Discovering the downstream targets of KLF-4 and Estrogen Receptor in primary plaque cells using ChIP-seq

Author: Roxy Finger (3587969)

Degree: Drug Innovation (MSc)

Daily supervisor: Dr. E. Diez Benavente & N. A. M. van den Dungen

Examinator: Dr. M. Mokry

Date: 03-06-2022

Abstract

Aim: Atherosclerosis is a progressive disease where lipids and fibrous elements accumulate in the walls of the arteries, and this is the primary cause of heart disease and stroke worldwide. Research has shown sex differences in the disease development of atherosclerosis between males and females, with males presenting more unstable lesions while females more often present stable lesions. These sex differences are known and accepted, but the underlying pathobiology of these differences are still largely unknown. The transcription factors KLF-4 and ER were found to play an important role in the disease development of atherosclerosis, so in this study ChIP-seq was performed to unravel the downstream genetic regulation of these transcription factors in plaque primary cells (PCs).

Methods: Human Aortic Smooth Muscle cells (HASMCs) and smooth muscle cell-like primary plaque cells (PCs) were cultured and the ChIP methods CUT&RUN and MAGnify were performed on these cell types. The immunoprecipitated DNA was purified and library construction was performed for Next Generation Sequencing. On top of this, qPCR was performed on the immunoprecipitated DNA to validate if the ChIP enriched the DNA sequences that are associated with the target genes. Next Generation Sequencing data was visualized in the IGV browser, analyzed with Model-based Analysis of ChIP-seq, after which the peaks were annotated, and pathway motif enrichment analysis was performed to characterize the candidate target genes.

Results: In this study, we showed that KLF-4 has different candidate target genes in PCs than in HASMCs and the KLF-4 candidate target genes show associations with RNA processes, while more associations with protein modifications was seen in HASMCs. Overlapping the KLF-4 candidate target genes with literature, showed 50 overlapping genes who might play an important role in the disease progression. The KLF-4 binding motif was found to closely resemble the classic KLF-binding consensus sequence (GGGCGTGGC). Furthermore, the enrichment analyses for the H3K4me2 modification in both PCs and HASMCs showed similar genes and enrichment analysis revealed similar biological processes. On top of this, the smooth muscle marker genes MYH11 and KLF-4 were found to be active (as represented by H3K) in PCs and HASMCs but not in HEK293 cells. Lastly, most PC specific genes found in literature are found in PCs and not in HASMCs validating the cell type of the cultured PCs.

Conclusion: This study points to KLF-4 as an important player in the disease progression of atherosclerosis by acting as a repressor. Furthermore, H3K4me2 modification were found associated with target genes involved in the same processes in both PCs and HASMCs and the smooth muscle cell marker genes MYH11 and KLF-4 are active in both cell lines but not in HEK293 cells pointing to a smooth muscle cell origin of PCs. Lastly, fourteen out of the nineteen pre-defined PC specific genes were found to be candidate target genes of either KLF-4 or H3K4me2 in PCs, validating the cultured cell line. These findings can be used in further research to validate the role of KLF-4 in the disease development of atherosclerosis and the role of the found candidate target genes could be further uncovered.

Keywords: Atherosclerosis, KLF-4, ER, CUT&RUN, MAGnify, Next Generation Sequencing

Layman's Summary

Atherosclerosis is a disease where lipids and fibrous elements accumulate in the walls of the arteries forming plaques. This causes the arteries to narrow, and less oxygen-rich blood will be delivered to tissues. The diseases with atherosclerosis as underlying mechanism such as myocardial infarction and stroke, are the main cause of death worldwide. Research has shown that men and women have plaques with different characteristics, but the mechanisms leading to these differences are poorly understood. However, research has unraveled that the transcription factor KLF-4 plays an important role in atherosclerosis as it regulates the switching of the smooth muscle cells (SMCs) to synthetic smooth muscle cells. Furthermore, as sex-differences are seen in the phenotypes of the plaques, the estrogen receptor might be involved in modulating these differences as estrogen is more abundant in woman. Therefore, the genes that are regulated downstream of these transcription factors were studied in this research by using chromatin immunoprecipitation sequencing methods (ChIP-seq). Chromatin immunoprecipitation uses an antibody against a specific target protein that is bound to DNA, in this study we targeted transcription factors. After this the antibody-protein complex is cut together with the bound DNA. The DNA is released from the antibody-protein complex and purified. As the targets of our antibodies are transcription factors, the purified DNA will most likely be promoter regions of genes. This allows us to find genes that may play an important role in atherosclerosis and could be targeted in new treatments.

Two ChIP-seq methods have been conducted called Cleavage Under Targets & Release Using Nuclease (CUT&RUN) and MAGnify, of which the first one is a newly developed ChIP method. The MAGnify method is more established but most often used to study histone markers. Using both allows us to compare the results to find out which one is better suitable to study transcription factors. The methods have been conducted on smooth muscle cell-like primary plaque cells (PCs) and human aortic smooth muscle cells (HASMCs). After conducting the ChIP-method, the DNA was used for library construction and send for Next Generation Sequencing. On top of this as a validation, the immunoprecipitated DNA was used for qPCR in order to validate if the target DNA was enriched. The sequencing data was analyzed in R and the genes found associated with the transcription factors were compared to literature.

The sequencing data of the transcription factor KLF-4 showed different enriched genes in plaque cells than in human aortic smooth muscle cells, which might indicate that this TF might play a different role in these two cell-types during atherosclerosis progression. Furthermore, most peaks from the sequencing data were found on the promoter of the genes indicating that KLF-4 might act as a repressor of those genes. The histone marker H2K4me2 in the promoter of MYH11 is an indicator for smooth muscle cell origin and as the origin of the plaque cells is still up for debate, this histone marker was used to find out if the plaque cells have a smooth muscle cell origin. The results of H3K4me2 in both cell lines show overlapping genes associated with the histone marker and a H3K4me2 mark was found proximal to the smooth muscle marker genes MYH11 and KLF-4 in both the cell lines. Lastly, in order to further validate the expression of relevant genes in plaque cells, we compared our results with a set of genes that have found specifically expressed in these cells using RNAseq. These genes are also found in the cultured plaque cells for KLF-4 and H3K4me2 but not in the human aortic smooth muscle cells so this might indicate that these cell lines studied indeed present different expression profiles. The results of the ER transcription factor were suboptimal as not many peaks could be called, so no conclusions could be drawn.

These results can be used in further research to validate the repression role of KLF-4, optimize the CUT&RUN method to find associated genes for ER and study the role of the genes found in this study to further elucidate the disease progression of atherosclerosis.

Table of contents

1. Introduction

1.1 Background; Atherosclerosis

Atherosclerosis is a progressive disease where lipids and fibrous elements accumulate in the walls of the arteries and this is the primary cause of heart disease and stroke worldwide[1]. Research has shown there are sex differences in the disease development of atherosclerosis between males and females, with males presenting more unstable lesions while females more often present stable lesions[2,3]. These sex differences are known and accepted, but the underlying pathobiology of these differences is still largely unknown[4].

1.2 Background; Using plaque cells as *in vitro* models

For some time, histological and flow cytometry studies together with cell-specific markers were the basis of our understanding of the cellular composition of the atherosclerotic plaque but we still did not know anything about cell functions and state[5]. This changed when advancements in single-cell transcriptomics were made as it further unraveled plaque composition on a cellular level[6,7]. Animal models have given us the opportunity to examine the molecular interactions underlying atherosclerosis but it is still challenging to mimic the complex human atherosclerotic disease[4]. In cancer research, immunology and other diseases, the use of primary cells from diseased tissue is a widely accepted approach, but it is not used as much in atherosclerotic research[8,9]. Using these primary plaque cells in *in vitro* models will give us a platform to mimic atherosclerotic disease as the donor characteristics will be maintained by epigenetic and transcriptomic features and gives us a model system which resembles function and cell biology and can be combined with animal models resembling human pathology and gene expression profiles of cell-based clusters in order to find (novel) therapies to treat and diagnose patients at risk[5].

1.3 Role of KLF-4 in atherosclerosis

Studies have shown that the klüngel-like factor 4 (KLF4*)* regulates the transition of smooth muscle cells (SMCs) into synthetic SMCs and plaque cells have similarities and characteristics of these synthetic SMCs but it is still up for debate if the plaque cells are indeed synthetic SMCs[2,11,12,]. During this switching process, the regulation of KLF4 inhibits the myocardin gene activation in response to downregulation of SMC-marker genes and the SMCs are differentiated to synthetic SMCs[11,12]. Loss of myocardin upregulates a variety of inflammatory pathways and activates downstream differentiation[13]. Furthermore, T. N. et al., showed that KLF-4 acts as a repressor of mesenchymal genes in atherosclerosis. They showed the loss of epithelial gene expression, and mesenchymal genes are activated after downregulation of KLF-4 during epithelial-mesenchymal transition (EMT)[32]. Therefore, KLF-4 might play an important role in atherosclerosis.

1.4 Role of ER-α in atherosclerosis

Not only the transcription factor KLF4 plays an important role in the disease progression of atherosclerosis, estrogens also plays a significant role in heart disease especially in women[14,15]. Research has shown that estrogen may prevent the atherosclerotic process in women before menopause[14]. Cardiovascular disease and the associated morbidity and mortality are far less in pre-menopausal women and this advantage disappears with increased age and reduced estrogen levels after menopause. Estrogen interacts with the Estrogen Receptor (ER) in the nucleus and estrogens exert their effect through the action of the ERs which regulate the transcriptional processes[16]. Estrogen has the ability to increase the high-density lipoprotein and decrease the low-density lipoprotein as well as preventing oxidative modification of this lipoprotein thereby preventing the uptake of LDL by macrophages[17]. This indicates that the estrogen receptor may also plays a very interesting role in the disease progression of atherosclerosis and potentially dictating some of the differences between male and female plaque composition.

1.5 Finding functional genome regions to unravel the patho-biology of atherosclerosis

In this study, the transcription factors (TFs) KLF4 and ER-α together with histone markers (H3K4me3 and H3K27Ac) antibodies will be studied using a specific method of chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-seq) in smooth muscle cell like primary plaque cells (PCs) and Human Aortic Smooth Muscle cells (HASMCs) as *in vitro* models to find functional genomic regions and compare the epigenetics. These PCs are derived from male patients and the HASMCs are derived from female patients, giving us the opportunity to better understand the patho-biology of atherosclerosis. Finding functional genomic regions in the HASMCs will give us the opportunity to use these genomic regions as a baseline and compare them to the functional genomic regions of the PCs and find out if there are genomic regions that are differentially regulated by these TFs.

1.6 Chromatin Immunoprecipitation

ChIP detects protein-DNA interactions and histone-modification sites throughout the entire genome, which enables us to study the transcription regulation and chromatin organization on a genome-wide level[18]. We can therefore study where the active promotor sites are in the PCs and thus which genes are activated in a plaque. However, chromatin immunoprecipitation has been first described 35 years ago and the basic protocol has not changed since[19,20]. Even though the readout technologies have improved tremendously over time from southern blotting to high-throughput sequencing, ChIP-seq is still not challenge-free[20]. It suffers from high background noise, low signals and requires a large number of cells[20,21]. Alternative methods have been developed to overcome these challenges and one of these methods is Cleavage Under Targets & Release Using Nuclease (CUT&RUN).

1.7 ChIP methods

CUT&RUN uses the Laemmli's Chromatin ImmunoCleavage (ChIC) method[22]. This method takes place in permeabilized cells and consists of the binding of a specific antibody to the target chromatin protein after which a Protein A/Micrococcoal Nuclease (pA-MNase) fusion protein tethers to the antibody. The tethered nuclease will be activated with calcium to introduce double-stranded DNA breaks after which the fragments are released into the supernatant. The extracted DNA can then be used for library preparation and paired-end sequencing. The advantages of CUT&RUN over ChIP are that the cell input requirements are as low as 5.000 cells, diverse sample inputs and targets, a cost-effective workflow and a high signal-to-noise ratio[23]. As the CUT&RUN method is a novel ChIP-method, another more established method called MAGnify has also been performed to compare both methods with each other. MAGnify has been proven to work very well on histone markers and can be implemented with lower cell numbers than conventional ChIP methods.

1.8 Data analysis of sequencing data

After CUT&RUN, the extracted DNA fragments were sequenced using Next Generation Sequencing (NGS)[24]. This allows us to find the active promotor sites across the whole genome for the different target chromatin proteins, following these bioinformatic pipelines can be used to identify gene enrichments and compare these between cell types to find out if there is a different epigenetic landscape to gain more insight in the pathobiology of these cells in atherosclerosis. Furthermore, new genes might be discovered that are regulated by ER or KLF-4 and could play a significant role in the disease development of atherosclerosis.

2. Materials and methods

2.1 Cell culture

Pieces of atherosclerotic plaques from patients who underwent carotid endarterectomy were cultured in a conditional medium and migrated cells were used for the culture of these PCs. In this study, male and female atherosclerotic plaque cell lines were cultured but only the male cell line could be used for the experiments. The cells were cultured using complete medium containing Basal HAM's F12K Nut Mix 1x Medium, Heat Inactivated-FCS (HI-FCS), Penicillin/Streptomycin (P/S), HEPES, TES solution, Insulin Transferrin Sodium Selenite (ITS), Endothelial cell growth supplement (ECGS) solution and vitamin C. The flasks were coated with Fibronectin solution in PBS 1x (2 ug/cm^3) and accutase was used for the detachment of the cells. Female Human Aortic Smooth muscle cells (HASMCs) were also cultured in the same way as the PCs.

2.2 CUT&RUN

The first six CUT&RUN experiments have been performed using the CUTANA™ ChIC/CHUT&RUN Kit version 2 (Epicypher, SKU: 14-1048) and the last three experiments have been performed using the CUTANA™ ChIC/CHUT&RUN Kit version 3.0 (Epicyper, SKU: 14-1048). The experiments were conducted as stated in the manufacturer's manual. The positive H3K4me3 and negative rabbit IgG control antibodies were provided with the kit. The histone H3K27Ac Antibody (Epicypher, SKU: 13-0045, SNAP-ChIP Certified), Estrogen Receptor Alpha (C-Terminal) antibody (Epicypher, SKU: 13-2012), Myc antibody (Produced in our lab with an hybridoma cell line, 9E10), H3K4me2 (Sigma-aldrich, 07030) and KLF-4 (Sigma Aldrich, 3822775) were purchased separately. The Magnetic Separation Rack for the 0.2 mL Tubes was used when working with the magnetic beads (Epicypher, SKU: 10-0008). Before the start of a CUT&RUN experiment, cells were first stimulated with 100 mmol of Estradiol to active the Estrogen Receptor for four hours. After the stimulation the cells were detached using 0.05% of Trypsin. The Qubit® 3.0 Fluorometer (ThermoFisher Scientific, Q33216) together with the Qubit™ dsDNA HS Assay kit (ThermoFisher Scientific, Q32854) was used to quantify the DNA concentration in each sample.

2.3 MAGnify

The MAGnify™ Chromatin Immunoprecipitation System (ThermoFisher Scientific, 492024) was conducted as stated in the manufacturers manual. In short, the Dynabeads were incubated with the antibodies, while in the meantime crosslinking of the cells was performed with 1% formaldehyde. The reaction was stopped with 1.25 M glycine. Cells were lysed and sonicated with the S2 sonicator (Covaris) to shear the chromatin into 100-300 bp fragments following the program stated in the protocol. The microTUBE AFA Fiber Crimp-Cap 6x16mm (Covaris, SKU: 520052) was used with the sonicator. 10 µl of the diluted chromatin was set aside for the input control samples. The diluted chromatin and the Dynabeads were added together, and several washing steps were performed followed by protein digestion using proteinase K. The crosslinking was reversed at 65 °C and the samples were purified using the ChIP DNA Clean and Concentrator kit (Zymo Research, D5205). The DynaMag™-2 Magnet (ThermoFisher Scientific, 12321D) was used when working with the magnetic beads.

2.4 qPCR

2.4.1 Designing the primers

The primers were designed using the UCSC genome browser (**Table 1**)[25]. Active promotor regions in the HEK293 cell lines were found looking at the promotor regions and the H3K27Ac track. If there was a peak in the H3K27Ac track at the same place of the promoter region, it was assumed that this was an active promoter region of a randomly selected gene. The negative region primers were designed outside of active promoter regions (**Table 1**). The Primer-Blast tool from the NCBI was used to design the primers and the UCSC In-Silico PCR tool was used to validate the primer pairs[26]. These primers were ordered from Integrated DNA Technologies.

Table 1. Designed primers for the qPCR targeting genes that are expressed in HASMCs. Six primers were designed in promoter regions called positive primers and six primers were designed outside of promoter regions called negative primers.

Primer	Region	Genes	Forward primer	Reverse primer
CCNA 1	Positive primer	CCNA	CTGCAGATATCCCGCATCCC	GGTACTGAAGTCCGGGAACC
CCNA 2	Positive primer	CCNA	CACCCTCCTGCAGATATCCC	GCTAGCATTGCAGCAGACG
RNF138 1	Positive primer	RNF138	TTCTACTGCCCCGTCTGTCA	CTAAGCGACAACCCGGCTC
RNF138 2	Positive primer	RNF138	GTCTGTCAGGAGGTGCTCAA	AGGGGGCGTCTACTCACAC
HCG11 1	Positive primer	HCG11	GTGACCACTCTGTCGCCATT	GTAGTACGGAGACCAACCGC
HCG11 2	Positive primer	HCG11	GGGCGAGACCTACCTAGTCC	GGTAGTACGGAGACCAACCG
Neg 1	Negative primer	Outside CCNA	AGGCATTCCCACAAAGGCTA	AGGCATTCCCACAAAGGCTA
Neg 1.1	Negative primer	Outside CCNA	TTATGGGTGTTGGCCTTTGC	TAGCCTTTGTGGGAATGCCT
Neg 2	Negative primer	Outside RNF138	ACTGCACTTTGGAGGGCATA	TCCCATTCCTCTACCTTTACCT
Neg 2.1	Negative primer	Outside RNF138	ATTACTGCACTTTGGAGGGCAT	AATGCTACGGTAGACTCAAGAAT
Neg 3	Negative primer	Outside HCG11	ATTGCCTCTCTCCTCATCTCC	GTCCTCTGCAAGGCCTCTAAAA
Neg 3.1	Negative primer	Outside HCG11	CCCTGTGAAAAGAGCAGAGGT	ATGAGGATAGAGTCCCACAGT

2.4.2 PCR to validate the primers

The primers were first validated on an 1.5% agarose-gel with 0.0001% GelRed (Biotium, 19G0326) using genomic DNA (gDNA), Phusion® High-Fidelity PCR Master Mix with HF Buffer (BioLabs, M0531S) and EvaGreen® Dye, 20X in Water (Biotium, 31000). A 100 bp – 4Kb FlashGel™ DNA Marker was used as ladder (Lonza, 50473). The PCR program used is stated in **Table 2.**

Table 2. The PCR program used to validate the primers.

Temperature (°C)	Time	# Cycles
95	3 minutes	1x
98	10 seconds	40 x
67	20 seconds	
72	20 seconds	
72	5 minutes	1x
55	5 seconds	1x
95	5 seconds	1x
4	Forever	1x

2.4.3 qPCR to find out if the CUT&RUN protocol worked

The qPCRs have been performed using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad, 1845097), Phusion® High-Fidelity PCR Master Mix with HF Buffer (BioLabs, M0531S) and EvaGreen® Dye, 20X in Water (Biotium, 31000). A 50 bp – 1.5 Kb FlashGel™ DNA Marker was used as ladder (Lonza, 57033). The same qPCR program was used for all conducted qPCRs. The qPCRs were performed in duplicate, and the means were analyzed (**Table 3**).

Table 3. The qPCR program used for all qPCRs.

Temperature (℃)	Time	# Cycles
95	3 minutes	1x
98	10 seconds	40 x
67	20 seconds	
72	20 seconds	
72	5 minutes	1x
55 -> 95	0.5 ℃ increase for 5 seconds	1x
4	Forever	1x

2.5 Library construction and Next Generation Sequencing

The immunoprecipitated DNA obtained from the CUT&RUN experiments were used for creating libraries which could then be sequenced. The libraries were created with the NEXTFLEX Rapid DNA-Seq Kit 2.0 (PerkinElmer, NOVA-5188-01) and the NEXTFLEX-HT Barcodes (PerkinElmer, NOVA-5188-12). The DNA in the libraries were visualized using the FlashGel™ DNA Cassette (Lonza, 57032) together with the Gel Loading Buffer II (Invitrogen, AM8546G) and the 50 bp – 1.5 Kb FlashGel™ DNA Marker as ladder (Lonza, 57033). The 2100 Bioanalyzer of Automated Droplet Generator (Agilent) has been used to visualize the fragment sizes in each sample. The Qubit® 3.0 Fluorometer (Thermo Fisher Scientific, Q33216) together with the Qubit™ dsDNA HS Assay kit (Thermo Fisher Scientific, Q32854) was used to quantify the DNA concentration of each library. The libraries were pooled and send to the Utrecht Sequencing Facility (USEQ). The first run was analyzed using the Illumina NextSeq500 platform with a 1 x 75 bp High Output Run type. The last two runs were analyzed using the Illumina NextSeq2000 with a 2 x 50 bp Run type.

2.6 Data analysis

The data analysis of the qPCR data has been analyzed using Microsoft Excel and R to create the fold enrichment plots and the boxplots. The mapped Next Generation Sequencing data against the human 19 genome has been first visualized using the Integrative Genomics Viewer. The data has been analyzed using Python for peak calling using Model-based Analysis of ChIP-seq (MACS) after which the peaks were annotated in R (Rstudio, version 1.3.1093) and the Enrichr tool has been used for enrichment analyses[27]. Motif enrichment analysis has been performed on the ChIP data of KLF-4[28].

3 Results

3.1 Only peaks could be called for the histone marker after conducting the CUT&RUN method on HEK293

As the CUT&RUN technology is a recently described new chromatin immunoprecipitation method, the kit had first to be validated to find out if the protocol works on the specific cell types, histone markers and transcription factors used during this study. To validate this, the CUT&RUN method was first performed on Human Embryonic Kidney 293 cells (HEK293) kindly provided by D. M. C. Kapteijn using 500.000 cells per experiment as these cells are easy to grow. The positive (H3K4me3), negative (rabbit IgG) control antibodies, H3K27Ac, ER and Myc were taken along, and the experiment has been replicated once. The number of reads for all samples was more than 13 million, but only peaks could be called for the positive control sample and the H3K27Ac antibody (**Table 4**). This might indicate that there was no enrichment for Myc and ER and the reads were spread over the genome.

Table 4. Number of reads and peaks found for the samples after conducting the CUT&RUN experiment. Only peaks could be called for the positive H3K4me3 antibody and H3K27Ac, even though the number of reads was in the same range for all antibodies.

Experiment	# Cells used	Sample	# Reads	# Peaks
CUT&RUN 1 on HEK293 cells	500.000/Antibody	H3K4me3	18674233	9102
		H3K27Ac	14290892	3748
		Myc	13787765	-
CUT&RUN 2 on HEK293 cells	500.000/Antibody	H3K4me3	22049917	13117
		H3K27Ac	17251311	-
		Myc	60157732	-
		ER	24467692	-

The pathway analyses do show similar RNA processes for both antibodies, which was expected as both the H3K4me3 and H3K27Ac histone markers are involved in transcription (**Figure 1**). As those results came back as expected, we determined that the CUT&RUN method worked on histone markers, but further optimization is still required in order to obtain data for the transcription factors.

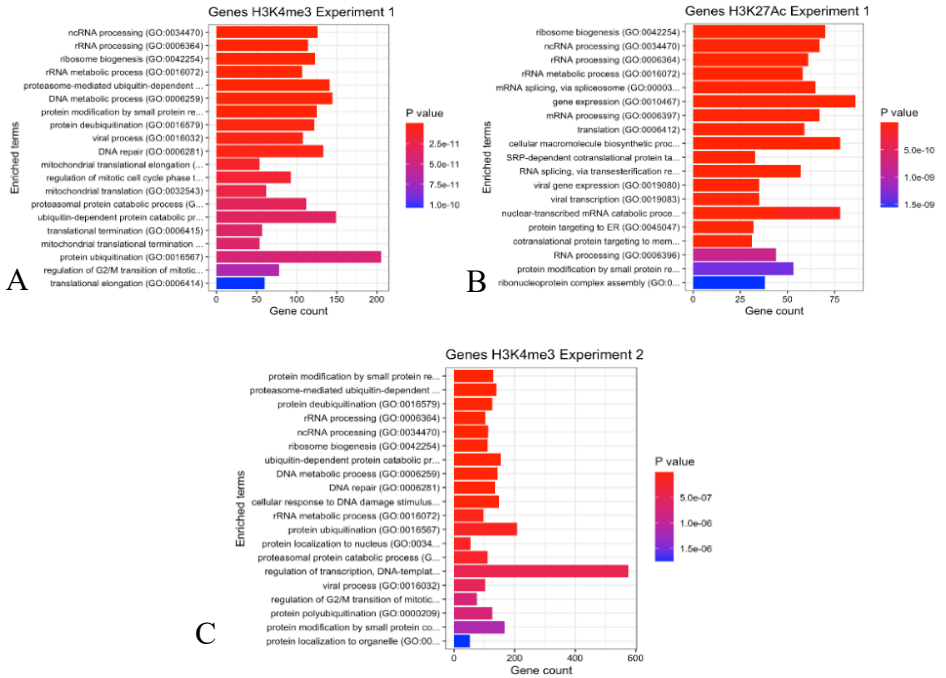

Figure 1. Pathway analysis show similar processes related to RNA for all three samples indicating that the CUT&RUN method worked for histone markers.
A) The positive control H3K4me3 antibody from the first experiment shows general processes related to RNA as this histone marker is associated with active genes and promotes transcription.
B) The histone marker H3K27Ac from the first experiment shows processes related to RNA as this marker is associated with higher activation of transcription.
C) The positive control H3K4me3 from the second experiment shows similar pathways as the positive control from the second experiment.

3.2 12377 genes were found associated with H3K4me2, 1812 genes with KLF-4 and 10 genes with ER in PCs

The transcription factors have been taken along in two CUT&RUN experiments and three MAGnify experiments (**Table 5**). Samples with more than 1000 peaks were manually inspected in the IGV browser and for H3K4me2 and ER performed on HASMCs, the two samples were merged. For KLF-4 performed on HASMCs, the sample with 407 peaks from the first MAGnify experiment was used for merging even though it has less than 1000 peaks, as most peaks where on the promoter regions and the other KLF-4 sample performed on HASMCs had less peaks (**Table 5**). The samples of histone marker

H3K4me2 performed on PCs were also merged. No peaks could be called for ER and KLF-4 from the fourth CUT&RUN experiment performed on PCs and the IGV browser also show little enrichment for the FAS and ACTA2 gene (**Figure 2**). Therefore, the results of ER and KLF-4 on PCs were based on one experiment.

Table 5. Experiments performed and number of reads and peaks per sample. Duplicate samples from H3K4me2, KLF-4 and ER performed on HASMCs were merged as well as the duplicates from H3K4me2 performed on PCs.

Experiment	# Cells used	Sample	# Reads	# Peaks	Merged samples
CUT&RUN 3 on HASMCs	500.000/Antibody	H3K4me3	36735232	11757	
		H3K27Ac	40504131	54653	
		ER	68840496	23	
		KLF-4	82002664	4635	x
CUT&RUN 4 on PCs	500.000/Antibody	H3K4me3	47230253	3300	
		H3K27Ac	43092542	23643	
		H3K4me2	33635396	14282	x
		ER	39964468	-	
		KLF-4	30740128	-	
MAGnify 1 on HASMCs	1.59 million	H3K27Ac	110262484	104823	
		H3K4me2	35101224	40063	x
		ER	84948595	60883	x
		KLF-4	12864086	407	x
MAGnify 2 on HASMCs	771.500	H3K27Ac	51514122	75335	
		H3K4me2	39096697	3672	x
		ER	52098363	1250	x
		KLF-4	70166714	88	
MAGnify 3 on PCs	1.85 million	H3K27Ac	69202792	117095	
		H3K4me2	84611619	38368	x
		ER	63650677	525	
		KLF-4	82130749	10657	

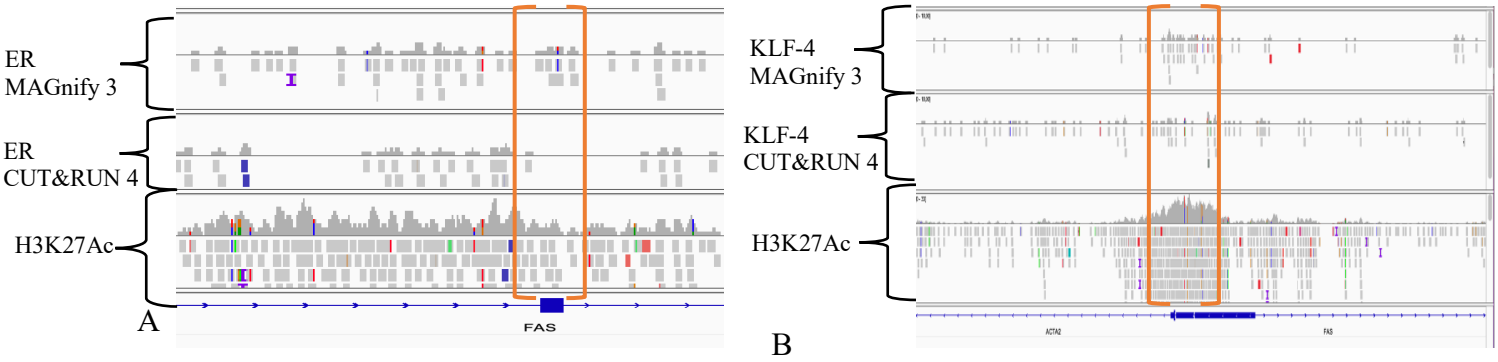

Figure 2. Comparison of the transcription factors from both ChIP methods on PCs in the IGV browser.
A). The ER antibody from the MAGnify experiment show more read alignments on the FAS gene than the ER antibody from the CUT&RUN experiment.
B) The KLF-4 antibody from the MAGnify experiment show more read alignments on the ACTA2 gene than the KLF-4 antibody from the CUT&RUN experiment.

For validating the peaks, (merged) peaks were overlapped with the peaks of the H3K27Ac sample to find the peaks on promoter regions. For the different cell types, the obtained data from the H3K27Ac antibody was visualized and the sample with the best read alignments has been chosen to validate the peaks (**Figure 3**). For the HASMC cell line, the H3K27Ac sample of the first MAGnify experiment was chosen as it showed the most enrichment on the promoter region of KLF-4 (**Figure 3A**) and for the PCs,

the H3K27Ac sample of the third MAGnify experiment was chosen as it showed to most enrichment on the promoter region of GAPDH (**Figure 3B**).

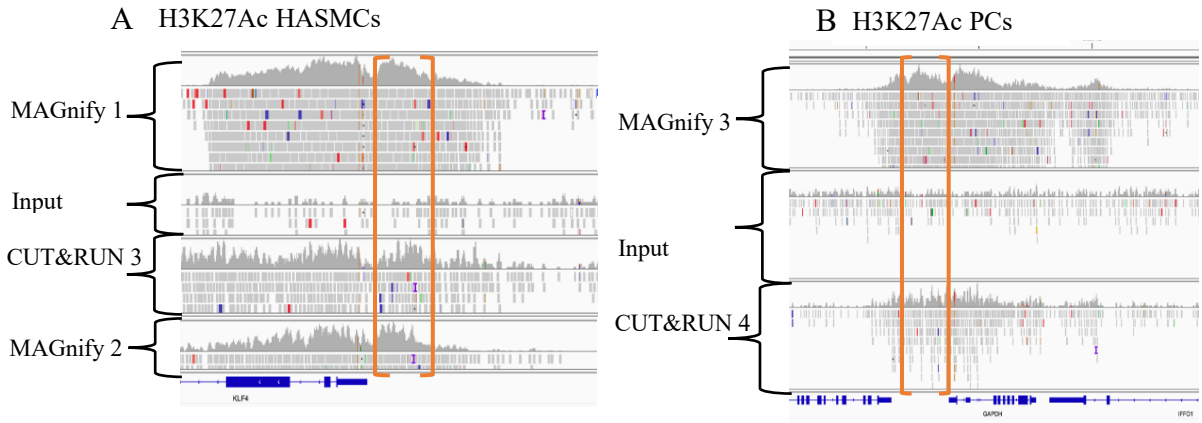

Figure 3. Comparison of the sequencing data of the H3K27Ac samples from the different experiments.
A) The first MAGnify expirement showed the most enrichment for the H3K27Ac sample in HASMCs on the promoter region of KLF-4.
B) The third MAGnify experiment showed the most enrichment for the H3K27Ac sample in PCs on the promoter region of GAPDH.

3.3 Enrichment analyses show similar processes associated with the H3K4me2 for both cell lines

After peak calling, the peaks were annotated, and enrichment analysis has been performed to find out in what processes the genes are involved in that are regulated by the transcription factors or histone marker. Most genes were found for the H3K4me2 histone marker in both cell types and the least genes were found for the ER transcription factor in both cell types (**Table 6).** This was expected as histone markers have more active epigenetic regions than transcription factors. As the ER transcription factor is only associated with 10 genes, the experiment should be repeated for ER as no conclusion can be drawn.

Table 6. Peaks called for H3K4me2, ER and KLF-4 after overlapping the peaks with the peaks of H3K27Ac in order to find the peaks on the promoter regions. The most genes were found for the histone marker H3K4me2 in both cell types and the least genes were found for ER in both cell types.

Sample	# Peaks called	# Genes called
ER in PCs	525	10
KLF-4 in PCs	10657	1812
H3K4me2 in PCs	16751	12377
ER in HASMCs	6235	301
KLF-4 in HASMCs	1566	1429
H3K4me2 in HASMCs	234589	13362

Comparing the processes of the ER regulated genes between the two cell types shows that the ER regulated genes in PCs are mostly involved in immune system processes, such as the regulation of T-cells and leukocytes (**Figure 4A**). The ER regulated genes in HASMCs also show processes involved in the immune system, such as the bone morphogenetic protein (BMP) signaling pathway but are also involved in cytosolic transport (**Figure 4B**). However, as only ten genes were called for the PCs, the enrichment analysis might not show a true representation of the cell line, but the associated processes are in line with the HASMCs. Looking at the genes involved in each process in PCs, the gene PDCD1LG2 is involved in the three processes negative regulation of activated T cell proliferation, interleukin-10 production, and interferon-gamma production. As all three processes are negatively regulated, it might indicate that this gene inhibits the immune system processes. However, the overlapping H3K27Ac track shows little read alignments for this gene so this gene might not be that strongly expressed (**Figure 4C**).

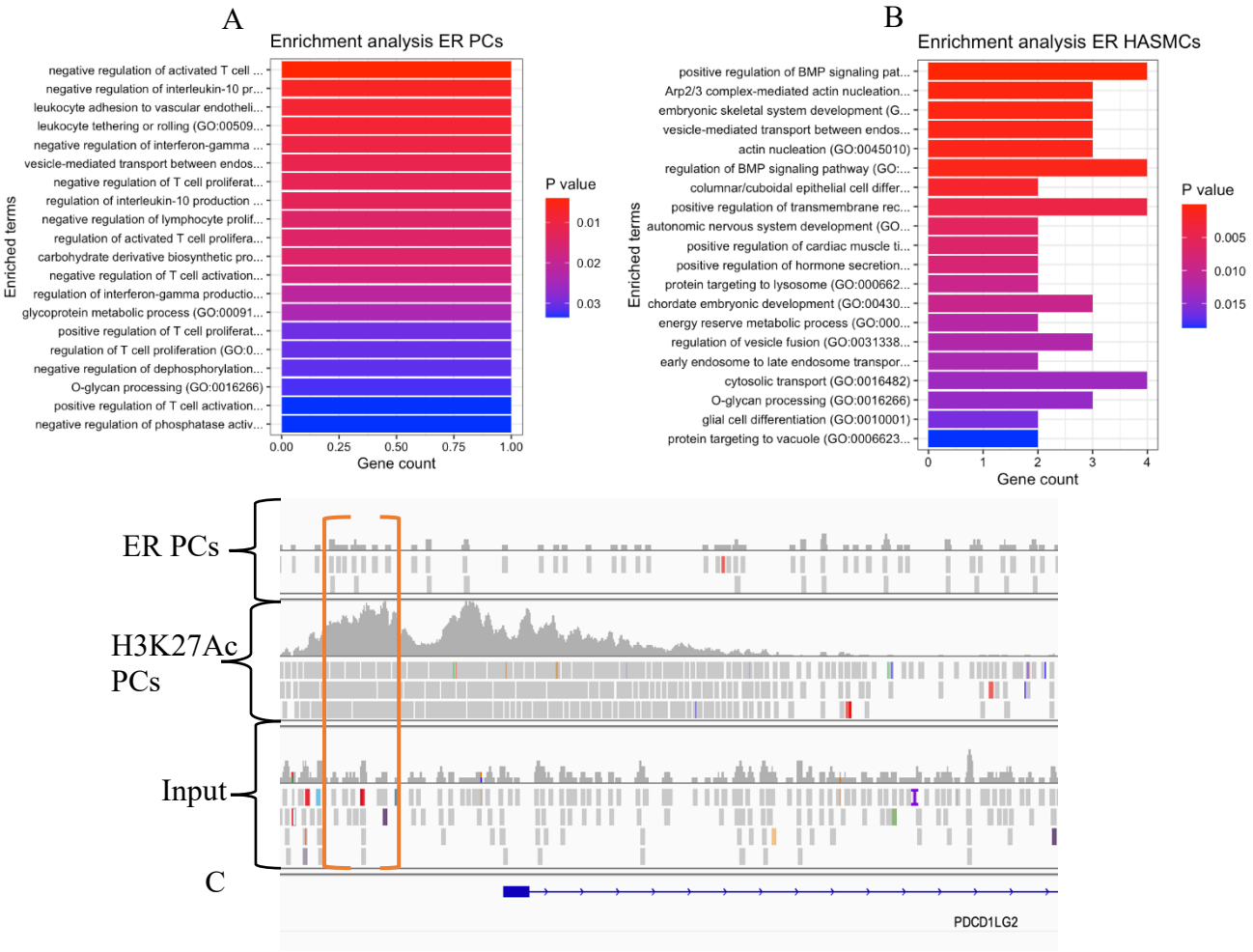

Figure 4. Enrichment analyses and involved gene for the ER antibody.
A) The ER regulated genes in PCs are mostly involved in immune system processes, such as the regulation of T cells and leukocytes.
B) The ER regulated genes in HASMCs also show processes involved in the immune system, such as the bone morphogenetic protein (BMP) signaling pathway but are also involved in cytosolic transport.
C) The gene PDCD1LG2 is involved in the three processes negative regulation of activated T cell proliferation, interleukin-10 production and interferon-gamma production, but little read alignments were found on the promoter region of this gene so this gene might not be that strongly regulated by ER.

The enrichment graphs of the KLF-4 antibody show abundant RNA related processes for the genes active in PCs and most genes are involved in regulation of transcription from RNA polymerase (**Figure 5A**). The KLF-4 regulated genes in HASMCs show more association with DNA-templated transcription, protein complex assembly and vesicle-mediated transport (**Figure 5B**). Furthermore, a paper by G.F. Alencar, et al[29]. found 336 differentially expressed genes in SMC-specific KLF-4 knockout versus WT mice in scRNA-seq data. In this dataset, 1750 genes were exclusively targeted in PCs by KLF-4 and even though the differentially expressed genes were found in mice, 50 of the 336 differently expressed genes were found in the KLF-4 candidate genes in PCs of this dataset. Two of these differently expressed genes are shown in the IGV browser and the overlapping H3K27Ac track showed that these genes are active in PCs (**Figure 5C, D**). Lastly, a search for the binding motifs from the KLF-4 peaks showed a binding motif closely resembling the classic KLF-binding consensus sequence (GGGCGTGGC) for KLF-4 and it was one of the most enriched binding motifs (**Figure 5E**).

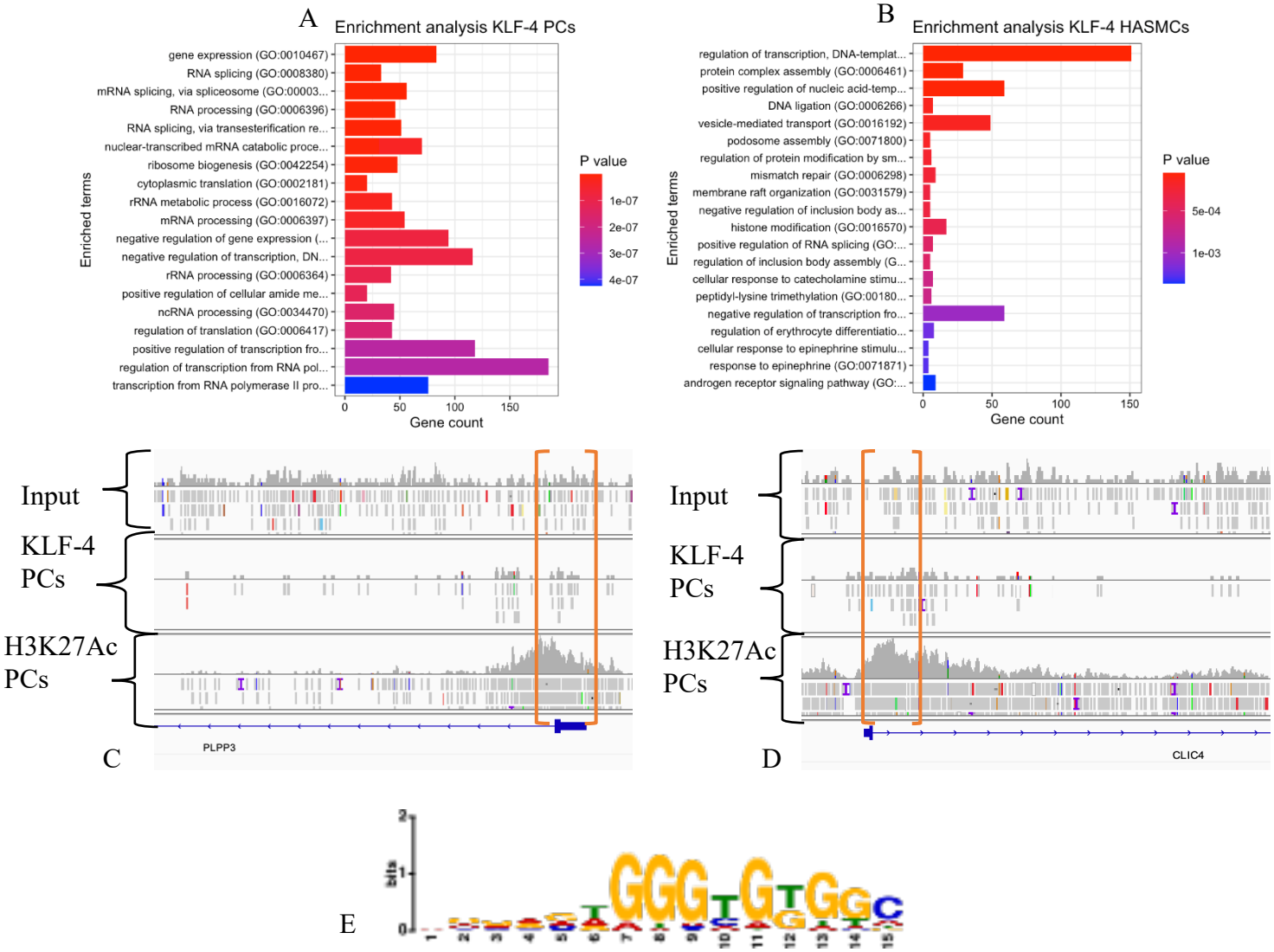

Figure 5. Enrichment analyses and involved genes for the KLF-4 antibody.
A). The KLF-4 antibody showed abundant RNA related processes for the genes active in PCs and most genes were involved in regulation of transcription from RNA polymerase.
B) The KLF-4 regulated genes in HASMCs showed more association with DNA-templated transcription, protein complex assembly and vesicle-mediated transport.
C) The gene PLPP3 was only regulated by KLF-4 in PCs and the H3K27Ac track showed that this gene was active.
D) The gene CLIC4 was only regulated by KLF-4 in PCs and the H3K27Ac track showed that this gene was active.
E) KLF-4 peaks showed a binding motif resembling closely resembling the classic KLF-binding consensus sequence (GGGCGTGGC) for KLF-4 and was one of the most enriched binding motifs.

The H3K4me2 histone marker in PCs shows an association with genes involved in regulation of gene expression and transcription, cellular protein modification and protein phosphorylation which is expected as the antibody regulates modifications to histone proteins (**Figure 6A**). The genes in HASMCs show similar associations involved in, regulation of transcription and gene expression, cellular protein modification and protein phosphorylation (**Figure 6B**). Furthermore, the same genes drive the processes associated with the H3K4me2 histone marker in both cell lines, e.g., the gene RBPJ is involved in DNA-methylated regulation of transcription and negative regulation of cellular macromolecule biosynthesis processes in PCs (**Figure 6C**) and the gene is involved in the same processes and in negative regulation of gene expression in HASMCs (**Figure 6D**). The gene GBP1 is involved in DNA-methylated regulation of transcription and negative regulation of cellular macromolecule biosynthesis processes in PCs, which are the same processes seen in HASMCs together with regulation of nucleic acid-templated transcription. For both genes the read alignments for the overlapping H3K27Ac target shows more reads for the H3K4me2 target in PCs than in HASMCs, which might indicate that the genes are more expressed in PCs than in HASMCs.

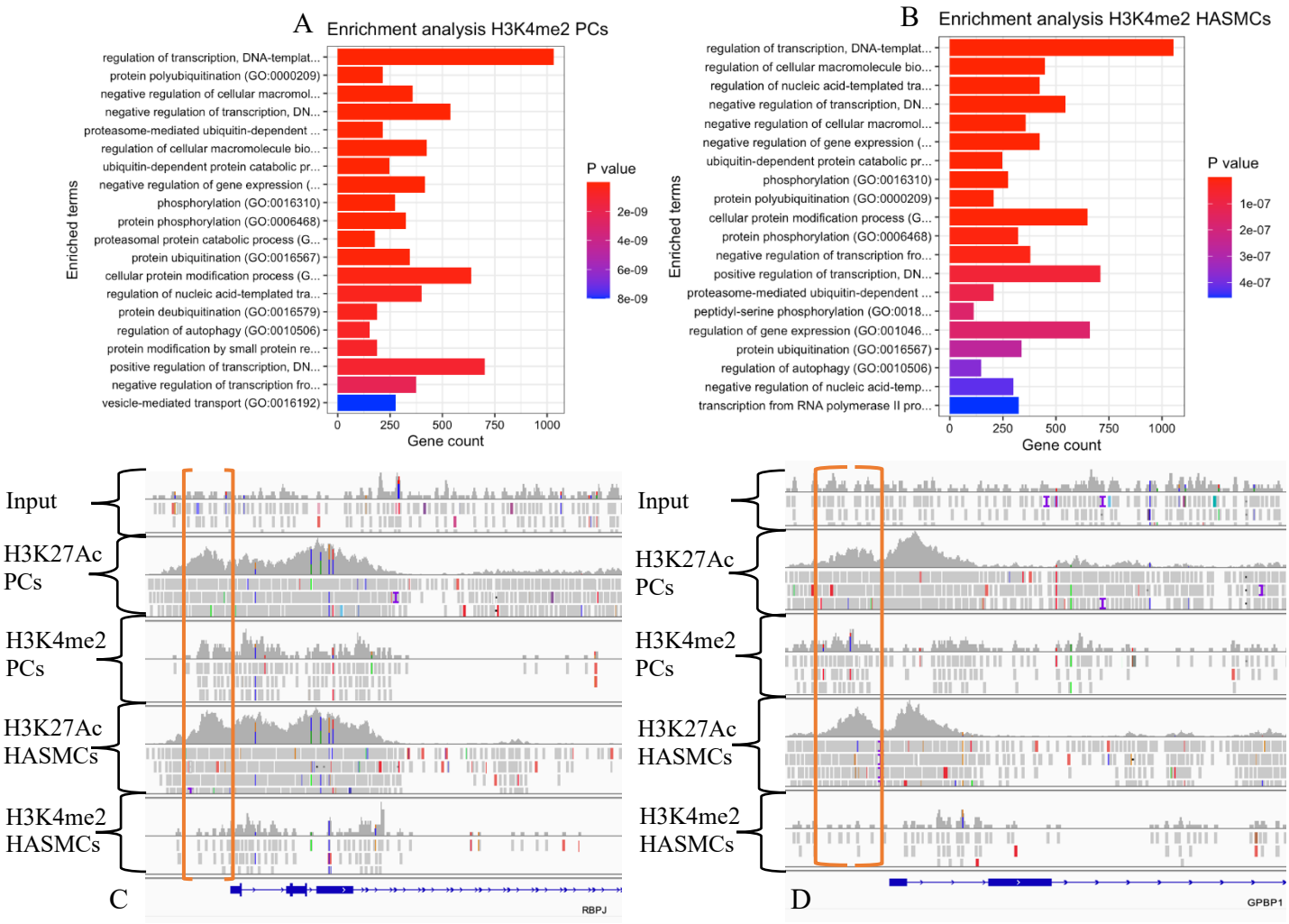

Figure 6. Enrichment analyses and involved genes for the H3K4me2 antibody.
A) The H3K4me2 histone marker in PCs shows an association with genes involved in regulation of gene expression and transcription, cellular protein modification and protein phosphorylation which is expected as the antibody regulates modifications to histone proteins.
B) The genes in HASMCs show similar associations as in PCs such as, regulation of transcription and gene expression, cellular protein modification and protein phosphorylation.
C) The gene RBPJ is involved in DNA-methylated regulation of transcription and negative regulation of cellular macromolecule biosynthesis processes in PCs and the gene is involved in the same processes and in negative regulation of gene expression in HASMCs.
D) The gene GBP1 is involved in DNA-methylated regulation of transcription and negative regulation of cellular macromolecule biosynthesis processes in PCs, which are the same processes seen in HASMCs together with regulation of nucleic acid-templated transcription.

3.4 The smooth muscle marker genes MYH11 and KLF-4 are active in PCs and HASMCs but not in HEK293 cells

The gene MYH11 is a well-known gene associated with smooth muscle cells and to further evaluate if PCs have a smooth muscle origin, the epigenetic regulation of this gene was studied in PCs, HASMCs and HEK293 cells. If this gene is found actively regulated (including marks of H3K4me2 in their promoter) in PCs and HASMCs but not in HEK293 cells, that will be a strong indication of the smooth muscle cell origin of the PCs. Therefore, the read alignments were visualized in the IGV browser and compared to the H3K27Ac target. As expected, the H3K27Ac track showed enrichment in both the PCs and HASMCs (**Figure 7A, B**). In the PCs showed H3K4me2 and ER some read alignments overlapping with the alignments of the H3K27Ac track indicating that these two might target MYH11, although very little reads were found for the ER track (**Figure 7A**). In HASMCs, both H3K27Ac and H3K4me2 target MYH11 indicating that MYH11 is expressed in the cultured HASMCs but not targeted by the transcription factors (**Figure 7C**). In HEK293 cells showed neither the positive control nor the H3K27Ac track enrichment in the promotor region of MYH11, indicating that this gene is not expressed in HEK293 cells.

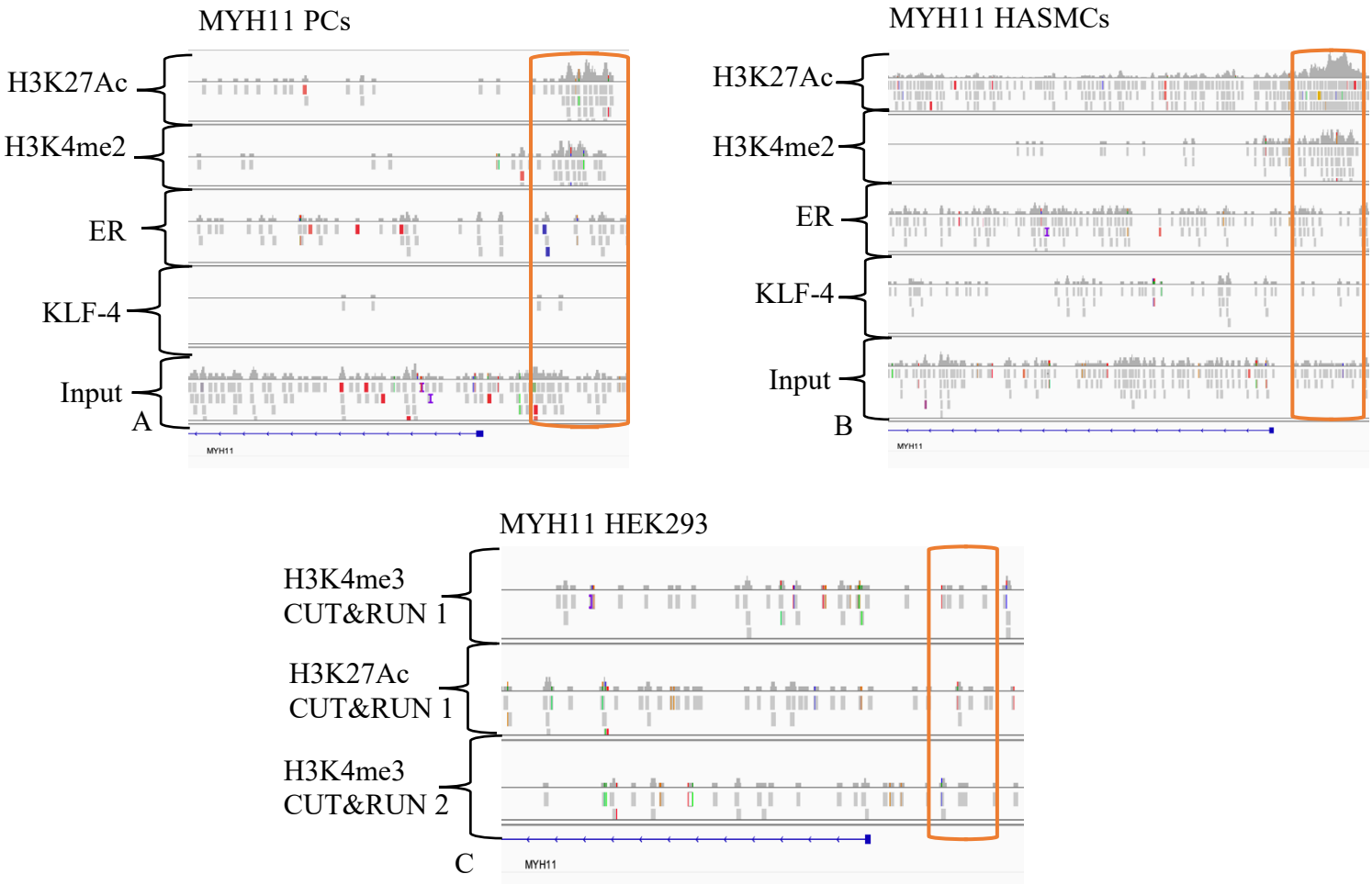

Figure 7. Comparison of the SMC-specific MYH11 gene expression in the three cell lines looking at H3K4me2 peaks.
A) In the PCs showed H3K4me2 and ER some read alignments overlapping with the alignments of the H3K27Ac track indicating that these two might target MYH11, although very little reads were found for the ER track.
B) In HASMCs, both H3K27Ac and H3K4me2 target MYH11 indicating that MYH11 is expressed in the cultured HASMCs but not targeted by the transcription factors.
C) In HEK293 cells showed neither the positive control nor the H3K27Ac track enrichment in the promotor region of MYH11, indicating that this gene is not expressed in HEK293 cells.

KLF-4 is another marker gene that plays a role in the regulation of smooth muscle cells genes is often expressed in smooth muscle cells and the visualization of the expression of this gene showed enrichments in the H3K27Ac track for both PCs and HASMCs but not for HEK293 cells (**Figure 8**). H3K4me2 also showed an enrichment in read alignments in the promoter region of KLF-4, indicating that this histone marker regulates KLF-4 in PCs. For the ER and KLF-4 track, no enrichment is seen in read alignments in the promoter region. (**Figure 8A**). In the HASMCs, enrichment in read alignments is also seen for H3K4me2, but not for ER and KLF-4 indicating that the transcription factors do not regulate KLF-4 (**Figure 8B**). For the HEK293 cells, both the positive control tracks showed some enrichment in read alignments for KLF-4 but no enrichment was found for H3K27Ac, indicating that the H3K4me3 histone code can be found in HEk293 cells, but KLF-4 is not active in HEK293 cells (**Figure 8C**).

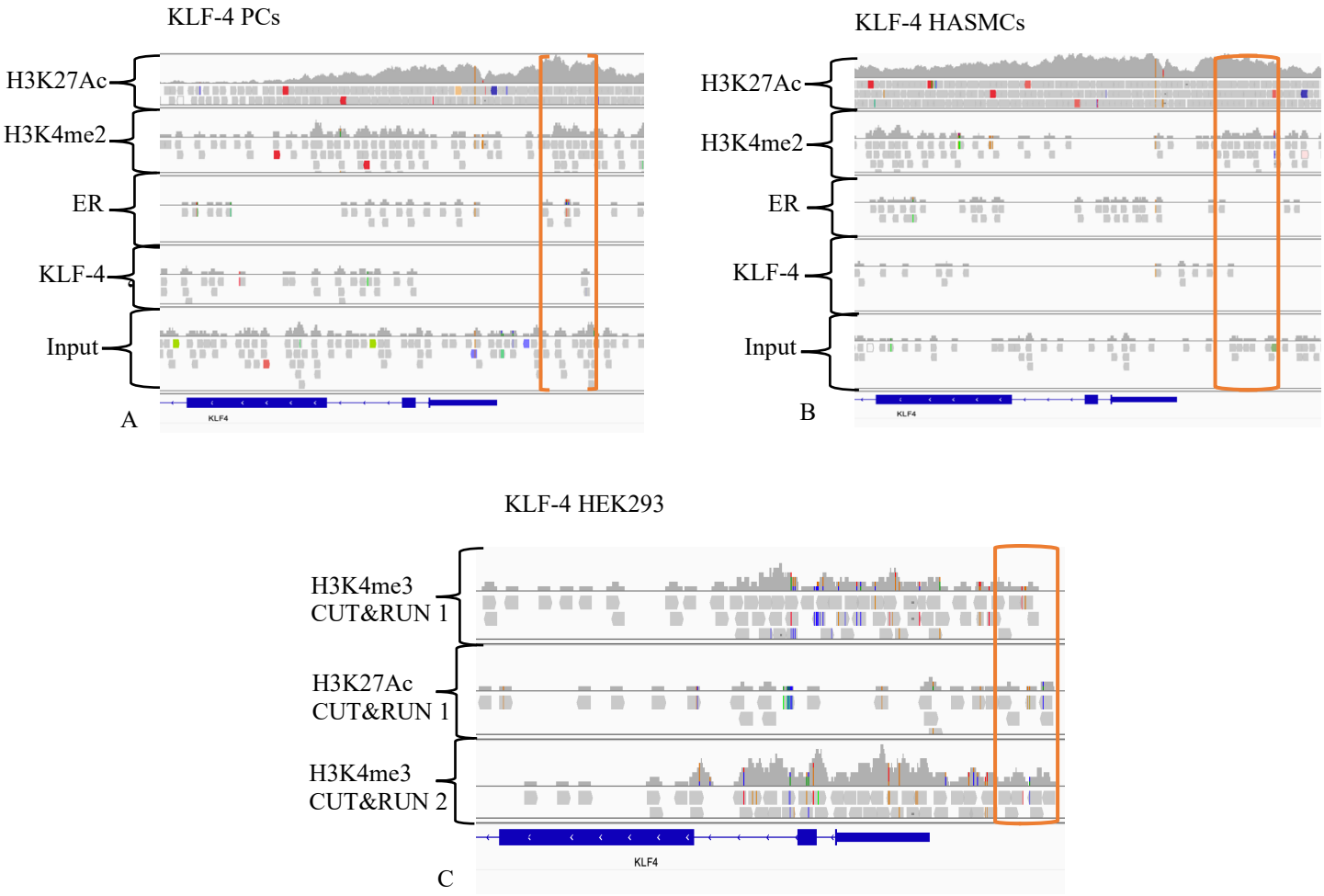

Figure 8. Comparison of the SMC marker gene KLF-4 in PCs, HASMCs and HEK293 cells.
A) H3K27Ac and H3K4me2 showed an enrichment in read alignments in the promoter region of KLF-4, indicating that these histone markers regulate KLF-4 in PCs. For the ER and KLF-4 track, no enrichment is seen in read alignments in the promoter region.
B) In the HASMCs, enrichment in read alignments is seen for H3K27Ac and H3K4me2, but not for ER and KLF-4 indicating that the transcription factors do not regulate KLF-4.
C) For the HEK293 cells, both the positive control tracks showed enrichment in read alignments for KLF-4 but no enrichment was found for H3K27Ac, indicating that the H3K4me3 histone code can be found in HEk293 cells, but KLF-4 is not active in HEK293 cells.

3.5 Pre-defined PC-specific genes by RNAseq were found in the cultured PCs and not in the HASMCs

To find out if there are genes that are specifically associated with the PC cell line, which might play a role in the disease progression of atherosclerosis, the candidate target genes found in PCs and HASMCs for the same transcription factor or histone marker were overlapped. For ER, the two genes EMBP1 and PDC1LG2 were only targeted in PCs and not in HASMCs by ER. For KLF-4, 1,751 genes were regulated by the transcription factor in PCs and 10,821 genes were regulated by the H3K4me2 histone marker in PCs and not in HASMCs. Furthermore, the expression of 19 PC-specific genes identified using RNAseq by E. Diez Benavente (data not provided) were overlapped with the associated genes in PCs to study if the cultured PC cell line are indeed PC cells (**Table 7**). ER only had two specific genes in PCs and none of the 19 PC-specific genes were found, but KLF-4 regulates fourteen out of the nineteen PC-specific genes and H3K4me2 regulated twelve out of the nineteen PC-specific genes, indicating that the cultured PC cell line are indeed PCs and that there is a different epigenetic landscape for PCs and HASMCs.

Table 7. Genes only expressed in PCs show overlap with known uniquely expressed genes in PCs.

Unique genes expressed in PCs	Genes associated with KLF-4 in PCs	Genes associated with H3K4me2 in PCs
FN1	FN1	**FN1**
COL1A1	**COL1A1**	**COL1A1**
POSTN	POSTN	**POSTN**
MTRNR2L12	MTRNR2L12	MTRNR2L12
VCAN	VCAN	**VCAN**
FP236383.3	FP236383.3	FP236383.3
THBS1	THBS1	THBS1
FBN1	**FBN1**	**FBN1**
NEAT1	NEAT1	**NEAT1**
EXT1	**EXT1**	**EXT1**
CR381670.1	CR381670.1	CR381670.1
SPARC	SPARC	**SPARC**
FNDC3B	FNDC3B	**FNDC3B**
ITGB1	ITGB1	**ITGB1**
COL4A1	**COL4A1**	**COL4A1**
CDH11	CDH11	**CDH11**
IGFBP7	IGFBP7	IGFBP7
ELN	ELN	**ELN**
COL5A2	COL5A2	**COL5A2**

Visualizing the genes in the IGV browser also shows the overlap of the read alignments from the PC-specific FBN1 and COL4A1 genes with the read alignments for the H3K27Ac (**Figure 9A, B**). This is not shown for the gene THBS1, which was discovered as PC-specific genes, but the enrichment analysis showed that this gene and four others are not targeted by KLF-4 or H3K4me2 in PCs (**Figure 9C**). This indicates that the four PC-specific genes targeted by KLF-4 might be active genes and could play a role in the disease development of atherosclerosis as they are only expressed in atherosclerotic plaque cells.

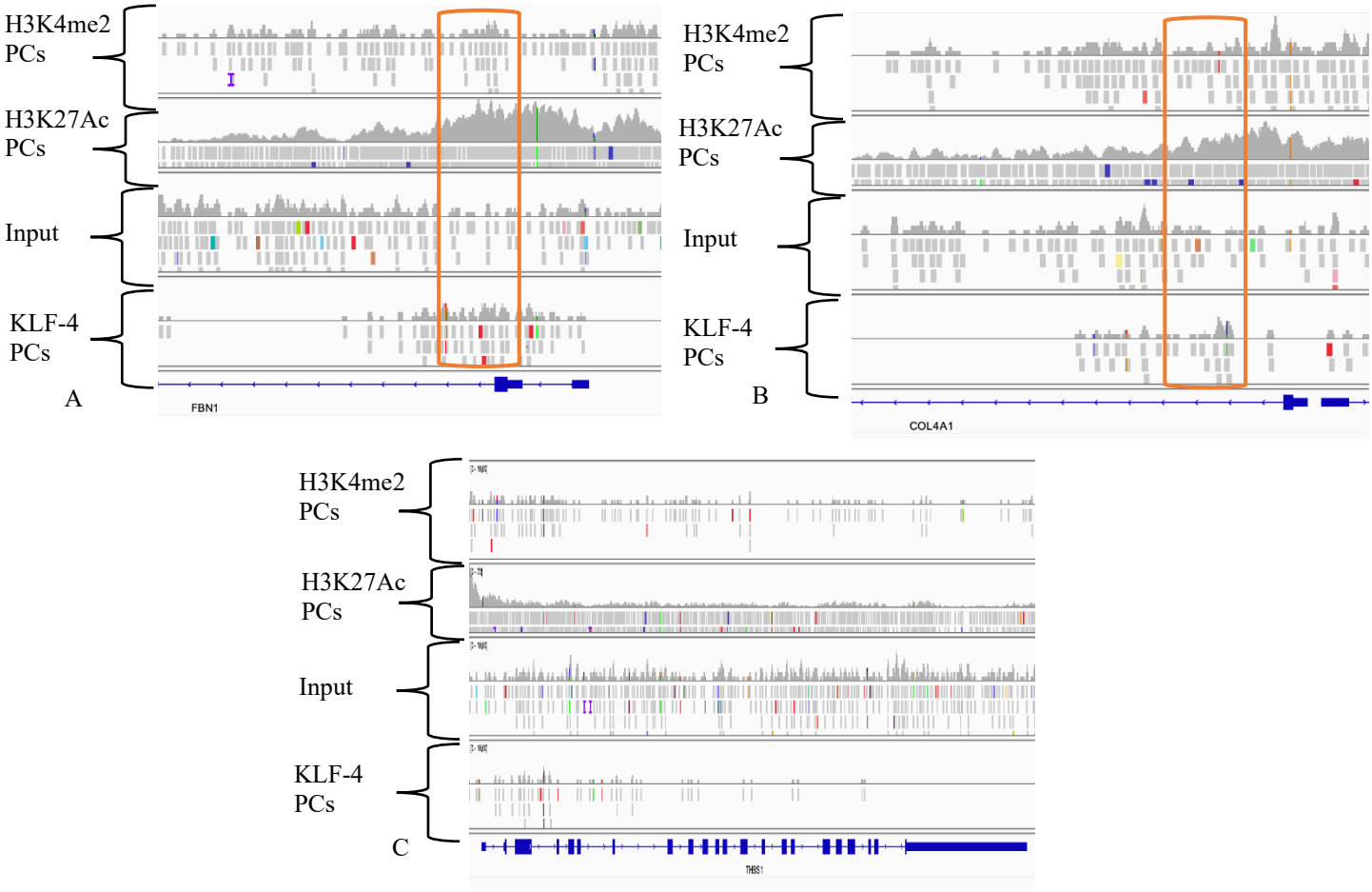

Figure 9. Visualization of two PC-specific genes associated with KLF-4 and H3K4me2.
A) The gene FBN1 shows read alignments that overlap with read alignments of H3K27Ac, indicating that this might be an expressed gene in PCs.
B) The gene COL4A1 shows read alignments that overlap with read alignments of H3K27Ac, indicating that this might be an expressed gene in PCs.
C) The gene THBS1 shows no read alignments that overlap with read alignments of H3K27Ac, indicating that this gene is not active and targeted by KLF-4 or H3K4me2.

4. Discussion

Atherosclerosis is a progressive disease where lipids and fibrous elements accumulate in the walls of the arteries and this is the primary cause of heart disease and stroke worldwide[1]. Research has shown there are sex differences in the disease development of atherosclerosis between males and females, with males presenting more unstable lesions while females more often present stable lesions[2,3]. These sex differences are known and accepted, but the underlying pathobiology of these differences is still largely unknown[4]. KLF-4 and ER were found to be transcription factors that might play an important role in the disease progression of atherosclerosis and therefore the downstream genetic targets of KLF-4 and ER were studied by performing ChIP-seq.

The results were derived from two different ChIP methods to compare to the methods with each other. The CUT&RUN method is recently developed ChIP method, which has the advantages of low cell input requirements, diverse sample inputs and targets, cost-effective workflow and a high signal-to-noise ratio[23]. The MAGnify method is an established ChIP method on histone markers and needs less cell input than other conventional methods. The MAGnify method gave better results for the transcription factors than the CUT&RUN method in this study, while other research showed good results for transcription factors using the CUT&RUN method so further research on ER could be performed after optimizing the method[30].

The sequencing results of the CUT&RUN method were not as conclusive as the MAGnify results but merging the two datasets still gave more peaks than only using the data from one of the methods. Optimizing the method will result in better results for the transcription factors but there was not enough time in this study to make the method work properly. On the other hand, the MAGnify method showed better results than expected as this method is only optimized for histone markers. The method has been performed on less cells than intended and even with less cells per experiment, the sequencing results from the MAGnify methods were more conclusive than the CUT&RUN results.

The results of KLF-4 from the MAGnify experiment shows candidate target genes that are all involved in RNA related processes and most peaks were found on transcription regions in the genome. Enhancer locations are often located at larger distances from the gene promoters, while for most peaks the distance to the transcription binding site (TSS) was quite close, indicating that KLF-4 might act as a repressor in PCs[31]. This is in line with literature, which has shown that KLF-4 acts as a repressor of mesenchymal genes. They showed that epithelial gene expression is lost, and mesenchymal genes are activated after downregulation of KLF-4 during epithelial-mesenchymal transition (EMT)[32]. These mesenchymal cells are one of the components of plaques and epithelial cells can lose their characteristics and gain mesenchymal characteristics[33]. In order to validate the findings, knock-down of KLF-4 can be performed to validate the repressor role of KLF-4 in atherosclerosis.

As G. F. Alencar, et al. have distinguished 336 differentially expressed genes upon KLF-4 knockdown and 50 of those 336 genes overlapped with the candidate target genes for KLF-4 in PCs. The peaks were further from the transcription start sites (TSS) in mice, while the peaks in our data were closer to the TSS, but the 50 genes could be conserved targets of KLF-4. These 50 genes can be taken along in further research to find genes that might play a vital role in the disease development of atherosclerosis. Furthermore, the found binding motive in KLF-4 was GGGTGTGGC which closely resembles the classic KLF-4 binding motif GGGCGTGGC where the C is either a T or a C. This indicates that KLF-4 binds to its classic binding motifs.

The enrichment analyses of ER showed that associated genes were predominantly involved in immune related processes in both PCs and HASMCs. The ER regulated genes in HASMCs also show processes involved in the immune system, such as the bone morphogenetic protein (BMP) signaling pathway but are also involved in cytosolic transport which might indicate that ER regulates calcification processes. However, as only ten genes were called for the PCs, the enrichment analysis might not show a true representation of the cell line, but the associated processes are in line with the HASMCs. Repeating this ChIP-seq method on ER, should give better results and give more insight in the role of ER in atherosclerosis.

The origin of PCs is still up for debate as the phenotype of SMCs in plaques show additional characteristics than SMCs in normal arteries and previous lineage tracking studies showed controversial results about the origin of the plaque SMCs[34,35]. As the characteristics of the plaque SMCs do not entirely overlap, some think that the plaque SMCs do not have a SMC lineage at all. Therefore, the SMC lineage was studied using the H3K4me2 histone marker as marked enrichment of H3K4me2 was found on SMC marker genes[36]. The enrichment analyses of H3K4me2 show similar candidate target genes and processes in both PCs and HASMCs. The candidate target genes are involved in regulation of gene expression and transcription, cellular protein modification and protein phosphorylation in both cell lines. Furthermore, the epigenetic marks around the promoter of the smooth muscle cell marker gene MYH11[37] is also seen in PCs and HASMCs but not in HEK293 cells. These results might give an indication of the smooth muscle cell origin of PCs, although further research should be done to confirm the origin of these cells.

PC specific genes found in literature were overlapped with the candidate gene targets and fourteen genes overlapped for H3K4me2, and four genes overlapped for KLF-4. As these genes were found in the cultured PCs, there can be concluded that these cells are indeed PCs and the found candidate target genes can be used in further research to investigate their roles in atherosclerosis.

This study is not without limitations as not all the PC cell lines grew to enough cell numbers in order to perform the experiments, so sex-specific differences in the epigenetic landscape could not be studied. However, as all cells were stimulated with estradiol before harvesting, the difference found should not be sex-related. Furthermore, as the CUT&RUN method could not be optimized for transcription factors, the experiments could not be repeated more than two times, therefore not enough data was generated and limiting the interpretation of the results. Nonetheless, this study points to KLF-4 as an important player in the disease progression of atherosclerosis by regulating gene expression in plaque cells. Furthermore, the H3K4me2 candidate target genes are involved in the same processes in both PCs and HASMCs and the smooth muscle cell marker genes MYH11 and KLF-4 are active in both cell lines but not in HEK293 cells pointing to a smooth muscle cell origin of PCs. Lastly, fourteen out of the nineteen pre-defined PC specific genes were found to be candidate target genes of either KLF-4 or H3K4me2 in PCs, validating the cultured cell line. These findings can be used in further research to validate the role of KLF-4 as repressor in the disease development of atherosclerosis and the role of the found candidate target genes could be further uncovered.

5. Conclusion

In this study, we showed that KLF-4 has a different epigenetic landscape in PCs than in HASMCs and points to KLF-4 as an important player in the disease progression of atherosclerosis by acting as a repressor. On the other hand, the H3K4me2 mark in the MYH11 promoter region in PCs points towards the smooth muscle origin of PCs. On top of this, most PC specific genes found in literature are found in PCs and not in HASMCs validating the cell type of the cultured PCs. Lastly, the ChIP methods conducted in this study can be both used for transcription factors, but the CUT&RUN method must be further adapted to obtain better results as it is a newly developed method. These findings can be used in further research to unravel the role of the transcription factors in the pathobiology of atherosclerosis.

6. References

1. Lusis AJ. Atherosclerosis. *Nature*. 2000;407(6801):233. doi:10.1038/35025203
2. Hartman RJG, Owsiany K, Ma L, et al. Sex-Stratified Gene Regulatory Networks Reveal Female Key Driver Genes of Atherosclerosis Involved in Smooth Muscle Cell Phenotype Switching. *Circulation*. 2021;143(7):713-726. doi:10.1161/CIRCULATIONAHA.120.051231
3. Hartman RJG, Siemelink MA, Haitjema S, et al. Sex-dependent gene regulation of human atherosclerotic plaques by DNA methylation and transcriptome integration points to smooth muscle cell involvement in women. *bioRxiv*. January 2021:2021.01.28.428414. doi:10.1101/2021.01.28.428414
4. Bennett BJ, Davis RC, Civelek M, et al. Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. Barsh GS, ed. *PLOS Genet*. 2015;11(12):e1005711. doi:10.1371/journal.pgen.1005711
5. Buono MF, Slenders L, Wesseling M, et al. The changing landscape of the vulnerable plaque: a call for fine-tuning of preclinical models. *Vascul Pharmacol*. 2021;141:106924. doi:10.1016/j.vph.2021.106924
6. Wirka RC, Wagh D, Paik DT, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med*. 2019;25(8):1280-1289. doi:10.1038/s41591-019-0512-5
7. Butcher MJ, Filipowicz AR, Waseem TC, et al. Atherosclerosis-Driven Treg Plasticity Results in Formation of a Dysfunctional Subset of Plastic IFNγ + Th1/Tregs. *Circ Res*. 2016;119(11):1190-1203. doi:10.1161/CIRCRESAHA.116.309764
8. Monaco C, Gregan SM, Navin TJ, Foxwell BMJ, Davies AH, Feldmann M. Toll-Like Receptor-2 Mediates Inflammation and Matrix Degradation in Human Atherosclerosis. *Circulation*. 2009;120(24):2462-2469. doi:10.1161/CIRCULATIONAHA.109.851881
9. Monaco C, Andreakos E, Kiriakidis S, et al. Canonical pathway of nuclear factor B activation

selectively regulates proinflammatory and prothrombotic responses in human atherosclerosis. *Proc Natl Acad Sci*. 2004;101(15):5634-5639. doi:10.1073/pnas.0401060101
10. van Keulen D, van Koeverden ID, Boltjes A, et al. Common Variants Associated With OSMR Expression Contribute to Carotid Plaque Vulnerability, but Not to Cardiovascular Disease in Humans. *Front Cardiovasc Med*. 2021;8:658915. doi:10.3389/fcvm.2021.658915
11. Alencar GF, Owsiany KM, Karnewar S, et al. Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. *Circulation*. 2020;142(21):2045-2059. doi:10.1161/CIRCULATIONAHA.120.046672
12. Alexander MR, Owens GK. Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease. *Annu Rev Physiol*. 2012;74:13-40. doi:10.1146/annurev-physiol-012110-142315
13. Bennett MR, Sinha S, Owens GK. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res*. 2016;118(4):692-702. doi:10.1161/CIRCRESAHA.115.306361
14. Xing D, Nozell S, Chen Y-F, Hage F, Oparil S. Estrogen and Mechanisms of Vascular Protection. *Arterioscler Thromb Vasc Biol*. 2009;29(3):289-295. doi:10.1161/ATVBAHA.108.182279
15. Kavanagh K, Davis MA, Zhang L, et al. Estrogen Decreases Atherosclerosis in Part by Reducing Hepatic Acyl-CoA:Cholesterol Acyltransferase 2 (ACAT2) in Monkeys. *Arterioscler Thromb Vasc Biol*. 2009;29(10):1471-1477. doi:10.1161/ATVBAHA.109.191825
16. Björnström L, Sjöberg M. Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes. *Mol Endocrinol*. 2005;19(4):833-842. doi:10.1210/me.2004-0486
17. Nathan and L, Chaudhuri G. ESTROGENS AND ATHEROSCLEROSIS. *Annu Rev Pharmacol Toxicol*. 1997;37(1):477-515. doi:10.1146/annurev.pharmtox.37.1.477
18. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform*. 2016;18(2):bbw023. doi:10.1093/bib/bbw023
19. Solomon MJ, Varshavsky A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci*. 1985;82(19):6470-6474. doi:10.1073/pnas.82.19.6470
20. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc*. 2020;15(10):3264-3283. doi:10.1038/s41596-020-0373-x
21. Kaya-Okur HS, Wu SJ, Codomo CA, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*. 2019;10(1):1930. doi:10.1038/s41467-019-09982-5
22. Schmid M, Durussel T, Laemmli UK. ChIC and ChEC. *Mol Cell*. 2004;16(1):147-157. doi:10.1016/j.molcel.2004.09.007
23. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*. 2017;6. doi:10.7554/eLife.21856
24. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child - Educ Pract Ed*. 2013;98(6):236-238. doi:10.1136/archdischild-2013-304340
25. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002;12(6):996-1006. doi:10.1101/gr.229102
26. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13(1):134. doi:10.1186/1471-2105-13-134
27. Kuleshov M V., Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90-W97. doi:10.1093/nar/gkw377
28. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*. 2010;11(1):165. doi:10.1186/1471-2105-11-165
29. Alencar GF, Owsiany KM, Karnewar S, et al. Stem Cell Pluripotency Genes Klf4 and Oct4

Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. *Circulation*. 2020;142(21):2045-2059. doi:10.1161/CIRCULATIONAHA.120.046672

30. Kong NR, Chai L, Tenen DG, Bassal MA. A modified CUT&RUN protocol and analysis pipeline to identify transcription factor binding sites in human cell lines. *STAR Protoc*. 2021;2(3):100750. doi:10.1016/j.xpro.2021.100750
31. Stadhouders R, van den Heuvel A, Kolovos P, et al. Transcription regulation by distal enhancers. *Transcription*. 2012;3(4):181-186. doi:10.4161/trns.20720
32. Tiwari N, Meyer-Schaller N, Arnold P, et al. Klf4 Is a Transcriptional Regulator of Genes Critical for EMT, Including Jnk1 (Mapk8). Katz E, ed. *PLoS One*. 2013;8(2):e57329. doi:10.1371/journal.pone.0057329
33. Wesseling M, Sakkers TR, de Jager SCA, Pasterkamp G, Goumans MJ. The morphological and molecular mechanisms of epithelial/endothelial-to-mesenchymal transition and its involvement in atherosclerosis. *Vascul Pharmacol*. 2018;106:1-8. doi:10.1016/j.vph.2018.02.006
34. Albarrán-Juárez J, Kaur H, Grimm M, Offermanns S, Wettschureck N. Lineage tracing of cells involved in atherosclerosis. *Atherosclerosis*. 2016;251:445-453. doi:10.1016/j.atherosclerosis.2016.06.012
35. Bentzon JF, Majesky MW. Lineage tracking of origin and fate of smooth muscle cells in atherosclerosis. *Cardiovasc Res*. 2018;114(4):492-500. doi:10.1093/cvr/cvx251
36. Gomez D, Swiatlowska P, Owens GK. Epigenetic Control of Smooth Muscle Cell Identity and Lineage Memory. *Arterioscler Thromb Vasc Biol*. 2015;35(12):2508-2516. doi:10.1161/ATVBAHA.115.305044
37. Wang C, Han M, Zhao X-M, Wen J-K. Kruppel-like Factor 4 is Required for the Expression of Vascular Smooth Muscle Cell Differentiation Marker Genes Induced by All-Trans Retinoic Acid. *J Biochem*. 2008;144(3):313-321. doi:10.1093/jb/mvn068