <i>Kdr-intron10</i> enhancer validation by epigenetic repression, using CRISPR/dCas9- <i>LSD1</i> : evaluation of transcriptional consequences during ECs differentiation.	164
5.4 <i>VE-Cadherin-intron1</i> enhancer validation by epigenetic repression, using CRISPR/dCas9- <i>LSD1</i> : evaluation of transcriptional consequences during ECs differentiation.	168
5.5 <i>Eng-intron2</i> enhancer validation by epigenetic repression, using CRISPR/dCas9- <i>LSD1</i> : evaluation of transcriptional consequences during ECs differentiation.	173
5.6 <i>Flt1-intron10</i> enhancer validation by epigenetic repression, using CRISPR/dCas9- <i>LSD1</i> : evaluation of transcriptional consequences during ECs differentiation.	179
5.7 Computational prediction of transcription factor motifs in EC enhancers.	184
DISCUSSION.....	188
CONCLUSIONS.....	194
LIST OF PUBLICATIONS.....	195
REFERENCES.....	196

ABSTRACT

The development of the cardiovascular system requires the coordinated differentiation of several cell types including endothelial cells (EC), smooth muscle cells and cardiomyocytes. This process involves the differentiation of cardiopharyngeal mesoderm (CPM) from which these cardiac cell types derive. The overall aim of my doctoral work is to understand better the genetic and epigenetic mechanisms responsible for cell fate transitions in multipotent cardiac progenitors to differentiate into endothelial cells (ECs).

Notably, I have developed a model for differentiation of CPM into ECs starting from engineered mouse embryonic stem cells (mESCs), using a serum-free protocol with the addition of specific growth factors that induce cardiac and endothelial differentiation.

The results showed that this procedure allows rapid vascular differentiation with high efficiency. I obtained approximately 91% CD144+ (VE-Cadherin) cells within 8 days. Then I performed RNA-seq and ATAC-seq on the early phases of CPM differentiation to define the transcriptomic and chromatin accessibility profile. First, I demonstrated that mESCs differentiation promoted the expression of EC-specific markers at day 4 of differentiation (d4), including *Pecam1*, *VE-Cadherin* (*Cdh5*), *Eng*, *Kdr*, *Gata2*, *Gata6*, *Ets1*, *Flt1* and others. RNAseq performed between d2 and d4 identified 1735 differentially expressed genes, many of which are involved in angiogenesis, indicating the activation of an EC transcription program. ATAC-seq revealed 6348 Differential Accessible Regions (DARs) that changed their chromatin accessibility during this time window. Most of them were located in intra- and inter-genic regions. Thanks to the integration of these two methods, I identified, at the first, 2 putative enhancers defined as regions of increased accessibility, associated with endothelial-specific genes: *Pecam1* and *Notch1*, both of which are critical for vascular development. Subsequently, I have extended the search of putative regulatory elements, identifying other 8 open chromatin regions, associated with *Kdr* (*Vegfr2*), *Cdh5* (*VE-Cadherin*), *CD34*, *Eng*, *Flt1* (*Vegfr1*), *Tall* (*Scl1*), *Dusp5* and *Gata6* endothelial genes.

To validate the putative regulatory regions, I followed two strategies: DNA editing (putative enhancers deletion) and epigenetic decommisioning.

For the first approach, I generated mESCs with deletion of *Pecam1*-enh.int2 and *Notch1*-enh.int15 (by CRISPR-Cas9 technology), which I then differentiated towards CM-EC lineages. Two *Notch1*- Δ enh.int15. independent mESC clones showed a significant reduction of *Notch1* expression during the later stages of EC differentiation (d6 and d8). Similarly, *Pecam1* expression was also downregulated in two independent *Pecam1*- Δ enh.int2. mESC clones at the same time points.

These results indicated that the regions deleted are required for appropriate expression of the respective genes during EC differentiation process.

The second validation strategy was based on epigenetic reprogramming by nuclease-deficient dCas9 fused with histone demethylase LSD1 (dCas9-LSD1). It removes mono and di-methylation of histone H3 lysine 4 (H3K4me1 and me2) to promote the change of chromatin shape into a repressive configuration. I generated mESC clones constitutively expressing dCas9-LSD1 and transfected these cells with gRNAs targeting the putative enhancers and then differentiated into endothelial cells (ECs). In particular, I have analyzed so far only five putative enhancer regions: *Notch1*-enh.intr15; *Kdr*-enh.intr10; *VE-Cadh*.-enh.intr1; *Eng*-enh.intr2; *Flt1*-enh.intr10. The targeted five loci resulted affected by dCas9-LSD1 epigenetically repression, giving rise to relative reduction of gene-related expression, specifically at day8 of differentiation.

Overall, 6 tested out of 10 identified putative enhancers seems to be regulatory elements and could be involved during later stages of EC differentiation.

Moreover, to predict, computationally, transcription factor motifs in EC enhancers, I performed a preliminary motif analysis of DARs regions related to endothelial cell fate specification. Sequence analyses of regions opened at d4 identified *Gata1*, *Gata2* and *JunB* transcription factors. They could regulate the differentiation of cardiopharyngeal mesoderm progenitors in derivative tissues, including EC.

In conclusion, the experimental model and methods used for differentiation of CPM into ECs allowed me to efficiently identify novel putative endothelial enhancers. Thanks to genetic and epigenetic manipulation of these sequences, I established their requirement for the transcription process during differentiation from cardiopharyngeal mesoderm to ECs.

ABBREVIATIONS

- Acetylation of lysine 27 of histone 3 → H3K27Ac
- Assay for Transposase-Accessible Chromatin → ATAC
- basic Helix Loop Helix → bHLH
- Bone Morphogenic Protein → BMP
- Cardiac fibroblasts → CF
- Cardiac Progenitors → CP
- Cardiomyocytes → CMs
- Cardiogenic Mesoderm Cells → CMCs
- Cardiopharyngeal Mesoderm Progenitors → CPM
- Clustered Regularly Interspaced Short Palindromic Repeats → CRISPR
- Complementary Deoxyribonucleic Acid → cDNA
- Congenital Heart Diseases → CHDs
- CRISPR activation → CRISPRa
- CRISPR interference → CRISPRi
- dead Cas9 → dCas9
- Deoxyribonucleic acid → DNA
- Destination Cell Types → DCTs
- Differentially Accessible Regions → DARs
- Differentially Expressed → DE
- DNase I hypersensitive sites sequencing → DNase-seq
- Double Stranded Breaks → DBS
- Embryoid Bodies → EBs
- Embryonic Stem Cells → ESCs
- Endothelial cells → EC
- Epicardial cells → Eps
- Formaldehyde-Assisted Isolation of Regulatory Elements → FAIRE
- Fetal Calf Serum → FCS
- Fetal Bovine Serum → FBS
- Fibroblast Growth Factors → FGFs
- Fluorescent Activated Cell Sorter → FACS
- Gene Ontology → GO
- H3K4 mono-methylation → H3K4Me1
- H3K27acetylation → H3K27ac
- H3K27 tri-methylation → H3K27Me3
- Hypergeometric Optimization of Motif Enrichment → HOMER
- Holt-Oram syndrome → HOS
- Human Pluripotent Stem Cell → hPSCs

- Kox1 Krüppel-associated box → KRAB
- Leukemia inhibitory factor → LIF
- Lymphatic Endothelial cell → LEC
- Lysine Specific Demethylase 1 → LSD1
- Micrococcal Nuclease Digestion With Deep Sequencing → MNase-seq
- Mitochondrial reads → M
- Monomethylation Of Lysine 4 Of Histone 3 → H3K4me1
- Murine Embryonic Stem Cells → mESC
- NK2 transcription factor related, locus 5 → Nkx.2.5
- Nonhomologous End-Joining → NHEJ
- Polymerase Chain Reaction → PCR
- Primitive streak → PS
- Proepicardium → PE
- Retinoic Acid → RA
- RNA polymerase II → RNAPII
- single cell RNA-sequencing → scRNA-seq
- Smooth Muscle Cells → SMCs
- Transcription Activator-Like Effector → TALE
- Transcription Factor Binding Sites → TFBS
- Transcriptional Start Site → TSS
- Vascular Endothelial Growth Factor → VEGF
- Wingless-Related MMTV Integration Sites → WNTs

LIST OF FIGURES

Figure 1 Signaling pathways for cardiac progenitor cells differentiation and their lineage specification (Brade T. <i>et al.</i> , (2013)).....	32
Figure 2 Violin plots showing the mean and variance difference between WT (red) and <i>Mesp1</i> KO (blue) cells of genes regulating pluripotency (<i>Nanog</i> , <i>Eras</i>) and epiblast (<i>Cdh1</i>) (Lescroart F. <i>et al.</i> , (2018)).....	33
Figure 3 Violin plots showing the mean and variance of EMT genes (<i>Snail</i>), cardiovascular markers (<i>Gata4</i> , <i>Etv2</i> , <i>Myl7</i> <i>Flk1/Kdr</i> and <i>Pdgfra</i>) expression in WT (red) and <i>Mesp1</i> KO (blue) (Lescroart F. <i>et al.</i> , (2018)).	34
Figure 4 Scheme of the different <i>Mesp1</i> subpopulations (named DCTs 1-4) and their localization during gastrulation (Lescroart F. <i>et al.</i> , (2018)).....	35
Figure 5 Vasculature formation during mouse embryo development (Park C. <i>et al.</i> , (2013)).....	36
Figure 6 Map of the selective cell surface markers expression during mesoderm specification and vascular cell derivation from ESCs (Ema M. <i>et al.</i> , (2006)).	37
Figure 7 Development of blood cells, endothelial cells, cardiomyocytes, vascular smooth muscle cell lineages from <i>Flk1</i> -expressing mesoderm (Park C. <i>et al.</i> , (2013)).....	38
Figure 8 Different stages in endothelial development, regulated by distinct group of transcription factors. Based on this scheme, <i>Fli-1</i> , GATA2, and <i>Tal1</i> control differentiation of hematopoietic cells from hemangioblasts, while <i>Etv2</i> and <i>FoxC</i> proteins control the differentiation of endothelial cells from that progenitor population and that <i>Etv2</i> likely sits at the top of this transcriptional cascade (De Val S. <i>et al.</i> , (2009)).....	42
Figure 9 Developmental of FLK1+ mesoderm. FLK1+ mesoderm originates from Brachyury/T-expressing cells in the primitive streak of the developing embryo and subsequently gives rise to not only endothelial and blood cells but also other mesodermal lineage cells, including vascular smooth muscle, cardiomyocyte, and skeletal muscle cells (Park C. <i>et al.</i> , (2013))	44
Figure 10 Schematic representation of VEGFR2 signaling pathways. VEGFR2 (https://www.tocris.com/pathways/vegf-signaling-pathway).	44
Figure 11 Signal transduction pathways from VEGFR-1 in vascular endothelial cells (Shibuya M. <i>et al.</i> , (2006)).....	45
Figure 12 Overview of the Notch signaling pathway and its role in regulation of target genes involved in biological processes including EndoMT, angiogenesis and CSC biology (Akil A. <i>et al.</i> , (2021)).....	47
Figure 13 The Waddington's epigenetic landscape model. A. Cellular properties (plasticity, potency), natural state changes (fate choice, lineages, cell types, differentiation), and molecular foundations (genes, gene expression); B. Cellular reprogramming, which changes the cell states; C. Transcription factors controlling	

cellular states. Master transcription factors influence cell fate decisions during development. Reprogramming factors can alter cell identities beyond the natural potency of the cell (Breunig CT. <i>et al.</i> , (2020)).	51
Figure 14 Scheme of early mouse development. ES cells differentiate in derivatives of the three embryonic germ layers: mesoderm, endoderm and ectoderm (Keller G. <i>et al.</i> , (2005)).	52
Figure 15 Three general approaches used to induce ES cell differentiation (Murry CE. <i>et al.</i> , (2008)).	53
Figure 16 Cardiac Potential of <i>Flk⁺/Pdgrf-α⁺</i> (F ⁺ P ⁺) mesoderm cell population (Kattman <i>et al.</i> , (2011)).	55
Figure 17 Cardiomyocyte differentiation of embryonic stem cells. a. Scheme of differentiation protocol, indicating the time points where differentiating cultures were treated with the specific growth factors listed below. b. cTnT (in red) immunofluorescence of day10 differentiated cells. c. Flow cytometry at day10 of cells stained for cTnT or isotype control (Wamstad <i>et al.</i> , (2012)).	56
Figure 18 Transcriptional profile during transition from pluripotent cells to cardiomyocytes. a. A representation of the four stages of cardiac differentiation. b. Expression of the main marker genes activated during cardiac differentiation (Wamstad <i>et al.</i> , (2012)).	57
Figure 19 a. Schematic illustration of ECs differentiation protocol from hPSCs. b. Schematic illustration of VSMCs differentiation strategy from hPSCs (Patsch <i>et al.</i> , (2015)).	58
Figure 20 a. Differentiation efficiency of hPSC ECs and hPSC VSMCs from four different hESC and induced PSC lines by evaluation of CD144 ⁺ (VE-Cadherin) and CD140b ⁺ positive cells. b. FACS sorting plots from the differentiation experiments in hPSCs (top panel), hPSC-derived ECs (middle panel) or hPSC-derived VSMCs (lower panel) stained for CD144 and CD140b. c. Immunostaining of EC-specific markers (VE-Cadherin, vWF, Pecam1) on hPSC ECs for both GSK3 inhibitors. d. Immunostaining of VSMC-specific markers on hPSC VSMCs (α SMA, myosin IIB, SM22a) (Patsch <i>et al.</i> , (2015)).	59
Figure 21 Flow cytometry analysis of CD144 ⁺ hPSC-ECs on day 10 of differentiation. Detection of endothelial- specific genes expression (KDR, CD31, CD34 and CD105), but no hematopoietic markers (CD43, CD45) (Patsch <i>et al.</i> , (2015)).	60
Figure 22 <i>In vitro</i> tube formation assay by plating hPSC ECs on Matrigel for 24 h (Patsch <i>et al.</i> , (2015)).	60
Figure 23 Schematic diagram representing enhancer activity for various biological processes governing development. Enhancers play a central role in cell fate determination; in regulating transcriptional programs that control development, cell identity and evolutionary processes. They also serve as docking sites for TFs, and the activity of enhancers is mostly based on the binding of these TFs.	

Enhancer reprogramming is an emerging area in developmental biology and cancer research: it could represent a hallmark of carcinogenesis, as it contributes to the deregulated expression of epigenetic modifiers, leading to abnormal cell growth (Maurya SS. <i>et al.</i> , (2021)).	61
Figure 24 Representation of enhancers loop: enhancers can regulate gene expression by recruiting TFs and the transcriptional machinery and subsequently forming a loop with the promoter region of the target gene (Claringbould A. <i>et al.</i> , (2021)).	62
Figure 25 Genomic methods in mapping chromatin accessibility. a. DNase-seq: enzymatic digestion to extract signal from open chromatin regions that are known as DNase I hypersensitive sites (DHSs); b. FAIRE-seq: method based on crosslinking of chromatin-interacting proteins to DNA using formaldehyde. Chromatin is then sheared, and regions that are unbound by proteins (e.g., histones) remain in the aqueous layer of a phenol-chloroform extraction, while crosslinked DNA remains in the organic layer; c. MNase-seq: enzymatic digestion to extract signal representing nucleosome positioning. After formaldehyde crosslinking, added MNase digests DNA that is unprotected by bound proteins, allowing one to infer increased accessibility; d. ATAC-seq based on the hyperactive Tn5 transposase to insert sequencing adapters at accessible regions of the genome. Following transposition, genomic DNA can be isolated and amplified by PCR, then subjected to deep sequencing (Klein DC. <i>et al.</i> , (2019)).	65
Figure 26 Epigenetic Features of active (A), primed (B), and poised enhancers (C). A. Active state: enhancers are associated with incorporation of hypermobile nucleosomes containing H3.3/H2A.Z histone variants, which compete for DNA binding with TFs. TFs in turn recruit coactivator proteins (e.g. p300 histone acetyltransferase) that can modify and remodel nucleosomes. H3K4me1 and H3K27ac are the predominant histone modifications deposited at nucleosomes flanking enhancer elements. In their active state, enhancers are also bound by general transcription factors (GTFs) and RNA polymerase II (Pol II), leading to the production of enhancer-originating RNAs termed eRNAs. B. Primed state (prior to activation): enhancers are characterized by the presence of H3K4me1. Other features that have been associated with enhancer priming are presence of pioneer TFs, hypermobile H3.3/H2A.Z nucleosomes, DNA 5mC hypomethylation, and hydroxylation (5hmC). C. Poised enhancer found in mouse and human ESCs: marked by H3K27me3 and associated with PRC2 (polycomb repressive complex 2); these enhancers are bound by TFs and coactivators and communicate with their target promoters. (Calo E. <i>et al.</i> , (2013)).	66
Figure 27 Chromatin enhancer signatures during cell differentiation (Spicuglia S. <i>et al.</i> , (2012)).	68

Figure 28 Schematic representation of traditional enhancer assay. Minimal promoter is indicated as red box; reporter gene is the green box: candidate enhancer element (test DNA) is indicated in blue box (Dailey L. <i>et al.</i> , (2015)).	69
Figure 29 a. Massively Parallel Reporter Assays (MPRA); b. Self-Transcribing Active Regulatory Region sequencing (STARR-seq)	70
Figure 30 CRISPR/Cas9 mediated gene-editing mechanism (created by Biorender).	72
Figure 31 Comparison between Cas9 nuclease (gene editing) and dCas9 nuclease-null (gene regulation). a. The <i>S. pyogenes</i> Cas9 endonuclease is targeted to specific DNA sequences by direct pairing of sgRNA with the target DNA. Two nuclease domains, RuvC1 and HNH, allow the cleavage of the target sequence. b. dCas9 protein contains mutations in its RuvC1 (D10A) and HNH (H841A) domains, which inactivate its nuclease function. dCas9 retains the ability to target specific sequences through the sgRNA and PAM. dCas9 binding downstream of the transcription start site (TSS) can block transcription elongation by blocking RNA polymerase II (Pol II) or the binding of important transcription factors (Txn) (Dominguez AA. <i>et al.</i> , (2016)).	73
Figure 32 Mechanism of CRISPR interference (CRISPRi) system (Qi LS. <i>et al.</i> , (2013)).	74
Figure 33 dCas9 fused to effector domains can efficiently activate or silence transcription (Gilbert LA. <i>et al.</i> , (2013)).	75
Figure 34 a. dCas9 fused with a VP64 activator. b. The SunTag activation system (left) consists of dCas9 fused to several tandem repeats of a short peptide sequence separated by linkers. The SunTag activator module (right) is an scFv, which specifically binds the SunTag peptide. The scFv is fused to sfGFP and VP64 (La Russa MF. <i>et al.</i> , (2015)).	76
Figure 35 a. The VPR activation system is dCas9 fused to VP64, p65, and Rta linked in tandem (La Russa MF. <i>et al.</i> , (2015)). b. RNA expression of target genes in HEK 293T cells transfected with gRNAs targeting the indicated genes along with the labelled dCas9-activator construct. Negative controls (Neg.) were transfected with the indicated guide RNAs alone. c. Analysis of mRNA expression levels of <i>NGN2</i> and <i>NEUROD1</i> in dCas9-Activator iPSC lines (Chavez A. <i>et al.</i> , (2015)).	77
Figure 36 a. SAM activation system (left): dCas9 is fused to VP64 and the sgRNA has been modified so as it contains two MS2 hairpins (green). An additional activator module (right) binds to an MS2 hairpin via the RNA-binding protein MCP. The MCP is fused to the activators p65 and HSF1 (La Russa MF. <i>et al.</i> , (2015)). b. A comparison of the activation efficiency between dCas9-VP64 alone (in yellow) and the SAM activator system (in green) in the activation of four different genes: HBG1, IL-1B, IL1R2, and ZFP42 (https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9).	78

Figure 37 a. dCas9 fused to KRAB, a transcriptional repressor (https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9). b. Stable suppression of CD71 and CXCR4 gene expression by dCas9 or dCas9-KRAB in HeLa cells (Gilbert LA. <i>et al.</i> , (2013)).	79
Figure 38 a. Schematic of dCas9-p300 system for epigenetic activation: dCas9-p300 acetylate target sites in the genome, resulting in transcriptional upregulation (https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9). b. mRNA expression level of <i>IL1RN</i> , <i>MYOD</i> and <i>OCT4</i> , using both dCas9 alone and dCas9 fused with different activators: VP64 activator, full-length p300 (dCas9 ^{FLp300}), HAT core domain of p300 (dCas9 ^{p300 Core}) and inactivated HAT core domain of p300 (dCas9 ^{p300 Core (D1399Y)}) (Hilton IB. <i>et al.</i> , (2015)).	80
Figure 39 a. Epigenetic gene repression by dCas9-LSD1 system: dCas9 is fused to the Lysine-specific histone demethylase 1 (LSD1). Demethylation of mono- and dimethyl-group on histone H3K4 lead to genes transcriptional repression (https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9). b. Left: Genomic organization of the targeted <i>Tbx3</i> locus. Right: Relative <i>Tbx3</i> expression in dCas9-repressor (LSD1 and KRAB) mESCs treated with sgRNAs specific to an unrelated control genomic region (Ctrl), the putative <i>Tbx3</i> distal enhancer (TDE) or the <i>Tbx3</i> promoter (TPP) (Kearns NA. <i>et al.</i> , (2015)).	81
Figure 40.a Representative pictures of differentiating EBs between d2 and d4;..	93
Figure 40.b Representative example of ECs-like in adhesion at day8 of differentiation.	93
Figure 41 Schematic illustration of both Cardiac (CM) / Endothelial (EC) differentiation strategies used for mESC.	94
Figure 42 Expression of pluripotency specific markers (<i>Nanog</i> , <i>Oct3/4</i> , <i>Rex1</i>), mesodermal (<i>Mesp1</i> , <i>Brachyury</i> , <i>Gata4</i> , <i>Pdgfra</i>) and endothelial genes (<i>VE-Cadherin</i> , <i>Pecam1</i> , <i>Eng</i> , <i>Kdr</i> , <i>Nos3</i>) during CPM/EC differentiation by RT-PCR. <i>Gapdh</i> is used as normalizer. Marker used 100bp.	95
Figure 43 Representative plot of the gating strategy used for immunophenotyping of cells during mES differentiation. The VE-Cadherin ⁺ (CD144 ⁺) subpopulation was identified at day 4-6-8 of differentiation by FACS using anti-CD144 antibody. Negative control is isotype control antibody-labeled differentiating cells.	96
Figure 44 <i>In vitro</i> tube formation assay of mESC ECs plated on Matrigel for 24h.	97
Figure 45 Alignment scores by STAR tool.	99
Figure 46 MA plot of all expressed genes dataset d2-d4. Log ratio (M) is log fold-change (y-axis); log average (A) is log mean of normalized counts (x-axis).	99
Figure 47 Intersection of DE genes at d4 vs d2 in Deseq2 and NOIseq methods.	101

Figure 48 Gene expression evaluation of representative marker genes of pluripotent stem cells, cardiopharyngeal mesoderm and endothelial cells. The results were indicated as FPKM average of two biological replicates.....	102
Figure 49 Gene ontology enrichment analysis for DEGs between d2 and d4. Terms related to biological processes, KEGG pathways, REACTOME pathways, Wiki pathways were considered. In red box there are some interesting and statistically significant (p-value) biological processes.....	105
Figure 50 ATAC-seq overview.	106
Figure 51 Quality check of aligned reads following removal of mitochondrial genome reads.....	107
Figure 52 Total accessible regions distribution in d2 (in blue) and d4 (in red) samples.....	108
Figure 53 Heat map of chromatin accessible regions distribution around TSS (from 1kb to 2kb) of d2 (2replicates) and d4 (2replicates).	109
Figure 54 Total accessible regions distribution in day2 and day4 mESC. The annotation pie shows accessible regions distribution around gene features (Promoter region= ± 3000 bp to TSS).....	110
Figure 55 Total accessible regions distribution in day2 and day4 mESC. The annotation pie shows accessible regions distribution around gene features (Promoter region= ± 1000 bp to TSS).....	110
Figure 56 Distribution of common_DARs in d4 vs d2. Annotation pie shows accessible regions distribution around gene features (TSS [± 3000] and [± 1000]).	112
Figure 57 Distance of the common_DARs to nearest TSS region (x-axis denotes the distance to TSS, from 0 bp to 50000bp; y-axis denotes the number of DARs).	112
Figure 58 Gene ontology enrichment analysis for genes annotated to common DARs between d2 and d4. Terms related to biological processes (GO:BP) were considered. In red there are some interesting and statistically significant (p-value) biological processes.	116
Figure 59 Volcano plot of intersection of DE and DARs genes in d4 vs d2 differentiation. In red were DARs and down-regulated DE genes (n.145); in dark blue were illustrated DARs and up-regulated DE genes (n.159); in orange were DARs and down-expressed genes, while in light blue DARs and up-expressed genes.....	117
Figure 60.a ATAC peaks and RNA-seq peaks coverage associated to <i>Kdr</i> . On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.	119

Figure 60.b ATAC peaks and RNA-seq peaks coverage associated to *Cdh5* (*VE-Cadherin*). On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 119

Figure 60.c ATAC peaks and RNA-seq peaks coverage associated to *CD34*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 120

Figure 60.d ATAC peaks and RNA-seq peaks coverage associated to *Eng*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 120

Figure 60.e ATAC peaks and RNA-seq peaks coverage associated to *Flt1*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 121

Figure 60.f ATAC peaks and RNA-seq peaks coverage associated to *Tall* (*Scl1*). On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 121

Figure 60.g ATAC peaks and RNA-seq peaks coverage associated to *Pecam1*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species. 122

Figure 60.h ATAC peaks and RNA-seq peaks coverage associated to *Notch1*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of

candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.	122
Figure 60.i ATAC peaks and RNA-seq peaks coverage associated to <i>Dusp5</i> . On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.	123
Figure 60.j ATAC peaks and RNA-seq peaks coverage associated to <i>Gata6</i> . On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.	123
Figure 61 Scheme of enhancer validation strategies. The first approach (left) is based on deletion of the putative enhancer by CRISPR-Cas9 technology; the second one (right) is using epigenetic reprogramming by fusing LSD1 demethylase with dCas9, which can erase methylation of histone H3 lysine 4 (H3K4me2) near the enhancer region to abrogate gene expression.	125
Figure 62 Experimental pipeline to generate mutant engineered mouse embryonic stem cell lines (mESC) using the CRISPR/Cas9 system (Mianné J. <i>et al.</i> , (2020).	127
Figure 63 Schematic example of T7 endonuclease 1 (T7E1) mismatch detection assays. Genomic DNA (in blue) from edited cells contains wild type and edited DNA (mutation in red). PCR amplification around the editing site generates wild type and edited PCR products (in black). Denaturing and reannealing of these products generates a fraction of heteroduplexes of mutant and wildtype PCR amplicons. Mismatches are cleaved by the T7E1 endonuclease. Running these PCR products on a gel resolves full length DNA and cleavage products. Gel shows untreated (-) and cells edited with Cas9 and crRNA:tracrRNA (+) (https://horizondiscovery.com/en/resources/featured-articles/proper-assessment-of-gene-editing-with-dna-mismatch-detection-assays).	128
Figure 64 Strategy to generate a knockout allele of <i>Tbx1</i> using CRISPR-Cas9 and homologous recombination. The bottom line indicates the DNA target: WT <i>Tbx1</i> -exon5 sequence (in light blue); in bold black the gRNA sequence; in bold red underlined the PAM sequence (CGG). The bottom line indicates the sequence of the recombinant allele (DNA donor, in bold red), inserted by homologous recombination. This includes a V5-Tag; a stop codon (TAG) and a diagnostic <i>EcoRI</i> digestion site (green line).	129
Figure 65 Evaluation of active Cas9 + gRNA efficacy using T7 Endonuclease 1 assay. Control cells (WT) show a single band corresponding to uncut amplicon.	

Amplicons from modified cells (<i>Tbx1</i> -KO) have 3 bands: 1 unmodified + 2 cleavage products.....	129
Figure 66.a PCR amplification of the targeted region from <i>Tbx1</i> homozygous clones 4D and 5H, and from WT digested with <i>EcoRI</i> . The restriction reaction produces two fragments, 161bp and 120bp, in mutant clones, while in WT is visible single PCR product at ~300bp.....	130
Figure 66.b Sequence of WT and homozygous clones (4D-5H). In bold black is indicated the WT sequence, that is replaced by the exogenous recombinant sequence (in bold red), in both samples 4D and 5H. <i>Tbx1</i> -exon5 sequence is in light blue.	130
Figure 67 <i>Tbx1</i> expression revealed by reverse transcription PCR. Left panel: PCR of samples collected at the differentiation stages indicated on WT mES cells. At day 4, the analysis was performed on total populations (T) and on FACS-purified subpopulations. The right panel shows the same experiment performed using the <i>Tbx1</i> ^{-/-} clone 5H. P, PDGFR α ; K, KDR (Cirino A. <i>et al.</i> , (2020)).....	131
Figure 68 Scheme of the steps of targeted <i>Pecam1</i> - enhancer deletion with CRISPR/Cas9. Red lines indicate the position of the two gRNAs used.....	132
Figure 69 Scheme of the CRISPR-Cas9 strategy used to delete the enhancer region, located in intron2 of <i>Pecam1</i> gene. On vertical axis there are the ATAC-seq peaks coverage of d2 (two replicates) and d4 (two replicates). Red lines indicate the gRNAs positions.....	132
Figure 70 T7 Endonuclease I digestion of genomic amplicons by WT cells and pool of <i>Pecam1</i> - Δ -enh. intron2 cells. Control cells (WT) have only a single band corresponding to uncut amplicon. Amplicons from modified cells (<i>Pecam1</i> - Δ -enh. intron2) transfected with active Cas9-gRNAs yield 4 bands: 1 unmodified + 3 cleavage products of predicted sizes.	133
Figure 71 Diagnostic PCR amplification of the targeted enhancer region from <i>Pecam1</i> - Δ -enh. intron2 clones and from WT cells. WT fragment = 569bp; Mutant fragment = ~250bp.	133
Figure 72 Sequence of WT <i>Pecam1</i> -enh. intron2 and mutated clones (5G-7G). In bold black is indicated the WT sequence, that is deleted in both samples 5G and 7G. In light blue are gRNAs sequences, while in red are indicated the PAM sequence (5'-NGG).	134
Figure 73 Scheme of the steps of targeted <i>Notch1</i> - enhancer deletion with CRISPR/Cas9. Red lines indicate the position of the two gRNAs used.....	135
Figure 74 Scheme of the CRISPR-Cas9 strategy used to delete the enhancer region, located in intron15 of <i>Notch1</i> gene. On vertical axis there are the ATAC-seq peaks coverage of d2 (two replicates) and d4 (two replicates). Red lines indicate the gRNAs position.	135
Figure 75 T7 Endonuclease I Assay to validate the efficacy of CRISPR/Cas9 + gRNAs strategy in mESC cells. Control cells (WT) have only a single band	

corresponding to uncut amplicon. Amplicons from modified cells (<i>Notch1</i> - Δ -enh. intron15) transfected with active Cas9-gRNAs yield 3 bands: 1 unmodified + 2 cleavage products of predicted sizes.	136
Figure 76 Diagnostic PCR amplification of the targeted enhancer region from <i>Notch1</i> - Δ -enh. intron15 clones and from WT cells. WT fragment = 664bp; Mutant fragment = ~280bp.	136
Figure 77 Sequencing of WT <i>Notch1</i> -enh. intron15 and mutated clones (7G-11B) to determine their exact sequences. In bold black is indicated the WT sequence, that is deleted in both samples 7G and 11B. In light blue are gRNAs sequences, while in red are indicated the PAM sequence (5'-NGG).	137
Figure 78 Full sequence map for p-dCas9-LSD1-Hygro (https://www.addgene.org/104406/sequences/).	138
Figure 79 Diagnostic PCR amplification of the p-dCas9-LSD1 expressing clones. WT sample is parental mESC (no targeted), CTR+ sample is parental mES cells transiently transfected with the plasmid p-dCas9-LSD1-Hygro. Red asterisk indicates the 3 mutant clones (A2; A8; B1).	139
Figure 80 Sequencing of plasmid p-dCas9LSD1 WT and mutated clones (A2, A8, B1) to determine the correct insertion of dCas9-LSD1 fusion protein. In violet is depicted the final part of dCas9 sequence, in bold black is indicated the 3xFLAG, in yellow is LSD1 sequence. All three clones show the correct insertion of the sequence.	140
Figure 81 Overview of the differentiation scheme. <i>Notch1</i> - Δ enh.intron15 clones (#7G and #11B) are differentiated in ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the <i>Notch1</i> expression.	141
Figure 82 Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in clone #11B <i>Notch1</i> - Δ enh.in15 during EC differentiation. <i>Notch1</i> is reduced in clone #11B sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).	142
Figure 83.a Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #11B <i>Notch1</i> - Δ enh.in15 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.	144
Figure 83.b Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #11B <i>Notch1</i> - Δ enh.in15 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the	

theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	145
Figure 83.c Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #11B <i>Notch1</i> - Δ enh.in15 at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	145
Figure 84 Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in clone #7G <i>Notch1</i> - Δ enh.in15 during EC differentiation. <i>Notch1</i> is reduced in clone #7G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).....	146
Figure 85.a Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #7G <i>Notch1</i> - Δ enh.in15 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	148
Figure 85.b Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #7G <i>Notch1</i> - Δ enh.in15 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	148
Figure 85.c Top: Quantitative real-time PCR analysis of <i>Notch1</i> mRNA expression level in WT and clone #7G <i>Notch1</i> - Δ enh.in15 at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	149
Figure 86 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce <i>Notch1</i> -enh.intron15 repression.....	150
Figure 87 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent <i>Notch1</i> _3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the <i>Notch1</i> expression.....	151
Figure 88 The dCas9-LSD1 transcriptional repression system, by chromatin inactivation. Specific gRNA target dCas9 fused with chromatin modifier LSD1 to	

a regulatory region. Lysine-specific demethylase 1A (LSD1) functions as transcriptional repressor: LSD1 binds to the CoREST or nucleosome remodeling and deacetylase repressive complex thus demethylating mono- and dimethyl-group on histone H3K4 and allowing genes transcriptional repression (Magliulo D. *et al.*, (2018)).

Figure 89 Quantitative real-time PCR analysis of *Notch1* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int15-*Notch1* during EC differentiation. *Notch1* is reduced in clone #B1p-dCas9LSD1+gEnh.int15-*Notch1*(red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and *Gapdh* expression as the normalizer. Values are the average of **four** (n=4) biological replicates \pm standard deviation (SD).

Figure 90.a Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-*Notch1* samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Figure 90.b Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-*Notch1* samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Figure 90.c Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-*Notch1* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Figure 91 Comparison of *Notch1*- Δ enh.in15 by CRISPR/Cas9 and *Notch1*-enh.in15 repression by dCas9LSD1.

Figure 92 Overview of the differentiation scheme. *Pecam1*- Δ enh.intron2 clones (#5G and #7G), as well as mESC WT cells, are differentiated in ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Pecam1* expression.

Figure 93 Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in clone #5G *Pecam1*- Δ enh.in2 during EC differentiation. *Pecam1* is reduced in clone #5G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression

level, evaluated using the $2^{-\Delta Ct}$ method, and <i>Gapdh</i> expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).....	157
Figure 94.a Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #5G <i>Pecam1</i> - Δ enh.in2 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	158
Figure 94.b Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #5G <i>Pecam1</i> - Δ enh.in2 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	159
Figure 94.c Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #5G <i>Pecam1</i> - Δ enh.in2 at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	159
Figure 95 Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in clone #7G <i>Pecam1</i> - Δ enh.in2 during EC differentiation. <i>Pecam1</i> is reduced in clone #7G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and <i>Gapdh</i> expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).....	160
Figure 96.a Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #7G <i>Pecam1</i> - Δ enh.in2 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	161
Figure 96.b Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #7G <i>Pecam1</i> - Δ enh.in2 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	162

Figure 96.c Top: Quantitative real-time PCR analysis of <i>Pecam1</i> mRNA expression level in WT and clone #7G <i>Pecam1</i> -Δ <i>enh.in2</i> at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	162
Figure 97 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce <i>Kdr</i> - <i>enh.intron10</i> repression.	164
Figure 98 Overview of the experimental plan. #B1 p-dCas9LSD1 cell clone is transfected with fluorescent <i>Kdr</i> _3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the <i>Kdr</i> expression.	165
Figure 99 Quantitative real-time PCR analysis for <i>Kdr</i> mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA- <i>Enh.int10-Kdr</i> during EC differentiation. <i>Kdr</i> is reduced in clone #B1p-dCas9LSD1+g <i>Enh.int10-Kdr</i> (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and <i>Gapdh</i> expression as the normalizer. Values are the average of four (n=4) biological replicates ± standard deviation (SD).....	165
Figure 100.a Top: Quantitative real-time PCR analysis of <i>Kdr</i> mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+g <i>Enh.int10-Kdr</i> samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	167
Figure 100.b Top: Quantitative real-time PCR analysis of <i>Kdr</i> mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+g <i>Enh.int10-Kdr</i> samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	167
Figure 100.c Top: Quantitative real-time PCR analysis of <i>Kdr</i> mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+g <i>Enh.int10-Kdr</i> samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.....	168
Figure 101 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce <i>VE-Cadherin</i> - <i>enh.int1</i> repression.	169

Figure 102 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent VE-Cadherin_3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the VE-Cadherin expression. 170

Figure 103 Quantitative real-time PCR analysis of *VE-Cadherin* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int1-*VE-Cadherin* during EC differentiation. *VE-Cadherin* is reduced in clone #B1p-dCas9LSD1+gEnh.int1-*VE-Cadherin* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=4) biological replicates \pm standard deviation (SD). ... 170

Figure 104.a Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 172

Figure 104.b Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 172

Figure 104.c Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 173

Figure 105 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce *Eng*-enh.int2 repression. 174

Figure 106 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent *Eng*_3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Eng* expression. 175

Figure 107 Quantitative real-time PCR analysis of *Eng* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int2-*Eng* during EC differentiation. *Eng* is reduced in clone #B1p-dCas9LSD1+ gRNA-Enh.int2-*Eng* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates

the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of five (n=4) biological replicates \pm standard deviation (SD)..... 176

Figure 108.a Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 177

Figure 108.b Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 178

Figure 108.c Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 178

Figure 109 Localization of the three gRNAs (in red) used to transfect clone #B1 p-Cas9LSD1 and to induce *Flt1*-enh.int10 repression..... 179

Figure 110 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent *Flt1*_3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Flt1* expression. 180

Figure 111 Quantitative real-time PCR analysis for *Flt1* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int10-*Flt1* during EC differentiation. *Flt1* is reduced in clone #B1p-dCas9LSD1+gEnh.int10- *Flt1* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of four (n=4) biological replicates \pm standard deviation (SD)..... 180

Figure 112.a Top: Quantitative real-time PCR analysis of *Flt1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Flt1* samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 182

Figure 112.b Top: Quantitative real-time PCR analysis of *Flt1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Flt1* samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 182

Figure 112.c Top: Quantitative real-time PCR analysis of *Flt1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Flt1* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear..... 183

Figure 113 Transcription factors' motifs enriched at peaks in day4 mESCs annotated to both expressed and DE genes from vascular-endothelial lineage cells expressed gene list. Here is used HOMER software for screening for Enrichment of Known Motifs. 186

LIST OF TABLES

Table 1 dCas9-epigenetic modifiers system.....	79
Table 2 gRNA sequences.	87
Table 3 Primer list used for RT-PCR and qRT-PCR.	91
Table 4 Total number of raw reads in d2-d4 differentiated mESCs.....	98
Table 5 Number of aligned reads of 2 replicates d2 and 2 replicates d4.....	98
Table 6 d4 vs d2 DE analysis using DESeq2 applying a threshold of P-adjust < 0.05 and < 0.01.....	100
Table 7 d4 vs d2 DE analysis using NOISeq (Posterior probability threshold 0.8 and 0.9).....	100
Table 8 Total number of ATAC-seq raw reads in d2-d4 differentiated mESCs.	106
Table 9 Number of aligned reads without chrM of 2 replicates d2 and 2 replicates d4.....	107
Table 10 Total peaks number (total accessible chromatin regions) in differentiating d2-d4 mESC. A consensus peak is a peak occurring in at least 2 replicates of the given replicates and it must pass the significance cutoff ($q < 0.001$).	108
Table 11 Total number of Differential Accessible Regions (DARs) between d2-d4 using DiffBind and DESeq2 tools.	111
Table 12 Annotation of common DARs in d4 vs d2. The DARs were annotated with two different promoter definition TSS [± 3000] and [± 1000].	111
Table 13 List of total 10 identified putative enhancers regions associated with endothelial-specific genes. The table contains: gene names (left column); position of the ATAC peaks increased at d4 (middle column); DE genes d4 vs d2 (right column) (gene expression variations during d2-d4 differentiation; + refers to genes whose expression increases at d4).....	118
Table 14 Two-way repeated measures ANOVA results, when compared WT and clone #11B <i>Notch1</i> - Δ enh.in15 dataset. p-value < 0.05 is considered significant.	143
Table 15 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #11B <i>Notch1</i> - Δ enh.in15 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	144
Table 16 Two-way repeated measures ANOVA results. Comparison between WT and #7G <i>Notch1</i> - Δ enh.in15 samples at d4, d6, d8, together, p-value < 0.05 is considered significant.....	147
Table 17 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #7G <i>Notch1</i> - Δ enh.in15 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	147

Table 18 Significant <i>Notch1</i> downregulation at day6 and day8 during EC differentiation in <i>Notch1</i> - Δ enh.intr15 cells (compared to WT). The asterisk (*) indicates the statistical significance of the data.	149
Table 19 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15- <i>Notch1</i> samples at d4, d6, d8, together. p-value < 0.05 is considered significant.	152
Table 20 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15- <i>Notch1</i> samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	153
Table 21 Two-way repeated measures ANOVA results, when compared WT and clone #5G <i>Pecam1</i> - Δ enh.in2 dataset. p-value < 0.05 is considered significant.	157
Table 22 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #5G <i>Pecam1</i> - Δ enh.in2 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	158
Table 23 Two-way repeated measures ANOVA results. Comparison between #7G <i>Pecam1</i> - Δ enh.in2 cells and #WT samples at d4, d6, d8, together. p-value < 0.05 is considered significant.....	161
Table 24 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between ##7G <i>Pecam1</i> - Δ enh.in2 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	163
Table 25 Significant <i>Pecam1</i> downregulation at day6 and day8 during EC differentiation in <i>Pecam1</i> - Δ enh.intr2 cells (compared to WT). The asterisk (*) indicates the statistical significance of the data.	163
Table 26 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10- <i>Kdr</i> samples at d4, d6, d8, together. p-value < 0.05 is considered significant.	166
Table 27 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10- <i>Kdr</i> samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	166
Table 28 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1- <i>VE-Cadherin</i> samples at d4, d6, d8, together. p-value < 0.05 is considered significant.....	171
Table 29 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1- <i>VE-Cadherin</i> samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	171

Table 30 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int2- <i>Eng</i> samples at d4, d6, d8, together. p-value < 0.05 is considered significant.	176
Table 31 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2- <i>Eng</i> samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	177
Table 32 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10- <i>Flt1</i> samples at d4, d6, d8, together. p-value < 0.05 is considered significant.	181
Table 33 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10- <i>Flt1</i> samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).	181
Table 34 Epigenetic enhancer decommissioning / repression of 5 regions by dCas9LSD1 induced significant enhancer-gene-related downregulation at later stages of EC differentiation (mainly at day8).	184

1. *Cardiogenesis in mouse development: signaling pathways from mesoderm induction to cardiogenic mesoderm progenitors*

The heart is the first organ to form during embryonic development and it has a vital role in the distribution of nutrients and oxygen.

Signaling from the surrounding microenvironment directs the transcriptional regulation of the developmental program of the heart, necessary for differentiation and proliferation (Brade T. *et al.*, (2013)). Heart development is a process that begins soon after embryo gastrulation at embryonic development day 6.5 (E6.5), when the mesoderm is formed between the ectoderm and the endoderm germ layer during ingression through the primitive streak (PS) (Aguilar-Sanchez C. *et al.*, (2018)). Later during development, early mesoderm-derived cardiac precursors undergo further lineage restriction and differentiate into progenitor pools that populate the First Heart Field (FHF) and Second Heart Field (SHF), respectively. In fact, after ingression through the PS, cardiogenic progenitor cells migrate to an anterior lateral position caudal to the headfolds and form the cardiac crescent. At this time in development (E 7.5), the first and second heart fields can be distinguished. FHF progenitor cells start to differentiate toward cardiomyocytes and smooth muscle cells when they are exposed to BMP and FGF cytokines as well as to inhibitors of the Wnt pathway. Consequentially, is induced the expression of key regulators of the cardiac lineage, like *Nkx2.5*, *Gata-4*, and *Tbx5*. Myocytic lineage commitment is associated with expression of contractile proteins including myosin light chain-2a (MLC2a) and sarcomeric myosin heavy chain (MHC) first in the cardiac crescent and then throughout the linear heart tube. Although these FHF progenitors already differentiate, Wnt/b-catenin, FGF, and endodermal Shh signaling keeps SHF progenitors in a proliferative state at the same time. SHF precursors are marked by the LIM-homeodomain transcription factor *Isl1* and its expression is absent in differentiated FHF derivatives (Kelly RG. *et al.*, (2012)). These SHF progenitors are defined by the molecular signature *Isl1*⁺/*Nkx2.5*⁺/*Flk1*⁺. Two subpopulations of SHF progenitors can be distinguished. One population *Isl1*⁺/*Flk1*⁺ which differentiates into endothelial cells and smooth muscle cells, whereas a second pool of *Isl1*⁺/*Nkx2.5*⁺ SHF precursors gives rise to smooth muscle cells and cardiomyocytes as well as contributing to the proepicardial lineages (*Wt1*⁺/*Tbx18*⁺ and *Scx*⁺/*Sema3D*⁺ populations), which later form cardiac fibroblasts (CF), smooth muscle cells (SMCs), endothelial cells (EC), and cardiomyocytes (CM). The fate of SHF progenitors is regulated by many different signaling pathways. Among them, FGF signaling within the SHF promotes progenitor cell proliferation; Shh-mediated signals from the endoderm and canonical Wnt signaling from the midline (neural tube) are important for the maintenance of a proliferative state and inhibition of

differentiation. On the contrary, BMPs secreted from lateral plate mesoderm as well as Notch and noncanonical Wnt signals promote cardiac differentiation of SHF progenitors. This highly complex network of spatiotemporal interactions of growth factors and transcriptional regulators (e.g., BMP, Wnt, FGF, Nkx2.5, and Gata4) governs both FHF and SHF development.

Mesoderm induction, a prerequisite for heart development, is evolutionarily conserved and regulated by numerous signaling pathways. The genetic pathways underlying cardiogenesis are complex and interconnected.

Key players are Nodal (including Activin) and bone morphogenetic protein (BMP) signals as well as Wnt and fibroblast growth factors (FGF) (Kimelman D. *et al.*, (2006)). Expression of the T-box transcription factor *Brachyury/T* (*Bry*), a direct target gene of Wnt/ β -catenin signaling, marks mesodermal cells. Commitment of nascent *Bry*⁺ mesodermal progenitors toward a cardiogenic fate requires inhibition of canonical Wnt/ β -catenin signaling and activation of noncanonical Wnt signaling (Gessert S. *et al.*, (2010)). Additionally, interactions between the mesodermal germ layer and the endoderm are necessary for the induction of a common cardiovascular progenitor population from *Bry*⁺ mesodermal precursors, which later form both FHF and SHF. Data from cell-tracing studies in mice showed that cardiac lineage arise from an intermediate mesodermal population that expresses vascular endothelial growth factor receptor 2 (*Vegfr2*, a.k.a *Flk-1* or *Kdr*) (Ema M. *et al.*, (2006)).

As depicted in Figure 1, during *in vitro* mESC differentiation and *in vivo* mouse development, BMP, Nodal, Wnt/ β -catenin and FGF pathways interact to induce mesoderm. *Bry*⁺ mesodermal precursors first differentiate through *Bry*⁺/*Flk-1*⁺ hemangioblasts toward endothelial and blood-cell lineages (around E5.5 during mouse development). After downregulation of Wnt/ β -catenin signaling and induction of noncanonical Wnt signals, a second wave of *Bry*⁺/*Flk-1*⁺ mesodermal progenitors appear. They downregulate *Bry* and activate the expression of mesoderm posterior 1 (*Mesp1*) gene, upon T-box transcription factor *Eomesodermin* action. Several studies have demonstrated that *Mesp1* is the earliest marker of cardiovascular development and are essential for cardiac mesoderm formation during embryonic development. It resides at the top of the cellular and transcriptional hierarchy of cardiovascular lineages during ESC differentiation (Bondue A. *et al.*, (2011)). *Mesp1*⁺ mesodermal progenitors contribute to paraxial mesoderm and skeletal muscle of the head as well as cardiac muscle.

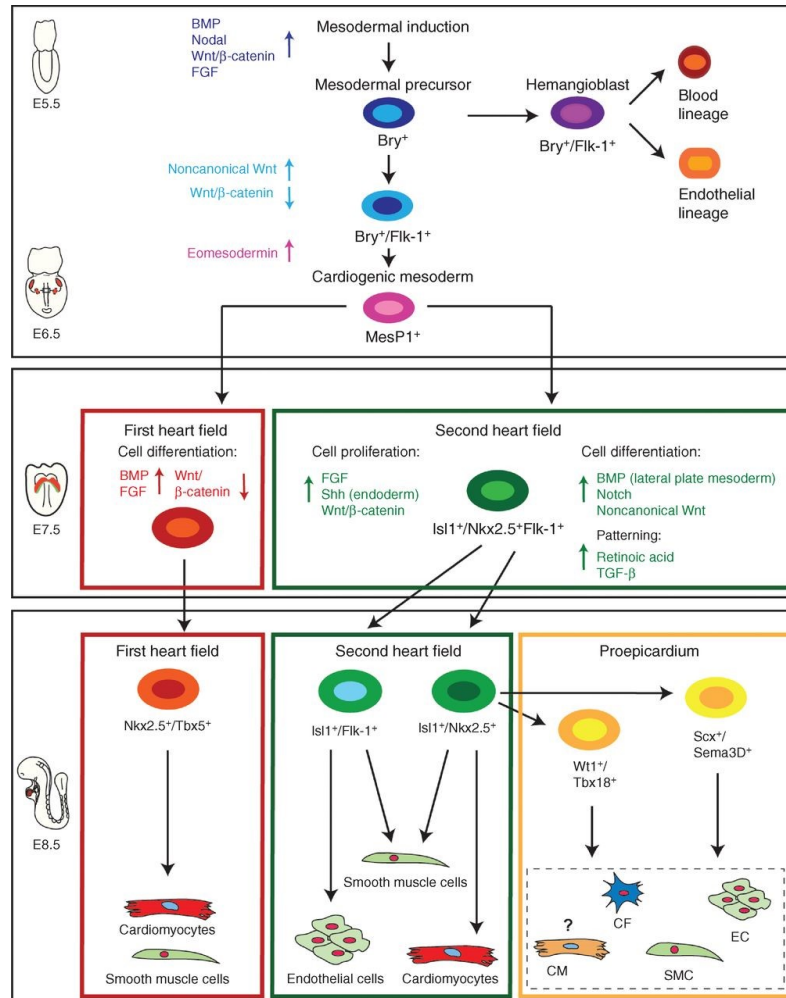


Figure 1 Signaling pathways for cardiac progenitor cells differentiation and their lineage specification (Brade T. *et al.*, (2013)).

Recently in 2018, Lescroart *et al.*, investigated the molecular and cellular basis of the earliest stages of cardiovascular lineage specification and diversification by single cell RNA-sequencing (scRNA-seq) analysis. They confirmed that mouse heart development arises from *Mesp1*-expressing cardiovascular progenitors (CPs) that are specified during gastrulation (Lescroart F. *et al.*, (2018)).

As described above, most of the *Mesp1* CPs differentiate into either CMs or ECs, suggesting that lineage segregation occurs early during gastrulation. *Mesp1*⁺ cells give rise to all heart cells, ECs of the aorta and brain, some muscles of the head and neck, as well as to few somitic derivatives and liver cells from its later expression (Devine WP. *et al.*, (2014)).

Since *Mesp1* marks the early CPs within the PS from embryonic day 6.25 (E6.25) to E7.25, Lescroart *et al.*, (2018) performed scRNA-seq of WT and *Mesp1* null CPs at the two stage of development.

Their scRNA-seq profiling showed that *Mesp1* CPs are molecularly distinct and make the continuum between epiblast and later mesodermal cells including hematopoietic progenitors. Single cell transcriptome analysis of *Mesp1*-deficient CPs indicated that *Mesp1* is required for the exit from the pluripotent state and the induction of the cardiovascular gene expression program. In fact, several well-known regulators of pluripotency (including *Nanog*; *Eras*), markers of the epiblast (such as *E-Cadherin/Cdh1*) were found up-regulated in single *Mesp1* KO cells (Figure 2). It is consistent with the defect of exiting the pluripotent epiblast stage.

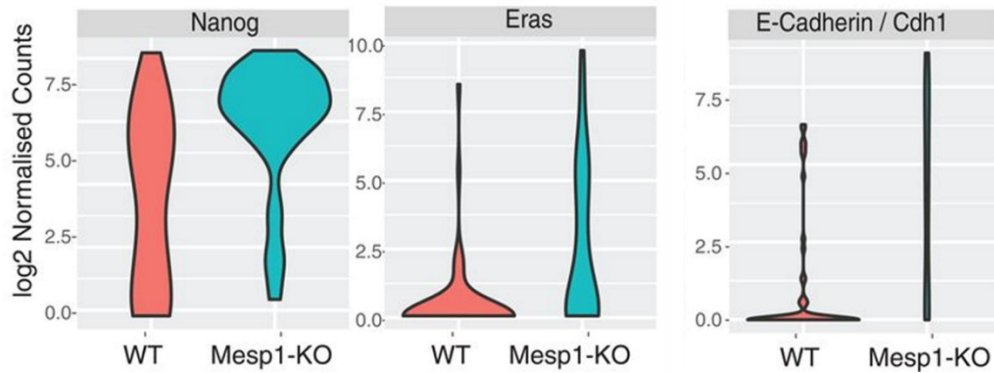


Figure 2 Violin plots showing the mean and variance difference between WT (red) and *Mesp1* KO (blue) cells of genes regulating pluripotency (*Nanog*, *Eras*) and epiblast (*Cdh1*) (Lescroart F. *et al.*, (2018)).

By contrast, the genes down-regulated in *Mesp1* KO cells were enriched for *Mesp1* target genes controlling epithelial-mesenchymal transition (EMT) (*Snail1*), and cardiovascular commitment (*Etv2*, *Hand1*, *Myl7*, *Gata4*, *Flk1*, and *Pdgfra*). *Pdgfra/Flk1*-expressing cells that mark *Mesp1* CPs in human and mouse ESC differentiation *in vitro* and during mouse gastrulation *in vivo* were also reduced in *Mesp1* KO cells, supporting the absence of CP specification (Figure 3).

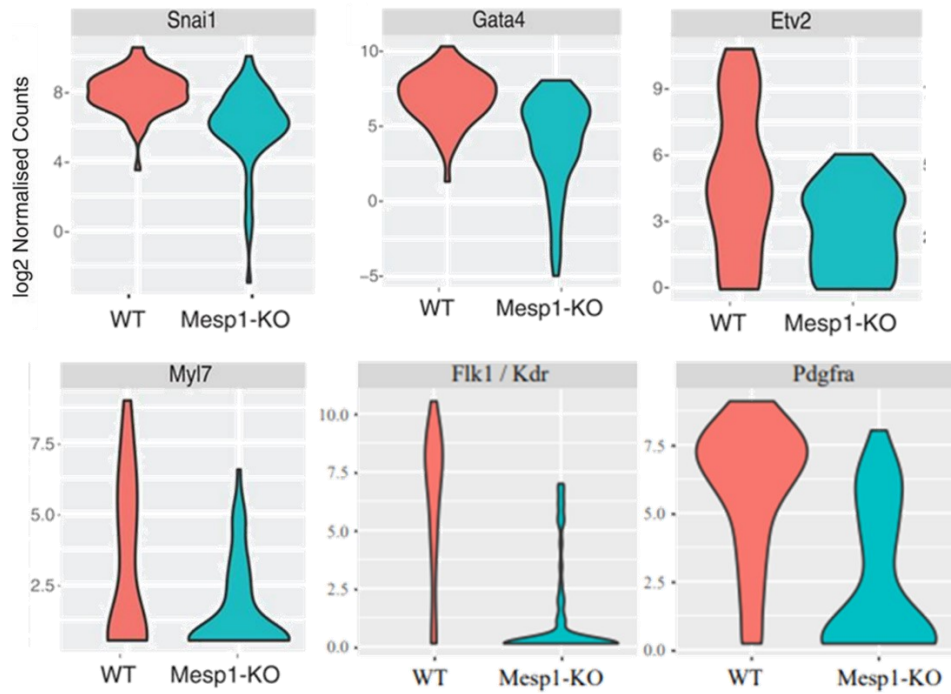


Figure 3 Violin plots showing the mean and variance of EMT genes (*Snai1*), cardiovascular markers (*Gata4*, *Etv2*, *Myl7* *Flk1/Kdr* and *Pdgfra*) expression in WT (red) and *Mesp1* KO (blue) (Lescroart F. *et al.*, (2018)).

Altogether, this single-cell profiling of early CPs shows that *Mesp1* CPs segregate rapidly from the epiblast into distinct cardiovascular lineages and it demonstrates that *Mesp1* CPs are also molecularly heterogeneous, as previously suggested by scRNA-seq during in vitro ESC differentiation (Chan SS. *et al.*, (2018)).

Moreover, SPRING analysis, which allows visualizing high-dimensional single-cell expression data, of WT *Mesp1*-expressing cells at E6.75 and E7.25 identified five distinct destination cell types (DCTs) protruding from a core of intermingled cells. Based on gene expression profiling, DCT1 (enriched in *Sox7*, *Etv2*, *Tall1*) and DCT2 cells (marked by *Hand1*, *Bmp4*, *Tnncl*, *Tbx3*, *Hand2*, *Tbx20*, *Gata4*, *Myl4*, and *Mef2C*) clustered respectively with EC and CM lineage. DCT3 (*Tbx1*+, *Foxc2*+) and DCT4 (*Wnt2b*+) correspond to *Mesp1* CPs committed to the anterior and posterior SHFs, respectively.

Otherwise, DCT5 expressed endoderm markers (such as *Sox17* and *Foxa2*) and may have no relation with cardiac development.

These findings suggest the existence of temporally and spatially distinct *Mesp1* subpopulations that likely correspond to CPs committed to the different

cardiovascular lineages and regions of the heart at the early stages of gastrulation. (Figure 4).

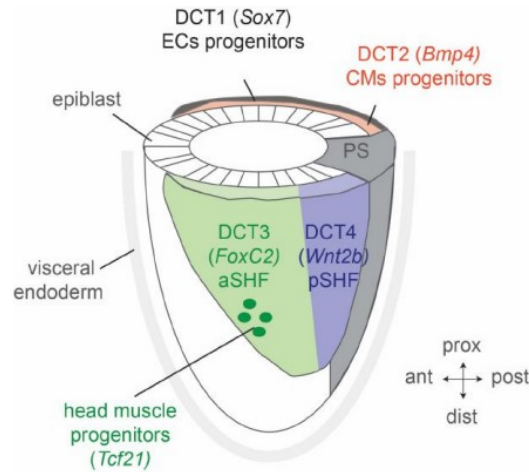


Figure 4 Scheme of the different *Mesp1* subpopulations (named DCTs 1-4) and their localization during gastrulation (Lescroart F. *et al.*, (2018)).

2. Specification of the endothelial cell lineage (EC) during mouse development

The development of functional vascular system is essential for embryonic development and adult life. Defects in endothelial cell development, vessel formation and function lead to embryonic lethality and are important in the pathogenesis of vascular diseases (Krüger-Genge A. *et al.*, (2019)). Endothelial cell, which line the inside of all blood and lymphatic vessels, have key roles in delivering oxygen and nutrients, regulating blood flow, modulating immune cell trafficking and maintaining tissue homeostasis.

The development of a mature vascular system is a complex process, which requires precise regulation and different types of cells: SCs, progenitor cells, vascular ECs, and mural cells. Mural cells are composed by VSMCs and pericytes and interact with ECs to form the complex network of capillaries, arterioles, arteries, and veins.

The vasculature consisting of arterial, venous, and lymphatic vessels form through two distinct processes during embryogenesis: (1) vasculogenesis, defined as de novo vessel formation induced by differentiation of mesodermal angioblasts, and (2) angiogenesis, the budding and branching of new vessels from pre-existing ones (Flamme I. *et al.*, (1997)). Although vasculogenesis is the major mechanism of formation of blood island vessels, dorsal aorta, endocardium, and vitelline vessels in the embryo, angiogenesis is the predominant means of vascularization of all

organs. In addition, vasculogenesis persists during vascular repair in the adult through differentiation of endothelial progenitor cells (Asahara T. *et al.*, (2011)). During embryonic development, the first sign of blood vessel formation occurs at the gastrulation stage, at mouse embryonic day E7.5 in the extraembryonic yolk sac blood island. In the extraembryonic yolk sac, mesodermal precursor cells aggregate to form blood island, the site of development of endothelial and primitive blood cells. Within the blood island, centrally located cells become primitive blood cells, whereas outer cells give rise to endothelial cells (ECs). Subsequently, ECs form the vascular primary plexus, which is then remodeled to form the yolk sac vasculature. In the embryo, mesodermal precursor cells differentiate into the vascular primary plexus and major vessels, aorta, and cardinal vein. After arterial and venous ECs are specified, the complex blood vasculature is formed via extensive remodeling. At embryonic day E9.5, a subset of ECs of the cardinal vein acquires lymphatic endothelial cell (LEC) fate and develops into lymphatic vessels (Figure 5).

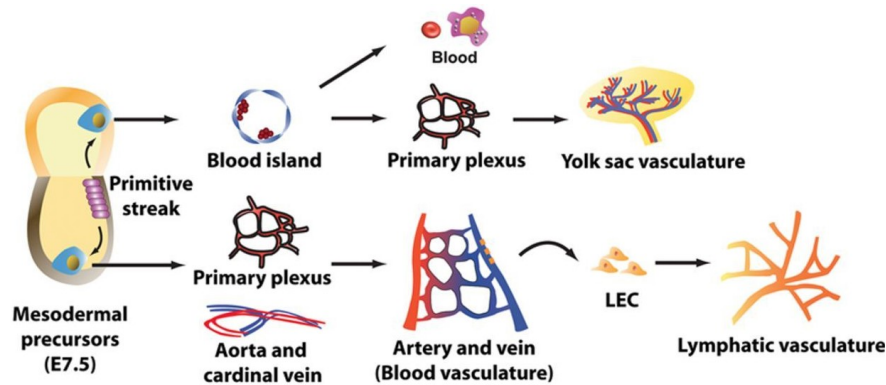


Figure 5 Vasculature formation during mouse embryo development (Park C. *et al.*, (2013)).

Within the embryo, endothelial precursor cells, also named angioblasts, appear at E7.5 and establish the vasculature of intra-embryonic regions including the dorsal aorta and vitelline vessels, and primary plexuses of lungs, spleen, and heart. The more complex phase of formation of the embryonic vascular networks occurs by angiogenesis during which newly formed vessels are stabilized through interactions of endothelial cells with each other via endothelial junction proteins and with recruited mural cells, the pericytes, and an ordered extracellular matrix. Endothelial precursor cells, mainly from mesodermal precursors, at early stage, already express markers such as CD31, CD34 and VEGFR-2 (Ema M. *et al.*, (2006)) (Figure 6).

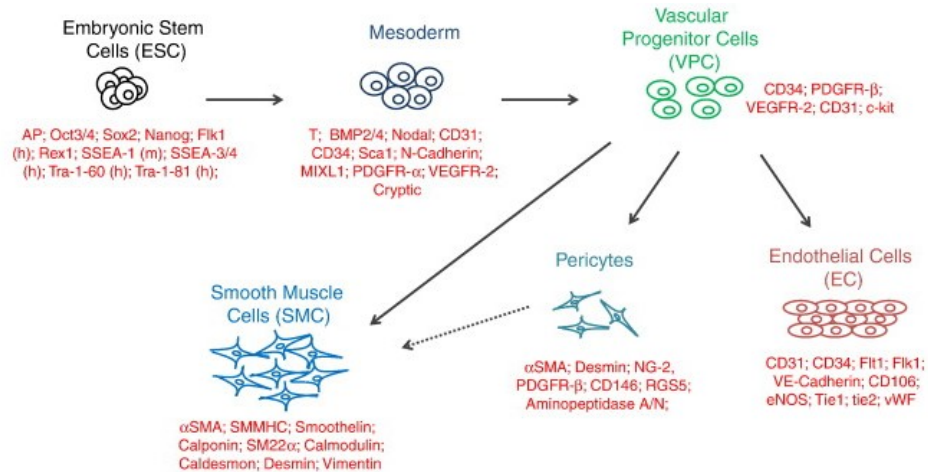


Figure 6 Map of the selective cell surface markers expression during mesoderm specification and vascular cell derivation from ESCs (Ema M. *et al.*, (2006)).

Several published studies have revealed that the endothelial cell-specific vascular endothelial growth factor (VEGF) and its two receptors, *Flt1* and *Flk1/Kdr* have been implicated in endothelial cell generation and formation of the embryonic vasculature. This is suggested by their colocalized expression during embryogenesis and by the impaired vessel formation in *Flk1* and *Flt1* deficient mouse embryos (Carmeliet P. *et al.*, (1996)).

Endothelial cells are derived from the mesoderm, which can be marked by the expression of *Brachyury/T box gene* (as previously described). Experiments in ES cells showed that *Brachyury*⁺ mesodermal cells were the first to appear and subsequently expressed *Flk1*, thus becoming double positive *Brachyury*⁺ *Flk1*⁺ cells during the differentiation *in vitro*. In developing mouse embryos, the expression of *Flk1* was first detected in the posterior portion of the primitive streak, followed by preferential expression in vascular endothelial cells of the yolk sac and embryonic vasculature including the endocardial tube. Deficiency of *Flk1* induced embryonic lethality due to lack of yolk sac blood island and blood vessel and endocardium formation (Shalaby F. *et al.*, (1995)). These findings define that *Flk1* is indispensable for the development of both blood and endothelial cell lineages. Further information indicated that *Flk1*⁺ cells obtained from differentiating ES cells were also shown to generate smooth muscle actin (SMA) cells (Yamashita, J. *et al.*, (2000)).

In addition, *Flk1*⁺ cells were also detected in skeletal muscles and cardiomyocyte of E10.5 embryos, besides the endothelial and blood cells (Figure 7).

Thus, the *Flk1*⁺ mesoderm likely represents a multipotent progenitor cell population in addition to blood and endothelial cell progenitors of the cardiovascular system.

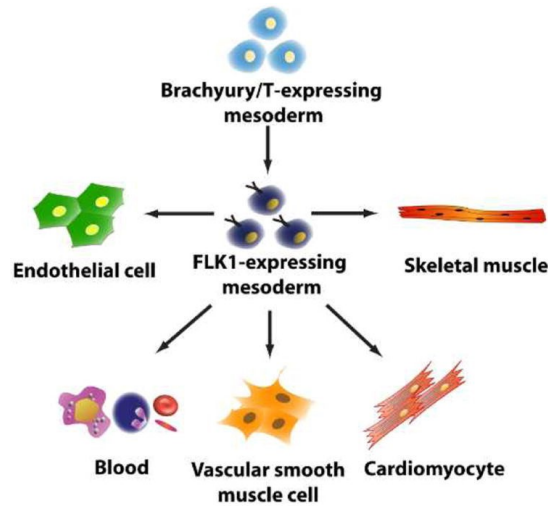


Figure 7 Development of blood cells, endothelial cells, cardiomyocytes, vascular smooth muscle cell lineages from *Flkl*-expressing mesoderm (Park C. *et al.*, (2013)).

3. *Transcriptional control and regulation of endothelial cell and vascular development*

Many transcription factors are known to play an important role in the activation and maintenance of endothelial gene expression. The specialization of all endothelial cell subtypes, from their mesodermal progenitors, requires extrinsic signals and intrinsic regulatory events.

3.1 *Tal1 (Scl1)*

SCL/TAL transcription factor belongs to the basic helix loop helix (bHLH) transcription factor family. Its early expression during embryogenesis is seen in yolk sac blood progenitors and endothelial cells. Disruption of *Tal1* in either mouse or zebrafish results in severe defects in the development of the vascular system. Blood vessels do form in absence of *Tal1*, suggesting that it may not be required for the initial induction of endothelial cells (Patterson LJ. *et al.*, (2005); Visvader JE. *et al.*, (1998)). Several endothelial-specific gene enhancers are activated by *Tal1* through essential E-box binding elements, such as *Flkl*, *Fli1*, *VE-Cadherin*, *Gata2*.

3.2 *GATA2 (GATA binding protein 2)*

The zinc finger transcription factor GATA2 is also implicated in regulation of both hematopoietic and endothelial genes. It is the most abundantly expressed GATA factors in endothelial cells and numerous endothelial enhancers contain GATA binding sites, like *Flkl*, *Tal1*, *Fli1*, *VE-cadherin*, *EPCR*, *ICAM2*, which are

bound directly by GATA2. The findings that several key endothelial genes contain GATA binding sites further suggest that GATA2 regulates vessel development through transcriptional activation of endothelial genes. In this regard, a recent report showed that GATA2 regulates endothelial specific gene expression and thus endothelial specificity through epigenetic modification. The authors found that *ENDOMUCIN*, an endothelial specific gene, contained preferential GATA2 binding sites in transcription start site (TSS) as well as -139 kb region of human dermal microvascular endothelial cells (HMVECs). An epigenetic experimental approach also demonstrated that both regions are in active chromatin state in HMVECs. Knockdown of GATA2 in HMVECs significantly reduced the expression of endothelial genes including *ENDOMUCIN* and *Kdr/VEGFR2* (Kanki Y. *et al.*, (2011)).

Moreover, mouse embryos deficient in *Gata2* died by E11.5 and exhibited anemia. *Gata2*^{-/-} ES cells and yolk sac cells generated significantly reduced number of multipotent progenitors (Tsai FY. *et al.*, (1994)). In addition, experiments in embryonic stem cells demonstrated the importance of GATA2 in the development of *Flk1*⁺ *Tali*⁺ hemangioblast-like cells and in the induction of endothelial-specific genes (Lugus JJ. *et al.*, (2007)). So, these studies support that GATA2 is an early regulator of hematopoietic and endothelial development, and that this transcription factor may be involved in the specification of hemangioblast progenitors from the mesoderm early in embryonic development.

Together, these results suggest that GATA2 plays an important role in mediating endothelial gene expression and the maintenance of endothelial cell fate.

3.3 ETS Transcription Factors in regulation of endothelial gene expression

The ETS transcription factors display a broad range of expression patterns and activities in developing mouse embryos as well as in adults. However, some of these factors show a preferential expression in endothelial cells and vessels during development. Numerous endothelial enhancers and promoters contain multiple, essential ETS binding sites, and ETS motifs are strongly associated with endothelial genes throughout the human genome (De Val S. *et al.*, (2009)). Within the Ets family, *Ets1*, *Elf1*, *Fli1*, *Tel*, and *Erg* each have well characterized roles in endothelial gene expression, and each bind to the enhancers and activate the expression of numerous endothelial genes.

Expression of the ETS transcription factors, *Ets1* and *Fli1*, can be detected in the yolk sac blood islands in the early-stage embryos and their expression is maintained in developing vessels. Whereas the ETS transcription factor, *Erg1*, is highly expressed in mesodermal lineages including endothelial cells, another ETS transcription factor *Etv6* shows ubiquitous expression pattern throughout embryogenesis and in adults.

Knockdown of either *Ets1* or *Erg* expression in endothelial cells in culture results in decreased endothelial cell migration and tube formation (Birdsey GM. *et al.*, (2008)).

The Ets transcription factor *Fli1* is expressed very early in cells of the hematopoietic and endothelial lineages in mice and zebrafish. *Fli1*^{-/-} embryos died around E11.5 with extensive hemorrhaging, indicating its dispensable role in endothelial cell specification (Spyropoulos DD. *et al.*, (2000)).

Interestingly, germline deletion or mutation of the majority of individual Ets genes in either mouse or zebrafish model systems has resulted in little or no vascular phenotype or has caused defects only in later vascular remodeling, while vasculogenesis remained largely intact (Pham VN. *et al.*, (2006)). An explanation may be due to redundancy among Ets factors in endothelial development. The exception to this apparent redundancy among Ets factors in endothelial development is observed when the function of the Ets protein *Etv2* (ER71) is removed in mice.

3.3.1 Role of *Etv2* in early vascular development

Several studies have shown that the *Etv2* is essential for the development of endothelial and blood lineages in the mouse. *Etv2* expression is present at the very early stages of vascular development in the mouse, with expression detected in the blood islands of the yolk sac and the earliest vessels in the embryo. Then, its expression begins decreasing within endothelial cell populations by E9.5 and is essentially extinguished in those lineages by E10.5, suggesting an involvement of this transcription factor in early vascular development. *Etv2* null mice have severe defects in vasculogenesis and hematopoiesis. *Etv2*^{-/-} embryos die at midgestation and lack any detectable embryonic vessels, blood islands in the yolk sac, or endothelial progenitors. Expression of early vascular markers, such as *Flk1*, *Pecam1*, and *Tie2* is almost completely abolished in the absence of *Etv2*, and endothelial cells are apparently not specified if *Etv2* is not present. In fact, *Etv2* is also a potent inducer of *Flk1*⁺ mesoderm in embryonic stem cells. Furthermore, *Etv2* is a potent activator of several early endothelial genes, including *Flk1*, *Tall*, *Mef2c*, *Pecam1*, and *Tie2*, and has been demonstrated to activate these genes through direct promoter or enhancer binding. Thus, inactivation of *Etv2* causes profound impairment of vasculogenesis, suggesting a central role for this factor in endothelial specification (Ferdous A. *et al.*, (2009)).

3.4 Forkhead Proteins are important regulators of endothelial transcription

Among the forkhead transcription factors, FoxC, FoxF, FoxH and FoxO are implicated in vascular development and endothelial transcription (Papanicolaou KN. *et al.*, (2008)). Targeted disruption of *FoxO1* in mice causes vascular remodeling defects and midgestational lethality, suggesting that FoxO1 is required

for vascular development. In fact, in mouse embryos, *FoxO1* is highly expressed in the developing vessels and *FoxO1*^{-/-} mice embryos failed to survive beyond E10.5 because of defective vascular development in the yolk sac and embryo (Furuyama T. *et al.*, (2004)).

The FoxC family of Forkhead proteins is essential for vascular development as *FoxC1*^{-/-}; *FoxC2*^{-/-} mice have severe vascular defects. Additionally, *FoxC1* and *FoxC2* have also been found to regulate arterio-venous specification and lymphatic vessel differentiation and may be important downstream effectors of *Notch* signaling (Hayashi H. *et al.*, (2008)). Promoters of multiple endothelial genes *Flk1*, *VEcadherin*, *Pecam1*, *Tie2* and *Scl*, contain evolutionarily conserved FOX:ETS binding motifs (De Val S. *et al.*, (2009)).

FoxF1 and *FoxH1* are also involved in endothelial gene regulation. Inactivation of *FoxF1* in mice results in a severe vascular phenotype and embryonic lethality. *FoxF1* is not expressed within endothelial cells of the differentiated embryonic vasculature, but it is expressed earlier in the splanchnic mesoderm prior to endothelial cell specification and may regulate BMP signaling, which is essential for vascular development (Astorga J. *et al.*, (2007)).

3.5 HEY1/HEY2 (Hairy/enhancer-of-split related with YRPW motif protein1/2)

HEY1 and HEY2, members of hairy and enhancer of split-related family of bHLH transcription factors, are direct transcriptional targets of the NOTCH pathway. It controls endothelial cell specification (arterial vs. venous endothelial cells) and plays an important role in endothelial cell sprouting (Swift MR. *et al.*, (2009)). *Hey1*^{-/-} and *Hey2*^{-/-} mouse embryos developed defective vessels and died at E9.5. In these mice, vasculogenesis was normal, but vessel remodeling in the yolk sac and placenta was impaired and some mice showed poorly developed dorsal aorta and cardinal veins (Fischer A. *et al.*, (2004)). In agreement with these findings, deficiency of *Rbpj* (recombination signal binding protein for immunoglobulin kappa J region) in mice, an important transcription factor of the NOTCH pathway, also induced defective arterial vessel formation. These studies showed that the NOTCH pathway in endothelial cells controls the development of arteries through RBPJ-HEY1/2 signaling.

3.6 SOX F group transcription factors: SOX7, SOX17 and SOX18

Among 20 different members of the SOX factors, the SOX F group (*Sox7*, *Sox17* and *Sox18*) has been found to play a pivotal role in cardio-vascular development. Expression of *Sox18* and *Sox7*, but not *Sox17*, was detected between

E7.5 and E8.5 in endothelial cells of the yolk sac vasculature and the dorsal aorta, and *Sox17* was expressed in endothelial cells of the dorsal aorta at E9.5.

According to Sakamoto Y. *et al.*, (2007), *Sox17* single-null embryos showed aberrant heart looping, enlarged cardinal vein and mild defects in anterior dorsal aorta formation; while the *Sox17/Sox18* double-null embryos showed more severe defects in formation of anterior dorsal aorta and head/cervical microvasculature, and in some cases, aberrant differentiation of endocardial cells and defective fusion of the endocardial tube. *Sox18* was also expressed in endothelial cells of the developing lymphatic vessels. Indeed, *Sox18*^{-/-} mouse embryos showed lethality at E14.5 and developed no *PROX1*⁺ (Prospero homeobox transcription factor 1) lymphatic endothelial cells, indicating a critical role of *Sox18* in lymphatic endothelial cell specification. Finally, taken together, these studies suggest that the SOX F group transcription factors function as an important regulator for endothelial cell specification; arterio-venous specification, and venous-lymphatic specification.

In Figure 8 is illustrated the different steps within endothelial cell development from mesodermal progenitors and hemangioblasts to differentiated arterial, venous, and lymphatic endothelium and the various transcription factors associated with their development.

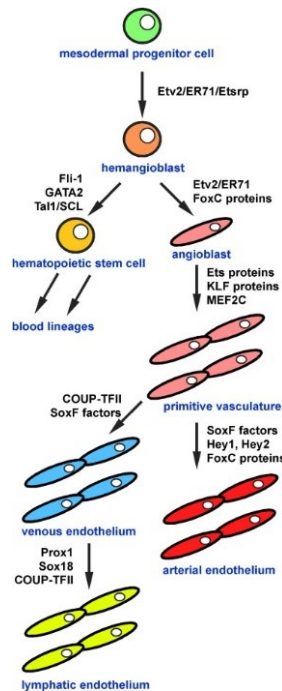


Figure 8 Different stages in endothelial development, regulated by distinct group of transcription factors. Based on this scheme, *Fli-1*, *GATA2*, and *Tal1* control differentiation of hematopoietic cells from hemangioblasts, while *Etv2* and *FoxC* proteins control the differentiation of endothelial cells from that progenitor population and that *Etv2* likely sits at the top of this transcriptional cascade (De Val S. *et al.*, (2009)).

4. *Signaling Pathways in Regulating Endothelial Transcription Factors and Vascular Development*

4.1 *VEGF signaling*

VEGF-activated signaling is the major pathway regulating multiple aspects of endothelial cell function including survival, proliferation, and vessel permeability (Bates DO. *et al.*, (2010)). Following the binding of VEGF, its receptor, *Flk1/Kdr*, transmit signals through several downstream molecules MAPK-ERK, p38-MAPK, phospholipase C, and phosphatidyl inositol 3-kinase (PI3K)/Akt/protein kinase B to regulate endothelial function. VEGF signaling plays a critical role in vessel development during embryogenesis. *Vegf*^{+/-} mouse embryos died due to defects in endothelial and hematopoietic cell development (Carmeliet P. *et al.*, (1996)). Further, VEGF signaling has a role in arteriovenous specification. It induced the expression of arterial markers *EphrinB2*, *Nrp1*, and *Gja5* in primary *EphrinB2*⁻ endothelial cells isolated from E10.5 mouse embryos. In addition, overexpression of VEGF in cardiomyocytes led to significant increase in number of cells expressing the arterial marker *EphrinB2* with decreased number of *EphB4*⁺ venous endothelial cells. Studies in literature suggest also that transcription factors *FOXC1/FOXC2* interact with the VEGF pathway components to promote arterial specification of endothelial cells through the NOTCH signaling pathway.

The receptor tyrosine kinase KDR (also known as VEGFR2/Flk1) and its high-affinity ligand vascular endothelial growth factor (VEGF) play a pivotal role in endothelial development, during embryonic vascular growth in the mouse. Expression of the *Kdr* gene and its ligand VEGF is restricted to endothelial cells and their embryonic precursors. In addition, during embryonic mouse development, *Kdr* also marks an early multipotent cardiovascular progenitor, which give rise to cardiomyocyte, endothelial and vascular smooth vascular lineage (Kattman *et al.*, (2011)). Highest levels of *Kdr* expression were observed during embryonic vasculogenesis and angiogenesis, as well as during pathological processes associated with neovascularization. In developing mouse embryos, the expression of *Kdr* was first detected in mesodermal yolk-sac blood-island progenitors as early as 7.0 days postcoitum, indicating that it marked the common embryonic endothelial and haematopoietic precursor, the haemangioblast, and thus is also involved in early haematopoiesis (as depicted in Figure 9) (Park C. *et al.*, (2013)). Some studies have demonstrated that *Kdr*^{-/-} mouse embryos die in utero between 8.5 and 9.5 days post-coitum, because of an early defect in the development of haematopoietic and endothelial cells. Yolk-sac blood islands were absent at 7.5 days, organized blood vessels were not observed in the embryo or yolk sac at any stage, and haematopoietic progenitors were severely reduced.

These results indicated that *Kdr* is essential for yolk-sac blood-island formation and vasculogenesis in the mouse embryo, as well as it is the most important effector in angiogenesis (Shalaby F. *et al.*, (1995)).

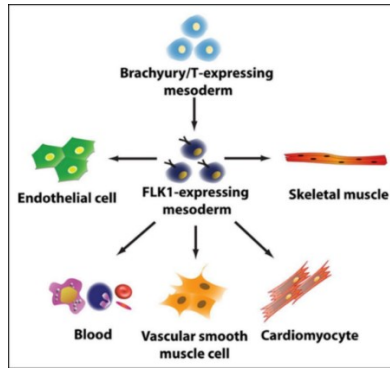


Figure 9 Developmental of FLK1+ mesoderm. FLK1+ mesoderm originates from Brachyury/T-expressing cells in the primitive streak of the developing embryo and subsequently gives rise to not only endothelial and blood cells but also other mesodermal lineage cells, including vascular smooth muscle, cardiomyocyte, and skeletal muscle cells (Park C. *et al.*, (2013))

The VEGFR2 signaling pathways are induced by its binding to VEGF-C, VEGF-E and VEGF-A. The binding of ligands to VEGFR2 results in the activation of different pathways such as SCR, phospholipase-C γ (PLC- γ)/Protein kinase C (PKC), Phosphoinositide-3-kinase (PI3K), p38-Mitogen-activated protein kinase (P38MAPK) and Focal Adhesion Kinase (FAK). These different pathways control cell shape, cell adhesion and permeability, proliferation, survival, vasodilatation, and migration of endothelial cells (Figure 10).

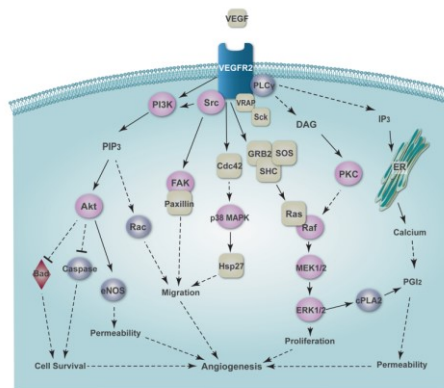


Figure 10 Schematic representation of VEGFR2 signaling pathways. VEGFR2 (<https://www.tocris.com/pathways/vegf-signaling-pathway>).

Flt-1 (fms-like tyrosine kinase-1), also known as vascular endothelial growth factor receptor 1 (Vegfr-1) encodes a high-affinity tyrosine kinase receptor for VEGF and is almost exclusively expressed in vascular endothelial cells, both in fetal and adult mouse tissues (161). Moreover, the *Flt-1* mRNA is also detected in both proliferating and quiescent endothelial cells, suggesting a role for *Flt-1* in the maintenance of endothelial cell (Peters KG. *et al.*, (1993)).

In 2002, Kearney JB *et al.*, reported that *Flt-1* mutant ES cell cultures had vascular overgrowth that was caused primarily by aberrant endothelial cell division; also *Flt-1* null mice died in the embryonic stage at E-8.5-9.0 due to an overgrowth of immature endothelial cells and disorganization of blood vessels, not due to a poor vascularization. Endothelial cells developed but failed to organize in vascular channels. These results strongly suggested that VEGFR-1 acts early in vascular development to modulate vessel formation by affecting the rate of cell division in embryonic endothelial cells and their precursors. It plays a negative role in angiogenesis by suppressing pro-angiogenic signals in the embryo to establish a critical balance essential for physiological vascular formation.

Flt-1 signalling pathway regulates normal endothelial cell-cell or cell-matrix interactions during vascular development. VEGFR-1 binds not only VEGF-A but also PlGF (placenta growth factor) and VEGF-B. It transduces several downstream signals including cell migration, vascular permeability and cell survival. The 1169-tyrosine is phosphorylated and activates PKC γ -PKC pathway towards mild activation of MAP-kinase (MAPK) and DNA synthesis. PI3-kinase (PI3K) and Akt are also activated upon the stimulation of VEGFR-1 (Figure 11).

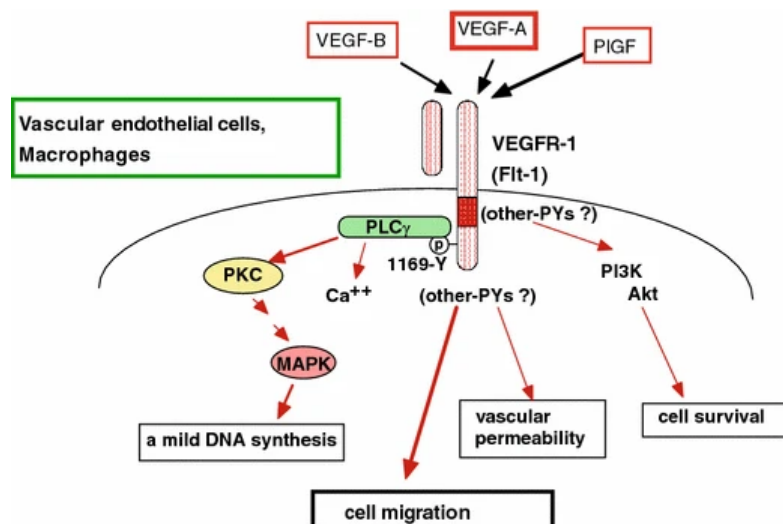


Figure 11 Signal transduction pathways from VEGFR-1 in vascular endothelial cells (Shibuya M. *et al.*, (2006))

4.2 NOTCH signaling

The NOTCH pathway is critical for arterial specification of embryonic vasculature. Four NOTCH receptors (1 through 4) and five ligands (jagged1, 2, Delta like ligand (Dll)1, 3, 4) have been identified in mammals. Binding between the ligand and the receptor induces proteolytic cleavage of NOTCH receptor, resulting in generation of intracellular form of the receptor (NICD, NOTCH intracellular domain) that translocates into the nucleus to induce its downstream targets HEY1 and HEY2 (Gridley T. *et al.*, (2010)). *Dll4*^{+/-} heterozygous mice exhibited severe remodeling defects in yolk sac vessels and smaller dorsal aorta consistent with the expression of components of NOTCH signaling, NOTCH1, NOTCH4 and Jagged1, Jagged2 and Dll4 selectively in arterial endothelial cells. Also, these embryos developed abnormal arteriovenous vessels due to fusion between the dorsal aorta and common cardinal vein. Moreover, *Dll4*^{-/-} mouse embryos completely lost arterial identity (Duarte A. *et al.*, (2004)). Mice deficient in both *Notch1* and *Notch4* died in utero with a severe vessel remodeling defects in both yolk sac and embryo (Krebs LT. *et al.*, (2000)). Thus, NOTCH, which is activated by VEGF, has an essential role in mediating arterial specification. In fact, it was shown that VEGF induced the expression of *Notch1* and *Dll4* through PI3K/AKT pathway in cultured endothelial cells. The VEGF-mediated NOTCH activation was specifically seen in arterial endothelial cells (as opposed to venous endothelial cells) *in vitro* (Liu ZJ. *et al.*, (2003)).

Another study has shown that the crosstalk between both pathways could be mediated by endothelial transcription factors, *FoxC1/FoxC2*. As mentioned above, deletion of *FoxC1/FoxC2* led to arterial defects in developing mouse embryo. In vitro analysis revealed that *FoxC1/FoxC2* upregulated the expression of the arterial endothelial markers *Notch1*, *Notch4*, *Dll4*, *Hey2* and *EphrinB2* through direct transcriptional activation. Furthermore, FOXC1/FOXC2-mediated promoter activation of *Dll4* and *Hey2* was augmented by VEGF treatment, and it was impaired by inhibiting PI3K. These interesting results suggest that the VEGF and NOTCH pathway promote arterial endothelial cell specification through FOX transcription factors.

Moreover, the Notch signaling pathway plays a critical role in cell fate determination during development and postnatally in continuously renewing tissues, such as the endothelium, the epithelium and in stem cells pool. Consequently, it can control the expression of a group of target genes involved in Endothelial-Mesenchymal Transition (EndoMT), angiogenesis and Cancer Stem Cell (CSC) biology, as shown in Figure 12 (Akil A. *et al.*, (2021)).

The Notch pathway plays also pivotal roles in the cardiovascular system, both during development and postnatal life. Some evidence show that the dysregulation of the Notch pathway is involved in the pathophysiology of cardiovascular diseases (CVDs) (Aquila G. *et al.*, (2019)).

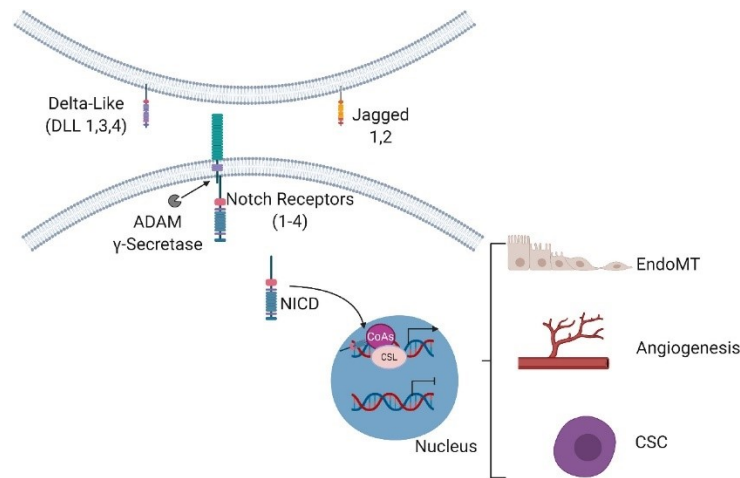


Figure 12 Overview of the Notch signaling pathway and its role in regulation of target genes involved in biological processes including EndoMT, angiogenesis and CSC biology (Akil A. *et al.*, (2021)).

Expression of Notch and ligand in vascular endothelium and defects in vascular phenotypes of targeted mutants in the Notch pathway have suggested a critical role for Notch signaling in vasculogenesis and angiogenesis. Loss of global Notch1 function results in early embryonic lethality, as the endothelium is the primary target tissue affected by Notch1 signaling (Limbourg FP. *et al.*, (2005)). General knockout strategies in the mouse have been used to show that the Notch1 signaling pathway regulates vascular morphogenesis and angiogenesis (Krebs LT. *et al.*, (2000)). Moreover, crosstalk between VEGF and Notch signaling is fundamental for angiogenic process, being both involved in the specification of the tip and stalk cell phenotype, as well as to sprout formation, vascular EC maintenance, and the establishment of EC heterogeneity (Blanco R. *et al.*, (2013)).

4.3 Signaling via Bone Morphogenetic Proteins (BMP Signaling)

Bone morphogenetic proteins (BMPs) belong to transforming growth factor (TGF) β superfamily that regulates a multitude of biological processes including embryonic vessel development (Cai J. *et al.*, (2012)). Upon binding to BMPs, the type II receptor, which is a constitutively active kinase, activates the type I receptor by phosphorylating specific serine and threonine residues. The activated type I receptor in turn phosphorylates SMADs (SMAD1, 5, and 8) to transmit BMP signaling. Subsequently, the phosphorylated SMADs interact with SMAD4, a common SMAD, and translocate into the nucleus to induce the expression of genes such as ID. Mice deficient in both *Id1* and *Id3*, the downstream

BMP/SMAD targets, showed vascular defects in the developing brain, while *Bmp4* deficient mice died without posterior mesodermal differentiation. Germ line deletion of *Alk-3*, a type IA BMP receptor, failed to survive up to E9.5 stage with defects in mesoderm formation (Mishina Y. *et al.*, (1995)).

Park *et al.*, (2004) demonstrated that BMP4 via SMAD1/5 signaling induced the generation of *Flk1*⁺ cells from mouse ES cells in serum-free differentiation condition. Blocking BMP4 activity by its antagonist, Noggin, reduced *Flk1*⁺ cell generation. Moreover, BMP4-mediated *Flk1*⁺ cell development was mediated by the transcription factor, ER71/*Etv2*. In addition, GATA2 together with BMP4 was shown to also promote mouse ES cell differentiation to *Flk1*⁺ cells.

4.4 WNT Signaling

WNT signaling is critical for embryogenesis and disease development (Nusse R. *et al.*, (2005)). WNT signaling is typically categorized into two pathways: the canonical (β -catenin)- and non-canonical (non- β -catenin-mediated) pathways. In summary, the non-canonical pathways involve all WNT pathways that do not lead to the stabilization of β -catenin and plays a role in processes including cell polarization, cell fate, inflammatory response, and cell migration. On the contrary, the canonical β -catenin-mediated WNT signaling pathway leads to the intracellular stabilization of β -catenin, resulting in its translocation to the nucleus and the transcription of numerous genes involved in cell proliferation, differentiation, tissue expansion, cell fate, and many more. The canonical WNT/ β -catenin pathway has been the best studied and has a crucial role in vascular development. In the absence of WNT proteins, β -catenin is phosphorylated by glycogen synthase kinase 3 β (GSK-3 β), and the “marked” β -catenin is ubiquitinated for degradation by proteasomes. Binding of WNT proteins to their receptor, Frizzled/Lrp (Low-density lipoprotein receptor), activates Dishevelled (Dvl) which inhibits GSK3 β and thus releases β -catenin from the degradation pathway, allowing the translocation of β -catenin to the nucleus.

Lindsley RC. *et al.*, (2006) demonstrate a requirement for WNT signaling during the earliest steps of mesendodermal differentiation of ES cells. Specifically, during ES cell differentiation, canonical WNT signaling is required for the expression of genes associated with the primitive streak and gastrulation in vivo, including *Brachyury*, *Mixl1* and *Evx1*. Furthermore, they showed that inhibition of WNT signaling abrogated the functional competence of differentiating ES cells, reflected by their failure to generate *Flk1*⁺ mesodermal precursors and subsequent mature mesodermal lineages.

Other studies in developing mouse embryos showed an important role of β -catenin, which can be stabilized downstream of WNT signaling, in vessel development. Inactivation of β -catenin in *Tie2*⁺ cells resulted in embryonic lethality with vascular remodeling defects and hemorrhages. The mutant embryos

also displayed defects in endocardial cushion and cardiac valve formation (Cattellino A. *et al.*, (2003)).

5. *In vitro* model to induce multipotent cardiac progenitors to differentiate into endothelial cells (ECs), starting from mESC

5.1 Undifferentiated ES cells

The discovery of mouse embryonic stem (ES) cells >25 years ago was a major advance in biology and experimental medicine, as ES cells represented an *in vitro* model of early mammalian development and a suitable source of differentiated cell types for cell replacement therapy.

Embryonic stem (ES) cells are pluripotent cells derived from the inner cell mass of blastocyst-stage embryos, which are able to divide without differentiating for a prolonged period in culture (Evans MJ. *et al.*, (1981)).

Their potential use in modern biology and medicine derives from two unique characteristics that distinguish them from all other organ-specific stem cells. First, they can be maintained and expanded as undifferentiated population cells for extended periods of time, possibly indefinitely, in culture. Moreover, ES cells can retain normal karyotypes following extensive passaging in culture. Second, they are pluripotent, possessing the capacity to generate every cell type in the body.

Efficient differentiation of embryonic stem cells (ESCs) to a variety of lineages requires stepwise approaches replicating the key commitment stages found during embryonic development.

The pluripotent nature of mouse ES cells was demonstrated by their ability to contribute to all tissues of adult mice, including the germline, following their injection into host blastocysts. Beyond their developmental capacity *in vivo*, ES cells display a remarkable potential to form differentiated cell types in culture. Under appropriate conditions *in vitro*, ES cells can differentiate and form embryoid bodies that have been shown to contain cells of the hematopoietic, endothelial, muscle and neuronal lineages (Keller GM. *et al.*, (1995)).

Comparable studies are difficult in the mouse embryo and impossible in the human embryo. Thus, the potential to generate virtually any differentiated cell type from embryonic stem cells (ESCs) offers the possibility to establish new models of mammalian development and to create new sources of cells for regenerative medicine.

The basic characteristics of ES cells, which include self-renewal, multilineage differentiation, clonogenicity, a normal karyotype, extensive proliferation, and the ability to be frozen and thawed, are all the fundamental properties of early

embryonic cells. External signaling pathways such as LIF-gp130-STAT3, BMP-TGF- β -Nodal -Smad, MAPK-ERK, and WNT as well as transcription factors *Oct3/4* and *Nanog* have been reported to play important roles in the self-renewal of mouse ES cells. Pluripotent mouse ES cells are usually propagated under complex culture conditions that are poorly defined because they include both growth-inactivated feeder cells and serum. The exposure of ES cells to these components has been a major concern for the potential therapeutic use of ES cells because of their immunogenicity and potential pathogenicity. After the discovery of LIF as a crucial factor produced by the feeder cells, mouse ES cells have been propagated in medium containing serum and LIF without feeder cells on gelatin-coated dishes, mainly (Ward CM. *et al.*, (2002)).

5.1.1 Cell states and identities

The human body is composed of over 37 trillion cells that can give rise to hundreds of different cell types despite sharing identical genotype (Han X. *et al.*, (2020)). This cellular diversity and morphological complexity have been achieved through cellular differentiation, as well as evolution and development. In addition to differentiation, cells are constantly challenged by external environment and thus must develop phenotypic plasticity. While some states are transitory and/or only present during embryonic development, others can become stable throughout the cellular lifetime. Cell states change not only during development, but also during disease. So, a great variety of cell states is present in each multicellular organism. Regarding the cell identity, many years ago Waddington CH. *et al.*, (1942) described the most famous and powerful metaphor in developmental biology, named as epigenetic landscape. It depicts how a cell progresses from an undifferentiated state to one of several discrete, distinct, differentiated cell fates during development. The cell is represented by a ball, and it starts out in a valley at the back of the landscape (Figure 13). As the ball rolls forward and downward, the valley splits or bifurcates into two new valleys separated by a ridge. These new valleys represent alternative cell fates. External stimuli, like inductive influences, or internal influences determine which of the two valleys a particular cell chooses. Likewise, the several and putative cell fates are dictated by the activity of specific genes, which support the landscape. The valleys continue to split, and eventually the cell ends up in one of many terminal sub-valleys, which represent terminally differentiated states. The cell is kept permanently in its terminally differentiated state by high valley walls. The steeper the walls, the more “canalized”, in Waddington’s terminology, the cell fate. The Waddington landscape highlights a key feature of epigenetic changes; once cells have undergone an epigenetic change (e.g., differentiation), they (and their progeny) have undergone a stable change that will not reverse. Transcription factors, which bind to DNA, are especially relevant for defining cellular states and to influence the expression of many genes. Some of

cell type-specific transcription factors can (through their absence or robust activity) affect cellular lineage choices during development and/or differentiation. In fact, all cell identities and lineage choices are regulated by one or a combination of master transcription factors (Morris SA. *et al.*, (2013)). A particular subset of master transcription factors is represented by reprogramming factors, which can alter identities of differentiated cells beyond their natural potency. Usually, their function is based on a genetic manipulation, thus the new acquired states are stable and heritable, even if the added reprogramming factors are removed. Interestingly, the reprogramming factors are classified according to their potency. They can: revert development (de-differentiation and induced pluripotency), induce alternative options for differentiation (trans-differentiation or trans-determination), or enforce alternative differentiated identities in already differentiated cells (Figure 13).

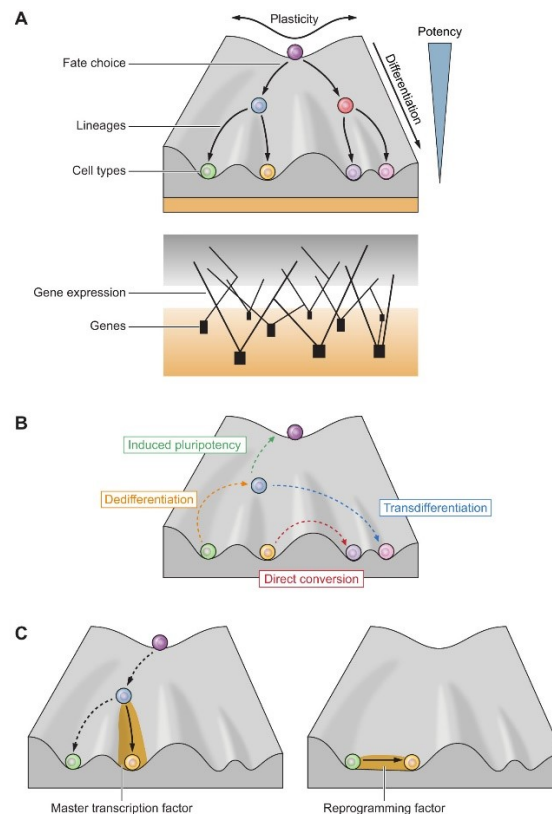


Figure 13 The Waddington's epigenetic landscape model. **A.** Cellular properties (plasticity, potency), natural state changes (fate choice, lineages, cell types, differentiation), and molecular foundations (genes, gene expression); **B.** Cellular reprogramming, which changes the cell states; **C.** Transcription factors controlling cellular states. Master transcription factors influence cell fate decisions during development. Reprogramming factors can alter cell identities beyond the natural potency of the cell (Breunig CT. *et al.*, (2020)).

5.2 Differentiation model of ES cells

By literature, when ES cells were removed from the factors that maintain them as stem cells, they are able to differentiate and, under appropriate conditions, generate progeny consisting of derivatives of the three embryonic germ layers: mesoderm, endoderm, and ectoderm (see Figure 14) (Keller G. *et al.*, (2005)).

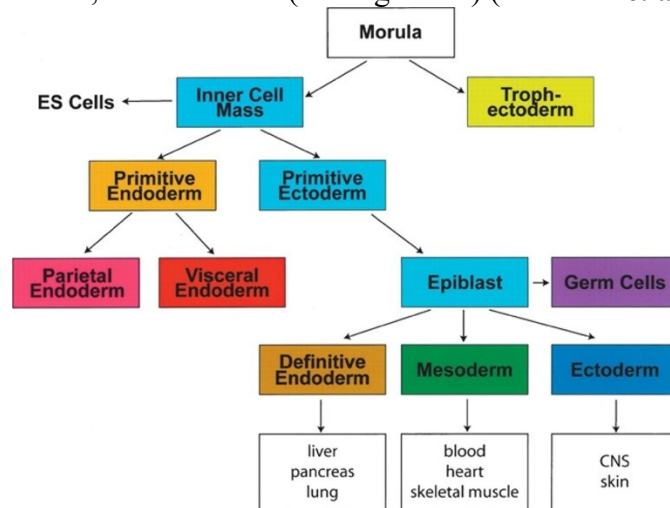


Figure 14 Scheme of early mouse development. ES cells differentiate in derivatives of the three embryonic germ layers: mesoderm, endoderm and ectoderm (Keller G. *et al.*, (2005)).

Furthermore, ES cells do not differentiate to troph-ectoderm *in vitro* and so they reflect the potential of their founder embryonic population, the inner cell mass. Contrary, human ES cells can also form the extra-embryonic tissues (troph-ectoderm) that differentiate from the embryo before gastrulation. The use of human ES cells to derive early human trophoblast is particularly valuable, because it is difficult to obtain from other sources and is significantly different from mouse trophoblast. According to Xu *et al.*, (2002) when hES cells are induced with bone morphogenetic protein 4 (BMP4), a member of the transforming growth factor-beta (TGF-beta) superfamily, they give rise to cells that display characteristics of the trophoblast lineage.

As outlined in Figure 15, three basic methods have been developed to promote differentiation of ESCs: (1) the formation of three-dimensional aggregates known as embryoid bodies (EBs); (2) the culture of ESCs directly on supportive stromal layers (the most commonly used stromal cell line for such differentiation studies is OP9); (3) the culture of ESCs as monolayers on extracellular matrix proteins (Murry CE. *et al.*, (2008)).

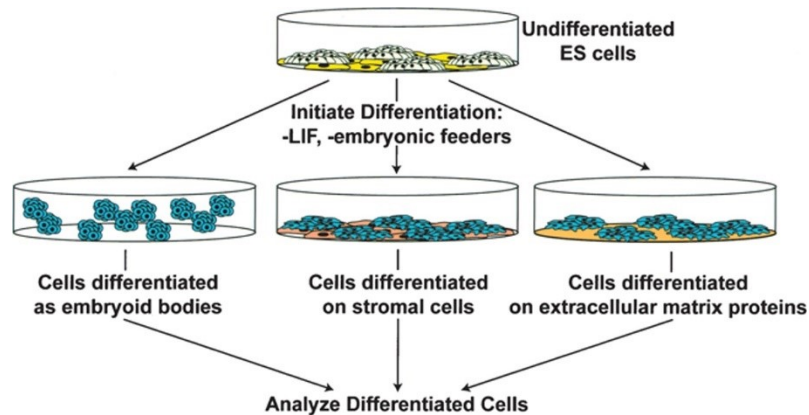


Figure 15 Three general approaches used to induce ES cell differentiation (Murry CE. *et al.*, (2008)).

Each method has been successful for the development of certain lineages. Many studies that investigated the differentiation capacity of ESCs included fetal calf or bovine serum (FCS or FBS) in the protocol. The use of FCS/FBS has several serious drawbacks that include batch-to-batch variability and the lack of identity of the inducing factors contained in it. Thus, these protocols were often difficult to reproduce, and most were not well optimized for the generation of any lineage. These obstacles were overcome by using serum-free media with specific inducers to direct differentiation and the development of reporter ESCs to monitor early differentiation steps have.

All three approaches to ES cell differentiation are valuable and have specific advantages and disadvantages. EBs offer the advantage of providing a three-dimensional structure that enhances cell–cell interactions that may be important for certain developmental programs. The complexity of the EBs can also be a disadvantage as the generation of cytokines and inducing factors within these structures can complicate interpretations of experiments in which one is trying to understand the signaling pathways involved in lineage commitment.

Coculture with stromal cells provides the beneficial growth promoting effects of the particular cell line used. However, undefined factors produced by these supportive cells may influence the differentiation of the ES cells to undesired cell types. Another problem is the difficulty that can be occurred when attempting to separate the ES-cell-derived cells from the stromal cells.

Lastly, differentiation in monolayers on known substrates can minimize the influence of neighboring cells and supportive stromal cells, but with this protocol the matrix proteins composition is critical. Different proteins may dramatically influence the generation and survival of the developing cell types.

Overall, three criteria should be considered when using the ES cell model for lineage-specific differentiation. First, protocols need to be established that promote

the efficient and reproducible development of the cell type of interest. If possible, selection strategies should be combined with optimal differentiation schemes to enable the isolation of highly enriched cell populations. Second, lineage development from ES cells should recapitulate the developmental program that establishes the lineage in the early embryo. Third, the mature cell populations that develop in these cultures must display appropriate functional properties both in culture and when transplanted into appropriate animal models.

This developmental biology approach has made it possible to recapitulate in ESC cultures the key events that regulate early lineage commitment in the embryo, resulting in the efficient and reproducible generation of highly enriched differentiated cell populations.

5.3 Cardiac differentiation of mouse embryonic stem cell lines

Heart development depends on precise temporal control of gene expression patterns, and disruption of transcriptional networks in heart development underlies congenital heart disease (CHD) (Bruneau BG. *et al.*, (2008)).

Cardiac differentiation is a dynamic process consisting of complex signaling network. In literature, numerous differentiation protocols have been described to generate cardiomyocytes from pluripotent stem cells. Overall, various cardiac differentiation studies were based on the induction of pluripotent stem cells with various growth factors, at an accurate timing and dose. It is essential for directing the differentiation process from early mesendoderm via mesoderm towards a more specific cardiac fate.

As demonstrated by Kattman *et al.*, (2011), it is possible to efficiently generate cardiovascular cells from different mouse ESCs. Using a serum-free differentiation procedure, they found that the coexpression of *Flk1/Kdr* and *Pdgfr- α* marked an early mesoderm population, which had a cardiac potential. Additionally, *Flk⁺/Pdgfr- α ⁺* (F⁺P⁺) cells developed under precise culture conditions. In fact, ES cells were able to generate a substantial F⁺P⁺ population when induced with 2, 4, 8, and 16 ng/ml Activin along with low concentrations of BMP4. Increasing the amount of BMP4 led to an increase in the size of the F⁺P⁺ populations in the presence of the different concentrations of activin tested. Subsequently, F⁺P⁺ cells induced with activin 8 ng/ml and BMP4 0.5 ng/ml (~ 50-60%) differentiated and give rise to a population consisting of 65% contracting *cTnT⁺* cells (see Figure 16). Moreover, mesodermal marker gene *Mesp1* was expressed in the F⁺P⁺ population and preceded expression of *Isl1* and *Nkx2.5*.

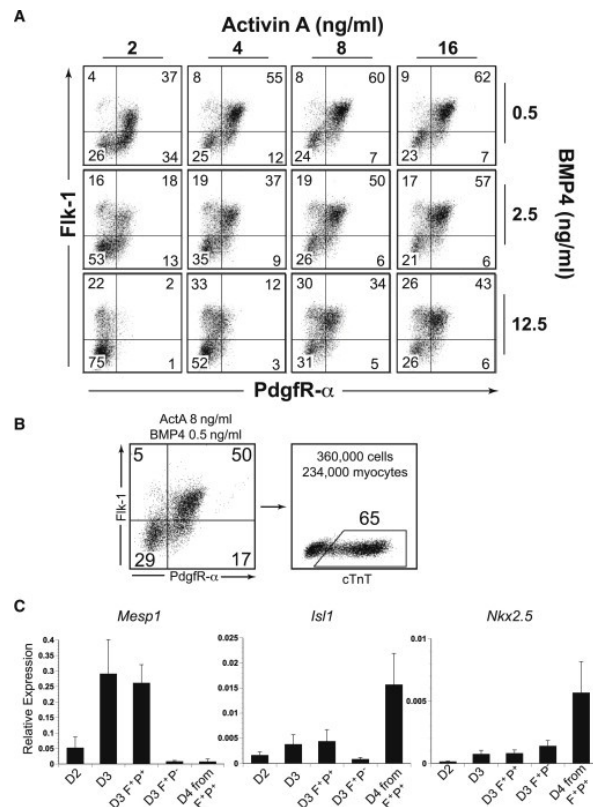


Figure 16 Cardiac Potential of *Flk*⁺/*Pdgfr*- α ⁺ (F⁺P⁺) mesoderm cell population (Kattman *et al.*, (2011)).

Subsequently, this type of serum-free protocol was also used by Wamstad *et al.*, (2012) to investigate how global patterns of gene expression and chromatin organization are coordinated in the cardiac lineage. Their results showed that the model system used reproduced normal cardiomyocyte differentiation and resulted in approximately 70% cardiac Troponin T (*cTnT*)-positive cardiomyocytes (Figure 17 a-b-c).

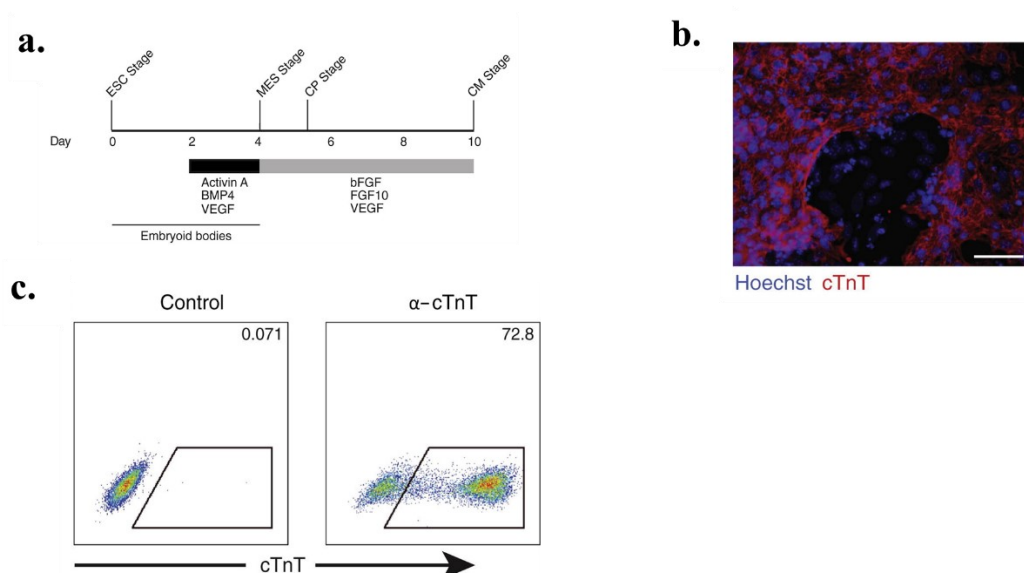


Figure 17 Cardiomyocyte differentiation of embryonic stem cells. **a.** Scheme of differentiation protocol, indicating the time points where differentiating cultures were treated with the specific growth factors listed below. **b.** cTnT (in red) immunofluorescence of day10 differentiated cells. **c.** Flow cytometry at day10 of cells stained for cTnT or isotype control (Wamstad *et al.*, (2012)).

Differentiating cultures were highly enriched at earlier stages for the cardiac transcription factors *Nkx2-5* and *Isl1*, indicating that these cells activated efficiently normal cardiac specific commitment. Based on the gene expression profile, the authors selected four stages of differentiation that represented key cell types in the transition from pluripotent cells to cardiomyocytes (Figure 18a): undifferentiated embryonic stem cells (ESC) expressing pluripotency genes (*Pou5f1/Oct4* and *Nanog*), cells expressing mesodermal markers (*Mesp1* and *Brachyury*) (MES), cells expressing cardiac transcription factors (*Nkx2-5*, *Tbx5*, and *Isl1*) but not yet beating (CP), and functional CM with cardiomyocyte-specific gene expression (*Myh6* and *Myh7*) (Figure 18b).

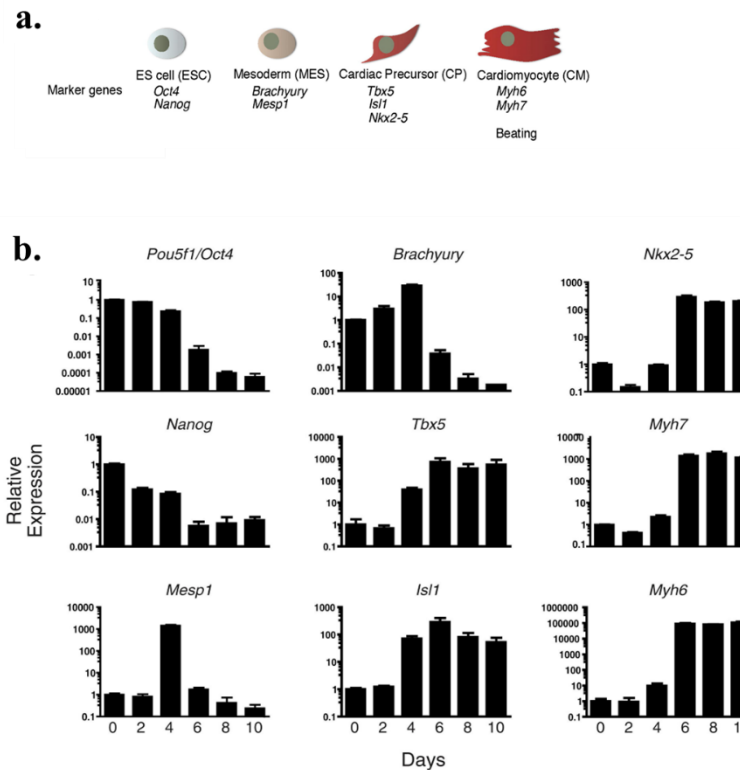


Figure 18 Transcriptional profile during transition from pluripotent cells to cardiomyocytes. **a.** A representation of the four stages of cardiac differentiation. **b.** Expression of the main marker genes activated during cardiac differentiation (Wamstad *et al.*, (2012)).

5.4 Vascular endothelial cell differentiation from ES cells

Embryonic stem cells (ESCs), as well as human pluripotent stem cell (hPSCs), represent a suitable *in vitro* model to study molecular events involved in vascular development. It has been demonstrated that ES cells can spontaneously differentiate into endothelial cell lineage, showing an increase in the expression of several endothelial cell-specific genes during differentiation and forming vascular structures in ES-derived EBs (Vittet D. *et al.*, (1998)). The vasculogenic potential of the embryonic cells could potentially lead to a variety of clinically relevant applications. In fact, vascular endothelial cells or endothelial progenitors cells could be used in therapeutic strategies for the repair and revascularization of ischemic tissue in patients exhibiting vascular defects (Kocher AA. *et al.*, (2001)). Murine embryonic stem (ES) cell lines can be maintained undifferentiated if cultured in the presence of leukemia inhibitory factor (LIF), which inhibits their differentiation. When LIF is removed, ES cells spontaneously differentiate into three-dimensional aggregates, termed embryoid bodies (EBs), which contain

derivatives of the three primitive germ layers (ectoderm, mesoderm, endoderm). The appearance of endothelial-specific markers occurred at different times during ES differentiation, recapitulating *in vivo* vasculogenesis steps. It suggested that endothelial cell commitment follows sequential maturation stages. During *in vivo* vascular development, *Flkl* is the earliest endothelial marker and is expressed in both extraembryonic and embryonic mesoderm at E7.0. *Pecam1* is detected from E7.0/E7.5 simultaneously with *Flkl*. At E7.5, in the yolk sac mesenchyme (from which will originate the yolk sac vasculature), is activated the expression of *Tie2* and *VE-Cadherin*, followed that of *Flkl*. Numerous studies in literature indicated that the addition of various growth factors (i.e., VEGF and GFs cocktail) to the culture medium could optimize the *in vitro* model, obtaining a higher level of endothelial cells differentiation from ES cells. These factors are known to be involved in the regulation of angiogenesis and/or vascular development. The two main cellular components of blood vessels are endothelial cells (ECs) and vascular smooth muscle cells (VSMCs). Both ECs and VSMCs are required for vascular function, including blood pressure control, interactions with immune cells, and the uptake of nutrients. In 2015, Patsch *et al.*, developed a rapid and efficient method to differentiate hPSCs into ECs and VSMCs. First, GSK3 β inhibition combined with BMP4 treatment enabled the commitment of hPSCs to mesoderm. Then, differentiation was induced by treatment with VEGF-A for ECs or PDGF-BB and ActivinA for VSMCs, similarly to vascular development *in vivo* (Figure 19a-b).

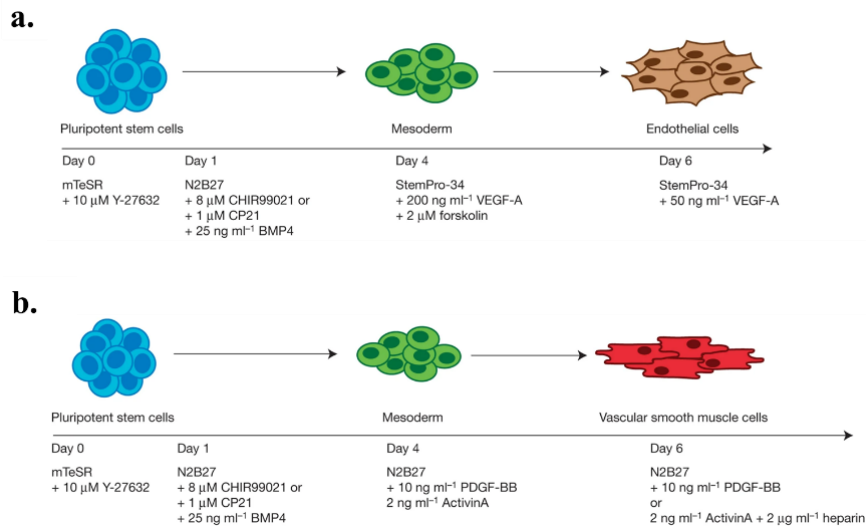


Figure 19 a. Schematic illustration of ECs differentiation protocol from hPSCs. **b.** Schematic illustration of VSMCs differentiation strategy from hPSCs (Patsch *et al.*, (2015)).

Overall, the GSK3 inhibition (with CP21 and CHIR molecules) and BMP4 treatment followed by a brief period of VEGF-A and forskolin growth factors exposure and continued VEGF-A treatment is sufficient to efficiently induce ECs from hPSCs. Forskolin is a cyclic-AMP signaling activator and data in literature demonstrated that cAMP enhance the vascular development (Yamamizu *et al.*, (2009)). Flow cytometry of VE-Cadherin⁺ (CD144⁺) cells, assessed that the protocol used promoted the differentiation of hPSCs into ECs with efficiencies of about 70% (Figure 20a-b). hPSC-derived ECs expressed endothelial markers, like *VE-Cadherin*, *vWF*, *Pecam1* (Figure 20c). FACS analysis of CD144⁺ hPSC-ECs on day 10 showed the expression of ECs-specific markers (KDR, CD31, CD34, CD105) and the absence of hematopoietic markers (CD43, CD45) (Figure 21). In the same way, ActivinA and PDGF-BB treatment following mesoderm induction resulted in the formation of almost exclusively CD140b⁺ (PDGFRB) cells (89.4%) with no CD144⁺ cells detectable (Figure 20a-b). The cells also expressed other markers of VSMCs, such as α SMA, myosin IIB, SM22a (Figure 20d).

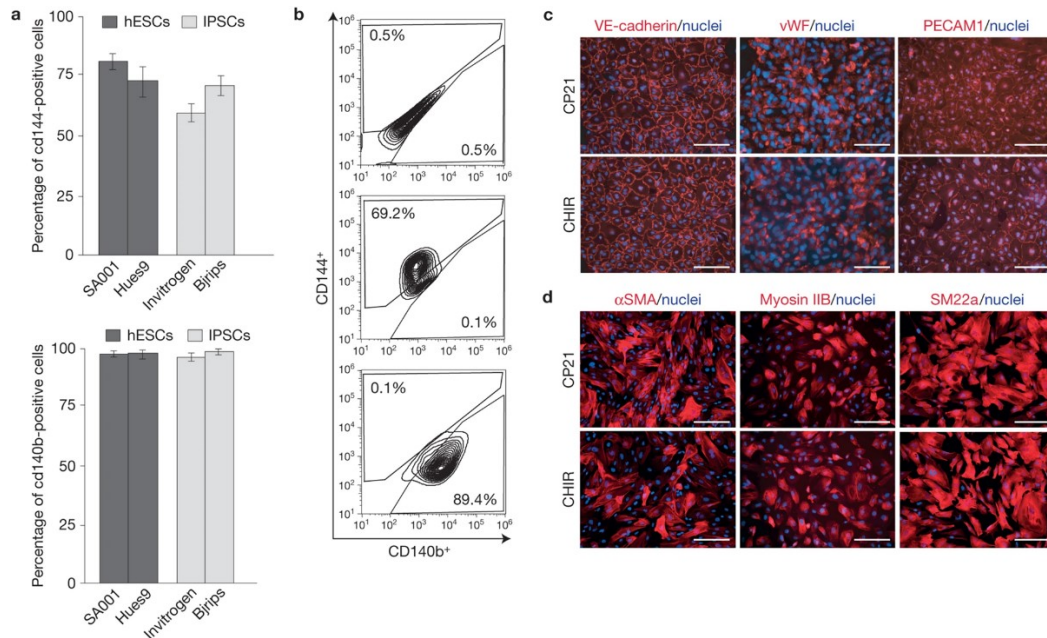


Figure 20 **a.** Differentiation efficiency of hPSC ECs and hPSC VSMCs from four different hESC and induced PSC lines by evaluation of CD144⁺ (VE-Cadherin) and CD140b⁺ positive cells. **b.** FACS sorting plots from the differentiation experiments in hPSCs (top panel), hPSC-derived ECs (middle panel) or hPSC-derived VSMCs (lower panel) stained for CD144 and CD140b. **c.** Immunostaining of EC-specific markers (VE-Cadherin, vWF, Pecam1) on hPSC ECs for both GSK3 inhibitors. **d.** Immunostaining of VSMC-specific markers on hPSC VSMCs (α SMA, myosin IIB, SM22a) (Patsch *et al.*, (2015)).

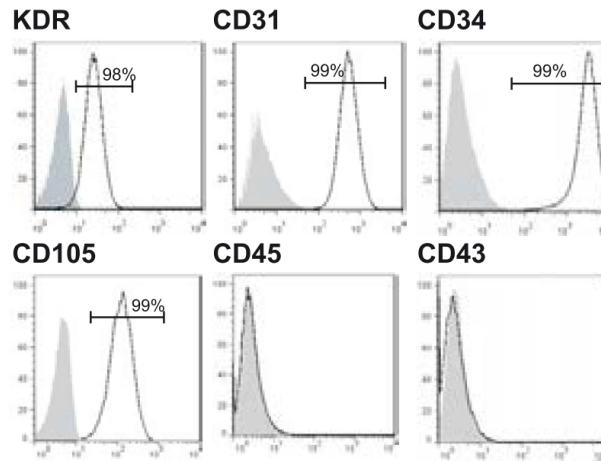


Figure 21 Flow cytometry analysis of CD144+ hPSC-ECs on day 10 of differentiation. Detection of endothelial- specific genes expression (KDR, CD31, CD34 and CD105), but no hematopoietic markers (CD43, CD45) (Patsch *et al.*, (2015)).

Interestingly, when hPSC ECs were plated on Matrigel, they were able to form vascular network-like structures within 24 h, suggesting that hPSC ECs have angiogenic potential *in vitro* (Figure 22).

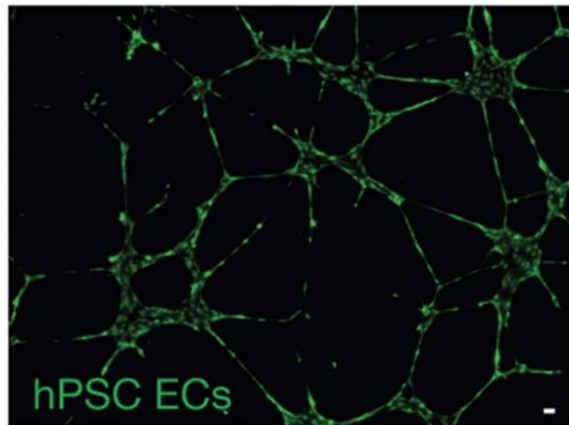


Figure 22 *In vitro* tube formation assay by plating hPSC ECs on Matrigel for 24 h (Patsch *et al.*, (2015)).

6. DNA-regulatory elements: Enhancers

Cellular differentiation requires precisely regulated tissue-specific and developmental stage-specific gene expression patterns. Numerous studies have determined that a key regulators of differential gene expression programs are the enhancers, cis-regulatory sequences that physically contact the target promoters and thus govern spatiotemporal and quantitative expression dynamics of target genes (Heinz S. *et al.*, (2018)). Consequently, enhancers contribute to determine the cell state identity during development. Accordingly, their aberrant activity could be involved in numerous complex human diseases (Figure 23).

Functional enhancers are non-coding sequences in the genome that activate the expression of target genes transcribed by the RNA polymerase II (RNAPII). They are discrete small DNA elements, with a size range of tens to hundreds of base pairs (bp), characterized by dense clusters of transcription factor binding sites (TFBS). Thus, enhancer is bound by cell type-specific TFs, coregulators, chromatin modifiers, architectural proteins like Cohesin, Condensin and CTCF (CCCTC-Binding Factor), other enzymes, and RNAPII. Owing to such large-scale protein assembly, enhancers are often nucleosome deficient, and are hypersensitive to nucleases reflecting DNA accessibility, a feature largely considered as a signature of enhancers. So, the enhancer complex loops over and physically contact the target promoter and activates gene transcription (Figure 24).

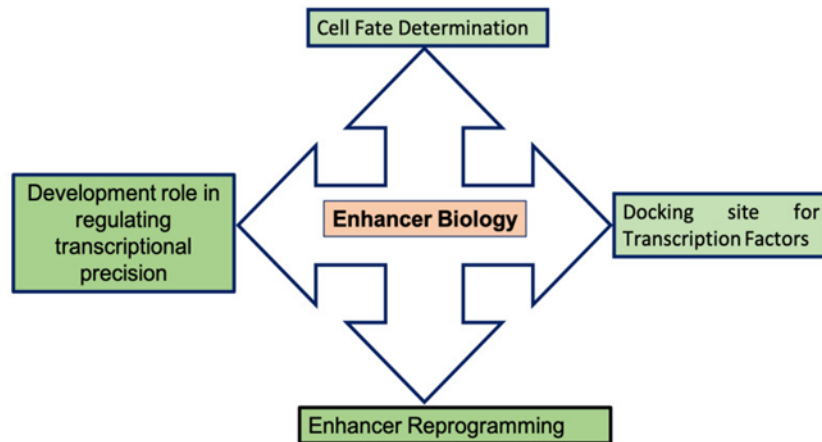


Figure 23 Schematic diagram representing enhancer activity for various biological processes governing development. Enhancers play a central role in cell fate determination; in regulating transcriptional programs that control development, cell identity and evolutionary processes. They also serve as docking sites for TFs, and the activity of enhancers is mostly based on the binding of these TFs. Enhancer reprogramming is an emerging area in developmental biology and cancer research: it could represent a hallmark of carcinogenesis, as it contributes to the deregulated expression of epigenetic modifiers, leading to abnormal cell growth (Maurya SS. *et al.*, (2021)).

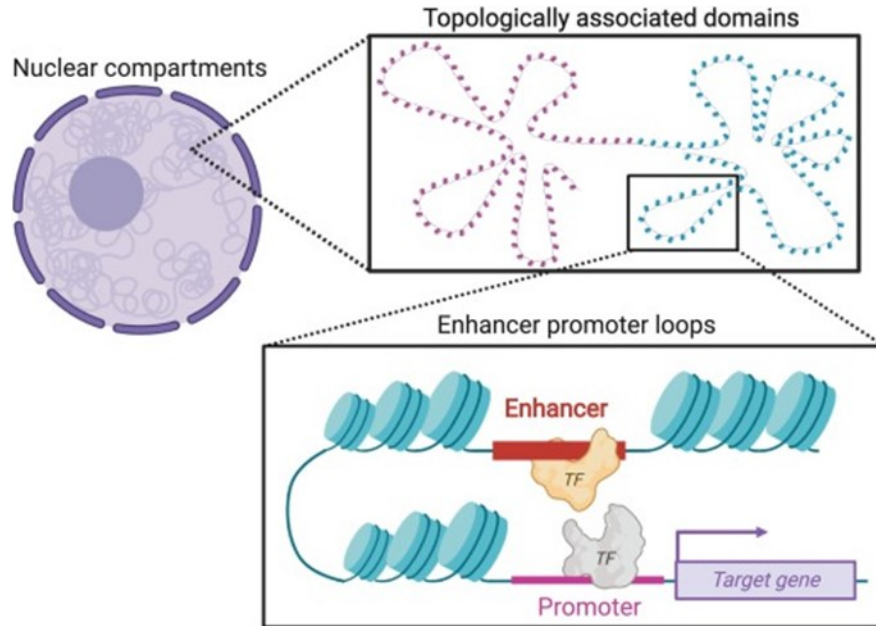


Figure 24 Representation of enhancers loop: enhancers can regulate gene expression by recruiting TFs and the transcriptional machinery and subsequently forming a loop with the promoter region of the target gene (Claringbould A. *et al.*, (2021)).

Moreover, enhancers can act independent of orientation, distance, and location with respect to the transcription start site (TSS) of the target genes and can be located over as much as a million base pairs away. In some cases, they can even activate transcription of genes located in a different chromosome (Lomvardas S. *et al.*, (2006)). Interestingly, they are found mostly in the intergenic and intronic regions, while a few enhancers have been found within exons.

6.1 Molecular assay to identify candidate enhancer regions

As mentioned above, enhancers are hypersensitive to nucleases and DNA accessibility is widely considered as a signature of enhancers. In order to an enhancer element to be bound TFs, the chromatin must be in an “open” state (with the exception of pioneer TFs). For this reason, mapping regions of open chromatin is a widely used way to identify enhancers, and other regulatory elements, across the genome of any cell type. In the past, isolation of enhancers has been based on laborious molecular approaches based on either chromatin structure (e.g., DNase I hypersensitivity assay) and/or transactivation activity (e.g., gene reporter assays) (Wu C. *et al.*, (1979)). Now, with the recent development of d high throughput sequencing technologies, it is possible to identify enhancers at the genome-wide scale. Chromatin accessibility approaches directly measure the effect of chromatin

structure modifications on gene transcription and do not require antibodies or epitope tags that can introduce potential bias, as in ChIP-seq. Current genome-wide high-throughput chromatin accessibility assays are: MNase-seq; DNase-seq, FAIRE-seq and ATAC-seq (Figure 25).

6.1.1 MNase-seq chromatin accessibility assay

MNase-seq (micrococcal nuclease digestion with deep sequencing) is used to map nucleosome positions in eukaryotic genomes to study the relationship between chromatin structure and DNA-dependent processes. This technique is based on the use of the non-specific endo-exonuclease micrococcal nuclease, an enzyme derived from the bacteria *Staphylococcus aureus*, to bind and cleave protein-unbound regions of DNA on chromatin. DNA bound to histones or other chromatin-bound proteins (e.g., transcription factors) remain undigested. The uncut DNA is then purified from the proteins and sequenced by Next-Generation sequencing methods (NGS) (Cui K. *et al.*, (2012)). In a typical MNase-seq experiment, mononucleosomes are extracted by extensive MNase treatment of chromatin crosslinked with formaldehyde. The nucleosomal population is subsequently submitted to NGS procedure. MNase-seq thus probes chromatin accessibility indirectly revealing the regions of the genome occupied by nucleosomes and other regulatory factors and providing information on TF occupancy (Figure 25). Overall, MNase-seq is a useful method for probing genome-wide nucleosome distributions and also provides an accurate way for assessing TF occupancy in many cell types. However, it requires a large number of cells and careful enzymatic titrations for accurate and reproducible evaluation of differential substrates.

6.1.2 DNase-seq chromatin accessibility assay

DNase-seq (DNase I hypersensitive sites sequencing) is a helpful approach to identify the location of regulatory regions through genome-wide sequencing of regions sensitive to cleavage by DNase I. Mapping DNase I hypersensitive (HS) sites has historically been used for identifying all different types of regulatory elements, including promoters, enhancers, silencers, insulators, and locus control regions. In DNase-seq method, DNA-protein complexes are treated with DNase I, which selectively digest nucleosome-depleted DNA (presumably by transcription factors), whereas DNA regions tightly wrapped in nucleosome structures are more resistant. Digested-DNA fragments are then sequenced to provide accurate representation of the location of regulatory proteins in the genome (Figure 25) (Song L. *et al.*, (2010)). DNase-seq has been extensively used by the ENCODE consortium and others to evaluate cell-specific chromatin accessibility and its relation to differential gene expression in various cell lines. The main controversy

over DNase-seq is the ability for DNase I to introduce cleavage bias, affecting its use as a reliable TFs detection assay. Moreover, DNase-seq requires many cells and involves many samples preparation and enzyme titration steps. Altogether, DNase-seq represents a reliable tool to identify active regulatory elements across the genome and in any cell type from a sequenced species, without a prior knowledge of additional epigenetic information.

6.1.3 FAIRE-seq chromatin accessibility assay

One of the easiest methods for directly probing nucleosome-depleted areas of a genome is FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements). FAIRE is based on the phenol-chloroform separation of nucleosome-bound and free areas of a genome in the interphase and aqueous phase respectively. The procedure involves the initial crosslinking of chromatin with formaldehyde to capture the protein-DNA interactions, and subsequent shearing of chromatin with sonication. Following phenol-chloroform extraction, nucleosome-depleted areas of the genome are released to the aqueous phase of the solution due to much higher crosslinking efficiency of histones to DNA, compared to other regulatory factors. This assay extracts the non-cross-linked DNA and only these nucleosome-depleted regions will be purified, enriched and sequenced in a high-throughput way using NGS (Figure 25) (Giresi PG. *et al.*, (2006)).

Thus, FAIRE-seq is a simple and high reproducible protocol, which does not require antibodies, enzymes (such as DNase or MNase) and does not require a single-cell suspension or nuclear isolation, so it is easily adapted for use on tissue samples.

6.1.4 ATAC-seq chromatin accessibility assay

ATAC-seq (Assay for Transposase Accessible Chromatin with high throughput sequencing) is the most current powerful approach for genome-wide chromatin accessibility profiling. This method probes DNA accessibility with hyperactive Tn5 transposase, which inserts sequencing adapters into accessible regions of chromatin. Sequencing reads can then be used to infer regions of increased accessibility, as well as to map regions of transcription-factor binding and nucleosome position. The method is a fast and sensitive alternative to DNase-seq for assaying chromatin accessibility genome-wide, or to MNase-seq for assaying nucleosome positions in accessible regions of the genome (Figure 25) (Buenrostro JD. *et al.*, (2013)). One of the main advantages of ATAC-seq over other methods, like DNase-Seq or FAIRE-Seq, is that ATAC-seq can be performed with significantly fewer cells (~ 50,000 cells for ATAC-seq compared to millions of cells for the other methods). Accordingly, ATAC-seq is fast, simple and sensitive approach, which works with many cell types and species and does

not require sonication, phenol-chloroform extraction (like FAIRE) or antibodies (ChIP-seq). Moreover, modifications have been made to the protocol in order to perform single-cell analysis. ATAC-seq is used for: nucleosome mapping (identification of changes in nucleosome position during differentiation or between experimental conditions); transcription factors occupancy analysis; identification of novel enhancers during development; deep study of the genomic profile associated to pathological conditions such as cancer.

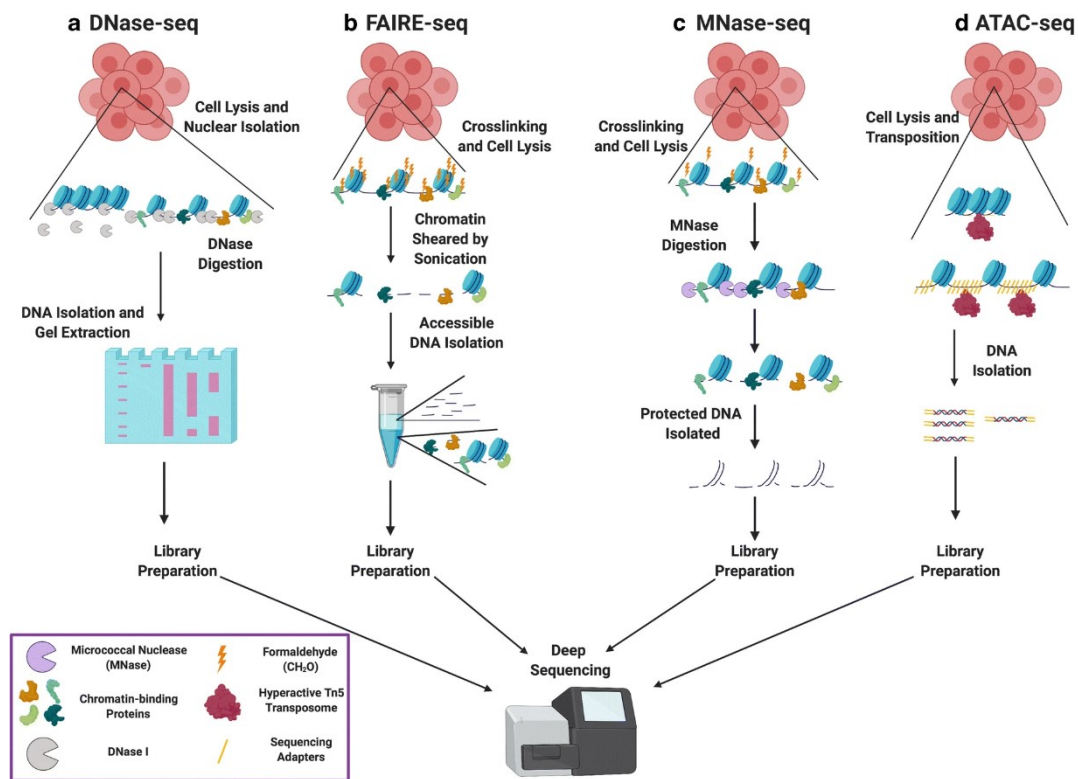


Figure 25 Genomic methods in mapping chromatin accessibility. **a.** DNase-seq: enzymatic digestion to extract signal from open chromatin regions that are known as DNase I hypersensitive sites (DHSs); **b.** FAIRE-seq: method based on crosslinking of chromatin-interacting proteins to DNA using formaldehyde. Chromatin is then sheared, and regions that are unbound by proteins (e.g., histones) remain in the aqueous layer of a phenol-chloroform extraction, while crosslinked DNA remains in the organic layer; **c.** MNase-seq: enzymatic digestion to extract signal representing nucleosome positioning. After formaldehyde crosslinking, added MNase digests DNA that is unprotected by bound proteins, allowing one to infer increased accessibility; **d.** ATAC-seq based on the hyperactive Tn5 transposase to insert sequencing adapters at accessible regions of the genome. Following transposition, genomic DNA can be isolated and amplified by PCR, then subjected to deep sequencing (Klein DC. *et al.*, (2019)).

6.2 Epigenetic features of enhancer

Enhancers are specified by distinct chromatin features that may contribute to the repertoire of epigenetic mechanisms responsible for cellular memory and cell type-specific gene expression. Recent advances in epigenomic profiling technologies allowed to determine that specific enhancer-associated chromatin features can be really used to annotate them (Maston GA. *et al.*, (2012)). Indeed, enhancers have been identified in the form of primed, active, and poised enhancers. In addition, each type of enhancer is characterized by specific histone modification patterns and can be easily identified by these signatures (Figure 26) (Zentner GE. *et al.*, (2011)).

Notably, active enhancers signatures are identified by H3K4 mono-methylation (H3K4Me1) and H3K27acetylation (H3K27ac). They are also associated with incorporation of hypermobile nucleosomes containing H3.3/H2A.Z histone variants. Prior to activation, enhancers can exist in a primed state, characterized by the presence of histone H3K4 mono-methylation (H3K4me1) only. Other features that have been associated with enhancer priming are presence of pioneer TFs, hypermobile H3.3/H2A.Z nucleosomes, DNA 5mC hypomethylation, and hydroxylation (5hmC). Finally, the signatures of poised enhancers are marked with H3K4Me1 with H3K27me3, but not H3K27ac (Figure 26).

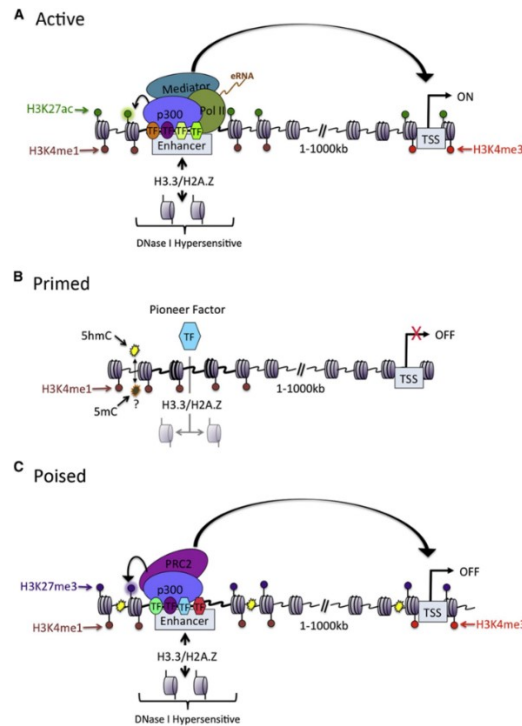


Figure 26 Epigenetic Features of active (A), primed (B), and poised enhancers (C). **A.** Active state: enhancers are associated with incorporation of hypermobile nucleosomes containing H3.3/H2A.Z histone variants, which compete for DNA binding with TFs. TFs in turn recruit coactivator proteins (e.g. p300 histone

acetyltransferase) that can modify and remodel nucleosomes. H3K4me1 and H3K27ac are the predominant histone modifications deposited at nucleosomes flanking enhancer elements. In their active state, enhancers are also bound by general transcription factors (GTFs) and RNA polymerase II (Pol II), leading to the production of enhancer-originating RNAs termed eRNAs. **B.** Primed state (prior to activation): enhancers are characterized by the presence of H3K4me1. Other features that have been associated with enhancer priming are presence of pioneer TFs, hypermobile H3.3/H2A.Z nucleosomes, DNA 5mC hypomethylation, and hydroxylation (5hmC). **C.** Poised enhancer found in mouse and human ESCs: marked by H3K27me3 and associated with PRC2 (polycomb repressive complex 2); these enhancers are bound by TFs and coactivators and communicate with their target promoters. (Calo E. *et al.*, (2013)).

The specific epigenetic modifications found at enhancers are derived from the recruitment of epigenetic writers and erasers. For instance, the myeloid/lymphoid or mixed-lineage leukemia methylases MLL2, MLL3 and MLL4 (also known as KMT2D, KMT2C and KMT2B, respectively) are histone methyltransferases responsible for deposition of the enhancer marks H3K4me1 and H3K4me2 (Kaikkonen MU. *et al.*, (2013)). Similarly, lysine acetyl transferases such as CBP (a.k.a. CREBBP) and P300 (a.k.a. EP300) bind enhancers to increase their activity through protein acetylation, inclusive of histones (Jin Q. *et al.*, (2011)). The EZH2 methyltransferase is responsible of H3K27me3 modification in silenced or poised enhancers. DNA methylation in some silent enhancers is established by DNA cytosine-5-methyltransferases DNMT1, DNMT3A and DNMT3B and the TET methylcytosine dioxygenases TET1, TET2, TET3 necessary for active removal of DNA methylation (Turek-Plewa J. *et al.*, (2005); Kohli RM. *et al.*, (2013)).

Normally, active enhancers are linked to expressed genes, while poised enhancers are always associated to developmental genes, which are inactive in embryonic stem cells or precursor cells and become expressed during different differentiation stages (Cruz-Molina S. *et al.*, (2017)).

During development and in terminally differentiated cells, enhancers can switch between states. Upon embryonic stem cell differentiation, for example, poised enhancers lose the repressive H3K27me3 mark and gain H3K27ac. The enhancer state change occurs concurrently with a change in the expression of the target gene from off to on state and the transition in cell state from undifferentiated to differentiated (Karnuta JM. *et al.*, (2018)).

Furthermore, during cell differentiation, in early precursors the enhancer region is covered by nucleosomes and often associated to repressive marks, such as H3K27me3 or DNA methylation. In this stage the enhancers are in an inactive state. Subsequently, lineage-specific TFs (also called master regulators or enhancer organizers) bind to the majority of the tissue-specific enhancers, which are nucleosome free regions and enriched for H3K4me1, but are generally in a poised state. Upon cell differentiation and/or external stimuli, induced or activated TFs bind to some of the accessible enhancers in order to activate gene expression. Thus, active enhancers are associated with additional cofactors such as BRG1, p300 and Pol II and correlate with further nucleosome remodeling, acquisition of

additional histone modifications, such as H3K27ac and H3K4me3, and local transcription (Figure 27) (Spicuglia S. *et al.*, (2012)).

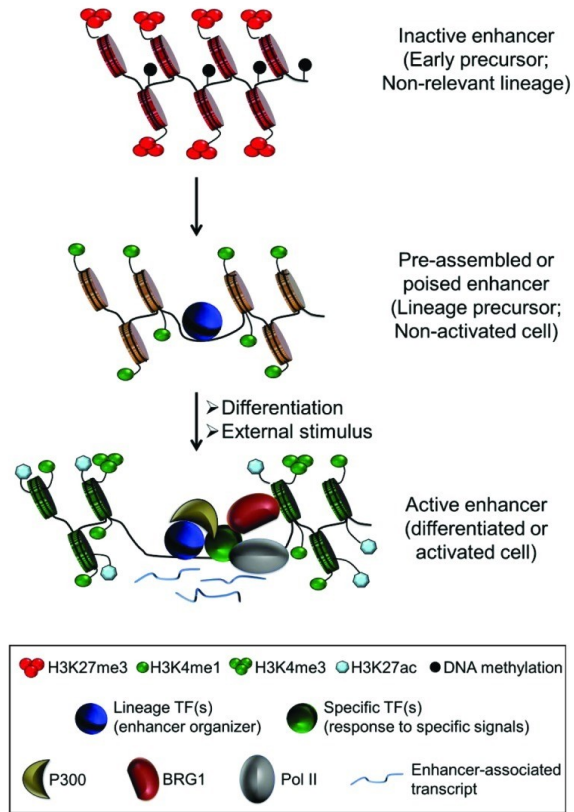


Figure 27 Chromatin enhancer signatures during cell differentiation (Spicuglia S. *et al.*, (2012)).

6.3 Functional validation enhancer assay

The first identified enhancers were defined by functional capacity to amplify transcriptional activity in reporter plasmids (Moreau P. *et al.*, (1981)). Putative enhancers have been predicted via DNA conservation using comparative genomics and, more recently, by epigenetic signatures such as open chromatin from DNaseI hypersensitive site sequencing (DNase-seq) or assaying for transposase-accessible chromatin using sequencing (ATAC-seq), histone tail modifications from chromatin immunoprecipitation sequencing (ChIP-seq), and 3D chromatin organization. However, these approaches are primarily descriptive and not directly evaluate whether a DNA sequence acts as a functional enhancer (Benton ML. *et al.*, (2019)). Enhancer reporter assays assess the ability of

candidate regulatory elements to drive expression of a reporter gene and have been widely used to functionally test activity of predicted enhancers.

6.3.1 Traditional reporter assays for enhancer activity

DNA sequences can be directly tested for their ability to activate or enhance transcription from a minimal core promoter. In fact, this activity, which is also independent of the sequence context of the enhancer and can be assessed in heterologous reporter systems, has been widely used to evaluate enhancer functionality. Generally, enhancer tests in whole developing embryos (for example, in flies, nematodes, zebrafish and mice) have readouts based on images, whereas tests *in vitro* cell culture typically either use luciferase or directly measure the abundance of reporter transcripts (Nord AS. *et al.*, (2015)).

Commonly, a reporter plasmid consists of a portion of a cellular- or viral promoter (i.e., minimal promoter) that will provide sequences required for transcriptional initiation. Enhancers are tested by placing the candidate DNA sequences upstream of a minimal promoter and a reporter gene. The enhancer activity is measured by evaluating the abundance and localization of the reporter transcript (i.e., *in situ* hybridization). Alternatively, the reporter gene is detected at the protein level, such as by enzymatic activities (for example, luciferase or β -galactosidase, encoded by lacZ), fluorescence (i.e., GFP) or specific antibodies. Many promoter and enhancer candidates have been tested in transgenic animals, like *D. melanogaster*, *C. elegans* and mouse embryos.

In addition, the reporter plasmids lacking or containing the test enhancer DNA can be also introduced into *in vitro* cultured cells by transfection and analyzed after 24–48 h using luciferase assays. The level of transcription detected in the absence of the test enhancer represents basal transcription level. The activity of the plasmid containing the candidate enhancer is measured relative to the basal level in order to define whether the putative regulatory element act as enhancer (Figure 28).

However, traditional transgenic approaches can test the activity of an individual enhancer sequence but are not recommended for validating putative enhancers at scale.

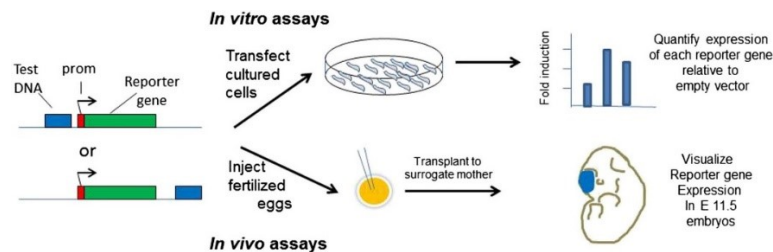


Figure 28 Schematic representation of traditional enhancer assay. Minimal promoter is indicated as red box; reporter gene is the green box; candidate enhancer element (test DNA) is indicated in blue box (Dailey L. *et al.*, (2015)).

6.3.2 Massively Parallel Reporter Assays (MPRA) to study enhancer function

Massively Parallel Reporter Assays (MPRAs) are an alternative and powerful technique for functional characterization of enhancer activity in a high-throughput manner. MPRAs can be used to functionally screen thousands of sequences for regulatory activity simultaneously. These methods are based on traditional reporter assays, in which a putative regulatory element is placed upstream of a reporter coding sequence (e.g., luciferase) in a vector, transfected into cells, and assayed for reporter activity. In MPRAs, libraries containing hundreds of thousands of candidate enhancers can be generated and tested for function in parallel. Each enhancer is cloned upstream of a minimal promoter and a reporter gene containing a unique sequence barcode in its 3' untranslated region. These libraries of barcoded reporter genes can then be introduced into cells or animals and expression quantified by RNA-seq of the unique sequence barcodes (Figure 29). Thus, MPRAs can be used to functionally validate numerous putative enhancer sequences in a single experiment with a quantitative readout (Ryan GE. *et al.*, (2019)). A variation of MPRA assay is Self-Transcribing Active Regulatory Region sequencing (STARR-seq), that exploits the characteristic that enhancers can function independently of their relative positions and places candidate sequences downstream of a minimal promoter and open reading frame. Enhancer activity is directly linked to the underlying DNA sequence and measured as presence of the resulting reporter transcripts among cellular RNA by deep sequencing. Specifically, DNA fragments are cloned downstream of a core promoter and into the 3' UTR of a reporter gene. Active enhancers will transcribe themselves and become part of the resulting reporter transcripts. They are isolated, sequenced and counted (Figure 29).

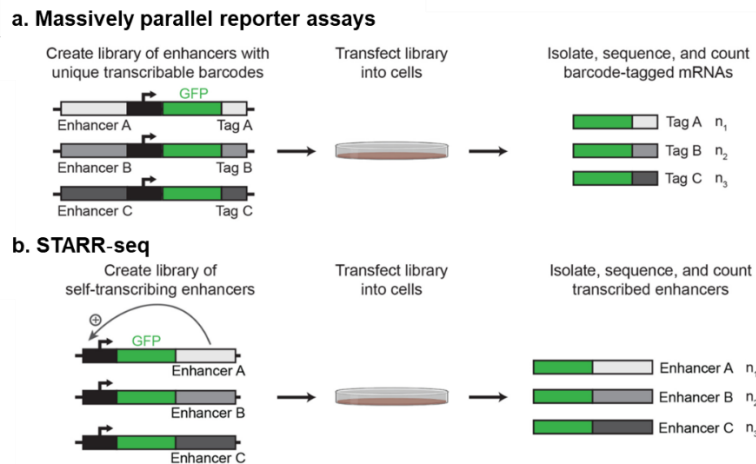


Figure 29 a. Massively Parallel Reporter Assays (MPRA); b. Self-Transcribing Active Regulatory Region sequencing (STARR-seq).

6.4 CRISPR/Cas9-based methods to functionally validate candidate enhancers

Although for more than 30 years, numerous transgenic animal studies have been used to provide *in vivo* evidence for DNA elements that function as transcriptional enhancers, it is also important to explore the functional activity of enhancers through enhancer deletion or enhancer modification experiments at their endogenous location. About this, the recent development of highly efficient CRISPR/Cas9 genome editing technology represents a powerful tool to target and manipulate individual genomic and epigenomic loci (Doudna JA. *et al.*, (2014)). In summary, the CRISPR/Cas9 technology originates from type II CRISPR/Cas systems, which provide bacteria with adaptive immunity to viruses and plasmids. The CRISPR-associated protein Cas9 is an endonuclease that uses a guide sequence within an RNA duplex, tracrRNA:crRNA, to form base pairs with DNA target sequences, enabling Cas9 to introduce a site-specific double-strand break (DSB) in the DNA. The dual tracrRNA:crRNA was engineered as a single guide RNA (sgRNA) that retains two critical features: a sequence at the 5' side that determines the DNA target site by Watson-Crick base-pairing and a duplex RNA structure at the 3' side that binds to Cas9. By changing the guide sequence of the sgRNA complex, Cas9 can target any DNA sequence of interest in the genome. Cas9-induced DSBs are typically repaired in two ways. First, the ends of the DNA breaks are re-joined by endogenous DNA repair pathways, known as non-homologous end-join-ing (NHEJ). However, this mechanism can introduce insertions or deletions of DNA (indels) that in turn may disrupt the translation of the targeted gene. Second, by providing a DNA donor template with homology to the target site, a homology-directed repair (HDR) may occur to repair the double-stranded break (Figure 30). The simplicity of CRISPR/Cas9 programming, together with a unique DNA cleaving mechanism, the capacity for multiplexed target recognition, and the existence of many natural type II CRISPR/Cas system variants, have allowed advances to target, edit, modify, regulate, and mark genomic loci of a numerous type of cells and organisms in precise and efficient manner.

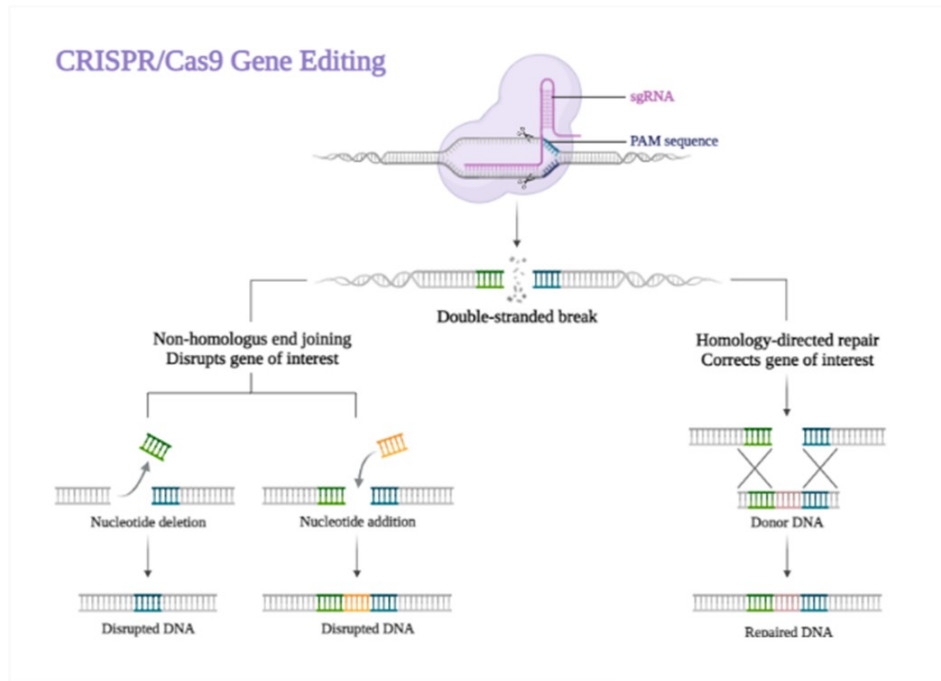


Figure 30 CRISPR/Cas9 mediated gene-editing mechanism (created by Biorender).

6.4.1 Genetic manipulation of enhancers

Genetic perturbation is a powerful approach to investigate links between genetic information and cellular functions. Loss-of-function experiments are essential for the functional investigation of cis-regulatory elements (CREs), such as transcriptional enhancers. CRISPR/Cas9 system allow to easily manipulate genomic sequences but, recently, it is also employed to target non-coding regulatory elements, such as enhancers (Lopes R. *et al.*, (2016)). It was found that a regulatory element can be fully inactivated by Cas9 nuclease through the generation of a specific deletion. This can be achieved using pairs of sgRNAs to target the flanking region of the candidate enhancer (Cebola I. *et al.*, (2021)).

Numerous scientists applied CRISPR/Cas9 strategy to successfully downregulated target gene expression by introducing mutations or deleting an enhancer. For example, Gröschel S. *et al.*, (2014) used this system to remove an enhancer that regulates GATA2 (GATA binding protein 2) expression. Furthermore, CRISPR/Cas9 gene editing was used to knockout two candidate forelimb *Tbx5* enhancers, demonstrating that deletion of the intron 2 and downstream elements, either singly or together in double knockouts, resulted in no effect on forelimb development (Cunningham TJ. *et al.*, (2018)).

Deletion of a 13kb-long super-enhancer (SE) located 100kb downstream of *Sox2* in mouse ESCs, using two sgRNAs flanking the *Sox2*-SE, have shown that the SE is responsible for over 90% of *Sox2* expression, and *Sox2* is the only target gene along the chromosome (Li Y. *et al.*, (2014)).

6.4.2 dCas9 system for gene regulation and epigenome control

A highly related approach to deletion scanning is to use CRISPR to modulate gene transcription processes and also modify the epigenetic landscape around candidate enhancer sequences. To adapt CRISPR/Cas9 for gene regulation studies, in 2013, Qi *et al.*, mutated the nuclease domains of Cas9 from *S. pyogenes* (making an H840A mutation in the HNH domain and a D10A mutation in the RuvC domain) to create a nuclease deficient “dCas9”, also called dCas9 null mutant (Qi LS. *et al.*, (2013)). Thus, dCas9 is unable to cleave DNA but retains the ability to specifically bind to DNA when guided by a sgRNA (Figure 31).

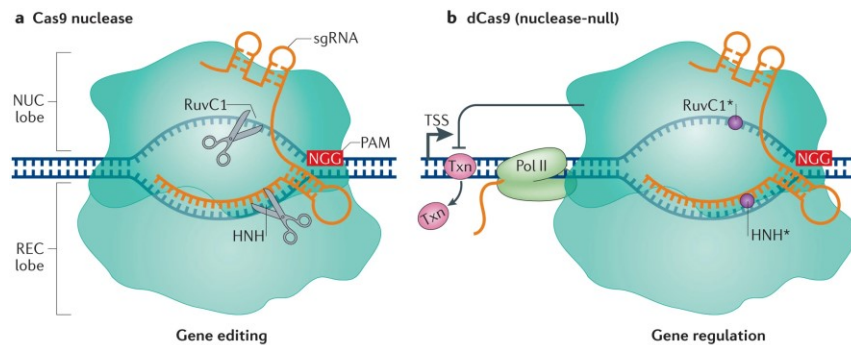


Figure 31 Comparison between Cas9 nuclease (gene editing) and dCas9 nuclease-null (gene regulation). **a.** The *S. pyogenes* Cas9 endonuclease is targeted to specific DNA sequences by direct pairing of sgRNA with the target DNA. Two nuclease domains, RuvC1 and HNH, allow the cleavage of the target sequence. **b.** dCas9 protein contains mutations in its RuvC1 (D10A) and HNH (H841A) domains, which inactivate its nuclease function. dCas9 retains the ability to target specific sequences through the sgRNA and PAM. dCas9 binding downstream of the transcription start site (TSS) can block transcription elongation by blocking RNA polymerase II (Pol II) or the binding of important transcription factors (Txn) (Dominguez AA. *et al.*, (2016)).

Qi *et al.* (2013), showed that dCas9, together with a target-specific sgRNA, can specifically interfere with transcriptional elongation, RNA polymerase binding, or transcription factor binding (Figure 32). This process is called CRISPR interference (CRISPRi). The utility of dCas9 for sequence-specific gene repression was first demonstrated in *E. coli*. In bacteria, the CRISPRi method using dCas9 is highly efficient in suppressing genes; is specific, with minimal off-target effects; and is multiplexable, such that several genes can be simultaneously controlled using multiple sgRNAs. Moreover, the introduction of CRISPRi into mammalian

cells using dCas9 alone achieved only modest repression of enhanced GFP (*egfp*) in the human HEK293T reporter cell line (Qi LS. *et al.*, (2013)). Targeting endogenous genes such as the transferrin receptor CD71, C-X-C chemokine receptor type 4 (CXCR4) and tumour protein 53 (TP53), showed significant gene expression repression (60%–80% repression) (Gilbert LA. *et al.*, (2013)). Hence, the CRISPRi system can be used as a general method for efficiently and specifically regulating gene transcription in many eukaryotes, as well as to characterize cis regulatory elements for transcription factor binding. Generally, the CRISPR/dCas9 system is a broadly applicable tool for genome-wide loss-of-function or gain-of-function screening, inducible and reversible gene regulation and cell fate modulation.

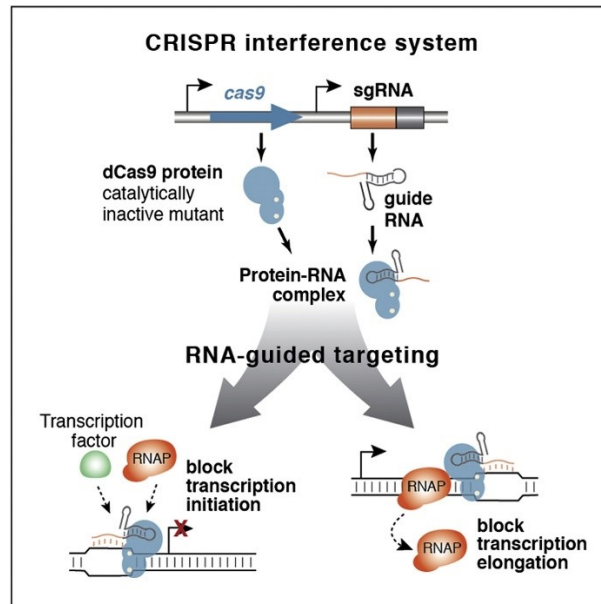


Figure 32 Mechanism of CRISPR interference (CRISPRi) system (Qi LS. *et al.*, (2013)).

6.4.3 CRISPR/Cas9-effector system mediated gene activation and repression

Coupling dCas9 to effector domains with distinct regulatory functions (transcription activators or repressors) converts the CRISPR/Cas9 technology into a site-specific programmable system, which can regulate gene expression in mammalian cells (Figure 33) (Gilbert LA. *et al.*, (2013)).

In CRISPR interference (CRISPRi), dCas9 is targeted to block transcription and thereby silence genes. The fusion of dCas9 to transcriptional repressors increased repression efficiency. CRISPR activation (CRISPRa) uses dCas9 fusion proteins to recruit transcription activators for targeted gene activation. The use of enhanced dCas9 activation systems allows recruitment of multiple activators with one sgRNA.

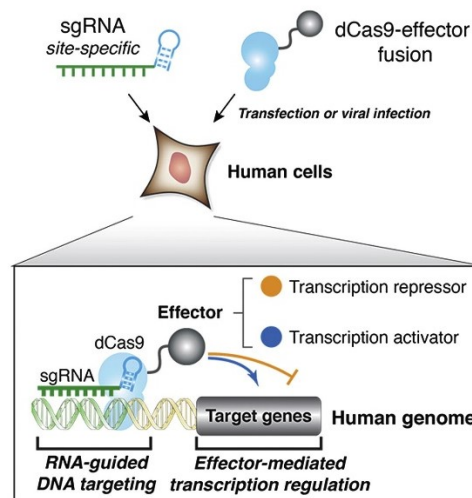


Figure 33 dCas9 fused to effector domains can efficiently activate or silence transcription (Gilbert LA. *et al.*, (2013)).

CRISPRa: dCas9 system for gene activation

dCas9/sgRNA complexes can be modified to activate gene expression in mammalian systems when targeted upstream of endogenous transcriptional start sites. One type of effector that can be fused to dCas9 is a transcriptional activator. There are different forms of these dCas9-activator fusions. For example, Bikard *et al.*, (2013) fused the ω subunit of RNA polymerase to dCas9 in *E. coli*. This fusion was able to activate reporter gene expression up to 3-fold (Bikard D. *et al.*, (2013)). In eukaryotic cells, instead, the first generation of dCas9 activators consisted of dCas9 fused to the activation domain of p65 (a transcription factor

involved in many cellular processes) or a VP64 activator (a synthetic tetramer of the Herpes simplex VP16) (Figure 34a). The dCas9-VP64 fusion was more effective than the p65 fusion and has been used more ubiquitously. Several studies have demonstrated that dCas9-VP64 is able to activate silent endogenous genes and reporters or to upregulate already active genes. Targeting dCas9-VP64 to the promoter of the developmentally relevant transcription factor *Sox17* induced its RNA and protein expression in hPSCs (Kearns NA. *et al.*, (2014)). Studies in literature showed that targeting of dCas9 fused to two VP64 domains to the *MyoD1* locus of mouse primary fibroblasts activated endogenous myogenic genes comparably to traditional *MyoD1* overexpression methods, resulting in the conversion of fibroblasts into skeletal myocytes (Chakraborty S. *et al.*, (2014)). However, the activation seen in mammalian cells was usually moderate, about 2-fold to 5-fold, on average. In order to enhance the activation, dCas9 was fused to a tandem array of peptides, called a SunTag array, which recruits many copies of the VP64 activator effector (Figure 34b) (Gilbert LA. *et al.*, (2014)). A 50-fold increase at the protein level with dCas9-SunTag for endogenous genes such as the CXCR4 chemokine receptor gene was observed in human erythroleukemia K562 cells. Activating endogenous CXCR4 using dCas9-SunTag was sufficient to produce significant increases in cell migration.

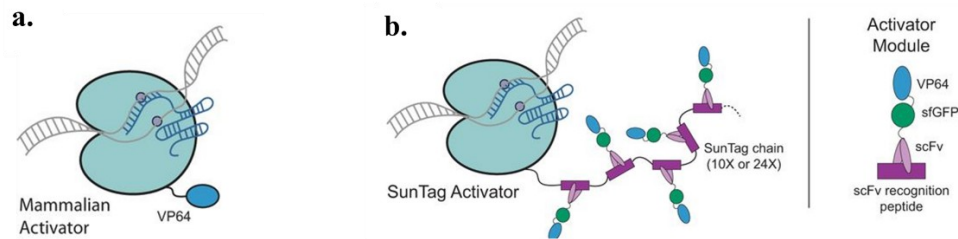


Figure 34 **a.** dCas9 fused with a VP64 activator. **b.** The SunTag activation system (left) consists of dCas9 fused to several tandem repeats of a short peptide sequence separated by linkers. The SunTag activator module (right) is an scFv, which specifically binds the SunTag peptide. The scFv is fused to sfGFP and VP64 (La Russa MF. *et al.*, (2015)).

Another strategy for CRISPR-dependent gene activation, reported by Chavez *et al.*, (2015) employs multiple different activators to synergistically amplify activation. The authors created a tripartite effector fused to dCas9, composed of activators VP64, p65, and Rta (VPR) linked in tandem (Figure 35a). These three activators were joined in a defined order to strongly activate genes, for example in this case was selected a set of genes related to cellular reprogramming, development and gene therapy. Additionally, it can upregulate endogenous gene

expression from 5- to 300-fold at the mRNA level compared to a single dCas9-VP64 fusion (Figure 35b). Furthermore, recruitment of dCas9-VPR to *NGN2* or *NEUROD1* induced neuronal differentiation, suggesting the potential use of the system to modify cell fate through gene activation (Figure 35c) (Chavez A. *et al.*, (2015)).

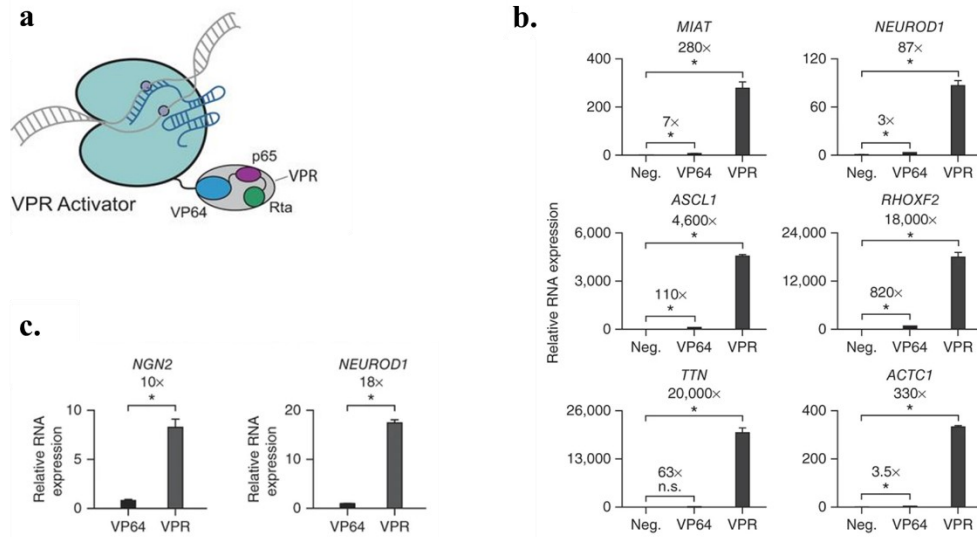


Figure 35 **a.** The VPR activation system is dCas9 fused to VP64, p65, and Rta linked in tandem (La Russa MF. *et al.*, (2015)). **b.** RNA expression of target genes in HEK 293T cells transfected with gRNAs targeting the indicated genes along with the labelled dCas9-activator construct. Negative controls (Neg.) were transfected with the indicated guide RNAs alone. **c.** Analysis of mRNA expression levels of *NGN2* and *NEUROD1* in dCas9-Activator iPSC lines (Chavez A. *et al.*, (2015)).

Subsequently, a third dCas9-activator approach was developed, called the synergistic activation mediator (SAM) system. This system is based on the previous dCas9-VP64 structure but includes an sgRNA modified to recruit additional transcriptional activators for a synergistic activation effect. This modified sgRNA incorporates two RNA hairpin aptamers that bind to dimers of the bacteriophage MS2 coat proteins. Fusion of the MS2 proteins to additional activators such as p65 and the human heat shock factor 1 (HSF1) results in the recruitment of 13 activation molecules per dCas9 molecule (Figure 36a) (Konermann S. *et al.*, (2015)). This dCas9-SAM system can amplify gene expression from 10 to multiple thousand-fold (Figure 36b).

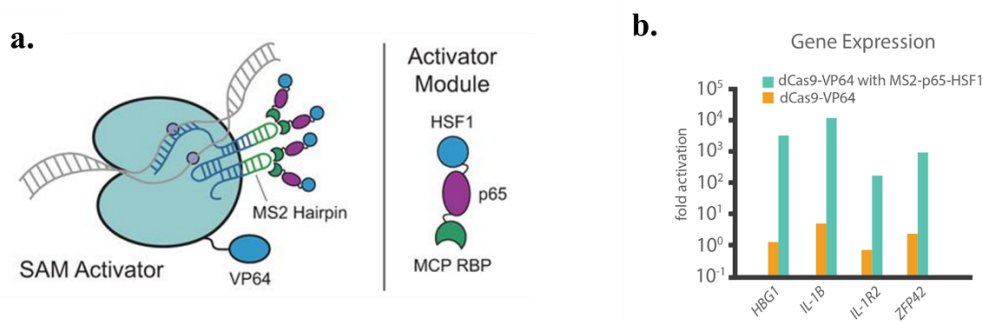


Figure 36 a. SAM activation system (left): dCas9 is fused to VP64 and the sgRNA has been modified so as it contains two MS2 hairpins (green). An additional activator module (right) binds to an MS2 hairpin via the RNA-binding protein MCP. The MCP is fused to the activators p65 and HSF1 (La Russa MF. *et al.*, (2015)). **b.** A comparison of the activation efficiency between dCas9-VP64 alone (in yellow) and the SAM activator system (in green) in the activation of four different genes: HBG1, IL-1B, IL1R2, and ZFP42 (<https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9>).

CRISPRi: dCas9-effector mediated transcriptional repression

CRISPR/dCas9 can be also combined with transcriptional repressors. As mentioned above, this was first demonstrated in bacterial cells, where dCas9 alone was able to act as a transcriptional repressor by sterically hindering the transcriptional activity of RNA polymerase (Qi LS. *et al.*, (2013)). This simple CRISPRi system can affect up to 1,000-fold repression, efficiently knocking down gene expression in cells. Although this system works very well in bacteria, yeast, and other prokaryotic cells, it is less effective in mammalian cells. This is likely because the binding of dCas9 to DNA is not sufficient to disrupt the action of eukaryotic RNA polymerases. A strategy to overcome this issue in mammalian cells has been to fuse the transcriptional repressor domain of Kox1 KRAB (Krüppel-associated box) to dCas9 (Gilbert LA. *et al.*, (2013)). This system is based on the ability of KRAB to recruit various types of histone modifiers that reversibly suppresses gene expression through the formation of heterochromatin (Figure 37a). CRISPR-repressor has been used in PSCs to probe the pluripotency network required to maintain the stem cell state. Targeting of regions upstream of the transcriptional start sites of *Oct4* and another pluripotency factor, *Tbx3*, using dCas9-KRAB in mouse embryonic stem cells (mESCs) effectively repressed their expression, resulting in spontaneous differentiation (Kearns NA. *et al.*, (2015)). Similarly, both dCas9 and dCas9-KRAB proteins could repress endogenous expression of CD71 and CXCR4 gene expression in HeLa cells (Figure 37b) (Gilbert LA. *et al.*, (2013)). Through CRISPR/dCas9-effector fusion mediated

transcriptional repression of pluripotency factors; these studies demonstrate the capability of the system to define the contribution of specific genes to a given cellular state.

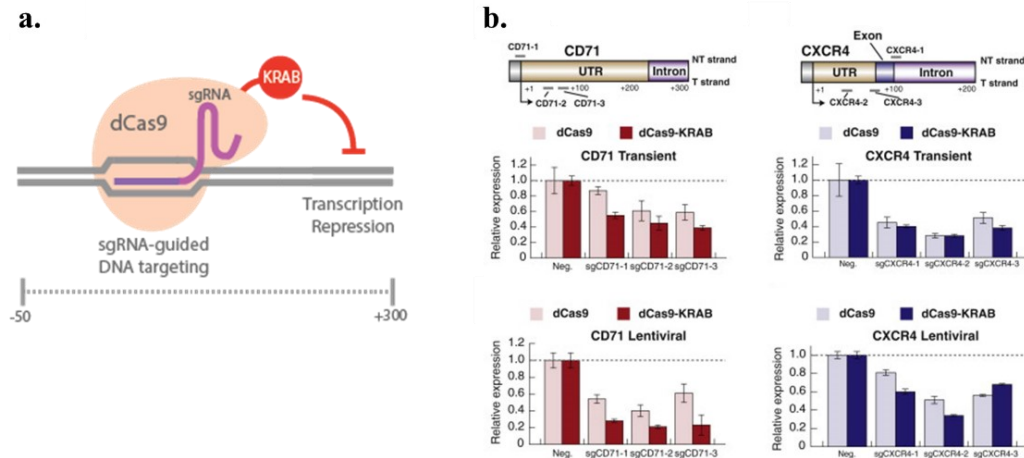


Figure 37 a. dCas9 fused to KRAB, a transcriptional repressor (<https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9>). b. Stable suppression of CD71 and CXCR4 gene expression by dCas9 or dCas9-KRAB in HeLa cells (Gilbert LA. *et al.*, (2013)).

dCas9-effector mediated epigenetic editing for activation and repression of gene expression

Epigenetic regulation works by affecting the structure of the chromatin, either by compressing it into a compact and transcriptionally inactive state (heterochromatin) or by opening it for active expression. Recently, fusion of dCas9 to various epigenetic modifiers has given a powerful tool to specifically target genomic sequences and modify local histone marks in order to study their effect on gene expression (Table 1).

	Construct	Function	Gene expression
Histone Modifications	dCas9-p300	Acetylation	Activation
	dCas9-LSD1	Demethylation	Repression
DNA Methylation	dCas9-TET1CD	Demethylation	Activation
	dCas9-DNMT3A	Methylation	Repression

Table 1 dCas9-epigenetic modifiers system.

Fusion of dCas9 to the catalytic core of the transcription activator acetyltransferase p300 (dCas9-p300^{CORE}) has been demonstrated to activate genes in human cells. Notably, the fusion protein catalyzes acetylation of histone H3 lysine 27 at its target sites, leading to robust transcriptional activation of target genes from promoters and both proximal and distal enhancers. When compared to dCas9-VP64, the dCas9-p300^{CORE} leads to higher levels of gene activation of endogenous genes, including *IL1RN*, *MYOD* and *OCT4*. Interestingly, a dCas9 fused with an inactive form of p300 was unable to activate target genes, indicating a crucial role of acetyltransferase activity for gene activation Figure 38a-b) (Hilton IB. *et al.*, (2015)).

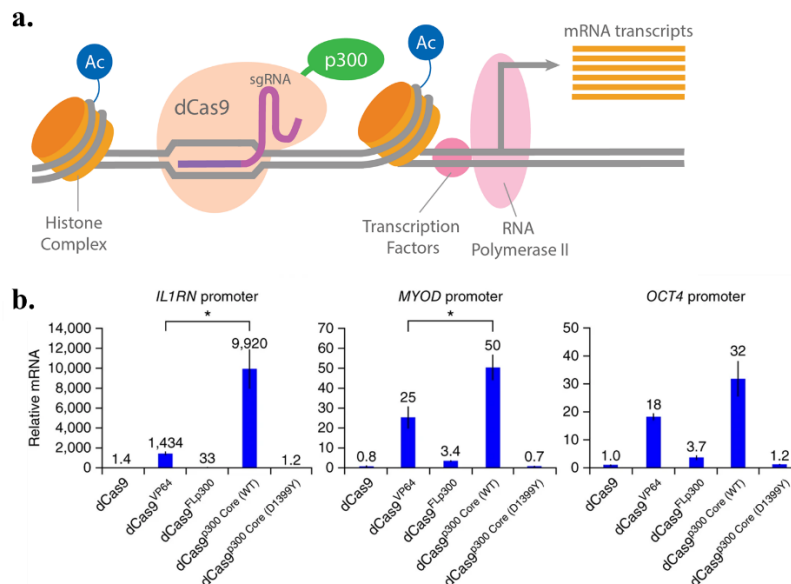


Figure 38 a. Schematic of dCas9-p300 system for epigenetic activation: dCas9-p300 acetylate target sites in the genome, resulting in transcriptional upregulation (<https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9>). **b.** mRNA expression level of *IL1RN*, *MYOD* and *OCT4*, using both dCas9 alone and dCas9 fused with different activators: VP64 activator, full-length p300 (dCas9^{FLp300}), HAT core domain of p300 (dCas9^{p300 Core}) and inactivated HAT core domain of p300 (dCas9^{p300 Core (D1399Y)}) (Hilton IB. *et al.*, (2015)).

dCas9-LSD1 is a complementary gene repressing strategy to the dCas9-p300 activating one. In this system dCas9 is fused to the histone demethylase LSD1 (KDM1A), which catalyze demethylation of H3K4 mono- and di-methylation, resulting in repression of target gene expression (Figure 39a). The histone demethylase LSD1 has been previously implicated in repression of enhancers and transcription activator-like effector (TALE)-LSD1 can target histone modifications that correlate with active enhancers (Mendenhall EM. *et al.*, (2013)). Kearns *et al.*, (2015) demonstrated that dCas9-LSD1 fusion protein reduces

endogenous gene expression in mESCs (that stably expressed dCas9-LSD1) when specifically targeted to active enhancer regions of pluripotency-associated genes, like *Oct4* and *Tbx3* (Figure 39b) (Kearns NA. *et al.*, (2015)). However, when dCas9-LSD1 was targeted at the *Oct4* promoter, no effect was observed. This suggests that the dCas9-LSD1 system can repress gene expression when targeted to only distal enhancers, while dCas9-KRAB repressed expression when targeted to promoters, proximal enhancers, and distal enhancers.

Based on these evidence, dCas9-LSD1 repressing system is able to functionally annotated cell type-specific enhancer elements that control cellular function in highly specific manner.

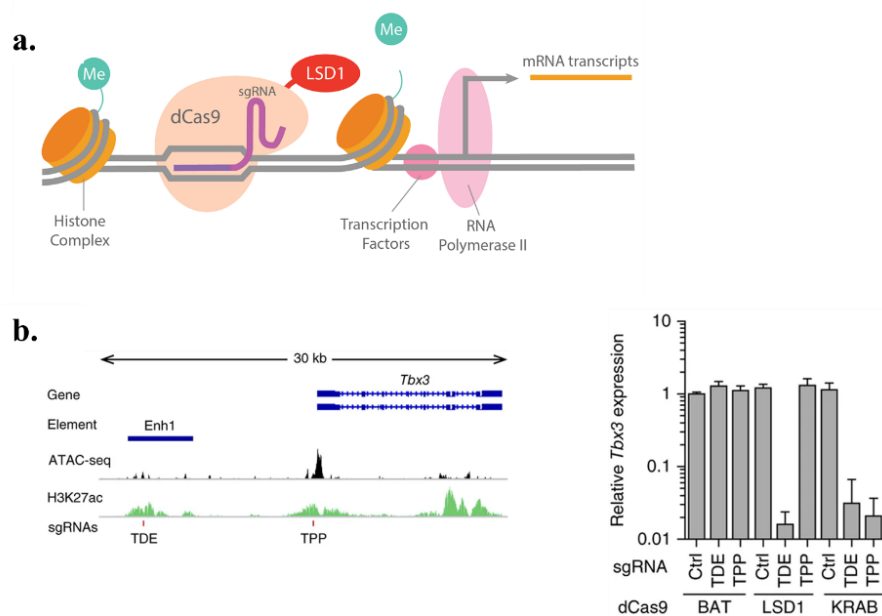


Figure 39 a. Epigenetic gene repression by dCas9-LSD1 system: dCas9 is fused to the Lysine-specific histone demethylase 1 (LSD1). Demethylation of mono- and dimethyl-group on histone H3K4 lead to genes transcriptional repression (<https://info.abmgood.com/crispr-cas9-gene-regulation-dCas9>). **b.** Left: Genomic organization of the targeted *Tbx3* locus. Right: Relative *Tbx3* expression in dCas9-repressor (LSD1 and KRAB) mESCs treated with sgRNAs specific to an unrelated control genomic region (Ctrl), the putative *Tbx3* distal enhancer (TDE) or the *Tbx3* promoter (TPP) (Kearns NA. *et al.*, (2015)).

Overall, CRISPR genome targeting, transcriptional engineering, and epigenome editing approaches can be used to mutate, activate, and repress individual genes, as well as to manipulate their epigenomic environment by editing DNA modifications, histone marks, and other chromatin features. Moreover, these technologies are applicable *in vivo* and can have strong potential for therapeutic approaches to alleviate disease states, or to manipulate disease-associated gene expression.

The overall aim of my doctoratal work is to understand better the genetic and epigenetic mechanisms responsible for cell fate transitions in multipotent cardiac progenitors to differentiate into endothelial cells (ECs).

In particular, I will address the following specific aims:

1. To develop a model for differentiation of cardiopharyngeal mesoderm (CPM) into ECs starting from engineered mouse embryonic stem cells (mESCs).

I will generate mutant mESCs and use a serum-free differentiation protocol with the addition of specific growth factors that induce cardiac and endothelial differentiation.

2. To identify putative enhancers regulated during cardiopharyngeal mesoderm (CPM) differentiation into endothelial cells (ECs).

I will perform RNA-seq to define the transcriptomic profile, and ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) to map chromatin accessibility genome-wide. I will use the integration of these two methods to identify putative enhancers defined as regions of increased accessibility associated with endothelial-specific genes.

3. To validate putative enhancers and test their requirement in EC differentiation.

To do this, I will use two validation strategies.

The first approach is based on deletion of the putative enhancer by CRISPR-Cas9 technology. I will generate the deletion in mESCs, and isolate stable mutant mESC clones. These will be differentiated in ECs and I will evaluate gene expression consequences.

The second one is using epigenetic reprogramming by nuclease-deficient dCas9 fused with histone demethylase LSD1 (dCas9-LSD1), that remove methylation of histone H3 lysine4 (H3K4me1 and me2) to de-commission the putative enhancers and to abrogate gene expression. I will generate stable mESC line expressing dCas9-LSD1, that will be transfected with a gRNA specific to the target region and differentiated, in order to evaluate the loss and/or repression of gene expression.

4. To predict, computationally, transcription factor motifs in EC enhancers.

I will perform a TF motif analysis of DARs region (Differential Accessible Regions) related to endothelial cell fate specification to identify the regulatory elements that are specifically enriched in my dataset. They could regulate the differentiation of cardiopharyngeal mesoderm progenitors in derivative tissues, including EC.

1. Cell biological methods

1.1 Mouse embryonic stem cells (mESC)

Mouse embryonic stem cells (mESC) derive from the inner cell mass of blastocyst. Due to their ability to differentiate in many cell types, mESCs are largely used in the research laboratories for *in vitro* studies and it is a very powerful tool used in development and stemness field. We have used the line ES-E14TG2a (ATCC CRL-1821), which were cultured without feeders and maintained undifferentiated on gelatin-coated dishes in GMEM (Sigma Cat# G5154) supplemented with 10^3 U/ml ESGRO LIF (Millipore, Cat# ESG1107), 15% fetal bovine serum (ES Screened Fetal Bovine Serum, US Euroclone Cat# CHA30070L), 0.1 mM non-essential amino acids (Gibco, Cat# 11140-035), 0.1 mM 2-mercaptoethanol (Gibco, Cat# 31350-010), 0.1 mM L-glutamine (Gibco, Cat# 25030081), 0.1 mM Penicillin/Streptomycin (Gibco, Cat# 10378016), and 0.1 mM sodium pyruvate (Gibco, Cat# 11360-070). The cells were passaged every 2–3 days using 0.25% Trypsin-EDTA (1X) (Gibco, Cat# 25200056) as the dissociation buffer.

1.2 In vitro differentiation protocol

E14-Tg2a mESCs were differentiated into the cardiopharyngeal mesoderm and endothelial lineage using a serum-free protocol (Kattman *et al.*, (2006); Patsch *et al.*, (2015)). mESCs were dissociated with Trypsin-EDTA and cultured at 75,000 cells/ml in serum-free media: 75% Iscove's modified Dulbecco's media (Cellgro Cat# 15-016-CV) and 25% HAM F12 media (Cellgro #10-080-CV), supplemented with N2 (GIBCO #17502048) and B27 (GIBCO #12587010) supplements, penicillin/streptomycin (GIBCO #10378016), 0.05% BSA (Invitrogen Cat#. P2489), L-glutamine (GIBCO #25030081), 5 mg/ml ascorbic acid (Sigma A4544) and 4.5×10^{-4} M monothioglycerol (Sigma M-6145). After 48h in culture, the EBs were dissociated using the Embryoid Body dissociation kit (cod. 130-096-348 Miltenyi Biotec) according to the manufacturer's protocol and reaggregated for 40h in serum-free differentiation media with the addition of 8 ng/ml human Activin A (R&D Systems Cat#. 338-AC), 0.5 ng/ml human BMP4 (R&D Systems Cat# 314-BP), and 5 ng/ml human VEGF (R&D Systems Cat#. 293-VE). The 2-day-old EBs were dissociated and 6×10^4 cells were seeded onto individual wells of a 24-well plate coated with 0.1% gelatin in EC Induction Medium consisting of StemPro-34 medium (Gibco #10639011), supplemented with SP34 supplement, L-glutamine, penicillin/streptomycin, 200 ng/ml human-VEGF, and 2 μ M forskolin (Abcam, ab120058). The Induction Medium was

renewed after one day. On day six of differentiation, the cells were dissociated and replated on gelatin 0.1% -coated dishes at a density of 25,000 cells/cm² in EC Expansion Medium, consisting of StemPro-34 supplemented with 50 ng/ml human-VEGF. Stem cell derived endothelial cells were maintained until they reached confluency (about 2-3 days). EC Expansion Medium was replaced every other day.

1.3 CRISPR-Cas9-Mediated Targeting of mESCs

1.3.1 Tbx1 knockout mESC line by CRISPR-Cas9 technology

Tbx1 knockout was induced in E14-Tg2a using Alt-R™ CRISPR-Cas9 System (IDT) following the manufacturer's specifications. This genome editing system is based on the use of a ribonucleoprotein (RNP) consisting of Alt-R S.p. Cas9 nuclease complexed with an Alt-R CRISPR-Cas9 guide RNA (crRNA:tracrRNA duplex). The crRNA is a custom synthesized sequence that is specific for the target (Tbx1KO:/AltR1/rUrG rGrCrC rGrArG rUrArC rArCrU rArCrC rArCrC rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/) and contains a 16 nt sequence that is complementary to the tracrRNA. Alt-R CRISPR-Cas9 tracrRNA-ATTO 550 (5 nmol catalog no. 1075927) is a conserved 67 nt RNA sequence that is required for complexing to the crRNA so as to form the guide RNA that is recognized by S.p. Cas9 (Alt-R S.p. Cas9 Nuclease 3NLS, 100 µg catalog no. 1081058). The fluorescently labeled tracrRNA with ATTO™ 550 fluorescent dye is used to FACS-purify transfected cells. The protocol involves three steps: (1) annealing of the crRNA and tracrRNA, (2) assembly of the Cas9 protein with the annealed crRNA and tracrRNAs, and (3) delivery of the ribonucleoprotein (RNP) complex into mESC by reverse transfection. Briefly, we annealed equimolar amounts of resuspended crRNA and tracrRNA to a final concentration (duplex) of 1 µM by heating at 95°C for 5 min and then cooling to room temperature. The RNA duplexes were then complexed with Alt-R S.p. Cas9 enzyme in OptiMEM media to form the RNP complex, which was then transfected into mESCs using the RNAiMAX transfection reagent (Invitrogen). After 48 h incubation, cells were trypsinized and ATTO 550 + (transfected) cells were purified by FACS. Fluorescent cells (approximately 65% of the total cell population) were plated at very low density to facilitate colony picking. We picked and screened by PCR 96 clones. Positive clones were confirmed by DNA sequencing.

PCR primers are:

Tbx1-KO FW: CTTCTGCCTTCTGCTCATGG

Tbx1-KO RV: CAGAGAAGGGTCGCCTACAT

1.3.2 Pecam1 intron2-enhancer deletion by CRISPR-Cas9 technology

Pecam1 intron2-enhancer deletion was induced in E14-Tg2a using Alt-R™ CRISPR-Cas9 System (IDT) following the manufacturer's specifications. This genome editing system is based on the use of a ribonucleoprotein (RNP) consisting of Alt-R S.p. Cas9 nuclease complexed with an Alt-R CRISPR-Cas9 guide RNA (crRNA:tracrRNA duplex). The crRNA is a custom synthesized sequence that is specific for the target (crRNA1:/AltR1/rCrU rGrUrC rUrCrC rArGrG rUrGrU rUrGrC rCrArA rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/; crRNA2:/AltR1/rGrU rUrCrC rGrArA rUrCrA rGrCrU rCrUrC rGrArG rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/) and contains a 16 nt sequence that is complementary to the tracrRNA. Alt-R CRISPR-Cas9 tracrRNA-ATTO 550 (5 nmol catalog no. 1075927) is a conserved 67 nt RNA sequence that is required for complexing to the crRNA so as to form the guide RNA that is recognized by S.p. Cas9 (Alt-R S.p. Cas9 Nuclease 3NLS, 100 µg catalog no. 1081058). The fluorescently labeled tracrRNA with ATTO™ 550 fluorescent dye is used to FACS-purify transfected cells. The protocol involves three steps: (1) annealing of the crRNA and tracrRNA, (2) assembly of the Cas9 protein with the annealed crRNA and tracrRNAs, and (3) delivery of the ribonucleoprotein (RNP) complex into mESC by reverse transfection. Briefly, we annealed equimolar amounts of resuspended crRNA and tracrRNA to a final concentration (duplex) of 1 µM by heating at 95°C for 5 min and then cooling to room temperature. The RNA duplexes were then complexed with Alt-R S.p. Cas9 enzyme in OptiMEM media to form the RNP complex, which was then transfected into mESCs using the RNAiMAX transfection reagent (Invitrogen). After 48 h incubation, cells were trypsinized and ATTO 550 + (transfected) cells were purified by FACS. Fluorescent cells (approximately 65% of the total cell population) were plated at very low density to facilitate colony picking. We picked and screened by PCR 96 clones. Positive clones were confirmed by DNA sequencing.

PCR primers are:

Pecam1-Δ enh.intr.2 (569 bp) FW: GGAGCGGAGGCCGTAGT

Pecam1-Δ enh.intr.2 (569 bp) RV: CCCTCTGTCCTAAGAGAGTAACA

1.3.3 Notch1 intron15-enhancer deletion by CRISPR-Cas9 technology

Notch1 intron15-enhancer deletion was induced in E14-Tg2a using Alt-R™ CRISPR-Cas9 System (IDT) following the manufacturer's specifications. This genome editing system is based on the use of a ribonucleoprotein (RNP) consisting of Alt-R S.p. Cas9 nuclease complexed with an Alt-R CRISPR-Cas9 guide RNA (crRNA:tracrRNA duplex). The crRNA is a custom synthesized sequence that is specific for the target (crRNA1: /AltR1/rArG rArGrU rCrArC rCrUrG rGrGrU

rUrArC rUrCrG rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/; crRNA2: /AltR1/rUrC rCrArG rUrGrU rUrCrC rUrArU rGrArU rGrCrC rGrUrU rUrUrA rGrArG rCrUrA rUrGrC rU/AltR2/) and contains a 16 nt sequence that is complementary to the tracrRNA. Alt-R CRISPR-Cas9 tracrRNA-ATTO 550 (5 nmol catalog no. 1075927) is a conserved 67 nt RNA sequence that is required for complexing to the crRNA so as to form the guide RNA that is recognized by S.p. Cas9 (Alt-R S.p. Cas9 Nuclease 3NLS, 100 µg catalog no. 1081058). The fluorescently labeled tracrRNA with ATTO™ 550 fluorescent dye is used to FACS-purify transfected cells. The protocol involves three steps: (1) annealing of the crRNA and tracrRNA, (2) assembly of the Cas9 protein with the annealed crRNA and tracrRNAs, and (3) delivery of the ribonucleoprotein (RNP) complex into mESC by reverse transfection. Briefly, we annealed equimolar amounts of resuspended crRNA and tracrRNA to a final concentration (duplex) of 1 µM by heating at 95°C for 5 min and then cooling to room temperature. The RNA duplexes were then complexed with Alt-R S.p. Cas9 enzyme in OptiMEM media to form the RNP complex, which was then transfected into mESCs using the RNAiMAX transfection reagent (Invitrogen). After 48 h incubation, cells were trypsinized and ATTO 550 + (transfected) cells were purified by FACS. Fluorescent cells (approximately 65% of the total cell population) were plated at very low density to facilitate colony picking. We picked and screened by PCR 96 clones. Positive clones were confirmed by DNA sequencing.

PCR primers are:

Notch1- Δ enh.intr.15 (664bp) FW: TGTGGCTACCCCAAGCATATC

Notch1- Δ enh.intr.15 (664bp) RV: AATGGCGAGAAATAGACCCCC

1.3.4 Generation of dCas9-LSD1 expressing mESC line by electroporation

20 µg of plasmid p-dCas9-LSD1-Hygro (kindly donated by Dr. Beck lab) was linearized with AhdI enzyme and electroporated in mESC (1×10^7 cells/10cm plate). The electroporation parameters used were 0.24V and 500 µF. The cells were maintained in Hygro-selection (500µg/ml) for 10 days. Individual colonies were isolated, expanded and screened by PCR for inserted sequence for both DNA and RNA.

PCR primers for DNA are:

CMVprom-dCas9 FW: GTAACAACTCCGCCCCATTG

CMVprom-dCas9 RV: TCGGTTATGACAGCCCATCC

LSD1-bGH polyA FW: CGTAACTACCCAGCCACAGT

LSD1-bGH polyA RV: GAGGGGCAAACAACAGATGG

HygrR FW: CCGTCAACCAAGCTCTGATAG

HygrR RV: GGCTCCAACAATGTCCTGAC

PCR primers for RNA are:

dCas9_start FW: GCAATGAGATGGCCAAAGTT
dCas9_start RV: TCCGAGTTGTCCGGATTTAG
dCas9_middle FW: AAGACCAACCGCAAAGTGAC
dCas9_middle RV: ATAAGTTTCCGCGACAATCG
dCas9_end FW: GCGAACAGGAGATAGGCAAG
dCas9_end RV: GGCCCTTATCCCATACGATT
dCas9_end-3XFLAG FW: TACTCTTACCAACCTCGGCG
dCas9_end-3XFLAG RV: ACTGTGAACTCGGTGGACAA

1.4 mESC reverse transfection protocol

Cells were plated at $5-8 \times 10^5$ per well in six-well plates and transfected with gRNA complex (crRNA:tracrRNA 10nM) in antibiotic-free medium using Lipofectamine RNA iMAX Reagent (Life Technology), according to the instructions. 24 hours after gRNA delivery, cells were harvested and processed. gRNA sequences are listed in Table 2.

Target	crRNA sequence (5'-3')
Non targeting	cccccgggggaataatttt
Kdr_int10-crRNA1	ttcaagtcacttagtcaa
Kdr_int10-crRNA2	acagaagatgatccgatgat
Kdr_int10-crRNA3	ttctgaaggacatctagact
VE-Cadh_int1-crRNA1	gataactgcccttctacact
VE-Cadh_int1-crRNA2	agttcttccccgagcaaatc
VE-Cadh_int1-crRNA3	tccttctaggacagacttgg
Eng_int2-crRNA1	tggtggcagccaagaacccc
Eng_int2-crRNA2	ccctcggcctgggcaaactc
Eng_int2-crRNA3	gttcttctctacctgaagag
Flt1_int10-crRNA1	accaaggctcaagccctagg
Flt1_int10-crRNA2	ggcgggaaggcgaacaacaag
Flt1_int10-crRNA3	gttcttctctgacacacag
Notch1_int15-crRNA1	agagtcacctgggttactcg
Notch1_int15-crRNA2	gttccgacaattgtgcaaa
Notch1_int15-crRNA3	tccagtgttctatgatgcc

Table 2 gRNA sequences.

1.5 Tube formation assay of mESC -ECs plated on Matrigel

To test the functionality of mESC-ECs *in vitro*, four hundred microlitres of Matrigel (BD Matrigel Basement Membrane Matrix Growth Factor Reduced, Phenol Red Free) was aliquoted into each well of a 12-well plate and incubated for 30–60 min at 37 °C to allow the gel to solidify. About 80-100,000 ECs were then added to the Matrigel-coated well and cultured for 24 h at 37 °C. Formation of tubular structures on a two-dimensional Matrigel surface was observed after 16 to 24h under an optical microscope.

1.6 Fluorescence Activated Cell Sorting (FACS)

1.6.1 Evaluation of endothelial cells (EC) from differentiated mESC

For flow cytometric analysis, we dissociated the cells with Trypsin-EDTA or with the Embryoid Body dissociation kit (cod. 130-096-348 Miltenyi Biotec). Dissociated cells (1×10^6 cells/100 μ l) were incubated with primary antibodies (CD144-APC, mouse cod.130-102-738) directly conjugated (1:10) in PBS-BE solution (PBS, 0.5%BSA, 5 mM EDTA) for 20 min on ice. Subsequently, cells were washed twice with 2 ml of PBS-BE. Cells were analyzed using the BD FACS ARIAIITM cell sorter. Negative controls were incubated with fluorochrome-labeled irrelevant isotype control antibody (REA Control APC, mouse cod. 130-113-446 Miltenyi Biotec).

1.6.2 gRNA transfection efficiency

To detect and visualize the fluorescently labeled gRNA complex (crRNA:tracrRNA-ATTOTM550), we dissociated mESC with Trypsin-EDTA 24h after transfection and analyze the % of fluorescence using the BD FACS ARIAIITM cell sorter. Cells containing the transfected gRNA complex were isolated and differentiated into ECs.

2. Molecular biological methods

2.1 Reverse transcription and PCR amplification (RT-PCR)

Total RNA was isolated from mouse ESCs with QIAzol lysis reagent (Qiagen #79306), according to the manufacturer's protocol. The isolated RNAs were quantified using a NanoDrop spectrophotometer 1000. Before reverse transcription, RNA samples were treated with DNase I to eliminate any contamination with genomic DNA (located in the interphase during extraction).

cDNA was transcribed using 1 or 2 µg total RNA with the High-Capacity cDNA reverse transcription kit (Applied Biosystem catalog. n. 4368814).

cDNAs were amplified using myTaq™ DNA polymerase (Meridian Bioscience) and a standard 3-step cycling PCR profile: 10 min at 94°C, 30 amplification cycles (denaturation at 94°C for 30sec, annealing at 60°C for 30sec, and extension at 72°C for 30sec), followed by a final extension at 72°C for 10 min. The amplified products were separated on agarose gels and visualized by ethidium bromide staining.

2.2 Quantitative real-time PCR (qRT-PCR)

Quantitative gene expression analyses (qRT-PCR) were performed using SYBR Green PCR master mix (Applied Biosystem). Relative gene expression was evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as normalizer (Livak KJ. *et al.*, (2001))

cDNA was amplified by qRT-PCR, using StepOnePlus™ Real-Time PCR System. The run used was similar to PCR default condition, but the number of cycles is increases up to 40 cycles. The cycle threshold (Ct) was determined during geometric phase of the PCR amplification plots, as illustrated in the manufacturer's protocol.

Primers are listed in Table 3. Expression data are shown as the mean ± SD.

2.3 Statistical Analysis

Graph Pad Prism software v8.00 (GraphPad) was used to analysis of data. Relative mRNA levels were analyzed in triplicate and data were presented as means ± SD.

Two-way repeated measures ANOVA test (ANOVA 2way-RM) was used to assess that there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a gene expression variable. Other two statistical methods between groups of data were used: nonparametric and parametric test. The first was nonparametric Wilcoxon matched-pairs signed rank test; the second statistical analysis were performed using the parametric Student's paired t- test. Differences were considered significant at p-value < 0.05. Shapiro-Wilk test was performed to determine the normality distribution of the dataset.

2.4 ATAC-seq

ATAC-seq assay on day2 and day4 mESCs was performed and sequencing library was prepared from the fragmented amplified tagmented DNA. Two biological replicates for each condition were sequenced using Illumina

NextSeq500 system to obtain paired-end (PE) reads of 60bp. The read quality was assessed using FastQC and multiQC before and after trimming for adaptors and low-quality reads with Cutadapt :-m 30, -q 30 (Martin M. *et al.*, (2011)). The PE reads were aligned to mm10 mouse reference genome using Bowtie2 (-q -t --end-to-end --very-sensitive -X 1000) (Langmead and Salzberg (2012)) and retain mapped reads with MAPQ < 20. PCR duplicates were marked (PICARD markDuplicates) and reads mapped to chrM were removed. BEDPE files were filtered using customized Rscript to remove mates mapping to different chromosomes or at distance > 2kb and discordant orientation of read-pairs. The ATAC-seq peaks were called for each replicate of day2 and day4 vs input with MACS2 using parameters --nomodel --shift 100 --extsize 200 -q 0.001 (Feng J.*et al.*, (2012)). The consensus peaks from replicates were identified and the blacklisted regions were removed. The Differential Accessible Regions (DARs) between day4 vs day2 were obtained using DEScan2 1.6.0 (Righelli D. *et al.*, (2018)) and using DiffBind (Stark *et al.*, (2012)) with FDR < 0.01. The DARs common to both analyses were retained. The consensus peaks and common DARs were annotated using ChIPseeker (annotatePeak) with parameter tssRegion: -1000,1000 (Yu G. *et al.*, (2015)). SE BED files from the BEDPE were used to generate BigWig and uploaded in Cyverse FTP and the coverage was visualized along with ATAC-seq peaks and DAR tracks in UCSC browser.

2.5 RNA-seq

The total RNA was isolated from day2 and day4 cells and sequenced using Illumina platform NextSeq 500 as PE reads of length 75bp. The read quality was assessed using multiQC prior and after quality and adaptor trimming (Cutadapt). The PE reads were aligned to mm10 genome using STAR aligner default parameters (Dobin A. *et al.*, (2013)). The read counts were obtained using featureCounts function from Rsubread package (strandSpecific=2, CountMultiMappingReads=FALSE). The expressed genes and DE genes list from day4 vs day2 were obtained using DEseq2 with FDR < 0.01 (Love MI. *et al.*, (2014)) and NOIseq applying posterior probability > 0.9 (Tarazona S. *et al.*, (2015)). The commonly expressed and DE genes from both analyses were retained.

Gene	Primer FW	Primer RV
Pecam1	tggttgctattggagtggc	ttctcgctgttgagttcag
Kdr	gtcgacatagcctccactgtt	gtgatgtacacgatgccatgct
Nos3	tctaccggcacgaggtactg	aggtcttgccactaggtcttg
VE-Cadherin	tcatcaaaccacgaagtcc	ggctctgtggcctcaatgtaga
Eng	gtgttctctggctctctgttc	tcttggctgtccttggaaga
CD-34	ctgcctggaactaagtgaagc	agcctctctcttttcacaca
Notch1	ctccgttacatgcagcagtt	ccaggatcagtgagggtgtg
Scf1	attgcacacacgggattctg	gaattcagggtcttcttag
Dusp5	gatcgaaggcgagagaagc	ggaagggaaggattcaacc
Gata6	ggctctacagcaagatgaatgg	tggcacaggacagccaag
Flt1	acctcgtgcatgtgtatga	catggacagccgataggac
Mesp1	gcgacatgctggctcttcta	tggatcactgccgcctcttc
Gata4	ctgtcatctcactatgggca	ccaagtccgagcaggaattt
Brachyury	gaacctcggattcacatcgt	ttcttggcatcaaggaagg
Isl1	gcctcagtcacagagtcac	agagcctggctctctctg
Oct3/4	tcagctgggctagagaagg	tgacgggaacagagggaaag
Tbx1	ttgtgcccgtagatgacaa	aatcggggctgatatctgtg
Pdgfra	ctggctgctgctctatgac	cacgatcgtttctctgccttat
Nanog	aagtacctcagcctccagca	gtgctgagcccttctgaatc
Gapdh	tgcaccaccaactgcttagc	tcttctgggtggcagtgatg
Tbx1-KO	cttctgccttctgctcatgg	cagagaagggtgcctacat
Notch1- Δenh. intron15	tgtggctaccccaagcatatc	aatggcgagaaatagaccccc
Pecam1- Δ enh. intron2	ggagcggaggccgtagt	ccctctgtcctaagagagtaaca
Tbx1-KO	cttctgccttctgctcatgg	cagagaagggtgcctacat

Table 3 Primer list used for RT-PCR and qRT-PCR.

2.6 Integrated analysis

The gene list from RNA-seq and DAR-annotated genes were visualized using Volcano plot. A subset of genes enriched in Angiogenesis and Endothelial differentiation- GO terms and genes expressed in cluster 7 of the single cell study on Tbx1 expressing E9.5 mouse embryos (Nomaru H. *et al.*, (2021)) were selected for Transcription Factor motif enrichment analysis. ATAC-seq peaks/DARs and expressed/DE genes obtained from day4 vs day2 study was subsetted for the above gene list of interest and motif enrichment study (Hypergeometric Optimization of Motif EnRichment: HOMER) was performed to identify transcriptional and epigenetic markers distinguishing day4 and day2 of the directed EC differentiation.

Chapter 1.

Establishment of a model for differentiation of cardiopharyngeal mesoderm (CPM) into endothelial cells (ECs) starting from mouse embryonic stem cells (mESCs).

For my experiments, I decided to use mouse embryonic stem cells (mESC). Due to their ability to differentiate in many cell types, mESCs are largely used in the research laboratories for *in vitro* studies and it is a very powerful tool for genetic disease and obviously are extremely used in development and stemness field. These undifferentiated cells can be induced to differentiate into cardiac mesodermal progenitors (Kattman *et al.*, (2011)) and endothelial cells (Patsch (2015)).

For this reason, I sought to develop and optimize an appropriate and well-defined mESC differentiation model, that allowed me to efficiently obtain cardiac mesoderm progenitors (CPM) and endothelial cells (EC) in the laboratory.

Studies in literature indicated that is necessary to define and control the signaling pathways that regulate the specification of the cardiovascular lineages during embryonic development.

Previous reports have shown that Activin/Nodal, BMP4 and VEGF signaling pathways have an essential role on the induction of cardiovascular mesoderm in mESC differentiation cultures (Laflamme *et al.*, (2007); Yang *et al.*, (2008)).

Studies in the mouse embryo and the mouse embryonic stem cell differentiation model identified *Mespl*-expressing cardiopharyngeal mesoderm (CPM) cells progenitors that were specified during gastrulation. *Mespl* is a mesoderm specific marker gene and *Mespl*⁺ CPM differentiate into cardiomyocytes (CMs), endothelial cells (ECs) and branchiomeric muscle (Lescroart *et al.*, (2018)).

In this scenario, looking for the most suitable differentiation protocols in literature, I understood that when mESC were differentiated by supplying a cocktail of specific growth factors, like Activin A, bone morphogenetic protein 4 (BMP4), vascular endothelial growth factor (VEGF, also known as VEGF-A) in serum-free media, at the appropriate stage of development, these cells can form cardiomyocyte, endothelial cells and vascular smooth muscle cells *in vitro*.

Based on this evidence, I have chosen two differentiation methods: the first, described in Kattman *et al.*, 2011, which generates differentiated cell populations highly enriched for cardiomyocytes. Here, I indicated it as “Cardiac protocol”. The second, described in Patsch *et al.*, 2015, allows a more specific endothelial cell differentiation. It produces approximately 90% mature ECs, using a treatment with high concentration of VEGF-A.

I compared the two protocols and found that the early common step was the cardiopharyngeal mesoderm induction, using a combination of BMP-4, Activin-A and VEGF-A growth factors. So, I decided to combine the procedures and develop an “hybrid” differentiation model, named as “EC protocol”.

In detail, the protocol used during my doctoral work consists of three steps:

- First, mESCs were plated in suspension as single cells in media without any factors to induce embryoid bodies (EBs) aggregation (Figure 40);
- Second, BMP4 and Activin A were added to specify cardiac mesoderm lineage and VEGF-A to promote the formation of cardiovascular progenitors;
- Third, cells were plated in adhesion plates and differentiated into a homogeneous endothelial cell population using combined VEGF-A treatment (high concentration) with forskolin for two days. Forskolin is a cyclic-AMP signaling activator and data in literature demonstrated that cAMP enhance the vascular development (Yamamizu *et al.*, (2009)) Then, the cells were maintained in media with VEGF-A alone for an additional two days (Figure 40.b).

The Figure 41 illustrates the differentiation scheme.

Examples of differentiating EBs and endothelial-like cells are showed in Figure 40.a and Figure 40.b.

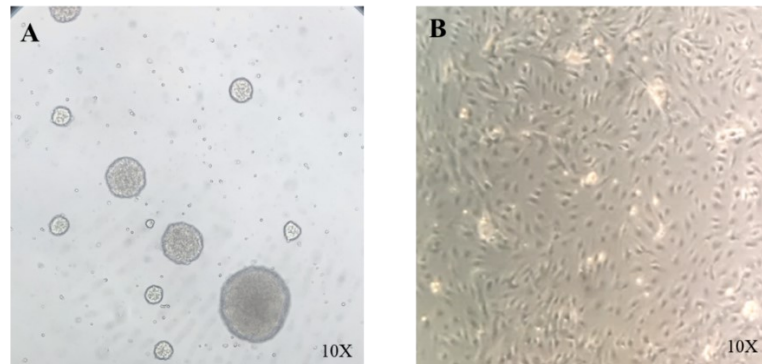


Figure 40.a Representative pictures of differentiating EBs between d2 and d4;

Figure 40.b Representative example of ECs-like in adhesion at day8 of differentiation.

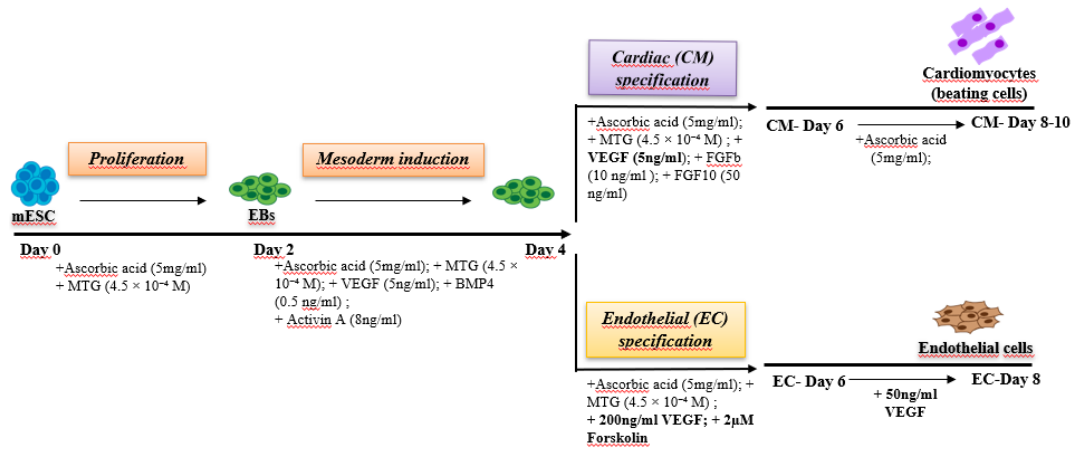


Figure 41 Schematic illustration of both Cardiac (CM) / Endothelial (EC) differentiation strategies used for mESC.

1.1 Characterization of mESC-derived differentiated cells

After establishment of a defined differentiation method, to assess the efficient differentiation of mesodermal progenitors to endothelial cells, I performed gene expression profiling at different time points during the procedure, as well as flow cytometric analysis and functional tests.

I harvested the cells during at days 0-2-4-6-8-10 of differentiation and extracted total RNA to analyze the expression of several endothelial as well as pluripotency and mesodermal marker genes involved in the vascular lineage commitment by RT-PCR (Figure 42).

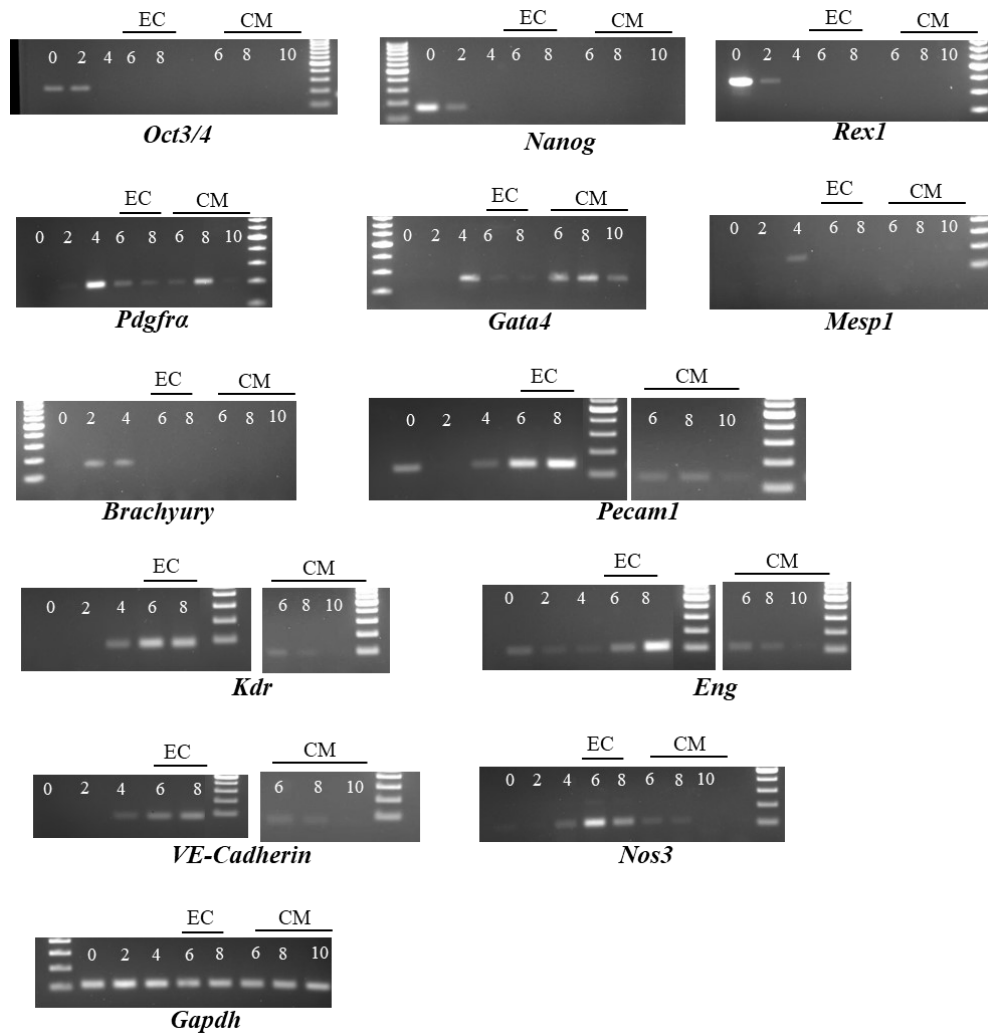


Figure 42 Expression of pluripotency specific markers (*Nanog*, *Oct3/4*, *Rex1*), mesodermal (*Mesp1*, *Brachyury*, *Gata4*, *Pdgfra*) and endothelial genes (*VE-Cadherin*, *Pecam1*, *Eng*, *Kdr*, *Nos3*) during CPM/EC differentiation by RT-PCR. *Gapdh* is used as normalizer. Marker used 100bp.

As Figure 42 shown, pluripotent cell marker genes such as *Nanog*, *Oct3/4* and *Rex1* were rapidly downregulated. Mesodermal genes, such as *Mesp1*, *Brachyury*, *Gata4*, *Pdgfra*, were activated between day2 and day4.

The expression of endothelial genes like *VE-Cadherin*, *Pecam1*, *Endoglin (Eng)*, *Kdr (Vegfr2)*, *Nos3*, were mostly activated at day4 and increased progressively during differentiation, starting from day 4 (d4) up until day 8 (d8). In addition, EC related markers were expressed at high level when I used the “EC protocol” than the “cardiac” one, as expected.

To confirm endothelial marker expression seen at mRNA level, I also evaluated the expression of VE-Cadherin (CD144) protein at day4-6-8 by flow cytometric analysis in both protocols.

I have observed that the percentage of VE-Cadherin⁺ cells increased progressively during differentiation, starting from day 4 (20%) up until day 8. More precisely, at d6 58.3% of total cell population in “EC protocol” was CD144⁺ (VE-Cadherin), compared to 39% in “Cardiac” conditions. Similarly, EC protocol induced a higher % of VE-Cadherin⁺ cells (91.3%) than the other one (58.7%) at d8. Data shown in Figure 43).

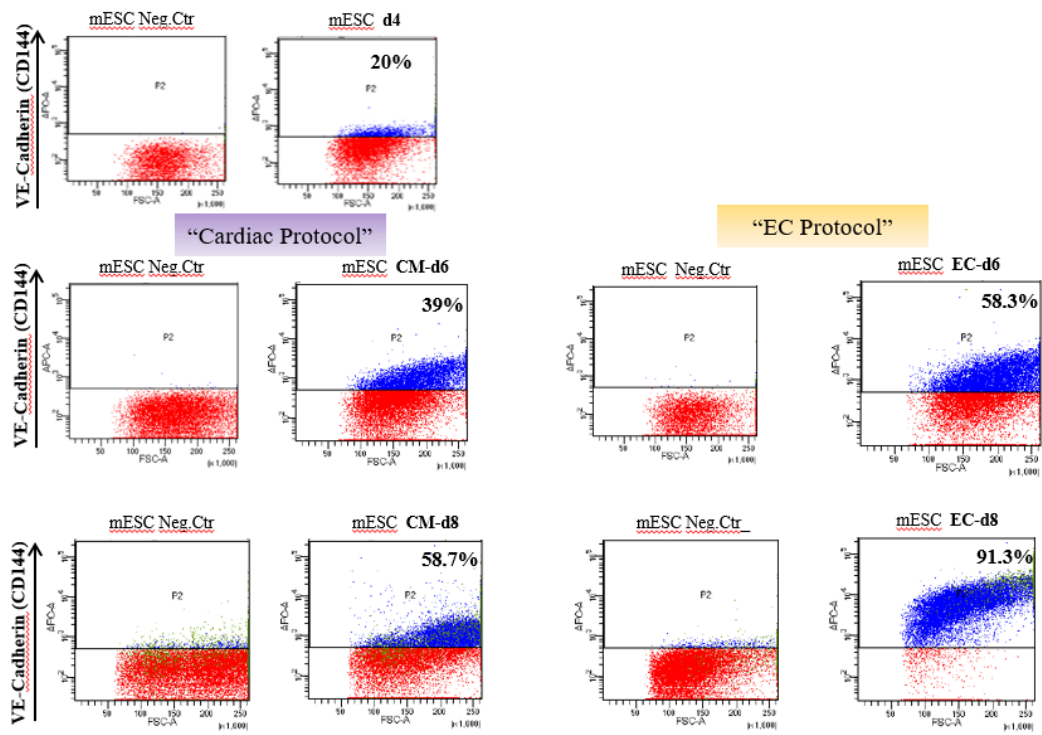


Figure 43 Representative plot of the gating strategy used for immunophenotyping of cells during mES differentiation. The VE-Cadherin⁺ (CD144⁺) subpopulation was identified at day 4-6-8 of differentiation by FACS using anti-CD144 antibody. Negative control is isotype control antibody-labeled differentiating cells.

1.2 mESC-derived endothelial cells form vascular network-like structures in vitro

Functional endothelial cells are able to build capillary-like structures when plated on a basement membrane matrix *in vitro* (Arnaoutova and Kleinman *et al.*, (2010)). To assess the ability of cells to form capillary-like structures, I performed a tube formation assay by plating mESC-derived ECs at d8 on Matrigel and observed the formation of vascular network-like structures after 16h to 24h. The results, showed in Figure 44, suggest that mESC-derived ECs can arrange in highly organized vessel-like structures and demonstrate that the cells have angiogenic potential *in vitro*.

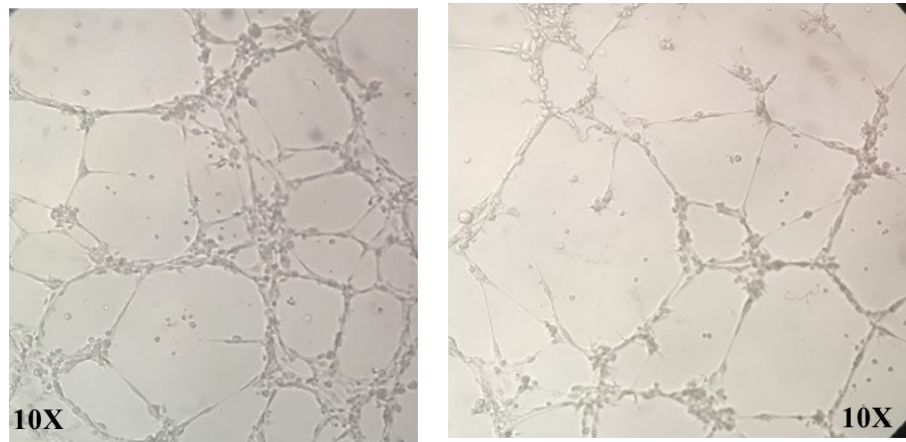


Figure 44 *In vitro* tube formation assay of mESC ECs plated on Matrigel for 24h.

These data demonstrate that, starting from mESC, the experimental model and methods used for differentiation of CPM into ECs allow me to efficiently produce differentiated endothelial cells (~91% VE-Cadherin⁺ cells), as assessed by marker gene expression and *in vitro* functional test.

Moreover, because of the goal of my doctoral work is to investigate the early phases of multipotent cardiac progenitors' differentiation into ECs, I decided to perform subsequent analyses between the critical time window day2-day4, based on the previous results described in Chapter 1.

Chapter 2.

Transcriptomic and chromatin accessibility profile in early phases of multipotent cardiac progenitors' differentiation into mature ECs.

To investigate the transcriptional profile of the early stages of CPM-EC differentiation, I decided to perform RNA-seq analysis between d2-d4. We have also defined chromatin state accessibility at the same time points during differentiation, using ATAC-seq assay, to provide insights into the coordinated regulation of gene expression programs.

2.1 Gene expression profile of multipotent cardiac progenitors at day2-day4 (RNA-seq analysis)

I have collected differentiating mESCs at d2 and d4 and processed them for RNAseq assay (two biological replicates for each condition). In collaboration with other colleagues in the laboratory, we carried out the bioinformatics analyses. The overall number of four paired-end reads from d2 and d4 differentiated mESC is illustrated in the Table 4.

Sample pairs	N° of raw reads
d2_rep1	15,060,453
d2_rep2	18,042,950
d4_rep1	14,104,249
d4_rep2	21,110,492

Table 4 Total number of raw reads in d2-d4 differentiated mESCs.

Next, we aligned RNA-seq paired-end reads to mouse genome (mm10 version), using STAR (Spliced Transcripts Alignment to a Reference) aligner tool, identifying both multi-mapped and unique mapped reads.

The Table 5 shows the percentage of alignment and the total number of aligned reads. Moreover, the quality check of aligned BAM reads is illustrated in Figure 45.

Sample	Input read pairs	Alignment %				Aligned reads
		Aligned	Unique	Multi	None	
d2_rep1	15,060,453	97.70%	87.42	10.28	2.31	14,714,063
d2_rep2	18,042,950	97.41%	85.44	11.97	1.30	17,575,638
d4_rep1	14,104,249	97.31%	88.60	8.71	3.65	13,724,846
d4_rep2	21,110,492	97.44%	88.25	9.19	2.56	20,561,619

Table 5 Number of aligned reads of 2 replicates d2 and 2 replicates d4.

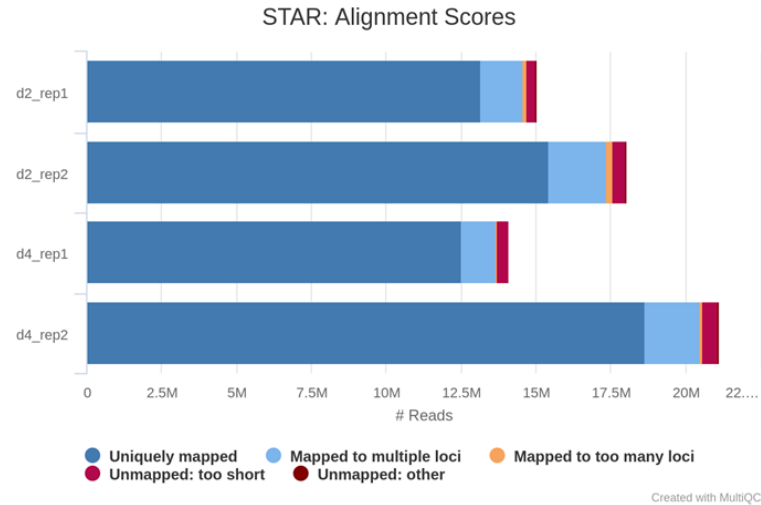


Figure 45 Alignment scores by STAR tool.

In Figure 46 is reported the MA plot, which shows the log average (A) in the x-axis and the log ratio (M) on the y-axis. This type of plot is used to represent the dataset distribution between the two conditions d2 and d4. It shows that few genes were highly expressed in d4 compared to d2 and vice versa and they were those with log2FC close to 10.

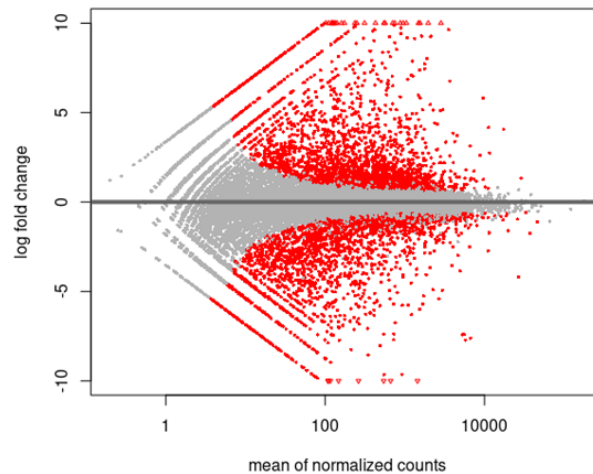


Figure 46 MA plot of all expressed genes dataset d2-d4. Log ratio (M) is log fold-change (y-axis); log average (A) is log mean of normalized counts (x-axis).

Auer PL. *et al.*, (2010) clearly demonstrated that it is important to have replications in designing RNA-seq experiments and understand the level of variation in the data. In practice, many RNA-seq experiments are done without or with only a few replicates. In order to accommodate such situations, Zheng X. *et al.*, (2013) showed that it is possible to obtain improved differential gene-calling results by combining the results obtained by two representative methods: one parametric, DESeq, and one nonparametric, NOISeq. They suggested strategies to use these two methods individually or combined according to the characteristics of the data. Both methods showed slightly but consistently higher sensitivities for the over-expressed genes than for the under-expressed genes. Over-expressed genes were slightly more likely to be called correctly as differentially expressed than under-expressed genes. Moreover, DESeq showed length-dependent results where longer transcripts were called more as differentially expressed, whereas NOISeq did not show this trend. It indicates that for longer genes, DESeq calls more genes as differentially expressed, but their results include more false positives. In contrast, NOISeq calls a smaller number of genes as positives, but with very high accuracy, regardless of the lengths.

For this reason, since the number of biological replicates (n.=2) was too small, we have determined differentially expressed genes (DEGs) between d2 and d4, by combining two representative differential gene-calling methods: one parametric method, DESeq2 (applying a threshold of p-adjust < 0.05 and p-adjust < 0.01), and one nonparametric method, NOISeq (posterior probability threshold > 0.8 and > 0.9).

The total number of expressed genes analyzed in d2 and d4 datasets is 13,841.

The Table 6 summarizes the results obtained using DESeq2 applying a threshold of p-adj < 0.05 and p-adj < 0.01; while the Table 7 shows DE analysis using NOISeq tool.

Comparison (DESeq2)	N° of DE genes (total)	N° up regulated genes	N° down regulated genes
d4 vs d2 DE genes (padj < 0.05)	5024	2618	2406
d4 vs d2 DE genes (padj < 0.01)	3790	2027	1763

Table 6 d4 vs d2 DE analysis using DESeq2 applying a threshold of P-adjust < 0.05 and < 0.01.

Comparison (NOISeq)	N° of DE genes	N° up-DE genes	N° down-DE genes
d4 vs d2 DE genes (pp>0.8)	2326	1139	1177
d4 vs d2 DE genes (pp>0.9)	970	512	458

Table 7 d4 vs d2 DE analysis using NOISeq (Posterior probability threshold 0.8 and 0.9).

To perform a stringent DEG analysis, we considered the threshold combination DEseq2 $\text{padj} < 0.01$ and NOIseq $\text{pp} > 0.8$.

By the intersection of DE genes in Deseq2 and NOIseq, we found that there were 1735 DEGs between d2-d4. In particular, 1057 genes were upregulated at d4 compared to d2, while 678 genes were downregulated, as illustrated in the Figure 47.

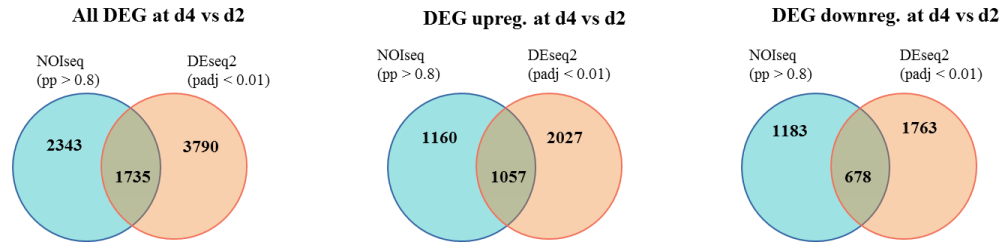
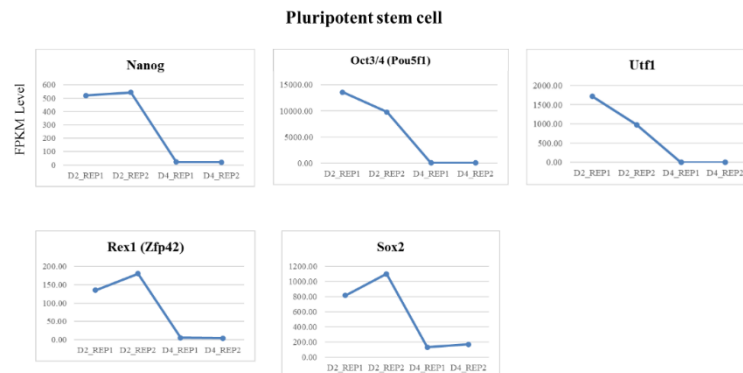


Figure 47 Intersection of DE genes at d4 vs d2 in Deseq2 and NOIseq methods.

Then, to confirm what I have seen by RT-PCR in Figure 42 (Chap.1), I plotted a time course of gene expression for representative genes from pluripotent cells, cardiopharyngeal mesoderm and endothelial cell differentiation (Figure 48).

As previously described, these analyses revealed that pluripotent cell markers such as *Nanog*, *Utf1*, *Sox2*, *Rex1*, *Oct3/4* were rapidly down regulated during differentiation. Next, I observed a higher expression level of the mesoderm-specific genes, like *Brachyury (T)*, *Eomes*, *Mixl1*, *Mef2c*, *Pdgfra* and *Gata4* at d4 compared to d2. Similarly, several well-known endothelial marker genes were highly expressed at day 4 (e.g., *Kdr*, *Pecam1*, *VE-Cadherin*, *Eng*, *Tal1*, *CD-34*, *Gata6*, *Gata1*, *Gata2*, *Flt1*, *Nos3*, *Ets1*, *Etv2*, *Vwf*).

These data support the idea that the endothelial lineage commitment is induced during the time window d2-d4.



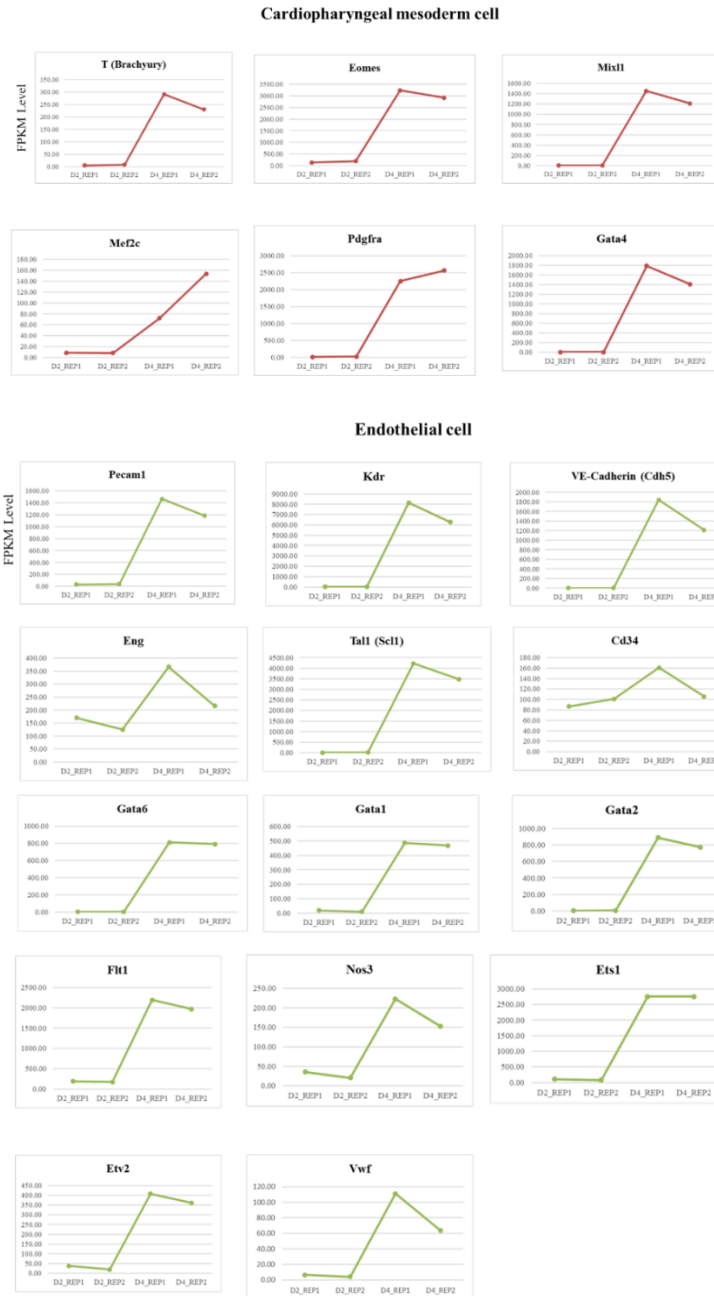


Figure 48 Gene expression evaluation of representative marker genes of pluripotent stem cells, cardiopharyngeal mesoderm and endothelial cells. The results were indicated as FPKM average of two biological replicates.

We performed a functional enrichment analysis to find within all 1735 DE genes the most over-represented Gene Ontology terms, related to biological processes (GO:BP). We used g:Profiler tool (Raudvere *et al.*, (2019)). The most enriched BP and Pathways terms are indicated in Figure 49.

I found that the most significant over-represented biological processes for differentiating mESCs d2-d4 included terms “developmental process”, “cellular developmental process”, “regulation of developmental process”, as well as “circulatory system development”, “vasculature development”, “blood vessel development”, “cell adhesion”, “blood vessel morphogenesis”, “regulation of cell migration”, “angiogenesis”, “embryo development” and others.

GO Biological Processes

GO:BP	stats	
Term name	Term ID	Padj
anatomical structure morphogenesis	GO:0009653	5.338×10 ⁻⁶⁰
system development	GO:0048731	1.491×10 ⁻⁵⁶
multicellular organism development	GO:0007275	7.813×10 ⁻⁵⁶
anatomical structure development	GO:0048856	1.245×10 ⁻⁵⁵
developmental process	GO:0032502	9.568×10 ⁻⁵⁵
cellular developmental process	GO:0048869	1.197×10 ⁻⁴⁷
cell differentiation	GO:0030154	1.032×10 ⁻⁴⁶
animal organ development	GO:0048513	2.903×10 ⁻⁴⁴
tube development	GO:0035295	3.389×10 ⁻⁴²
anatomical structure formation involved in morphogenesis	GO:0048646	2.982×10 ⁻⁴¹
regulation of developmental process	GO:0050793	4.004×10 ⁻⁴¹
positive regulation of cellular process	GO:0048522	6.126×10 ⁻⁴¹
circulatory system development	GO:0072359	6.635×10 ⁻⁴¹
positive regulation of biological process	GO:0048518	2.300×10 ⁻⁴⁰
tube morphogenesis	GO:0035239	9.517×10 ⁻⁴⁰
vasculature development	GO:0001944	1.722×10 ⁻³⁶
blood vessel development	GO:0001568	1.820×10 ⁻³⁵
negative regulation of cellular process	GO:0048523	3.196×10 ⁻³⁵
tissue development	GO:0009888	5.227×10 ⁻³⁵
negative regulation of biological process	GO:0048519	7.256×10 ⁻³³
regulation of multicellular organismal process	GO:0051239	9.215×10 ⁻³³
locomotion	GO:0040011	1.450×10 ⁻³²
biological adhesion	GO:0022610	1.542×10 ⁻³²
cell adhesion	GO:0007155	1.934×10 ⁻³²
blood vessel morphogenesis	GO:0048514	4.701×10 ⁻³²
movement of cell or subcellular component	GO:0006928	7.147×10 ⁻³²

cell morphogenesis	GO:0000902	1.222×10 ⁻³¹
regulation of cell differentiation	GO:0045595	2.531×10 ⁻³¹
regulation of multicellular organismal development	GO:2000026	4.365×10 ⁻³¹
regulation of cellular component movement	GO:0051270	1.899×10 ⁻³⁰
cell migration	GO:0016477	2.177×10 ⁻³⁰
cellular component organization	GO:0016043	2.223×10 ⁻³⁰
positive regulation of developmental process	GO:0051094	2.662×10 ⁻³⁰
regulation of cell motility	GO:2000145	5.979×10 ⁻³⁰
regulation of locomotion	GO:0040012	8.360×10 ⁻³⁰
nervous system development	GO:0007399	8.913×10 ⁻³⁰
cellular component organization or biogenesis	GO:0071840	1.145×10 ⁻²⁹
regulation of anatomical structure morphogenesis	GO:0022603	2.384×10 ⁻²⁹
regulation of cell migration	GO:0030334	1.389×10 ⁻²⁸
cell motility	GO:0048870	1.509×10 ⁻²⁸
localization of cell	GO:0051674	1.509×10 ⁻²⁸
embryo development	GO:0009790	2.765×10 ⁻²⁸
angiogenesis	GO:0001525	2.969×10 ⁻²⁸
regulation of cell population proliferation	GO:0042127	4.106×10 ⁻²⁸
cell surface receptor signaling pathway	GO:0007166	5.438×10 ⁻²⁸
cell population proliferation	GO:0008283	8.131×10 ⁻²⁸
cell development	GO:0048468	1.480×10 ⁻²⁷
positive regulation of cell differentiation	GO:0045597	2.688×10 ⁻²⁷
enzyme linked receptor protein signaling pathway	GO:0007167	3.208×10 ⁻²⁷
regulation of signaling	GO:0023051	1.189×10 ⁻²⁶
cellular response to chemical stimulus	GO:0070887	1.336×10 ⁻²⁶
regulation of cell communication	GO:0010646	1.556×10 ⁻²⁶
regulation of signal transduction	GO:0009966	1.980×10 ⁻²⁶
tissue morphogenesis	GO:0048729	3.415×10 ⁻²⁶

KEGG Pathways

KEGG	stats	
Term name	Term ID	Padj
Pathways in cancer	KEGG:05200	6.221×10 ⁻¹⁰
Focal adhesion	KEGG:04510	2.012×10 ⁻⁸
Proteoglycans in cancer	KEGG:05205	2.328×10 ⁻⁷
Signaling pathways regulating pluripotency of stem cells	KEGG:04550	2.638×10 ⁻⁷
MicroRNAs in cancer	KEGG:05206	5.100×10 ⁻⁶
PI3K-Akt signaling pathway	KEGG:04151	1.794×10 ⁻⁵
Hippo signaling pathway	KEGG:04390	2.261×10 ⁻⁵
Rap1 signaling pathway	KEGG:04015	2.336×10 ⁻⁵
Ras signaling pathway	KEGG:04014	3.745×10 ⁻⁵
MAPK signaling pathway	KEGG:04010	1.546×10 ⁻⁴
ECM-receptor interaction	KEGG:04512	1.546×10 ⁻⁴
Regulation of actin cytoskeleton	KEGG:04810	2.588×10 ⁻⁴
Wnt signaling pathway	KEGG:04310	1.045×10 ⁻³
Human papillomavirus infection	KEGG:05165	1.177×10 ⁻³
Arrhythmogenic right ventricular cardiomyopathy	KEGG:05412	1.322×10 ⁻³
Axon guidance	KEGG:04360	1.678×10 ⁻³
AGE-RAGE signaling pathway in diabetic complications	KEGG:04933	1.927×10 ⁻³
p53 signaling pathway	KEGG:04115	3.519×10 ⁻³
Prostate cancer	KEGG:05215	3.919×10 ⁻³
EGFR tyrosine kinase inhibitor resistance	KEGG:01521	5.333×10 ⁻³
Renal cell carcinoma	KEGG:05211	6.290×10 ⁻³
Dilated cardiomyopathy	KEGG:05414	7.603×10 ⁻³

REACTOME Pathways

REAC		stats
Term name	Term ID	Padj
Axon guidance	REAC:R-MMU-4...	1.028×10 ⁻⁸
Nervous system development	REAC:R-MMU-9...	1.028×10 ⁻⁸
Extracellular matrix organization	REAC:R-MMU-1...	2.339×10 ⁻⁷
Hemostasis	REAC:R-MMU-1...	1.071×10 ⁻⁵
Platelet activation, signaling and aggregation	REAC:R-MMU-7...	3.971×10 ⁻⁵
Platelet degranulation	REAC:R-MMU-1...	5.893×10 ⁻⁵
Response to elevated platelet cytosolic Ca2+	REAC:R-MMU-7...	1.095×10 ⁻⁴
Cell junction organization	REAC:R-MMU-4...	1.124×10 ⁻⁴
Signaling by Receptor Tyrosine Kinases	REAC:R-MMU-9...	1.443×10 ⁻⁴
Integrin cell surface interactions	REAC:R-MMU-2...	1.601×10 ⁻⁴
Regulation of Insulin-like Growth Factor (IGF) transport an...	REAC:R-MMU-3...	1.601×10 ⁻⁴
Developmental Biology	REAC:R-MMU-1...	1.601×10 ⁻⁴
Cell-Cell communication	REAC:R-MMU-1...	1.601×10 ⁻⁴
Elastic fibre formation	REAC:R-MMU-1...	1.758×10 ⁻⁴
ECM proteoglycans	REAC:R-MMU-3...	3.823×10 ⁻⁴
Post-translational protein phosphorylation	REAC:R-MMU-8...	7.349×10 ⁻⁴
L1CAM interactions	REAC:R-MMU-3...	1.241×10 ⁻³
Molecules associated with elastic fibres	REAC:R-MMU-2...	1.403×10 ⁻³
MAPK family signaling cascades	REAC:R-MMU-5...	2.198×10 ⁻³
Non-integrin membrane-ECM interactions	REAC:R-MMU-3...	2.837×10 ⁻³
Cell-extracellular matrix interactions	REAC:R-MMU-4...	3.717×10 ⁻³
Semaphorin interactions	REAC:R-MMU-3...	8.977×10 ⁻³
G alpha (12/13) signalling events	REAC:R-MMU-4...	9.364×10 ⁻³

Wiki Pathways

WP		stats
Term name	Term ID	Padj
PluriNetWork: mechanisms associated with pluripotency	WP:WP1763	1.663×10 ⁻⁸
ESC Pluripotency Pathways	WP:WP339	1.888×10 ⁻⁶
Focal Adhesion	WP:WP85	2.408×10 ⁻⁴
TGF Beta Signaling Pathway	WP:WP113	1.465×10 ⁻³
Heart Development	WP:WP2067	1.465×10 ⁻³
Integrin-mediated Cell Adhesion	WP:WP6	1.465×10 ⁻³
Neural Crest Differentiation	WP:WP2074	5.291×10 ⁻³

Figure 49 Gene ontology enrichment analysis for DEGs between d2 and d4. Terms related to biological processes, KEGG pathways, REACTOME pathways, Wiki pathways were considered. In red box there are some interesting and statistically significant (p-value) biological processes.

Overall, these results are consistent with the *in vitro* induced developmental progression from pluripotent stem cells to endothelial cells.

Many of the 1735 DEGs between d2 and d4 were involved in angiogenesis and endothelial related processes, indicating the activation of an EC transcription program.

2.2 Chromatin accessibility profile in early phases of multipotent cardiac progenitors' differentiation into mature ECs.

To generate maps of chromatin accessibility during the early phases of differentiation, I collected differentiating mESCs at d2 and d4 and they were processed for ATAC-seq assay (two biological replicates for each condition). ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) is a method for determining chromatin accessibility across the genome. It utilizes a hyperactive Tn5 transposase to insert sequencing adapters into open chromatin regions (Figure 50). High-throughput sequencing then yields reads that indicate these regions of increased accessibility.

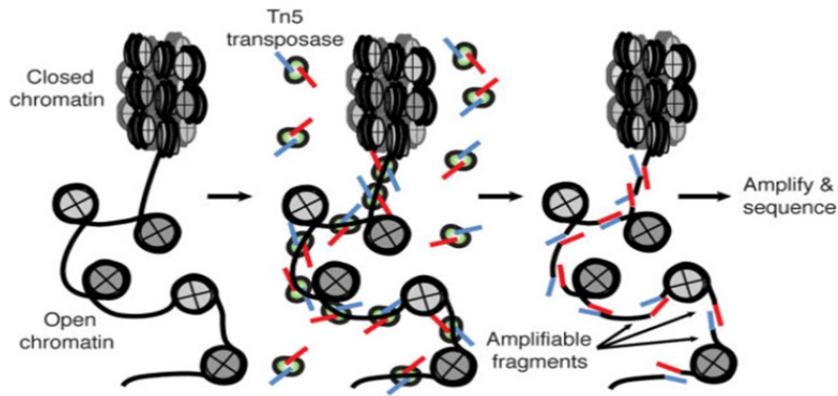


Figure 50 ATAC-seq overview.

The overall number of raw reads is illustrated in the Table 8.

Sample pairs	N° of raw reads
d2_rep1	20,294,457
d2_rep2	23,034,093
d4_rep1	32,221,749
d4_rep2	30,364,839

Table 8 Total number of ATAC-seq raw reads in d2-d4 differentiated mESCs.

Next, we aligned the BAM (Binary Alignment Map) reads of d2-d4 samples to mouse genome mm10, using Bowtie2 (Langmead B *et al.*, (2012)).

To reduce mitochondrial noise from samples (Montefiori L. *et al.*, (2017)), mitochondrial reads (chrM) were removed from aligned SAM (Sequence Alignment Map) files. The Sam file is a Tab-delimited text format divided into two parts: the first one includes header and second one which has details about

alignment, for example mapping position (Li H. *et al.*, (2009)). Also unmapped and random reads were discarded from the analysis.

Table 9 shows the total number of aligned reads without chrM. Moreover, quality check of these aligned reads (using FastQC tool) is depicted in Figure 51.

Sample	Input read pairs	Concordant Alignment %				Aligned read pairs	Aligned pairs with no chrM
		Aligned	Unique	Multi	None		
d2_rep1	20,294,457	96.45%	64.01	28.54	7.46	19,574,004	17,097,584
d2_rep2	23,034,093	97.72%	53.70	40.58	5.72	22,508,916	13,192,117
d4_rep1	32,221,749	97.15%	56.66	37.32	6.03	31,303,430	21,539,055
d4_rep2	30,364,839	97.55%	57.54	36.82	5.64	29,620,900	20,566,625

Table 9 Number of aligned reads without chrM of 2 replicates d2 and 2 replicates d4.

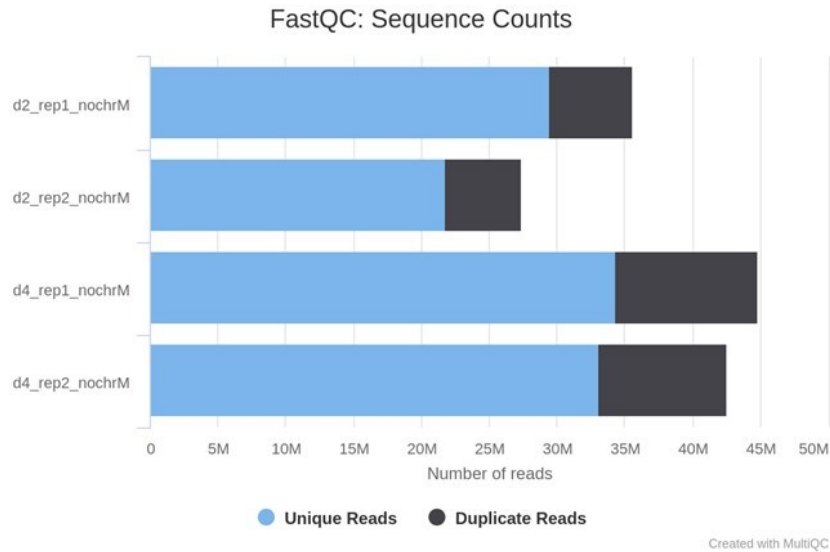


Figure 51 Quality check of aligned reads following removal of mitochondrial genome reads.

Subsequently, we carried out ATAC peak calling, which quantifies how many open regions (peaks) there were in our samples. The algorithm used is MACS2 (Model-Based Analysis of ChIP-seq) (Feng J. *et al.*, (2012)). This tool identifies statistically enriched accessible regions in the genome.

Overall, we identified 16952 chromatin accessible peaks in differentiating d2 mESC and 13330 in d4 samples. The total number of the called and consensus (a consensus peak is a peak occurring in at least 2 replicates of the given replicates and it must pass the significance cutoff $q < 0.001$) peaks identified between d2-d4 were shown in Table 10.

Sample pairs	Peaks called ($q < 0.001$)	Consensus peaks ($q < 0.001$)
d2_rep1	19887	16952
d2_rep2	22592	
d4_rep1	16612	13330
d4_rep2	15627	

Table 10 Total peaks number (total accessible chromatin regions) in differentiating d2-d4 mESC. A consensus peak is a peak occurring in at least 2 replicates of the given replicates and it must pass the significance cutoff ($q < 0.001$).

We found that overall ATAC-seq peaks of d2 and d4 replicates were located mostly around the transcription start site (TSS) of genes, as seen in two different graphs in Figure 52-Figure 53, suggesting that the technique was working well.

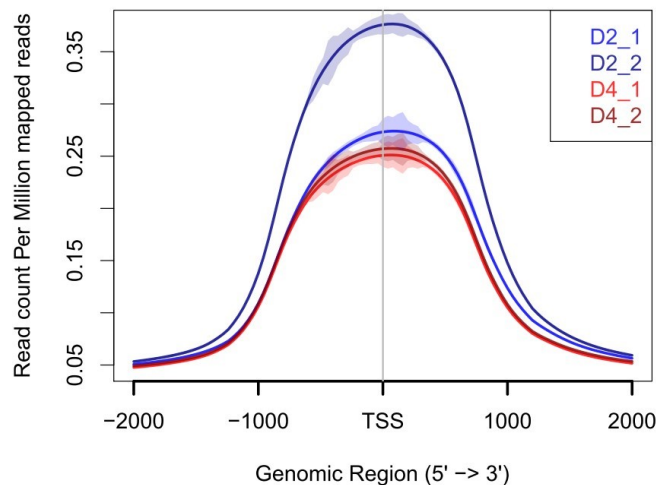


Figure 52 Total accessible regions distribution in d2 (in blue) and d4 (in red) samples.

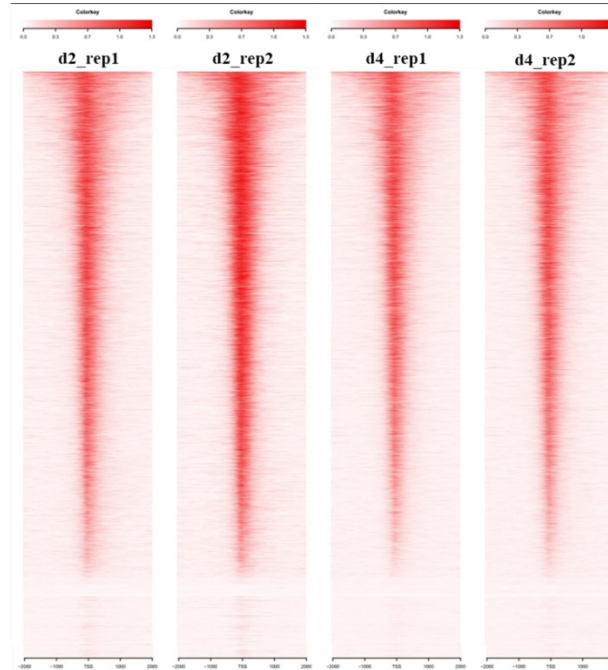


Figure 53 Heat map of chromatin accessible regions distribution around TSS (from 1kb to 2kb) of d2 (2replicates) and d4 (2replicates).

We annotated the consensus MACS2 peaks common to d2 replicates (16952) and d4 replicates (13330), using ChIP-seeker (Yu G. *et al.*, (2015)). Two different promoter definition were used: one was defined as 3000bp upstream to 3000bp downstream to TSS, the other was defined as 1000bp upstream to 1000bp downstream to TSS. We evaluated the distribution of peaks relative to gene features defined as Promoter 1kb and 3 kb (from 1 kb to 3kb from the transcription start site), 5'-UTR, 3'-UTR, First Exon, First Intron, Other Exon, Other Intron, Downstream, Distal Intergenic. The results shown that there were no relevant differences between the two promoter definitions at d2 and d4 samples, as seen in Figure 54-Figure 55.

In summary, at day2, 59% of regions were localized around promoter, 22.7% were intragenic and 18.23% were distal intergenic; while at day 4, 66.5% of regions were localized around promoter, 21% were intragenic and 12.9% were distal intergenic. In both cases, peaks were distributed mostly at the promoter regions of the genes and in the intragenic regions, as expected. Overall distribution suggested no differences in chromatin accessibility regions between the two different conditions.

Promoter region= $\pm 3000\text{bp}$ to TSS

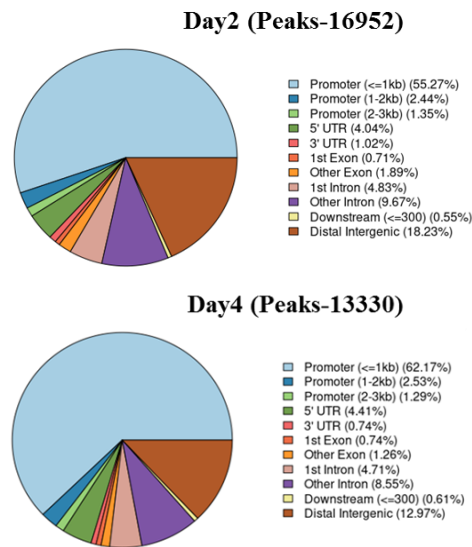


Figure 54 Total accessible regions distribution in day2 and day4 mESC. The annotation pie shows accessible regions distribution around gene features (Promoter region= $\pm 3000\text{bp}$ to TSS).

Promoter region= $\pm 1000\text{bp}$ to TSS

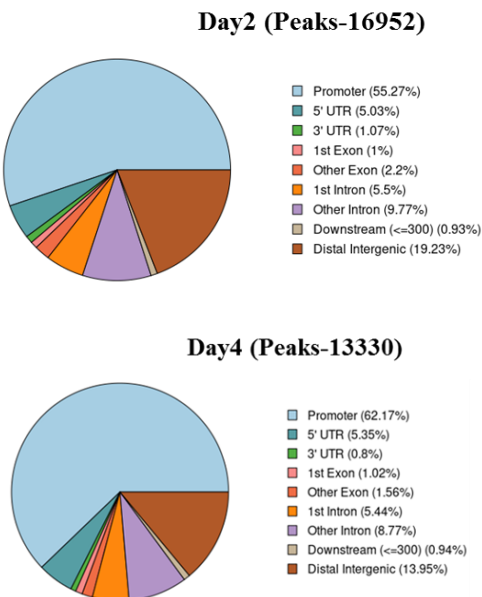


Figure 55 Total accessible regions distribution in day2 and day4 mESC. The annotation pie shows accessible regions distribution around gene features (Promoter region= $\pm 1000\text{bp}$ to TSS).

Next, we performed a search of differentially accessible regions (DARs), using the d2 and d4 replicate peaks, with two different methods: DEScan2 and DiffBind. DARs are regions that increase/decrease their chromatin accessibility during the differentiation procedure in a statistically significant measure.

For DiffBind and DEScan2 analysis, MACS2 peaks with $q < 0.001$ (q-value statistical cutoff) from the replicates were given as input and regions with $FDR < 0.01$ were considered as DARs. Total number of DARs regions identified in d4 vs d2 comparison are shown in Table 11.

I found 6348 regions identified as DARs with both methods (common DARs). In particular, there were 2409 DARs with increased accessibility at d4 vs d2, and 3939 with decreased accessibility (Table 12).

Sample pairs (d2-d4 peaks)	DiffBind	DEScan2	Common DARs (intersection DEScan2 with DiffBind)
Total DARs	11798	9650	6348
Increased DARs	4237	3977	2409
Decreased DARs	7561	5673	3939

Table 11 Total number of Differential Accessible Regions (DARs) between d2-d4 using DiffBind and DEScan2 tools.

Common DARs were then annotated using ChIPSeeker and two different promoter definitions TSS [± 3000] and TSS [± 1000]. Each DAR was annotated to at least one gene and, in some cases, more than one. Hence, the number of genes was less than the number of DARs, as seen in Table 12.

Classification	Number of DARs	Number of genes in DARs
DARs	6348	5054
Increased DARs	2409	2080
Decreased DARs	3939	3394

Table 12 Annotation of common DARs in d4 vs d2. The DARs were annotated with two different promoter definition TSS [± 3000] and [± 1000].

We evaluated genomic feature annotation and observed that while the majority of day4 peaks were located in the promoters, the DARs in day4 were primarily annotated to intra- (43.8%) and inter-genic (37-39%) regions, implying that the chromatin landscape is mainly altered within gene body or at enhancers located far

from the target genes. Only about 15-18% of them were around promoter regions (Figure 56).

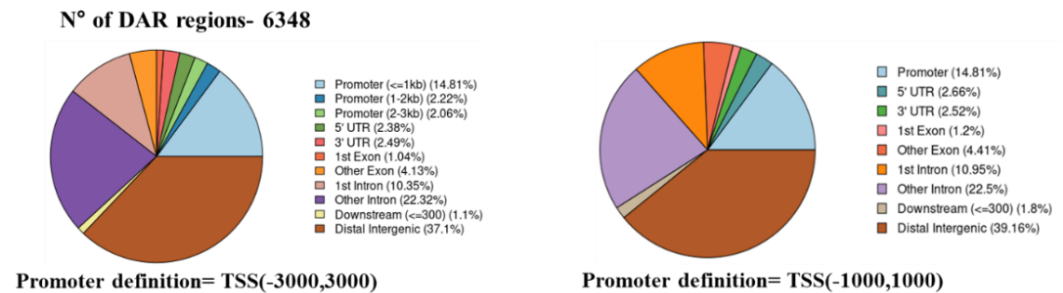


Figure 56 Distribution of common DARs in d4 vs d2. Annotation pie shows accessible regions distribution around gene features (TSS [± 3000] and [± 1000]).

The distance between each common DARs to its nearest TSS region was calculated and plotted in Figure 57 (x-axis denotes the distance to TSS, from 0 bp to 50000bp; y-axis denotes the number of DARs).

940 out 6348 DARs lie within 1kb distance from TSS and 2082/6348 DARs within 10kb distance from TSS.

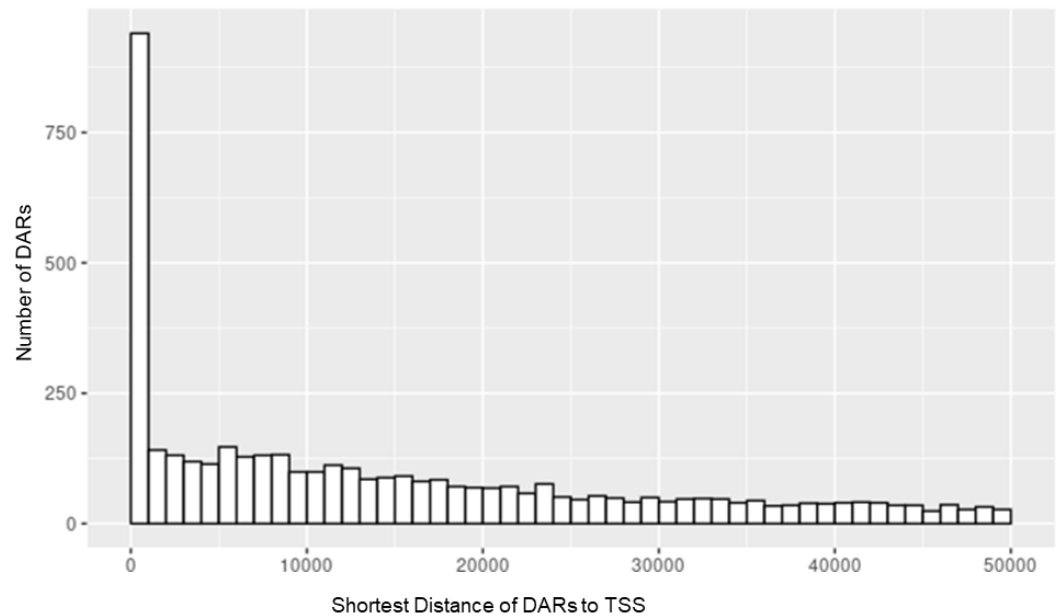


Figure 57 Distance of the common DARs to nearest TSS region (x-axis denotes the distance to TSS, from 0 bp to 50000bp; y-axis denotes the number of DARs).

Gene ontology analysis (g:Profiler2) of genes annotated to the common DARs (5054) was performed and the most enriched BP and Pathways terms are depicted in the Figure 58. Many genes associated with new accessible regions are involved in cardiac and endothelial cell fate specification. In particular, I found that the most significant over-represented biological processes included terms “developmental process”, “cell development”, “regulation of developmental process”, “stem cell differentiation” as well as “vasculature development”, “blood vessel development”, “blood circulation”, “blood vessel morphogenesis”, “heart development”, “angiogenesis”, “embryo development”, “regulation of vasculature development”, “cardiomyocyte differentiation”, “regulation of endothelial cell migration”, “endothelial cell migration”, “regulation of vascular endothelial growth factor signaling pathway” and so on.

GO Biological Processes

source	term_name	term_id	adjusted_p_value
GO:BP	anatomical structure morphogenesis	GO:0009653	5.56E-62
GO:BP	system development	GO:0048731	7.49E-60
GO:BP	anatomical structure development	GO:0048856	2.68E-59
GO:BP	multicellular organism development	GO:0007275	2.68E-59
GO:BP	developmental process	GO:0032502	4.72E-51
GO:BP	nervous system development	GO:0007399	6.75E-51
GO:BP	generation of neurons	GO:0048699	9.79E-44
GO:BP	neurogenesis	GO:0022008	6.67E-43
GO:BP	localization	GO:0051179	2.72E-37
GO:BP	cell morphogenesis	GO:0000902	2.72E-37
GO:BP	neuron differentiation	GO:0030182	8.93E-37
GO:BP	animal organ development	GO:0048513	4.86E-36
GO:BP	cell development	GO:0048468	5.21E-36
GO:BP	regulation of localization	GO:0032879	2.41E-35
GO:BP	cell-cell signaling	GO:0007267	3.58E-33
GO:BP	tissue development	GO:0009888	5.95E-33
GO:BP	regulation of biological quality	GO:0065008	6.01E-32
GO:BP	plasma membrane bounded cell projection organization	GO:0120036	1.84E-31
GO:BP	cell projection organization	GO:0030030	3.85E-31
GO:BP	cell differentiation	GO:0030154	1.06E-30
GO:BP	regulation of signaling	GO:0023051	1.20E-30
GO:BP	regulation of cell communication	GO:0010646	1.20E-30
GO:BP	cellular developmental process	GO:0048869	4.20E-30
GO:BP	regulation of developmental process	GO:0050793	1.58E-29
GO:BP	neuron projection development	GO:0031175	7.39E-29
GO:BP	movement of cell or subcellular component	GO:0006928	8.49E-29
GO:BP	regulation of multicellular organismal process	GO:0051239	1.07E-28
GO:BP	cell morphogenesis involved in differentiation	GO:0000904	1.86E-28
GO:BP	cell projection morphogenesis	GO:0048858	5.20E-28
GO:BP	multicellular organismal process	GO:0032501	1.46E-27
GO:BP	plasma membrane bounded cell projection morphogenesis	GO:0120039	1.46E-27
GO:BP	tube development	GO:0035295	2.49E-27
GO:BP	neuron projection morphogenesis	GO:0048812	3.21E-27

GO:BP	positive regulation of developmental process	GO:0051094	1.70E-15
GO:BP	regulation of cell motility	GO:2000145	2.05E-15
GO:BP	regulation of cell migration	GO:0030334	2.43E-15
GO:BP	regulation of multicellular organismal development	GO:2000026	3.40E-15
GO:BP	dendrite development	GO:0016358	4.14E-15
GO:BP	axon guidance	GO:0007411	5.70E-15
GO:BP	neuron projection guidance	GO:0097485	9.82E-15
GO:BP	cell-cell adhesion	GO:0098609	1.10E-14
GO:BP	circulatory system process	GO:0003013	1.21E-14
GO:BP	regulation of cell population proliferation	GO:0042127	1.33E-14
GO:BP	regulation of membrane potential	GO:0042391	1.36E-14
GO:BP	signaling	GO:0023052	1.36E-14
GO:BP	blood vessel development	GO:0001568	1.58E-14
GO:BP	regulation of small GTPase mediated signal transduction	GO:0051056	1.67E-14
GO:BP	sensory organ development	GO:0007423	2.09E-14
GO:BP	vasculature development	GO:0001944	2.21E-14
GO:BP	positive regulation of GTPase activity	GO:0043547	2.35E-14
GO:BP	small GTPase mediated signal transduction	GO:0007264	2.43E-14
GO:BP	cell population proliferation	GO:0008283	7.05E-14
GO:BP	blood circulation	GO:0008015	7.24E-14
GO:BP	biological regulation	GO:0065007	7.63E-14
GO:BP	actin cytoskeleton organization	GO:0030036	8.80E-14
GO:BP	regulation of cellular process	GO:0050794	1.05E-13
GO:BP	regulation of neuron projection development	GO:0010975	1.85E-13
GO:BP	cation transport	GO:0006812	2.08E-13
GO:BP	positive regulation of multicellular organismal process	GO:0051240	2.30E-13
GO:BP	regulation of cell adhesion	GO:0030155	2.51E-13
GO:BP	forebrain development	GO:0030900	2.64E-13
GO:BP	transmembrane transport	GO:0055085	2.82E-13
GO:BP	ion transport	GO:0006811	3.51E-13
GO:BP	regulation of cell differentiation	GO:0045595	4.53E-13
GO:BP	blood vessel morphogenesis	GO:0048514	5.58E-13
GO:BP	regulation of intracellular signal transduction	GO:1902531	5.68E-13
GO:BP	regulation of biological process	GO:0050789	6.03E-13
GO:BP	cell junction assembly	GO:0034329	6.16E-13
GO:BP	positive regulation of response to stimulus	GO:0048584	7.11E-11
GO:BP	regulation of synapse organization	GO:0050807	7.86E-11
GO:BP	urogenital system development	GO:0001655	9.74E-11
GO:BP	regulation of synapse structure or activity	GO:0050803	1.25E-10
GO:BP	heart development	GO:0007507	1.40E-10
GO:BP	regulation of transmembrane transport	GO:0034762	1.56E-10
GO:BP	embryonic morphogenesis	GO:0048598	4.65E-10
GO:BP	eye development	GO:0001654	4.86E-10
GO:BP	regulation of developmental growth	GO:0048638	4.96E-10
GO:BP	angiogenesis	GO:0001525	4.98E-10
GO:BP	epithelial tube morphogenesis	GO:0060562	5.21E-10
GO:BP	positive regulation of cellular component organization	GO:0051130	6.60E-10
GO:BP	positive regulation of cell differentiation	GO:0045597	1.29E-09
GO:BP	response to endogenous stimulus	GO:0009719	1.36E-09
GO:BP	positive regulation of cell population proliferation	GO:0008284	1.57E-09
GO:BP	cytoskeleton organization	GO:0007010	2.13E-09
GO:BP	developmental growth involved in morphogenesis	GO:0060560	2.14E-09
GO:BP	cellular response to organic substance	GO:0071310	2.24E-09
GO:BP	negative regulation of cellular component movement	GO:0051271	2.28E-09
GO:BP	skeletal system development	GO:0001501	2.44E-09
GO:BP	negative regulation of cell motility	GO:2000146	2.45E-09
GO:BP	dendrite morphogenesis	GO:0048813	3.41E-09
GO:BP	regulation of secretion by cell	GO:1903530	4.07E-09
GO:BP	embryo development	GO:0009790	4.25E-09
GO:BP	vascular process in circulatory system	GO:0003018	4.38E-09
GO:BP	regulation of secretion	GO:0051046	4.44E-09
GO:BP	cellular localization	GO:0051641	4.52E-09
GO:BP	kidney development	GO:0001822	6.58E-09

GO:BP	cardiac muscle tissue development	GO:0048738	9.86E-07
GO:BP	gland development	GO:0048732	9.86E-07
GO:BP	cardiac chamber development	GO:0003205	1.14E-06
GO:BP	divalent metal ion transport	GO:0070838	1.17E-06
GO:BP	divalent inorganic cation transport	GO:0072511	1.17E-06
GO:BP	regulation of cell morphogenesis	GO:0022604	1.20E-06
GO:BP	epithelial cell differentiation	GO:0030855	1.22E-06
GO:BP	canonical Wnt signaling pathway	GO:0060070	1.24E-06
GO:BP	gland morphogenesis	GO:0022612	1.27E-06
GO:BP	positive regulation of metabolic process	GO:0009893	1.27E-06
GO:BP	regulation of anion transport	GO:0044070	1.31E-06
GO:BP	regulation of Wnt signaling pathway	GO:0030111	1.41E-06
GO:BP	cellular cation homeostasis	GO:0030003	1.55E-06
GO:BP	cellular calcium ion homeostasis	GO:0006874	1.59E-06
GO:BP	activation of GTPase activity	GO:0090630	1.71E-06
GO:BP	regulation of multicellular organism growth	GO:0040014	1.06E-05
GO:BP	anion transport	GO:0006820	1.10E-05
GO:BP	skeletal system morphogenesis	GO:0048705	1.17E-05
GO:BP	mesenchymal cell development	GO:0014031	1.22E-05
GO:BP	regulation of vascular permeability	GO:0043114	1.25E-05
GO:BP	postsynapse assembly	GO:0099068	1.26E-05
GO:BP	calcium ion transmembrane transport	GO:0070588	1.27E-05
GO:BP	respiratory tube development	GO:0030323	1.29E-05
GO:BP	glutamate receptor signaling pathway	GO:0007215	1.30E-05
GO:BP	transmembrane receptor protein serine/threonine kinase signaling pathway	GO:0007178	1.34E-05
GO:BP	potassium ion transport	GO:0006813	1.44E-05
GO:BP	stem cell development	GO:0048864	6.11E-05
GO:BP	positive regulation of neurogenesis	GO:0050769	6.65E-05
GO:BP	kidney epithelium development	GO:0072073	6.67E-05
GO:BP	regulation of programmed cell death	GO:0043067	6.72E-05
GO:BP	regulation of angiogenesis	GO:0045765	6.81E-05
GO:BP	regulation of cellular protein localization	GO:1903827	6.83E-05
GO:BP	long-term synaptic depression	GO:0060292	6.83E-05
GO:BP	blood vessel diameter maintenance	GO:0097746	6.83E-05
GO:BP	regulation of tube diameter	GO:0035296	6.83E-05
GO:BP	ERK1 and ERK2 cascade	GO:0070371	6.98E-05
GO:BP	regulation of vasculature development	GO:1901342	0.000114022
GO:BP	cartilage development	GO:0051216	0.000114753
GO:BP	regulation of actin filament organization	GO:0110053	0.000117265
GO:BP	regulation of neurotransmitter receptor activity	GO:0099601	0.000121139
GO:BP	neural tube development	GO:0021915	0.00012149
GO:BP	neuron projection organization	GO:0106027	0.000124606
GO:BP	macromolecule modification	GO:0043412	0.0001247
GO:BP	associative learning	GO:0008306	0.000126183
GO:BP	neural crest cell development	GO:0014032	0.000130658
GO:BP	positive regulation of gene expression	GO:0010628	0.00013218
GO:BP	regulation of cardiac muscle cell proliferation	GO:0060043	0.000133964
GO:BP	heart growth	GO:0060419	0.002545166
GO:BP	smooth muscle cell proliferation	GO:0048659	0.002620802
GO:BP	regulation of membrane protein ectodomain proteolysis	GO:0051043	0.002634641
GO:BP	regulation of amine transport	GO:0051952	0.00281806
GO:BP	cardiocyte differentiation	GO:0035051	0.002905528
GO:BP	response to oxygen levels	GO:0070482	0.002905528
GO:BP	positive regulation of blood-brain barrier permeability	GO:1905605	0.003541484
GO:BP	regulation of protein tyrosine phosphatase activity	GO:1903613	0.003541484
GO:BP	synaptic membrane adhesion	GO:0099560	0.003584233
GO:BP	regulation of endothelial cell migration	GO:0010594	0.003593871
GO:BP	cellular protein complex disassembly	GO:0043624	0.003604298
GO:BP	neural tube formation	GO:0001841	0.003720596
GO:BP	positive regulation of vasculature development	GO:1904018	0.003723555
GO:BP	positive regulation of angiogenesis	GO:0045766	0.003723555
GO:BP	behavioral fear response	GO:0001662	0.003723555
GO:BP	regulation of positive chemotaxis	GO:0050926	0.003746944
GO:BP	branching involved in salivary gland morphogenesis	GO:0060445	0.003746944
GO:BP	hypothalamus development	GO:0021854	0.003746944
GO:BP	regulation of AMPA receptor activity	GO:2000311	0.003746944
GO:BP	fatty acid transport	GO:0015908	0.00386346
GO:BP	positive regulation of multicellular organism growth	GO:0040018	0.00386346
GO:BP	vascular endothelial growth factor signaling pathway	GO:0038084	0.00386346
GO:BP	cell fate determination	GO:0001709	0.00386346

GO:BP	positive regulation of endothelial cell migration	GO:0010595	0.004758447
GO:BP	endothelial cell migration	GO:0043542	0.005455253
GO:BP	protein glycosylation	GO:0006486	0.005455253
GO:BP	salivary gland morphogenesis	GO:0007435	0.005593742
GO:BP	negative regulation of actin filament bundle assembly	GO:0032232	0.005593742
GO:BP	regulation of biomineral tissue development	GO:0070167	0.005632109
GO:BP	chloride transmembrane transport	GO:1902476	0.005632109
GO:BP	positive regulation of biomineral tissue development	GO:0070169	0.005666867
GO:BP	stress-activated protein kinase signaling cascade	GO:0031098	0.005884021
GO:BP	protein localization to cell junction	GO:1902414	0.005911937
GO:BP	regulation of receptor internalization	GO:0002090	0.005943248
GO:BP	SMAD protein signal transduction	GO:0060395	0.006276349
GO:BP	regulation of heart growth	GO:0060420	0.006276349
GO:BP	regulation of presynapse organization	GO:0099174	0.006283557
GO:BP	positive regulation of calcium ion transmembrane transporter activity	GO:1901021	0.006283557
GO:BP	embryonic hindlimb morphogenesis	GO:0035116	0.006327184
GO:BP	regulation of ossification	GO:0030278	0.006412351
GO:BP	regulation of cardiac muscle tissue growth	GO:0055021	0.006556352
GO:BP	regulation of vascular endothelial growth factor signaling pathway	GO:1900746	0.007531054

Figure 58 Gene ontology enrichment analysis for genes annotated to common DARs between d2 and d4. Terms related to biological processes (GO:BP) were considered. In red there are some interesting and statistically significant (p-value) biological processes.

2.3 Intersection of DE genes and DARs regions between day2 and day4 of CPM-EC differentiation.

We compared DE genes list and genes associated with DARs. The volcano plot in Figure 59 shows the intersection of DE and DARs genes in d4 vs d2 differentiation. By this data integration, we obtained a subset of genes that were both altered in terms of chromatin accessibility and its expression.

The X and Y-axis are in terms of expressed genes. The X-axis denotes the fold change of expressed genes, while the Y-axis denotes the posterior probability (pp). Genes with $pp > 0.8$ ($-\log_{10}(1-pp) > 1$) were considered DEGs.

In red were DARs and down-regulated DE genes (n.145); in dark blue were illustrated DARs and up-regulated DE genes (n.159); in orange were DARs and down-expressed genes, while in light blue DARs and up-expressed genes.

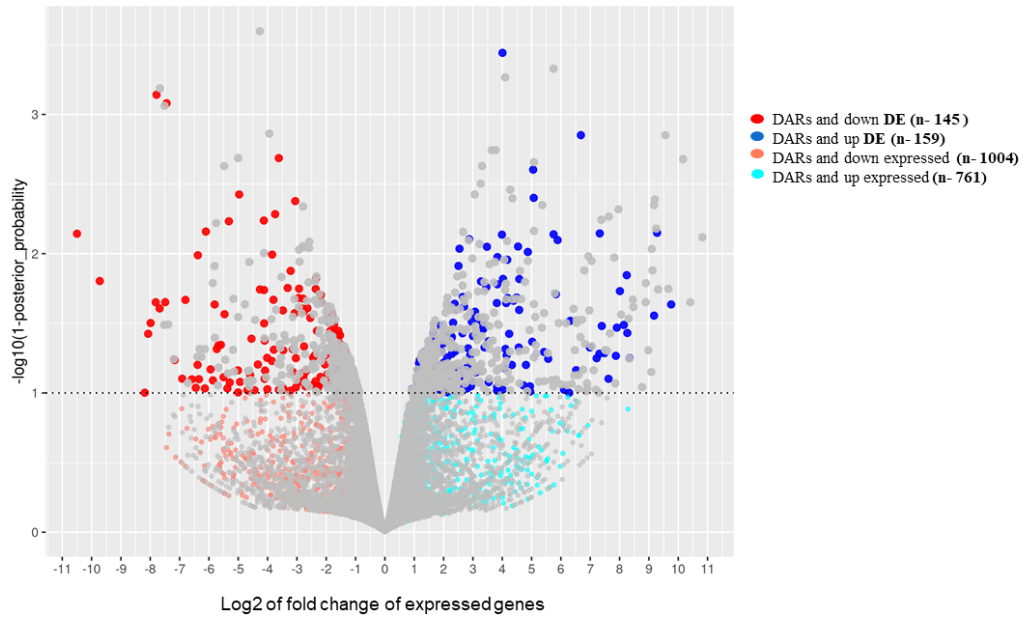


Figure 59 Volcano plot of intersection of DE and DARs genes in d4 vs d2 differentiation. In red were DARs and down-regulated DE genes (n.145); in dark blue were illustrated DARs and up-regulated DE genes (n.159); in orange were DARs and down-expressed genes, while in light blue DARs and up-expressed genes.

Chapter 3.

Identification of putative endothelial regulatory elements (enhancers) regulated during CPM differentiation into endothelial cells (EC) by integration of RNA-seq and ATAC-seq data.

In general, enhancers are distal regulatory elements that can impact gene expression regardless of their position, orientation, and distance relative to a target promoter. These elements are essential for precise spatiotemporal regulation of gene expression, which is required for proper cell development and differentiation. Assays of chromatin accessibility, which provide an indication of how “open” a region is, can be used to identify enhancer elements. Moreover, enhancers are found mostly in the intergenic and intronic regions, while a few enhancers have been found within exons (Panigrahi *et al.*, (2021)).

Since most of DARs found in d4 vs d2 samples are located in intra- (43.8%) and inter-genic (37-39%) regions (see Figure 17, Cap2), I have searched for DARs opening at d4 and located within 10 kb endothelial-specific genes activated at d4.

We uploaded RNA-seq and ATAC-seq data on the USCS genome browser (<https://genome.ucsc.edu/>) and by comparison of the two methods, I have identified in total 10 putative endothelial-specific enhancer regions, which increased their chromatin accessibility and are associated with endothelial genes activated at day 4 of differentiation, compared to day2, as shown in Table 13.

Gene	ATAC peaks position (increased at d4)	DE genes d4 vs d2
<i>Kdr (Vegfr2)</i>	Intron10	+
<i>Cdh5 (VE-Cadherin)</i>	Intron1	+
<i>CD34</i>	10 kb upstream to TSS	+
<i>Eng</i>	Intron2	+
<i>Flt1 (Vegfr1)</i>	Intron10	+
<i>Tal1 (Scl1)</i>	5 kb upstream to TSS	+
<i>Pecam1</i>	Intron2	+
<i>Notch1</i>	Intron15	+
<i>Dusp5</i>	5 kb upstream to TSS	No DEG but involved in angiogenesis
<i>Gata6</i>	Intron6	+

Table 13 List of total 10 identified putative enhancers regions associated with endothelial-specific genes. The table contains: gene names (left column); position of the ATAC peaks increased at d4 (middle column); DE genes d4 vs d2 (right column) (gene expression variations during d2-d4 differentiation; + refers to genes whose expression increases at d4).

Examples of ATAC-seq and RNA-seq peaks coverage related to the 10 regions, which show different chromatin accessibility and gene expression at these two stages of differentiation (d2-d4), are depicted in Figure 60.a-j.

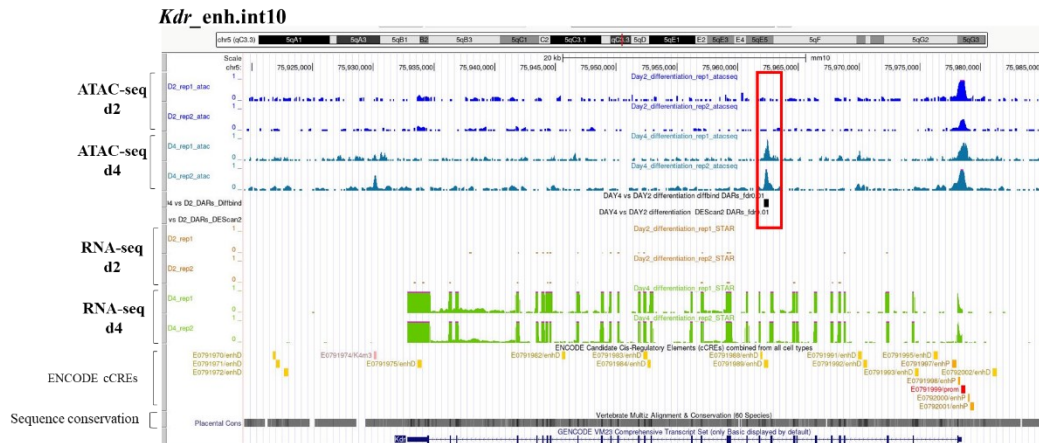


Figure 60.a ATAC peaks and RNA-seq peaks coverage associated to *Kdr*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.

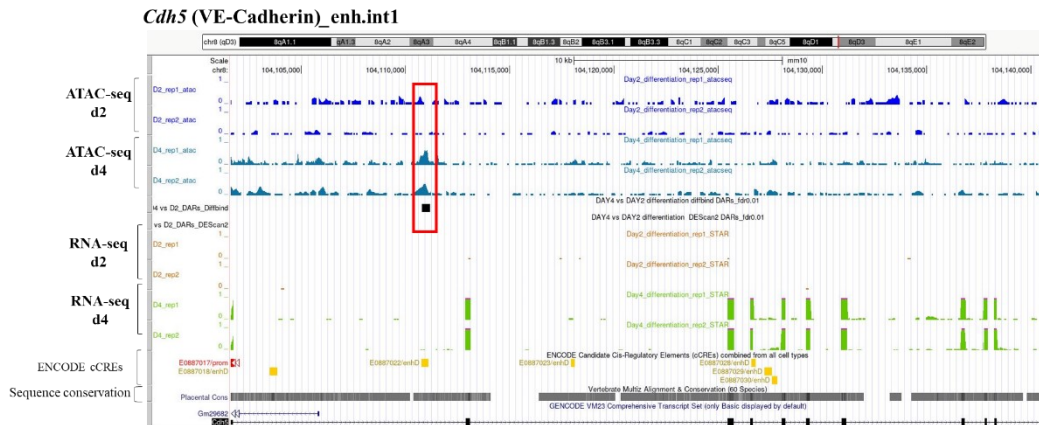


Figure 60.b ATAC peaks and RNA-seq peaks coverage associated to *Cdh5 (VE-Cadherin)*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.

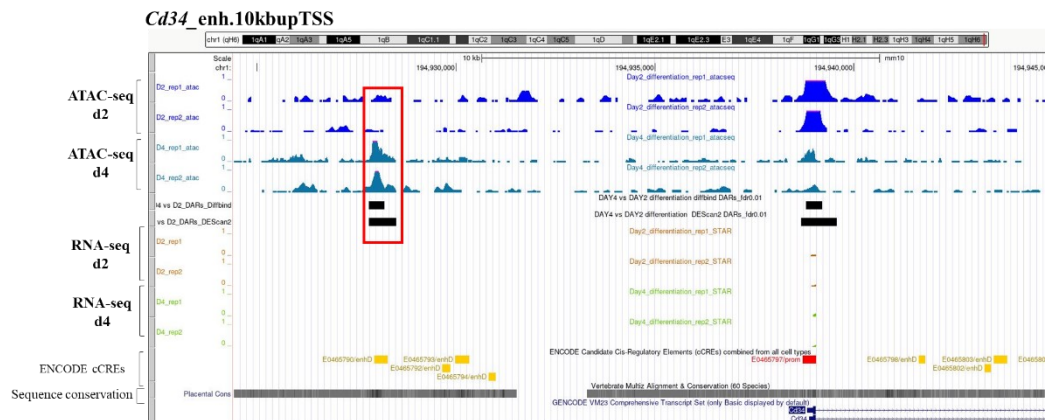


Figure 60.c ATAC peaks and RNA-seq peaks coverage associated to *CD34*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.



Figure 60.d ATAC peaks and RNA-seq peaks coverage associated to *Eng*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.



Figure 60.g ATAC peaks and RNA-seq peaks coverage associated to *Pecam1*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.

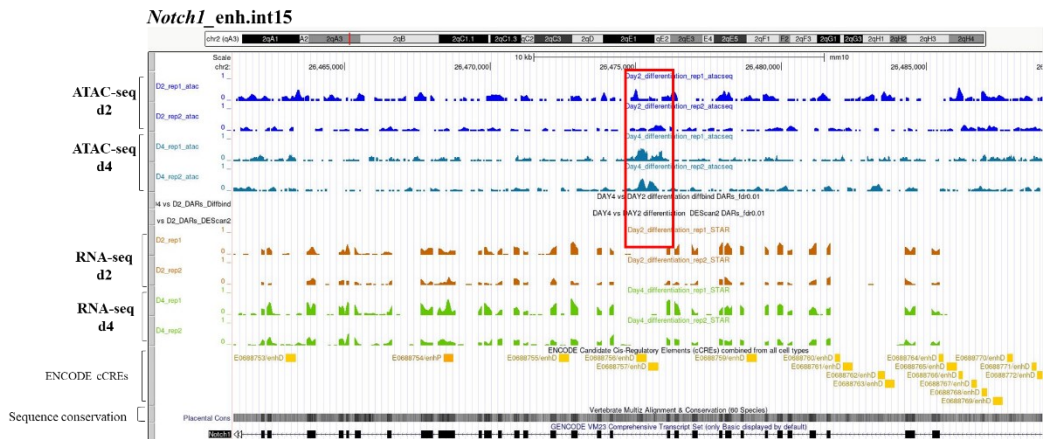
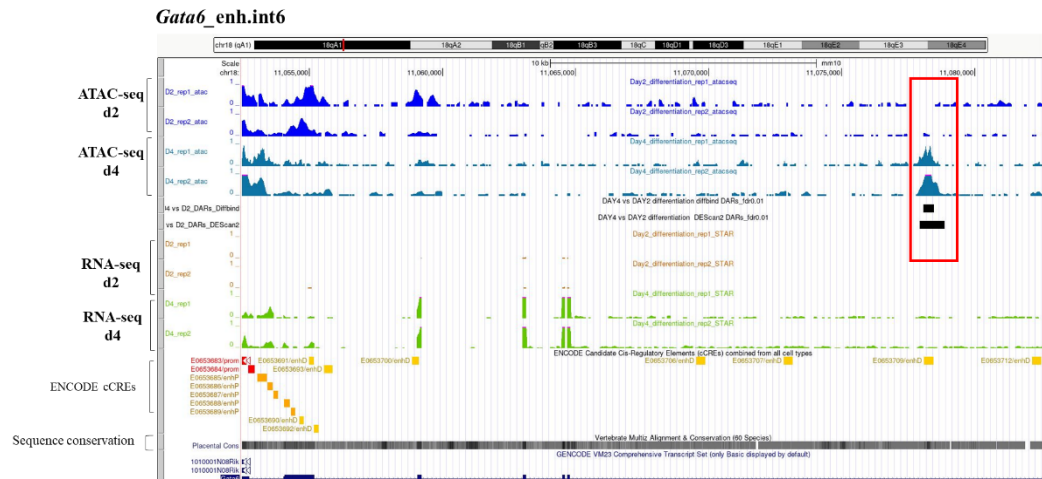
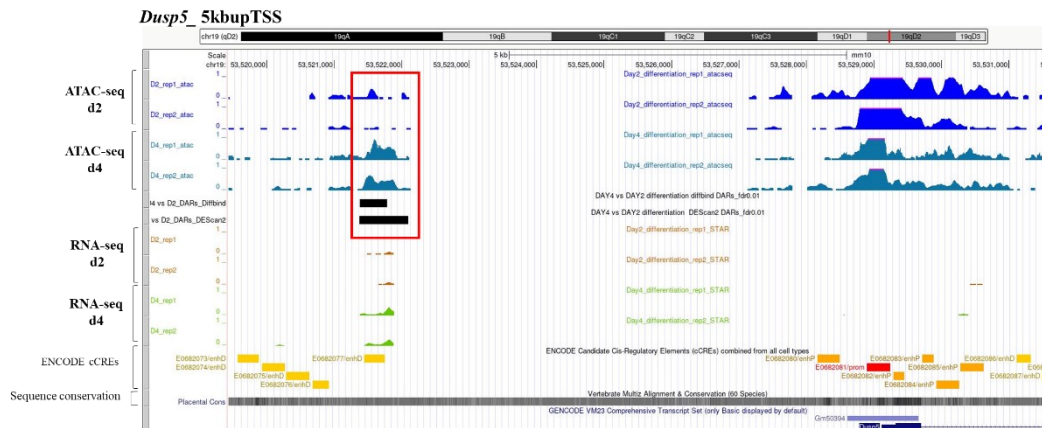


Figure 60.h ATAC peaks and RNA-seq peaks coverage associated to *Notch1*. On vertical axis there are the genome coverage of d2 first replicate, d2 second replicate, d4 first replicate and d4 second replicate. Red box indicates the open chromatin at d4 compared to d2. Bottom of figure, the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the mouse genome is showed; black parallel lines indicate detected conservation between species.



3.1 Putative endothelial enhancers validation strategies

Although enhancers can be predicted based on genome-wide datasets associated with open chromatin and other epigenomic features, the function of predicted enhancers needs to be experimentally validated.

For this reason, one of the aims of my PhD work was to determine whether any of these 10 putative regulatory elements was required for transcriptional regulation of the associated gene during CPM differentiation into ECs.

To validate the putative endothelial enhancers, I followed two strategies:

- 1) Genetic deletion of putative enhancers, using CRISPR-Cas9 technology;
- 2) Epigenetic repression/decommission, by targeting the deactivated Cas9 (dCas9) coupled with the repressive effector LSD1 (Lysine-specific histone demethylase 1) (dCas9-LSD1 system) onto the putative enhancer.

During the first period of my PhD project, I only focused on 2 putative enhancer regions, associated with *Pecam1* and *Notch1* genes, both of which are critical for vascular development. For this reason, I have focused on their validation using gene editing (by CRISPR-Cas9 system).

The Cas9 nuclease was directed by two sgRNA, designed to target the extremities of each specific putative enhancer region. Thus, the Cas9 will induce two double-strand breaks (DSB) at the targeted genomic region. The DSB is generally repaired by the error-prone repair pathway non-homologous end joining (NHEJ), resulting in the formation of insertion and deletion (indel) mutations. In detail, I have generated mESCs with deletion of putative *Pecam1*-enhancer and *Notch1*-enhancer separately, which I then differentiated towards CPM-EC lineages.

Subsequently, during the PhD course, I have extended the search of putative regulatory elements, identifying other 8 open chromatin regions, associated with *Kdr* (*Vegfr2*), *Cdh5* (*VE-Cadherin*), *CD34*, *Eng*, *Flt1* (*Vegfr1*), *Tall* (*Scl1*), *Dusp5* and *Gata6* endothelial genes. For these putative enhancers I have used the second validation strategy, based on epigenetic reprogramming by nuclease-deficient dCas9 fused with effectors, which could be transcription activator or repressor. In general, dCas9 fusion proteins can target and alter epigenetic marks in enhancers and promoters, thereby modulating gene expression. Kearns *et al.* fused dCas9 with the histone demethylase LSD1 (lysine-specific histone demethylase 1), which was implicated in enhancer repression (Kearns NA. *et al.*, (2015)). LSD1 removes mono and di-methylation of histone H3 lysine 4 (H3K4me1 and me2) to repress putative enhancers.

In their study, they demonstrated that epigenetic alterations in regulatory elements were sufficient to cause strong changes in gene expression. Such tools should therefore enable the functional annotation of regulatory elements.

Genome-wide mapping of histone post-translational modifications revealed that H3K4me1 (histone H3 Lys4 monomethylation) and H3K27ac (H3K27 acetylation) are enriched at active-enhancer regions (Ernst J. *et al.*, (2011)), whereas active-

promoter regions are marked by H3K4me3 and H3K27ac. This 'histone code' is widely used to annotate regulatory elements.

So, I decided to generate mESC clones constitutively expressing dCas9-LSD1 and delivered into them guide RNAs (gRNAs) targeting the putative enhancers. Then the cells were differentiated into endothelial cells (ECs).

Figure 61 illustrates the scheme of enhancer validation strategies.

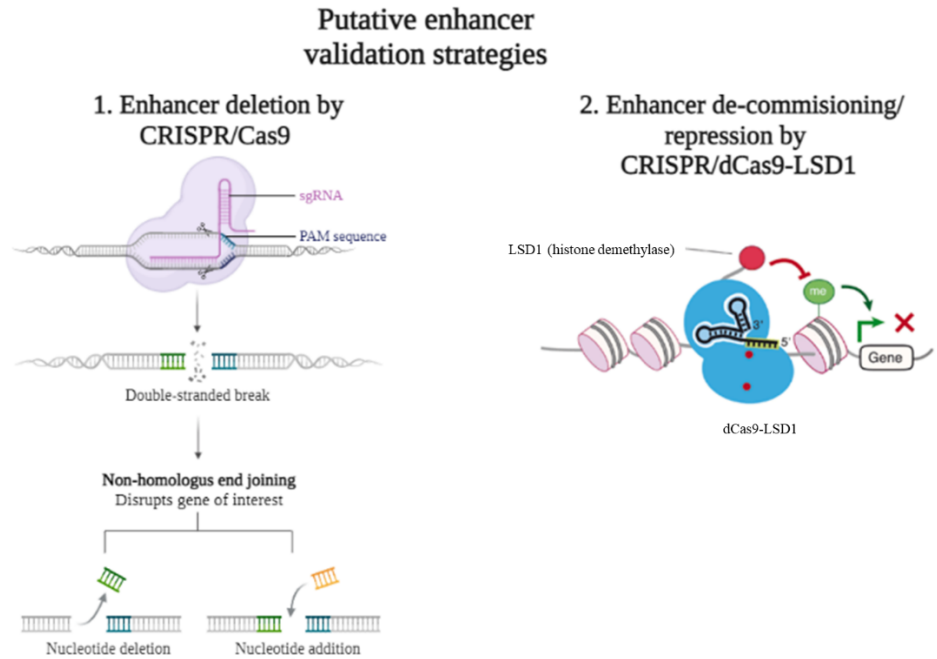


Figure 61 Scheme of enhancer validation strategies. The first approach (left) is based on deletion of the putative enhancer by CRISPR-Cas9 technology; the second one (right) is using epigenetic reprogramming by fusing LSD1 demethylase with dCas9, which can erase methylation of histone H3 lysine 4 (H3K4me2) near the enhancer region to abrogate gene expression.

Chapter 4.

Generation of mutant engineered mouse embryonic stem cell lines (mESC) using the CRISPR/Cas9 system.

During my PhD training, I had the opportunity to develop the CRISPR-Cas9 technique to induce precise genome editing in mouse embryonic stem cells (mESC). CRISPR/Cas9 is the most widely used tool for genome engineering projects due to its simplicity, versatility, efficacy, and low cost (Hsu *et al.*, (2014)). CRISPR-Cas9 system relies on a universal Cas9 nuclease, that can generate a DNA double-strand break (DSB) when combined with a single-guide RNA (sgRNA) to form a ribonucleoprotein complex (RNP). In this RNP complex, sgRNA will guide the Cas9 to a specific locus in the genome, thus allowing nuclease activity and cleavage of the target site. The sgRNA can be designed to target any 20-nucleotide-long sequence that must be followed in the targeted sequence by a 5'-NGG tri-nucleotide recognition site, called protospacer adjacent motif (PAM).

In particular, the experimental process to generate mutant mESC lines via CRISPR-Cas9 includes four main phases (schematic workflow is shown in Figure 62).

(1) Transfection of CRISPR reagents: delivering the reagents (single guide RNA, Cas9, and if required, a DNA donor template for homologous recombination) in the parental mESC line to introduce a targeted DNA double strand break (DSB). The DSB will be repaired by the endogenous DNA repair pathways. The non-homologous end-joining (NHEJ) mechanism leads to the introduction of small insertions/deletions (indels), while the Homology-directed repair (HDR) pathway introduces precise exogenous nucleotides (by providing a DNA donor template with homology to the target site).

(2) Clone isolation: it can be facilitated by adding a positive selection marker (e.g., sgRNA fluorescent complex, or antibiotic resistance gene) to specifically select the transfected cells. Cells that express a fluorescent reporter are selected by fluorescence-activated cell sorting (FACS), whereas cells that include an antibiotic resistance gene will survive antibiotic selection. Then, a pool of transfected cells is plated at low density and later, colonies are picked and manually transferred in individual wells.

(3) Screening: is performed to identify the correctly modified clones. At the first, a PCR-based screen (using primers that surround the target region) is performed to quickly detect clones with the desired modification. Then, positive clones are checked by Sanger sequencing to determine their exact sequence.

Another important point is this step is gene editing efficiency by T7 endonuclease 1 (T7E1) mismatch detection assay (Vouillot L. *et al.*, (2015)). T7 endonuclease 1 recognizes and cleaves structural deformities in DNA heteroduplexes. Genomic

DNA of a pool of transfected cells is amplified and PCR product is then denatured and reannealed to produce fraction of heteroduplexes of mutant and wildtype PCR amplicons. DNA mismatches in those heteroduplexes are recognized and cleaved by T7E1, and the cleavage products are easily detectable by an agarose gel analysis. A schematic example of a DNA mismatch detection T7 assay is shown in Figure 63.

(4) *Molecular characterization*: the selected clones are finally characterized and validated using additional analyses. For example, the detection of pluripotency markers (e.g., *Nanog*, *Oct3/4*, *Rex1*) by qPCR to assess the pluripotency status of the selected clones; RNA expression of the modified target gene in clones harboring the mutation; the ability of mutated mESC to differentiate.

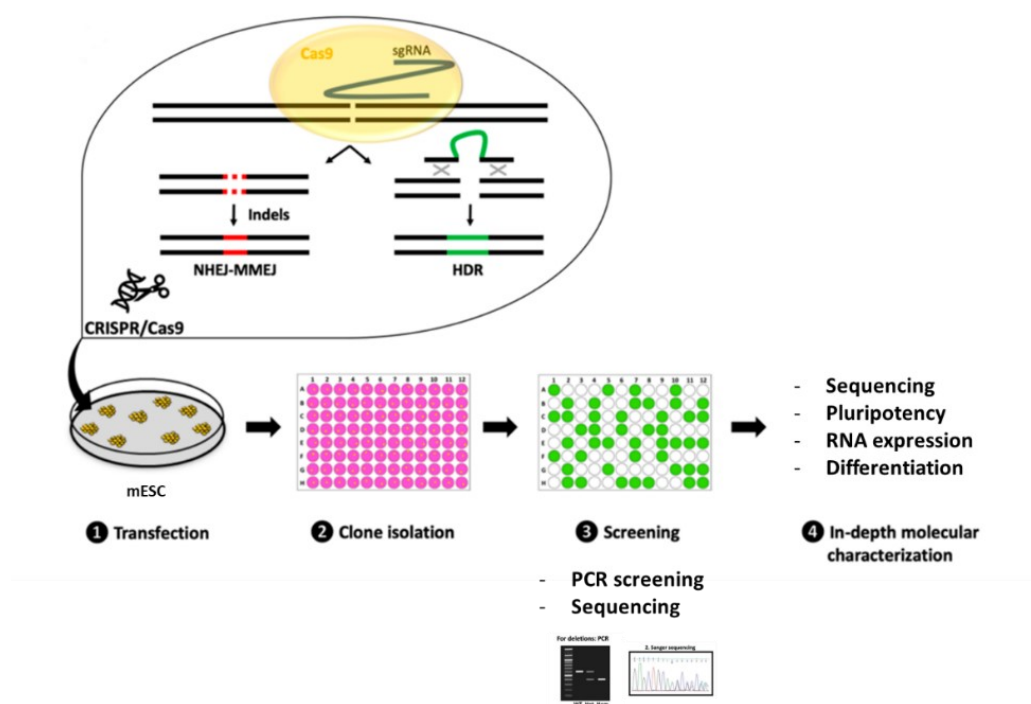


Figure 62 Experimental pipeline to generate mutant engineered mouse embryonic stem cell lines (mESC) using the CRISPR/Cas9 system (Mianné J. *et al.*, (2020).

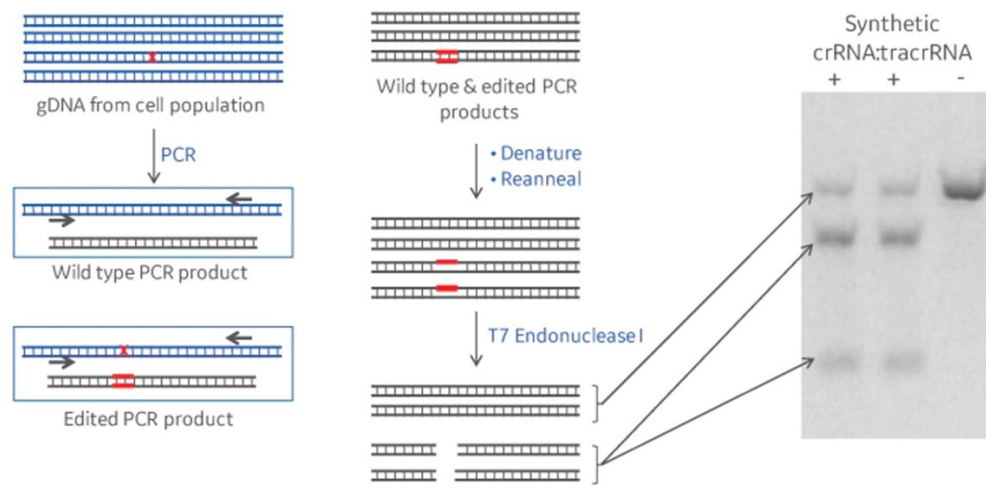


Figure 63 Schematic example of T7 endonuclease 1 (T7E1) mismatch detection assays. Genomic DNA (in blue) from edited cells contains wild type and edited DNA (mutation in red). PCR amplification around the editing site generates wild type and edited PCR products (in black). Denaturing and reannealing of these products generates a fraction of heteroduplexes of mutant and wildtype PCR amplicons. Mismatches are cleaved by the T7E1 endonuclease. Running these PCR products on a gel resolves full length DNA and cleavage products. Gel shows untreated (-) and cells edited with Cas9 and crRNA:tracrRNA (+) (<https://horizondiscovery.com/en/resources/featured-articles/proper-assessment-of-gene-editing-with-dna-mismatch-detection-assays>).

4.1 *Tbx1* knockout (*Tbx1*^{-/-}) mESC lines

The first CRISPR-Cas9 strategy used in our laboratory was performed to induce the knockout of the gene *Tbx1* in mESC. Our goal was to produce a cell-based model of the mutation in exon 5 of the murine gene *Tbx1*, previously generated *in vivo*.

In details, mouse ES cells were targeted in order to generate a homozygous *Tbx1* loss of function mutation by inserting multiple stop codons and polyA signals into exon 5 by homologous recombination of a DNA donor sequence. The single guide RNA (sgRNA) used was complementary to a specific region of *Tbx1*-exon5. The DNA donor sequence was an exogenous template, containing homology arms for the DNA target and also carried out multiple stop codons, polyA signals, enzymatic restriction site (*EcoRI*) and V5-Tag. A scheme of the strategy is depicted in Figure 64.

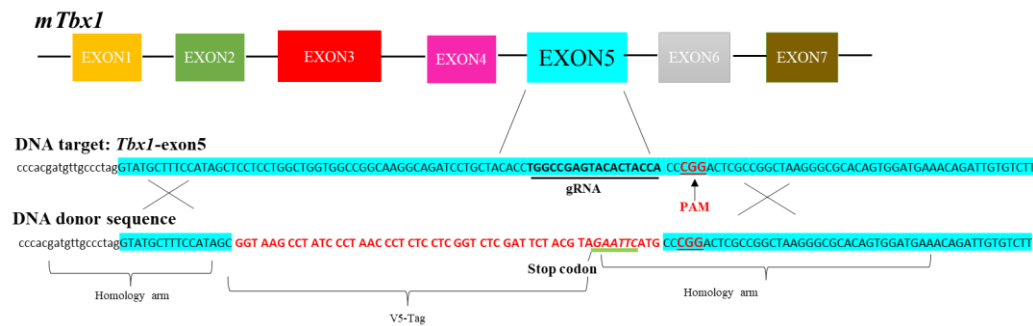


Figure 64 Strategy to generate a knockout allele of *Tbx1* using CRISPR-Cas9 and homologous recombination. The bottom line indicates the DNA target: WT *Tbx1*-exon5 sequence (in light blue); in bold black the gRNA sequence; in bold red underlined the PAM sequence (CGG). The bottom line indicates the sequence of the recombinant allele (DNA donor, in bold red), inserted by homologous recombination. This includes a V5-Tag; a stop codon (TAG) and a diagnostic *EcoRI* digestion site (green line).

To estimate the CRISPR-Cas9-mediated editing efficiency in a pooled cell population, I used T7 endonuclease 1 (T7E1) mismatch detection assays, as previously described. Overall, I observed high frequency of mutations detected by the T7E1 test. In Figure 65, the arrows indicate bands related to the cleavage products of expected sizes (range 100-200bp). As can be seen, Cas9 nuclease, when combined with specific gRNA, provides consistent and effective gene editing.

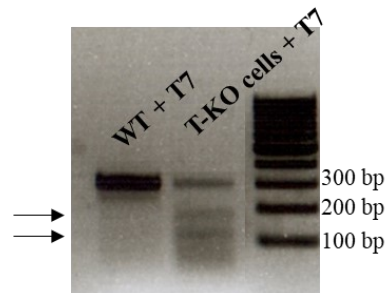


Figure 65 Evaluation of active Cas9 + gRNA efficacy using T7 Endonuclease 1 assay. Control cells (WT) show a single band corresponding to uncut amplicon. Amplicons from modified cells (*Tbx1* -KO) have 3 bands: 1 unmodified + 2 cleavage products.

I obtained two correctly targeted homozygous mutant clones, named 4D and 5H, confirmed by diagnostic PCR screening (using primers that surround the target region), followed by *EcoRI* restriction reaction, and by sequencing. The results showed the correct insertion of the precise homologous nucleotides in the target site (Figure 66.a-b).

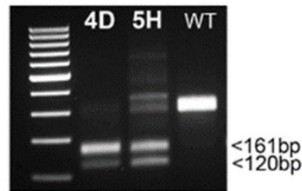


Figure 66.a PCR amplification of the targeted region from *Tbx1* homozygous clones 4D and 5H, and from WT digested with *EcoRI*. The restriction reaction produces two fragments, 161bp and 120bp, in mutant clones, while in WT is visible single PCR product at ~300bp.

> *Tbx1* WT exon5
AGGTAGTCTGTGGGAGTCTTCTGGTTACCTATATTGTATGACCCCCCTCCCACGATGTTGCCCTAG
GTATGCTTCCATAGC
TCCTCCTGGCTGGTGGCCGGCAAGCAGATCCTGCTACACCTGGCCGAGTACACTACCA
CCCGGACTCGCCGGCTAAGGGCGCACAGTGGATGAAACAGATTGTGTCTTTTCGACAAGCTGAAACTGACCAATAA
CCTGCTGGATGACAATGGCCATGTAGGCGACCCTTCTCTG

> *Tbx1-KO* clone 4D
GACGTAGGTGGGTGGGAGGCTTCTGGTTACCTATATTGTATGACCCCCCTCCCACGATGTTGCCCTAG
GTATGCTTCCATAGC
GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGTAGAATTCATG
CCCGGACTCGCCGGCTAAGGGCGCACAGTGGATGAAACAGATTGTGTCTTTTCGACAAGCTGAAACTGACCAATAA
CCTGCTGGATGACAATGGCCATGTAGGCGACCCTTCTCTG

> *Tbx1-KO* clone 5H
ACGGAGGTGGCTGGTCAGTCTTCTGGTTACCTATATTGTATGACCCCCCTCCCACGATGTTGCCCTAG
GTATGCTTCCATAGC
GGTAAGGGTATCCCGAACCCTCTGCTCGGTCTCGATTCTACGTAGAATTCATG
CCCGGACTCGCCGCGAAGGGTGCACAGTGGATGAAACAGATTGTGTCTTTTCGACAAGCTGAAACTGACCAATAA
CCTGCTGGATGACAATGGCCATGTAGGCGACCCTTCTCTG

Figure 66.b Sequence of WT and homozygous clones (4D-5H). In bold black is indicated the WT sequence, that is replaced by the exogenous recombinant sequence (in bold red), in both samples 4D and 5H. *Tbx1*-exon5 sequence is in light blue.

In laboratory, clone 5H *Tbx1*-KO was selected and used for experiments of my colleague's paper, to explore genome-wide gene expression and chromatin remodeling in cellular models of *Tbx1* loss of function (Cirino A. *et al.*, (2020)). Clone 5H did not express any *Tbx1* mRNA, when it and the parental cell line (WT) were subjected to a differentiation protocol. WT cells expressed *Tbx1* at day 4, while no expression was detected in *Tbx1*^{-/-} cells (Figure 67).

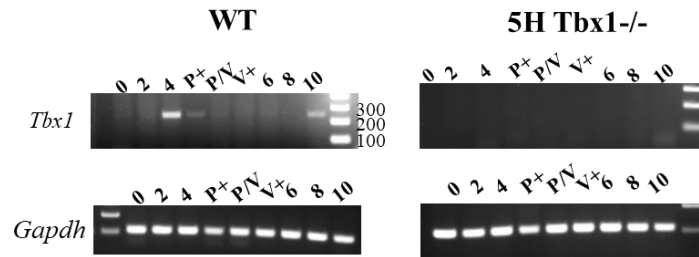


Figure 67 *Tbx1* expression revealed by reverse transcription PCR. Left panel: PCR of samples collected at the differentiation stages indicated on WT mES cells. At day 4, the analysis was performed on total populations (T) and on FACS-purified subpopulations. The right panel shows the same experiment performed using the *Tbx1*^{-/-} clone 5H. P, PDGFR α ; K, KDR (Cirino A. *et al.*, (2020)).

TBX1 encodes a T-box transcription factor, a family of DNA binding proteins that has important roles in development and their mutation is associated with developmental disorders in humans and mice. *Tbx1* is expressed in several tissues, but its mesodermal domain is critical for heart development, suggesting that the major role of *Tbx1* in heart development is effected in precursors destined to populate the heart (Zhang Z. *et al.*, (2006)). Previous data obtained in my laboratory demonstrated that *Tbx1* is expressed in multipotent cardiac progenitors cells (CPCs) that, in clonal assays, can give rise to 3 heart lineages expressing endothelial, smooth muscle and cardiomyocyte markers (Chen L. *et al.*, (2009)). Moreover, *Tbx1* expression in endothelial cells (ECs) is essential for vascular development in mice. EC-specific inactivation of *Tbx1* leads to lymphatic and brain vessel anomalies. *Tbx1* regulates two essential vascular genes, *Vegfr2* and *Vegfr3*, in ECs mutants (Chen L. *et al.*, (2010); Cioffi S. *et al.*, (2014)). Based on these data, in my future experiments, I would use clone 5H *Tbx1*-KO to determine whether *Tbx1* affects lymphatic EC differentiation and by what mechanism.

4.2 *Pecam1*- Δ *enh. intron2* mESC lines

I used CRISPR/Cas9 technology to induce the deletion of the putative enhancer identified in intron2 of *Pecam1* gene during differentiation (as previously described). For the deletion of a large segment of genomic DNA (569bp), I used two guide RNAs against the targeted locus, which recognized the extremities of the segment (as indicated in Figure 68-Figure 69) (Zheng Q. *et al.*, (2018)). In this case, the DBS generated by CRISPR/Cas9 are joined through the error-prone nonhomologous end-joining (NHEJ), which results in the loss of the segment between the two DBSs.

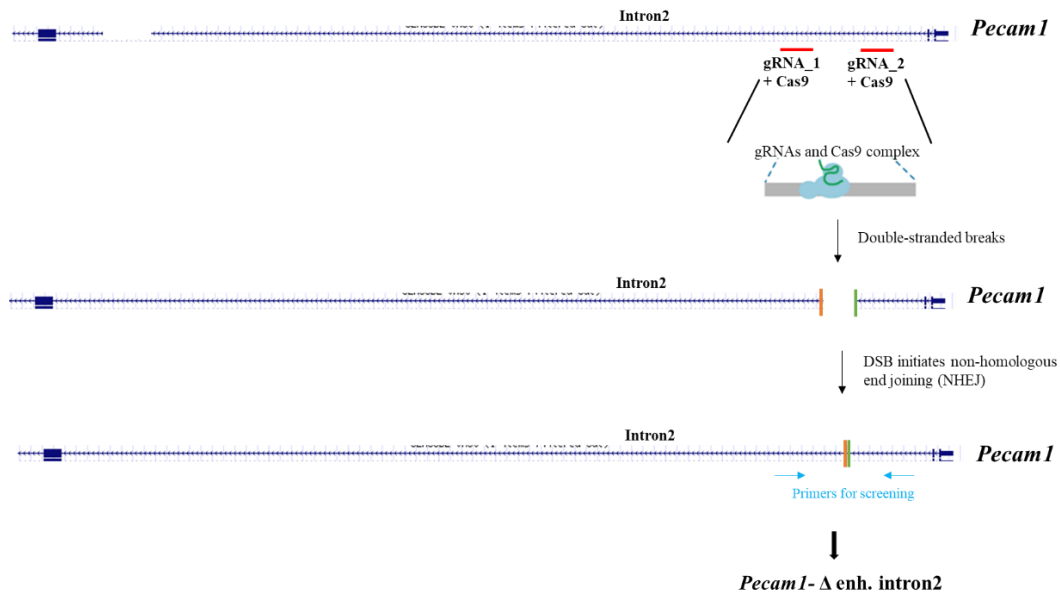


Figure 68 Scheme of the steps of targeted *Pecam1*- enhancer deletion with CRISPR/Cas9. Red lines indicate the position of the two gRNAs used.

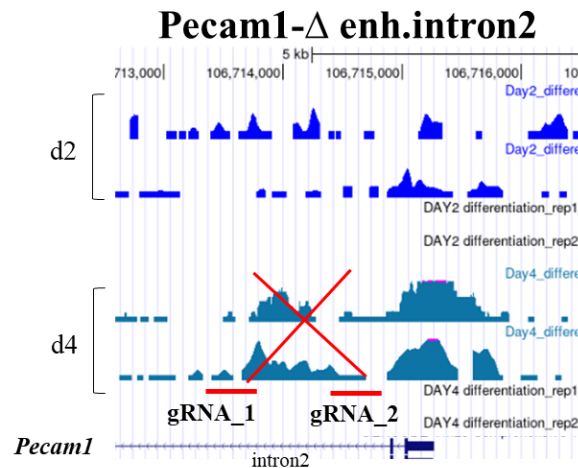


Figure 69 Scheme of the CRISPR-Cas9 strategy used to delete the enhancer region, located in intron2 of *Pecam1* gene. On vertical axis there are the ATAC-seq peaks coverage of d2 (two replicates) and d4 (two replicates). Red lines indicate the gRNAs positions.

I harvested a pool of cells transfected with RNP complex and used the genomic DNA to amplify by PCR the region flanking the target site. The PCR product was denatured, followed by re-annealing, leading to a mix of double strand fragments, some of which contain mismatches. These mismatches were detected by the T7 endonuclease 1 test and resolved on an agarose gel. I used WT cells as control.

As showed in Figure 70, the presence of three shorter bands, compared to WT sample, of predicted sizes suggested that CRISPR/Cas9 editing had successfully introduced mutations at the targeted site.

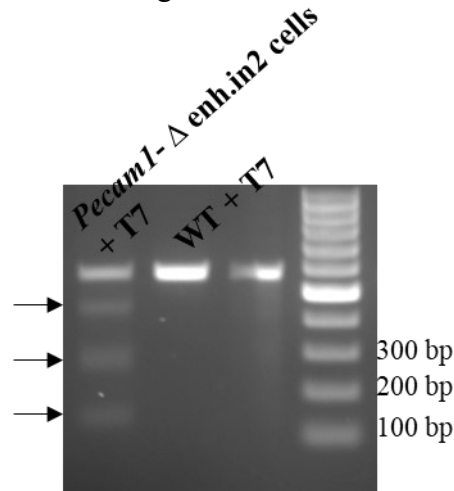


Figure 70 T7 Endonuclease I digestion of genomic amplicons by WT cells and pool of *Pecam1*- Δ enh. intron2 cells. Control cells (WT) have only a single band corresponding to uncut amplicon. Amplicons from modified cells (*Pecam1*- Δ enh. intron2) transfected with active Cas9-gRNAs yield 4 bands: 1 unmodified + 3 cleavage products of predicted sizes.

I obtained 17 correctly targeted mutant clones, confirmed by diagnostic PCR screening, using primers that surround the target region (see Figure 71).

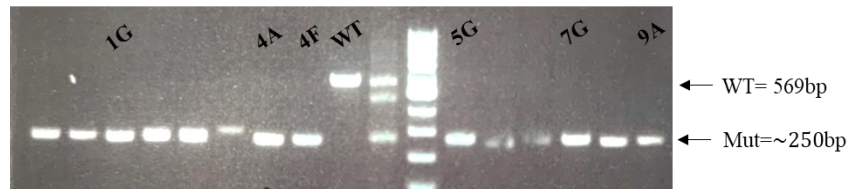


Figure 71 Diagnostic PCR amplification of the targeted enhancer region from *Pecam1*- Δ enh. intron2 clones and from WT cells. WT fragment = 569bp; Mutant fragment = ~250bp.

6 out of 17 positive clones (*Pecam1*- Δ enh. intron2) are checked by sequencing to determine their exact sequences. Subsequently, for my experiments, I selected only two mutant *Pecam1*- Δ enh. intron2 clones, named 5G and 7G (Figure 71). In Figure 72 is depicted the sequence of the two clones 5G and 7G, confirming the deletion of the entire WT target region (in bold black) located between the two specific gRNAs (sequence in light blue).

> ***Pecam1* WT enh. intron2** [chr11:106575022-10657559]

CCTGACATTCTAGGCTGAGTCATGTCTTTCTGAGGCTCACCCATCCCACCCTTTAGTGGGGATGACTTGAAGACAATA
CACAGTCTGTCTCCAGGTGTTGCCAAGGAGCCAGAACCTAAGGAATCACAGACTCACAGACCACAGCCCTCC
CCACCTTGGGGCCACAGCTCCATCCCTGGCTCAACTCAGCCCTGTCACGTTCTTCAGCTCAAAAATGCCTGCA
GCCCAGCATCCCTAGTCTTCCACACTGGAATAACTTTTGACTGAGCCAGTGAATCTGTGGCCTTCTTCCTG
ACCCACAATCCCTGGGCTGGGCTGAGCCTCTTCCTTCAGAGTCTCTGATGGACGGAGCGAGACAGATCCTCT
TAATCTGAAAAATTCGGAGGTTCCGAATCAGCTCTC

GAGTGGATGATAGGTCTGACTTTAAAGACAGGCAGAACCCTTCCCCACAAGCAAGCTCAGATAGGAGCGGAGCCTCC
CCGGTGGGAGGGAAAGGGGATCTGGCTGGTCCTTTTGTACTCTCTTAGGACAGAGGGA

> ***Pecam1* Δ enh. intron2 clone 5G**

CCTGACTCAGGCTGAGTCATGTCTTTCTGAGGCTCACCCATCCCACCCTTTAGTGGGGATGACTTGAAGACAATA
CACAGTCTGTCTCCAGGTGTTGTC

GAGTGGATGATAGGTCTGACTTTAAAGACAGGCAGAACCCTTCCCCACAAGCAAGCTCAGATAGGAGCGGAGCCTCC
CCGGTGGGAGGGAAAGGGGATCTGGCTGGTCCTTTTGTACTCTCTAGGTACAGAGGGA

> ***Pecam1* Δ enh. intron2 clone 7G**

CGAACTACTCTAGGCTGAGTCATGTCTTTCTGAGGCTCACCCATCCCACCCTTTAGTGGGGATGACTTGAAGACAATA
CACAGTCTGTCTCCAGGTGTTGCC

GAGTGGATGATAGGTCTGACTTTAAAGACAGGCAGAACCCTTCCCCACAAGCAAGCTCAGATAGGAGCGCACTCC
CCCGTGGGAGGGAAAGGGGATCTGGCTGGTCCTTTTGTACTCTCTAGAGGTTTCCTCCTT

Figure 72 Sequence of WT *Pecam1*-enh. intron2 and mutated clones (5G-7G). In bold black is indicated the WT sequence, that is deleted in both samples 5G and 7G. In light blue are gRNAs sequences, while in red are indicated the PAM sequence (5'-NGG).

4.3 *Notch1*-Δ enh. intron15 mESC lines

As described in the previous paragraph about the putative enhancer in intron2 of *Pecam1* gene, I used CRISPR/Cas9 technology to also induce the deletion of the putative enhancer identified in intron15 of *Notch1*. For the deletion of a large segment of genomic DNA (664bp), I used two guide RNAs against the targeted locus, which recognized the extremities of the segment (as indicated in Figure 73 - Figure 74). Also in this case, the two DBSs generated by CRISPR/Cas9 are joined through the error-prone nonhomologous end-joining (NHEJ), which results in a deletion.

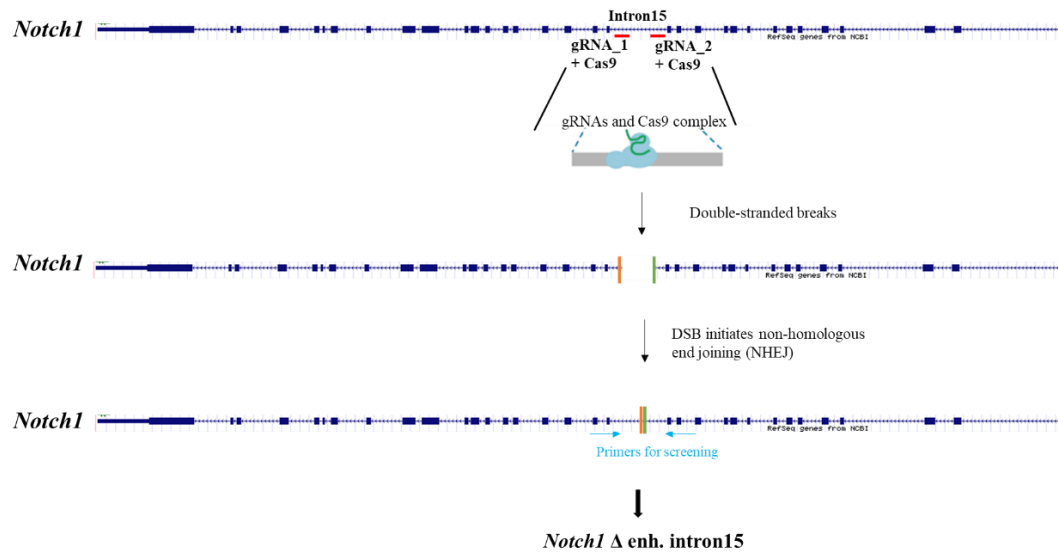


Figure 73 Scheme of the steps of targeted *Notch1*- enhancer deletion with CRISPR/Cas9. Red lines indicate the position of the two gRNAs used.

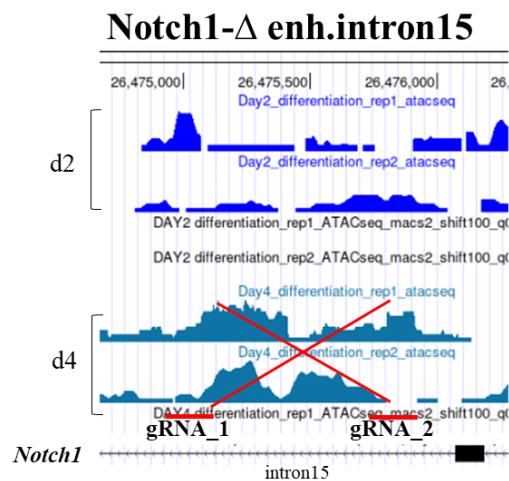


Figure 74 Scheme of the CRISPR-Cas9 strategy used to delete the enhancer region, located in intron15 of *Notch1* gene. On vertical axis there are the ATAC-seq peaks coverage of d2 (two replicates) and d4 (two replicates). Red lines indicate the gRNAs position.

I carried out T7 endonuclease 1 assay to evaluate the efficiency of CRISPR/Cas9 genome editing. As can be seen in Figure 75, modified cells (*Notch1*-Δ-enh. intron15) transfected with active Cas9-gRNAs showed 3 different PCR products: 1

unmodified + 2 cleavage products of predicted sizes. The results suggested that indel mutations were successfully introduced in the target region.

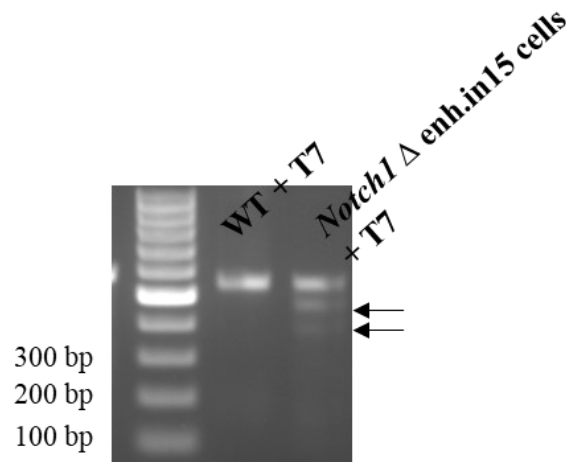


Figure 75 T7 Endonuclease I Assay to validate the efficacy of CRISPR/Cas9 + gRNAs strategy in mESC cells. Control cells (WT) have only a single band corresponding to uncut amplicon. Amplicons from modified cells (*Notch1*- Δ -enh. intron15) transfected with active Cas9-gRNAs yield 3 bands: 1 unmodified + 2 cleavage products of predicted sizes.

Two correctly targeted mutant clones, named 7G and 11B, were confirmed by diagnostic PCR screening, using primers that surround the target region and by sequencing to determine the exact sequences (see Figure 76).

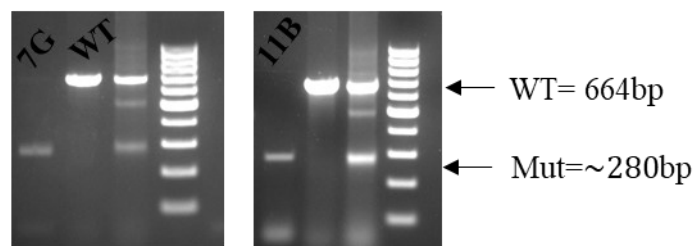


Figure 76 Diagnostic PCR amplification of the targeted enhancer region from *Notch1*- Δ -enh. intron15 clones and from WT cells. WT fragment = 664bp; Mutant fragment = ~280bp.

In Figure 77 is depicted the sequence of the two clones 7G and 11B, confirming the deletion of the entire WT target region (in bold black) located between the two specific gRNAs (sequence in light blue).

> *Notch1* WT enh. intron15 [chr2:26330427-26331154]

ATAACACTAGGGCAGTGAAATCTGATAACCG

CTTGGGGGAGCCA**CGAGTAACCCAGGTGACTCT****CCT**TGCAATCCCAGTTAGCTTTCCTGTTTATGGCCATGGTTG
GCTCCAGGAGCTAAGCAAGGAAACCTTATCTTCCTTGTCTCCACCATGTCCCAGCCCACTGGTGAGACTCT
CCAGGGCCTCTCTCTATCGGTTTCCACGGTGACCTGGGCAGACAGGAACCTTTGACAGAGTTTCCGACAATT
GTGCAAAGGGAAGCAGGAAGCTGGTCAGCGCGGCTTGACTCTCCCGCTAGCCCCGATAGCAGCAAGACCAA
AAGGAAACATGGGGTGTGCCTGGCTGCTGACAGTTTC

CTCACTGACCAACACAGGAGACAAAATCCAGGCCCTTTAACTTTGTCA**AAGAAAACAGCATCCTGCCAGG**

CATCATAGGAACACTGGAAGGGCCTCCACCTTCACAGACCCACTTGCTATGTCCAGAACCTAGCAGCAGCAGCATGA
AGCCACCAGTACTCTGTGGCAGTATTTGGAGCTGCAGGGTTGGGGGCCATGTGGTCACGGCTTTCATGTTCAATGCAC
ATCTCTCTGCAGGGGCA

> *Notch1* Δ enh. intron15 clone 7G

ATATCACACTAGGGCAGTGAAATCTGATAACCG

CATAGGAACACTGGAAGGGCCTCCACCTTCACAGACCCACTTGCTATGTCCAGAACCTAGCAGCAGCAGCATGAAGC
CACCAGTACTCTGTGGCAGTATTTGGAGCTGCAGGGTTGGGGGCCATGTGGTCACGGCTTTCATGTTCAA
TGCACATCTCTCTGCAGGGGTCTATTTCTCGCCATTGACCA

> *Notch1* Δ enh. intron15 clone 11B

TATTAGCTTCTGCTTCTTGTCTCTGCAGTCATAGTTAGTGAAATCTGATAACCG

CTTGGGGGAGTATTTT

CATAGGAACACTGGAAGGGCCTCCACCTTCACAGACCCACTTGCTATGTCCAGAACCTAGCAGCAGCAGCATGAAGC
CACCAGTACTCTGTGGCAGTATTTGGAGCTGCAGGGTTGGGGGCCATGTGGTCACGGCTTTCATGTTCAATGCACATC
TCTCTGCAGGGGTCTATTTCTCGCCATT

Figure 77 Sequencing of WT *Notch1*-enh. intron15 and mutated clones (7G-11B) to determine their exact sequences. In bold black is indicated the WT sequence, that is deleted in both samples 7G and 11B. In light blue are gRNAs sequences, while in red are indicated the PAM sequence (5'-NGG).

4.4 *dCas9-LSD1 mESC lines*

To generate mESC lines expressing dCas9 fused with LSD1 repressor, I decided to deliver the plasmid p-dCas9-LSD1-Hygro (Addgene plasmid # 104406, kindly donated by Dr. Beck lab) into the cells by electroporation. As shown in plasmid map, in Figure 78, it dCas9 from the *Streptococcus pyogenes* Type II CRISPR/Cas system; three tandem FLAG epitope tags (3xFLAG) and LSD1A isoform B catalytic domain (a.a. 171-852). Moreover, a HygR cassette confers resistance to hygromycin.

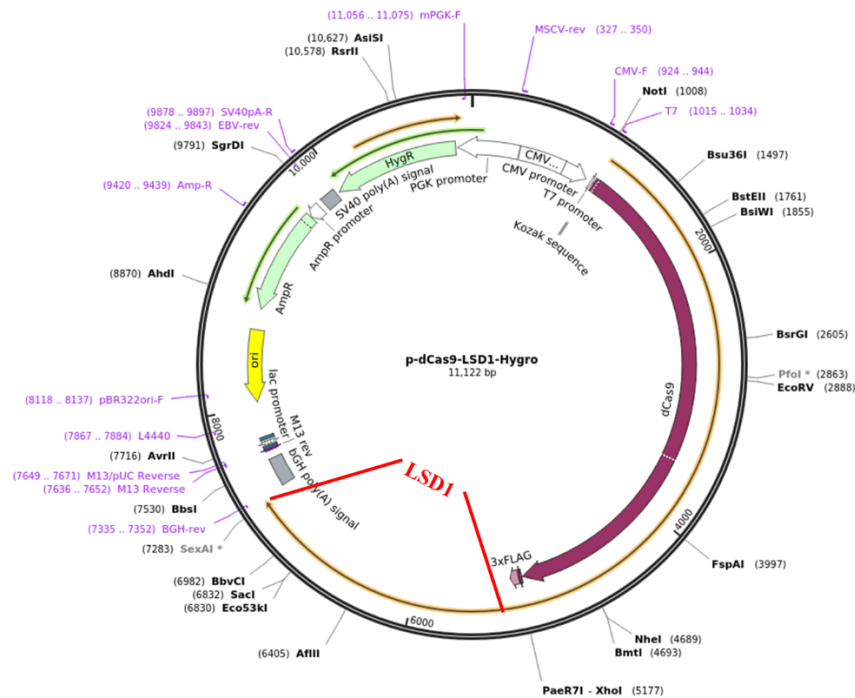


Figure 78 Full sequence map for p-dCas9-LSD1-Hygro (<https://www.addgene.org/104406/sequences/>).

24 hours after electroporation, cells were maintained in Hygromycin B selection for 10 days. Individual colonies were isolated (n.25), expanded and screened by PCR for inserted sequence in the DNA, and for RNA expression. I designed specific primers for different fragments of the plasmid: CMVpromoter-dCas9; dCas9start_sequence; dCas9middle_sequence; dCas9end_sequence; dCas9end-3xFLAG-LSD1 and LSD1-bGH poly(A) signal.

I obtained 3 correctly targeted mutant clones, named A2; A8, B1 (indicated with red asterisk in Figure 79).

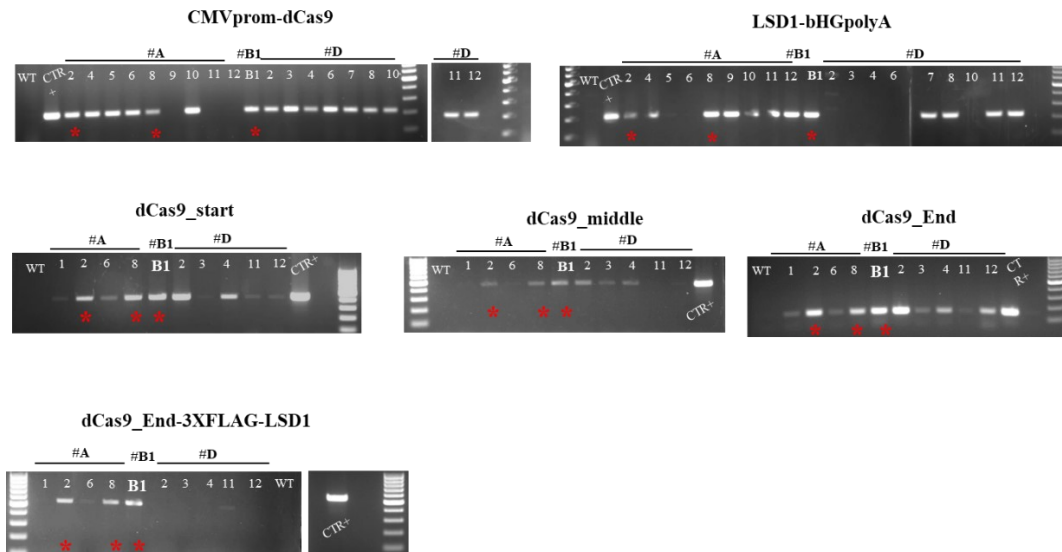


Figure 79 Diagnostic PCR amplification of the p-dCas9-LSD1 expressing clones. WT sample is parental mESC (no targeted), CTR+ sample is parental mES cells transiently transfected with the plasmid p-dCas9-LSD1-Hygro. Red asterisk indicates the 3 mutant clones (A2; A8; B1).

To determine if they expressed the sequence region between dCas9 - 3xFLAG and LSD1 fusion protein, the three clones were sequenced using primers specific for this region (indicated previously as dCas9_End-3xFLAG-LSD1).

In Figure 80, in violet is depicted the final part of dCas9 sequence, in bold black is indicated the 3xFLAG, in yellow is LSD1 sequence. The results suggested that A2, A8, B1 clones contained in their sequences the correct fragment, carrying dCas9_end part (in violet in Figure 80); 3xFLAG (in bold black in Figure 80) and LSD1 (in yellow in Figure 80). The three clones were compared with the sequence of the plasmid p-dCas9-LSD1, available on the Addgene site (<https://www.addgene.org/104406/sequences/>). I selected clone B1 p-dCas9LSD1 for all my experiments.

>p-dCas9LSD1 plasmid sequence WT

GATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATAT
GAAACTCGGATAGATTGTCACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACT
ACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGAGG
CGGAGGTAGCCCATCGGGTGTGGAGGGCGCAGCTTCCAGAGCCGACTTCCTCATGACCGGATGACTTCTCAAGAA
GCAGCCTGTTTTCCAGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTCTTTTCATTAGAAACCGCACACTG
CAGTTGTGGTTGGATAATCCAAAGATTCAGCTGACATTTGAGGCTACTCTCCAACAATTAGAAGCACCTTATAACAGT
GATACTGTGCTTGTC

>clone A2 p-dCas9LSD1

GATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATAT
GAAACTCGGATAGATTGTCACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACT
TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGG
AGGCGGAGGTAGCCCATCGGGTGTGGAGGGCGCAGTTTTCCAGAGCCGACTTCCTCATGACCGGATGACTTCTC
AAGAAGCAGCCTGTTTTCCAGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTCTTTTCATTAGAAACC
GCACACTGCAGTTGTGGTTGGATAATCCAAAGATTCAGCTGACATTTGAGGCTACTCTCCAACAATTAGAAGCAC
CTTATAACAGTGATACTGTGCTTGTC

>clone A8 p-dCas9LSD1

GATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATG
AAACTCGGATAGATTGTCACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAAGCGACT
ACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA
GGCGGAGGTAGCCCATCGGGTGTGGAGGGCGCAGCTTCCAGAGCCGACTTCCTCATGACCGGATGACTTCTCA
AGAAGCAGCCTGTTTTCCAGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTCTTTTCATTAGAAACCG
CACACTGCAGTTGTGGTTGGATAATCCAAAGATTCAGCTGACATTTGAGGCTACTCTCCAACAATTAGAAGCAC
TTATAACAGTGATACTGTGCTTGTC

>clone B1 p-dCas9LSD1

GATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATG
AAACTCGGATAGATTGTCACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACT
ACAAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA
GGCGGAGGTAGCCCATCGGGTGTGGAGGGCGCAGCTTCCAGAGCCGACTTCCTCATGACCGGATGACTTCTCAA
GAAGCAGCCTGTTTTCCAGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTCTTTTCATTAGAAACCGCA
CACTGCAGTTGTGGTTGGATAATCCAAAGATTCAGCTGACATTTGAGGCTACTCTCCAACAATTAGAAGCACCTTA
TAACAGTGATACTGTGCTTGTC

Figure 80 Sequencing of plasmid p-dCas9LSD1 WT and mutated clones (A2, A8, B1) to determine the correct insertion of dCas9-LSD1 fusion protein. In violet is depicted the final part of dCas9 sequence, in bold black is indicated the 3xFLAG, in yellow is LSD1 sequence. All three clones show the correct insertion of the sequence.

Chapter 5.

Validation of putative endothelial enhancers and verification of their requirement during EC differentiation

5.1 *Notch1-intron15 enhancer validation: evaluation of gene expression consequences during ECs differentiation.*

Based on literature data described above (Section 4.2), I decided to investigate the function of the putative enhancer region identified in intron15 of *Notch1* gene in my experiments, assessing its requirement during mESC differentiation in ECs (Figure 60.h, Chap.3).

To validate this regulatory element, I followed the two previously described approaches:

- 1) Enhancer deletion, using CRISPR-Cas9 technology;
- 2) Epigenetic enhancer repression/decommission, by dCas9-LSD1 system.

5.1.1 *Assessment of Notch1-intron15 putative enhancer using CRISPR/Cas9 deletion strategy*

As previously mentioned in Chap. 4.3, one of the first putative enhancer identified and tested was in intron15 of *Notch1*.

I used CRISPR/Cas9 technology to induce the deletion of the specific target region, generating two *Notch1*- Δ enh.intron15 mESC lines, named #7G and #11B clones, which were differentiated into ECs. I collected the cells at day4, day6 and day8 of differentiation and evaluated the expression level of the *Notch1* gene. A schematic experimental workflow is illustrated in Figure 81.

My hypothesis was that if the regulatory element in *Notch1*-intron15 acts as an enhancer element, its deletion should cause dysregulation of *Notch1* gene expression consequences during EC differentiation.

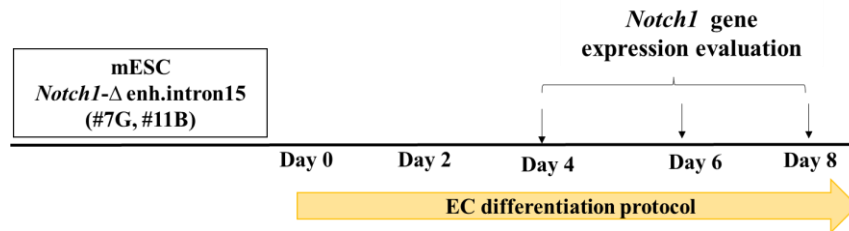


Figure 81 Overview of the differentiation scheme. *Notch1*- Δ enh.intron15 clones (#7G and #11B) are differentiated in ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Notch1* expression.

The first clone analyzed was **#11B *Notch1*- Δ enh.intron15**.

Results showed that *Notch1* expression is significantly reduced in clone #11B *Notch1*- Δ enh.in15 at d6 and d8 during EC differentiation (Figure 82). The experiments were done on five biological replicates (n. =5).

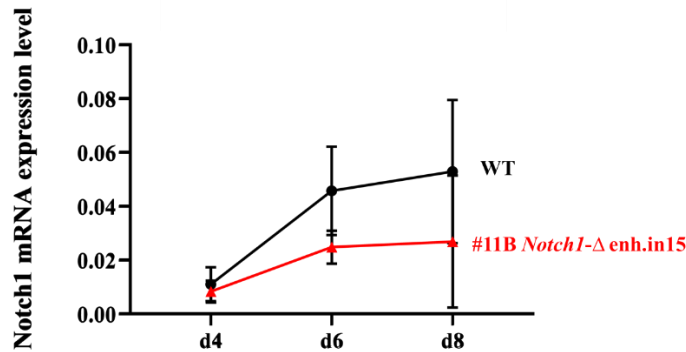


Figure 82 Quantitative real-time PCR analysis of *Notch1* mRNA expression level in clone #11B *Notch1*- Δ enh.in15 during EC differentiation. *Notch1* is reduced in clone #11B sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).

Then, I carried out statistical analysis, using GraphPad Prism8 software, choosing at first **two-way repeated measures ANOVA test** (ANOVA 2way-RM).

A two-way repeated measures ANOVA (also known two-factor or two-way ANOVA with repeated measures) compares the mean differences between groups that have been split on two within-subjects factors (also known as independent variables). This test is often used in studies where you have measured a dependent variable (in this case “gene expression level”) over two or more conditions (i.e., the two factors are “time points” and “genotype”). The primary purpose of a two-way repeated measures ANOVA is to understand if there is a statistically significant interaction between these two factors on the dependent variable.

Results from ANOVA 2way-RM analysis suggested that there was a significant change in all dataset population across the two factors “genotype” and “time” (Table 14).

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.0215	*	yes
Genotype	0.0220	*	yes
Interaction: time x genotype	0.0204	*	yes

Table 14 Two-way repeated measures ANOVA results, when compared WT and clone #11B *Notch1*-Δ *enh.in15* dataset. p-value < 0.05 is considered significant.

However, two-way repeated measures ANOVA cannot tell which specific groups within each factor were significantly different from each other.

In order to find out exactly which groups are different from each other, I conducted subsequent tests, by pairwise comparisons of the two samples, WT and #11B *Notch1*-Δ *enh.in15* cells, at different time points (d4, d6, d8), separately.

Since the number of biological replicates (n.=5) was too small, I selected in GraphPad Prism8, together with the statistician, two types of paired test: (1) Wilcoxon matched-pairs signed rank test; (2) Paired t-test.

(1) The Wilcoxon matched-pairs signed rank test is a nonparametric test that compares two paired groups. That means that the test does not assume any properties regarding the distribution of the variables in the analysis. GraphPad Prism8 first computes the differences between each set of pairs and ranks the absolute values of the differences from low to high. Then, it sums the ranks of the differences where column A was higher (positive ranks), sums the ranks where column B was higher (it calls these negative ranks), and reports the two sums. If the average sums of ranks are very different in the two groups, the p-value will be small.

(2) Paired t-test is a parametric test commonly used to test whether there is a statistical difference between two paired measurements. It assumes that the distribution of the data follows a Gaussian distribution. In particular, I tested for normality using the Shapiro-Wilk test, especially recommended for small samples. Furthermore, normality is the assumption that also the underlying residuals are normally distributed, or approximately so. A residual plot helped me to assess this assumption and QQ plot was the most useful way.

When I run a normality test on each paired dataset or on residual, Prism8 created a QQ plots, where the X-axis is the actual values, the Y-axis is the predicted ones (assuming sampling from a Gaussian distribution). Figure 83.a-b-c shows the graph related to *Notch1* mRNA expression level (at the top) and the corresponding normal QQ plot (at the bottom).

In the QQ graphs, the points followed a straight line that matched the line of identity (shown by Prism), suggesting that the data had a Gaussian (normal)

distribution. In this way, I could also apply a parametric statistical t-test on my dataset, in addition to the previous nonparametric Wilcoxon matched-pairs signed rank test.

The following Table 15 indicates the statistical analysis results, when the two samples (WT and #11B *Notch1*- Δ enh.in15) were compared at the three different time points (d4, d6, d8), separately.

Day of differentiation (#11B <i>Notch1</i> - Δ enh.in15 vs WT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	* p-value= 0.0313	ns p-value= 0.0566
d6	* p-value= 0.0313	* p-value= 0.0481
d8	* p-value= 0.0313	*** p-value= 0.0008

Table 15 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #11B *Notch1*- Δ enh.in15 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

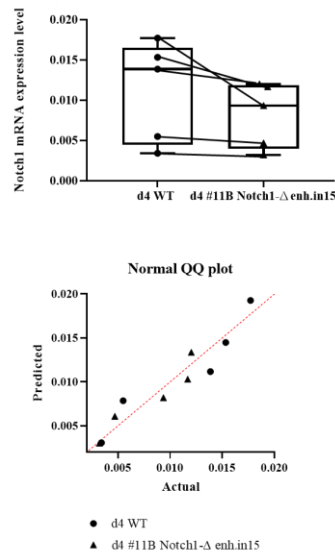


Figure 83.a Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #11B *Notch1*- Δ enh.in15 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

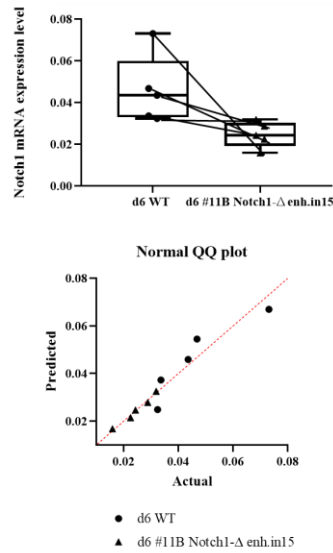


Figure 83.b Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #11B *Notch1*-Δ enh.in15 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

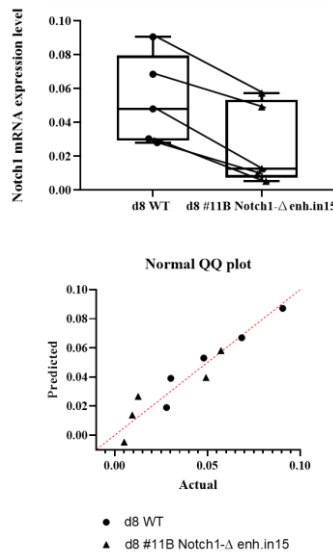


Figure 83.c Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #11B *Notch1*-Δ enh.in15 at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

The second mutant cell line analyzed during differentiation was clone #7G *Notch1*- Δ enh.intron15.

The enhancer deletion induced a statistically significant reduction of *Notch1* in #7G *Notch1*- Δ enh.in15 clone, at d4- d6-d8 during EC differentiation (Figure 84). Also in this case the experiments were done on five biological replicates (n=5).

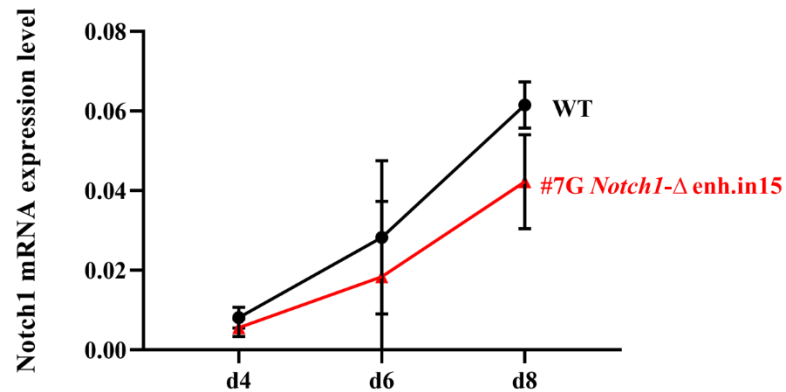


Figure 84 Quantitative real-time PCR analysis of *Notch1* mRNA expression level in clone #7G *Notch1*- Δ enh.in15 during EC differentiation. *Notch1* is reduced in clone #7G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta\text{Ct}}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).

As described above in detail, I performed statistical analysis, using GraphPad Prism8 software. The first test used was two-way repeated measures ANOVA test (ANOVA 2way-RM), to determine if there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a *Notch1* expression variable.

Data on Table 16 suggests that there was a significant change in *Notch1* mRNA expression level across the two factors “genotype” and “time”.

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.0015	**	yes
Genotype	0.0072	**	yes
Interaction: time x genotype	0.0265	*	yes

Table 16 Two-way repeated measures ANOVA results. Comparison between WT and #7G *Notch1*-Δ enh.in15 samples at d4, d6, d8, together, p-value < 0.05 is considered significant.

Then, I used other two test (Wilcoxon matched-pairs signed rank test and Paired t-test) to determine which groups are different from each other. I compared the two samples, WT and #7G *Notch1*-Δ enh.in15 cells, at different time points (d4, d6, d8), separately. I could apply the t-test as I demonstrated that the data had a Gaussian (normal) distribution.

Table 17 summarizes the Wilcoxon matched-pairs and Paired t-test results, after comparison of WT and #7G *Notch1*-Δ enh.in15 at d4, d6, d8, separately.

Day of differentiation (#7G <i>Notch1</i> -Δ enh.in15 vs WT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	* p-value= 0.0313	** p-value= 0.0014
d6	* p-value= 0.0313	* p-value= 0.0121
d8	* p-value= 0.0313	* p-value= 0.0115

Table 17 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #7G *Notch1*-Δ enh.in15 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

Notch1 mRNA expression level graph (at the top) and the corresponding normal QQ plot (at the bottom) are illustrated in Figure 85.a-b-c. It demonstrated that the distribution of the data follows a Gaussian distribution.

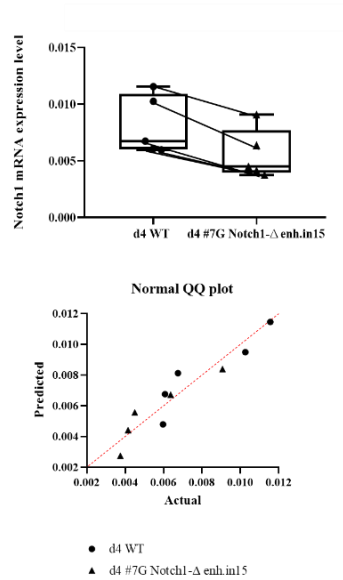


Figure 85.a Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #7G *Notch1*-Δ enh.in15 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

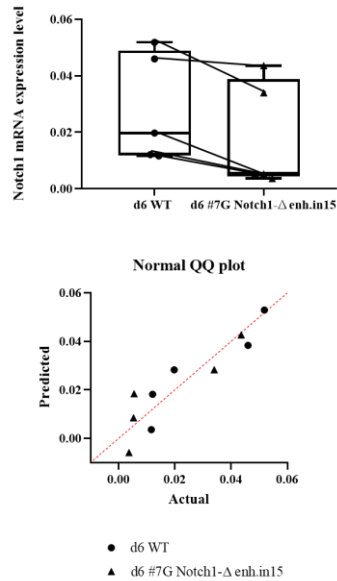


Figure 85.b Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #7G *Notch1*-Δ enh.in15 at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

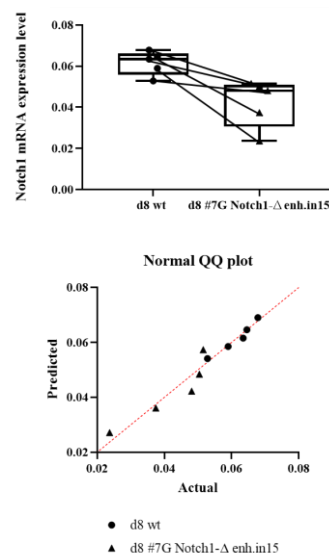


Figure 85.c Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level in WT and clone #7G *Notch1*-Δ enh.in15 at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Taken together these data offered the first evidence that CRISPR/Cas9-based *Notch1*-enh.in15 deletion resulted in statistically significant *Notch1* downregulation at day6 and day8 during EC differentiation process, in two *Notch1*-Δ enh.in15. independent mESC clones. I have summarized the results in Table 18.

Day of differentiation	Downregulation of gene expression in <i>Notch1</i> - Δ enh.intr15 (#11B, #7G)
d4	-
d6	*
d8	*

Table 18 Significant *Notch1* downregulation at day6 and day8 during EC differentiation in *Notch1*- Δ enh.intr15 cells (compared to WT). The asterisk (*) indicates the statistical significance of the data.

5.1.2 Assessment of *Notch1*-intron15 putative enhancer using CRISPR/dCas9 – LSD1 epigenetic repression strategy

A second enhancer validation strategy was based on epigenetic repression/decommissioning of the putative regulatory element, using inactive “dead” version of Cas9 (dCas9) fused with the histone demethylase LSD1, in order to endow dCas9 with gene repression abilities (previously described in detail in Chap.3). dCas9 is no longer able to cleave DNA but can still target and bind DNA with the same precision when guided by gRNAs. Instead, LSD1 regulates chromatin accessibility through its demethylating activity on histone H3 Lys4 residues, allowing chromatin inactivation and transcriptional repression. I induced mESC to express dCas9 fused with LSD1 repressor, generating stable mutant cell lines. I decided to proceed with validation of *Notch1*-intron15 putative enhancer, using clone #B1 p-dCas9LSD1 (see Chap. 4.4). To do this, I transfected three gRNAs against the targeted locus, which recognized the extremities (the same gRNAs used for enhancer deletion) and the middle part of the of the segment (as indicated in Figure 86) in the cells. I also transfected clone #B1p-dCas9LSD1 with a non-targeting control gRNA (gNT), which do not recognize any sequence in the genome. It was used as control sample. The experiments were done on four biological replicates (n=4).

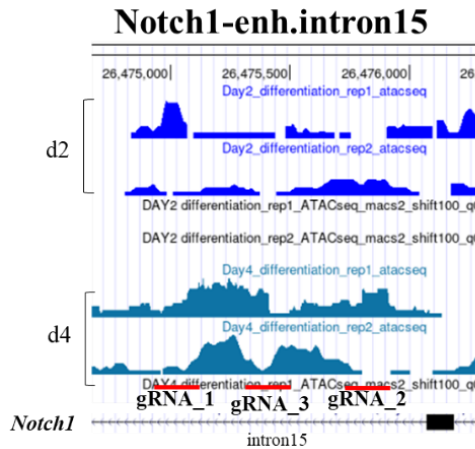


Figure 86 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce *Notch1*-enh.intron15 repression.

24h after the transfection, I visualized the fluorescently labeled gRNAs complex (crRNA:tracrRNA-ATTO™550), using the BD FACS ARIAI™ cell sorter to evaluate the % of fluorescent cells and the transfection efficiency. Cells containing the transfected gRNAs complex were isolated and differentiated into ECs. The experiments were done on four biological replicates (n=4).

A schematic experimental workflow is illustrated in Figure 87.

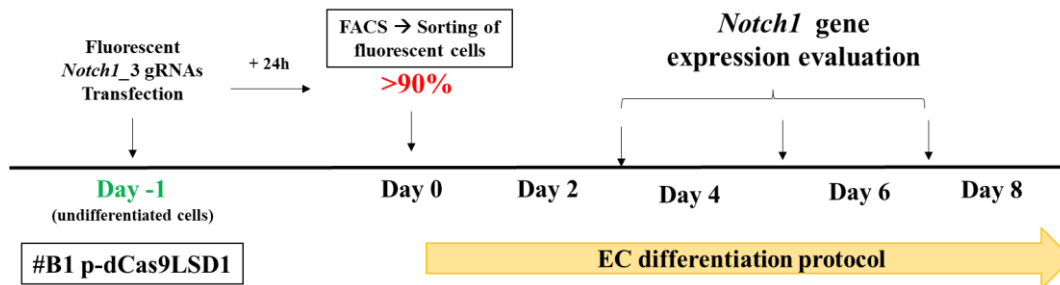


Figure 87 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent *Notch1_3* gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Notch1* expression.

My hypothesis was that the guide RNAs (gRNAs) target dCas9-LSD1 repressor, to the putative enhancer in intron15 of *Notch1*, changing its chromatin marks. LSD1 binds to several proteins, such as the CoREST transcriptional repressor complex and the Mi-2/nucleosome remodeling and deacetylase (NuRD) complex, to promote H3K4 demethylation and shape chromatin into a repressive configuration (Figure 88) (Magliulo D. *et al.*, (2018)).

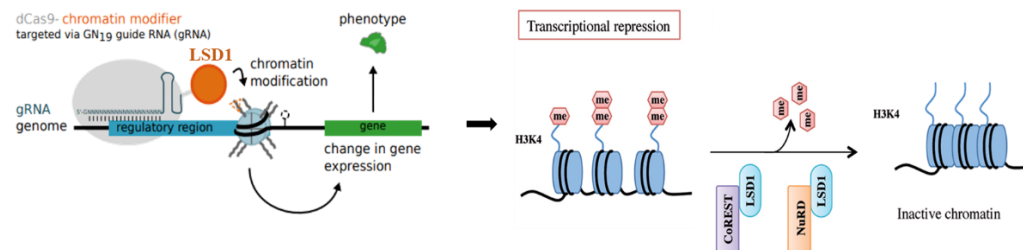


Figure 88 The dCas9-LSD1 transcriptional repression system, by chromatin inactivation. Specific gRNA target dCas9 fused with chromatin modifier LSD1 to a regulatory region. Lysine-specific demethylase 1A (LSD1) functions as transcriptional repressor: LSD1 binds to the CoREST or nucleosome remodeling and deacetylase repressive complex thus demethylating mono- and dimethyl-group on histone H3K4 and allowing genes transcriptional repression (Magliulo D. *et al.*, (2018)).

Targeting of dCas9-LSD1 to the *Notch1*-enh.intr15 resulted in repression of *Notch1* expression during endothelial differentiation, specifically at day 6 and day8 (Figure 89).

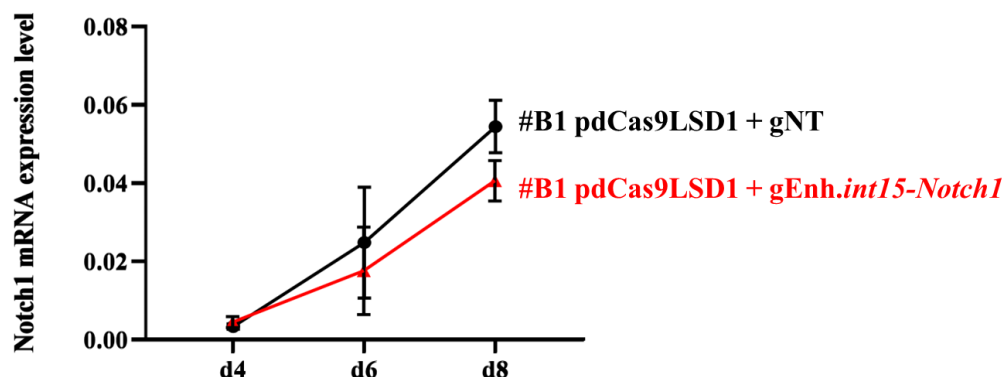


Figure 89 Quantitative real-time PCR analysis of *Notch1* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int15-*Notch1* during EC differentiation. *Notch1* is reduced in clone #B1p-dCas9LSD1+gEnh.int15-*Notch1*(red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of four (n=4) biological replicates \pm standard deviation (SD).

I performed statistical analysis, using GraphPad Prism8 software. The first test used was two-way repeated measures ANOVA test (ANOVA 2way-RM), to determine if there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a *Notch1* expression variable. The Table 19 suggested that there was a significant change in *Notch1* mRNA expression level across the two factors “genotype” and “time”.

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.0004	***	yes
Genotype	0.0013	**	yes
Interaction: time x genotype	0.0018	**	yes

Table 19 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-*Notch1* samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

I also used Wilcoxon matched-pairs signed rank test and Paired t-test to determine which groups are different from each other, comparing #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-*Notch1* samples at each different time points

(d4, d6, d8). I demonstrated that the data had a Gaussian (normal) distribution, so that I could perform a parametric statistical analysis (Figure 90.a, b, c).

The statistical results are showed in Table 20.

Day of differentiation (#B1p- dCas9LSD1+gEnh.int15- <i>Notch1</i> vs gNT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.0625	ns p-value= 0.0506
d6	ns p-value= 0.0625	* p-value= 0.0111
d8	ns p-value= 0.0625	** p-value= 0.0024

Table 20 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-Notch1 samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

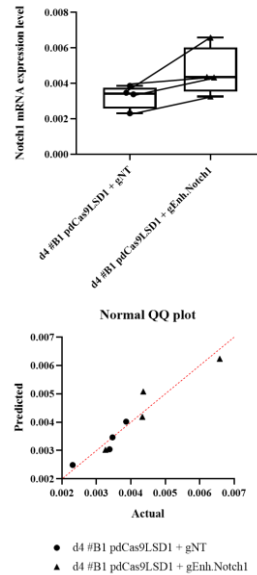


Figure 90.a Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-Notch1 samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

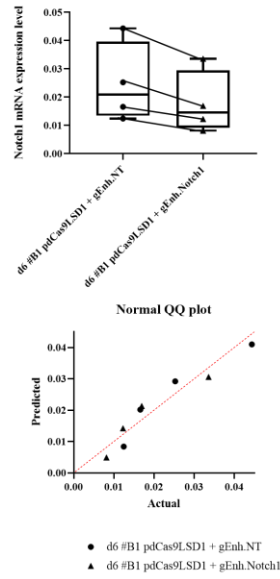


Figure 90.b Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-Notch1 samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

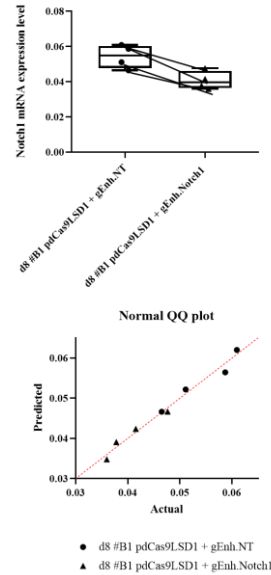


Figure 90.c Top: Quantitative real-time PCR analysis of *Notch1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int15-Notch1 samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

In conclusion, I detected similar transcriptional effect in *Notch1* expression during EC differentiation, using both enhancer validation strategies (enh.intr15 deletion via CRISPR/Cas9 and enh.intr15 epigenetic repression by dCas9LSD1). These results indicate that *Notch1* enh.intr15, which increased its chromatin accessibility at day4 of differentiation, may function as a regulatory elements. I observed a relative reduction of *Notch1* mRNA expression upon both deletion of the enh.intr15 and targeting with dCas9LSD1, at day6 and day8 of differentiation (Figure 91).

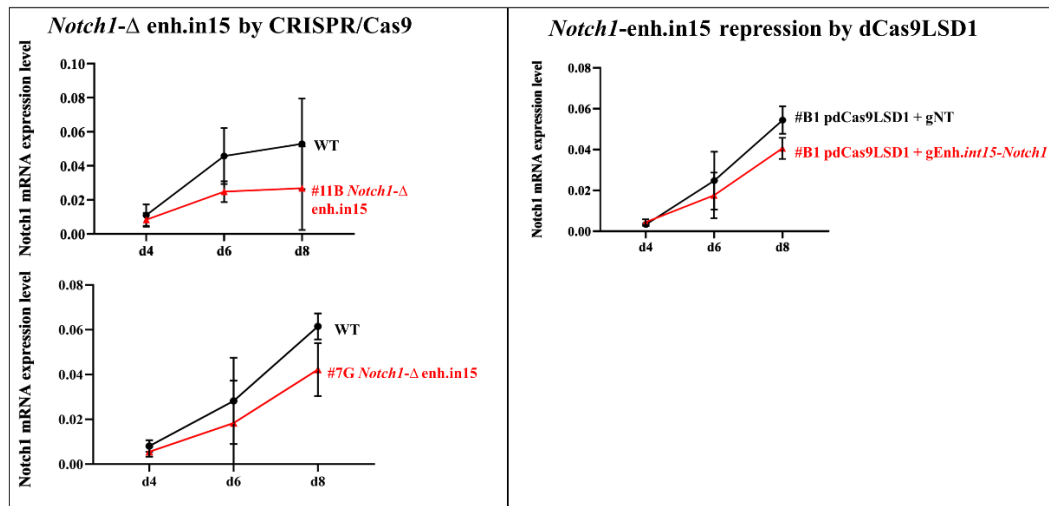


Figure 91 Comparison of *Notch1*-Δ enh.in15 by CRISPR/Cas9 and *Notch1*-enh.in15 repression by dCas9LSD1.

5.2 *Pecam1*-intron2 enhancer validation: evaluation of gene expression consequences during ECs differentiation, using CRISPR/Cas9 deletion strategy

Platelet endothelial cell adhesion molecule-1 (PECAM-1, also known as CD31) is a 130 kDa transmembrane glycoprotein expressed by endothelial cells and a variety of hematopoietic cells, as well as by platelets and some leucocytes. It is thought to be a sensitive and specific marker for vascular differentiation. In fact, *Pecam1* is widely used as a marker during vasculogenesis, angiogenesis, endothelial cell migration from embryonic stem (ES) cells (Li *et al.*, (2005)). Some studies have shown that mouse inner cell mass (ICM) of blastocyst expresses PECAM1 where embryonic stem (ES) cells derived from (Redick SD. *et*

al., (1999)). PECAM1 is expressed in early endothelial precursors first within the yolk sac and subsequently within the embryo itself. Furthermore, PECAM1 is also found on the entire vascular endothelium in the adult mouse (Baldwin HS. *et al.*, (1994)); Garlanda and Dejana (1997)).

According to Cao G. *et al.*, (2002), the endothelial cells cultured in the presence of anti-PECAM-1 antibodies fail to form normal cell-cell contacts, not only *in vitro* but also *in vivo*, in PECAM1-null mice. This inhibition is reversed by removal of the antibody, suggesting that PECAM1 is involved in angiogenesis and in endothelial cell-cell interactions, important in the formation of new vessels.

Based on this evidence, I decided to investigate the function of the putative enhancer region identified in intron2 of *Pecam1* gene at day4 of EC differentiation, assessing its requirement during mESC differentiation. (Figure 60.g, Chap.3).

To validate this regulatory element, I used the approach based on enhancer deletion, by CRISPR/Cas9 system.

As previously mentioned in Chap. 4.2, I induced the deletion of the putative enhancer region in *Pecam1*-intr2, by CRISPR/Cas9 technology, generating two *Pecam1*- Δ enh.intron2 mESC lines, named #5G and #7G clones, which were differentiated into ECs. I collected the cells at day4, day6 and day8 during the differentiation process and evaluated the expression level of *Pecam1* gene. A schematic experimental workflow is illustrated in Figure 92.

The hypothesis was that if the regulatory element in *Pecam1*-intron2 acts as an enhancer element, its deletion should induce an alteration of *Pecam1* expression during EC differentiation.

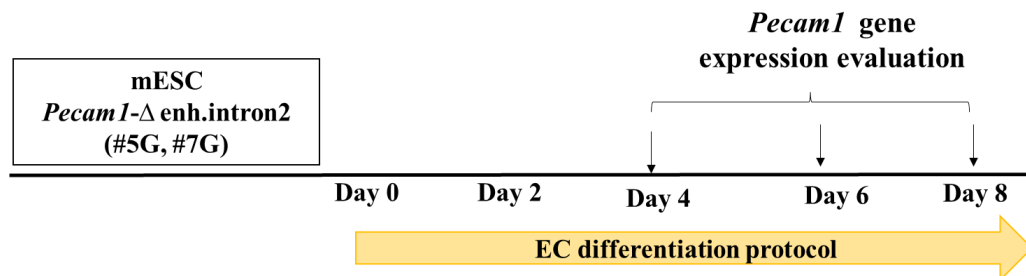


Figure 92 Overview of the differentiation scheme. *Pecam1*- Δ enh.intron2 clones (#5G and #7G), as well as mESC WT cells, are differentiated in ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Pecam1* expression.

At first, I analyzed the clone **#5G *Pecam1*- Δ enh.intron2**.

Pecam1 expression is significantly reduced in this mutant cell line, compared to WT mESC, at d6-d8 during EC differentiation (Figure 93). The experiments were done on five biological replicates (n. = 5).

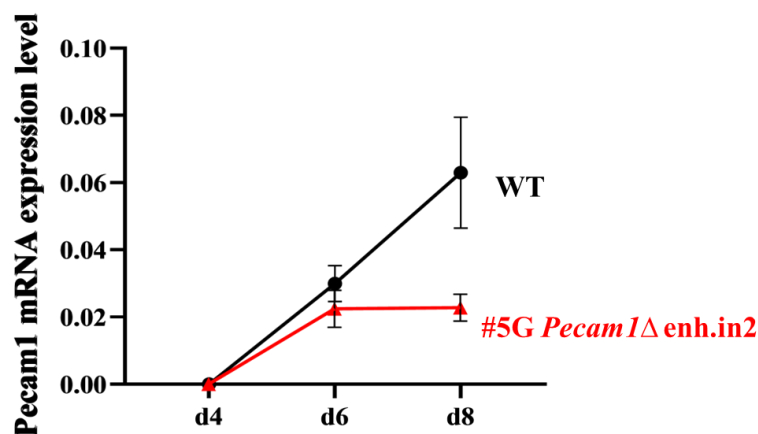


Figure 93 Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in clone #5G *Pecam1*-Δ *enh.in2* during EC differentiation. *Pecam1* is reduced in clone #5G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and *Gapdh* expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).

Then, I performed statistical analysis, using GraphPad Prism8 software, choosing two-way repeated measures ANOVA test (ANOVA 2way-RM), to assess that there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a *Pecam1* expression variable (Table 21).

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	<0.0001	****	yes
Genotype	0.0010	**	yes
Interaction: time x genotype	<0.0001	****	yes

Table 21 Two-way repeated measures ANOVA results, when compared WT and clone #5G *Pecam1*-Δ *enh.in2* dataset. p-value < 0.05 is considered significant.

To also determine which groups are different from each other, I have performed two subsequent tests (Wilcoxon matched-pairs signed rank test and Paired t-test), by pairwise comparisons of the two samples, WT and #5G *Pecam1*-Δ *enh.in2* cells, at different time points (d4, d6, d8), separately.

Table 22 summarizes the statistical comparison results. The Gaussian distribution of the data was confirmed by normality test and QQ plots (Figure 94.a,b,c).

Day of differentiation (#5G <i>Pecam1</i> - Δ enh.in2 vs WT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.3125	ns p-value= 0.2663
d6	* p-value= 0.0313	**** p-value= <0.0001
d8	* p-value= 0.0313	*** p-value= 0.001

Table 22 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #5G *Pecam1*- Δ enh.in2 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

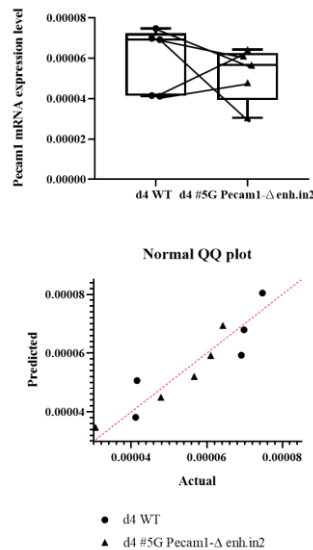


Figure 94.a Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #5G *Pecam1*- Δ enh.in2 at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

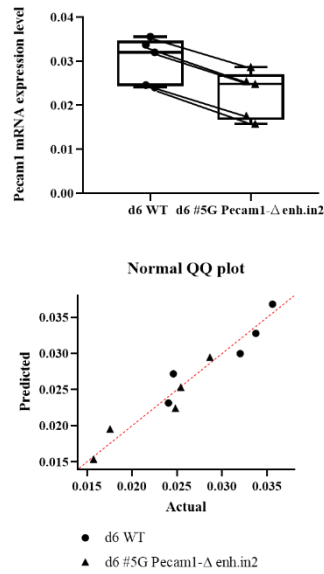


Figure 94.b Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #5G *Pecam1*-Δ *enh.in2* at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

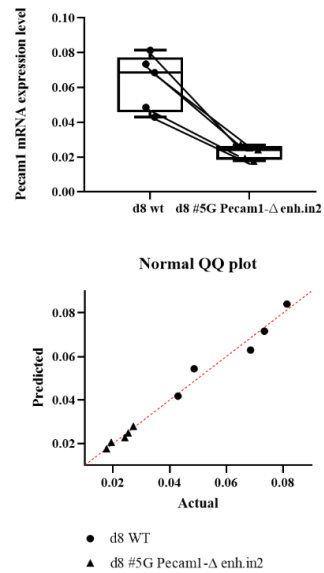


Figure 94.c Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #5G *Pecam1*-Δ *enh.in2* at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Subsequently, I induced the second mutant cell line, clone **#7G *Pecam1*- Δ enh.intron2**, to differentiate into EC and then I have evaluated the transcriptional effects.

The enhancer targeting induced a statistically significantly reduction of *Pecam1* in #7G *Pecam1*- Δ enh.intron2, at d4-d6-d8 during EC differentiation (Figure 95). Also, in this case the experiments were done on five biological replicates (n=5).

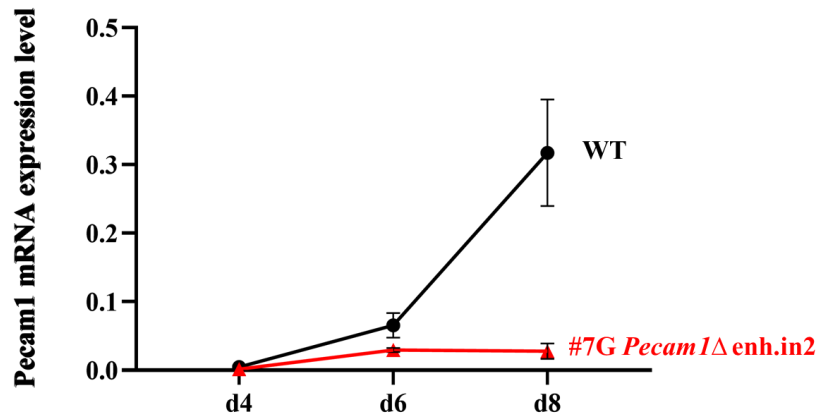


Figure 95 Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in clone #7G *Pecam1*- Δ enh.in2 during EC differentiation. *Pecam1* is reduced in clone #7G sample (red) compared to WT cells, used as control (black). X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of five (n=5) biological replicates \pm standard deviation (SD).

The data were then analyzed to evaluate the statistical significance of the differences between the two conditions (WT vs #7G *Pecam1*- Δ enh.in2 cells) during the three time points (d4-d6-d8).

ANOVA 2way-RM suggested that there was a significant correlation between the two factors (“time” and “genotype”) on a *Pecam1* expression variable (Table 23).

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	<0.0001	****	yes
Genotype	0.0008	***	yes
Interaction: time x genotype	<0.0001	****	yes

Table 23 Two-way repeated measures ANOVA results. Comparison between #7G *Pecam1*-Δ *enh.in2* cells and #WT samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

As previously mentioned in details, I applied other two statistical tests, by paired comparisons of the two samples, WT and #7G *Pecam1*-Δ *enh.in2*, at different time points (d4, d6, d8), separately.

The two types of paired test were: (1) Wilcoxon matched-pairs signed rank test; (2) Paired t-test. In the QQ graphs (Figure 96.a,b,c), the points followed a straight line that matched the line of identity (showed by Prism), demonstrating that the data had a Gaussian (normal) distribution. I could so apply a parametric statistical t-test on my dataset, in addition to the previous nonparametric Wilcoxon matched-pairs signed rank test.

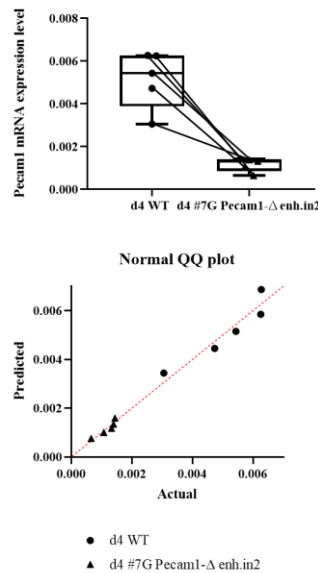


Figure 96.a Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #7G *Pecam1*-Δ *enh.in2* at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

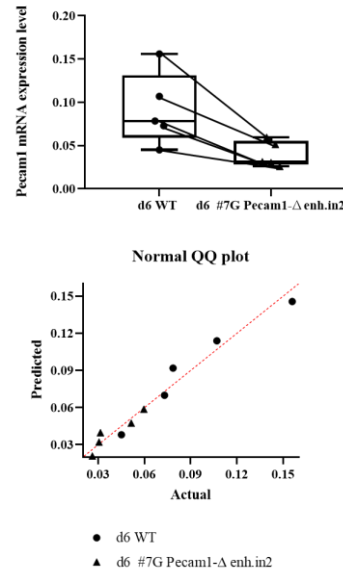


Figure 96.b Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #7G *Pecam1*-Δ *enh.in2* at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

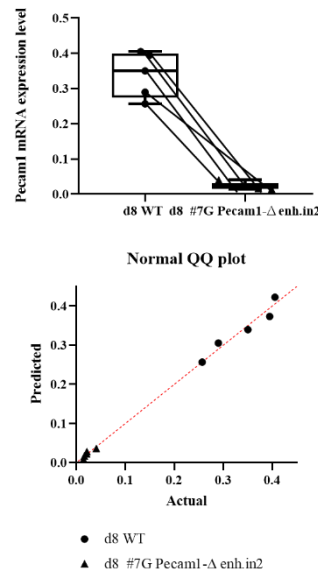


Figure 96.c Top: Quantitative real-time PCR analysis of *Pecam1* mRNA expression level in WT and clone #7G *Pecam1*-Δ *enh.in2* at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal

line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

The following Table 24 indicates the statistical analysis results, when the two samples (WT and #7G *Pecam1*- Δ enh.in2) were compared at the three different time points (d4, d6, d8), separately.

Day of differentiation (#7G <i>Pecam1</i>-Δ enh.in2 vs WT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	* p-value= 0.0313	** p-value= 0.0020
d6	* p-value= 0.0313	** p-value= 0.0073
d8	* p-value= 0.0313	*** p-value= 0.0003

Table 24 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #7G *Pecam1*- Δ enh.in2 vs WT samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

In summary, the obtained results indicated that *Pecam1* enh.intr2, which increased its chromatin accessibility at day4 of differentiation, could act as a regulatory elements. The precise enhancer deletion impaired *Pecam1* gene expression, resulting in statistically significant *Pecam1* downregulation at day6 and day8 during EC differentiation process, in two independent *Pecam1*- Δ enh.intr2. mESC clones. (Table 25).

Day of differentiation	Downregulation of gene expression in <i>Pecam1</i>- Δ enh.intr15 (#5G, #7G)
d4	-
d6	*
d8	*

Table 25 Significant *Pecam1* downregulation at day6 and day8 during EC differentiation in *Pecam1*- Δ enh.intr2 cells (compared to WT). The asterisk (*) indicates the statistical significance of the data.

5.3 *Kdr-intron10* enhancer validation by epigenetic repression, using CRISPR/dCas9-LSD1: evaluation of transcriptional consequences during ECs differentiation.

An in-silico search of the mouse *Kdr* locus for enriched chromatin accessibility at day4 during EC differentiation, allowed me to identify a region in the 10th intron of *Kdr* (*Kdr*-enh.int10) as a putative regulatory element (Figure 60.a).

For this putative enhancer, I have used the second validation strategy (described in Chap. 3), based on epigenetic reprogramming by dCas9 fused with LSD1, by transfection of ES p-dCas9LSD1 (clone#B1) with three gRNAs against the targeted locus, which recognized the extremities and the middle part of the of the segment (as indicated in Figure 97) and with a gRNA specific to an unrelated control genomic region, as control sample (gNT).

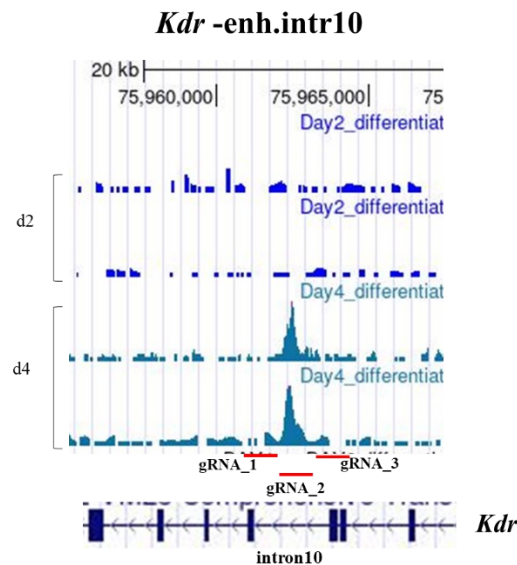


Figure 97 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce *Kdr*-enh.intron10 repression.

24h after the transfection, I visualized the fluorescently labelled gRNAs complex (crRNA:tracrRNA-ATTOTM550), using the BD FACS ARIAIITM cell sorter to evaluate the % of fluorescent cells and the transfection efficiency. Cells containing the transfected gRNAs complex were isolated and differentiated into ECs. The experiments were done on four biological replicates (n=4).

A schematic experimental workflow is illustrated in Figure 98.

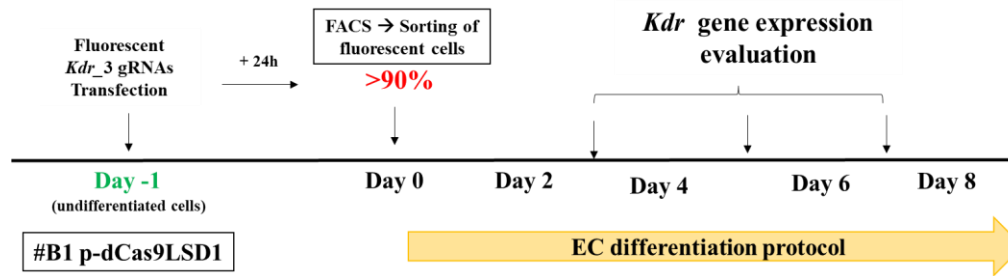


Figure 98 Overview of the experimental plan. #B1 p-dCas9LSD1 cell clone is transfected with fluorescent *Kdr_3* gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Kdr* expression.

The hypothesis was that the guide RNAs (gRNAs) target dCas9-LSD1 fusion protein to the putative enhancer in intron10 of *Kdr*, modifying its chromatin state. Overall, results (showed in Figure 99) indicated that targeting of *Kdr*-enh.int10 by dCas9-LSD1 resulted in repression of *Kdr* expression during endothelial differentiation, specifically at day8.

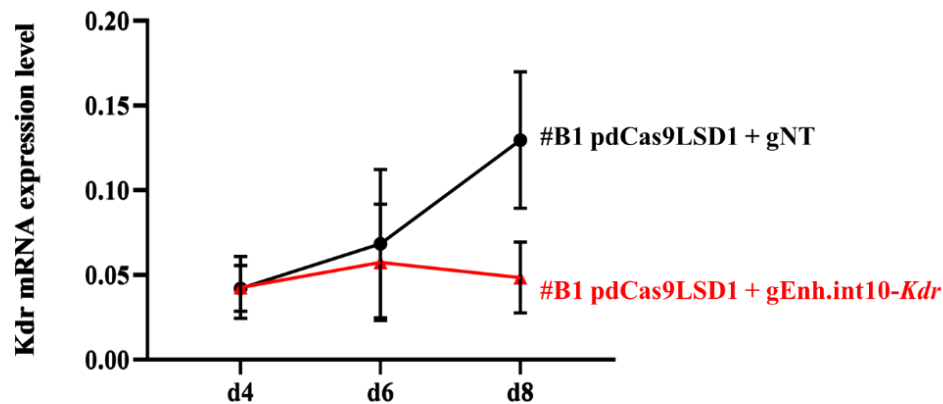


Figure 99 Quantitative real-time PCR analysis for *Kdr* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int10-*Kdr* during EC differentiation. *Kdr* is reduced in clone #B1p-dCas9LSD1+gEnh.int10-*Kdr* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of four (n=4) biological replicates \pm standard deviation (SD).

According to ANOVA 2way-RM statistical analysis performed, there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a *Kdr* expression variable, indicated in Table 26.

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.1314	ns	no
Genotype	0.0404	*	yes
Interaction: time x genotype	<0.0001	****	yes

Table 26 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-Kdr samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

I also used Wilcoxon matched-pairs signed rank test and Paired t-test to determine which groups are different from each other, comparing #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-Kdr samples at each different time points (d4, d6, d8). I demonstrated the Gaussian (normal) distribution of the data, so that I performed a parametric statistical analysis (Figure 100.a, b, c). The statistical results are showed in Table 27.

Day of differentiation (#B1p-dCas9LSD1+gEnh.int10-Kdr vs gNT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.5000	ns p-value= 0.7490
d6	ns p-value= 0.1250	ns p-value= 0.1225
d8	ns p-value= 0.0625	** p-value= 0.0043

Table 27 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-Kdr samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

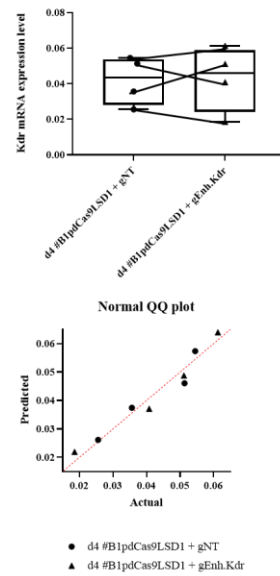


Figure 100.a Top: Quantitative real-time PCR analysis of *Kdr* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-Kdr samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

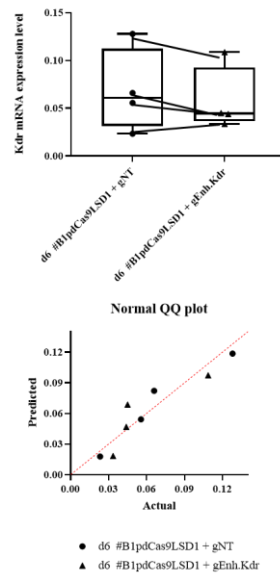


Figure 100.b Top: Quantitative real-time PCR analysis of *Kdr* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-Kdr samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper

quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

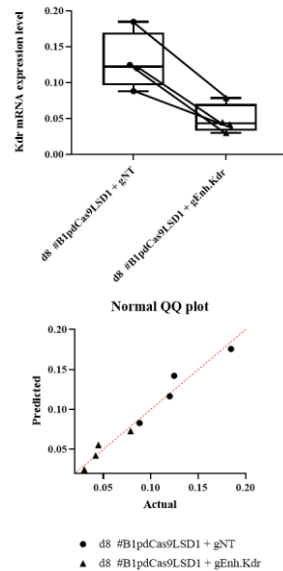


Figure 100.c Top: Quantitative real-time PCR analysis of *Kdr* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-*Kdr* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

In conclusion, results showed that *Kdr*-enh.intr10, which increased its chromatin accessibility at day4 of differentiation, may function as a regulatory elements. The targeted locus was affected by dCas9-LSD1 epigenetic repression, giving rise to relative reduction of *Kdr* mRNA expression at day8 of differentiation.

5.4 *VE-Cadherin-intron1 enhancer validation by epigenetic repression, using CRISPR/dCas9-LSD1: evaluation of transcriptional consequences during ECs differentiation.*

VE-Cadherin (*Cdh5*) is a component of endothelial cell-to-cell adherens junctions, and it has a key role in the maintenance of vascular integrity. During embryo development, *VE-Cadherin* is one of the first cell-specific markers to be expressed by endothelial cells in embryo vasculature. *VE-Cadherin* transcripts have been detected at very early stages of vascular development (embryonic day

7.5 [E7.5]) in mesodermal cells of the yolk-sac mesenchyme. Consequently, *VE-Cadherin* is essential during embryonic endothelial differentiation. At later embryonic stages, *VE-Cadherin* was expressed in vascular structures of organs, which included, for example, the ventricles of the heart, the inner cell lining of the atrium, the dorsal aorta, the intersomitic vessels, and capillaries entering the brain. This suggested that *VE-Cadherin* is needed for the correct organization of a stable vascular system and in the adult, it controls vascular permeability and inhibits unrestrained vascular growth (Giannotta M. *et al.*, (2013)).

Looking for genomic regions with increased chromatin accessibility during cardiopharyngeal mesoderm (CPM) differentiation into endothelial cells (ECs), I identified a region of the 1st intron of the *VE-Cadherin* (*Cdh5*) gene (*VE-Cadherin*-enh.int1) as a putative enhancer and tested it during *in vitro* EC differentiation (Figure 60.b).

In particular, I used the validation strategy based on dCas9LSD1 enhancer repressing system. I transfected into stable clone #B1 p-dCas9LSD1 (see Chap. 4.4) three gRNAs against the targeted locus, which recognized the extremities and the middle part of the of the segment (as indicated in Figure 101). As control sample, I used a non-targeting control gRNA (gNT). The experiments were done on four biological replicates (n=4).

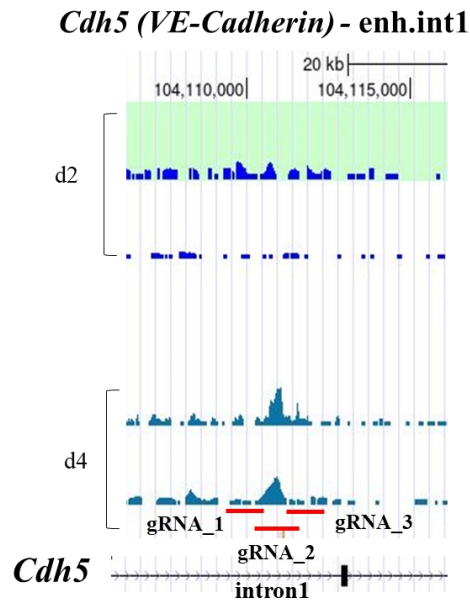


Figure 101 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce *VE-Cadherin*-enh.int1 repression.

24h after the transfection, I visualized the fluorescently labeled gRNAs complex, using the BD FACS ARIAM™ cell sorter to evaluate the % of fluorescent cells and the transfection efficiency. Cells containing the transfected gRNAs complex were isolated and differentiated into ECs. The experiments were done on four biological replicates (n=4). An experimental workflow is depicted in Figure 102.

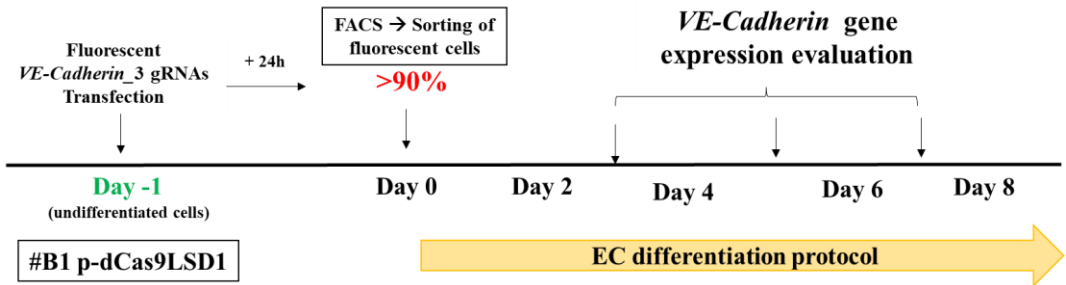


Figure 102 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent VE-Cadherin_3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the VE-Cadherin expression.

The idea was that the gRNAs could specifically direct dCas9LSD1 fusion protein to the enhancer region in intron1 of *VE-Cadherin*, modulating the chromatin landscape and achieving targeted gene repression in differentiating cells. Indeed, when dCas9-LSD1 activity was targeted to the *VE-Cadherin*-enh.int1, I observed a significant downregulation of *VE-Cadherin* expression at day8 of EC differentiation (Figure 103).

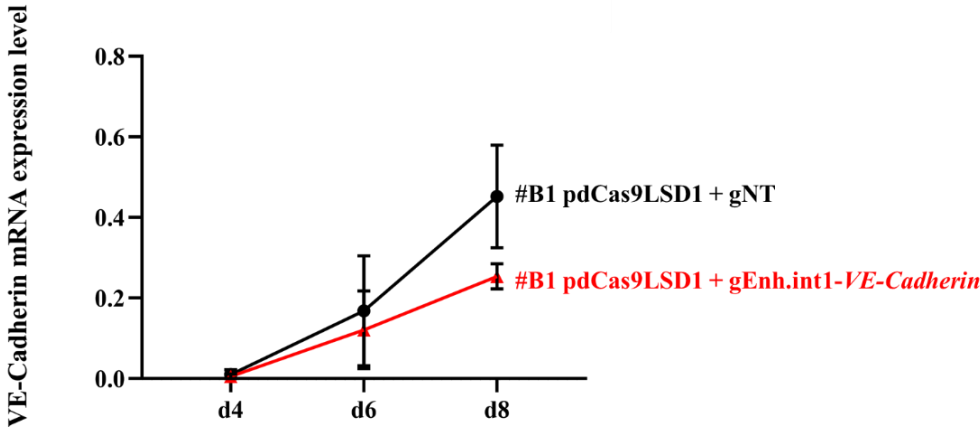


Figure 103 Quantitative real-time PCR analysis of *VE-Cadherin* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int1-*VE-Cadherin* during EC differentiation. *VE-Cadherin* is reduced in clone #B1p-dCas9LSD1+gEnh.int1-*VE-Cadherin* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates

the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=4) biological replicates \pm standard deviation (SD).

ANOVA 2way-RM test (by GraphPad Prism8) suggested that there was a significant change in *VE-Cadherin* mRNA expression level across the two factors “genotype” and “time” (Table 28).

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.0010	**	yes
Genotype	0.0575	ns	no
Interaction: time x genotype	0.0143	*	yes

Table 28 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-*VE-Cadherin* samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

I also used Wilcoxon matched-pairs signed rank test and Paired t-test to determine which groups are different from each other, comparing #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-*VE-Cadherin* samples at each different time points (d4, d6, d8). The statistical results are showed in Table 29. The Gaussian distribution of the data was confirmed by normality test and QQ plots (Figure 104.a,b,c).

Day of differentiation (#B1p-dCas9LSD1+gEnh.int1- <i>VE-Cadherin</i> vs gNT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.3125	ns p-value= 0.7454
d6	ns p-value= 0.0625	ns p-value= 0.0829
d8	ns p-value= 0.0625	** p-value= 0.0246

Table 29 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-*VE-Cadherin* samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

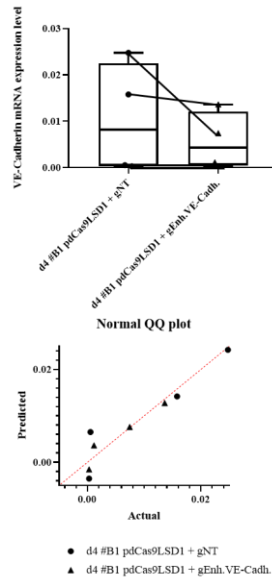


Figure 104.a Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

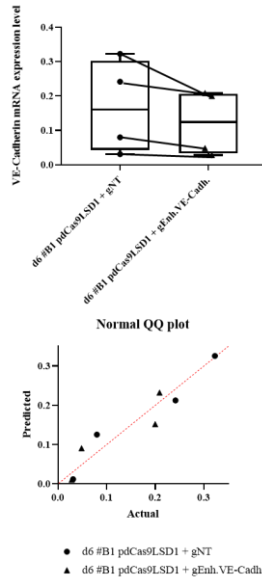


Figure 104.b Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

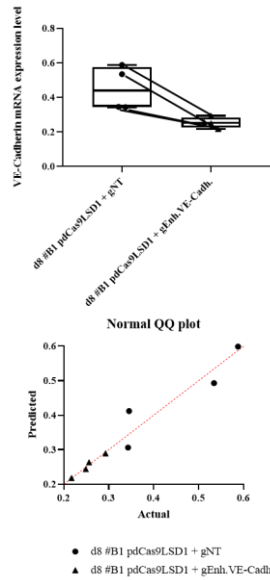


Figure 104.c Top: Quantitative real-time PCR analysis of VE-Cadherin mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int1-VE-Cadherin samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Taken together these data offer evidence that dCas9LSD1/gRNAs system is able to affect the activity of the *VE-Cadherin*-Enh.int1 region, leading to gene expression downregulation. This support the assumption that *VE-Cadherin*-Enh.int1 region can be a regulatory element and may have an involvement during EC differentiation.

5.5 *Eng*-intron2 enhancer validation by epigenetic repression, using CRISPR/dCas9-LSD1: evaluation of transcriptional consequences during ECs differentiation.

Endoglin (*Eng*) is predominantly expressed by activated endothelial cells and plays a crucial role in developmental angiogenesis (Arthur HM. *et al.*, (2000); Wikström P. *et al.*, (2002)). In mice, a complete loss of Endoglin is embryonically lethal around E10.5, primarily due to impaired development of the vascular plexus into a mature vascular network, causing hampered low and osmotic imbalance, disturbing normal cardiac development (Goumans MJ. *et al.*, (2018)). Published work suggest that *Eng* is involved in angiogenesis and vascular development, and in maintenance of vessel wall integrity (Kopczyńska E. *et al.*,

(2012)). Lee NY, *et al.*, (2012) compared control wild-type ($Endo^{+/+}$) and endoglin-null ($Endo^{-/-}$) murine embryonic endothelial cells (MEECs), showing that $Endo^{+/+}$ MEECs formed a significantly greater number of capillaries than did $Endo^{-/-}$ MEECs when plated on Matrigel.

Endoglin is overexpressed on proliferating endothelial cells in tissues undergoing active angiogenesis, such as regenerating and inflamed tissues or tumours; the inhibition of *Eng* expression not only enhances the ability of TGF- β to suppress growth and migration of endothelial cells but also enhances apoptosis induced by hypoxia and TGF- β (Li C. *et al.*, (1999)).

The importance of endoglin for normal vascular architecture is further indicated by the association of mutations in the endoglin gene with the inherited disorder Hereditary Haemorrhagic Telangiectasia Type 1 (HHT1), a disease characterized by bleeding from vascular malformations (McAllister, K. *et al.*, (1994)).

As previously mentioned in Chap. 3 (Figure 60.d), I identified a DAR opening region at day4 of differentiation, located in intron2 of *Eng*.

I decided to investigate the function of this putative enhancer region, assessing its requirement during mESC differentiation, and using the approach based on enhancer decommissioning, by dCas9LSD1/gRNAs system.

I transfected into ES dCas9LSD1 three gRNAs against the targeted locus, which recognized the extremities and the middle part of the of the segment (as indicated in Figure 105). I also used a gRNA specific to an unrelated control genomic region, as control sample (gNT).

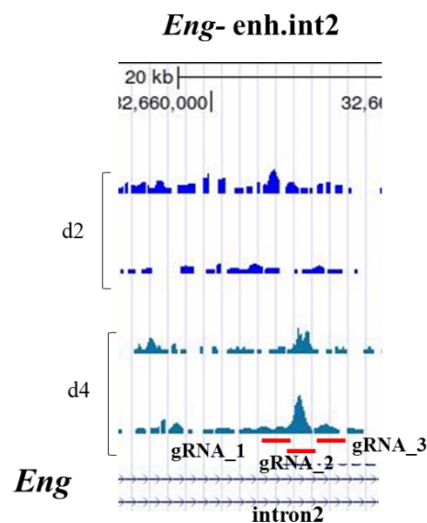


Figure 105 Localization of the three gRNAs (in red) used to transfect clone #B1 p-dCas9LSD1 and to induce *Eng*-enh.int2 repression.

24h after the transfection, I visualized the fluorescently labeled gRNAs complex (crRNA:tracrRNA-ATTO™550), using the BD FACS ARIAI™ cell sorter to evaluate the % of fluorescent cells and the transfection efficiency. Cells containing the transfected gRNAs complex were isolated and differentiated into ECs. I collected the cells at day4, day6 and day8 during the differentiation process and evaluated the expression level of *Eng* gene. The experiments were done on four biological replicates (n=4). A schematic workflow is illustrated in Figure 106.

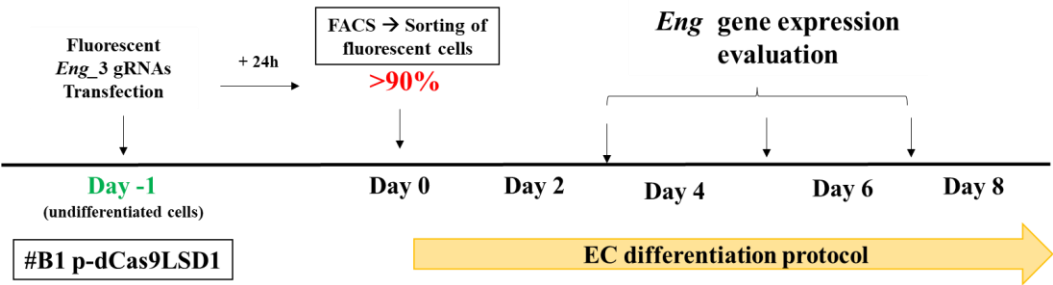


Figure 106 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent *Eng_3* gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Eng* expression.

The hypothesis was that if the regulatory element in *Eng*-intron2 acts as an enhancer element, its epigenetic repression by dCas9LSD1 protein should induce an alteration of *Eng* expression during EC differentiation. Overall, results (shown in Figure 107) indicated that epigenetic silencing of *Eng*-enh.int2 by dCas9-LSD1 resulted in repression of *Eng* expression at day6 and day8 during endothelial differentiation.

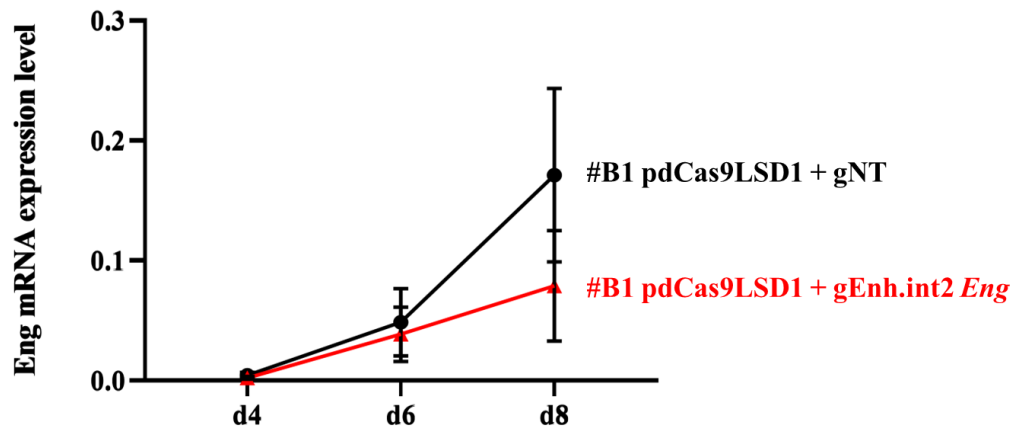


Figure 107 Quantitative real-time PCR analysis of *Eng* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int2-*Eng* during EC differentiation. *Eng* is reduced in clone #B1p-dCas9LSD1+gRNA-Enh.int2-*Eng* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta Ct}$ method, and Gapdh expression as the normalizer. Values are the average of five (n=4) biological replicates \pm standard deviation (SD).

ANOVA 2way-RM statistical analysis indicated that there was a statistically significant interaction effect between the two specific factors, “time” and “genotype” on an *Eng* expression variable (Table 30).

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	0.0065	**	yes
Genotype	0.0645	ns	no
Interaction: time x genotype	<0.0366	*	yes

Table 30 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int2-*Eng* samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

Besides, I demonstrated that the data followed a Gaussian distribution, so that I could perform a parametric statistical analysis (Figure 108.a, b, c). Indeed, to know which groups are different from each other, I used Wilcoxon matched-pairs signed rank test and Paired t-test, comparing #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int2-*Eng* samples at each different time points (d4, d6, d8). The statistical results are showed in Table 31.

Day of differentiation (#B1p-dCas9LSD1+ gEnh.int2- <i>Eng</i> vs gNT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.1875	ns p-value= 0.1476
d6	ns p-value= 0.0625	* p-value= 0.0458
d8	ns p-value= 0.0625	* p-value= 0.0401

Table 31 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

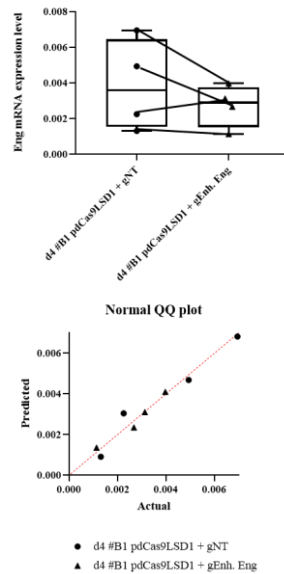


Figure 108.a Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

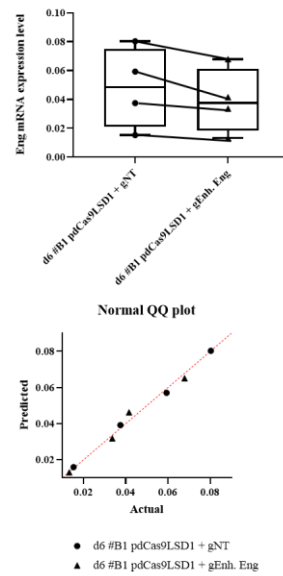


Figure 108.b Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

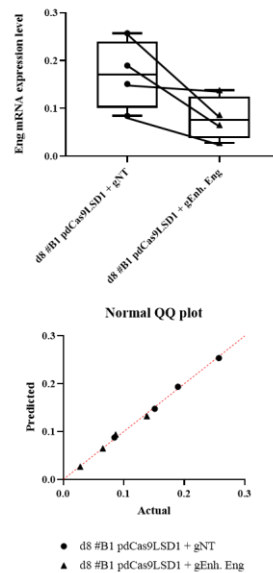


Figure 108.c Top: Quantitative real-time PCR analysis of *Eng* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int2-*Eng* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper

24h after the transfection, I visualized the fluorescently labeled gRNAs complex, using the BD FACS ARIAM™ cell sorter to evaluate the % of fluorescent cells and the transfection efficiency. Cells containing the transfected gRNAs complex were isolated and differentiated into ECs. An experimental workflow is depicted in Figure 110.

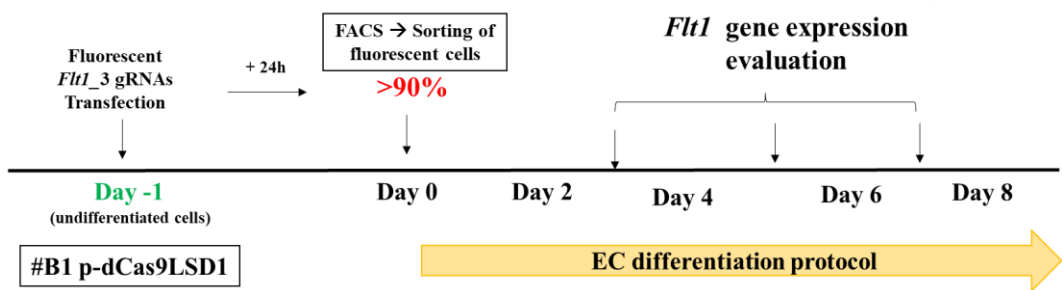


Figure 110 Overview of the experimental plan. #B1 p-dCas9LSD1 clone are transfected with fluorescent *Flt1*_3 gRNAs. Fluorescent sorted cells are differentiated into ECs from day0 to day8. Samples are collected at day4, day6 and day8 to analyze the *Flt1* expression.

The hypothesis was that gRNAs would target dCas9-LSD1 fusion protein to the putative enhancer in intron10 of *Flt1*, modifying its chromatin state. Overall, results (shown in Figure 111) indicated that targeting of *Flt1*-enh.int10 by dCas9-LSD1 resulted in repression of *Flt1* expression during EC differentiation, at day8.

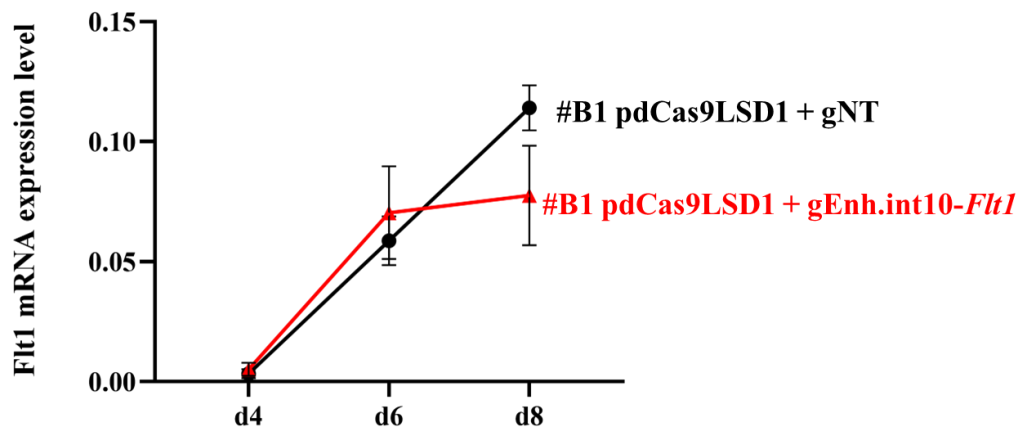


Figure 111 Quantitative real-time PCR analysis for *Flt1* mRNA expression level in clone #B1p-dCas9LSD1 transfected with gRNA-Enh.int10-*Flt1* during EC differentiation. *Flt1* is reduced in clone #B1p-

dCas9LSD1+gEnh.int10- *Flt1* (red) compared to control #B1p-dCas9LSD1+gNT (black), where gNT is a non-targeting gRNA. X-axis denotes the three time points (d4-d6-d8); y-axis indicates the gene expression level, evaluated using the $2^{-\Delta C_t}$ method, and *Gapdh* expression as the normalizer. Values are the average of four (n=4) biological replicates \pm standard deviation (SD).

According to ANOVA 2way-RM statistical analysis performed, there is a statistically significant interaction effect between the two specific factors, “time” and “genotype” on a *Flt1* expression variable, indicated in Table 32.

ANOVA 2way-RM

Source of variation	p-Value	p-Value summary	Significant?
Time	<0.0001	****	yes
Genotype	0.2100	ns	no
Interaction: time x genotype	0.0403	*	yes

Table 32 Two-way repeated measures ANOVA results. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-*Flt1* samples at d4, d6, d8, together. p-value < 0.05 is considered significant.

I demonstrated the Gaussian (normal) distribution of the data, so that I could perform a parametric statistical analysis (Figure 112.a, b, c).

I also used Wilcoxon matched-pairs signed rank test and Paired t-test to determine which groups are different from each other, comparing #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-*Flt1* samples at each different time points (d4, d6, d8).

The statistical results are shown in Table 33.

Day of differentiation (#B1p-dCas9LSD1+gEnh.int10- <i>Flt1</i> vs gNT)	Wilcoxon matched-pairs signed rank test (nonparametric)	Paired t-test (parametric)
d4	ns p-value= 0.0625	* p-value= 0.0386
d6	ns p-value= 0.1875	ns p-value= 0.1413
d8	ns p-value= 0.0625	* p-value= 0.0441

Table 33 Statistical analysis of data, using Wilcoxon matched-pairs signed rank test and paired t-test. Comparison between #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+gEnh.int10-*Flt1* samples at d4, d6, d8, separately. p-value < 0.05 was considered significant and indicated with an asterisk (*).

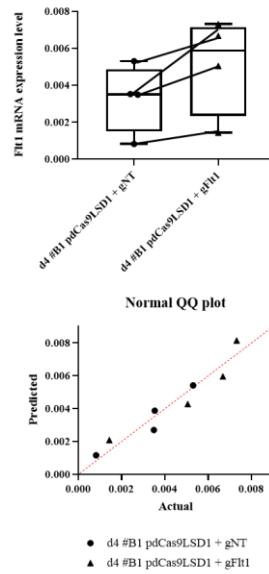


Figure 112.a Top: Quantitative real-time PCR analysis of *Ftl* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Ftl* samples at day4 of differentiation. Bottom: Normal QQ plot related to d4 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

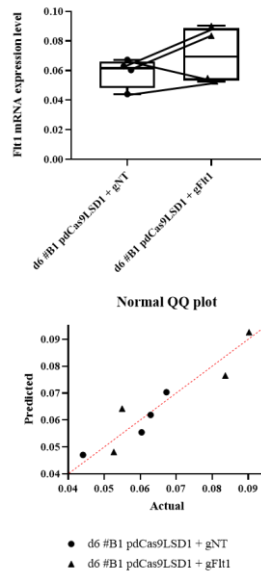


Figure 112.b Top: Quantitative real-time PCR analysis of *Ftl* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Ftl* samples at day6 of differentiation. Bottom: Normal QQ plot related to d6 dataset. The diagonal line (in dotted red), which passes through the lower and

upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

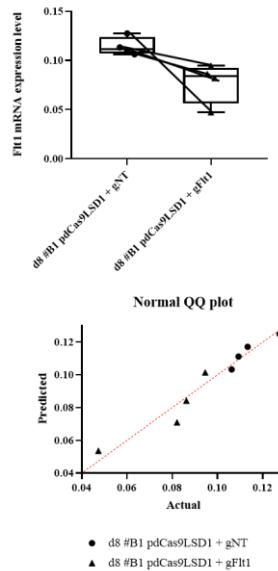


Figure 112.c Top: Quantitative real-time PCR analysis of *Fhl1* mRNA expression level #B1p-dCas9LSD1+gNT and #B1p-dCas9LSD1+ gEnh.int10-*Fhl1* samples at day8 of differentiation. Bottom: Normal QQ plot related to d8 dataset. The diagonal line (in dotted red), which passes through the lower and upper quartiles of the theoretical distribution, is helpful to assess that the relationship between the theoretical and sample percentiles is linear.

Overall, I followed two experimental systems for candidate enhancers validation, to examine their activity and requirement during cardiopharyngeal mesoderm (CPM) differentiation into endothelial cells (ECs). Thanks to the first strategy, based on CRISPR-Cas9 enhancer deletion, I investigated 2 putative enhancers, associated with endothelial-specific genes: *Pecam1* and *Notch1*, both of which are critical for vascular development. Two *Notch1*-Δ enh.int15 independent mESC clones (#7G, #11B) showed a significant reduction of *Notch1* expression during the later stages of EC differentiation (d6 and d8). Similarly, *Pecam1* expression was also downregulated in two independent *Pecam1*-Δ enh.int2 mESC clones (#5G, #7G) at the same time points. These results indicated that the regions deleted are required for appropriate expression of the respective genes during EC differentiation process.

Subsequently, using epigenetic enhancer reprogramming by dCas9LSD1 fusion protein, I analyzed five putative enhancer regions: *Notch1*-enh.intr15; *Kdr*-enh.intr10; *VE-Cadherin*-enh.intr1; *Eng*-enh.intr2; *Flt1*-enh.intr10. The targeted region resulted affected by dCas9-LSD1 repression, giving rise to relative reduction of gene-related expression, specifically at day8 of differentiation (Table 34)

Day of EC differentiation	#B1 <i>pdCas9-LSD1</i> + <i>gEnh.Notch1 intr15</i>	#B1 <i>pdCas9-LSD1</i> + <i>gEnh.Kdr intr10</i>	#B1 <i>pdCas9-LSD1</i> + <i>gEnh.VE-Cadh. intr1</i>	#B1 <i>pdCas9-LSD1</i> + <i>gEnh.Eng intr2</i>	#B1 <i>pdCas9-LSD1</i> + <i>gEnh.Flt1 intr10</i>
d4	ns	ns	ns	ns	ns
d6	*	ns	ns	*	ns
d8	*	*	*	*	*

Table 34 Epigenetic enhancer decommissioning / repression of 5 regions by dCas9LSD1 induced significant enhancer-gene-related downregulation at later stages of EC differentiation (mainly at day8).

In conclusion, these data offered evidence that 6 tested out of 10 total candidate enhancers seems to be regulatory elements and could be involved during later stages of EC differentiation. The other 4 putative enhancer regions have not yet been tested.

5.7 Computational prediction of transcription factor motifs in EC enhancers.

To identify the regulatory elements that are specifically enriched in my dataset, I and our bioinformatician colleague performed a preliminary, computational Transcription Factors (TFs) motif analysis in DAR regions related to endothelial cell fate specification.

We employed the integration of ATAC-seq and RNA-seq datasets to identify markers distinguishing day4 and day2 of CMP into EC differentiation. We investigated this association by studying genes that were known to be involved in angiogenesis and/or expressed in vascular-endothelial-lineage (Nomaru H. *et al.*, (2021)).

Indeed, we selected genes expressed in EC-cluster (#7) from single cell experiment performed by our collaborators (Nomaru *et al.*, (2021)) on *Tbx1* expressing cells in heterozygous *Tbx1*^{Cre/+} mouse embryos at stage E9.5.

Tbx1 is expressed in multipotent heart progenitors that, in clonal assays, can give rise to three heart lineages expressing endothelial, smooth muscle and cardiomyocyte markers (Chen L. *et al.*, (2009)). Since *Tbx1* is involved in the cardiopharyngeal mesoderm differentiation, we thought that EC-cluster dataset (named as sc-EC gene list) from *Tbx1*^{Cre/+} mouse embryos at stage E9.5 (when *Tbx1* is highly expressed) might be more similar to my dataset genes from day2 and day4 mESC differentiation.



We made some comparisons between sc-EC gene list (from Nomaru work) and my d4 vs d2 ATAC-seq/RNA-seq dataset: (1) intersection of DARs regions & DE genes with sc-EC gene list; (2) ATAC-seq peaks & DE genes with sc-EC gene list; (3) DARs regions & expressed genes at day4 with sc-EC gene list. Then, we looked for the most enriched motifs in each of these comparisons.

To identify enriched transcription factors' motifs, we used Hypergeometric Optimization of Motif EnRichment (HOMER) software, choosing as background the peakome, which is composed of all ATAC peaks between d2 and d4, annotated to expressed genes. Peaks annotated to promoters were removed to avoid canonical TSS- enriched TFs and looked for TF motifs enriched in gene body/distal regions.

Overall, we identified *Gata1*, *Gata2* and *JunB* TF motifs to be selectively enriched in day4 mESCs compared to day2 (Figure 113).


Intersection of DARs regions & DE genes with sc-EC gene list

Total Target Sequences = 42, Total Background Sequences = 12139

Rank	Motif	Name	P-value	log P-value
1		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-2	-4.993e+00
2		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-2	-4.901e+00

Intersection of ATAC-seq peaks & DE genes with sc-EC gene list

Total Target Sequences = 61, Total Background Sequences = 12157

Rank	Motif	Name	P-value	log P-value
1		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-2	-4.801e+00

Intersection of DARs regions & d4-expressed genes with sc-EC gene list

Total Target Sequences = 79, Total Background Sequences = 12412



Rank	Motif	Name	P-value	log P-value
1		Gata1(Zf)/K562-GATA1-ChIP-Seq(GSE18829)/Homer	1e-2	-6.624e+00
2		Gata2(Zf)/K562-GATA2-ChIP-Seq(GSE18829)/Homer	1e-2	-5.216e+00

Figure 113 Transcription factors' motifs enriched at peaks in day4 mESCs annotated to both expressed and DE genes from vascular-endothelial lineage cells expressed gene list. Here is used HOMER software for screening for Enrichment of Known Motifs.

These transcription factors seem to be involved in vascular development, angiogenesis, and endothelial cell function.

Gata1 is specifically expressed during hematopoietic development of erythroid and megakaryocytic cell lineages. *Gata1* knockouts (KO) (*Gata1*^{-/-}) mice die between E9.5 to E10 due to a block of differentiation at the proerythroblast stage, leading to the absence of mature red blood cell (Ferreira R. *et al.*, (2005)). Moreover, *Gata1* promotes angiogenesis by transcriptional activation of the AGGF1 gene. Indeed, knockdown of *Gata1* expression reduced expression of AGGF1 and resulted in inhibition of endothelial vessel formation *in vitro*. Endothelial cell migration is also inhibited, and endothelial cell apoptosis is induced (Fan C. *et al.*, (2009)).

Gata2 is an important regulator of both hematopoietic and endothelial genes. It is the most abundantly expressed GATA factor in endothelial cells. Experiments in embryonic stem cells demonstrated the importance of *Gata2* in the development of *Flk-1*⁺/*Tal1*⁺ hemangioblast-like cells and in the induction of endothelial-specific genes. *Gata2* was a direct target of BMP4 and the activation of *Gata2* upregulated

Bmp4, *Flk1* and *Scl* gene expression. Thus, it plays an important role in mediating endothelial gene expression and the maintenance of endothelial cell fate (Lugus JJ. *et al.*, (2007)).

JunB (encoding a member of the AP-1 transcription factor family) has been implicated in angiogenesis, and its expression is induced by hypoxia and VEGF. In endothelial cells, *JunB* regulates endothelial cell functions as a downstream factor of VEGF signaling (Jia J. *et al.*, (2016)). It is upregulated in tip cells and contributes to vascular development in mouse embryonic skin and retinal vasculatures. The VEGF signals induce *JunB* expression resulting in the conversion of endothelial cells to tip cells. Thus, *JunB* seems to be an angiogenic factor that induces endothelial cell migration and sprouting during vascular development (Yoshitomi Y. *et al.*, (2021)).

In future, it will be necessary to validate the binding of *Gata1*, *Gata2* and *JunB* at the relevant ATAC peaks to confirm their localization by ChIP. Furthermore, it will be interesting to determine whether the alteration of their expression might have a role in early stage-specific EC differentiation on our system.

A highly efficient EC differentiation protocol

The ability of mouse embryonic stem cells (mESCs) to differentiate to the cardiovascular lineages has opened new approaches to study the earliest stages of heart development, as well as molecular events involved in vascular endothelial lineage. The development of the cardiovascular system requires the coordinated differentiation of several cell types including endothelial cells (ECs), smooth muscle cells and cardiomyocytes. This process involves the differentiation of cardiopharyngeal mesoderm (CPM) from which these cardiac cell types derive.

Some studies in literature indicated that during mouse development as well as *in vitro* differentiation of embryonic stem cells, several signaling pathways (BMP, Nodal, Wnt/ β -catenin, FGF) interact to induce first *Brachyury* (*Bry*) positive mesodermal precursors, which express also vascular endothelial growth factor receptor 2 (*Vegfr2*, a.k.a *Flk1* or *Kdr*) (Ema M. *et al.*, (2006)).

Later, these initial findings were confirmed by Kattman *et al.*, (2011), demonstrating that the cardiac lineage arises from a *Flk1*⁺ multipotent cardiovascular progenitor, that can develop into cardiomyocyte, endothelial and vascular smooth vascular lineage. In addition, *Mesp1* is also one of the earliest markers of cardiovascular development and are essential for the specification and differentiation of cardiovascular lineages during embryonic development. Recently in 2018, Lescroart *et al.*, corroborate that mouse heart development arises from *Mesp1*-expressing cardiovascular progenitors (CPs) by single cell RNA-sequencing (scRNA-seq) analysis of the earliest stages of cardiovascular lineage specification. Most of these *Mesp1*⁺ CPs are able to differentiate into either CMs or ECs during mouse embryonic development.

Many transcription factors are involved in the activation and maintenance of endothelial gene expression, and it is necessary to define and control the signaling pathways that regulate the specification of the cardiovascular lineages during embryonic development.

In this scenario, in my doctoral work I tried to understand better the genetic and epigenetic mechanisms responsible for multipotent cardiac progenitors (CPM) differentiation into endothelial cells (ECs).

For my experiments, I decided to use mouse embryonic stem cells (mESC) to establish a suitable *in vitro* model for differentiation of cardiopharyngeal mesoderm (CPM) into endothelial cells (ECs).

As described above, when mESC were differentiated *in vitro* by supplying a cocktail of specific growth factors, like Activin A, BMP4, VEGF (a.k.a. VEGF-A) in serum-free media, they can form several cell types, including cardiac

mesodermal progenitors and mature cardiomyocytes (Kattman *et al.*, (2011)), as well as endothelial cells and vascular smooth muscle cells (Patsch *et al.*, (2015)).

By comparison of two published differentiation protocols (the first, called as “CM protocol”, which allows to obtain cardiomyocytes; the second, named as “EC protocol”, which induces more specific endothelial cell differentiation), I developed an “hybrid” endothelial differentiation model. It consists of two main stages, which recapitulate the developmental processes occurring during embryogenesis: (1) induction of common cardiopharyngeal mesoderm progenitors, using a combination of BMP-4, Activin-A and VEGF-A growth factors.; (2) addition of specific growth factors (VEGF-A and forskolin) to differentiate the cells into a homogeneous endothelial cell population.

The obtained results showed that this procedure efficiently generate vascular endothelial cells starting from mESC, in a well-defined process closely resembling *in vivo* development.

Gene expression profiling during the differentiation revealed that pluripotent cell marker genes (such as *Nanog*, *Oct3/4* and *Rex1*) were rapidly downregulated, while *Mespl*, *Brachyury*, *Gata4*, *Pdgfra* mesodermal genes were activated between day2 and day4. Similarly, the expression of endothelial genes like *VE-Cadherin*, *Pecam1*, *Endoglin (Eng)*, *Kdr (Vegfr2 or Flk1)*, *Nos3*, *Gata2*, *Gata6*, *Ets1*, *Flt1* and others were mostly activated at day4 and increased progressively during differentiation, up until day 8 (d8), the end of differentiation procedure. Thanks to this *in vitro* model, I obtained approximately 91% CD144⁺ (VE-Cadherin) cells within 8 days, demonstrating that this approach provides a suitable tool for generating differentiated populations highly enriched for ECs. In addition, assessing the *in vitro* functional features of these mESC-derived ECs, I observed the formation of vascular network-like structures when plated at d8 on Matrigel. Taken together, my experiments confirm the identity and maturity of the mESC-derived ECs. Overall, the choice to use serum-free protocol avoids the FBS-related batch-to-batch variability, ameliorating the experimental reproducibility. Moreover, using well-defined conditions, based on the addition of cocktails of growth factors, allows to induce rapid cardiac mesoderm and endothelial differentiation with high efficiency, resembling *in vivo* development.

Transcription and chromatin changes on a key d2-d4 mesoderm differentiation step identifies putative enhancers for EC differentiation

RNAseq data and ATAC-seq identified 1735 DEGs between d2 and d4. The most enriched terms by GO analysis included angiogenesis, blood vessel development, vasculature development and others, implying the activation of an EC transcription program in the selected interval.

ATAC-seq revealed 6348 DARs indicating broad chromatin remodeling. While the majority of peaks were located in the promoters, the DARs were primarily

annotated to intra- (43.8%) and inter-genic (37-39%) regions, indicating that during the differentiation window observed, chromatin changes occur within the gene body (promoter excluded) or at distal enhancers. Interestingly, enhancers were found mostly in the intergenic and intronic regions, rather than exons.

GO analysis indicated that genes associated with regions that open at d4 are enriched with genes involved in cardiac and endothelial cell fate specification.

By integration of RNA-seq and ATAC-seq analyses we identified candidate cis-regulatory elements that may be responsible for the regulation of the early phases of CPM differentiation into mature ECs. Thus, I have obtained a subset of differentially expressed genes that were also differentially accessible.

Among these, I have found a total of 10 endothelial genes associated to a d4-opened ATAC peak: *Kdr* (*Vegfr2*), *Cdh5* (*VE-Cadherin*), *CD34*, *Eng*, *Flt1* (*Vegfr1*), *Tall* (*Sc11*), *Dusp5*, *Gata6*, *Pecam1* and *Notch1*. Thus, the approach allowed me to identify putative enhancers in some of the most critical endothelial genes.

In 2016, Becker PW. *et al.*, also identified a region of the 10th intron of the *Flk1* gene (*Flk1in10*) as a putative enhancer and tested it in mouse and zebrafish transgenic models. This region robustly directed reporter gene expression in arterial endothelial cells. To verify and characterize the activity of this region, they cloned the 825-bp mouse *Flk1in10* sequence upstream of the silent hsp68 minimal promoter and LacZ reporter gene and used to generate the stable transgenic mouse line *Flk1in10:LacZ*. Analysis of embryonic *Flk1in10:LacZ* mice clearly demonstrated enhancer activity in the developing vasculature (including both venous and arterial compartments) and heart. Furthermore, transgene expression decreased after birth and was absent in adult organs, confirming that the *Flk1in10* sequence represents a developmental endothelial enhancer which becomes restricted to the arterial compartment. The identity percentage between *Flk1in10* (by Becker PW. *et al.*) & *enh.int10Kdr* is 100%.

Besides, Chiang IK *et al.*, (2017) conducted a detailed *in silico* analysis of the human *NOTCH1* locus with the aim of identifying novel, arterial-specific enhancers, using publicly available information describing chromatin modifications in human endothelial cell lines. They were able to pinpoint four regions of DNA rich in endothelial cell-specific H3K4me1 and H3K27ac histone modifications and DNaseI digital genomic footprints, all marks closely associated with enhancer activity. In particular, the human *NOTCH1* locus contains multiple putative endothelial enhancers, named NOTCH1+33, NOTCH1+16, NOTCH1+3/5 and NOTCH1-68. These regions were tested for their ability to drive reporter gene expression specifically in arterial endothelial cells of transient transgenic mice at embryonic day (E) 12-13. Only the NOTCH1+33 and NOTCH1+16 enhancers were able to direct expression in

endothelial cells. A comparison between hNOTCH1+33 (by Chiang IK *et al.*) & enh.int15*Notch1* showed an identity/conservation percentage of 72.5%.

How to validate putative enhancers?

Among the numerous strategies described in literature to measure the ability and the requirement of candidate sequences to drive endogenous gene transcription, I decided to use two CRISPR/Cas9-based approaches to manipulate and characterize candidate transcriptional enhancers: (1) Genetic deletion of putative enhancers; (2) Epigenetic repression/decommission.

CRISPR/Cas9 technology provides a powerful tool to perturb and test enhancer elements in their genomic context, through both mutation and epigenetic modulation.

I found that two *Notch1*-Δ enh.int15. independent mESC clones (#7G and #11B) showed a significant reduction of *Notch1* expression during the later stages of EC differentiation (d6 and d8). Similarly, *Pecam1* expression was also downregulated in two independent *Pecam1*-Δ enh.int2 mESC clones (#5G and #7G) at the same time points. Overall, these results indicated that the regions deleted are required for appropriate expression of the respective genes during EC differentiation process.

Concerning the dCas9-LSD1 enhancer repression, the idea was that gRNAs could specifically direct dCas9-LSD1 repressor to the enhancer region. LSD1 would then demethylate H3K4me and me2, modulating the chromatin landscape and achieving targeted gene repression in differentiating cells.

I have analyzed so far only five putative enhancer regions: *Notch1*-enh.intr15; *Kdr*-enh.intr10; *VE-Cadh*.-enh.intr1; *Eng*-enh.intr2; *Flt1*-enh.intr10. The targeted five loci resulted affected by dCas9-LSD1 repression, indicating that epigenetic enhancer silencing gives rise to relative reduction of gene-related expression, specifically at day8 of differentiation. I plan to complete the validation of the other identified enhancers (*Pecam1*-enh.int2; *Gata6*-enh.int6; *Tal1*-enh.5kb_upTSS; *CD-34*-enh.10kb_upTSS; *Dusp5*-enh.5kb_upTSS).

It remains to understand whether the targeted loci can synergize to regulate endothelial differentiation. For this reason, I would like to carry out experiments where I will transfect simultaneously a pool of gRNAs, targeting different putative enhancers (previously validated) with the dCas9-LSD1 method and then evaluate transcriptional and functional consequences.

The *Notch1*-enh.intr15 is the only enhancer for which I have used both validation strategies. Interestingly, I demonstrated that the two independent validation approaches cause similar transcriptional effect in *Notch1* expression during EC differentiation, by both genomic enhancer deletion (CRISPR/Cas9) and remodeling of enhancer-associated H3K4me1/2 (dCas9-LSD1).

These results suggest that the candidate enhancer in intron15 of *Notch1* gene, which increased its chromatin accessibility at day4 of differentiation, may function as a regulatory element. I observed a relative reduction of *Notch1* mRNA expression upon both deletion of the enh.intr15 and targeting with dCas9LSD1, at day6 and day8 of differentiation.

These exciting findings, demonstrate that the transcriptional effects of the deletion of *Notch1*-enh.intr15, which is intragenic, is not due to alteration of gene body integrity

Collectively, 6 tested out of 10 identified putative enhancers (*Notch1*-enh.intr15; *Kdr*-enh.intr10; *VE-Cadh*.-enh.intr1; *Eng*-enh.intr2; *Flt1*-enh.intr10; *Pecam1*-enh.int2) seems to be regulatory elements and could be involved during later stages of EC differentiation. The other 4 putative enhancer regions have not yet been tested.

Identification of candidate transcription factors regulating EC differentiation

I performed a preliminary, computational Transcription Factors (TFs) motif analysis in DARs related to endothelial cell fate specification. The most common tool used is HOMER software, choosing as background the peakome, which in my experiments is composed of all ATAC peaks from d2 and d4 experiments, associated to expressed genes.

Overall, we identified *Gata1*, *Gata2* and *JunB* TF motifs to be selectively enriched in day4 mESCs compared to day2. Consistently, literature data show that these transcription factors are involved in vascular development, angiogenesis, and endothelial cell function.

It will be necessary to validate the binding of *Gata1*, *Gata2* and *JunB* at the relevant ATAC peaks to confirm their localization, for example by ChIP. Furthermore, it will be interesting to determine whether the alteration of their expression might have a role in early stage-specific EC differentiation on our system.

Future perspectives

In the future, I would like to investigate the involvement of each enhancer region that I have identified, individually or together, in endothelial-specific functional tests. Matrigel assay is widely used in literature to evaluate the angiogenic activity of potential endothelial cells to form vascular tubules. Once the endothelial cells are plated on gelled Matrigel, they attach within an hour, migrate toward each other, form tubes with a lumen, and develop tight cell-cell and cell-matrix contacts generally within 16-24h, mimicking in vivo angiogenesis process (Kleinman HK. *et al.*, (2005)).

In addition, it would be interesting to employ single-cell multiomics technologies, that typically measure multiple types of molecules from the same individual cell, enabling more profound biological insight than can be inferred by analyzing each molecular layer from separate cells. Multiomics analyses enable systematic, high-resolution profiling of DNA, RNA and proteins in individual cells and they have also proven beneficial for obtaining a comprehensive understanding of cellular events (Lee J. *et al.*, (2020)). In particular, we could apply multiome single-cell RNA sequencing (scRNA-seq) and assay for transposase-accessible chromatin sequencing (scATAC-seq) techniques in order to simultaneously profile gene expression and chromatin accessibility from the same cell during EC differentiation. This approach combines discovery of regulatory elements with gene expression to explore gene regulatory interactions driving cell differentiation. The advancement of single-cell technologies, such as scRNA-seq and scATAC-seq, provides powerful tools to uncover complex and dynamic gene regulatory networks during development across different cell types (Li G. *et al.*, (2022)). Recently, Rossi *et al.*, (2021) developed powerful models for studying cardiogenesis using gastruloids. The protocol has now been established in the laboratory and it is clearly visible that these 3-Dimensional gastruloids develop an extensive endothelial network. Therefore, it would be interesting to evaluate the requirement of the candidate enhancers in the correct formation of the vascular-like EC network in gastruloids. It is possible that inactivation or repression of the candidate enhancers could affect a 3D development of the EC network.

Finally, I am interested in exploring further the *Notch1* enhancer that I have identified in the context of *Notch* signaling involvement in EC differentiation. Specifically, it would be interesting to perform experiments to rescue the observed reduced expression caused by enhancer manipulation. We could use specific antagonist and agonist of *Notch1* signaling. In a gain-of-function experiment, I could use a synthetic peptide corresponding to the δ /serrate/Lag-2 domain of Jagged1 (Jag1) that has proven Notch agonistic activity (Weijzen S. *et al.*, (2002)). On the other hand, DAPT (N-[N-(3,5-difluorophenacetyl)-l-alanyl]-s-phenylglycine-butylester) is often used as a specific inhibitor of γ -secretase, which is a blocking agent of Notch pathway (Dorneburg C. *et al.*, (2016)), thus it works as Notch1 signaling antagonist. I will treat wild type mESC with DAPT and evaluate transcriptional and functional consequences. I assume that the chemical disruption of the *Notch1* pathway will lead to similar results obtained after *Notch1*-enh.int15 targeting. Conversely, the two independent *Notch1*- Δ enh.int15. mESC clones and dCas9-LSD1 expressing cells targeted to *Notch1*-enh.int15 could be treated with Jag1 to test whether the agonist will rescue the consequences of enhancer inactivation.

CONCLUSIONS

Collectively, the results obtained during my PhD work indicated that the *in vitro* serum-free differentiation model employed efficiently generate vascular endothelial cells starting from mESC with high efficacy. It is a well-defined process, which closely resembles *in vivo* development. Gene expression profiling confirmed that CPM and EC lineage commitment is induced at the early stages of differentiation, between day2 and day4. My data have proven that this is a good model to study EC differentiation and search for cis-regulatory elements relevant for the differentiation process. There are many EC-related diseases and my model may help studying them.

Moreover, my approach to putative enhancer identification using the model above has demonstrated effective for the identification of 10 novel enhancer. Of these, I was able to validate each one of those that I have tested. Thus, the approach is effective and could be scaled-up for larger projects. Furthermore, the validation approaches used here avoid (albeit do not replace completely) the use of transgenic mouse models, and it could be applied to human cell models to validate human enhancers.

Lastly, I have identified *Gata1*, *Gata2* and *JunB* as potential regulators of my EC enhancers, providing an additional, upstream regulatory layer in the EC differentiation.

In conclusion, the experimental model and methods used for differentiation of CPM into ECs allowed me to efficiently identify novel putative endothelial enhancers. By genetic and epigenetic perturbations of these sequences, I established their requirement for the transcription process during differentiation. The new knowledge and reagents produced during my thesis work would be instrumental for future research into EC differentiation.

LIST OF PUBLICATIONS

Cirino A, Aurigemma I, Franzese M, Lania G, Righelli D, Ferrentino R, Illingworth E, Angelini C, Baldini A. Chromatin and Transcriptional Response to Loss of TBX1 in Early Differentiation of Mouse Cells. *Front Cell Dev Biol.* 2020 Sep 8;8:571501. doi: 10.3389/fcell.2020.571501. PMID: 33015063; PMCID: PMC7505952.

REFERENCES

- Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab.* 2016 Jan 11;5(3):233-244. doi: 10.1016/j.molmet.2016.01.002. PMID: 26977395; PMCID: PMC4770267.
- Akil A, Gutiérrez-García AK, Guenter R, et al. Notch Signaling in Vascular Endothelial Cells, Angiogenesis, and Tumor Progression: An Update and Prospective. *Frontiers in Cell and Developmental Biology.* 2021; 9:642352. DOI: 10.3389/fcell.2021.642352. PMID: 33681228; PMCID: PMC7928398.
- Aguilar-Sanchez C, Michael M, Pennings S. Cardiac Stem Cells in the Postnatal Heart: Lessons from Development. *Stem Cells Int.* 2018 Jun 24; 2018:1247857. doi: 10.1155/2018/1247857. PMID: 30034478; PMCID: PMC6035836.
- Aquila G, Kostina A, Vieceli Dalla Sega F, Shlyakhto E, Kostareva A, Marracino L, Ferrari R, Rizzo P, Malaschicheva A. The Notch pathway: a novel therapeutic target for cardiovascular diseases? *Expert Opin Ther Targets.* 2019 Aug;23(8):695-710. doi: 10.1080/14728222.2019.1641198. Epub 2019 Jul 14. PMID: 31304807.
- Arnaoutova, I. & Kleinman, H. K. 2010. In vitro angiogenesis: endothelial cell tube formation on gelled basement membrane extract. *Nat Protoc*, 5, 628-35.
- Arthur HM, Ure J, Smith AJ, Renforth G, Wilson DI, Torsney E, Charlton R, Parums DV, Jowett T, Marchuk DA, Burn J, Diamond AG. Endoglin, an ancillary TGFbeta receptor, is required for extraembryonic angiogenesis and plays a key role in heart development. *Dev Biol.* 2000 Jan 1;217(1):42-53. doi: 10.1006/dbio.1999.9534. PMID: 10625534.
- Asahara T, Kawamoto A, Masuda H. Concise review: Circulating endothelial progenitor cells for vascular medicine. *Stem Cells.* 2011 Nov;29(11):1650-5. doi: 10.1002/stem.745. PMID: 21948649.
- Astorga J, Carlsson P. Hedgehog induction of murine vasculogenesis is mediated by Foxf1 and Bmp4. *Development.* 2007 Oct;134(20):3753-61. doi: 10.1242/dev.004432. Epub 2007 Sep 19. PMID: 17881493.
- Auer PL, Doerge RW. Statistical design and analysis of RNA sequencing data. *Genetics.* 2010 Jun;185(2):405-16. doi: 10.1534/genetics.110.114983. Epub 2010 May 3. PMID: 20439781; PMCID: PMC2881125.

Baldwin HS, Shen HM, Yan HC, DeLisser HM, Chung A, Mickanin C, Trask T, Kirschbaum NE, Newman PJ, Albelda SM, et al. Platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31): alternatively spliced, functionally distinct isoforms expressed during mammalian cardiovascular development. *Development*. 1994 Sep;120(9):2539-53. doi: 10.1242/dev.120.9.2539. PMID: 7956830.

Bates DO. Vascular endothelial growth factors and vascular permeability. *Cardiovasc Res*. 2010 Jul 15;87(2):262-71. doi: 10.1093/cvr/cvq105. Epub 2010 Apr 16. PMID: 20400620; PMCID: PMC2895541.

Becker PW, Sacilotto N, Nornes S, Neal A, Thomas MO, Liu K, Preece C, Ratnayaka I, Davies B, Bou-Gharios G, De Val S. An Intronic Flk1 Enhancer Directs Arterial-Specific Expression via RBPJ-Mediated Venous Repression. *Arterioscler Thromb Vasc Biol*. 2016 Jun;36(6):1209-19. doi: 10.1161/ATVBAHA.116.307517. Epub 2016 Apr 14. PMID: 27079877; PMCID: PMC4894770.

Beerli RR, Barbas CF 3rd. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol*. 2002 Feb;20(2):135-41. doi: 10.1038/nbt0202-135. PMID: 11821858.

Benton ML, Talipineni SC, Kostka D, Capra JA. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC Genomics*. 2019 Jun 20;20(1):511. doi: 10.1186/s12864-019-5779-x. PMID: 31221079; PMCID: PMC6585034

Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, Witherspoon K, Gerdtts J, Baker C, Vulto-van Silfhout AT, Schuurs-Hoeijmakers JH, Fichera M, Bosco P, Buono S, Alberti A, Failla P, Peeters H, Steyaert J, Vissers LELM, Francescatto L, Mefford HC, Rosenfeld JA, Bakken T, O'Roak BJ, Pawlus M, Moon R, Shendure J, Amaral DG, Lein E, Rankin J, Romano C, de Vries BBA, Katsanis N, Eichler EE. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell*. 2014 Jul 17;158(2):263-276. doi: 10.1016/j.cell.2014.06.017. Epub 2014 Jul 3. PMID: 24998929; PMCID: PMC4136921.

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res*. 2013 Aug;41(15):7429-37. doi: 10.1093/nar/gkt520. Epub 2013 Jun 12. PMID: 23761437; PMCID: PMC3753641.

Birdsey GM, Dryden NH, Amsellem V, Gebhardt F, Sahnun K, Haskard DO, Dejana E, Mason JC, Randi AM. Transcription factor Erg regulates angiogenesis and endothelial apoptosis through VE-cadherin. *Blood*. 2008 Apr 1;111(7):3498-506. doi: 10.1182/blood-2007-08-105346. Epub 2008 Jan 14. PMID: 18195090; PMCID: PMC2275018.

Blanco R, Gerhardt H, Gerhardt H. VEGF and Notch in tip and stalk cell selection. *Cold Spring Harbor Perspectives in Medicine*. 2013 Jan;3(1):a006569. DOI: 10.1101/cshperspect.a006569. PMID: 23085847; PMCID: PMC3530037.

Bondue A, Tännler S, Chiapparo G, Chabab S, Ramialison M, Paulissen C, Beck B, Harvey R, Blanpain C. Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. *J Cell Biol*. 2011 Mar 7;192(5):751-65. doi: 10.1083/jcb.201007063. PMID: 21383076; PMCID: PMC3051813.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013 Dec;10(12):1213-8. doi: 10.1038/nmeth.2688. Epub 2013 Oct 6. PMID: 24097267; PMCID: PMC3959825.

Brade T, Pane LS, Moretti A, Chien KR, Laugwitz KL. Embryonic heart progenitors and cardiogenesis. *Cold Spring Harb Perspect Med*. 2013 Oct 1;3(10):a013847. doi: 10.1101/cshperspect.a013847. PMID: 24086063; PMCID: PMC3784811.

Breunig CT, Köferle A, Neuner AM, Wiesbeck MF, Baumann V, Stricker SH. CRISPR Tools for Physiology and Cell State Changes: Potential of Transcriptional Engineering and Epigenome Editing. *Physiol Rev*. 2021 Jan 1;101(1):177-211. doi: 10.1152/physrev.00034.2019. Epub 2020 Jun 11. PMID: 32525760.

Bruneau BG. The developmental genetics of congenital heart disease. *Nature*. 2008 Feb 21;451(7181):943-8. doi: 10.1038/nature06801. PMID: 18288184.

Brutsaert DL. Cardiac endothelial-myocardial signaling: its role in cardiac growth, contractile performance, and rhythmicity. *Physiol Rev*. 2003 Jan;83(1):59-115. doi: 10.1152/physrev.00017.2002. PMID: 12506127.

Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources of myocardial cells. *Nat Rev Genet*. 2005 Nov;6(11):826-35. doi: 10.1038/nrg1710. PMID: 16304598.

Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell*. 2003 Dec;5(6):877-89. doi: 10.1016/s1534-5807(03)00363-0. PMID: 14667410; PMCID: PMC5578462.

Cai J, Pardali E, Sánchez-Duffhues G, ten Dijke P. BMP signaling in vascular diseases. *FEBS Lett*. 2012 Jul 4;586(14):1993-2002. doi: 10.1016/j.febslet.2012.04.030. Epub 2012 May 3. PMID: 22710160.

Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013 Mar 7;49(5):825-37. doi: 10.1016/j.molcel.2013.01.038. PMID: 23473601; PMCID: PMC3857148.

Cao G, Fehrenbach ML, Williams JT, Finklestein JM, Zhu JX, Delisser HM. Angiogenesis in platelet endothelial cell adhesion molecule-1-null mice. *Am J Pathol*. 2009 Aug;175(2):903-15. doi: 10.2353/ajpath.2009.090206. Epub 2009 Jul 2. PMID: 19574426; PMCID: PMC2716984.

Cao G, O'Brien CD, Zhou Z, Sanders SM, Greenbaum JN, Makrigiannakis A, DeLisser HM. Involvement of human PECAM-1 in angiogenesis and in vitro endothelial cell migration. *Am J Physiol Cell Physiol*. 2002 May;282(5):C1181-90. doi: 10.1152/ajpcell.00524.2001. PMID: 11940533.

Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoek A, Harpal K, Eberhardt C, Declercq C, Pawling J, Moons L, Collen D, Risau W, Nagy A. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature*. 1996 Apr 4;380(6573):435-9. doi: 10.1038/380435a0. PMID: 8602241.

Cebola I. Deletion of Regulatory Elements with All-in-One CRISPR-Cas9 Vectors. *Methods Mol Biol*. 2021;2351:321-334. doi: 10.1007/978-1-0716-1597-3_18. PMID: 34382198.

Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N, Leong KW. A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Reports*. 2014 Dec 9;3(6):940-7. doi: 10.1016/j.stemcr.2014.09.013. Epub 2014 Oct 23. PMID: 25448066; PMCID: PMC4264059.

Chakraborty S, Yutzey KE. Tbx20 regulation of cardiac cell proliferation and lineage specialization during embryonic and fetal development in vivo. *Dev Biol*. 2012 Mar 1;363(1):234-46. doi: 10.1016/j.ydbio.2011.12.034. Epub 2011 Dec 29. PMID: 22226977; PMCID: PMC3296120.

Chan SS, Chan HHW, Kyba M. Heterogeneity of *Mesp1*⁺ mesoderm revealed by single-cell RNA-seq. *Biochem Biophys Res Commun*. 2016 Jun 3;474(3):469-475. doi: 10.1016/j.bbrc.2016.04.139. Epub 2016 Apr 27. PMID: 27131741; PMCID: PMC4873396.

Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, Lin S, Kiani S, Guzman CD, Wiegand DJ, Ter-Ovanesyan D, Braff JL, Davidsohn N, Housden BE, Perrimon N, Weiss R, Aach J, Collins JJ, Church GM. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods*. 2015 Apr;12(4):326-8. doi: 10.1038/nmeth.3312. Epub 2015 Mar 2. PMID: 25730490; PMCID: PMC4393883.

Chen L, Fulcoli FG, Tang S, Baldini A. *Tbx1* regulates proliferation and differentiation of multipotent heart progenitors. *Circ Res*. 2009 Oct 23;105(9):842-51. doi: 10.1161/CIRCRESAHA.109.200295. Epub 2009 Sep 10. PMID: 19745164; PMCID: PMC2796444.

Chen L, Mupo A, Huynh T, Cioffi S, Woods M, Jin C, McKeehan W, Thompson-Snipes L, Baldini A, Illingworth E. *Tbx1* regulates *Vegfr3* and is required for lymphatic vessel development. *J Cell Biol*. 2010 May 3;189(3):417-24. doi: 10.1083/jcb.200912037. PMID: 20439995; PMCID: PMC2867300.

Chiang IK, Fritzsche M, Pichol-Thievend C, Neal A, Holmes K, Lagendijk A, Overman J, D'Angelo D, Omini A, Hermkens D, Lesieur E, Liu K, Ratnayaka I, Corada M, Bou-Gharios G, Carroll J, Dejana E, Schulte-Merker S, Hogan B, Beltrame M, De Val S, Francois M. *SoxF* factors induce *Notch1* expression via direct transcriptional regulation during early arterial development. *Development*. 2017 Jul 15;144(14):2629-2639. doi: 10.1242/dev.146241. Epub 2017 Jun 15. Erratum in: *Development*. 2017 Oct 15;144(20):3847-3848. PMID: 28619820; PMCID: PMC5536923.

Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at *BRCA1* promoter. *Oncotarget*. 2016 Jul 19;7(29):46545-46556. doi: 10.18632/oncotarget.10234. PMID: 27356740; PMCID: PMC5216816.

Chung YS, Zhang WJ, Arentson E, Kingsley PD, Palis J, Choi K. Lineage analysis of the hemangioblast as defined by *FLK1* and *SCL* expression. *Development*. 2002 Dec;129(23):5511-20. doi: 10.1242/dev.00149. PMID: 12403720.

Cirino A, Aurigemma I, Franzese M, Lania G, Righelli D, Ferrentino R, Illingworth E, Angelini C, Baldini A. Chromatin and Transcriptional Response to Loss of *TBX1* in Early Differentiation of Mouse Cells. *Front Cell Dev Biol*. 2020

Sep 8; 8:571501. doi: 10.3389/fcell.2020.571501. PMID: 33015063; PMCID: PMC7505952.

Claesson-Welsh, L, Welsh, M Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden and Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden VEGFA and tumour angiogenesis (Review). *J Intern Med* 2013; 273: 114– 127

Claringbould Annique, Zaugg Judith B., Enhancers in disease: molecular basis and emerging treatment strategies, *Trends in Molecular Medicine*, Volume 27, Issue 11, 2021, Pages 1060-1073, ISSN 1471-4914, <https://doi.org/10.1016/j.molmed.2021.07.012>.

Cruz-Molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R, van Ijcken WFJ, Grosveld F, Frommolt P, Bazzi H, Rada-Iglesias A. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*. 2017 May 4;20(5):689-705.e9. doi: 10.1016/j.stem.2017.02.004. Epub 2017 Mar 9. PMID: 28285903.

Cui K, Zhao K. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol Biol*. 2012;833:413-9. doi: 10.1007/978-1-61779-477-3_24. PMID: 22183607; PMCID: PMC3541821.

Cunningham TJ, Lancman JJ, Berenguer M, Dong PDS, Duester G. Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. *Cell Rep*. 2018 Jun 12;23(11):3146-3151. doi: 10.1016/j.celrep.2018.05.052. PMID: 29898387; PMCID: PMC6034701.

Dailey L. High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics. *Genomics*. 2015 Sep;106(3):151-158. doi: 10.1016/j.ygeno.2015.06.004. Epub 2015 Jun 10. PMID: 26072436.

DeLisser HM, Christofidou-Solomidou M, Strieter RM, Burdick MD, Robinson CS, Wexler RS, Kerr JS, Garlanda C, Merwin JR, Madri JA, Albelda SM. Involvement of endothelial PECAM-1/CD31 in angiogenesis. *Am J Pathol*. 1997 Sep;151(3):671-7. PMID: 9284815; PMCID: PMC1857836.

De Val S, Black BL. Transcriptional control of endothelial cell development. *Dev Cell*. 2009 Feb;16(2):180-95. doi: 10.1016/j.devcel.2009.01.014. PMID: 19217421; PMCID: PMC2728550.

Devine WP, Wythe JD, George M, Koshiba-Takeuchi K, Bruneau BG. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *Elife*.

2014 Oct 8;3:e03848. doi: 10.7554/eLife.03848. PMID: 25296024; PMCID: PMC4356145.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol*. 2016 Jan;17(1):5-15. doi: 10.1038/nrm.2015.2. Epub 2015 Dec 16. PMID: 26670017; PMCID: PMC4922510.

Dorneburg C, Goß AV, Fischer M, Roels F, Barth TF, Berthold F, Kappler R, Oswald F, Siveke JT, Molenaar JJ, Debatin KM, Beltinger C. γ -Secretase inhibitor I inhibits neuroblastoma cells, with NOTCH and the proteasome among its targets. *Oncotarget*. 2016 Sep 27;7(39):62799-62813. doi: 10.18632/oncotarget.11715. PMID: 27588497; PMCID: PMC5325329.

Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014 Nov 28;346(6213):1258096. doi: 10.1126/science.1258096. PMID: 25430774.

Duarte A, Hirashima M, Benedito R, Trindade A, Diniz P, Bekman E, Costa L, Henrique D, Rossant J. Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev*. 2004 Oct 15;18(20):2474-8. doi: 10.1101/gad.1239004. Epub 2004 Oct 1. PMID: 15466159; PMCID: PMC529534.

Ehling M, Adams S, Benedito R, Adams RH. Notch controls retinal blood vessel maturation and quiescence. *Development*. 2013 Jul;140(14):3051-61. doi: 10.1242/dev.093351. Epub 2013 Jun 19. PMID: 23785053.

Ema M, Takahashi S, Rossant J. Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood*. 2006 Jan 1;107(1):111-7. doi: 10.1182/blood-2005-05-1970. Epub 2005 Sep 15. PMID: 16166582.

Ema M, Yokomizo T, Wakamatsu A, Terunuma T, Yamamoto M, Takahashi S. Primitive erythropoiesis from mesodermal precursors expressing VE-cadherin, PECAM-1, Tie2, endoglin, and CD34 in the mouse embryo. *Blood*. 2006 Dec 15;108(13):4018-24. doi: 10.1182/blood-2006-03-012872. Epub 2006 Aug 22. PMID: 16926294.

Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011 May 5;473(7345):43-9. doi: 10.1038/nature09906. Epub 2011 Mar 23. PMID: 21441907; PMCID: PMC3088773.

Esnault C, Magat T, García-Oliver E, Andrau JC. Analyses of Promoter , Enhancer, and Nucleosome Organization in Mammalian Cells by MNase-Seq. *Methods Mol Biol*. 2021;2351:93-104. doi: 10.1007/978-1-0716-1597-3_5. PMID: 34382185.

Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 1981 Jul 9;292(5819):154-6. doi: 10.1038/292154a0. PMID: 7242681.

Fan C, Ouyang P, Timur AA, He P, You SA, Hu Y, Ke T, Driscoll DJ, Chen Q, Wang QK. Novel roles of GATA1 in regulation of angiogenic factor AGGF1 and endothelial cell function. *J Biol Chem*. 2009 Aug 28;284(35):23331-43. doi: 10.1074/jbc.M109.036079. Epub 2009 Jun 25. PMID: 19556247; PMCID: PMC2749107.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012 Sep;7(9):1728-40.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012 Sep;7(9):1728-40.

Ferdous A, Caprioli A, Iacovino M, Martin CM, Morris J, Richardson JA, Latif S, Hammer RE, Harvey RP, Olson EN, Kyba M, Garry DJ. Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. *Proc Natl Acad Sci U S A*. 2009 Jan 20;106(3):814-9. doi: 10.1073/pnas.0807583106. Epub 2009 Jan 7. PMID: 19129488; PMCID: PMC2630085.

Ferreira R, Ohneda K, Yamamoto M, Philipsen S. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol Cell Biol*. 2005 Feb;25(4):1215-27. doi: 10.1128/MCB.25.4.1215-1227.2005. PMID: 15684376; PMCID: PMC548021.

Fischer A, Schumacher N, Maier M, Sendtner M, Gessler M. The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. *Genes Dev*. 2004 Apr 15;18(8):901-11. doi: 10.1101/gad.291004. PMID: 15107403; PMCID: PMC395849.

Flamme I, Frölich T, Risau W. Molecular mechanisms of vasculogenesis and embryonic angiogenesis. *J Cell Physiol.* 1997 Nov;173(2):206-10. doi: 10.1002/(SICI)1097-4652(199711)173:2<206::AID-JCP22>3.0.CO;2-C. PMID: 9365523.

Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, Ho TW, Kattamis A, Kernytzsky A, Lekstrom-Himes J, Li AM, Locatelli F, Mapara MY, de Montalembert M, Rondelli D, Sharma A, Sheth S, Soni S, Steinberg MH, Wall D, Yen A, Corbacioglu S. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. *N Engl J Med.* 2021 Jan 21;384(3):252-260. doi: 10.1056/NEJMoa2031054. Epub 2020 Dec 5. PMID: 33283989.

Fong GH, Rossant J, Gertsenstein M, Breitman ML. Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. *Nature.* 1995 Jul 6;376(6535):66-70. doi: 10.1038/376066a0. PMID: 7596436.

Furuyama T, Kitayama K, Shimoda Y, Ogawa M, Sone K, Yoshida-Araki K, Hisatsune H, Nishikawa S, Nakayama K, Nakayama K, Ikeda K, Motoyama N, Mori N. Abnormal angiogenesis in Foxo1 (Fkhr)-deficient mice. *J Biol Chem.* 2004 Aug 13;279(33):34741-9. doi: 10.1074/jbc.M314214200. Epub 2004 Jun 7. PMID: 15184386.

Garlanda C, Dejana E. Heterogeneity of endothelial cells. Specific markers. *Arterioscler Thromb Vasc Biol.* 1997 Jul;17(7):1193-202. doi: 10.1161/01.atv.17.7.1193. PMID: 9261246.

George V, Colombo S, Targoff KL. An early requirement for nkx2.5 ensures the first and second heart field ventricular identity and cardiac function into adulthood. *Dev Biol.* 2015 Apr 1;400(1):10-22. doi: 10.1016/j.ydbio.2014.12.019. Epub 2014 Dec 20. PMID: 25536398; PMCID: PMC4361364.

Gessert S, Kühl M. The multiple phases and faces of wnt signaling during cardiac differentiation and development. *Circ Res.* 2010 Jul 23;107(2):186-99. doi: 10.1161/CIRCRESAHA.110.221531. PMID: 20651295.

Giannotta M, Trani M, Dejana E. VE-cadherin and endothelial adherens junctions: active guardians of vascular integrity. *Dev Cell.* 2013 Sep 16;26(5):441-54. doi: 10.1016/j.devcel.2013.08.020. PMID: 24044891.

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.

Cell. 2013 Jul 18;154(2):442-51. doi: 10.1016/j.cell.2013.06.044. Epub 2013 Jul 11. PMID: 23849981; PMCID: PMC3770145.

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014 Oct 23;159(3):647-61. doi: 10.1016/j.cell.2014.09.029. Epub 2014 Oct 9. PMID: 25307932; PMCID: PMC4253859.

Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 2007 Jun;17(6):877-85. doi: 10.1101/gr.5533506. Epub 2006 Dec 19. PMID: 17179217; PMCID: PMC1891346.

Gougos A, Letarte M. Primary structure of endoglin, an RGD-containing glycoprotein of human endothelial cells. J Biol Chem. 1990 May 25;265(15):8361-4. PMID: 1692830.

Goumans MJ, Ten Dijke P. TGF- β Signaling in Control of Cardiovascular Function. Cold Spring Harb Perspect Biol. 2018 Feb 1;10(2):a022210. doi: 10.1101/cshperspect. a022210. PMID: 28348036; PMCID: PMC5793760.

Gourdie RG, Wei Y, Kim D, Klatt SC, Mikawa T. Endothelin-induced conversion of embryonic heart muscle cells into impulse-conducting Purkinje fibers. Proc Natl Acad Sci U S A. 1998 Jun 9;95(12):6815-8. doi: 10.1073/pnas.95.12.6815. PMID: 9618495; PMCID: PMC22646.

Gridley T. Notch signaling in the vasculature. Curr Top Dev Biol. 2010;92:277-309. doi: 10.1016/S0070-2153(10)92009-7. PMID: 20816399; PMCID: PMC3659771.

Gröschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BAM, Erpelinck C, van der Velden VHJ, Havermans M, Avellino R, van Lom K, Rombouts EJ, van Duin M, Döhner K, Beverloo HB, Bradner JE, Döhner H, Löwenberg B, Valk PJM, Bindels EMJ, de Laat W, Delwel R. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell. 2014 Apr 10;157(2):369-381. doi: 10.1016/j.cell.2014.02.019. Epub 2014 Apr 3. PMID: 24703711.

Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, Wu R, Chen C, Li X, Zhou L, He M, Li Z, Sun X, Jia W, Chen J, Yang S, Zhou F, Zhao X, Wan S, Ye R, Liang C, Liu Z, Huang P, Liu C, Jiang H, Wang Y, Zheng H, Sun L, Liu X, Jiang Z, Feng D, Chen J, Wu S, Zou J, Zhang Z, Yang R, Zhao J, Xu C, Yin W, Guan Z, Ye J,

Zhang H, Li J, Kristiansen K, Nickerson ML, Theodorescu D, Li Y, Zhang X, Li S, Wang J, Yang H, Wang J, Cai Z. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet.* 2011 Aug 7;43(9):875-8. doi: 10.1038/ng.907. PMID: 21822268; PMCID: PMC5373841.

Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y, Chen H, Wang J, Tang H, Ge W, Zhou Y, Ye F, Jiang M, Wu J, Xiao Y, Jia X, Zhang T, Ma X, Zhang Q, Bai X, Lai S, Yu C, Zhu L, Lin R, Gao Y, Wang M, Wu Y, Zhang J, Zhan R, Zhu S, Hu H, Wang C, Chen M, Huang H, Liang T, Chen J, Wang W, Zhang D, Guo G. Construction of a human cell landscape at single-cell level. *Nature.* 2020 May;581(7808):303-309. doi: 10.1038/s41586-020-2157-4. Epub 2020 Mar 25. PMID: 32214235.

Hannon GJ. RNA interference. *Nature.* 2002 Jul 11;418(6894):244-51. doi: 10.1038/418244a. PMID: 12110901.

Happe CL. and Engler A. J., “Mechanical forces reshape differentiation cues that guide cardiomyogenesis,” *Circulation Research*, vol. 118, no. 2, pp. 296–310, 2016.

Hayashi H, Kume T. Foxc transcription factors directly regulate Dll4 and Hey2 expression by interacting with the VEGF-Notch signaling pathways in endothelial cells. *PLoS One.* 2008 Jun 11;3(6):e2401. doi: 10.1371/journal.pone.0002401. PMID: 18545664; PMCID: PMC2398774.

Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol.* 2015 Mar;16(3):144-54. doi: 10.1038/nrm3949. Epub 2015 Feb 4. PMID: 25650801; PMCID: PMC4517609.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol.* 2015 May;33(5):510-7. doi: 10.1038/nbt.3199. Epub 2015 Apr 6. PMID: 25849900; PMCID: PMC4430400.

Hsu, P.D.; Lander, E.S.; Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014, 157, 1262–1278

Ishii Y, Langberg J, Rosborough K, Mikawa T. Endothelial cell lineages of the heart. *Cell Tissue Res.* 2009 Jan;335(1):67-73. doi: 10.1007/s00441-008-0663-z. Epub 2008 Aug 6. PMID: 18682987; PMCID: PMC2729171.

Jia J, Ye T, Cui P, Hua Q, Zeng H, Zhao D. AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvasc Res.* 2016

May;105:103-8. doi: 10.1016/j.mvr.2016.02.004. Epub 2016 Feb 6. PMID: 26860974; PMCID: PMC4836857.

Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SY, Ge K. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J*. 2011 Jan 19;30(2):249-62. doi: 10.1038/emboj.2010.318. Epub 2010 Dec 3. PMID: 21131905; PMCID: PMC3025463.

Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell*. 2013 Aug 8;51(3):310-25. doi: 10.1016/j.molcel.2013.07.010. PMID: 23932714; PMCID: PMC3779836.

Kattman SJ, Huber TL, Keller GM. Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev Cell*. 2006 Nov;11(5):723-32. doi: 10.1016/j.devcel.2006.10.002. PMID: 17084363.

Kearney JB, Ambler CA, Monaco KA, Johnson N, Rapoport RG, Bautch VL. Vascular endothelial growth factor receptor Flt-1 negatively regulates developmental blood vessel formation by modulating endothelial cell division. *Blood*. 2002 Apr 1;99(7):2397-407. doi: 10.1182/blood.v99.7.2397. PMID: 11895772.

Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development*. 2014 Jan;141(1):219-23. doi: 10.1242/dev.103341. PMID: 24346702; PMCID: PMC3865759.

Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods*. 2015 May;12(5):401-403. doi: 10.1038/nmeth.3325. Epub 2015 Mar 16. PMID: 25775043; PMCID: PMC4414811.

Keller GM. In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol*. 1995 Dec;7(6):862-9. doi: 10.1016/0955-0674(95)80071-9. PMID: 8608017.

Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev*. 2005 May 15;19(10):1129-55. doi: 10.1101/gad.1303605. PMID: 15905405.

Kelly RG, Brown NA, Buckingham ME. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev Cell*. 2001 Sep;1(3):435-40. doi: 10.1016/s1534-5807(01)00040-5. PMID: 11702954.

Kelly RG. 2012. The second heart field. *Curr Top Dev Biol* 100: 33–65.

Kimelman D. Mesoderm induction: from caps to chips. *Nat Rev Genet*. 2006 May;7(5):360-72. doi: 10.1038/nrg1837. PMID: 16619051.

Klein DC, Hainer SJ. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res*. 2020 Mar;28(1):69-85. doi: 10.1007/s10577-019-09619-9. Epub 2019 Nov 27. PMID: 31776829; PMCID: PMC7125251.

Kleinman HK, Martin GR. Matrigel: basement membrane matrix with biological activity. *Semin Cancer Biol*. 2005 Oct;15(5):378-86. doi: 10.1016/j.semcancer.2005.05.004. PMID: 15975825.

Kocher AA, Schuster MD, Szabolcs MJ, Takuma S, Burkhoff D, Wang J, Homma S, Edwards NM, Itescu S. Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. *Nat Med*. 2001 Apr;7(4):430-6. doi: 10.1038/86498. PMID: 11283669.

Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013 Oct 24;502(7472):472-9. doi: 10.1038/nature12750. PMID: 24153300; PMCID: PMC4046508.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015 Jan 29;517(7536):583-8. doi: 10.1038/nature14136. Epub 2014 Dec 10. PMID: 25494202; PMCID: PMC4420636.

Kopczyńska E, Makarewicz R. Endoglin - a marker of vascular endothelial cell proliferation in cancer. *Contemp Oncol (Pozn)*. 2012;16(1):68-71. doi: 10.5114/wo.2012.27340. Epub 2012 Feb 29. PMID: 23788858; PMCID: PMC3687377.

Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development*. 1989 Apr;105(4):707-14. doi: 10.1242/dev.105.4.707. PMID: 2557196.

Krebs LT, Xue Y, Norton CR, Shutter JR, Maguire M, Sundberg JP, Gallahan D, Closson V, Kitajewski J, Callahan R, Smith GH, Stark KL, Gridley T. Notch

signaling is essential for vascular morphogenesis in mice. *Genes Dev.* 2000 Jun 1;14(11):1343-52. PMID: 10837027; PMCID: PMC316662.

Krüger-Genge A, Blocki A, Franke RP, Jung F. Vascular Endothelial Cell Biology: An Update. *Int J Mol Sci.* 2019 Sep 7;20(18):4411. doi: 10.3390/ijms20184411. PMID: 31500313; PMCID: PMC6769656.

La Russa MF, Qi LS. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol Cell Biol.* 2015 Nov;35(22):3800-9. doi: 10.1128/MCB.00512-15. Epub 2015 Sep 14. PMID: 26370509; PMCID: PMC4609748.

Laflamme, M., Chen, K., Naumova, A. et al. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotechnol* 25, 1015–1024 (2007). <https://doi.org/10.1038/nbt1327>

Laforest B, Andelfinger G, Nemer M. Loss of Gata5 in mice leads to bicuspid aortic valve. *J Clin Invest.* 2011 Jul;121(7):2876-87. doi: 10.1172/JCI44555. PMID: 21633169; PMCID: PMC3223824.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 Mar 4;9(4):357-9.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 Mar 4;9(4):357-9.

Lawson ND, Vogel AM, Weinstein BM. sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. *Dev Cell.* 2002 Jul;3(1):127-36. doi: 10.1016/s1534-5807(02)00198-3. PMID: 12110173.

Lederer D, Shears D, Benoit V, Verellen-Dumoulin C, Maystadt I. A three generation X-linked family with Kabuki syndrome phenotype and a frameshift mutation in KDM6A. *Am J Med Genet A.* 2014 May;164A(5):1289-92. doi: 10.1002/ajmg.a.36442. Epub 2014 Mar 24. PMID: 24664873.

Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med.* 2020 Sep;52(9):1428-1442. doi: 10.1038/s12276-020-0420-2. Epub 2020 Sep 15. PMID: 32929225; PMCID: PMC8080692.

Lee NY, Golzio C, Gatz CE, Sharma A, Katsanis N, Blobel GC. Endoglin regulates PI3-kinase/Akt trafficking and signaling to alter endothelial capillary stability during angiogenesis. *Mol Biol Cell.* 2012 Jul;23(13):2412-23. doi:

10.1091/mbc.E11-12-0993. Epub 2012 May 16. PMID: 22593212; PMCID: PMC3386206.

Léna Vouillot, Aurore Thélie, Nicolas Pollet, Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases, *G3 Genes|Genomes|Genetics*, Volume 5, Issue 3, 1 March 2015, Pages 407–415, <https://doi.org/10.1534/g3.114.015834>

Lepore JJ, Mericko PA, Cheng L, Lu MM, Morrissey EE, Parmacek MS. GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J Clin Invest*. 2006 Apr;116(4):929-39. doi: 10.1172/JCI27363. Epub 2006 Mar 23. PMID: 16557299; PMCID: PMC1409743.

Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, Auer H, Achouri Y, Dubois C, Bondue A, Simons BD, Blanpain C. Early lineage restriction in temporally distinct populations of *Mesp1* progenitors during mammalian heart development. *Nat Cell Biol*. 2014 Sep;16(9):829-40. doi: 10.1038/ncb3024. Epub 2014 Aug 24. PMID: 25150979; PMCID: PMC6984965.

Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sánchez-Dànes A, Moignard V, Dubois C, Paulissen C, Kinston S, Göttgens B, Blanpain C. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*. 2018 Mar 9;359(6380):1177-1181. doi: 10.1126/science.aao4174. Epub 2018 Jan 25. PMID: 29371425; PMCID: PMC6556615.

Li C, Hampson IN, Hampson L, Kumar P, Bernabeu C, Kumar S. CD105 antagonizes the inhibitory signaling of transforming growth factor 1 on human vascular endothelial cells. *FASEB J*. 1999; 14:55–64.

Li G, Fu S, Wang S, Zhu C, Duan B, Tang C, Chen X, Chuai G, Wang P, Liu Q. A deep generative model for multi-view profiling of single-cell RNA-seq and ATAC-seq data. *Genome Biol*. 2022 Jan 12;23(1):20. doi: 10.1186/s13059-021-02595-6. PMID: 35022082; PMCID: PMC8756637.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–9, 2009.

Li Y, Rivera CM, Ishii H, Jin F, Selvaraj S, Lee AY, Dixon JR, Ren B. CRISPR reveals a distal super-enhancer required for *Sox2* expression in mouse embryonic stem cells. *PLoS One*. 2014 Dec 8;9(12):e114485. doi: 10.1371/journal.pone.0114485. PMID: 25486255; PMCID: PMC4259346.

Li, Z.J., Wang, Z.Z., Zheng, Y.Z., Xu, B., Yang, R.C., Scadden, D.T. and Han, Z.C. (2005), Kinetic expression of platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31) during embryonic stem cell differentiation. *J. Cell. Biochem.*, 95: 559-570. <https://doi.org/10.1002/jcb.20436>

Limbourg FP, Takeshita K, Radtke F, Bronson RT, Chin MT, Liao JK. Essential role of endothelial Notch1 in angiogenesis. *Circulation*. 2005 Apr 12;111(14):1826-32. doi: 10.1161/01.CIR.0000160870.93058.DD. Epub 2005 Apr 4. PMID: 15809373; PMCID: PMC2633594.

Lindsay EA. Chromosomal microdeletions: dissecting del22q11 syndrome. *Nat Rev Genet*. 2001 Nov;2(11):858-68. doi: 10.1038/35098574. PMID: 11715041.

Lindsley RC, Gill JG, Kyba M, Murphy TL, Murphy KM. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development*. 2006 Oct;133(19):3787-96. doi: 10.1242/dev.02551. Epub 2006 Aug 30. PMID: 16943279.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001 Dec;25(4):402-8. doi: 10.1006/meth.2001.1262. PMID: 11846609.

Liu ZJ, Shirakawa T, Li Y, Soma A, Oka M, Dotto GP, Fairman RM, Velazquez OC, Herlyn M. Regulation of Notch1 and Dll4 by vascular endothelial growth factor in arterial endothelial cells: implications for modulating arteriogenesis and angiogenesis. *Mol Cell Biol*. 2003 Jan;23(1):14-25. doi: 10.1128/MCB.23.1.14-25.2003. PMID: 12482957; PMCID: PMC140667.

Liu C, Wang M, Wei X, Wu L, Xu J, Dai X, Xia J, Cheng M, Yuan Y, Zhang P, Li J, Feng T, Chen A, Zhang W, Chen F, Shang Z, Zhang X, Peters BA, Liu L. An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Sci Data*. 2019 May 20;6(1):65. doi: 10.1038/s41597-019-0071-0. PMID: 31110271; PMCID: PMC6527694.

Lopes R, Korkmaz G, Agami R. Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat Rev Mol Cell Biol*. 2016 Sep;17(9):597-604. doi: 10.1038/nrm.2016.79. Epub 2016 Jul 6. PMID: 27381243.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.

Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013 Apr 11;153(2):320-34. doi: 10.1016/j.cell.2013.03.036. PMID: 23582323; PMCID: PMC3760967.

Lu CX, Gong HR, Liu XY, Wang J, Zhao CM, Huang RT, Xue S, Yang YQ. A novel HAND2 loss-of-function mutation responsible for tetralogy of Fallot. *Int J Mol Med*. 2016 Feb;37(2):445-51. doi: 10.3892/ijmm.2015.2436. Epub 2015 Dec 15. PMID: 26676105.

Lugus JJ, Chung YS, Mills JC, Kim SI, Grass J, Kyba M, Doherty JM, Bresnick EH, Choi K. GATA2 functions at multiple steps in hemangioblast development and differentiation. *Development*. 2007 Jan;134(2):393-405. doi: 10.1242/dev.02731. Epub 2006 Dec 13. PMID: 17166922.

Ma L, Wang J, Li L, Qiao Q, Di RM, Li XM, Xu YJ, Zhang M, Li RG, Qiu XB, Li X, Yang YQ. ISL1 loss-of-function mutation contributes to congenital heart defects. *Heart Vessels*. 2019 Apr;34(4):658-668. doi: 10.1007/s00380-018-1289-z. Epub 2018 Nov 2. PMID: 30390123.

Mack JJ, Iruela-Arispe ML. NOTCH regulation of the endothelial cell phenotype. *Current Opinion in Hematology*. 2018 May;25(3):212-218. DOI: 10.1097/moh.0000000000000425. PMID: 29547401; PMCID: PMC5902133.

Magliulo D, Bernardi R and Messina S (2018) Lysine-Specific Demethylase 1A as a Promising Target in Acute Myeloid Leukemia. *Front. Oncol*. 8:255. doi: 10.3389/fonc.2018.00255

Martin, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, [S.l.], v. 17, n. 1, p. pp. 10-12, may 2011. ISSN 2226-6089. Available at: <<https://journal.embnnet.org/index.php/embnnetjournal/article/view/200>>. Date accessed: 26 nov. 2022. doi: <https://doi.org/10.14806/ej.17.1.200>.

Maston GA, Landt SG, Snyder M, Green MR. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet*. 2012;13:29-57. doi: 10.1146/annurev-genom-090711-163723. Epub 2012 Jun 11. PMID: 22703170.

Maurya SS. Role of Enhancers in Development and Diseases. *Epigenomes*. 2021 Oct 4;5(4):21. doi: 10.3390/epigenomes5040021. PMID: 34968246; PMCID: PMC8715447.

McAllister, K. A., Grogg, K. M., Johnson, D. W., Gallione, C. J., Baldwin, M. A., Jackson, C. E., Helmbold, E. A., Markel, D. S., McKinnon, W. C., Murrell, J., et al. (1994). Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nature Genet.* 8, 345–351

McDonald JL, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open.* 2016 Jun 15;5(6):866-74. doi: 10.1242/bio.019067. PMID: 27170255; PMCID: PMC4920199.

McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E. NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol.* 2003 Nov 5;42(9):1650-5. doi: 10.1016/j.jacc.2003.05.004. PMID: 14607454.

Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol.* 2013 Dec;31(12):1133-6. doi: 10.1038/nbt.2701. Epub 2013 Sep 8. PMID: 24013198; PMCID: PMC3858395.

Mianné J, Bourguignon C, Nguyen Van C, Fieldès M, Nasri A, Assou S, De Vos J. Pipeline for the Generation and Characterization of Transgenic Human Pluripotent Stem Cells Using the CRISPR/Cas9 Technology. *Cells.* 2020 May 25;9(5):1312. doi: 10.3390/cells9051312. PMID: 32466123; PMCID: PMC7290981.

Mishina Y, Suzuki A, Ueno N, Behringer RR. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes Dev.* 1995 Dec 15;9(24):3027-37. doi: 10.1101/gad.9.24.3027. PMID: 8543149.

Montefiori L, Hernandez L, Zhang Z, Gilad Y, Ober C, Crawford G, Nobrega M, Jo Sakabe N. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci Rep.* 2017 May 26;7(1):2451.

Moretti A, Caron L, Nakano A, Lam JT, Bernshausen A, Chen Y, Qyang Y, Bu L, Sasaki M, Martin-Puig S, Sun Y, Evans SM, Laugwitz KL, Chien KR. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell.* 2006 Dec 15;127(6):1151-65. doi: 10.1016/j.cell.2006.10.029. Epub 2006 Nov 22. PMID: 17123592.

Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* 1981 Nov 25;9(22):6047-68. doi: 10.1093/nar/9.22.6047. PMID: 6273820; PMCID: PMC327583

Mori AD, Bruneau BG. TBX5 mutations and congenital heart disease: Holt-Oram syndrome revealed. *Curr Opin Cardiol.* 2004 May;19(3):211-5. doi: 10.1097/00001573-200405000-00004. PMID: 15096952.

Morris SA, Daley GQ. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Res.* 2013 Jan;23(1):33-48. doi: 10.1038/cr.2013.1. Epub 2013 Jan 1. PMID: 23277278; PMCID: PMC3541663.

Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell.* 2008 Feb 22;132(4):661-80. doi: 10.1016/j.cell.2008.02.008. PMID: 18295582.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet.* 2010 Sep;42(9):790-3. doi: 10.1038/ng.646. Epub 2010 Aug 15. PMID: 20711175; PMCID: PMC2930028.

Nomaru, H., Liu, Y., De Bono, C. et al. Single cell multi-omic analysis identifies a Tbx1-dependent multilineage primed population in murine cardiopharyngeal mesoderm. *Nat Commun* 12, 6645 (2021). <https://doi.org/10.1038/s41467-021-26966-6>

Nord AS. Learning about mammalian gene regulation from functional enhancer assays in the mouse. *Genomics.* 2015 Sep;106(3):178-184. doi: 10.1016/j.ygeno.2015.06.008. Epub 2015 Jun 14. PMID: 26079655.

Nusse R. Wnt signaling in disease and in development. *Cell Res.* 2005 Jan;15(1):28-32. doi: 10.1038/sj.cr.7290260. PMID: 15686623.

Oliver G, Alitalo K. The lymphatic vasculature: recent progress and paradigms. *Annu Rev Cell Dev Biol.* 2005;21:457-83. doi: 10.1146/annurev.cellbio.21.012704.132338. PMID: 16212503.

Olley G, Ansari M, Bengani H, Grimes GR, Rhodes J, von Kriegsheim A, Blatnik A, Stewart FJ, Wakeling E, Carroll N, Ross A, Park SM; Deciphering Developmental Disorders Study, Bickmore WA, Pradeepa MM, FitzPatrick DR. BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange-like syndrome. *Nat Genet.* 2018 Mar;50(3):329-332. doi: 10.1038/s41588-018-0042-y. Epub 2018 Jan 29. Erratum in: *Nat Genet.* 2018 Feb 12;; Erratum in: *Nat Genet.* 2019 Jul;51(7):1192. PMID: 29379197; PMCID: PMC6469577.

Ottaviani, G. & Buja, L.M. Chapter 15 - congenital heart disease: Pathology, natural history, and interventions. Cardiovascular Pathology. 4th Ed. 611-647. Academic Press (2016).

Panigrahi, A., O'Malley, B.W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol* 22, 108 (2021). <https://doi.org/10.1186/s13059-021-02322-1>

Pardali E, Goumans MJ, ten Dijke P. Signaling by members of the TGF-beta family in vascular morphogenesis and disease. *Trends Cell Biol.* 2010 Sep;20(9):556-67. doi: 10.1016/j.tcb.2010.06.006. Epub 2010 Jul 23. PMID: 20656490.

Park C, Afrikanova I, Chung YS, Zhang WJ, Arentson E, Fong Gh Gh, Rosendahl A, Choi K. A hierarchical order of factors in the generation of FLK1- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells. *Development.* 2004 Jun;131(11):2749-62. doi: 10.1242/dev.01130. PMID: 15148304.

Park C, Kim TM, Malik AB. Transcriptional regulation of endothelial cell and vascular development. *Circ Res.* 2013 May 10;112(10):1380-400. doi: 10.1161/CIRCRESAHA.113.301078. PMID: 23661712; PMCID: PMC3730491.

Patan S. Vasculogenesis and angiogenesis. *Cancer Treat Res.* 2004; 117:3-32. doi: 10.1007/978-1-4419-8871-3_1. PMID: 15015550.

Patsch, C., Challet-Meylan, L., Thoma, E. et al. Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. *Nat Cell Biol* 17, 994–1003 (2015). <https://doi.org/10.1038/ncb3205>

Patterson LJ, Gering M, Patient R. Scl is required for dorsal aorta as well as blood formation in zebrafish embryos. *Blood.* 2005 May 1;105(9):3502-11. doi: 10.1182/blood-2004-09-3547. Epub 2005 Jan 11. PMID: 15644413.

Peterkin T, Gibson A, Loose M, Patient R. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Semin Cell Dev Biol.* 2005 Feb;16(1):83-94. doi: 10.1016/j.semcdb.2004.10.003. Epub 2004 Dec 15. PMID: 15659343.

Peters KG, deVries C, Williams LT 1993 Vascular endothelial growth factor receptor expression during embryogenesis and tissue repair suggests a role in endothelial differentiation and blood vessel growth. *Proc Natl Acad Sci USA* 90:8915–8919

Pham VN, Lawson ND, Mugford JW, Dye L, Castranova D, Lo B, Weinstein BM. Combinatorial function of ETS transcription factors in the developing vasculature. *Dev Biol.* 2007 Mar 15;303(2):772-83. doi: 10.1016/j.ydbio.2006.10.030. Epub 2006 Oct 25. PMID: 17125762; PMCID: PMC1859867.

Puri MC, Partanen J, Rossant J, Bernstein A. Interaction of the TEK and TIE receptor tyrosine kinases during cardiovascular development. *Development.* 1999 Oct;126(20):4569-80. doi: 10.1242/dev.126.20.4569. PMID: 10498691.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013 Feb 28;152(5):1173-83. doi: 10.1016/j.cell.2013.02.022. Erratum in: *Cell.* 2021 Feb 4;184(3):844. PMID: 23452860; PMCID: PMC3664290.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.

Redick SD, Bautch VL. Developmental platelet endothelial cell adhesion molecule expression suggests multiple roles for a vascular adhesion molecule. *Am J Pathol.* 1999 Apr;154(4):1137-47. doi: 10.1016/S0002-9440(10)65366-7. PMID: 10233852; PMCID: PMC1866548.

Righelli D, Koberstein J, Zhang N, Angelini C, Peixoto L, Risso D. 2018. Differential Enriched Scan 2 (DEScan2): a fast pipeline for broad peak analysis. *PeerJ Preprints* 6:e27357v1 <https://doi.org/10.7287/peerj.preprints.27357v1>

Robertson KD. DNA methylation and human disease. *Nat Rev Genet.* 2005 Aug;6(8):597-610. doi: 10.1038/nrg1655. PMID: 16136652.

Rossi G, Broguiere N, Miyamoto M, Boni A, Guet R, Girgin M, Kelly RG, Kwon C, Lutolf MP. Capturing Cardiogenesis in Gastruloids. *Cell Stem Cell.* 2021 Feb 4;28(2):230-240.e6. doi: 10.1016/j.stem.2020.10.013. Epub 2020 Nov 10. PMID: 33176168; PMCID: PMC7867643.

Ryan GE, Farley EK. Functional genomic approaches to elucidate the role of enhancers during development. *Wiley Interdiscip Rev Syst Biol Med.* 2020 Mar;12(2):e1467. doi: 10.1002/wsbm.1467. Epub 2019 Dec 5. PMID: 31808313; PMCID: PMC7027484.

Sara Cioffi, Stefania Martucciello, Filomena Gabriella Fulcoli, Marchesa Bilio, Rosa Ferrentino, Edoardo Nusco, Elizabeth Illingworth, Tbx1 regulates brain

vascularization, *Human Molecular Genetics*, Volume 23, Issue 1, 1 January 2014, Pages 78–89, <https://doi.org/10.1093/hmg/ddt400>

Sakamoto Y, Hara K, Kanai-Azuma M, Matsui T, Miura Y, Tsunekawa N, Kurohmaru M, Saijoh Y, Koopman P, Kanai Y. Redundant roles of Sox17 and Sox18 in early cardiovascular development of mouse embryos. *Biochem Biophys Res Commun*. 2007 Aug 31;360(3):539-44. doi: 10.1016/j.bbrc.2007.06.093. Epub 2007 Jun 25. PMID: 17610846.

Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet*. 2010 Jul 15;6(7):e1001023. doi: 10.1371/journal.pgen.1001023. PMID: 20657823; PMCID: PMC2904778.

Shalaby F, Rossant J, Yamaguchi TP, Gertsenstein M, Wu XF, Breitman ML, Schuh AC. Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature*. 1995 Jul 6;376(6535):62-6. doi: 10.1038/376062a0. PMID: 7596435.

Shibuya M. Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis. *Angiogenesis*. 2006;9(4):225-30; discussion 231. doi: 10.1007/s10456-006-9055-8. Epub 2006 Nov 16. PMID: 17109193.

Shorstova T, Foulkes WD, Witcher M. Achieving clinical success with BET inhibitors as anti-cancer agents. *Br J Cancer*. 2021 Apr;124(9):1478-1490. doi: 10.1038/s41416-021-01321-0. Epub 2021 Mar 15. PMID: 33723398; PMCID: PMC8076232.

Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*. 2010 Feb;2010(2):pdb.prot5384. doi: 10.1101/pdb.prot5384. PMID: 20150147; PMCID: PMC3627383.

Spicuglia S, Vanhille L. Chromatin signatures of active enhancers. *Nucleus*. 2012 Mar 1;3(2):126-31. doi: 10.4161/nucl.19232. Epub 2012 Mar 1. PMID: 22555596; PMCID: PMC3383566.

Spyropoulos DD, Pharr PN, Lavenburg KR, Jackers P, Papas TS, Ogawa M, Watson DK. Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor. *Mol Cell Biol*. 2000 Aug;20(15):5643-52. doi: 10.1128/MCB.20.15.5643-5652.2000. PMID: 10891501; PMCID: PMC86032.

Stark, Rory and Gord Brown. “DiffBind: Differential binding analysis of ChIP-Seq peak data.” (2012).

Steven J. Kattman, Alec D. Witty, Mark Gagliardi, Nicole C. Dubois, Maryam Niapour, Akitsu Hotta, James Ellis, Gordon Keller, Stage-Specific Optimization of Activin/Nodal and BMP Signaling Promotes Cardiac Differentiation of Mouse and Human Pluripotent Stem Cell Lines, *Cell Stem Cell*, Volume 8, Issue 2, 2011, Pages 228-240, ISSN 1934-5909,

Sun Y, Liang X, Najafi N, Cass M, Lin L, Cai CL, Chen J, Evans SM. Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Dev Biol*. 2007 Apr 1;304(1):286-96. doi: 10.1016/j.ydbio.2006.12.048. Epub 2006 Dec 29. PMID: 17258700; PMCID: PMC2582044.

Swift MR, Weinstein BM. Arterial-venous specification during development. *Circ Res*. 2009 Mar 13;104(5):576-88. doi: 10.1161/CIRCRESAHA.108.188805. PMID: 19286613.

Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*. 2015 Dec 2;43(21):e140. doi: 10.1093/nar/gkv711. Epub 2015 Jul 16. PMID: 26184878; PMCID: PMC4666377.

Terada R, Warren S, Lu JT, Chien KR, Wessels A, Kasahara H. Ablation of Nkx2-5 at mid-embryonic stage results in premature lethality and cardiac malformation. *Cardiovasc Res*. 2011 Jul 15;91(2):289-99. doi: 10.1093/cvr/cvr037. Epub 2011 Feb 1. PMID: 21285290; PMCID: PMC3125071.

Tirosh-Finkel L, Zeisel A, Brodt-Ivenshitz M, Shamai A, Yao Z, Seger R, Domany E, Tzahor E. BMP-mediated inhibition of FGF signaling promotes cardiomyocyte differentiation of anterior heart field progenitors. *Development*. 2010 Sep;137(18):2989-3000. doi: 10.1242/dev.051649. Epub 2010 Aug 11. PMID: 20702560.

Tough DF, Prinjha RK. Immune disease-associated variants in gene enhancers point to BET epigenetic mechanisms for therapeutic intervention. *Epigenomics*. 2017 Apr;9(4):573-584. doi: 10.2217/epi-2016-0144. Epub 2016 Dec 7. PMID: 27925476.

Tsai FY, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, Alt FW, Orkin SH. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*. 1994 Sep 15;371(6494):221-6. doi: 10.1038/371221a0. PMID: 8078582.

Tsurusaki Y, Okamoto N, Ohashi H, Mizuno S, Matsumoto N, Makita Y, Fukuda M, Isidor B, Perrier J, Aggarwal S, Dalal AB, Al-Kindy A, Liebelt J, Mowat D, Nakashima M, Saitsu H, Miyake N, Matsumoto N. Coffin-Siris syndrome is a SWI/SNF complex disorder. *Clin Genet*. 2014 Jun;85(6):548-54. doi: 10.1111/cge.12225. Epub 2013 Jul 23. PMID: 23815551.

Turek-Plewa J, Jagodziński PP. The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett*. 2005;10(4):631-47. PMID: 16341272.

Visvader JE, Fujiwara Y, Orkin SH. Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. *Genes Dev*. 1998 Feb 15;12(4):473-9. doi: 10.1101/gad.12.4.473. PMID: 9472016; PMCID: PMC316527.

Vittet D, Prandini MH, Berthier R, Schweitzer A, Martin-Sisteron H, Uzan G, Dejana E. Embryonic stem cells differentiate in vitro to endothelial cells through successive maturation steps. *Blood*. 1996 Nov 1;88(9):3424-31. PMID: 8896407.

Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, Klasić M, Zoldoš V. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res*. 2016 Jul 8;44(12):5615-28. doi: 10.1093/nar/gkw159. Epub 2016 Mar 11. PMID: 26969735; PMCID: PMC4937303.

von Gise A, Pu WT. Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ Res*. 2012 Jun 8;110(12):1628-45. doi: 10.1161/CIRCRESAHA.111.259960. PMID: 22679138; PMCID: PMC3427736.

Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA, Bruneau BG. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell*. 2012 Sep 28;151(1):206-20. doi: 10.1016/j.cell.2012.07.035. Epub 2012 Sep 12. PMID: 22981692; PMCID: PMC3462286.

Ward CM, Stern P, Willington MA, Flenniken AM. Efficient germline transmission of mouse embryonic stem cells grown in synthetic serum in the absence of a fibroblast feeder layer. *Lab Invest*. 2002 Dec;82(12):1765-7. doi: 10.1097/01.lab.0000043123.37057.f6. PMID: 12480926.

Weijzen S, Velders MP, Elmishad AG, Bacon PE, Panella JR, Nickoloff BJ, Miele L, Kast WM. The Notch ligand Jagged-1 is able to induce maturation of monocyte-derived human dendritic cells. *J Immunol*. 2002 Oct 15;169(8):4273-8. doi: 10.4049/jimmunol.169.8.4273. PMID: 12370358.

Wikström P, Lissbrant IF, Stattin P, Egevad L, Bergh A. Endoglin (CD105) is expressed on immature blood vessels and is a marker for survival in prostate cancer. *Prostate*. 2002 Jun 1;51(4):268-75. doi: 10.1002/pros.10083. PMID: 11987155.

Wu SM, Chien KR, Mummery C. Origins and fates of cardiovascular progenitor cells. *Cell*. 2008 Feb 22;132(4):537-43. doi: 10.1016/j.cell.2008.02.002. PMID: 18295570; PMCID: PMC2507768.

Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol*. 2002 Dec;20(12):1261-4. doi: 10.1038/nbt761. Epub 2002 Nov 11. PMID: 12426580.

Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu GL, Hu R. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov*. 2016 May 3;2:16009. doi: 10.1038/celldisc.2016.9. PMID: 27462456; PMCID: PMC4853773.

Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, Ichida F, Joo K, Kimura M, Imamura S, Kamatani N, Momma K, Takao A, Nakazawa M, Shimizu N, Matsuoka R. Role of TBX1 in human del22q11.2 syndrome. *Lancet*. 2003 Oct 25;362(9393):1366-73. doi: 10.1016/s0140-6736(03)14632-6. PMID: 14585638.

Yamamizu, K., Kawasaki, K., Katayama, S., Watabe, T. & Yamashita, J. K. Enhancement of vascular progenitor potential by protein kinase A through dual induction of Flk-1 and Neuropilin-1. *Blood* 114, 3707–3716 (2009).

Yamashita JK, Takano M, Hiraoka-Kanie M, Shimazu C, Peishi Y, Yanagi K, Nakano A, Inoue E, Kita F, Nishikawa S. Prospective identification of cardiac progenitors by a novel single cell-based cardiomyocyte induction. *FASEB J*. 2005 Sep;19(11):1534-6. doi: 10.1096/fj.04-3540fje. Epub 2005 Jul 20. PMID: 16033809.

Yamashita, J., Itoh, H., Hirashima, M. et al. Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* 408, 92–96 (2000). <https://doi.org/10.1038/35040568>

Yang, L., Soonpaa, M., Adler, E. et al. Human cardiovascular progenitor cells develop from a KDR⁺ embryonic-stem-cell-derived population. *Nature* 453, 524–528 (2008). <https://doi.org/10.1038/nature06894>

Yoshitomi Y, Ikeda T, Saito-Takatsuji H, Yonekura H. Emerging Role of AP-1 Transcription Factor JunB in Angiogenesis and Vascular Development. *Int J Mol Sci.* 2021 Mar 10;22(6):2804. doi: 10.3390/ijms22062804. PMID: 33802099; PMCID: PMC8000613.

Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015 Jul 15;31(14):2382-3.

Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015 Jul 15;31(14):2382-3.

Waddington CH. The epigenotype. 1942. *Int J Epidemiol.* 2012 Feb;41(1):10-3. doi: 10.1093/ije/dyr184. Epub 2011 Dec 20. PMID: 22186258.

Weijzen S, Velders MP, Elmishad AG, Bacon PE, Panella JR, Nickoloff BJ, Miele L, Kast WM. The Notch ligand Jagged-1 is able to induce maturation of monocyte-derived human dendritic cells. *J Immunol.* 2002 Oct 15;169(8):4273-8. doi: 10.4049/jimmunol.169.8.4273. PMID: 12370358.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013 Apr 11;153(2):307-19. doi: 10.1016/j.cell.2013.03.035. PMID: 23582322; PMCID: PMC3653129.

Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SC. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell.* 1979 Apr;16(4):797-806. doi: 10.1016/0092-8674(79)90095-3. PMID: 455449.

Zaffran S, Kelly RG, Meilhac SM, Buckingham ME, Brown NA. Right ventricular myocardium derives from the anterior heart field. *Circ Res.* 2004 Aug 6;95(3):261-8. doi: 10.1161/01.RES.0000136815.73623.BE. Epub 2004 Jun 24. PMID: 15217909.

Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 2011 Aug;21(8):1273-83. doi: 10.1101/gr.122382.111. Epub 2011 Jun 1. PMID: 21632746; PMCID: PMC3149494.

Zhang Z, Huynh T, Baldini A. Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development.

Development. 2006 Sep;133(18):3587-95. doi: 10.1242/dev.02539. Epub 2006 Aug 16. PMID: 16914493; PMCID: PMC1850622.

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol.* 2011 Feb;29(2):149-53. doi: 10.1038/nbt.1775. Epub 2011 Jan 19. PMID: 21248753; PMCID: PMC3084533.

Zheng Q, Cai X, Tan MH, Schaffert S, Arnold CP, Gong X, Chen CZ, Huang S. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques.* 2014 Sep 1;57(3):115-24. doi: 10.2144/000114196. PMID: 25209046.

Zheng, X., Moriyama, E.N. Comparative studies of differential gene calling using RNA-Seq data. *BMC Bioinformatics* 14 (Suppl 13), S7 (2013). <https://doi.org/10.1186/1471-2105-14-S13-S7>