

Université de Montréal

Investigation des variants génétiques dans la dysfonction endothéliale et le risque de maladies
cardiovasculaires

par Valérie-Anne Codina-Fauteux

Département de biochimie, faculté de médecine

Mémoire présenté en vue de l'obtention du grade de M.Sc en biochimie option génomique humaine

Août 2018

© Valérie-Anne Codina-Fauteux, 2018

RÉSUMÉ FRANÇAIS

Les maladies cardiovasculaires (MCV) représentent un fléau mondial pour l'économie et la santé dont la prévalence ne cesse de croître. Au cours des dernières années, il y a eu un effort considérable de compréhension de l'aspect génétique de ces maladies avec les GWAS, associant ainsi près de 500 SNPs. Un de ceux-ci, rs9349379, dans l'intron de *PHACTRI* brise le motif de liaison du facteur MEF2 et représente un eQTL de sorte à ce qu'il soit considéré comme un variant possiblement causal. Alors que les pistes potentielles se multiplient, il demeure impossible d'inférer une causalité par le manque d'informations biologiques. Vis-à-vis cette problématique, il est possible de combiner des expériences de transcriptomique et conformation chromosomique pour avoir une concordance d'informations permettant de relever l'importance d'un locus dans son implication dans le risque des MCV. Des expériences de RACE et PacBio ont permis de valider 6 transcrits différents issus du gène *PHACTRI* : un long, quatre intermédiaires avec deux codons d'initiation différents et avec l'épissage alternatifs de l'exon interne 10.11 et un court. Par qPCR, il a été montré que rs9349379 présente un eQTL spécifique aux transcrits intermédiaires avec l'exon supplémentaire 10.11. Une priorisation des variants est également réalisée par la cumulation de données d'ATACseq, de RNAseq et de Hi-C dans un contexte de dysfonction endothéliale, soit un phénotype à la base de l'athérosclérose et caractérisant les MCV. Ainsi, quatre loci d'intérêt dans les MCV sont sélectionnés puis délétés par le système CRISPR-Cas9 et sont présentement en cours de caractérisation dans le laboratoire. Ces expériences ont permis de construire une cartographie exhaustive à l'échelle du génome sous un phénotype pathogénique et s'avère une mine d'informations pour la compréhension des mécanismes reliant les SNPs au risque des MCV.

Mots clés : Maladies cardiovasculaires (MCV/CVD), dysfonctions endothéliales, études d'association pangénomique (GWAS), Polymorphismes nucléotidiques (SNPs), phosphatase and actin regulator 1 (*PHACTRI*), transcriptome, épissage, loci d'expression quantitatifs (Expression quantitative trait loci; eQTLs), conformation de la chromatine, molécules d'adhésion.

RÉSUMÉ ANGLAIS

CVD is a global burden for the economy and health, whose prevalence continues to grow. In the last few years, there has been a considerable effort to understand the genetic aspect of those diseases with GWAS, associating nearly 500 SNPs. One of these, rs9349379, in the intron of *PHACTR1* breaks the MEF2 factor binding motif and is an eQTL. As potential leads multiply, it is impossible to pinpoint causality by the lack of biological information. With regard to this problem, it is possible to combine transcriptomic and chromosomal conformation experiments to obtain a concordance of information in order to explain SNP's involvement in the risk of CVD. RACE and PacBio experiments allowed to validate 6 different transcripts of the *PHACTR1* gene: one long, four intermediate with two different initiation codons and with the alternative splice of internal exon 10.11 and one short. By qPCR, it was shown that eQTL was specific to intermediate transcripts with the additional exon 10.11. Variant prioritization is also performed by aggregating ATACsec, RNAsec and Hi-C data in a context of endothelial dysfunction, a phenotype underlying atherosclerosis and characterizing CVD. Thus, four loci of interest in the CVD are selected and edited by the CRISPR-Cas9 system and are under process of characterization in the laboratory. These experiments have made it possible to build an exhaustive mapping at the genome scale under a pathogenic phenotype and provides a wealth of information for understanding the mechanisms linking SNPs to CVD risk.

Key words : Cardiovascular disease (MCV / CVD), endothelial dysfunction, genome-wide association studies (GWAS), single nucleotide polymorphisms (SNPs), phosphatase and actin regulator 1 (*PHACTR1*), transcriptome, splicing, expression quantitative trait loci (eQTLs), chromatin conformation, adhesion molecules.

Table des matières

RÉSUMÉ FRANÇAIS	II
RÉSUMÉ ANGLAIS	III
TABLE DES MATIÈRES	IV
LISTE DES TABLEAUX.....	VII
LISTE DES FIGURES.....	VIII
LISTE DES ANNEXES	IX
LISTE DES ABRÉVIATIONS	X
REMERCIEMENTS.....	XII
1 - Introduction	1
1.1 - Maladies cardiovasculaires (MCV).....	1
1.2 - Athérosclérose	2
1.3 - Facteurs de risque traditionnel.....	3
1.4 - L'endothélium et la dysfonction endothéliale	4
1.5 - Association génétique.....	8
1.6 - Le SNP rs9349379	13
1.7 - <i>PHACTR1</i>	15
1.7.1 Structure.....	15
1.7.2 Interaction avec PP1 et rôle du NO dans l'IM.....	16
1.7.3 Isoformes.....	17
1.7.4 Induction et répression de l'expression.....	18
1.8 - Ère des omics	19
1.8.1 Bases de données publiques.....	20
1.8.2 Transcriptome	21

1.8.3	Conformation chromosomique	21
1.9 -	Système d'édition du génome CRISPR-Cas9.....	22
1.10 -	Hypothèses et objectifs de recherche.....	24
2 -	Caractérisation des isoformes de PHACTR1	25
2.1 -	Methods	26
2.1.1	Cell lines and tissues.....	26
2.1.2	Rapid Amplification of cDNA Ends (RACE)	26
2.1.3	Long-read Pacific Biosciences (PacBio) sequencing library preparation.....	27
2.1.4	Reverse transcription polymerase chain reaction (RT-PCR).....	28
2.1.5	Quantitative polymerase chain reaction (qPCR).....	29
2.1.6	eQTL analyses using GTEx data	30
2.1.7	Statistical analyses	31
2.1.8	Western blot	31
2.2 -	Results.....	32
2.2.1	Six PHACTR1 transcripts in human cells and tissues	32
2.2.2	PHACTR1 transcripts are differentially expressed.....	34
2.2.3	CAD-associated rs9349379 is an eQTL for PHACTR1 intermediate transcripts A.....	36
2.3 -	Discussion.....	40
2.4 -	Conclusions.....	41
3 -	Inactivation de <i>PHACTR1</i> par le système CRISPR-Cas9	42
3.1 -	Méthodes.....	43
3.1.1	Culture cellulaire.....	43
3.1.2	Édition du génome par le système CRISPR-Cas9	43
3.1.3	Dériver des populations en clones	44
3.1.4	Validation des indels générées par CRISPR-Cas9.....	45
3.1.5	Quantification des ARNm de PHACTR1 par qPCR	45
3.1.6	Quantification protéique de PHACTR1 par immunobuvardage.....	47
3.1.7	Visualisation et quantification de la F-actine par immunohistochimie et microscopie :	47

3.1.8	RNAseq.....	49
3.2 -	Résultats.....	49
3.2.1	Inactivation de PHACTR1 par édition du génome avec le système CRISPR-Cas9.....	49
3.2.2	Structure de l'actine	52
3.2.3	Analyse différentielle du transcriptome des clones avec et sans TNF α par RNAseq	55
3.3 -	Discussion.....	57
4 -	Priorisation de variants	60
4.1 -	Méthodes.....	60
4.1.1	Culture cellulaire.....	60
4.1.2	Séquençage du génome entier.....	60
4.1.3	ATAC-seq.....	61
4.1.4	RNAseq.....	63
4.1.5	Hi-C.....	63
4.1.6	Édition du génome par le système CRISPR-Cas9	64
4.1.7	Dériver des populations en sous-populations et en clones.....	64
4.1.8	Validation des délétions générées par le système CRISPR-Cas9.....	64
4.2 -	Résultats.....	65
4.2.1	Enrichissement de variants associés aux MCV dans des régions de chromatine ouverte pour les EC	65
4.2.2	Débalancements alléliques dans l'ouverture de la chromatine	67
4.2.3	Régions de chromatine différemment ouvertes avec un traitement au TNF α	68
4.2.4	Enrichissement de motifs liés par des facteurs de transcription	68
4.2.5	Cartographie de la conformation chromosomique et du transcriptome des teloHAEC et priorisation de loci d'intérêt pour les MCV.	72
4.2.6	Délétion de 4 loci d'environ 2kb par CRISPR-Cas9 avec double évènement de coupure Cas9	76
5 -	Conclusion	79
6 -	Références	87

Liste des tableaux

Tableau 1 :	Résumé des SNPs au locus de <i>PHACTR1</i> et des maladies auxquelles ils sont associés.	12
Tableau 2 :	List of primers used in this study.....	28
Tableau 3 :	GTF file.....	30
Tableau 4 :	Transcript and protein IDs of <i>PHACTR1</i> gene (221692; ENSG00000112137).....	34
Tableau 5 :	rs9349379 is an eQTL for <i>PHACTR1</i> intermediate transcripts A+ and B+ in hCA.....	37
Tableau 6 :	Replication of rs9349379 eQTL effects in hCA from GTEx.	39
Tableau 7 :	Séquences des sgRNA utilisés pour l'édition de génome.....	44
Tableau 8 :	Amorces utilisées pour la validation des indels par PCR	45
Tableau 9 :	Amorces utilisées pour la quantification de <i>PHACTR1</i> par qPCR	46
Tableau 10 :	Expression de <i>PHACTR1</i> et des molécules d'adhésion.....	55
Tableau 11 :	Calcul de l'expression différentielle des 55972 gènes et transcrits par Hisat2.....	56
Tableau 12 :	Analyse de DAVID des voies biologiques des gènes différentiellement exprimés.....	57
Tableau 13 :	Séquences des sgRNA utilisés pour l'édition de génome.....	64
Tableau 14 :	Séquences des amorces utilisées pour la validation des délétions par PCR	65
Tableau 15 :	Association de SNPs à un débalancement allélique de chromatine ouverte.....	67
Tableau 16 :	Analyse de voies biologiques de DAVID des TFs différentiellement liés.....	71
Tableau 17 :	Résumé des 20 SNPs dans les pics de chromatine différentiellement ouverts.....	73
Tableau 18 :	Description des gènes DE dans les 4 régions sélectionnées	76
Tableau 19 :	64 motifs différentiellement liés par des TFs.	84

Liste des figures

Figure 1 :	Schématisation de la cascade menant aux MCV.	5
Figure 2 :	Effets des forces hémodynamiques sur l'endothélium.	7
Figure 3 :	Déséquilibre de liaison.	9
Figure 4 :	Corrélation inverse entre l'effet et la fréquence des SNPs.	11
Figure 5 :	Éléments de la séquence protéique des PHACTR1-4.	16
Figure 6 :	Modèle de PHACTR1 dans les EC.	17
Figure 7 :	Schématisation de la Cas9 ciblant une séquence d'ADN reconnue par le sgRNA.	23
Figure 8 :	Six <i>PHACTR1</i> transcripts are expressed in human samples.	33
Figure 9 :	Zoom-in representation of <i>PHACTR1</i> exon 14.	34
Figure 10 :	<i>PHACTR1</i> transcripts differential expression in human tissues and cell types.	35
Figure 11 :	Presence of PHACTR1 protein in vasclar smooth muscle cells.	36
Figure 12 :	Associations between rs9349379 genotypes and <i>PHACTR1</i> expression in hCA.	38
Figure 13 :	Effet des indels sur la séquence protéique des clones sélectionnés.	51
Figure 14 :	Validation de la diminution d'expression de <i>PHACTR1</i> dans les clones édités.	52
Figure 15 :	Maintien du cytosquelette d'actine dans la 1 ^{ière} série de clones <i>PHACTR1</i> ^{-/-}	53
Figure 16 :	Coloration de l'actine et des noyaux de la 2 ^e série de clones <i>PHACTR1</i> ^{-/-}	54
Figure 17 :	Chevauchement des gènes différentiellement exprimés.	56
Figure 18 :	Répartition des SNPs tombant dans des pics de chromatine ouverte selon les groupes.	66
Figure 19 :	Enrichissement de SNPs dans des pics de chromatine ouverte.	66
Figure 20 :	Répartition des SNPs tombant dans des motifs liés par des TFs.	70
Figure 21 :	Priorisation de 4 loci associés aux maladies coronariennes par GWAS.	75
Figure 22 :	Réponse des teloHAEC au stimulus de TNF α selon le temps d'exposition.	82
Figure 23 :	Répartition des pics de chromatine ouverte selon les chromosomes.	83
Figure 24 :	Exemple du navigateur au loci de SELE.	85
Figure 25 :	Validation des délétions par PCR.	86

Liste des annexes

Annexe 1 -	Observed <i>PHACTR1</i> transcripts	81
Annexe 2 -	Traîtement au TNF α	82
Annexe 3 -	Répartition uniforme de la chromatine ouverte	83
Annexe 4 -	Motifs différentiellement liés par des TFs	84
Annexe 5 -	Navigateur WashU Epigenome Browser	85
Annexe 6 -	Validation de la délétion des 4 loci.....	86

Liste des abréviations

ADMA	Asymmetric dimethylarginine
ATACseq	Assay for Transposase-Accessible Chromatin with highthroughput sequencing
Cas9	CRISPR associated protein 9
CCL	C-C motif ligand
CD31	cluster of differentiation 31 / PECAM-1
COX	Cyclooxygenase
CRISPR	clustered regularly interspaced short palindromic repeats / Courtes répétitions palindromiques groupées et régulièrement espacées
DNase-seq	séquençage des sites hypersensibles à la DNase I
EC	Cellules endothéliales (endothelial cells)
<i>EDN1</i>	Endothelin 1 gene
eQTL	locus d'expression quantitatif (Expression quantitative trait locus)
ET-1	protéine endothéline 1
<i>GAPDH</i>	Glyceraldehyde 3-phosphate dehydrogenase
GWAS	Étude d'association pangénomique (Genome-wide association study)
HAEC	cellules endothéliales aortiques humaines immortalisées (Immortalized human aortic endothelial cells)
HCASMC	cellules musculaires lisses d'artère coronaire humaine (Human Coronary Artery Smooth Muscle Cells)
HDL	Lipoprotéines de haute densité (High-density lipoprotein)
<i>HPRT1</i>	Hypoxanthine Phosphoribosyltransférase 1
hTERT	human telomerase reverse transcriptase
HUVEC	cellules endothéliales humaines de veine ombilicale
HYP	Hypertension
ICAM-1	molécules-1 d'adhésion intracellulaires (intracellular adhesion molecule 1)
IM	infarctus du myocarde
JAM	Junctional adhesion molecule
LD	déséquilibre de liaison (linkage disequilibrium)
LDL	Lipoprotéines de basse densité (Low-density lipoprotein)
LFA-1	Lymphocyte function-associated antigen 1
MCV	Maladies cardiovasculaires / CAD = Coronary artery disease
NO	oxyde nitrique (Nitric oxide)
OR	rapports de chances (odd ratios)
oxLDL	LDL oxydées
PECAM-1	Platelet endothelial cell adhesion molecule
<i>PHACTR1</i>	Phosphatase and actin regulator 1
PSGL-1	P-selectin glycoprotein ligand-1
ROS	Reactive oxygen species
SELE	Sélectine E

SELP	Sélectine P
sgRNA	ARN guides
siRNA	ARN interférents
SMC	cellules musculaires lisses (Smooth muscle cells)
SNPs	polymorphisme nucléotidique (single-nucleotide polymorphism)
TADs	topologically associated domains
<i>TBP</i>	TATA-binding protein
teloHAEC	cellules endothéliales aortiques humaines (human aortic endothelial cells)
TNF α	facteur de nécrose tumorale (tumor necrosis factor)
VCAM-1	molécules-1 d'adhésions de cellules vasculaires (vascular cell adhesion molecule 1)
VE-cadherin	vascular endothelial cadherin
VLA-4	Very Late Antigen-4
VSMC	cellules musculaires lisses vasculaires (Vascular smooth muscle cells)

Remerciements

J'aimerais remercier tous les membres de mon laboratoire qui ont été présents pour moi lors de mon cheminement et avec qui j'ai été amenée à collaborer sur plusieurs aspects de mon projet de maîtrise. Je remercie particulièrement Mélissa Beaudoin et Simon Lalonde pour les modèles cellulaires fournis, l'appui et les efforts conjoints dans l'avancement du projet ainsi qu'à Ken Sin Lo pour les différentes analyses bio-informatiques réalisées en parallèle. Je porte également un merci spécial à Mélissa Beaudoin pour son encadrement et ses encouragements ainsi qu'à Ken Sin Lo pour sa patience et persévérance à expliquer les bases de la bio-informatique à la néophyte que j'étais. Je remercie Guillaume Lettre pour l'incroyable opportunité qu'il m'a fournie à travailler dans son laboratoire. Son positivisme à toute épreuve est apprécié de tous et contribue grandement à l'harmonie au sein du laboratoire. Finalement, je remercie l'ensemble de mes collègues pour l'excellente équipe qu'ils forment et l'environnement de travail favorable qu'ils entretiennent. Lors de ma maîtrise, c'était un milieu où il faisait bon travailler, mais bon vivre aussi! Ce fut une expérience des plus enrichissantes!

1 - Introduction

1.1 - Maladies cardiovasculaires (MCV)

Les progrès médicaux ont permis d'allonger la durée de vie moyenne des gens. Toutefois, la population vieillissante vient avec une augmentation du taux des maladies reliées au cœur, reconnues comme principales causes de décès à l'échelle planétaire. La prévalence mondiale des maladies cardiovasculaires (MCV) a connu une hausse de 24,8% entre 2005 et 2015, passant à près de 423 millions de personnes touchées en 2015 (1). L'infarctus du myocarde (IM) affecte à lui seul environ 16 millions d'individus (1). Cela est en partie dû à l'importance des facteurs de risque traditionnels tels que l'obésité, le diabète et l'hypertension (HYP) (1). Au Canada, ce sont 22,2 milliards de dollars qui sont dépensés chaque année, ce qui représente 14,2% du fardeau économique (2). Cela positionne les MCV au deuxième rang des maladies les plus coûteuses, immédiatement après les maladies musculo-squelettiques (2). Malgré l'existence de traitements ciblant ceux-ci (ex : médicaments contre la haute pression et le cholestérol) et la sensibilisation quant au besoin d'une bonne hygiène de vie (non-fumeur, activité physique régulière, alimentation saine), cette maladie demeure un fléau mondial autant économiquement qu'au niveau de la santé (1). Il y a donc un besoin urgent de développer de nouveaux traitements contre celle-ci.

Les MCV sont des traits complexes avec une composante environnementale et une composante génétique. Bien qu'une portion de celles-ci soient expliquées par les facteurs de risques traditionnels, il demeure une portion héritable estimée à 40% (3) qui est encore inexpliquée et qui requiert une investigation génomique. Avec l'avancement des technologies dans le séquençage de nouvelle génération et la chute des coûts requis pour ces expériences, les études d'associations pangénomiques (GWAS) se multiplient. Celles-ci utilisent de larges cohortes avec des cas contrôles et des individus atteints dont le génome a été séquencé et tentent d'associer les variants avec le risque des maladies afin d'expliquer leur aspect génétique. À ce jour, près de 500 polymorphismes nucléotidiques (Single-nucleotide polymorphism, SNPs) sont associés avec les MCV ou l'HYP et ce nombre ne cesse de croître. La majorité des SNPs associés par GWAS pour les traits complexes tombent dans des régions non codantes de l'ADN de sorte qu'ils influenceraient le risque de maladies en modifiant l'expression génique plutôt qu'en altérant directement la fonction protéique. Il est cependant ardu de déterminer lequel des gènes à un locus donné est impliqué dans le phénotype puisque la distance effective du SNP

par rapport au gène est variable et que la proximité n'est pas automatiquement le mécanisme emprunté. De plus, malgré qu'un SNP soit associé, il se pourrait que ce soit un SNP en déséquilibre de liaison, donc non indépendant des SNPs avoisinants, qui soit à la source de l'effet observé. Ces questionnements s'appliquent à chaque association, ce qui multiplie grandement le nombre de gènes candidats potentiels. Cela engendre un problème au niveau de la compréhension biologique, car il est difficile d'inférer une causalité. Cette problématique tente d'être adressée à notre ère des omics afin d'apporter une compréhension approfondie des associations pangénomiques. Dans le cadre de ma maîtrise, je me suis appliquée à investiguer l'ensemble des SNPs associés aux MCV et à l'HYP.

1.2 - Athérosclérose

L'athérosclérose, soit la principale cause de MCV, est une maladie chronique affectant les parois des artères (4). Les artères sont formées de trois couches circulaires concentriques, soit l'intima, la média et l'adventice en partant de la lumière artérielle vers la paroi externe (4). L'intima comprend une monocouche de EC (endothelial cells, EC), appelée endothélium, qui sert de barrière sélective entre le sang et les couches sous-jacentes (4). L'athérosclérose correspond à la formation d'une plaque située sous l'endothélium qui est constituée entre autres de macrophages, de lipides et de cristaux de cholestérol (4). À long terme, celle-ci peut graduellement boucher une artère (sténose) (4). Entre temps, l'endothélium peut éclater de manière à recruter des plaquettes au site de rupture ce qui forme un caillot qui obstrue rapidement l'artère (thrombose) (4). Cela peut mener à des complications telles qu'un IM ou un accident vasculaire cérébral (4). L'IM, aussi connu sous le nom de crise cardiaque, consiste en l'obstruction des artères coronaires, soit les vaisseaux qui procurent un apport sanguin au muscle cardiaque afin d'assurer son oxygénation (5).

Les composantes et causes de l'athérosclérose sont controversées, mais il est bien connu qu'une réponse inflammatoire soit impliquée. Dans la phase initiale de l'athérosclérose, il y a un dysfonctionnement des EC. Le cisaillement oscillatoire du flux sanguin, les lipoprotéines de basse densité (Low-density lipoprotein, LDL) oxydé (oxLDL), les molécules inflammatoires (ex.: facteur de nécrose tumorale, tumor necrosis factor, TNF α) ou encore les radicaux libres sont des stress pouvant nuire à l'intégrité de l'endothélium (6). De plus, une accumulation subendothéliale d'apolipoprotéines B en raison de lipoprotéines oxydées contribuerait aussi dans l'initiation de la réponse inflammatoire (7). Lors d'une lésion, les EC activées sécrèteront des molécules d'adhésion ainsi que des chimiokines qui auront pour effet d'attirer les monocytes contenus dans le sang et permettre leur passage au travers

de l'endothélium vers l'intima, processus nommé transmigration (7). Ce dernier est possible par le biais de l'interaction entre le ligand PSGL-1 (P-selectin glycoprotein ligand-1) des monocytes avec les molécules P- et E-sélectine des EC qui sollicite le roulement des monocytes. La transmigration se fait également par l'interaction entre les intégrines VLA-4 (Very Late Antigen-4) et LFA-1 (Lymphocyte function-associated antigen 1) des monocytes avec les molécules-1 d'adhésions de cellules vasculaires (vascular cell adhesion molecule-1; VCAM-1) et les molécules-1 d'adhésion intracellulaires (intracellular adhesion molecule 1, ICAM-1) des EC qui permet l'arrêt des monocytes (8). La voie de signalisation des intégrines et chimiokines (CCL2/MCP1, CCL4, CCL5, C-C motif ligand) sécrétées par les EC participe également en activant les monocytes (8). De plus, l'adhésion de monocytes agit au niveau des protéines de jonctions serrées (JAM1-3, Junctional adhesion molecule) et adhérentes (VE-cadherin, vascular endothelial cadherin) qui maintiennent les EC collées de sorte que cela modifie la perméabilité de l'endothélium (8). Une fois traversés dans l'intima, les monocytes se différencient en macrophages et provoquent une réponse immunitaire, soit une marque typique de l'athérosclérose (7). Les macrophages phagocytent les LDL et deviennent des cellules spumeuses (9). Il s'en suit une accumulation de cellules apoptotiques, de lipides, de cristaux de cholestérol et de débris qui cause l'expansion de la couche subendothéliale (7, 9). Une nécrose subséquente et un amincissement de la chape fibreuse, soit la fine couche de EC entourant la plaque, risquent alors d'occasionner une thrombose (7, 9).

1.3 - Facteurs de risque traditionnel

Les facteurs de risque cardiovasculaire expliqueraient près de 90% du risque d'IM (10). La majorité de ceux-ci impliquerait un dysfonctionnement de l'endothélium vasculaire (6). Entre autres, il y a le tabagisme, l'obésité, la dyslipidémie, l'hyperhomocystéinémie, le diabète sucré, l'HYP, les maladies cérébro-vasculaires, la coronaropathie et l'insuffisance cardiaque (6). Concernant le **tabagisme**, cela constitue un apport important de radicaux libres qui endommage les EC (6). Cela occasionne un stress physiologique qui diminue la pression sanguine dans les coronaires et l'apport en oxygène au cœur en plus d'activer les plaquettes sanguines et les monocytes (6). Toutefois, il semble que certains individus soient moins susceptibles d'être affectés par ce facteur essentiellement environnemental (6). L'**obésité** est également associée aux MCV en ce sens où elle permet la production de ROS (dérivés réactifs de l'oxygène, Reactive oxygen species), l'hyperactivité de COX (Cyclooxygenase), l'activation de ET-1 (endothéline 1) et la résistance à l'insuline dans la voie du PI3K/Akt (phosphatidylinositol 3-kinase / serine/threonine kinase). Tout comme le tabagisme, il y a un facteur environnemental impliquer de

sorte que l'activité physique pourrait améliorer l'aspect de l'endothélium (6). En effet, l'exercice augmente normalement le diamètre des artères coronaires, mais l'augmentation de la pression sanguine augmente le relâchement d'oxyde nitrique (Nitric oxide, NO) qui mène à une vasoconstriction des artères coronaires (6). Au niveau de la **dyslipidémie** qui correspond à un changement de la biodisponibilité du cholestérol sanguin et de oxLDL (6). Il est question d'hypercholestérolémie lorsque ce taux est supérieur à la normale (6). Parallèlement, les lipoprotéines de haute densité permettraient de prévenir le dysfonctionnement endothélial (6). Par rapport à l'**hyperhomocystéinémie**, il y aurait une augmentation d'homocystéines causée principalement par un faible apport en vitamines et une prédisposition génétique qui provoquerait un stress oxydatif par des ROS (6). Pour ce qui est du **diabète (principalement le diabète de type 2)**, les dommages à l'endothélium proviennent d'une diminution de l'activité du NO pouvant être causée par la détérioration de la voie métabolique du L-arginine/NO (par une diminution de production ou une augmentation de la dégradation du NO), par une diminution de la sensibilité des cellules musculaires lisses vasculaires (Vascular smooth muscle cells, VSMC) au NO ou encore par l'inhibiteur de NO ADMA (Asymmetric dimethylarginine) (6). Quant à l'**hypertension artérielle**, elle engendrerait une vasodilatation affectant l'endothélium en plus d'interférer avec l'activité du NO sollicitant alors des voies alternatives incluant l'hyperpolarisation qui augmenterait ce premier effet (6). Pour les **maladies cérébro-vasculaires**, les réponses vasodilatatrices de l'endothélium des artères seraient affectées (6). Un dysfonctionnement endothélial au niveau des coronaires semble reproduire celui cérébral, d'où l'association (6). En ce qui a trait à la **coronaropathie** et l'**insuffisance cardiaque**, ce sont des maladies qui impliquent aussi le niveau d'activité du NO et la vasodilatation de l'endothélium (6). Dans tous ces cas, il est important de noter le dysfonctionnement au niveau de l'endothélium.

1.4 - L'endothélium et la dysfonction endothéliale

La dysfonction endothéliale est en cause dans les facteurs de risques traditionnels et elle initie l'athérosclérose qui est la principale cause des MCV (Figure 1). Tel que mentionné précédemment, l'endothélium est la monocouche de EC formant la paroi interne des artères et servant de **barrière perméable sélective**. Ces cellules sont de forme allongée (20-40µm de long X 10-15µm de large X 0,1-0,5µm d'épaisseur) suivant l'axe des artères en raison des forces hémodynamiques de la circulation sanguine (11). Elles sont liées étroitement entre elles par des jonctions serrées limitant le passage des macromolécules (11). Les mécanismes possibles pour traverser la barrière endothéliale étant de passer au travers des EC elles-mêmes, de traverser les jonctions serrées ou de voyager par des vésicules de

transport (11). La majorité des macromolécules utiliseront les deux dernières options puisque leurs propriétés hydrophiles, leurs charges et leur grande taille les empêchent de diffuser au travers de la membrane des EC (11). Pour passer les jonctions serrées, soit la principale voie, il doit y avoir une réorganisation de celles-ci via l'actine, la myosine et les points de jonction (11). Des molécules inflammatoires telles que l'histamine et la thrombine peuvent promouvoir l'ouverture des jonctions serrées par la voie de signalisation des protéines kinases C qui phosphorylent les protéines de jonction ce qui remanie l'actine et arrondie les EC laissant plus d'espace pour traverser l'endothélium (11). Les facteurs de croissance de l'endothélium vasculaire (Vascular endothelial growth factors; VEGFs) jouent un rôle important dans la perméabilité de l'endothélium via la régulation de la synthèse d'oxyde nitrique (nitric oxide synthase; NOS) (11). Les ROS causent un stress oxydatif néfaste favorisant l'athérosclérose (11). Un équilibre est donc nécessaire entre le NO et les ROS (11).

Risk factors for CVD

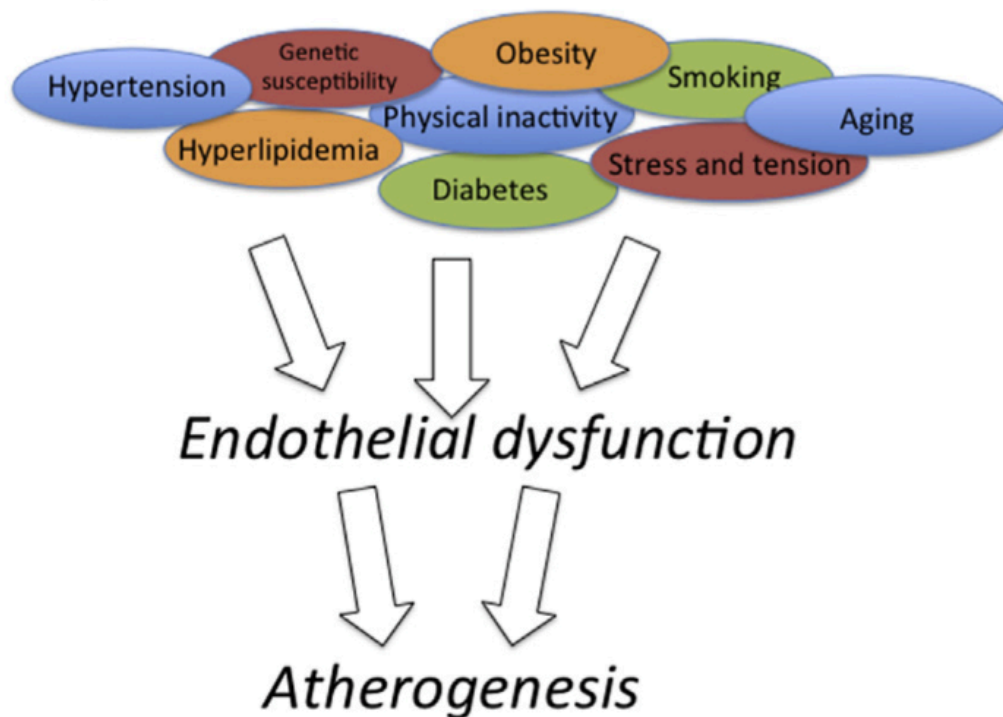


Figure 1 : Schématisation de la cascade menant aux MCV (11).

Les **forces hémodynamiques**, soient les forces de cisaillement et la déformation cyclique dues au flux sanguin, réduisent la prolifération cellulaire qui retarde l'apoptose et promeuvent la migration cellulaire qui aide à la réparation de la monocouche (11). Elles agissent en régulant la production de substances vasoactives. En effet, les forces de cisaillement constantes et élevées favorisent l'intégrité de

l'endothélium par la sécrétion des **vasodilatateurs** dont l'oxyde nitrique (nitric oxide; NO) et la prostacycline (PGI₂), de **vasoconstricteurs** dont l'endothéline 1 (ET-1), de thrombomoduline, de sulfate d'héparine et d'activateurs tissulaires du plasminogène (tPA) (11). NO et PGI₂ inhibent la croissance et la migration des cellules musculaires lisses (Smooth muscle cells, SMC) alors que ET-1 les stimule (11). De plus, NO et PGI₂ inhibent l'agglomération des plaquettes (11). À l'opposé, les forces de cisaillement faibles, perturbées ou oscillantes de la circulation sanguine affectent également la perméabilité de la barrière et s'accompagnent généralement d'une morphologie cellulaire davantage circulaire et non orientée, d'une accumulation de LDL et une infiltration accrue de ces particules, soit une étape qui enclenche le développement de l'athérosclérose (11). L'exercice préviendrait les MCV à ce niveau en augmentant le flux sanguin et donc les forces de cisaillement qui aident au maintien de la barrière endothéliale (11). L'effet de la force de cisaillement proviendrait du fait de son impact sur l'activité des ADN méthyltransférases dont DNMT1 qui assure le maintien des profils de méthylation de l'ADN de sorte à modifier le phénotype de l'endothélium lorsque la circulation sanguine est irrégulière (11). Ainsi, la méthylation de l'ADN aurait un rôle majeur dans l'homéostasie de l'endothélium (11). À noter que l'HYP augmente la déformation cyclique et nuit donc à l'intégrité de l'endothélium (11). Les forces hémodynamiques contrôlent aussi la production de **facteurs de croissance** et de **cytokines**, telles que des facteurs de croissance dérivés des plaquettes (platelet-derived growth factor; PDGF), des facteurs de croissance des fibroblastes (basic fibroblast growth factor; bFGF), des facteurs de croissance transformants β (transforming growth factor beta; TGF-β) et des interleukines 1 et 6 (Il-1 et Il-6), qui stimulent les SMC à migrer et proliférer, soit une seconde étape cruciale dans le développement de l'athérosclérose (11). Elles influencent aussi l'expression de **molécules d'adhésion**, telles que VCAM-1, ICAM-1, des protéines-1 chimioattractrices de monocytes (monocyte chemotactic protein 1; MCP-1 / Chemokine ligand 2, CCL2), des sélectines P (SELP) et des sélectines E (SELE), à la surface membranaire des EC ce qui contribue à l'agglomération des plaquettes de même qu'à l'adhésion et la migration des leucocytes dont les monocytes (11). Ainsi, les forces hémodynamiques ont un rôle important dans l'intégrité de l'endothélium (Figure 2). Il va de soi que l'altération de l'endothélium et de ses fonctions peut avoir des conséquences majeures sur l'intégrité de l'endothélium et ainsi promouvoir les MCV. De plus, la dysfonction endothéliale étant un facteur de risque peu étudié pour les MCV, elle représente une bonne piste pour développer un traitement et donc mérite d'être investiguée.

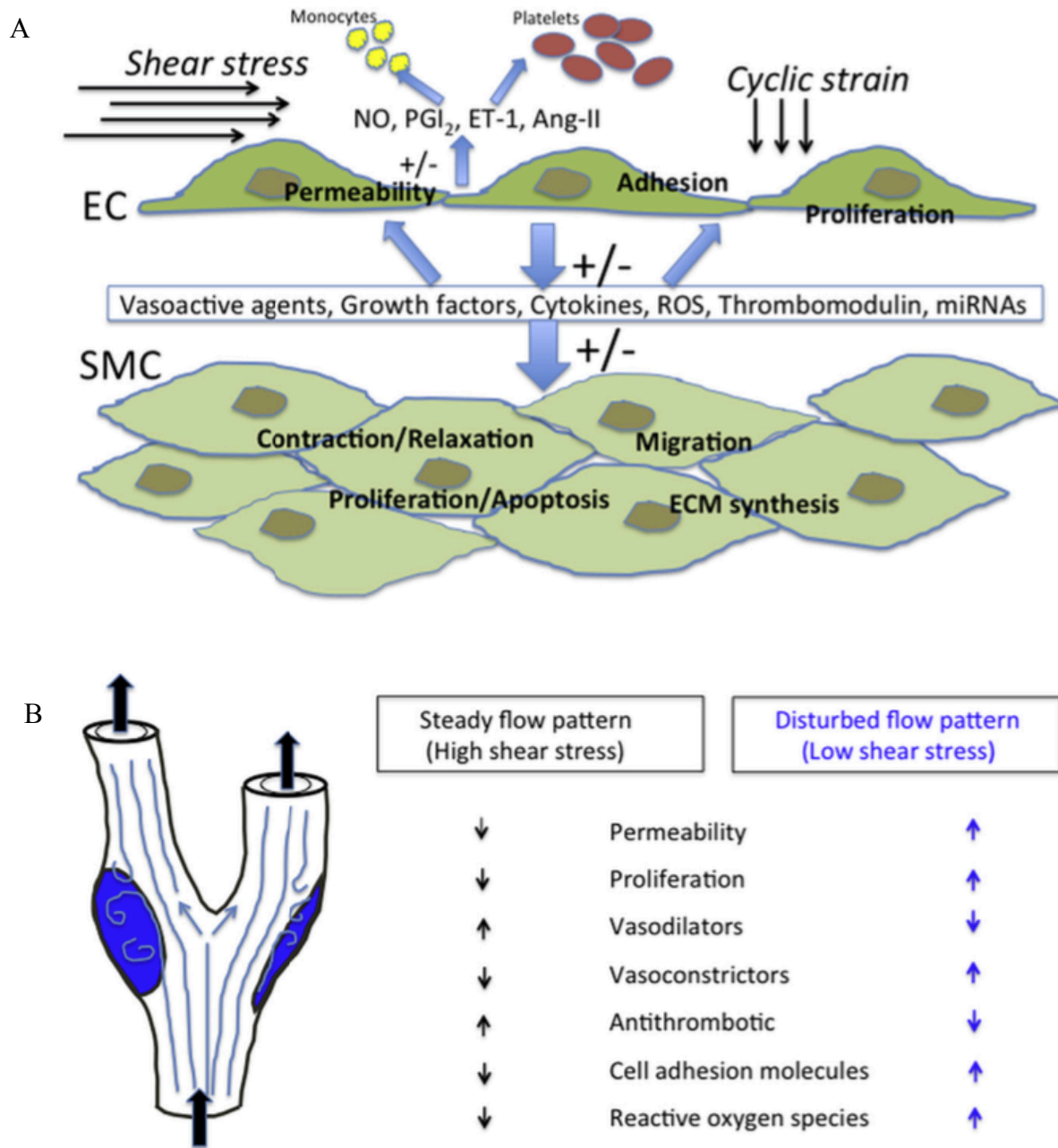


Figure 2 : Effets des forces hémodynamiques sur l'endothélium. A) Les forces de cisaillement et la déformation cyclique contrôlent l'expression de vasodilatateurs, de vasoconstricteurs, de dérivés réactifs de l'oxygène, de facteurs de croissance, de cytokines et de molécules d'adhésion ce qui module les interactions avec les SMC, les cellules immunitaires et les plaquettes et peut enclencher l'athérosclérose. B) Comparaison des forces de cisaillement fortes et constantes par rapport à celles faibles et perturbées sur le phénotype de l'endothélium. Il y a la formation d'une plaque d'athérosclérose dans les régions où le flux est affecté et les forces de cisaillement sont moindres (régions en bleu) (11).

1.5 - Association génétique

Bien que plusieurs facteurs de risque soient établis tels que l'HYP, l'hypercholestérolémie, le diabète et la cigarette, cela ne parvient pas à expliquer l'ensemble des cas (12). Cela provient du fait que les MCV dont l'IM ont des caractères héréditaires, c'est-à-dire que l'historique familial peut prévenir ou prédisposer à ces maladies (12). Les facteurs de risque connus ont eux-mêmes une proportion héréditaire à considérer dans le calcul, de sorte que même s'ils expliquent 90% du risque de MCV, l'aspect génétique explique possiblement plus que le faible 10% restant (6). Il est estimé que la proportion héréditaire de ces maladies est de 40% (3). La compréhension des traits génétiques impliqués dans les MCV pourrait ainsi mener à la découverte de voies biologiques n'impliquant pas les facteurs de risques traditionnels et pouvant être exploitées pour le développement de nouveaux traitements. En ce sens, les **GWAS** permettent d'associer des variants génétiques avec des phénotypes.

Dans ces GWAS, le génome de milliers d'individus, cas ou contrôles, est séquencé puis des analyses statistiques sont réalisées sur les millions de SNPs retrouvés. Des **rapports de chances** (odd ratios; OR), soit le risque d'avoir la maladie pour un allèle donné, sont ainsi calculés pour chaque SNP par comparaison entre les cas et les contrôles.

	Allèle A	Allèle B
Cas (maladie)	a	b
Contrôles	c	d

$$OR = \frac{P(\text{cas} \mid \text{allèle A})}{P(\text{cas} \mid \text{allèle B})} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Ainsi, un OR de 1 indique que le SNP n'influence pas le phénotype. Un OR > 1 signifie que l'allèle augmente la probabilité du phénotype, donc le risque de maladie, tandis qu'un allèle avec un OR < 1 serait protecteur (13). Pour chaque SNP, une **p-valeur** est également calculée, soit la probabilité d'obtenir un résultat égal ou supérieur que celui observé sous l'hypothèse nulle. Un test est généralement considéré significatif et l'hypothèse nulle est rejetée lorsque la p-valeur est $\leq 0,05$, de sorte que 5% du temps il s'agira d'un faux positif. Toutefois, la probabilité d'obtenir un faux positif augmente avec le nombre de tests. Du fait que les GWAS combinent une multitude de tests indépendants (Un test par SNP), la correction de Bonferroni est apportée pour en tenir compte, soit la division du niveau de confiance par le nombre de tests (14). Ainsi, une p-valeur de 5×10^{-8} est généralement considéré comme un seuil conservateur pour être significatif dans les GWAS (14). Autrement, il faut tenir compte de la non-indépendance de certains de ces tests. Le **déséquilibre de**

liaison (linkage disequilibrium; LD) est une mesure pour déterminer la dépendance entre les allèles de deux SNPs dans une séquence qui donne une idée de la variation génétique d'une population dans le temps (14). La dépendance provient du lien chromosomal entre deux marqueurs sur un chromosome qui ségrègent dans une famille avec les générations jusqu'à ce qu'une recombinaison homologe se produise entre ces deux marqueurs (Figure 3). Plus les deux marqueurs sont près l'un de l'autre et plus rare sera la recombinaison homologe brisant le lien entre les deux. Avec le temps et les événements de recombinaison homologe, les SNPs tendent vers un équilibre de liaison, c'est-à-dire où la fréquence d'une observation simultanée de deux allèles est proportionnelle à leur fréquence allélique. Le LD témoigne de la probabilité que l'un soit vu en présence de l'autre et se calcule par la corrélation (r^2) :

$$r^2 = \frac{(\pi_{AB}\pi_{ab}\pi_{Ab}\pi_{aB})^2}{\pi_A\pi_B\pi_a\pi_b}$$

où plus la valeur est près de 1 et plus le LD est grand (14). Afin d'éviter la redondance dans les SNPs associés par GWAS, ceux ayant un $r^2 > 0,8$ sont généralement considérés en déséquilibre de liaison et seulement un (celui avec la plus petite p-valeur) est conservé. Toutefois, cela engendre deux alternatives sur le résultat des GWAS : soit le SNP conservé peut être celui impliqué dans la fonction biologique de sorte qu'il est un SNP fonctionnel et que l'association est directe, soit il est en LD et donc indirectement associé (14). Il est à noter que les SNPs en LD atteignent généralement 10–100kb (15), ce qui couvre une grande région et implique de nombreux SNPs. Cela complexifie grandement l'interprétation biologique des GWAS et multiplie les pistes potentielles.

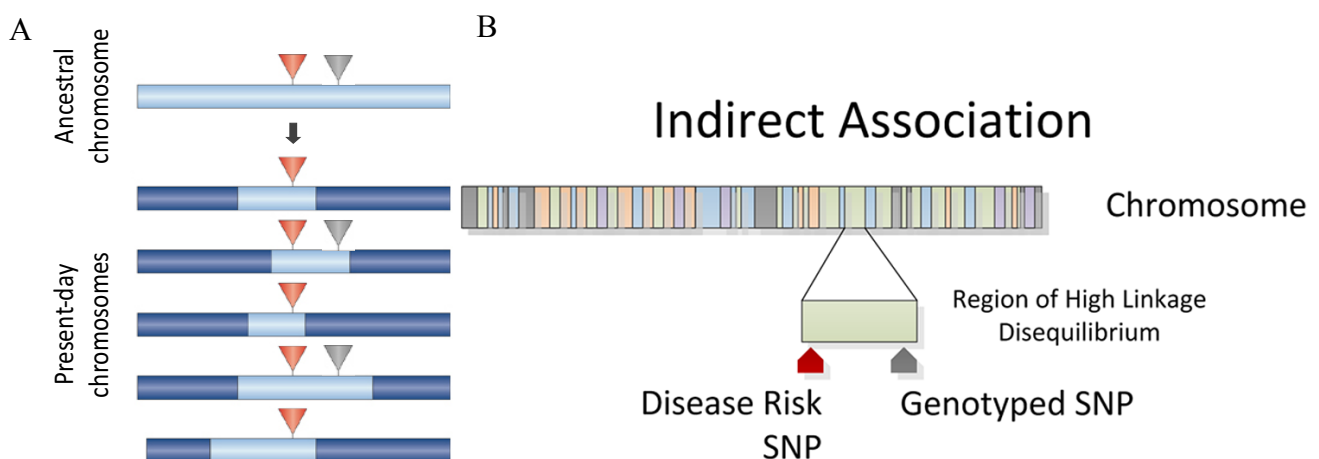


Figure 3 : Déséquilibre de liaison. A) Évolution d'un chromosome dans le temps. Les régions en bleu pâle correspondent à la séquence ancestrale et celles en bleu foncé résultent de recombinaisons homologues avec les générations. Pour un variant donné (triangle rouge), un marqueur à proximité

(triangle gris) ne se retrouvera pas chez tous les individus (ici, il est présent dans 2 cas sur 5). (16) B) L'association génétique indirecte d'un SNP en LD (gris) dans une région de haut LD contenant le SNP fonctionnel (rouge) est problématique pour comprendre l'aspect biologique (14). Chaque bloc de couleur sur le chromosome représente une région de fort déséquilibre de liaison et ses délimitations correspondent à des recombinaisons homologues ayant eu lieu à un moment donné au cours des générations.

D'un point de vue biologique, les maladies mendéliennes sont caractérisées par des variants rares ayant un impact majeur tandis que les traits complexes sont considérés pour avoir des **variants communs** (fréquence allélique dans la population >1-5%) avec des effets individuels mineurs et des pénétrances variables (Figure 4) (17). Cela provient du fait que les variants rares tombent généralement dans une région codante de manière à altérer directement la fonction de la protéine tandis que les variants communs sont non-codants et affectent plutôt l'expression génique (18-20). Les traits complexes sont déterminés par une multitude de loci qui ont individuellement un effet minime, mais qui parviennent collectivement à expliquer l'héritabilité et donc le risque de maladie (21). Étant donné leur faible impact individuel, ces variants difficilement identifiables nécessitent de travailler sur de larges cohortes pour assurer une bonne puissance des tests statistiques et ainsi parvenir à obtenir un résultat significatif permettant de les distinguer. La reproduction des associations dans des populations de différentes ethnies assure aussi la validité des associations. De plus, il est important de tenir compte de l'influence des facteurs de risque traditionnels qui varient entre les individus et donc de corriger pour ceux-ci ainsi que d'autres co-variables (âge, sexe, ethnie) pour ne pas introduire de biais. Au niveau des MCV et de l'HYP, de nombreuses associations ont été réalisées et reproduites dans différentes populations. Il y a actuellement **161 et 357 SNPs associés avec les MCV et l'HYP** respectivement (22-30). Il est estimé que près de 75% des associations génétiques impliquent des loci qui ne sont pas liés aux facteurs de risques traditionnels des MCV (31) et donc qui ne sont pas encore exploités pour des traitements et demeurent des pistes potentielles.

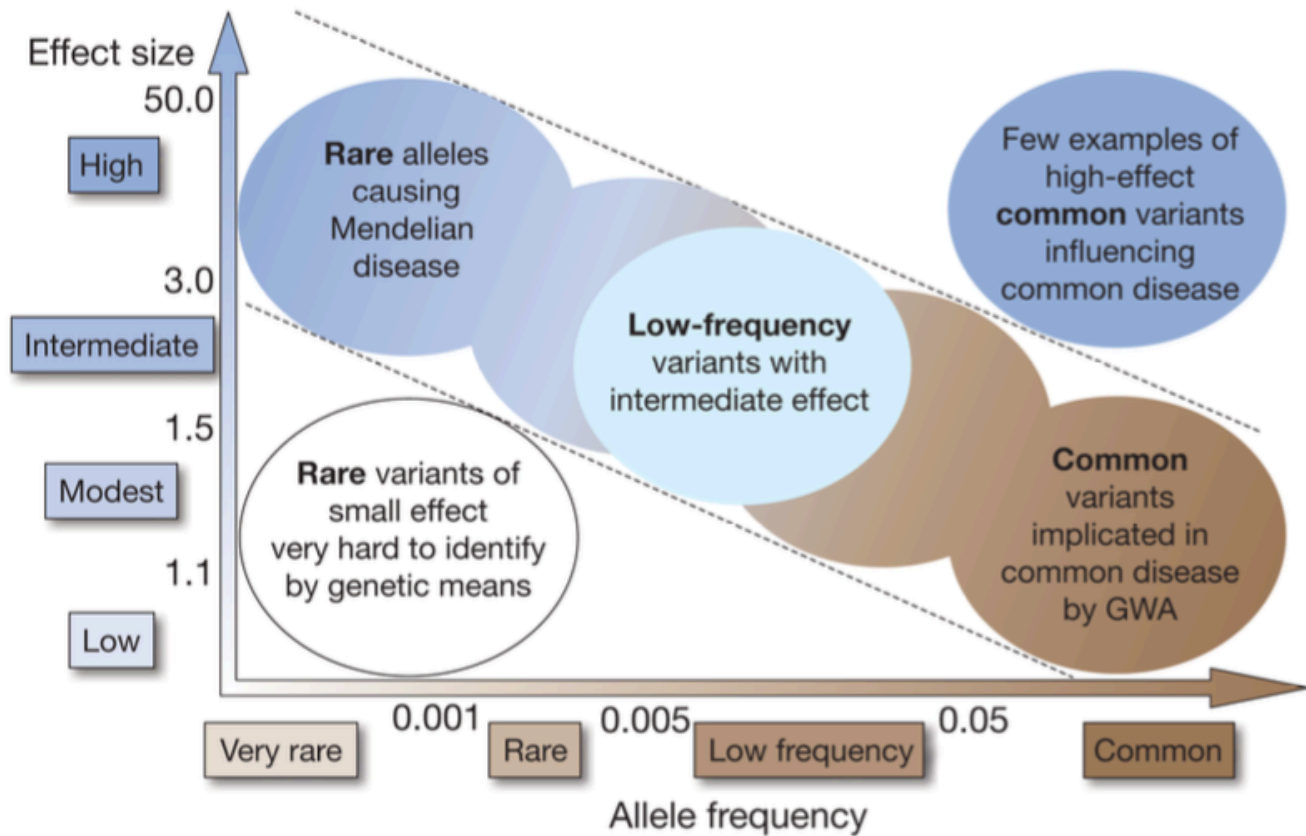


Figure 4 : Corrélation inverse entre l'effet et la fréquence des SNPs. Les variants impliqués dans les traits complexes tels que les MCV sont donc communs et avec un faible impact. (17)

Les MCV ont été associées au locus du chromosome 6p24 pour la première fois avec le SNP rs12526453, situé dans une région intronique du gène *PHACTR1* (Phosphatase and actin regulator 1), dans une cohorte de 2967 cas et 3075 contrôles provenant des États-Unis, de la Suisse, de l'Italie, de la Finlande et de l'Espagne (L'allèle de référence G est remplacé par l'allèle alternatif C : fréquence = 0,65; OR = 1,13; $P = 1 \times 10^{-9}$) (32). Ce résultat a été reproduit à plus grande échelle dans une cohorte européenne de 22 233 cas et 64 762 contrôles (Allèle alternatif C : fréquence = 0,67; OR = 1,10; $P = 1,15 \times 10^{-9}$) (33). Depuis, ce loci de *PHACTR1* et des SNPs non codants (Tableau 1) qui y sont présents ont été associés à plusieurs problèmes liés au système vasculaire dont les MCV (34-36), les migraines (37, 38), la dissection de l'artère cervicale (39), la dysplasie fibromusculaire (40), l'HYP (41), la sténose de l'artère coronaire (42), la calcification de l'artère coronaire (43), les accidents vasculaires cérébraux ischémiques (44), les indices hémodynamiques centraux altérés (45), l'hypercholestérolémie familiale (46) et le diabète de type 2 (36). Les associations avec ce loci ont été réalisées dans des cohortes avec différentes ethnies dont caucasienne, européenne, mexicaine,

libanaise, chinoise et plusieurs autres (32, 33, 42, 43, 47-51). Considérant que les différences génétiques au niveau des fréquences allélique et du déséquilibre de liaison entre les populations peuvent influencer le résultat de GWAS (52), la réplification dans plusieurs groupes ajoute une valeur à l'association de ce loci avec les MCV.

Tableau 1 : Résumé des SNPs au locus de *PHACTR1* et des maladies auxquelles ils sont associés.

SNPs	OR	p-valeurs	Maladies	Populations
rs12526453 G/C	1,13	1×10^{-9}	IM	États-Unis, Suisse, Italie, Finlande et Espagne (32)
	1,10	$1,15 \times 10^{-9}$	MCV	Europe (33)
	1,09	$9,72 \times 10^{-9}$	MCV	Europe (35)
	0,94	$8,18 \times 10^{-6}$	Migraine	Europe (35)
	0,82	$8,9 \times 10^{-4}$	Sténose de l'artère coronaire	Liban (42)
		$2,63 \times 10^{-6}$	accidents vasculaires cérébraux ischémiques	Europe (44)
	0,48	0,001	Hypercholestérolémie familiale	Canada (46)
	1,25	0,0002	Diabète de type 2 avec MCV	Europe (36)
	β : 3,50	0,001	Indices hémodynamiques centraux altérés	
	β : 1,10	0,005	HYP	Europe (45)
β : -4,30	0,002	Temps d'onde réfléchie		
rs9349379 A/G	β : -0,211	$2,65 \times 10^{-11}$	IM et calcification des artères coronaires	Europe (43)
	1,09	$5,9 \times 10^{-9}$	migraine	Europe (35)
	1,1	$8,97 \times 10^{-8}$	MCV	Europe (35)
	1,15	$1,76 \times 10^{-9}$	MCV	Chine (34)
	0,86	$3,20 \times 10^{-8}$	migraine	Allemagne, Pays-Bas (37)
	0,86	$2,81 \times 10^{-10}$	migraine	(Multiple 29 groupes) (38)
	0,77	$1,00 \times 10^{-11}$	dissection de l'artère cervicale	Europe (39)
	1,32	$7,21 \times 10^{-4}$	dysplasie fibromusculaire	France, État-Unis (EUR) (40)
	0,82	0,0119	HYP	Japon (41)
	1,338	$8,02 \times 10^{-10}$	sténose de l'artère coronaire	Liban (42)
rs4714955 C/T		$9,8 \times 10^{-11}$	MCV	Europe (35)
rs9369640 A/C	1,29	$8,45 \times 10^{-4}$	dysplasie fibromusculaire	France, État-Unis (EUR) (40)
rs1332844 C/T	1,27	$1,72 \times 10^{-3}$	dysplasie fibromusculaire	France, État-Unis (EUR) (40)
rs2327620 A/G	0,806	$2,39 \times 10^{-6}$	sténose de l'artère coronaire	Liban (42)
rs2026458 T/C	β : 0,162	$1,78 \times 10^{-7}$	calcification de l'artère coronaire	Europe (43)

* Allèle à risque pour lequel un OR est calculé est mis en gras.

Le SNP rs9349379, se retrouvant aussi dans la région intronique du gène *PHACTR1* au locus 6p24, a été identifié dans une cohorte européenne de 9961 participants comme étant impliqué dans l'IM et la calcification des artères coronaires (Fréquence A = 0,59; $P = 2,65 \times 10^{-11}$) (43). Dans une cohorte libanaise de 1949 individus, puis répliqué dans un groupe de 2547 individus, rs9349379 a été associé à la sténose des artères coronaires, soit un facteur de risque important dans l'IM (OR = 1.34, $P = 8.02 \times 10^{-10}$) (42). Ce SNP a également été reporté comme étant associé à la calcification des artères coronaires, l'athérosclérose étant impliqué dans les MCV, dans une population allemande de 4329 individus (Allèle G : fréquence = 0,39 ; OR = 1,26, $P = 1,47 \times 10^{-5}$) (49) ainsi que dans une population afro-américaine de 5186 personnes (50). Cela est compatible avec une étude qui a associé la calcification des coronaires et de l'aorte avec le SNP rs12526453 dans la population aux Pays-Bas (48). Parallèlement, rs9349379 a été associé avec les migraines (37) et la dissection de l'artère cervicale (39) de telle sorte que l'allèle alternatif G du SNP correspondrait à une diminution du risque (OR = 0,86; $P = 3,20 \times 10^{-8}$ et OR = 0,75; $P = 4,46 \times 10^{-10}$ respectivement). Des résultats préliminaires du laboratoire Lettre confirment ces associations dans une population canadienne-française. À partir d'une cohorte de 1176 cas d'IM et 1996 contrôles provenant de la biobanque de l'Institut de cardiologie de Montréal (ICM), 35 des 45 SNPs génotypés avait des OR concordant avec la littérature (53). La plus forte association était avec le SNP rs12526453 (allèle de référence G : OR = 0.79, $P = 8.4 \times 10^{-4}$) (53). Pour avoir une meilleure caractérisation de cette association à ce locus, treize SNPs supplémentaires retrouvés dans des individus de descendance européenne ainsi que 387 variants identifiés à partir d'haplotypes, tous provenant du projet 1000 Genomes, ont été analysés (53). rs9349379 est le SNP ressortant avec la plus forte association (allèle alternatif G : OR = 1,37, $P = 8,4 \times 10^{-6}$) (53). De plus, ce dernier n'est pas relié aux facteurs de risque traditionnels de l'HYP, de la haute pression sanguine, du diabète de type 2, des niveaux de LDL et de lipoprotéines de haute densité (High-density lipoprotein, HDL) et du tabagisme (53). Ainsi, la réplification dans différentes populations malgré leur génétique distinctive et les styles de vie variés souligne *PHACTR1* comme une piste potentielle à étudier pour développer des traitements alternatifs.

1.6 - Le SNP rs9349379

Bien que le SNP rs9349379 se situe dans le gène *PHACTR1*, cela ne signifie pas que *PHACTR1* soit un gène causal, donc qu'il influence le risque de développer l'IM. En effet, il se pourrait que le SNP régule l'expression d'un autre gène à proximité et que ce soit ce dernier qui soit causal. rs9349379 est localisé dans une région avec un fort signal H3K27ac et une absence de signaux pour H3K4me3,

H3K36me3, H3K27me3 et H3K9me3 signifiant une activité activatrice dans quatre échantillons d'aorte dans la base de données ENCODE (54). Une étude montre que la délétion de cette région par le système CRISPR-Cas9 (clustered regularly inter-spaced short palindromic repeats) dans des EC et VSMC augmente l'expression de l'endothéline 1 (*EDNI*), un gène à 600kb du SNP (54). Toutefois, dans GTEx, soit une base de données avec des expériences de RNAseq dans une multitude de tissus, le génotype du SNP n'affecte pas l'expression de ce gène, mais influence l'expression de *PHACTRI* dans l'artère tibiale, l'artère coronaire et l'aorte. De plus, le laboratoire du Dr Lettre a démontré qu'il y a une forte corrélation ($r^2=0,52$; $P=0,008$) entre le génotype de rs9349379 et le niveau d'expression de *PHACTRI* dans l'artère coronaire droite (RCA) (53). Ce SNP représente donc un locus d'expression quantitatif (eQTL; expression quantitative trait locus), ce qui signifie que *PHACTRI* est un gène à ce locus impliqué dans la maladie. Comme l'allèle alternative G associé à l'IM diminue l'expression, une faible expression de *PHACTRI* serait un facteur de risque pour l'IM (53). Ce résultat n'est pas répliqué dans des études testant la corrélation entre les SNPs de *PHACTRI* et son niveau d'expression dans les leucocytes, le foie, les cellules adipeuses, la peau, l'épiploon, la média et l'adventice de l'aorte et les cellules lymphoblastoïdes (33, 55). Cela pourrait signifier une transcription spécifique au tissu de la RCA, un tissu mixte composé entre autres de EC (EC) et de SMC. Toutefois, les SMC expriment peu *PHACTRI* (53). Il faut donc regarder l'expression de *PHACTRI* au niveau des EC, là où l'expression de *PHACTRI* est mesurable, pour étudier le mécanisme qui influence la probabilité de développer l'IM.

Autrement, le SNP rs9349379 se situerait dans un site prédit de liaison pour le facteur de transcription MEF2 (myocyte enhancer factor 2) qui a un rôle dans la différenciation cellulaire, la prolifération, l'apoptose, la morphogénèse et la réponse quant à un stress (56). Il a également un rôle quant aux mécanismes épigénétiques tels que le changement de conformation de la chromatine et l'expression de microARN. Ainsi, il permet le développement vasculaire de même que le maintien de l'intégrité vasculaire dans l'endothélium (56). L'allèle G de rs9349379, soit celle qui augmente le risque de développer l'IM, brise le motif de liaison de MEF2 (53). Autrement, une délétion hétérozygote de 34 paires de bases au site du SNP rs9349379 réalisée avec le système CRISPR-Cas9 provoque une diminution de 34% de l'expression de *PHACTRI* comparativement à la souche sauvage homozygote A au rs9349379 (53). Même si les résultats sont controversés (57, 58), une étude montre qu'une variation génétique de MEF2A prédispose aux MCV (59). Ainsi, l'effet eQTL du SNP et le bris du motif de liaison de MEF2 appuient le rôle du gène *PHACTRI* dans l'IM.

1.7 - PHACTR1

Il existe quatre protéines dans la famille de PHACTR, soit *PHACTR1* à 4, dont l'expression est différentielle (60). Le gène *PHACTR1* est principalement exprimé au niveau du cerveau et du cœur (53). Son expression est d'ailleurs plus élevée dans les lésions athérosclérotiques qui est fortement régulée par des stimuli inflammatoires et athérogènes (61). Parallèlement, il a été démontré que *PHACTR1* joue un rôle important dans l'angiogenèse (62, 63). De plus, la protéine qu'il encode se lie à l'actine et interagit avec la protéine phosphatase 1 (PP1) (60). Son interaction avec PP1, une enzyme ubiquitaire impliquée dans de nombreux processus biologiques tels que l'expression génique, la contraction musculaire et le métabolisme du glycogène, se ferait via un site sur PP1 distinct du domaine conventionnel R/K-R/K-hydrophobique-X-F/W permettant possiblement ainsi la formation de complexes multiprotéiques (60). Quant à sa liaison avec la G-actine cytoplasmique cela inhiberait la polymérisation d'actine et pourrait ainsi réguler la dynamique du cytosquelette d'actine (60). Par ailleurs, *PHACTR1* est impliqué dans la stabilité des plaques via la régulation des métalloprotéinases matricielles de manière à assurer le maintien des tissus conjonctifs (64).

1.7.1 Structure

Concernant sa structure, la protéine contient quatre motifs RPEL (Arginine, Proline, Glutamine, Leucine), un au N-terminal et trois répétés avant le C-terminal, qui sont liés par la G-actine (Figure 5) (65, 66). Elle possède également une séquence conservée au C-terminal qui est liée par PP1 et deux signaux d'import nucléaire au N- et C-Terminal qui interagissent avec les importines α - β (Figure 5) (65, 66). La superposition des motifs RPEL avec la séquence conservée et les signaux d'import nucléaire occasionne une compétition de l'actine avec PP1 et les importines α - β respectivement pour la liaison à la protéine PHACTR1, ce qui permet de contrôler sa localisation cellulaire (65, 66). En effet, une faible concentration de G-actine permet une accumulation nucléaire de PHACTR1 par rapport à une grande concentration qui inhibe cette importation (65, 66). Il a d'ailleurs été démontré que l'altération de ces domaines interrompt le contrôle de la localisation cytoplasmique et nucléaire de PHACTR1 de manière à altérer l'assemblage d'actomyosine (65, 66). En liant la G-actine cytoplasmique, PHACTR1 pourrait débalancer l'équilibre de la réaction de polymérisation d'actine de manière à déstabiliser le cytosquelette et ainsi altérer la perméabilité de l'endothélium (Figure 6). Inversement, en étant lié par les importines α - β , PHACTR1 est importé au noyau et pourrait moduler l'expression génique (Figure 6).



Figure 5 : Éléments de la séquence protéique des PHACTR1-4 (65). Rouge : motifs RPEL; Jaune : signaux d'import nucléaire; Vert : séquence conservée en C-Terminal.

1.7.2 Interaction avec PP1 et rôle du NO dans l'IM

PHACTR1 inhibe l'activité de PP1 in vitro, une enzyme qui active l'eNOS (endothelial nitric oxide synthase; enzyme qui produit de l'oxyde nitrique (NO) dans les vaisseaux sanguins) via la déphosphorylation de son site d'inhibition à la Thr⁴⁹⁵ (67). Ainsi, avec PP1, eNOS est active et produit du NO (68), soit un vasodilatateur ayant des propriétés cardioprotectrices dont antioxydant, anticoagulant et antithrombogène (69). En effet, le NO prévient l'adhésion des monocytes et donc empêche leur transmigration qui occasionnerait une accumulation de VSMC (6). De plus, il inhibe l'activation des plaquettes sanguine de sorte à limiter leur agrégation et la formation subséquente de thrombose (6). Une répression de *PHACTR1* par ARN interférents (siRNA) n'affecte pas le niveau d'expression de PP1 et de eNOs, mais réduit considérablement le taux de Thr⁴⁹⁵ phosphorylées et augmente significativement la production de NO (70). Donc, puisque PHACTR1 est un inhibiteur endogène de PP1 qui active eNOS, la réduction du niveau de protéines de PHACTR1 cause une augmentation de l'activité de PP1 et l'activité subséquente de eNOS qui mène à une production supérieure de NO (Figure 6). Ces résultats sont cependant contradictoires, car l'allèle qui prédispose à la maladie permet également la production du NO protecteur. Deux explications pourraient justifier ce phénomène. La première étant que l'endothélium vasculaire s'adapte à cette surexpression de NO relié au génotype de rs9349379 en diminuant la production de NO (68). La seconde serait une production élevée de peroxy-nitrite, un radical libre toxique et proathérogène dommageable pour l'endothélium, résultant de l'excès de NO (68). Dans les faits, le peroxy-nitrite aurait été impliqué dans les MCV incluant l'IM (71).

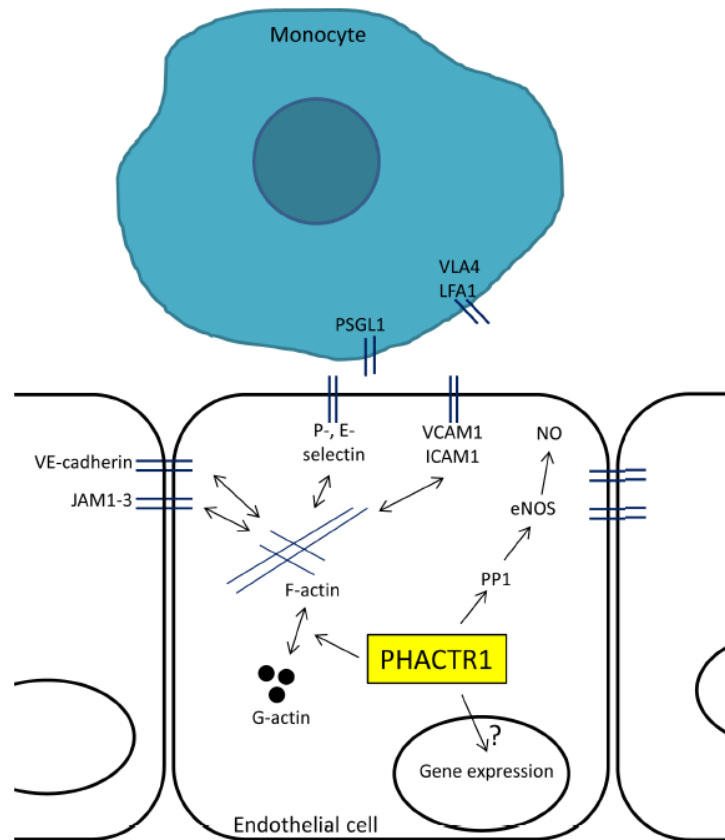


Figure 6 : Modèle de PHACTR1 dans les EC (68). Les protéines transmembranaires sur lesquelles repose le cytosquelette d'actine déterminent l'adhésion cellulaire des SMC et les jonctions serrées entre les EC. Le cytosquelette d'actine sert ainsi de support dans plusieurs voies de signalisation. Entre autres, il permet de réguler la perméabilité de l'endothélium lors de la transmigration des monocytes. PP1 déphosphoryle eNOS qui synthétise alors de l'NO. PHACTR1 peut être lié par les importines α - β via ses deux signaux d'import nucléaire et ainsi être recruté dans le noyau. PHACTR1 pourrait donc agir sur le cytosquelette d'actine, la production de NO ou l'expression génique de manière à moduler le risque de MCV.

1.7.3 Isoformes

Lors du processus de maturation des ARN pré-messagers, il y a l'épissage des introns pour ne conserver que les exons. Toutefois, il est possible d'obtenir différents ARNm issus d'un même gène en raison d'une variabilité contrôlée durant cet épissage. Chaque isoforme est obtenue à partir d'un épissage alternatif, c'est-à-dire que certains exons de l'ARNm sont coupés de sorte à être inclus ou exclus de la séquence finale et qu'il est ainsi possible de produire différentes protéines. Dans les bases de données, il y a près d'une dizaine de transcrits répertoriés (72), mais ceux-ci ne sont pas tous

caractérisés et leur patron d'expression n'est pas connu. Toutefois, dans le cadre de mon projet, il est important de déterminer quelles isoformes sont présentes dans les tissus d'intérêts pour les MCV, c'est-à-dire au niveau du cœur et plus précisément dans les cellules de l'intima (ECs, monocytes, SMC...). En plus du transcrit **long (1743pb, 580a.a.)** connu pour être exprimé uniquement au niveau du cerveau, une isoforme plus **courte (1674pb, 557a.a.)** a été identifiée dans le même cadre de lecture dont le codon d'initiation se situe en aval de la forme longue et ayant un exon supplémentaire. Il est à noter que sa séquence peptidique a une masse moléculaire apparente de 75kDa au lieu de celle prédite de 66kDa qui serait potentiellement due à des modifications post-traductionnelles (60). Celle-ci est exprimée au niveau du cœur, de l'artère coronaire droite et de EC. Elle serait donc importante pour le système cardiaque.

Une publication récente démontre qu'il existerait trois isoformes, soit une longue (1743pb, 580a.a.) exclusive au système neuronal, une intermédiaire (1674pb, 557a.a.) présente dans les cellules immunitaires et endothéliales et une **courte (435pb, 144a.a.)** présente uniquement dans les monocytes (61). Dans cette étude, il est démontré que l'expression de *PHACTRI* est augmentée dans les lésions athérogènes (61). Toutefois, des stimuli inflammatoires (lipopolysaccharides, TNF α), utilisés pour reproduire les conditions d'athérosclérose, augmenteraient l'expression du transcrit intermédiaire alors qu'ils diminueraient celle du transcrit court(61). L'inverse serait observé pour des stimuli de oxLDL (61). De plus, le SNP rs9349379 a été identifié comme un eQTL pour les MCV dans les macrophages (61). Ainsi une diminution de l'expression du transcrit court dans les macrophages serait associée avec les MCV (61).

1.7.4 Induction et répression de l'expression

Pour induire l'expression de *PHACTRI*, des stimuli endothéliaux peuvent être utilisés tels que le VEGF (vascular endothelial growth factor), le TNF α et la contrainte de cisaillement. Dans les EC humaines de veine ombilicale (HUVEC), l'expression de *PHACTRI* est induite par un traitement de VEGF, un facteur de perméabilité vasculaire ainsi qu'un facteur proangiogénique (63). Cependant, ce phénomène n'a pas été observé dans le laboratoire Lettre lorsque l'expérience fut reproduite (53). Le TNF α est une molécule inflammatoire qui engendre des dommages à l'endothélium et la contrainte de cisaillement est un stimulus qui prévient l'athérosclérose dans les vaisseaux sanguins (53). Ceux-ci ne semblent pas affecter l'expression de *PHACTRI* selon les résultats obtenus dans le laboratoire Lettre (53). De manière concordante avec l'enclenchement de la dysfonction endothéliale, un traitement au TNF α

induit une augmentation de l'expression de *SELE*, soit une molécule d'adhésion impliquée dans le phénotype endothélial.

Dans les HUVEC, sa répression enclenche l'apoptose et réduit la formation de tubes capillaires (62). Une expérience préliminaire dans le laboratoire Lettre consistait à inhiber *PHACTRI* avec des siRNA. Cela a permis d'observer par qPCR une réduction de 60-70% du nombre de transcrits de sorte que la protéine n'est plus visible par immunobuvardage (68). Afin de comparer l'expression de plusieurs gènes avec ou sans la diminution de l'expression de *PHACTRI*, l'expérience de RNAseq menée montre une répression de *PHACTRI* par des siRNA. Cette répression affecte de manière significative les voies biologiques d'adhésion et de migration, ce qui est concordant avec l'importance de l'intégrité vasculaire et l'attachement et la migration de monocytes dans l'athérosclérose (68).

1.8 - Ère des omics

Avec l'accessibilité des technologies de nouvelle génération, les données génétiques se font nombreuses et de multiples variants sont associés avec les maladies. La difficulté étant que les MCV sont des traits multigéniques, que la majorité des SNPs impliqués sont non-codants et influencent de manière minime le risque de maladies via la régulation de l'expression génique et qu'une association ne permet pas d'inférer une causalité en raison du LD. Cela rend ainsi la compréhension du rôle biologique de ces SNPs fastidieuse. De plus, l'effet d'un SNP peut également être spécifique à un tissu, un type cellulaire ou une condition donnée de sorte que les GWAS peinent à repérer une telle association en raison de la dilution du signal. Sur ce point, plusieurs des SNPs associés par GWAS modulent l'expression génique que lorsque les cellules sont activées. (73). J'ai mentionné qu'à ce jour, **161 et 357 SNPs sont associés aux MCV et à l'HYP** respectivement. En considérant les SNPs en LD ($r^2 \geq 0,8$), il y a un total de **5274 et 12443 SNPs** respectivement (SNPs trouvés à partir de la population européenne du 1000 Genomes Project). Près de la totalité de ceux-ci sont non codants, car ils n'affectent pas la séquence protéique (mutations non-sens, mutations faux-sens, mutation causant un changement du cadre de lecture ou site d'épissage) (outil ENSEMBL's Variant Effect Predictor). De plus, environ 50% des SNPs associés par GWAS aux MCV ne sont pas associés aux facteurs de risque traditionnels dont l'hypercholestérolémie, l'HYP, le diabète de type 2 et l'obésité (outil PhenoScanner (74)) sous-entendant un effet via des mécanismes alternatifs à ceux connus. Toutefois, même ceux qui sont liés peuvent impliquer la dysfonction endothéliale puisque c'est un phénotype relié à ces facteurs de risque.

Ces problématiques ne peuvent être adressées exclusivement par les GWAS et donc il y a un besoin de compléter l'information de ces études avec d'autres expériences pour avoir une vision d'ensemble de l'effet des SNPs dans l'organisme. Pour comprendre le rôle de chacun, il s'agit de distinguer le SNPs causaux des SNPs en LD, de les relier avec les gènes dont l'expression est affectée et de déterminer dans quel sens le génotype applique son effet (augmentation ou diminution de l'expression avec l'allèle de risque). Un SNPs peut occasionner un effet de plusieurs façons tel que via la modulation de l'ouverture de la chromatine ou l'altération d'un motif de liaison. En effet, les SNPs associés à des traits complexes sont enrichis dans les régions de chromatine ouverte dans les tissus pertinents (75-77). De plus, bien qu'il soit coutume de considérer le gène le plus près comme étant celui affecté (dbSNPs utilise des seuils de 2kb et 0,5kb en amont et aval du gène respectivement), l'effet d'un SNP peut se faire via une interaction allant souvent jusqu'à 2mb (78). Les SNPs peuvent également être enrichis dans des motifs de liaison particuliers et influencer l'expression génique, donc être des eQTLs, en modulant l'affinité de liaison de ces TFs (79). De plus, bien que la séquence du génome est sensiblement identique dans toutes les cellules d'un organisme, l'organisation de la chromatine varie considérablement entre les types cellulaires et les tissus. Cela permet une régulation de l'expression génique différentielle spécifique à un tissu, un type cellulaire ou une condition. C'est donc à ce niveau que les SNPs non codants identifiés par GWAS vont avoir un effet sur les activateurs et répresseurs (75, 80).

1.8.1 Bases de données publiques

Dans cette initiative de compréhension de causalité, plusieurs bases de données publiques, telles que GTEx et ENCODE, ont été développées pour fournir des données de transcriptomique, de conformation chromosomique et d'accessibilité de la chromatine dans plusieurs tissus et lignées cellulaires. **GTEx** est une base de données du transcriptome obtenues par RNAseq dans 53 tissus (81-83). Des analyses d'eQTLs dans les différents tissus et pour les différents transcrits sont réalisées pour mettre en évidence la spécificité des SNPs dans les tissus et sur l'épissage alternatif. **ENCODE** (Encyclopedia of DNA Elements) a pour mission d'établir une liste exhaustive de tous les éléments régulateurs de l'ADN via des expériences de DNase-seq (séquençage des sites hypersensibles à la DNase I) ou d'ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing), de méthylation de l'ADN, d'immunoprécipitation de diverses protéines (modificateurs d'histone, facteurs de transcription, régulateurs de la chromatine) et de conformation chromosomique (84). Celles-ci représentent une source d'information précieuse pour investiguer la causalité des SNPs

dans un tissu ou un type cellulaire spécifique. Toutefois, une limitation de ces outils est que les études sont exclusivement réalisées dans des tissus sains. Dans l'optique où il devient nécessaire d'investiguer davantage le rôle des variants dans un contexte de maladie, ces bases de données manquent d'information. La solution pour pallier cette lacune étant d'utiliser un tissu malade ou des cellules sous un stimulus engendrant la pathologie. Quelques études récentes démontrent le rôle de SNPs non codant dans les MCV et témoignent de la pertinence de caractériser un SNP dans un modèle cellulaire impliqué : le SNPs rs9349379 régule l'expression de *EDNI* dans les EC (54) et rs17114036 contrôle l'expression de *PLPP3* dans les EC (85). Une autre étude témoigne de la pertinence de combiner des expériences de génomique, de transcriptomique et d'épigénomique pour interpréter la multitude des loci identifiés par GWAS : les gènes *SIPA1*, *TCF21*, *SMAD3*, *FES* et *PDGFRA* dans les cellules musculaires lisses d'artère coronaire humaine (Human Coronary Artery Smooth Muscle Cells, HCASMC) affectent le risque de MCV (31).

1.8.2 Transcriptome

Il a été mentionné précédemment que les SNPs peuvent exercer un effet sur l'expression génique même s'ils sont loin des promoteurs de gènes (78). L'analyse du transcriptome permet donc de corréliser les SNPs avec l'expression de gènes spécifiques (eQTLs) et donc de comprendre en partie leur mécanisme d'action de par la fonction exercée par ces gènes. D'ailleurs, il y a un enrichissement de SNPs associés par GWAS pour des eQTLs (81). Les SNPs étant des eQTLs sont enrichis dans les régions de chromatine ouverte ($P = 2,9 \times 10^{-8}$), les motifs de liaison ($P = 4,3 \times 10^{-18}$), les promoteurs ($P = 1,7 \times 10^{-19}$) et les régions activatrices ($P = 0,003$) (81). Il y a aussi un enrichissement des sQTL dans les sites d'initiation de la transcription (TSS) (81). Il faut également considérer que la variance dans l'expression génique diverge davantage entre les tissus (47% de la variance totale) qu'entre individus (4% de la variance totale) (83) soulignant ainsi l'importance d'étudier un mécanisme biologique dans un tissu relié au phénotype analysé. Dans le cas de la dysfonction endothéliale, les expériences réalisées dans des EC sont donc toutes indiquées.

1.8.3 Conformation chromosomique

La chromatine est divisée en compartiments ouverts « A » et fermés « B » correspondant à des régions actives et réprimées respectivement (86). Cette configuration peut être analysée par ATAC-seq qui permet de mesurer l'ouverture de la chromatine. Les compartiments sont divisés en plus petits domaines appelés TADs (topologically associated domains), soit un segment linéaire de l'ADN

comportant de nombreuses interactions à même la région, mais très peu avec les régions voisines (86). Une expérience de Hi-C permet de les identifier. Les TADs sont généralement délimités par la cohésine et les CTCF (CCCTC-binding factor), ce qui facilite les interactions longue distance entre activateurs et promoteurs (86). Ils sont maintenus avec les divisions cellulaires, les types cellulaires et les espèces, mais leur état ouvert ou fermé est modulé par le type cellulaire et le stade de développement (86). Bien que cela reste à prouver, plusieurs résultats indiquent que les gènes impliqués dans une voie biologique en particulier tendent à être réunis au sein d'un TAD (86). La conformation ouverte et fermée et les interactions activateurs-promoteurs étant hautement spécifiques aux types cellulaires et aux conditions données de manière à assurer la spécificité de l'expression génique, il est probable que la dysfonction endothéliale ait un impact sur ces régions de la chromatine. La variation de la chromatine entre les individus est souvent causée par l'altération d'un unique motif de liaison d'un TFs qui enclenche des modifications dans les interactions de longue distance perturbant ainsi tout le mécanisme de régulation (86-88). En effet, une modification des sites liés par les CTCF aux extrémités des TADs pourrait altérer la conformation tridimensionnelle de la chromatine et avoir un impact majeur sur le phénotype cellulaire (86). En appui avec cette notion, les motifs CTCF sont enrichis de SNPs associés par GWAS (86).

1.9 - Système d'édition du génome CRISPR-Cas9

Le système CRISPR (clustered regularly interspaced short palindromic repeats / Courtes répétitions palindromiques groupées et régulièrement espacées) provient du système immunitaire adaptatif des bactéries (89). Lorsqu'un virus infecte une bactérie, cette dernière incorpore un fragment de cet ADN étranger à son locus CRISPR comme « espaceur » entre des séquences répétées (89). Par la suite, la bactérie transcrit ces « espaceurs » du locus CRISPR en ARN qui sont utilisés comme séquences guides par les nucléases pour identifier et couper l'ADN étranger de nouveaux virus et ainsi éviter les infections (89). Ce système a été exploité par l'emploi des nucléases bactériennes combinées avec des ARN guides (sgRNA) synthétiques ciblant des séquences d'intérêt afin d'éditer le génome. La Cas9 (CRISPR associated protein 9), une endonucléase causant une cassure double brin, nécessite un sgRNA qui permet de reconnaître la séquence d'ADN à cliver (Figure 7) (90). Une fois la cassure réalisée, la cellule enclenche un mécanisme de réparation de l'ADN pour réparer cette cassure hautement délétère via la jonction des deux extrémités ou par recombinaison homologue lorsqu'une matrice d'ADN est présente (Figure 7) (90). Le premier cause une insertion ou délétion aléatoire de généralement quelques paires de bases (indel). Le second permet d'obtenir exactement la séquence désirée, mais est moins

fréquent puisqu'il se produit lors de la division cellulaire et que son efficacité dépend du type cellulaire, du locus et de l'ADN matrice (90). Depuis sa découverte, le système ne cesse d'évoluer pour améliorer l'efficacité et la spécificité. De nouvelles techniques ont également été développées pour contrôler l'expression (Cas9 inactivée; Cas9 couplée avec un activateur ou répresseur transcriptionnel), modifier l'épigénétique (Cas9 couplée avec une enzyme : histone méthyltransférase, déméthylase, acétyltransférase, désacétylase), visualiser des régions chromosomiques (Cas9 couplée à un fluorophore), identifier les interactions de protéines avec l'ADN (Cas9 couplé à un tag pour une immunoprécipitation) et cibler des ARNm (89).

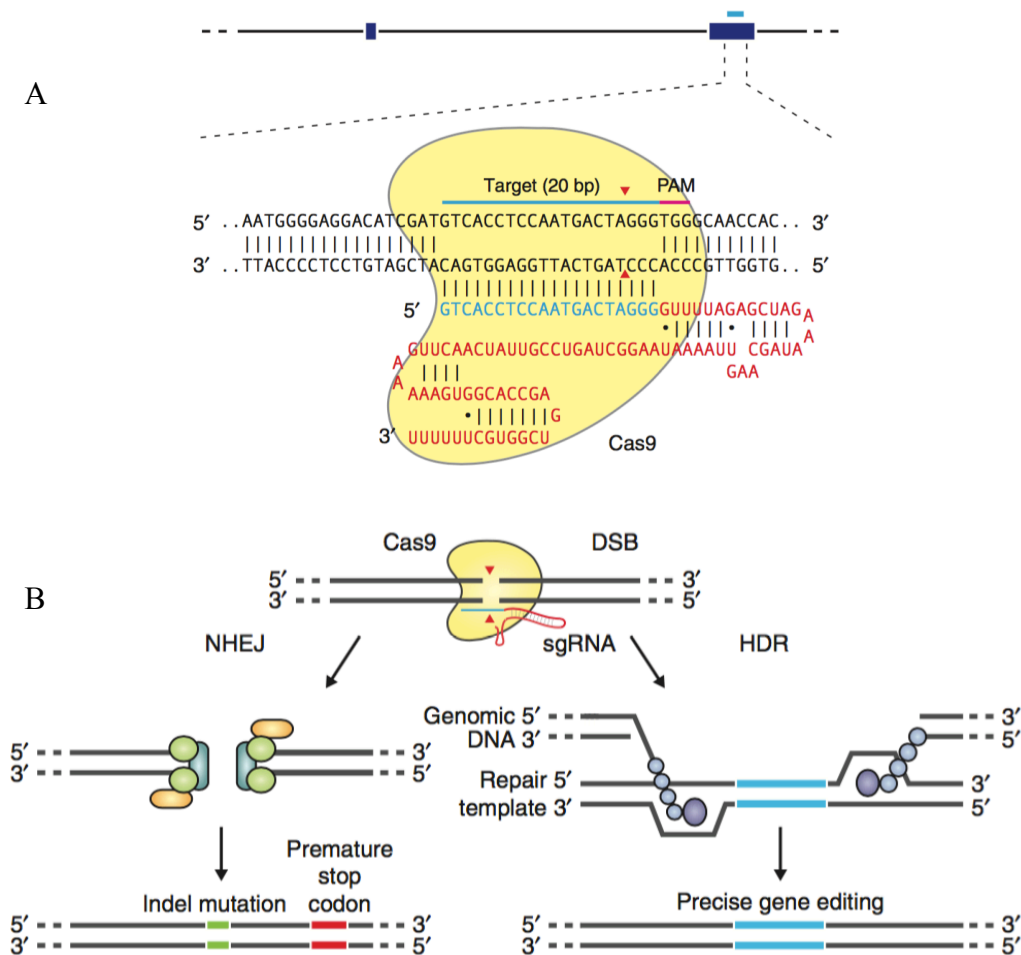


Figure 7 : Schématisation de la Cas9 ciblant une séquence d'ADN reconnue par le sgRNA. A) La portion rouge de le sgRNA est fixe et est reconnue par la Cas9. La portion bleue de le sgRNA est la séquence modifiée générée synthétiquement pour cibler une région d'ADN d'intérêt. B) Après une cassure double brin, la cellule enclenche le mécanisme de réparation de l'ADN qui peut être soit la jonction des deux extrémités (NHEJ : nonhomologous end joining) de manière à générer une indel, soit la recombinaison homologue avec une matrice d'ADN (HDR : homology-directed repair) (90).

1.10 - Hypothèses et objectifs de recherche

Les MCV sont reconnues pour être une cause importante de mortalité dans le monde. Des nombreuses GWAS effectuées afin de déterminer l'aspect héréditaire de la maladie ont permis d'identifier près de 500 SNPs. Toutefois, la compréhension biologique de ceux-ci demeure limitée. L'un d'entre eux, le SNP rs9349379, identifié et répliqué dans plusieurs cohortes et populations, est un eQTL du gène *PHACTRI* et n'est pas relié aux facteurs de risques traditionnels pour les MCV suggérant l'implication d'une nouvelle voie biologique influençant ces maladies.

Mon projet de maîtrise se divise en deux parties. Dans un premier temps, le but est de **déterminer le rôle de *PHACTRI* dans l'endothélium vasculaire**. L'hypothèse est que l'expression réduite de *PHACTRI* déstabilise le cytosquelette d'actine dans les EC et altère l'intégrité de l'endothélium vasculaire de manière à promouvoir l'athérosclérose. Pour cela, il s'agira de [1] valider les différentes isoformes exprimées, de [2] caractériser l'effet phénotypique de l'inactivation de *PHACTRI* sur la polymérisation d'actine et le transcriptome. Puisqu'il est estimé que le phénotype ne sera pas blanc ou noir, des expériences quantitatives (quantification d'actine par microscopie, mesure de l'expression de molécule d'adhésion par qPCR) seront réalisées afin de tester cette hypothèse. Toutefois, bien que *PHACTRI* demeure une bonne piste d'investigation, les MCV sont des traits complexes modulés par plusieurs gènes de sorte que les autres variants identifiés par GWAS demeurent des pistes potentielles d'investigation pour inférer la causalité de gènes impliqués dans le risque de maladies.

Dans un deuxième temps, le but est de **prioriser des variants potentiellement causaux parmi l'ensemble de ceux associés aux MCV et à l'HYP par GWAS**. L'hypothèse est que certains de ces SNPs modulent le risque de maladies en affectant directement les fonctions endothéliales. Il s'agit donc de considérer ces variants dans un contexte de dysfonction endothéliale induit via un stimulus inflammatoire afin de [1] réaliser une cartographie exhaustive du transcriptome, de l'ouverture de la chromatine et de la conformation chromosomique dans des EC, [2] sélectionner les variants présentant une cumulation de données indiquant leur rôle potentiel, et [3] déléter ces régions afin de pouvoir caractériser leur effet au niveau des fonctions endothéliales et identifier les gènes impliqués. Pour générer les données complémentaires aux GWAS, des expériences comparatives de RNAseq, ATAC-seq et Hi-C seront réalisées avec et sans stimulus inflammatoire pour mettre en évidence la régulation dynamique des fonctions endothéliales. Puis pour étudier le rôle des loci sélectionnés, une édition de génome par CRISPR-cas9 sera menée suivie d'une caractérisation phénotypique.

2 - Caractérisation des isoformes de PHACTR1

(Ce chapitre réalisé en collaboration avec Mélissa Beaudoin, Simon Lalonde, Ken Sin Lo et Guillaume Lettre est un article publié dans la revue BMC Medical Genetics en date du 8 juin 2018 (91).)

Genome-wide association studies (GWAS) have identified a robust association between coronary artery disease (CAD) and the *PHACTR1* locus on chromosome 6p24 (32). This locus is highly pleiotropic as the same allele associated with increased risk of CAD is protective of migraine (37), cervical artery dissection (39), and fibromuscular dysplasia (40). Previously, we fine-mapped the CAD association to the candidate causal variant rs9349379 located in an intron of *PHACTR1* (53). This SNP disrupts a MEF2 binding site and is an expression quantitative trait locus (eQTL) for *PHACTR1* in human coronary arteries (hCA) (53). The eQTL result was replicated in the Genotype-Tissue Expression (GTEx) dataset (81). *PHACTR1* encodes proteins of unclear biological functions: it was discovered in a yeast two-hybrid screen because it interacts with protein phosphatase-1 (PP1) and can also bind actin through its RPEL domains (60, 65). Over-expression experiments have also implicated *PHACTR1* in apoptosis (62), angiogenesis (63), cellular matrix remodelling (64), and cell motility (65). Although these processes are important for atherogenesis and CAD, the precise role(s) that *PHACTR1* plays at the site of atherosclerotic lesions in coronary arteries is unknown.

Gupta et al. used CRISPR-Cas9 genome editing experiments in endothelial cells (ECs) derived in vitro from induced pluripotent stem cells to test the effect of genotypes at rs9349379 on the expression of nearby genes (54). In this cellular system, the regulatory element around rs9349379 did not modulate the expression of *PHACTR1*, but controlled the expression of the endothelin-1 (*EDNI*) gene located 600 kb upstream. *EDNI* encodes ET-1, a 21 amino acids peptide with a potent vasoconstrictor effect in humans and clear implications for cardiovascular diseases. Paradoxically, rs9349379 is not an eQTL for *EDNI* in our hCA samples nor in the GTEx dataset (53, 81). Together, these results raise the possibility that rs9349379 might control the expression of both *PHACTR1* and *EDNI* in different cell types and/or upon different environmental stimuli (e.g. inflammation).

For these reasons, we continued to investigate the regulation of *PHACTR1* expression in human tissues and cell types. Consistent with a recent study (61), we report here a long *PHACTR1* transcript expressed in the brain and a short transcript expressed in the brain, the heart, and monocytes. We also confirmed four intermediate transcripts that are different due to alternative splicing of two exons. They are co-expressed in tissues and cells tested. The shorter intermediate transcripts have a higher expression in the brain, the heart, and monocytes, whereas the longer intermediate transcripts are more expressed in ECs and vascular smooth muscle cells (VSMCs). Finally, we tested genetic associations between genotypes at rs9349379 and the expression of *PHACTR1* transcripts (or exons) in hCA samples. Our results strengthen the link between CAD-associated rs9349379 and *PHACTR1* expression in non-immune cells found in hCA.

2.1 - Methods

2.1.1 Cell lines and tissues

We purchased pooled human umbilical vein endothelial cells (HUVEC; C2519A), human aortic endothelial cells (HAEC; CC-2535), and human coronary artery endothelial cells (HCAEC; CC-2585; Lonza) and cultured them in EBM-2 media supplemented with EGM-2 or EGM2-MV SingleQuot kit. We purchased immortalized HAEC (teloHAEC; CRL-4052; ATCC) and cultured them in vascular cell basal medium supplemented with vascular endothelial cell growth kit and puromycin at 0.3µg/mL. Human coronary artery smooth muscle cells (HCASMC) and human aortic smooth muscle cells (HASMC) were obtained from Dr. Tardif's lab and cultured in medium 231 (M-231-500) supplemented with smooth muscle growth supplement (S-007-25; Gibco). Monocytes were obtained from Dr. Rioux's lab and we cultured them in RPMI with 10% fetal bovine serum. RNA from tissues was extracted either with the Ribopure Kit (Ambion) or using EZ1-XL Advance and Qiagen EZ1 RNA Tissue Mini Kit. hCA were obtained from the "Réseau d'Échanges de Tissus et d'Échantillons Biologiques" (RÉTEB) biorepository at the Montreal Heart Institute and Quebec Heart and Lung Institute. We purchased adult brain total RNA (540005–41, lot#6048990) and adult heart total RNA (540011–41, lot#6056165; Agilent Technologies).

2.1.2 Rapid Amplification of cDNA Ends (RACE)

We extracted total RNA from HUVEC, teloHAEC, HCAEC, HAEC, HCASMC and HASMC P20 (with a passage (P) lower than 8 for primary cells unless specified) cells grown to confluence in

100mm dishes with the Rneasy Plus Mini Kit (Qiagen). For the RACE experiments, we also added RNA from one hCA donor (#ICM167), adult brain, and adult heart. RNA integrity and concentration were measured by Agilent RNA 6000 Nano II assays (Agilent Technologies) on an Agilent 2100 Bioanalyzer. We reverse transcribed 1µg of total RNA using a modified oligo (dT) primer and the SMARTScribe Reverse Transcriptase (Clontech). We specifically amplified cDNA of *PHACTRI* using the SMARTer RACE 5'/3' Kit (Clontech). The gene specific primers are in Table 2. The RACE products were purified by gel extraction with the NucleoSpin Gel and PCR Clean-Up Kit, cloned in pRACE vector with the In-Fusion HD Cloning Kit and transformed with Stellar Competent Cells (Clontech). After overnight incubation, plasmids were extracted from single bacterial colonies with QIAprep Spin miniprep kit (Qiagen). We sequenced *PHACTRI* inserts by Sanger sequencing using M13 primers (Table 2).

2.1.3 Long-read Pacific Biosciences (PacBio) sequencing library preparation

(This section was made by MéliSSa Beaudoin)

We produced libraries using 1 µg of total RNA (HUVEC, HAEC, HCAEC, HASMC, HCASMC and brain) and 1µL of 12µM dT-Barcode (dT-BC), according to the manufacturer's protocol (Procedure and checklist-Isoform sequencing using the SMARTer PCR cDNA synthesis kit (Clontech) and no size selection). Each library had its own dT-BC (Table 2). Briefly, we synthesized the first strand of cDNA with the following thermal profile: 72°C for 3 min, slow ramp to 42°C at 0.1°C/sec and 42°C for 2 min. We synthesized cDNA using the SMARTer PCR cDNA synthesis kit at 42°C for 90 min. We performed a large-scale PCR reaction on each cDNA library with optimization for the numbers of cycle using the KAPA Hi-Fi PCR kit. As determined with a High Sensitivity DNA Chip (Agilent Technologies), we needed 11 to 13 PCR cycles to obtain 250ng of each cDNA. Before capture, we washed cDNA with AMPure PB beads. To enrich for *PHACTRI* transcripts, we designed biotinylated capture probes as Ultramers from IDT DNA according to the manufacturer's protocol (full-length cDNA target sequence capture using SeqCap EZ libraries; Table 2). We used a total of 1.5µg of cDNA (250ng from each barcoded cDNA prep) to capture *PHACTRI* transcripts. We used six different blocking oligos, corresponding to the reverse complement sequence of the dT-BC with a 3' modification (3SpC3) (Table 2). After capture, we amplified the library with 14 cycles of PCR using Takara LA Taq DNA polymerase (Clontech). The long-read PacBio sequencing was performed on one SMRTcell with Pacbio RS II at the Genome Quebec/McGill Innovation Center.

Tableau 2 : List of primers used in this study. The numbers in the “Name” column refers to the primers shown in Figure 8. All sequences are given in the 5’ to 3’ orientation.

Forward primers		Reverse primers	
Name	Sequence	Name	Sequence
RACE	1		gattacccaagcttGGACTCCCTAGAAAATGGGCAGTCCCTG
	2		gattacccaagcttGAGGAGGACGAAGACGACGACAGCTCATTA
	3		gattacccaagcttGACCGCAGGGCAGATAAGCCGTGGACC
	MI3F		GTAAAACGACGGCCAGT
		MI3R	GGAAACAGCTATGACCATG
PacBio	dT-BC1		AAGCAGTGGTATCAACGCAGAGTACtcagacgatgcgtcatTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN,
	dT-BC2		AAGCAGTGGTATCAACGCAGAGTACctatacagactctgcTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN,
	dT-BC3		AAGCAGTGGTATCAACGCAGAGTACtactagagtagcactcTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN,
	dT-BC4		AAGCAGTGGTATCAACGCAGAGTACtgtgtatcagtagcatgTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN,
	dT-BC5		AAGCAGTGGTATCAACGCAGAGTACgatctactatagcTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN,
	dT-BC6		AAGCAGTGGTATCAACGCAGAGTACacagctatactgctgTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTVN
RT-PCR	6		GCGGCATCCTCGGAGGATG
	7		GGTTGCCTCCAATGTCAAAGT
	8		AGGTTTGCCTCCGTTCCG
	10		ATGTATCTGCAAGGGCCGAG
qPCR	12		AGAGAGGCGGATGCATGTG
	8		AGGTTTGCCTCCGTTCCG
	14		ACCATGAAGAGACCCAGTG
	15		GGATGGCCAGATCCTGGCG
	10		ATGTATCTGCAAGGGCCGAG
	TBP_F		CGAATATAATCCAAGCGGTTT
	HPRT_F		TGGCGTCGTGATTAGTGATG
	GapD_H_F		GACAGTCAGCCGCATCTTC
		9	TCACTGGCAGACAAGGCAAT
		9	TCACTGGCAGACAAGGCAAT
		9	TCACTGGCAGACAAGGCAAT
		11	CTCTTGATCTCTCTCTCTCCTCCTG
		13	AGGGAGTCAGAACGCATCG
		13	AGGGAGTCAGAACGCATCG
		9	TCACTGGCAGACAAGGCAAT
		16	CCTGCTGGCCACTCTCTGG
		11	CTCTTGATCTCTCTCTCTCCTCCTG
		TBP_R	GTGGTTCGTGGCTCTCTTATCC
		HPRT_R	CAGAGGGCTACAATGTGATGG
		GapDH_R	GCAACAATATCCACTTACCAGAG

2.1.4 Reverse transcription polymerase chain reaction (RT-PCR)

We extracted RNA from HUVEC, teloHAEC (treated with tumor necrosis factor (TNF)- α at 10ng/ml for 4h), monocytes, HCASMC, HASMC, HCAEC, and HAEC as described above. We analyzed unstimulated monocytes, as well as monocytes stimulated with 10ng/ml of phorbol myristate acetate (PMA) for 48h. Three hours later, some of the PMA-treated monocytes were also treated with lipopolysaccharides (LPS) (100ng/ml) for one hour. We used the same adult brain total RNA and adult heart total RNA as for the RACE (see above). We measured RNA quality and concentration with

Agilent RNA 6000 Nano II assays (Agilent Technologies) on an Agilent 2100 Bioanalyzer. We reverse transcribed 1µg or less of total RNA (with RNA integrity number of 10 for all cells and above 7 for all tissues) using random primers and 1U of the MultiScribe Reverse Transcriptase (Applied Biosystems) in a 20µL reaction volume at 100mM dNTPS and 20U of Rnase inhibitor with these three steps: 10 min at 25°C, 120 min at 37°C and 5 min at 85°C. We mixed an equal volume of cDNA from each sample to create a pool of cDNA to use as positive control. We amplified fragments of *PHACTRI* transcripts from these tissues and cell lines with GoTaq DNA polymerase (Promega) with the following thermal cycle: 95°C for 2 min; 94°C for 30s, optimal temperature for 30s and 72°C for 1 min and 20s (35 times); 72°C for 5 min. We also amplified full *PHACTRI* open reading frame (ORF) with SeqAmp DNA polymerase (Clontech) with the following thermal cycle: 94°C for 1 min; 98°C for 10s, optimal temperature for 15s and 68°C for 3 min (30 times); 68°C for 5 min. The sequences of the primers are in Table 2. We characterised the PCR products by agarose gel electrophoresis.

2.1.5 *Quantitative polymerase chain reaction (qPCR)*

(The extractions were made by Mélissa Beaudoin or at the Pharmacogenomics Centre)

We extracted DNA and RNA from 36 hCA samples from the RÉTEB. Of these 36 samples, 14 are new and 22 were previously analyzed (53). Genotyping of *PHACTRI*-rs9349379 was performed at the Beaulieu-Saucier Pharmacogenomics Centre of the Montreal Heart Institute on the Illumina Infinium MEGA_Consortium_v2 BeadChip: nAA = 15, nAG = 13, nGG = 8. We measured RNA integrity and concentration with Agilent RNA 6000 Nano II assays (Agilent Technologies) on an Agilent 2100 Bioanalyzer. We reverse transcribed exactly 1µg of total RNA as for the RT-PCR experiments (with RNA integrity number of 6 or above for all samples). We followed the MIQE guidelines to assess quality and reproducibility of our qPCR results (92). We performed qPCR in triplicates for all samples using: 1.25µL of cDNA (1/50 dilution), 5µL of Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies) and 3.75µL of primer pair mix at 0.8µM on a CFX384 from Biorad or Eco Illumina qPCR system (Montreal Biotech). We used the following thermal profile: 10 min at 95°C, and 40 cycles of 30s at 95°C, 30s at 55°C and 45s at 72°C. We carried out melting curve analyses after the amplification process to ensure the specificity of the amplified products. We also simultaneously performed qPCR reactions with no template controls for each gene to test the absence of non-specific products. Cq values were determined with the CFX Manager 3.1 (Bio-Rad) software and expression levels were normalized on the expression levels of the housekeeping genes TATA-box binding protein

(TBP), hypoxanthine-guanine phosphoribosyltransferase (HPRT), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) using the $\Delta\Delta C_t$ method. Based on geNORM principles for accurate normalization of qPCR data by geometric averaging of multiple internal control genes (93), mean M values of 0.5845 and 0.5356 (for two experiments) were generated from the GAPDH, HPRT1, and TBP genes. The primer sequences are in Table 2.

2.1.6 eQTL analyses using GTEx data

Tableau 3 : GTF file. Genomic coordinates of exons based on the reference human genome sequence GRCh37/hg19.

seqname	source	feature	start	end	score	strand	frame	attribute
6	Ensembl	exon	12957636	12957776	.	+	.	transcript_id "with"; exon_id "06"
6	Ensembl	exon	12958214	12958269	.	+	.	transcript_id "with"; exon_id "07"
6	Ensembl	exon	13013852	13014371	.	+	.	transcript_id "with"; exon_id "78"
6	Ensembl	exon	13053597	13053761	.	+	.	transcript_id "with"; exon_id "08"
6	Ensembl	exon	13160436	13160516	.	+	.	transcript_id "with"; exon_id "09"
6	Ensembl	exon	13182751	13182918	.	+	.	transcript_id "with"; exon_id "10"
6	Ensembl	exon	13185031	13185237	.	+	.	transcript_id "with"; exon_id "1011"
6	Ensembl	exon	13206047	13206368	.	+	.	transcript_id "with"; exon_id "11"
6	Ensembl	exon	13228048	13228295	.	+	.	transcript_id "with"; exon_id "12"
6	Ensembl	exon	13230269	13230425	.	+	.	transcript_id "with"; exon_id "13"
6	Ensembl	exon	13273092	13273147	.	+	.	transcript_id "with"; exon_id "14"
6	Ensembl	exon	13278500	13278561	.	+	.	transcript_id "with"; exon_id "15"
6	Ensembl	exon	13283654	13283794	.	+	.	transcript_id "with"; exon_id "16"
6	Ensembl	exon	13286378	13286454	.	+	.	transcript_id "with"; exon_id "17"
6	Ensembl	exon	13287295	13288073	.	+	.	transcript_id "with"; exon_id "18"
6	Ensembl	exon	12957636	12957776	.	+	.	transcript_id "without"; exon_id "06"
6	Ensembl	exon	12958214	12958269	.	+	.	transcript_id "without"; exon_id "07"
6	Ensembl	exon	13013852	13014371	.	+	.	transcript_id "without"; exon_id "78"
6	Ensembl	exon	13053597	13053761	.	+	.	transcript_id "without"; exon_id "08"
6	Ensembl	exon	13160436	13160516	.	+	.	transcript_id "without"; exon_id "09"
6	Ensembl	exon	13182751	13182918	.	+	.	transcript_id "without"; exon_id "10"
6	Ensembl	exon	13206047	13206368	.	+	.	transcript_id "without"; exon_id "11"
6	Ensembl	exon	13228048	13228295	.	+	.	transcript_id "without"; exon_id "12"
6	Ensembl	exon	13230269	13230425	.	+	.	transcript_id "without"; exon_id "13"
6	Ensembl	exon	13273092	13273147	.	+	.	transcript_id "without"; exon_id "14"
6	Ensembl	exon	13278500	13278561	.	+	.	transcript_id "without"; exon_id "15"
6	Ensembl	exon	13283654	13283794	.	+	.	transcript_id "without"; exon_id "16"
6	Ensembl	exon	13286378	13286454	.	+	.	transcript_id "without"; exon_id "17"
6	Ensembl	exon	13287295	13288073	.	+	.	transcript_id "without"; exon_id "18"
6	Ensembl	exon	13272908	13273091	.	+	.	transcript_id "short"; exon_id "av14"
6	Ensembl	exon	13273092	13273147	.	+	.	transcript_id "short"; exon_id "14"
6	Ensembl	exon	13278500	13278561	.	+	.	transcript_id "short"; exon_id "15"
6	Ensembl	exon	13283654	13283794	.	+	.	transcript_id "short"; exon_id "16"
6	Ensembl	exon	13286378	13286454	.	+	.	transcript_id "short"; exon_id "17"
6	Ensembl	exon	13287295	13288073	.	+	.	transcript_id "short"; exon_id "18"

We downloaded from dbGaP genotypes at *PHACTR1*-rs9349379 and RNA-sequencing (RNA-seq) data from 122 hCA samples generated by the GTEx Project (81). We used Trimmomatic (V0.36) on reads with the paired-end (PE) function to remove adapter sequences (TruSeq3-PE.fa:2:30:10) and to trim low-quality bases with the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. We then performed quality control checks using FastQC (V0.11.5). We aligned reads to the reference human genome sequence (GRCh37) using HISAT2 (V2.0.5) (94). We selected reads that mapped to the *PHACTR1* locus using BamTools (V2.4.1) with the parameters `-region "6:12717037..13288073" -mapQuality ">=30"`, and used SortSam and BuildBamIndex tools from Picard (V1.130) to convert .sam files into .bam files and to build their index files. We quantified *PHACTR1* expression with Stringtie (V1.3.3b) (94) with parameters `-G -e -b`; for these analyses we provided our own GTF file with genomic coordinates corresponding to the three *PHACTR1* transcripts expressed in hCA (Table 3). Finally, we used Ballgown (V2.8.4) (94) to extract exon-level expression data for eQTL analyses with genotypes at *PHACTR1*-rs9349379.

2.1.7 Statistical analyses

We normalized *PHACTR1* expression data (for full transcripts or exons) from our qPCR experiments or GTEx using inverse normal transformation. Using linear regression in R, we tested the association between *PHACTR1* expression levels and genotypes at *PHACTR1*-rs9349379 (additive coding), correcting for sex, RNA integrity number (RIN), and the recruitment center (Montreal or Quebec city) when available.

2.1.8 Western blot

(This section was made by Simon Lalonde)

To detect *PHACTR1* protein expression in VSMC, 0.75×10^6 and 1×10^6 of cryopreserved HASMC and HCASMC cells, respectively, were pelleted and directly lysed in RIPA buffer (50mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% Na-deoxycholate, 0.1% SDS, 150mM NaCl) containing protease inhibitor cocktail (Sigma), phosphatase inhibitor cocktail 2 and 3 (Sigma), and 1mM PMSF. As controls, teloHAEC were transfected with 20nM of scramble (Santa Cruz, sc-37,007) or *PHACTR1* (Santa Cruz, sc-95,456) small interfering RNA (siRNA) in 2mL of siRNA transfection medium (Santa Cruz, sc-36,868) for 4h using siRNA Transfection Reagent (Santa Cruz, sc-29,528), and then grown for 48h before cell lysis and protein extraction. Homogenized lysates were clarified by centrifugation to

eliminate insoluble cell debris. We determined protein concentration with a Pierce BCA Protein Assay Kit (ThermoFisher) and prepared samples containing 30µg of proteins in a 600mM DTT reducing buffer that was denatured 5 min at 95°C prior to loading on a 8% polyacrylamide gel (running buffer: 25mM Tris base; 192mM glycine; 0.1% SDS). We transferred proteins onto nitrocellulose membranes in Towbin buffer (25mM Tris base; 192mM glycine, 20% (v/v) methanol) and incubated them for 1.5h in the blocking buffer (1× TBS, 0.1% Tween-20, 5% milk) at room temperature. We then incubated membranes with primary antibodies against PHACTR1 (1:1000; custom antibody generated by Biomatik) or GAPDH (1:10,000; Cell Signaling, #2118) at 4°C overnight. For the secondary antibody, we incubated membranes with horseradish peroxidase-conjugated anti-rabbit IgG (1:10,000; GE Healthcare, NA934) at room temperature for 1h. The immunoblotting signal was revealed with the SuperSignal west Pico PLUS chemiluminescent substrate (Thermo scientific, 34,580) and visualized with the ChemiDoc Touch system (Bio-Rad).

2.2 - Results

2.2.1 Six *PHACTR1* transcripts in human cells and tissues

Public databases (e.g. ENSEMBL, Gencode, GTEx) report > 10 different *PHACTR1* mRNA transcripts expressed in human samples. To confirm these results and comprehensively characterize the different *PHACTR1* transcripts expressed in humans, we combined RACE and long-read next-generation (PacBio) DNA sequencing experiments in many cell types and tissues. These two complementary approaches generated highly concordant results that we also validated by sequencing of the whole ORF. Although we did identify a large number of *PHACTR1* transcripts due to alternative splicing of non-coding 5' and 3' untranslated exons (Annexe 1), our transcriptomic profiling indicates that *PHACTR1* can give rise to six main different transcripts that would encode six proteins based on their coding sequences (Figure 8, Tableau 9). The long transcript (1743bp) encodes a protein of 580 amino acids (Figure 8). We confirmed four intermediate transcripts that are different due to the inclusion of the alternatively spliced exons 7.8 and 10.11 (Figure 8). Both alternative exons are in-frame and add respectively 279- (exon 7.8) and 207-bp (exon 10.11) to the ORF. Intermediate transcripts A+ (1953bp) and B+ (1674-bp) include exon 10.11 and encode proteins of 650 and 557a.a.. Intermediate transcripts A- (1746bp) and B- (1467bp) exclude exon 10.11 and encode proteins of 581 and 488a.a.. Finally, we confirmed a short transcript (435bp) encoding a protein of 144aa (Figure 8). Its start codon is located in exon 14, 83-bp upstream of the 3' splice site for the long and intermediate transcripts (Figure 9).

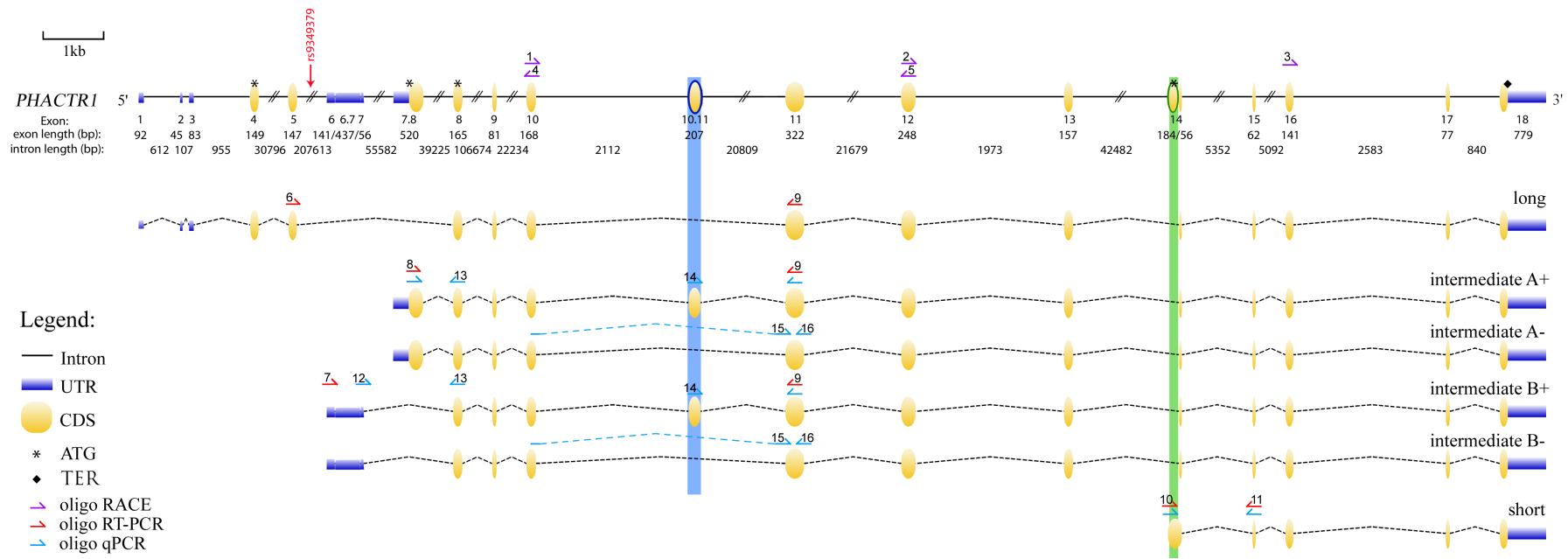


Figure 8 : Six *PHACTR1* transcripts are expressed in human samples. We combined Rapid Amplification of cDNA Ends (RACE) and long-read DNA sequencing (Pacific Biosciences) to identify all *PHACTR1* mRNA transcripts expressed in human samples. Although we identified multiple transcripts, they give rise to six main transcripts. The schematic diagram shows the *PHACTR1* gene (top) and the six main transcripts below. The coronary artery disease-associated SNP rs9349379 (red arrow) is located in a large intron (207,613-bp) between exons 5 and 6. The start codons for the long, intermediate A, intermediate B and short transcripts are located, respectively, in exons 4, 7.8, 8, and 14. The stop codon for all six transcripts is in exon 18. In blue, we highlight the in-frame, 207-bp long, exon 10.11 that is included in intermediate transcripts A+ and B+. In green, we highlight part of exon 14 that is specific to the short transcript. We also indicated the different oligonucleotides used in our experiments; numbers referred to their sequences in Table 2. UTR, untranslated transcribed region; CDS, coding DNA sequence; ATG, start codon; TER, terminator codon. The figure was drawn to scale using GSDS2.0 (95).

Tableau 4 : Transcript and protein IDs of *PHACTR1* gene (221692; ENSG00000112137).

	Transcript		Protein	
	Length (bp)	IDs	Length (aa)	IDs
Long	1743	NM_030948.3, ENST00000332995.11	580	NP_112210.1
Intermediate A-	1746	XM_017010464.1	581	XP_016865953.1
Intermediate A+	1953	NM_001322314.1	650	NP_001309243.1
Intermediate B-	1467	NM_001322312.1	488	NP_001309241.1
Intermediate B+	1674	NM_001322313.1	557	NP_001309242.1
Short	435	ENST00000379335.7	144	EAW55320.1

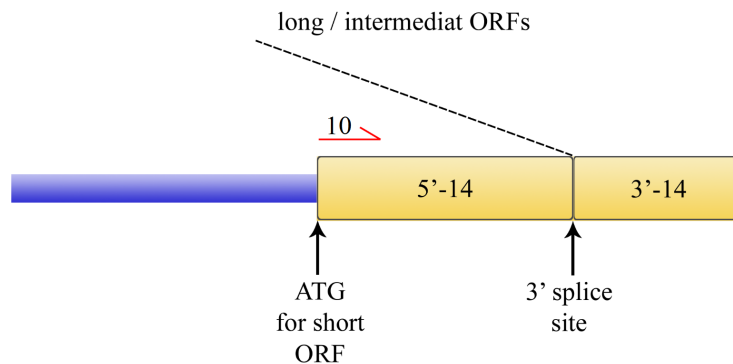


Figure 9 : Zoom-in representation of *PHACTR1* exon 14. The 5'-14 fraction of exon 14 is specific to the short immune-specific *PHACTR1* transcript. We also illustrate the 3' splice site used for the long and intermediate *PHACTR1* transcripts. In blue is the intron between exons 13 and 14. The red arrow indicates a primer used to detect the short *PHACTR1* transcript. The sequence of this primer #10 is in Table 2.

2. 2. 2 *PHACTR1* transcripts are differentially expressed

We used the sequence of the six *PHACTR1* transcripts to design transcript-specific pairs of primers and carried out RT-PCR experiments on mRNA extracted from human brain, heart, monocytes, ECs, and VSMCs (Figure 10). The long transcript, which never included exons 7.8 nor 10.11, was only expressed in the brain (60). In contrast, both intermediate transcripts B were expressed in all samples tested, although we noted that the inclusion of exon 10.11 varied from one sample to the other. Intermediate transcript B+ was more abundant in ECs and VSMCs, whereas intermediate transcript B-

was the major intermediate transcript expressed in the brain, the heart, and in monocytes (Figure 10, third panel from top). Both intermediate transcripts A were expressed in the brain, the heart, untreated monocytes, and ECs. Similar to what we saw for the intermediate transcripts B, the inclusion of exon 10.11 in the intermediate transcripts A was variable from one tissue to the other (Figure 10, second panel from top). We validated that the short transcript was expressed in monocytes (61). Stimulation with PMA or LPS to differentiate monocytes into macrophages, which are of interest regarding atherosclerosis, did not induce the expression of the long transcript, but seems to weakly increase the expression of intermediate transcripts B (Figure 10). We detected expression of the short transcript in the heart and the brain, although this is consistent with the presence of immune cells within these tissues.

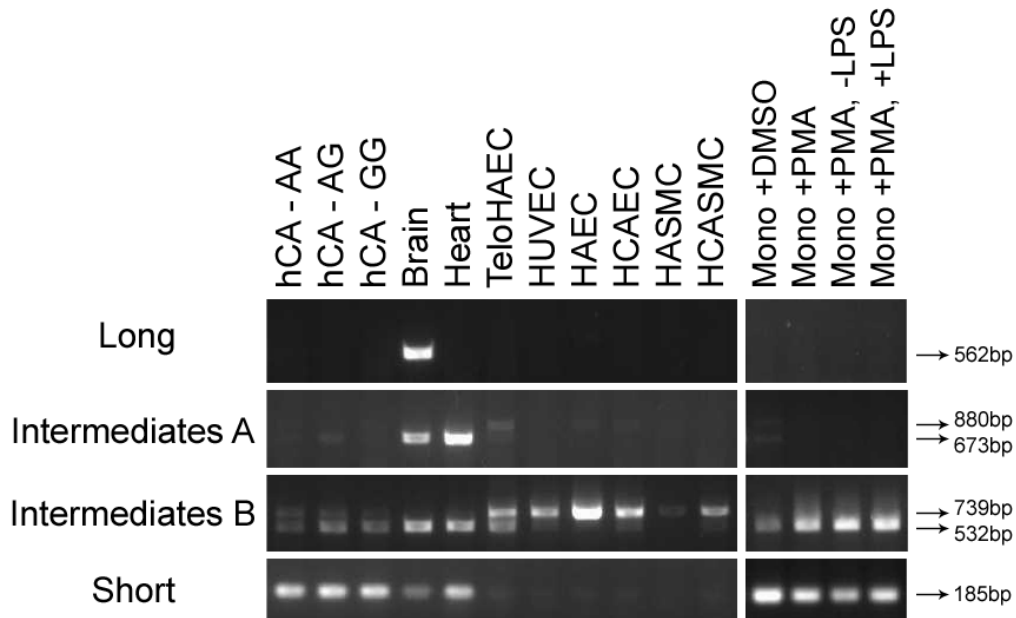


Figure 10 : *PHACTR1* transcripts differential expression in human tissues and cell types. We designed transcript-specific primers and amplified *PHACTR1* by RT-PCR. The PCR fragment for the long transcript is 562bp. We did not amplify a 769bp fragment using this set of primers, indicating that exon 10.11 is absent from the long transcript. The PCR fragments for the intermediate transcripts A+, A-, B+ and B- are 880bp, 673bp, 739bp and 532bp, respectively. The PCR fragment for the short transcript is 185bp. HUVEC: human umbilical vein endothelial cell; teloHAEC: immortalized human aortic endothelial cells; HCASMC: human coronary artery smooth muscle cells; HAEC: human aortic endothelial cells; HCAEC: human coronary artery endothelial cells; HASMC: human aortic smooth muscle cells; Mono: monocytes; PMA: phorbol myristate acetate; LPS: lipopolysaccharide.

Although *PHACTR1* transcripts had previously been detected in VSMCs, immunoblotting with a commercially available anti-PHACTR1 antibody failed to detect the corresponding protein (61). We purchased and tested all commercially available anti-PHACTR1 antibodies, but none showed specificity in short interfering RNA (siRNA)-mediated knockdown experiments. For this reason, we developed our own polyclonal anti-PHACTR1 antibody (Materials and methods) and have shown its specificity using siRNA against *PHACTR1* (96). Using this antibody, we detected weak PHACTR1 protein expression in VSMCs from hCA (Figure 11).

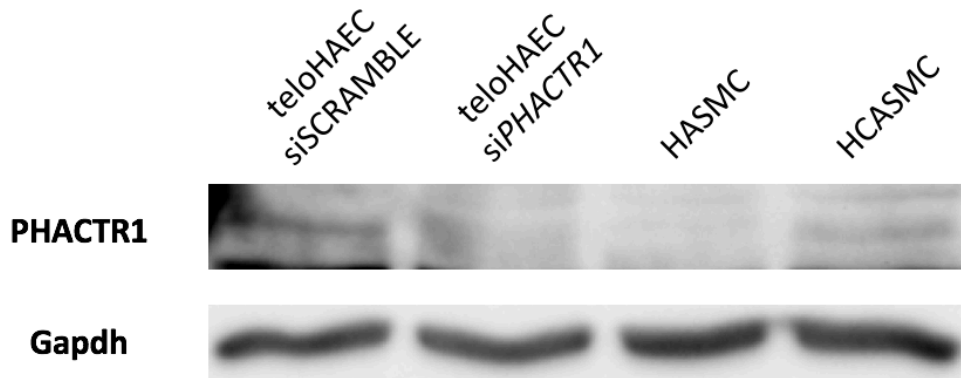


Figure 11 : Presence of PHACTR1 protein in vascular smooth muscle cells. We tested for the presence of the PHACTR1 protein by immunoblotting using a custom anti-PHACTR1 antibody. The first and second lane, using teloHAEC protein extracts treated or not with a short interfering RNA (siRNA) against *PHACTR1*, serve as control to demonstrate the specificity of this anti-PHACTR1 antibody. Although we did not detect PHACTR1 in human aortic smooth muscle cells (HASMC), there is weak but specific band corresponding to PHACTR1 in human coronary artery smooth muscle cells (HCASMC). The expected 64 kDa PHACTR1 protein migrated with an apparent molecular weight of 74kDa. GAPDH was used as loading control (Figure made by Simon Lalonde)

2. 2. 3 *CAD-associated rs9349379 is an eQTL for PHACTR1 intermediate transcripts A*

Using primers that capture the long and intermediate *PHACTR1* transcripts, we previously showed that genotypes at rs9349379 are associated with *PHACTR1* expression levels in hCA (53). That result was subsequently replicated in the GTEx database. Given that hCA are composed of different cell types (ECs, VSMCs, and immune cells) and that *PHACTR1* transcripts are differentially expressed across tissues and cell types, we tested if rs9349379 is associated with the levels of all or specific *PHACTR1* transcripts. We obtained DNA and RNA from 36 hCA, including 14 new samples not analyzed in our previous study (53). We genotyped rs9349379 and measured by qPCR the levels of each of the five

PHACTR1 transcripts expressed in hCA. We found a significant association between rs9349379 and intermediate transcripts A+ and B+, whereas the SNP was not significantly associated with intermediate transcripts A- and B- (Table 5, Figure 12). Although we detected strong expression of the short *PHACTR1* transcript in hCA, genotypes at rs9349379 were not associated with its levels (Table 5, Figure 12).

Tableau 5 : rs9349379 is an eQTL for *PHACTR1* intermediate transcripts A+ and B+ in human coronary arteries (hCA). We tested by linear regression the association between genotypes at rs9349379 (additive model) and normalized expression levels for the different *PHACTR1* transcripts expressed in hCA. The long *PHACTR1* transcript is not expressed in hCA. Effect sizes (Beta) and standard errors (SE) are in standard deviation units. The direction of the Beta is for the G-allele. We analyzed 36 hCA samples (NAA=15, NAG=13, NGG=8). The frequency of the G-allele was 40%. The G-allele, associated with increased coronary artery disease risk, is associated with reduced *PHACTR1* expression. When significant (P-value < 0.05), we provided the percentage of *PHACTR1* expression variation explained by genotypes at rs9349379.

transcripts	Beta (SE)	P-value	Variance explained (%)
Intermediates A+ and A-	-0.49 (0.21)	0.026	7.8
Intermediates B+ and B-	-0.54 (0.17)	0.0040	17.5
Intermediates A+ and B+	-0.56 (0.19)	0.0073	16.4
Intermediates A- and B-	-0.35 (0.21)	0.10	-
Short	-0.068 (0.20)	0.73	-

Because our hCA sample size remains small, we attempted to replicate these eQTL results in the larger GTEx dataset with genotype and gene expression data available from 122 hCA. Since it is computationally challenging to assemble specific mRNA transcripts out of short-read RNA-seq results, we opted to test the association between rs9349379 and the expression levels of specific *PHACTR1* exons. Overall, we found nominally significant associations (P-value < 0.05) with exons common to the intermediate and short transcripts (Table 6). However, the association between rs9349379 and the expression levels of the 5' end of exon 14 that is specific to the short transcript (Figure 9) was not significant (P-value = 0.47) (Table 6, Figure 12). Thus, using qPCR and RNA-seq gene expression data, we could not detect an eQTL effect between rs9349379 and the expression of the immune-specific short *PHACTR1* transcript in 158 hCA samples.

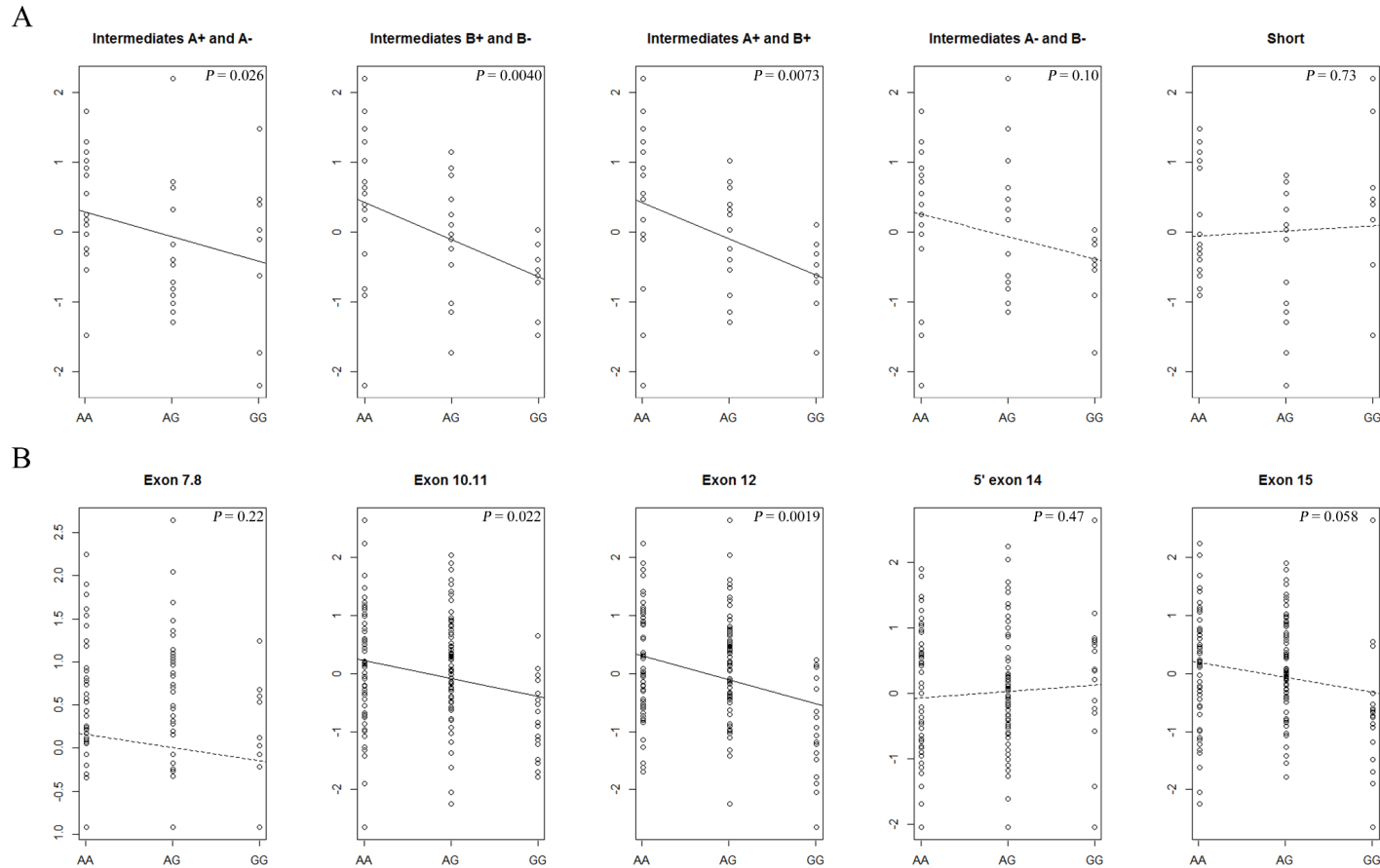


Figure 12 : Associations between genotypes at rs9349379 and *PHACTR1* expression levels in human coronary arteries. (A) By qPCR using transcript-specific primers, we measured the expression of *PHACTR1* transcripts in 36 human coronary arteries (hCA)(NAA=15, NAG=13, NGG=8). The long *PHACTR1* transcript is not expressed in hCA. (B) We used GTEx data to test the associations between rs9349379 and the expression levels of *PHACTR1* exons in 122 hCA (NAA=48, NAG=57, NGG=17). Here, we only show results for four exons, but association results for all *PHACTR1* exons are available in Table 2. Suppl. exon corresponds to the alternatively spliced exon located between exons 10 and 11. 5' exon 14 corresponds to part of exon 14 that is specific to the short *PHACTR1* transcript.

Tableau 6 : Replication of rs9349379 eQTL effects in human coronary arteries (hCA) from GTEx. We analysed by linear regression the association between genotypes at rs9349379 (additive model) and normalized levels of *PHACTR1* exon expression. Expression levels were corrected for the length of the exons (Materials and methods). Effect sizes (Beta) and standard errors (SE) are in standard deviation units. The direction of the effect size is for the G-allele, associated with increased coronary artery disease risk. We analyzed data from 122 hCA samples (NAA=48, NAG=57, NGG=17). The frequency of the G-allele was 37%. When significant (P-value < 0.05), we provided the percentage of *PHACTR1* exon expression variation explained by genotypes at rs9349379. ^aThe full exon 14; ^bpart of exon 14 which is specific to the short *PHACTR1* transcript; ^cpart of exon 14 which is present in all *PHACTR1* transcripts (Figure 9).

Exon	Exon length (bp)	Beta (SE)	P-value	Variance explained (%)
6 (5'UTR)	141	-0.084 (0.13)	0.51	-
7 (5'UTR)	56	-0.11 (0.13)	0.39	-
7.8	520	-0.15 (0.12)	0.22	-
8	165	-0.31 (0.13)	0.017	3.4
9	81	-0.19 (0.13)	0.14	-
10	168	-0.35 (0.13)	0.0065	3.1
10.11	207	-0.30 (0.13)	0.022	3.9
11	322	-0.35 (0.13)	0.0074	2.8
12	248	-0.40 (0.13)	0.0019	5.0
13	157	-0.28 (0.13)	0.029	3.0
14 ^a	240	-0.17 (0.13)	0.20	-
5'-14 ^b	184	0.095 (0.13)	0.47	-
3'-14 ^c	56	-0.25 (0.13)	0.059	-
15	62	-0.25 (0.13)	0.058	-
16	141	-0.27 (0.13)	0.042	2.3
17	77	-0.31 (0.13)	0.017	3.5
18 (3'UTR)	779	-0.25 (0.13)	0.055	-

2.3 - Discussion

Despite the recent report linking *PHACTR1*-rs9349379 with *EDNI* in ECs (54), *PHACTR1* itself remains a strong causal candidate gene at this CAD-associated locus because of the eQTL effect in hCA (53, 81). Further, it was recently shown in an in vitro cell culture system that *PHACTR1* contributes to VSMC calcification, a hallmark of CAD (97). Using two different methods, we confirmed that the human *PHACTR1* gene encodes six main transcripts that are differentially expressed. Although our results are largely consistent with a study by Reschen et al. (61), we note three important differences: First, we confirmed that there are six, and not three, *PHACTR1* transcripts. Indeed, whereas Reschen and colleagues reported only one intermediate transcript (transcript B+ in Figure 8), we found that four intermediate transcripts (with or without exons 7.8 and 10.11) are co-expressed in the heart (Figure 10). Whether these four *PHACTR1* isoforms, of 650, 581, 557 and 488 amino acids, have the same functions is currently unknown. Second, we confirmed using our own custom polyclonal antibody against *PHACTR1* that the protein is present in VSMCs. In contrast, we have not been able to detect by immunoblotting the protein encoded by the short transcript. And third, our eQTL analyses in 158 hCA validated an effect of genotypes at rs9349379 on the expression of intermediate transcripts A+ and B+, but not the short immune-specific transcript. One possible interpretation of this result is that rs9349379 mediates in part its effect on CAD risk by regulating the expression or splicing of these *PHACTR1* intermediate transcripts in ECs and/or VSMCs within hCA, and not through the short immune-specific *PHACTR1* transcript. In support of this hypothesis, we also know that rs9349379 is associated with fibromuscular dysplasia (40), a disease without an inflammatory component. Precisely how, at the molecular level, genotypes at rs9349379 modulate cell type-specific transcriptional expression and/or splicing remains to be determined.

Our study presents with some limitations. As for every transcriptomic profiling experiments, weakly expressed transcripts might have missed detection. We tried to account for this by combining two complementary methods and profiling *PHACTR1* transcripts in high quality mRNA prepared from many tissues and cell types. The number of available hCA samples for our eQTL analyses was small (N = 36). To increase statistical power, we also analyzed RNA-seq data from the GTEx Project (N = 122). Because short-reads cannot be used unambiguously to reconstruct specific transcripts, we decided to analyze the expression levels of specific exons. This analysis generated results that were highly concordant with our transcript-specific qPCR results. Finally, although we did not detect an eQTL effect for the short *PHACTR1* transcript in hCA, it does not rule out a potential role for *PHACTR1* in

immune cells in the context of atherosclerosis progression and CAD. Indeed, the short *PHACTR1* protein might play a critical atherosclerotic function in circulating monocytes, which would not have been captured in hCA samples. However, it is critical to confirm that the short *PHACTR1* transcript can be translated into an active protein in human cells.

2.4 - Conclusions

In this article, we provide a comprehensive survey of *PHACTR1* transcripts in tissues and cells that are relevant to atherosclerosis initiation and progression. Importantly, we confirmed that the *PHACTR1* protein is present in VSMCs and that the CAD-associated variant rs9349379 associates with the expression of specific *PHACTR1* transcript in hCA. Gupta and colleagues have argued that the molecular link they identified between rs9349379 and *EDNI* expression largely accounts for the association of this locus with several vascular diseases (54). One natural extension of this model, which would be consistent with the pleiotropy observed as well as the *PHACTR1* eQTL we detected in hCA, is that the same SNP influences the expression of two genes – *EDNI* and *PHACTR1* – which are expressed in CAD-relevant cells and could act together to contribute to vascular disease risk.

3 - Inactivation de *PHACTR1* par le système CRISPR-Cas9

Il a été démontré que *PHACTR1* jouerait un rôle dans les MCVs de par le fait que plusieurs SNPs à ce loci sont associés avec ce phénotypes et plusieurs autres connexes, que rs9349379 est un eQTL dans les EC, et que rs9349379 est également un QTL pour l'inclusion de l'exon 10.11 par épissage alternatif dans les EC. Toutefois, bien que le lien de causalité entre les MCVs et *PHACTR1* semble présent, le mécanisme par lequel ce gène modulerait le risque de maladies demeurent inconnus. Maintenant que les transcrits de *PHACTR1* et leur expression au niveau du cœur sont caractérisés, il est possible de pousser l'analyse plus loin pour déterminer l'implication de ce gène dans la maladie. Pour ce faire, il est possible d'observer l'effet d'une perte de fonction par édition du génome et puis d'étudier la réponse phénotypique par la réexpression des différents transcrits individuellement pour voir la modulation exercée par chacun. Des cellules teloHAEC éditées par le système CRISPR-Cas9 ont été utilisées à cette fin. À noter que deux séries de clones ont été réalisées, une première par l'insertion d'indels dans les exons 8, 9 et 10 (96) puis une seconde avec des indels dans l'exon 11. Tous ceux-ci ayant des indels homozygotes confirmées au niveau génomique et donc inactivant théoriquement *PHACTR1*, seule la seconde édition de l'exon 11 semble fonctionnelle en pratique.

Les teloHAEC constituent une lignée de cellules immortalisées exprimant de manière stable la sous-unité catalytique hTERT (human telomerase reverse transcriptase) de la télomérase humaine. Bien que celle-ci soit modifiée et donc sans sénescence, elle maintient un phénotype normal. En effet, le modèle cellulaire a été validé. Il a été démontré qu'elles possèdent un karyotype diploïde normal, expriment le marqueur CD31 (cluster of differentiation 31 ou Platelet endothelial cell adhesion molecule (PECAM 1)), incorporent du AcLDL, montrent une réponse inflammatoire normale avec un traitement au TNF α , augment leur prolifération avec une stimulation au VEGF et forment des tubules lors de l'angiogenèse in vitro. Ce modèle a ses imperfections étant donné que les cellules sont moins représentatives de ce qui est biologiquement observé. Toutefois, elles s'avèrent indispensables pour les expériences d'édition de génome. En effet, cela requiert plusieurs dédoublements de population pour parvenir à ce que les cellules expriment stablement l'enzyme Cas-9, à éditer leur génome puis à dériver des clones de sorte que cela serait difficilement réalisable dans une lignée cellulaire primaire.

Le TNF α est une molécule proinflammatoire couramment utilisée pour activer les cellules endothéliales et représente donc un bon moyen pour étudier la dysfonction endothéliale par comparaison des cellules stimulées et non stimulées au TNF α . Les temps d'exposition sont variables dans la littérature. Un test avec plusieurs temps a été réalisé pour déterminer le temps à utiliser (Annexe 2). Afin de causer une dysfonction endothéliale, les teloHAEC sont soumises à un traitement au TNF α pendant 0, 4 ou 24 heures, soit le contrôle, la réponse maximale et le retour vers la normale respectivement. Les échantillons sont faits en *duplicata* pour s'assurer que les signaux observés sont réels biologiquement et ne proviennent pas de la technique ni de l'échantillon.

3.1 - Méthodes

3.1.1 Culture cellulaire

Les EC aortiques humaines immortalisées (Immortalized human aortic endothelial cells, teloHAEC ; ATCC, CRL-4052) sont maintenues en culture à 5% CO₂ et 37°C dans le milieu VCBM (vascular cell basal media ; ATCC, PCS-100-030) avec des suppléments (endothelial cell growth kit-VEGF ; ATCC, PCS-100-041) et des antibiotiques contre les bactéries (200 U/mL pénicilline et 200 μ g/mL of streptomycine ; ThermoFisher, 15140122) et passées avant d'atteindre une confluence de 90%.

3.1.2 Édition du génome par le système CRISPR-Cas9

(Section réalisée par Simon Lalonde)

Des paires de sgrNA ciblant les trois premiers exons (exons 8, 9 et 10) de *PHACTRI* (Tableau 7) ont été modélisées, puis de nouvelles paires ciblant l'exon 11 de *PHACTRI* par un outil en ligne CRISPR Design (version 1, Zhang Lab, MIT 2013, <http://crispr.mit.edu/>), générées par IDT DNA puis clonées par digestion avec BbsI-HF (NEB, R3539) et ligation avec T4 DNA ligase (NEB, M0202) dans le vecteur pCas9-Néo (vecteur pSpCas9(BB)-2A-Puro (PX459) V2.0 (<https://www.addgene.org/62988>) dont la résistance a été remplacée par la séquence synthétique de résistance à la néomycine (GeneArt de ThermoFisher)). La séquence des vecteurs a été validée par séquençage de type Sanger. Les cellules préalablementensemencées à 2x10⁵ cellules par puits dans des 6 puits sont transfectées durant quatre heures dans le milieu Opti-MEM I Reduced Serum media (ThermoFisher, 31985062) avec 0,5 ou 1 μ g de vecteur avec le transfectant Lipofectamine 3000 Transfection Reagent (ThermoFisher, L3000001). Les cellules sont ensuite retournées dans du milieu VCBM et mises sous sélection positive 48h post-

transfection pour l'intégration du vecteur avec G418 (généticine, analogue à la néomycine, Life technologies, 10131035) à 200µg/mL pour 7 jours, soit jusqu'à la mortalité des cellules non transfectées, puis à 100µg/mL par la suite pour maintenir la sélection sans nuire à la viabilité. Une validation des indels générées est réalisée par une amplification par PCR avec des amorces encadrant le site de coupure de la Cas9 puis par séquençage Sanger et décomposition des chromatogrammes d'un produit non pur par l'outil en ligne TIDE (Tracking of Indels by Decomposition ; version 2.0.1 ; <https://tide.deskgen.com>) (98).

Tableau 7 : Séquences des sgRNA utilisés pour l'édition de génome par le système CRISPR-Cas9

Exon	sgRNA sens	sgRNA anti
8	5'-caccgAACTTACTTCTCCTGCGGAT-3'	5'-aaacATCCGCAGGAGAAGTAAGTTc-3'
8	5'-caccgTGGGAAGGATTTTCAAGCCT-3'	5'-aaacAGGCTTGAAAATCCTTCCCAc-3'
9	5'-caccGATAAAGCGAGGAGTCCTGA-3'	5'-aaacTCAGGACTCCTCGCTTTATC-3'
10	5'-caccGACTGGCTCCATTTCGGACA-3'	5'-aaacTGTCCGAAATGGAGCCAGTC-3'
10	5'-caccgCAGGGATCCCTGCTCATATG-3'	5'-aaacCATATGAGCAGGGATCCCTGc-3'
11 (Sg1)	5'-caccgGTCATGATCTGTATGCCCGT-3'	5'-aaacACGGGCATACAGATCATGACc-3'
11 (Sg2)	5'-caccgAGACGAGCTCAACAAAACGC-3'	5'-aaacGCGTTTTGTTGAGCTCGTCTc-3'
11 (Sg3)	5'-caccgCATGATCTGTATGCCCGTGG-3'	5'-aaacCCACGGGCATACAGATCATGc-3'
Non-spécifique	5'-caccgCCGATATCGCGATCGTCGGT-3'	5'-aaacACCGACGATCGCGATATCGGc-3'

3.1.3 Dériver des populations en clones

(Section réalisée par Simon Lalonde)

Les cellules sont passées dans des 96 puits à une confluence de 0,3 cellule par puit afin d'éviter d'avoir plus d'une cellule par puit. Les puits contenant une cellule sont mis en expansion pour environ deux semaines ou jusqu'à l'obtention d'une bonne confluence. À la suite de quoi elles sont passées en faisant un rinçage avec 25µl de trypsine, puis en les décollant avec 25µl de trypsine de deux à trois minutes, en

les reprenant dans 180µl de VCBM pour les transférer en 48 puits dans lequel 100µl de milieu est ajouté. À nouveau, elles sont mises en expansion pour environ 1 semaine, puis sont passées en 6 puits de la même façon (Rinçage avec 60µl trypsine ; 60µl trypsine 2-3 minutes + 460µl de milieu transféré dans un 6 puits ; ajout de 1,5ml de milieu). Lorsque les 6 puits sont rendus à confluence, elles sont passées (Rinçage avec 300µl trypsine ; 300µl trypsine 2-3 minutes + 2ml de milieu ; centrifugation et reprise dans 500µl de milieu) et une portion de 100 000 cellules est conservée pour l'extraction d'ADN génomique à des fins de caractérisation et le restant est congelé.

3.1.4 Validation des indels générées par CRISPR-Cas9

(Section réalisée par Simon Lalonde)

L'ADN génomique est extrait de 1×10^5 cellules avec le kit « QuickExtract DNA Extraction Solution » (Epicentre, QE0905). Une PCR avec des amorces introniques (Tableau 8) de part et d'autre de l'exon ciblé et 100 ou 200ng d'ADNg est réalisée pour chaque clone obtenu ainsi que pour la population de cellules non modifiées servant de contrôle. Le résultat est ensuite observé sur gel d'agarose 1% par électrophorèse puis séquencé par Sanger. Le choix est ensuite effectué en fonction de si l'indel cause une modification dans le cadre de lecture de la séquence protéique et si elle est homozygote afin d'obtenir une inactivation complète de *PHACTR1* et d'éliminer la possibilité de recombinaison homologue.

Tableau 8 : Amorces utilisées pour la validation des indels par PCR

Exon	Amorce sens	Amorce anti
8	5'-TGAGAAATTGCCGGACCAGT-3'	5'-ATAAACGCATCATCAACGCCA-3'
9	5'-CTACCCTGTCATAGGTGGTAATTTG-3'	5'-AAGGGCCTAACAGGTGACTG-3'
10	5'-CGTGATGGATTCGGGTTAGAAA-3'	5'-TGGCATACTGGGACTAAGACT-3'
11	5'-TCCTTCCTCCTGCCTCACTA-3'	5'-GGATGACCCCTTGTTACAT-3'

3.1.5 Quantification des ARNm de *PHACTR1* par qPCR

Les clones sélectionnés sont remis en culture puis leur ARN total est extrait avec le kit Rneasy Plus Mini kit (Qiagen), analysé par spectrophotométrie (Take3 Micro-Volume plates ; Biotek) et avec le kit RNA 6000 Nano kit (Agilent Technologies) sur un bioanalyseur Agilent 2100 pour mesurer la pureté et

l'intégrité de l'ARN respectivement puis reverse transcrits à partir de 1µg d'ARN total et 1U de l'enzyme MultiScribe Reverse Transcriptase dans une réaction de 20µl à 100Mm dNTPS, 20U d'inhibiteurs de Rnase et 1X d'amorces aléatoires (Applied Biosystems, #4374966) selon le cycle thermal suivant : 25°C pour 10 minutes, 37°C pour 2 heures et 85°C pour 5 minutes. Une qPCR avec 1,25µl d'ADN dilué 1/50, 5µl de Platinum SYBR Green qPCR SuperMix-UDG (ThermoFisher) et 3,75µL de la paire d'amorces à 0,8µM (Tableau 9) est réalisée en triplicata sur le thermocycleur CFX384 Touch Real-Time PCR Detection System (Bio-Rad, #1855485) avec le profil thermal suivant : 50°C pour 2 minutes, 95°C pour 15 minutes puis 40 cycles d'amplification de 95°C pour 10 secondes, 55°C pour 15 secondes et 72°C pour 15 secondes. Les amorces utilisées pour *PHACTRI* tombent à la fin de l'exon 12 et au début de l'exon 13, soit deux exons présents dans tous les transcrits connus. Les produits amplifiés sont analysés avec une courbe de fusion pour s'assurer de leur pureté puis séquencés par Sanger pour valider le produit amplifié. Des contrôles sans ADN matrice sont effectués en parallèle et n'amplifient aucun produit ce qui témoigne de la l'absence de contamination. La quantification de l'expression relative de *PHACTRI* et *SELE* normalisée avec les trois gènes de référence *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase), *HPRT1* (Hypoxanthine Phosphoribosyltransferase 1) and *TBP* (TATA-binding protein) est faite selon la méthode des $\Delta\Delta CT$ (99) avec le logiciel CFX Manager 3.1 (Bio-Rad). À noter que toutes ces étapes suivent les lignes directrices établies par les MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines) pour que l'expérience soit fiable et reproductible (92).

Tableau 9 : Amorces utilisées pour la quantification de *PHACTRI* par qPCR

Gène	Amorce sens	Amorce anti
<i>PHACTRI</i> (Ex.7-8)	5'-AGAGAGGCGGATGCATGTG-3'	5'-AGGGAGTCAGAACGCATCG-3'
<i>PHACTRI</i> (Ex.12-13)	5'-GAGGAGGAAGAGGAGGAGGA-3'	5'-GGCCTGTTGCTGAGTTTGAT-3'
<i>SELE</i>	5'-CCTACAAGTCCTCTTGTGCCTTC-3'	5'-ACAGGCGAACTTGCACACA-3'
<i>HPRT1</i>	5'-TGGCGTCGTGATTAGTGATG-3'	5'-CAGAGGGCTACAATGTGATGG-3'
<i>GAPDH</i>	5'-GACAGTCAGCCGCATCTTC-3'	5'-GCAACAATATCCACTTTACCAGAG-3'
<i>TBP</i>	5'-CGAATATAATCCCAAGCGGTTT-3'	5'-GTGGTTCGTGGCTCTCTTATCC-3'

3.1.6 *Quantification protéique de PHACTR1 par immunobuvardage*

Les clones sont mis en culture puis une lyse cellulaire est réalisée avec une solution tampon RIPA (50mM Tris-HCl, pH 7.4 ; 1% NP-40 ; 0,25% Na-deoxycholate ; 0,1% SDS ; 150mM NaCl) avec les cocktails d'inhibiteurs de protéases et d'inhibiteurs de phosphatases (Sigma) et du PMSF à 1mM. Du lysat est centrifugé et le surnageant contenant les protéines est prélevé. Une quantification protéique totale avec le kit Pierce BCA Protein Assay Kit (ThermoFisher). 25µg de protéines sont mis dans un tampon à 600mM DTT et dénaturés à 95°C pour 5 minutes puis chargés sur un gel d'acrylamide à 8% dans un tampon à 25mM Tris base ; 192mM glycine ; 0,1% SDS et séparées par SDS-Page. Elles sont ensuite transférées sur une membrane de nitrocellulose dans un tampon Towbin à 25mM Tris ; 192mM glycine, 20% (v/v) méthanol. La membrane est d'abord submergée dans un tampon de blocage (TBS, 0.1% Tween-20, 5% lait) pour 2 heures à température ambiante puis dans une solution contenant les anticorps primaires contre PHACTR1 (1 :1000 ; anticorps maison générés par Biomatik) et GAPDH (1:10,000 ; Cell Signaling, #2118) à 4°C pour la nuit. Il est à noter que la spécificité des anticorps anti-PHACTR1 a préalablement été vérifiée avec une expérience de siRNA. Une immersion d'une heure à température ambiante dans une solution d'anticorps secondaires anti-lapin Ig4G couplés à une peroxydase de raifort (1:10,000 ; GE Healthcare, NA934) puis une réaction de chimioluminescence avec le kit Western Lightning Plus-ECLSubstrate (PerkinElmer, NEL105001EA) sont réalisées pour révéler les protéines qui sont ensuite observées avec l'appareil ChemiDoc Touch system (Bio-Rad, #1708370).

3.1.7 *Visualisation et quantification de la F-actine par immunohistochimie et microscopie :*

Afin d'évaluer la polymérisation d'actine, un marquage de la F-actine à la phalloïdine est réalisé. La phalloïdine est une toxine provenant de champignon qui lie l'actine et empêche sa dépolymérisation (100). Lorsque couplée à un fluorophore, il est donc possible de visualiser les filaments d'actine par microscopie confocale (100). De plus, en saturant le système de ce composé fluorescent, il est possible de quantifier les filaments d'actine (100). Les cellules sont mises en culture sur des lamelles jusqu'à confluence. Les lamelles sont lavées deux fois avec le milieu de culture, les cellules sont ensuite fixées avec du PBS / 4% PFA durant 10 minutes et rincées trois fois dans du PBS pendant 5 minutes, elles sont ensuite perméabilisées et bloquées avec du PBS / 2% BSA / 0,1% triton X-100 durant 30 minutes et rincées deux fois dans du PBS pendant 5 minutes. Pour la coloration de la F-actine, les lamelles sont

couvertes de 200µl de PBS / 1% BSA / Phalloïdine-alexa555 diluée 1/200 (A555 ; Invitrogen, A34055) à la noirceur pour une heure à la température ambiante et rincées trois fois avec du PBS pendant 5 minutes. Les noyaux sont ensuite marqués avec du PBS / DAPI 0,1µg/ml (Sigma, D9542) pendant 1 minute à la noirceur puis rincés avec du PBS durant 5 minutes. Les lamelles sont montées avec 10µl de milieu de montage.

L'acquisition d'images est réalisée avec le lecteur Cytation5 (Biotek) par microscopie à épifluorescence à un objectif de 20X. Le DAPI est excité et mesuré à des longueurs d'onde de 377nm et 447nm et le A555 est excité et mesuré à 531nm et 593nm respectivement. Les acquisitions sont réalisées via un protocole automatisé pour assurer une sélection aléatoire de trois champs par condition. L'intensité des lasers et l'exposition des échantillons sont ajustées en fonction de la condition la plus fluorescente à chaque expérience pour éviter de saturer les capteurs ce qui correspond à une perte d'information. Des volumes sont générés par une acquisition séquentielle (z-stack) de 50 couches distancées de 2µm de manière à passer verticalement au travers de la cellule puis une déconvolution avec le logiciel Huygens Pro (Scientific Volume Imaging) et l'aide de Louis Villeneuve à l'Institut de Cardiologie de Montréal est réalisée pour combiner l'ensemble de ces acquisitions. Pour cette étape, une solution de billes fluorescentes (10µl de chacune des quatre couleurs dans 200µl total ; ThermoFisher, P7220) est appliquée sur des lamelles avec des cellules fixées tel que décrit précédemment puis les lamelles sont séchées avant d'être montées sur lames. Ces dernières sont utilisées pour permettre au logiciel de déterminer la déformation obtenue sur l'axe des Z (imprécision attendue de la technique) et ainsi de corriger les acquisitions. Par la suite, un seuil d'intensité de fluorescence est déterminé pour chaque expérience pour retirer le bruit de fond de manière uniforme entre les conditions. Le calcul suivant :

$$\frac{\sum \text{intensité}}{\sum \text{volume } (\mu\text{m}^3)}$$

soit la somme de l'intensité des voxels au-dessus du seuil normalisée par le volume occupé par ces mêmes voxels permet d'obtenir une quantification de la fluorescence et donc de l'actine.

3.1.8 RNAseq

(Traitement des cellules et préparation des bibliothèques réalisés par Mélissa Beaudoin)

Les telloHAEC ont été traitées pendant 24h au TNF α à 10ng/ μ l pour induire un phénotype de réponse inflammatoire. L'ARN des cellules traitées et non traitées au TNF α est extrait avec le kit commercial RNEasy (Qiagen) et la préparation des bibliothèques est faite avec le kit « Illumina's TruSeq rRNA-depleted Stranded RNA ». 6 bibliothèques indexées sont séquencées par ligne sur l'appareil HiSeq4000 par « paired-end » (2x101-bp), résultant à 50 millions de séquences par bibliothèque. Pour la partie informatique, l'analyse se fait via la suite HISAT2 et StringTie (94). Les séquences (fichier .fastq) sont alignées sur le génome de référence avec Hisat2 (version 2.1.0). Cela génère un fichier .sam avec les positions des séquences qui est ensuite converti en .bam avec Samtools (version 1.5). L'appel des transcrits est réalisé par Stringtie (version 1.3.4b) en considérant ceux décrits dans la base de données Ensembl et en en générant *de novo*. Le logiciel calcule également l'abondance et la couverture des transcrits et de chaque exon individuellement. Plutôt que d'utiliser Deseq2 qui permet une comparaison simple, l'outil FeatureCounts de Subread (Version 1.6.0) permettant de compter le nombre absolu de séquences couvrant le transcrit est utilisé et les tests de Student suivants sont réalisés pour permettre une double comparaison de l'effet de l'édition du génome et de l'effet du traitement au TNF α :

- Comparaison des deux clones sans indel avec les trois clones avec une indel **sans** TNF α
- Comparaison des deux clones sans indel avec les trois clones avec une indel **avec** TNF α
- Soustraction entre l'expression avec TNF α et celle sans TNF α , puis comparaison entre les deux clones sans indel avec les trois clones avec une indel.

3.2 - Résultats

3.2.1 Inactivation de PHACTR1 par édition du génome avec le système CRISPR-Cas9

Par édition de génome avec le système CRISPR-Cas9, des indels ont été générés dans un premier temps dans les trois premiers exons des transcrits intermédiaires, soit les exons 8, 9 et 10, de manière à modifier le cadre de lecture de la séquence protéique (96). Dans un deuxième temps, l'expérience a été refaite en ciblant l'exon 11, celui-ci n'étant pas un multiple de trois à l'opposé des premiers. La création d'indels par édition de génome, via le mécanisme de liaison des brins non homologues après une cassure double brins de l'ADN, est devenue une technique courante pour générer des pertes de

fonction afin d'étudier le rôle des gènes. Ces indels insèrent un codon stop prématuré qui en théorie forme des protéines tronquées non fonctionnelles ou enclenche le mécanisme de dégradation des ARNm non-sens. Toutefois, les indels introduites dans *PHACTR1* lors de la première tentative sollicitent un troisième mécanisme qui est l'épissage alternatif des exons portant l'indel. En effet, ces derniers ayant des séquences dont le nombre de nucléotides est un multiple de trois, leur excision raccourcit la séquence de la protéine sans toutefois altérer le cadre de lecture de sorte que cela permet possiblement à la protéine de maintenir en partie sa fonction si ce n'est pas entièrement (96). À la suite de la deuxième expérience, en plus du contrôle WT, deux contrôles avec des sgRNA non spécifiques sont utilisés et 3 clones avec *PHACTR1* édité (Figure 13) sont sélectionnés. Cette seconde tentative semble avoir fonctionné et réduit significativement l'expression de *PHACTR1*. En effet, une diminution au niveau des ARNm et des protéines est observée par qPCR et par immunobuvardage respectivement (Figure 14). À noter qu'une expression est mesurée pour les échantillons sg1n6 et sg2n6 sans TNF α (Figure 14-B), mais il semble que cela soit du bruit de fond sur la membrane (Figure 14-A) tel que le laisse supposer la diminution d'expression de PHACTR1 pour ces clones avec TNF α . D'autre part, SELE, une molécule d'adhésion exprimée lors d'une dysfonction endothéliale, a également été quantifiée (Figure 14). La diminution de *PHACTR1* et l'augmentation de *SELE* concordent avec la validité des clones. Les clones teloHAEC *PHACTR1*^{-/-} sont donc jugés utilisables pour la suite des expériences.

Transcrit intermédiaire

MRSDSLVPGTHTPPIRRRSKFANLGRIFKPWKWRKKKSEKFKHTSAALERKISMR
QSREELIKRQVLKEIYDKDGELSI SNEEDSLENGQSLSSSQLSLPALSEMPEVPM
PRDPCSYEVLQPSDIMDGPVSEESPSASESGVLLSQDPSAKPVLLLPPKKPAAFP
GDHEETPVKQLPLLKQPPALPPKPTTRIANHLTD**PGAPVKLPCLPVKLS PPLPPK**
KVMICMPVGGPDL SLVSYTAQKSGQQGVAQHHTVLP S QIQHQLQYGSHGQHLPS
TTGSLPMHPSGCRMIDELNKTLAMTMQRLESSEQRVPCSTSYHSSGLHSGDGVTK
AGPMGLPEIRQVPTVVEICDDNKENVP HESDYEDSSCLYTREEEEEDEDDSS
LYTSSLAMKVCRKDSLAIKLSNRPSKRELEEKNILPRQTDEERLELRQQIGTKLT
RRLSQRPTAELEQRNILKPRNEQEEQEEKREIKRRLTRKLSQRPTVEELRERKI
LIRFSDYVEVADAQDYDRRADKPWTRLTAADKAAIRKELNEFKSTEMEVHEL SRH
LTRFHRP*

Sg1n1

Indel : -CC

Effet : ATGCCCGTG : MPV... → MR...Stop

PGAPVKLPCLPVKLS PPLPPKKVMICMRGGARPLTGVLHSPEEWPAGCGPAPPHC
PALPDPAPAAVRQPRPAPPLHHRLPPHAPLGLQNDRRAQQNAGHDHAEAGKL*

Sg1n6

Indel : +C

Effet : ATCCCCGTGGG : MPVGG... → MPRG...Stop

PGAPVKLPCLPVKLS PPLPPKKVMICMPRGGARPLTGVLHSPEEWPAGCGPAPPH
CPALPDPAPAAVRQPRPAPPLHHRLPPHAPLGLQNDRRAQQNAGHDHAEAGKL*

Sg2n6 (hétérozygote composé)

Indel : +T

Effet : TAAAACGCTG : KTL → Stop

PGAPVKLPCLPVKLS PPLPPKKVMICMPVGGPDL SLVSYTAQKSGQQGVAQHHT
VLPSQIQHQLQYGSHGQHLPS TTGSLPMHPSGCRMIDELN*

Indel : +CC

Effet : AAAACCGCTG : KTL → KTRWP Stop

PGAPVKLPCLPVKLS PPLPPKKVMICMPVGGPDL SLVSYTAQKSGQQGVAQHHT
VLPSQIQHQLQYGSHGQHLPS TTGSLPMHPSGCRMIDELNKTRWP*

Figure 13 : Effet des indels sur la séquence protéique des clones sélectionnés. Rouge = exon 11, Souligné = endroit ciblé par le sgRNA, Gras = séquence modifiée.

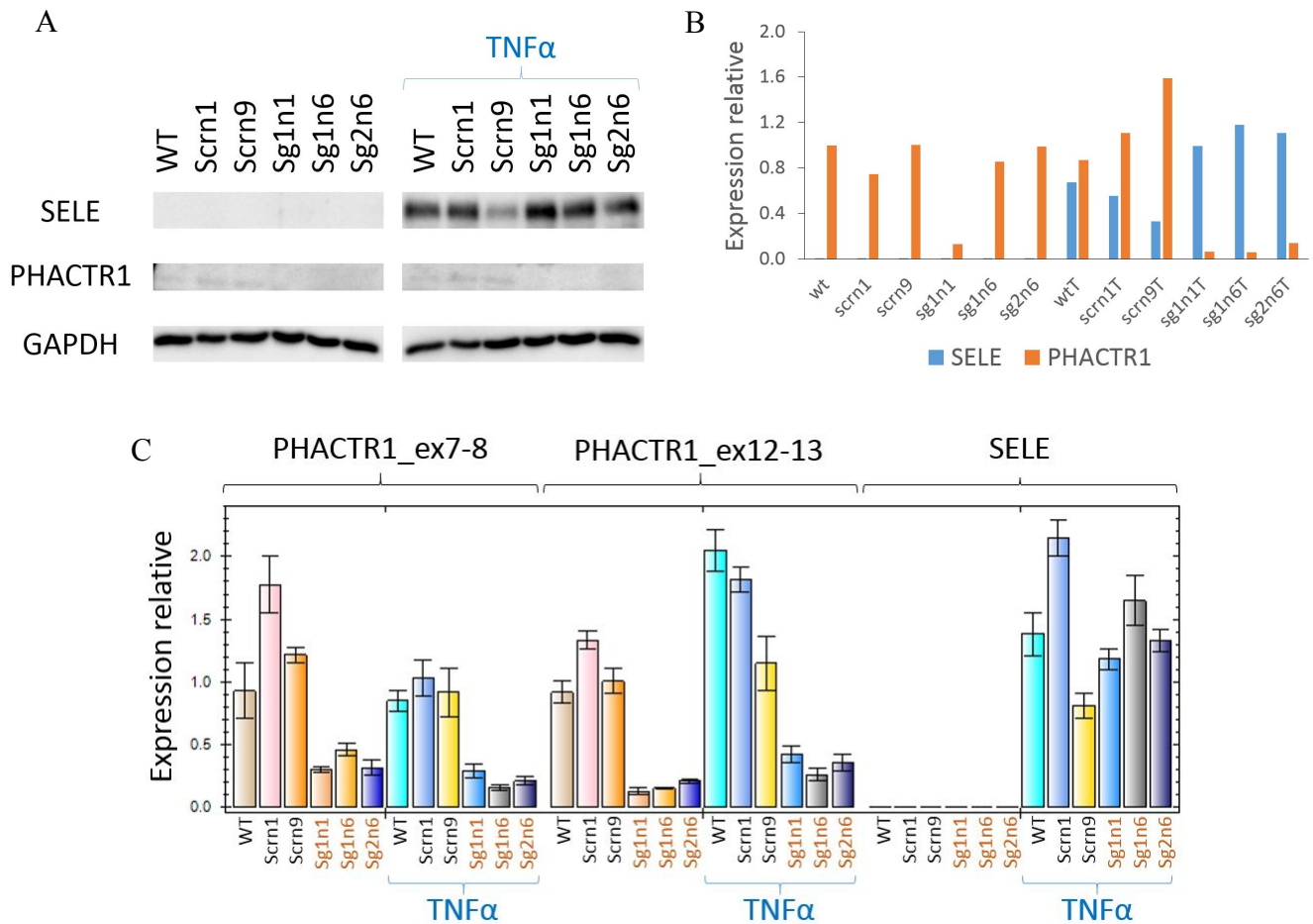


Figure 14 : Validation de la diminution d'expression de *PHACTR1* dans les clones édités par le système CRISPR-Cas9. Les cellules sont soumises à un traitement au $TNF\alpha$ pendant 24h avant d'extraire les protéines et l'ARNm. A) Quantification au niveau protéique par immunobuvardage. B) analyse de l'intensité des bandes de l'immunobuvardage. C) Quantification au niveau des ARNm par qPCR. Deux paires d'amorces distinctes sont utilisées pour quantifier *PHACTR1*, soit dans les exons 7 et 8 et les exons 12 et 13 pour amplifier une région de *PHACTR1* en amont et en aval de l'indel dans l'exon 11.

3.2.2 Structure de l'actine

Puisque PHACTR1 interagit avec l'actine monomérique (G-actine : actine globulaire) dans le cytoplasme, il était attendu que l'inactivation du gène module la polymérisation d'actine. En effet, l'actine existe sous deux forme dans le cytoplasme des cellules, soit la G-actine et la F-actine (actine filamenteuse ; polymère). Les filaments d'actine qui constituent le cytosquelette sont formés par la polymérisation des molécules de G-actine. PHACTR1 pourrait donc médier la polymérisation en

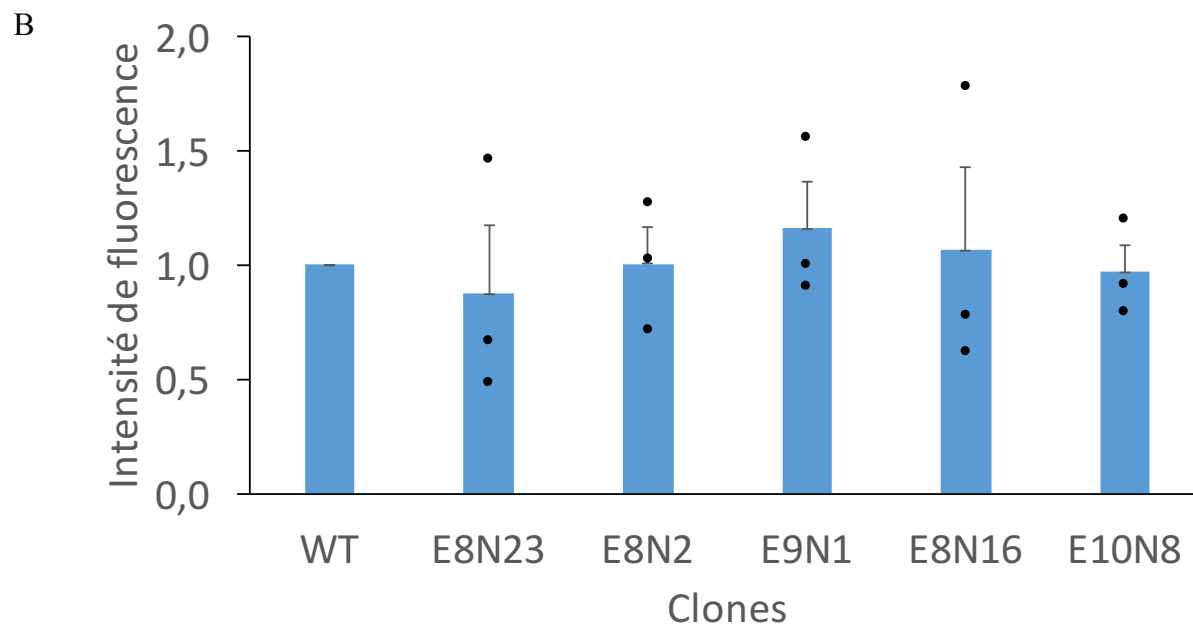
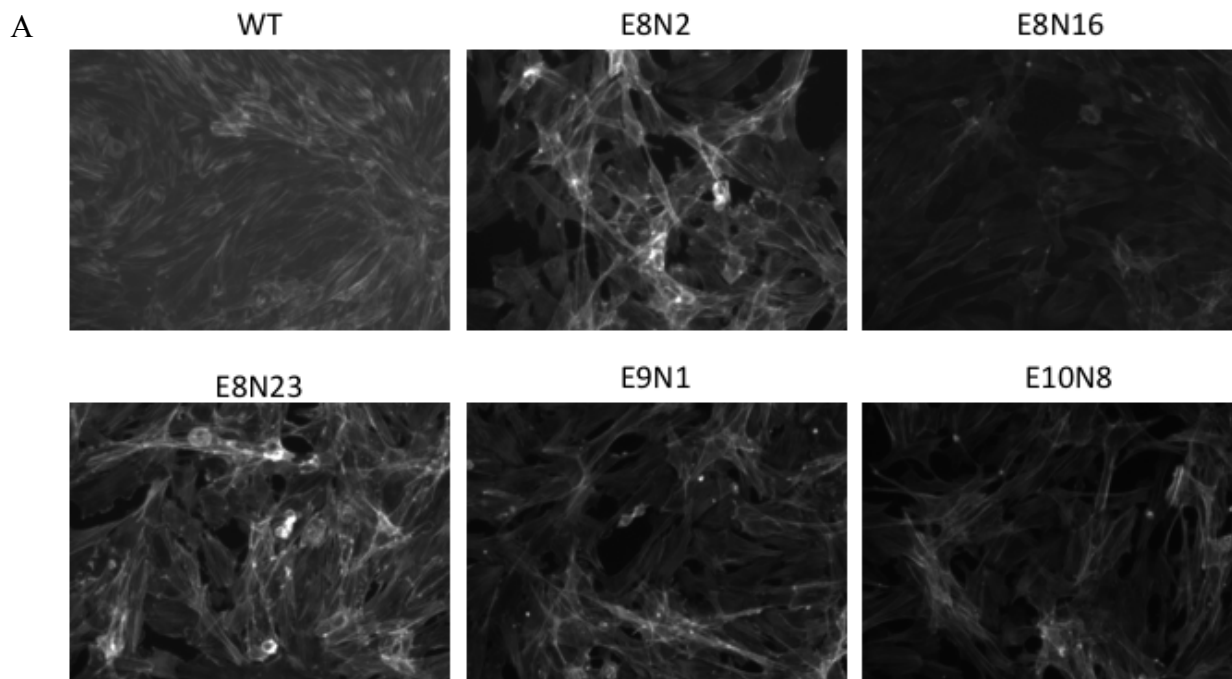


Figure 15 : Maintien de la structure du cytosquelette d'actine dans la première série de clones teloHAEC Cas9 avec *PHACTR1* inactivé. A) Visualisation de l'actine marquée à la phalloïdine-A555 à 20X. B) Quantification de l'actine via l'intensité de fluorescence chez les clones par rapport aux cellules WT. Trois expériences indépendantes ont été réalisées avec trois acquisitions pour chaque condition. Nomenclature des clones : WT = échantillon teloHAEC Cas-9 n'ayant pas subi l'édition de génome de *PHACTR1*; E = # exon ciblé; N = # clone.

limitant la quantité de G-actines libres. Pour ce faire, les filaments d'actine ont été marqués à la phalloïdine-A555 puis observé par microscopie afin d'observer la structure du cytosquelette et de quantifier l'actine. Un changement morphologique des cellules est observé entre les clones et les teloHAEC non édités. Les cellules dans ce contrôle forment une monocouche uniforme à confluence alors que les cellules des clones adoptent une conformation comparable à un « filet » de sorte qu'elles sont très confluentes et se grimpent les unes sur les autres à certains points alors qu'elles laissent d'autres régions inoccupées (Figure 15-A). Cette différence pourrait avoir été occasionnée par l'expérience d'édition du génome et non par l'inactivation de *PHACTR1*. Toutefois, au sein d'une cellule, les filaments d'actine semblent demeurer constants et transversaux (Figure 15-A). De plus, aucune variabilité significative n'est mesurée lors de la quantification de l'actine via un calcul d'intensité de fluorescence pour un volume occupé (Figure 15-B). Dans les faits, bien que les clones présentent des indels altérant le cadre de lecture au niveau génomique, ils expriment tout de même une protéine PHACTR1 tronquée sans modification du cadre de lecture grâce à un mécanisme d'épissage alternatif de l'exon comportant l'indel puisque celui-ci portent un multiple de trois nucléotides (96). Une deuxième série de clones a alors été générée de la même façon en ciblant l'exon 11 qui n'a pas un nombre de nucléotides correspondant à un multiple de trois. Après la validation des clones, le marquage de l'actine et son observation par microscopie a été répétée, mais la structure de celles-ci étant encore une fois similaire (Figure 16), la quantification n'a donc pas été reprise.

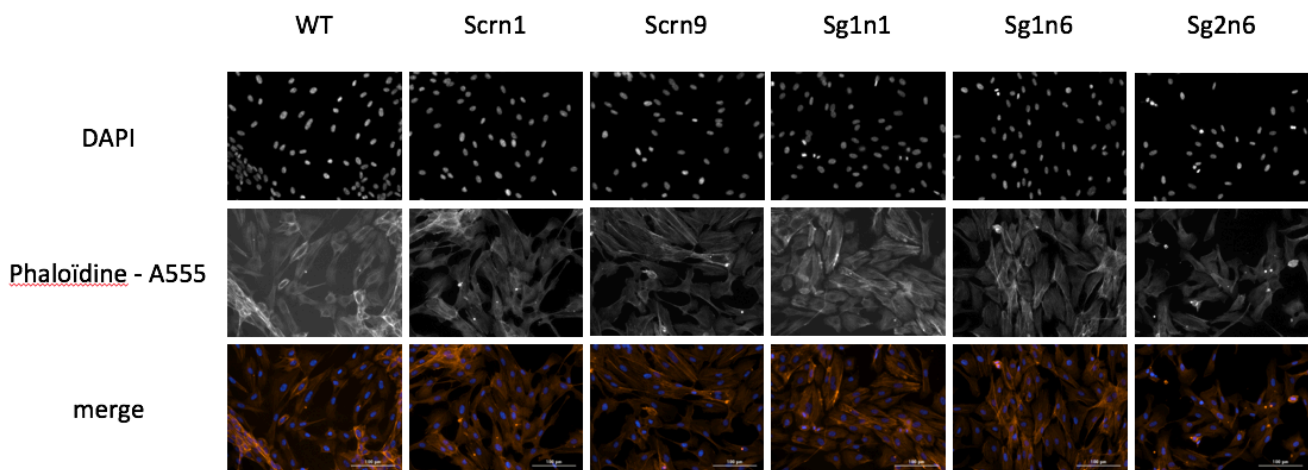


Figure 16 : Coloration de l'actine (phalloïdine-A555) et des noyaux (DAPI) de la seconde série de clones teloHAEC Cas9 avec *PHACTR1* inactivé. Nomenclature des clones : WT = échantillon teloHAEC Cas-9 n'ayant pas subi l'édition de génome de *PHACTR1*; Scrn = scramble - contrôle avec un sgrNA ne ciblant pas de région génomique; Sg = sgrNA utilisé; n = # clone.

3.2.3 Analyse différentielle du transcriptome des clones avec et sans TNF α par RNAseq

L'inactivation de *PHACTRI* pourrait affecter l'expression génique des voies biologique dans lesquelles il est impliqué. Afin de pouvoir identifier celles-ci, une expérience de RNAseq a été menée sur les teloHAEC avec et sans stimulus inflammatoire pour les contrôles et les clones. Tel qu'attendu, l'expression de *PHACTRI* est moindre dans les clones et l'expression des molécules d'adhésion est supérieure lorsque les cellules sont stimulées au TNF α (Tableau 10). Cela est consistant avec la validité du modèle cellulaire généré.

Tableau 10 : Expression de *PHACTRI* et des molécules d'adhésion selon le traitement au TNF α et l'inactivation de *PHACTRI*

Gènes	Sans TNF α		Avec TNF α	
	Contrôles	Clones	Contrôles	Clones
<i>PHACTRI</i>	282	92,7	437,5	132,3
<i>SELE</i>	27	32,3	135908	152790
<i>SELP</i>	8	5,7	5,5	4,3
<i>VCAM-1</i>	10	21,3	65062	61611,3
<i>ICAM-1</i>	4883,5	5168,7	181809,5	158018

Pour faire une analyse différentielle d'expression, DESeq2 est l'outil par excellence. Toutefois, il ne peut être employé dans le cas présent en raison d'une double comparaison (avec et sans TNF α et avec et sans indel). À la place, la couverture des transcrits est mesurée avec l'outil FeatureCounts et une analyse différentielle est effectuée via trois tests de t comparant l'expression des transcrits entre les clones et les contrôles selon les trois contextes suivant : sans TNF α , avec TNF α et la différence entre ces conditions. À noter qu'il n'y a nul besoin de normaliser la longueur des transcrits puisque chacun est analysé individuellement et que la comparaison se fait entre conditions. Seuls les transcrits couverts par des séquences sont analysés. Dans le cas présent, environ 0,8-2,0% des transcrits sont différentiellement exprimés (p-valeur $\leq 0,05$) (Tableau 11). Le nombre absolu de gènes différentiellement exprimés, soit de 531, 231 et 289 pour les conditions avec TNF α , sans TNF α et la différence respectivement (Tableau 11), demeure considérable et l'interprétation complexe. Pour simplifier l'interprétation, une analyse de voies biologiques est réalisée via l'outil en ligne DAVID

(Database for Annotation, Visualization and Integrated Discovery; Version 6.7) qui recherche l'enrichissement d'annotations fonctionnelles pour une liste de gènes donnée (101, 102). Puisque les gènes différentiellement exprimés sont majoritairement exclusifs entre les conditions (Figure 17), ils ont été analysés séparément dans DAVID. Dans cette analyse, les gènes avec une expression différentielle (p-valeur $\leq 0,05$) sont donnés comme liste et l'ensemble des gènes dont l'expression a été quantifiée par RNAseq est utilisé comme arrière-plan. Il en ressort que les voies impliquant les cytokines sont impliquées, ce qui était attendu étant donné leur rôle dans la réponse inflammatoire. Autrement, il y a une redondance des gènes entre les voies biologiques impliquées et il y a peu de gènes impliqués proportionnellement à la totalité de ceux différentiellement exprimés, soit environ 1-5% (Tableau 12). De plus, les quelques voies identifiées ne ressortent que marginalement significatives (Tableau 12). Il est donc difficile d'identifier les voies biologiques d'intérêt impliquant *PHACTR1*.

Tableau 11 : Calcul de l'expression différentielle des 55972 gènes et transcrits connus ou générés *de novo* par Hisat2.

	Avec TNF α	Sans TNF α	Différence entre avec et sans TNF α
# comparaisons	27504	28034	30168
# p-valeur $\leq 0,05$	531	231	289

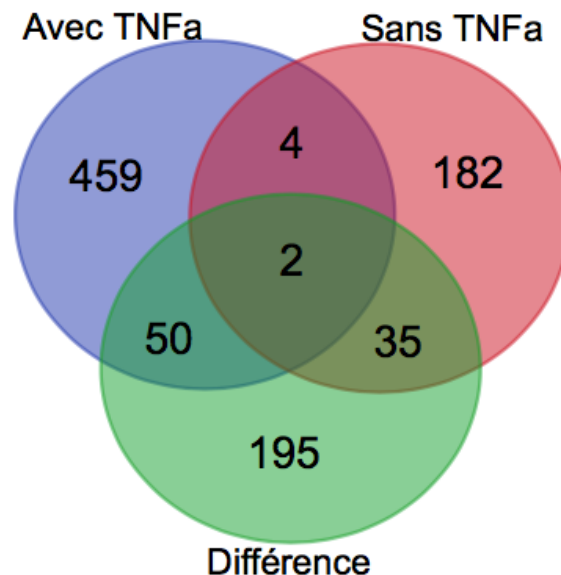


Figure 17 : Chevauchement des gènes différentiellement exprimés (p-valeur $\leq 0,05$) selon la condition pour laquelle la comparaison contrôles-clones est effectuée.

Tableau 12 : Analyse de DAVID des voies biologiques impliquées pour tous les gènes avec une expression différentielle (p-valeur $\leq 0,05$).

Comparaison (Contrôles – Clones)	Termes	Gènes impliqués (Compte absolu)	Gènes impliqués (%)	P-Valeur	Score Benjamini
Sans TNF α	Voie de signalisation du TGF- β	3	3.8	2.60 ^E -02	4.40 ^E -01
	Interaction cytokine – récepteur	10	3.6	9.20 ^E -03	6.30 ^E -01
Avec TNF α	Cascades de complémentation et de coagulation	5	1.8	1.50 ^E -02	5.50 ^E -01
	Métabolisme de l’amidon et du saccharose	4	1.5	2.00 ^E -02	5.10 ^E -01
	Maladies à prions	3	1.1	8.30 ^E -02	9.10 ^E -01
Avec – sans TNF α	Interaction cytokine – récepteur	6	5.5	1.30 ^E -02	5.00 ^E -01
	Voie de signalisation des chimiokines	4	3.7	8.20 ^E -02	9.00 ^E -01

3.3 - Discussion

Dans un premier temps, afin de déterminer le rôle de PHACTR1 dans les MCVs, une imagerie des filaments d’actine par microscopie a été réalisée pour analyser l’effet de l’inactivation de la protéine sur la polymérisation d’actine. Parmi les clones de *PHACTR1* édités, il a été noté une conservation de l’intégralité du cytosquelette d’actine. D’une part, Ces résultats ne sont pas concordants avec ceux obtenus dans la littérature. Deux études ont regardé l’effet d’une inactivation de *PHACTR1* par siRNA sur le cytosquelette d’actine par microscopie. L’une note une désorganisation et une réduction des filaments d’actine générales, mais plus prononcés en périphérie des cellules HUVEC (63). En microscopie, les longs filaments d’actine parallèles traversant la cellule d’une extrémité à l’autre pour les cellules sans traitement et pour le contrôle avec des siRNA sans cible se voient concentrés partiellement en points dans l’échantillon avec siRNA contre *PHACTR1* (63). L’autre montre une diminution du nombre de filaments d’actine transversaux et une conservation de ceux en périphérie dans les cellules de mélanome CHL-1, soit des cellules cancéreuses de la peau (65). De plus, les points d’adhésion focaux observés via un marquage de la paxilline, protéine à la jonction entre les filaments d’actine et les protéines d’ancrage avec la matrice extracellulaire, se divisent et se répartissent en périphérie de la cellule (65). Cela pourrait provenir de la lignée cellulaire utilisée ou de l’expérience d’édition de génome qui diffèrent de ces expériences.

D'autre part, la quantification n'est pas optimale puisque la technique de marquage de l'actine est hautement variable en raison de la confluence des cellules, de la dilution de la phalloïdine concentrée, du petit volume à appliquer sur les lamelles lors de l'incubation pour le marquage et de l'évaporation de celui-ci. Le marquage est cependant supposé être proportionnel à la quantité de F-actine (100) et la quantification est faite avec les conditions optimales, soit par l'imagerie d'un volume qui prend en considération la cellule dans sa totalité plutôt que sur un seul plan focal et une normalisation sur le volume occupé pour prendre en compte la différence de confluence des cellules entre les conditions. L'inconvénient est que si l'inactivation de *PHACTRI* n'induit qu'une faible variation dans la quantité d'actine, il se pourrait que l'incertitude reliée avec la technique surpasse la variation biologique qui serait alors non détectable. Un immunobuvardage serait une solution pour une quantification adéquate. Cependant, les anticorps contre l'actine ne distinguent pas les formes monomérique et filamenteuse de sorte que la mesure serait appliquée à la combinaison de la G-actine et la F-actine. Toutefois, ayant des masses différentes, ces deux formes peuvent être séparée par centrifugation préalablement à l'expérience d'immunobuvardage (103, 104). La technique de marquage a été sélectionnée de par le fait qu'elle apporte une information supplémentaire sur la structure qui est hautement pertinente et parce qu'elle permet une comparaison avec ce qui est observé dans la littérature bien qu'il été considéré que cela ne soit pas la meilleure méthode de quantification. Malgré les difficultés à obtenir des clones valides de *PHACTRI* et la méthode semi-quantitative, les résultats négatifs pourraient être représentatifs de la réalité. En effet, il se pourrait que la quantité d'actine ne soit pas affectée par l'inactivation de *PHACTRI* si sa concentration cytoplasmique est élevée et non limitante dans la réaction de polymérisation. Dans les faits, de nombreux processus biologiques sont régis non par les composés, mais par les enzymes et l'énergie requises pour la transformation. Pour l'actine, il se peut que cela soit au niveau des protéines agissant sur l'initiation de la polymérisation ou sur la régulation du niveau cytosolique de G-actine telles que l'Arp2/3, la thymosine et la profiline (105).

Dans un second temps, une expérience de RNAseq a été menée afin de caractériser l'expression différentielle à la suite de l'inactivation de *PHACTRI* avec et sans traitement au TNF α . Le transcriptome semble représentatif de la réponse inflammatoire et de l'effet de l'indel introduit dans *PHACTRI* tel que montré par l'expression de *PHACTRI* et des molécules d'adhésion. Toutefois, l'expression différentielle ne permet pas de faire ressortir des voies biologiques majeures. C'est résultats non concluants semblent indiquer que *PHACTRI* ne module pas le risque de maladie en empruntant une voie principale. Considérant le fait que les MCV sont multigéniques et que les variants

dans *PHACTRI* ne modulent qu'une faible portion du risque des maladies, il est toutefois possible que l'effet de *PHACTRI* soit léger de manière à être difficilement mesurable.

Par ailleurs, l'édition de génome s'est avéré plus complexe qu'attendue, l'inactivation de *PHACTRI* étant palliée par un épissage alternatif des exons avec une indel, de sorte qu'il est possiblement essentiel et que c'est un moyen pour contourner l'inactivation. Il ne faut pas oublier qu'il existe quatre gènes *PHACTR* de sorte que bien que *PHACTRI* ait été inactivé, il est possible qu'un mécanisme de compensation par les autres gènes de cette famille soit mis en place. En effet, il existe de nombreuses redondances génétiques pour assurer la fonction de gènes essentiels. Même si cela n'est pas optimal, un tel mécanisme permet parfois de pallier les mutations délétères au niveau génomique ou d'amoindrir leur effet sur le phénotype.

4 - Priorisation de variants

Dans cette ère des omics, nombreuses sont les GWAS qui associent des SNPs avec un phénotype de maladie. C'est ainsi que *PHACTR1* avait initialement été sélectionné comme candidat d'intérêt potentiellement causal dans les MCV. Toutefois, bien qu'il y ait une corrélation, il n'y a pas nécessairement de causalité. Les scientifiques sont vis-à-vis la problématique d'avoir une multitude de SNPs influençant potentiellement le risque de maladie sans toutefois pouvoir expliquer l'effet de chacun. Au niveau biologique, il demeure primordial de justifier le rôle d'un SNP pour en inférer sa causalité. Considérant l'impossibilité de valider le rôle de *PHACTR1* et la diversité des pistes encore inexplorées, la suite du projet a été orientée vers la génération et l'analyse de données génomiques permettant d'obtenir un aperçu du rôle biologique des variants associés pour prioriser ceux pertinents. Étant donné le contexte des MCV et l'implication des dysfonctions endothéliales dans le développement de l'athérosclérose qui aboutit en MCV, les SNPs responsables des MCV via cette voie devraient ressortir par une analyse comparative des EC avec et sans dysfonctions endothéliales. En effet, il a été démontré que les SNPs associés aux maladies sont enrichis dans les gènes exprimés dans les tissus pertinents pour les maladies (106). Pour l'étude, des cellules teloHAEC sont utilisées avec un traitement au TNF α pour enclencher les dysfonctions endothéliales. Des expériences d'ATAC-seq, de RNAseq et de Hi-C ont été réalisées sous un stimulus inflammatoire pour caractériser la conformation chromosomique et le transcriptome lors d'une dysfonction endothéliale.

4.1 - Méthodes

4.1.1 Culture cellulaire

Voir la section 3.1.1.

Les cellules teloHAEC ont été modifiées par Mélissa Beaudoin pour exprimer la Cas9 de manière stable. Celles-ci sont utilisées pour l'expérience d'édition du génome par le système CRISPR-Cas9 (section 4.1.6).

4.1.2 Séquençage du génome entier

(Section réalisée par Ken Sin Lo)

Un séquençage du génome entier avec une couverture de 43X a été réalisé : 4 131 557 SNPs (7% nouveau), 820,895 indels, 176,543 variants multi-allelic, 814 inversions, 1434 duplications, 31 180 délétions). Les variants trouvés ont été pris en compte lors de la conception des sgRNA pour l'expérience d'édition de génome.

4.1.3 ATAC-seq

(Traitement des cellules et préparation des bibliothèques réalisés par Mélissa Beaudoin)

Les teloHAEC ont été traitées en *duplicata* pendant 0h, 4h et 24h au TNF α à 10ng/ml pour induire un phénotype de réponse inflammatoire. L'expérience d'ATAC-seq est réalisée selon le protocole du projet ENCODE. Les noyaux sont purifiés et perméabilisés, puis traités avec la transposase Tn5 qui digère l'ADN proportionnellement à l'ouverture de la chromatine. La réaction de digestion est arrêtée et l'ADN est purifié. La construction de bibliothèques est réalisée avec le kit d'Illumina. À noter que lors de l'amplification par PCR, le nombre de cycles est limité au maximum pour réduire le nombre de *duplicata* de PCR (amplicons identiques), soit des artefacts de la préparation de la bibliothèque biaisant le signal biologique voulant être mesuré. Une sélection de taille est faite et un séquençage des 6 bibliothèques est effectué par extrémités appariées (« paire-end ») sur l'appareil HiSeq2500. Le séquençage par extrémités appariées permet un meilleur alignement de par le fait que les deux extrémités sont séquencées. Cela permet aussi une meilleure reconnaissance des *duplicata* de PCR à retirer. Une fois les amplicons séquencés, le traitement informatique est réalisé selon les lignes directrices établies avec les paramètres par défaut (https://github.com/kundajelab/atac_dnase_pipelines). Brièvement, un contrôle de qualité est réalisé, les adaptateurs de séquençage illumina sont coupés et les *duplicata* de PCR sont retirés. Les séquences restantes sont alignées au génome de référence avec le logiciel HISAT2 (version 2.0.5) et l'appel des pics de chromatine ouverte est fait par le logiciel MACS2 (fonction callpeak). Cela génère un fichier de positions chromosomiques correspondant aux pics.

Analyse : Enrichissement de variants dans les régions de chromatine ouverte selon les tissus

Des données d'ATAC-seq en format .bigwig provenant de ENCODE ont été téléchargées. Cela comprend 48 échantillons dans 23 tissus différents. Ces fichiers ont été convertis en .bedgraph avec l'outil bigWigToBedGraph puis l'appel des pics est réalisé avec MACS2 (fonction bdgpeakcall). Pour

nos échantillons et pour ceux d'ENCODE, le nombre de SNPs tombant dans des pics de chromatine ouverte est fait avec la fonction intersect de bedtools.

Analyse : Régions de chromatine différenciellement ouverte

À partir des fichiers de pics de chromatine ouverte des trois conditions (fichiers .bed), un fichier combiné a été généré à l'aide de l'outil « merge » du logiciel Bedtools (version 2.17.0). Ainsi, dès qu'une paire de bases se retrouvait dans au moins l'une des trois conditions, elle était incluse dans le fichier combiné. En générant ainsi des pics consensus, la taille des pics est normalisée ce qui simplifie grandement l'analyse. La couverture de ces « pics combinés » est calculée pour chaque condition et une analyse de variance (distribution binomiale négative) par le module DESeq2 sur R est réalisée et permet de calculer un ratio logarithmique de changement (logFC : log fold change). Le logFC correspond au logarithme en base 2 du ratio de la valeur finale sur la valeur initiale.

$$\log FC = \log_2 \frac{X_{\text{final}}}{X_{\text{initial}}}$$

Il est considéré qu'un pic avec une p-valeur ajustée inférieure à 0,001 et un logFC inférieur à -0,3 ou supérieur à 0,3 (soit un ratio de changement d'ordre de grandeur de deux) est significativement plus fermé ou plus ouvert que le groupe comparatif respectivement.

Analyse : Déséquilibre allélique de l'ouverture de la chromatine

Un script en langage Ruby a été écrit pour faire l'analyse. Celui-ci regarde toutes les positions hétérozygotes tombant dans un pic de chromatine ouverte. Pour chacune de celles-ci, il filtre les séquences selon une qualité d'appel des bases de 9 et une couverture de 9 puis compte le nombre de séquences obtenues pour chaque allèle. Il calcule ensuite le ratio qu'il compare avec le ratio attendu, calculé préalablement à partir de la fréquence allélique moyenne de toutes les positions hétérozygotes de même génotype, et exécute un test de t dans R pour déterminer si cette différence est significative (P = 0,05 ajusté pour le nombre de tests pour chaque condition avec la correction de Bonferroni : $P_{WT} = 0,001852$, $P_{4h} = 0,001923$, $P_{24h} = 0,0025$). En résumé, l'analyse permet d'attribuer les séquences obtenues par séquençage de nouvelle génération superposant un SNP hétérozygote à l'un ou l'autre des chromosomes pour établir le signal propre à chaque allèle et déterminer si l'ouverture de la chromatine est différentielle entre ces allèles.

Analyse : Motifs différenciellement liés par des facteurs de transcription

Pour cette analyse, l'outil CENTIPEDE a été utilisé. Dans un premier temps, il recherche les séquences liées par des facteurs de transcription (Base de données de motifs JASPAR 2018 pour les vertébrés incluant les motifs redondants) dans les régions de chromatine ouverte et attribue un score *a priori* d'être lié pour chacun des motifs trouvés. Ce score est basé uniquement sur les données génomiques. Dans un second temps, il calcule un score *a posteriori* en fonction des données expérimentales, soit celles du degré d'ouverture de la chromatine provenant du ATAC-seq. Un décompte du nombre de motifs liés pour chaque facteur de transcription est effectué pour chaque échantillon et une normalisation est faite en fonction du nombre de paires de bases ouvertes, puis une analyse différentielle entre condition est réalisée par un test de t dans R.

4.1.4 RNAseq

(Traitement des cellules et préparation des bibliothèques réalisés par Mélissa Beaudoin ;
Analyses réalisées par Ken Sin Lo)

Voir la section 3.1.8.

Les teloHAEC ont été traitées en *duplicata* pendant 0h, 4h et 24h au TNF α à 10ng/ml pour induire un phénotype de réponse inflammatoire. Pour la portion informatique, l'analyse d'expression différentielle se fait ensuite avec l'outil Deseq2 dans R, soit l'outil normalement utilisé. Le LogFC est calculé et la correction de Bonferroni est apportée aux p-values. Les transcrits avec un $|\logFC|$ supérieur à 0,3 et une p-valeur ajustée inférieure à 0,001 sont considérés.

4.1.5 Hi-C

(Traitement des cellules et préparation des bibliothèques réalisés par Simon Lalonde;
Analyses réalisées par Ken Sin Lo)

Les teloHAEC ont été traitées pendant 0h, 4h au TNF α à 10ng/ml pour induire un phénotype de réponse inflammatoire. Les bibliothèques sont préparées tel que mentionné dans (107). Celles-ci sont séquencées sur deux lignes du HiSeq4000 (2x101-bp). L'analyse des séquences brutes et leur alignement sur le génome sont faits avec HiCUP (108), puis l'analyse statistique des interactions est réalisée avec CHiCAGO (109). Cela génère plus de 150M "di-tags" reliant physiquement deux séquences non adjacentes. Les interactions présentées sous forme de « heatmap » sont ensuite visualisées en ligne via le navigateur « 3D genome browser » (<http://promoter.bx.psu.edu/hi-c>).

4.1.6 Édition du génome par le système CRISPR-Cas9

Voir la section 3.1.2.

Un programme en ruby généré par Ken Sin Lo permet de rechercher les séquences PAM dans le génome et de générer les séquences de sgRNA.

Tableau 13 : Séquences des sgRNA utilisés pour l'édition de génome par le système CRISPR-Cas9.

Loci	sgRNA sens	sgRNA anti
Chr1 (amont)	5'-caccGTTACATTTAGTGGACACGG-3'	5'-aacCCGTGTCCACTAAATGTAAC-3'
Chr1 (aval)	5'-caccgAGATTTTCCAGGTTTGTTC-3'	5'-aacGAACAAAACCTGGAAAATCTc-3'
Chr5 (amont)	5'-caccgAAAGTGGAGGAAGGGGTTGC-3'	5'-aacGCAACCCCTTCCTCCACTTTC-3'
Chr5 (aval)	5'-caccgCCAGTCTGGTTGGCGAGATT-3'	5'-aacAATCTCGCCAACCAGACTGGc-3'
Chr9 (amont)	5'-caccGTGGGGTTTGCCGCAAACAT-3'	5'-aacATGTTTGCGGCAAACCCAC-3'
Chr9 (aval)	5'-caccgAGGGCCGGTTTTTGGACCAC-3'	5'-aacGTGGTCCAAAACCGGCCCTc-3'
Chr20 (amont)	5'-caccGCCAGGTTCTCCAACCTAAA-3'	5'-aacTTAAGTTGGAGAACCTGGC-3'
Chr20 (aval)	5'-caccgAGTCCTTGGAGTGTCTCGG-3'	5'-aacCGAGGACACTCCAAGGACTc-3'

4.1.7 Dériver des populations en sous-populations et en clones

Pour les sous-populations, les cellules sont mises en 96 puits à 50 cellules par puit et sont gardées en culture jusqu'à confluence. Un passage est alors effectué pour prélever l'ADNg du $\frac{3}{4}$ des cellules à des fins d'analyse par PCR et le $\frac{1}{4}$ restant est remis en culture. Les sous-populations positives pour la délétion sont ensuite dérivées pour obtenir des clones tel que précédemment.

Voir la section 3.1.3 pour dériver des clones.

4.1.8 Validation des délétions générées par le système CRISPR-Cas9

L'ADNg des sous-populations est extrait avec le kit « QuickExtract DNA Extraction Solution » (Epicentre, QE0905). Une PCR de 10µl est effectuée avec des amorces (Tableau 14) de part et d'autre de la délétion, 10ng d'ADNg et la polymérase GoTaq (Promega) avec le profil thermal suivant : 94°C pour 2 minutes, 94°C pour 30 secondes, 56°C pour 30 secondes et 72°C pour 1 minute pour 30 cycles d'amplification, puis 72°C durant 5 minutes. Les produits sont ensuite analysés sur un gel d'agarose 1% par électrophorèse et les sous-populations avec les plus grands ratios d'amplicons avec et sans délétion sont sélectionnées.

Voir section 3.1.4.

Tableau 14 : Séquences des amorces utilisées pour la validation des délétions par PCR

Chr	Amorce sens	Amorce anti
1	5'-TGATGGATGTCTCTCGGGTA-3'	5'-CTGAAGCAGGACAGGCATTT-3'
5	5'-GAGCTGCATTTCTTCCTGGG-3'	5'-CCTCCCTTCTCTGACCAGTG-3'
9	5'-GACTGGAAAACGACTGGCTC-3'	5'-CCCTGACCTCCATCCTGTTT-3'
20	5'-TCACAGTGATCATTGGGCCA-3'	5'-AAGGGGAACAGAGGGAAACG-3'

4.2 - Résultats

4.2.1 Enrichissement de variants associés aux MCV dans des régions de chromatine ouverte pour les EC

Pour cette partie, il est considéré que les SNPs d'intérêt pour les MCV se situent sur de la chromatine ouverte dans les tissus reliés aux dysfonctions endothéliales. Il est attendu que la proportion de SNPs associés aux MCV tombant dans une région de chromatine ouverte soit plus élevée dans un tissu et modèle cellulaire impliqué dans les dysfonctions endothéliales. En effet, un SNP situé dans une région de chromatine active est plus à même d'influencer la régulation génique et donc de moduler l'effet d'un phénotype. Dans cette optique, une quantification des ratios du nombre de SNPs parmi les 161 associés aux MCV et les 357 associés à l'HYP ainsi que ceux en LD positionnés dans de la chromatine ouverte proportionnellement au nombre total de SNPs a été réalisée pour notre modèle cellulaire de teloHAEC ainsi que pour 23 tissus différents provenant de la base de données publique ENCODE. Le chevauchement des SNPs tombant dans des pics de chromatine ouverte est montré dans un diagramme de Venn (Figure 18). Il est à considérer que les expériences d'ATAC-seq n'ont pas été effectuées de la même façon, ce qui influence le nombre et la qualité des "reads" obtenus. Cependant l'analyse bio-informatique se fait de la même façon, donc suivent les mêmes contrôles de qualités et seuils de discrimination de séquences, et les données sont ensuite normalisées en fonction du nombre de paires de bases ouvertes total pour chaque échantillon. Selon les résultats, il semble y avoir plus de pics de chromatine ouverte comportant des SNPs dans notre modèle cellulaire de teloHAEC, ainsi que dans le cœur, ce qui était attendu étant donné leur implication dans les MCV (Figure 19). Le tissu de la muqueuse musculaire de l'œsophage comportait également une proportion similaire de SNPs dans des

MCV	# SNPs	HYP
5 274	total	12 443
274	dans pics ATACseq	606

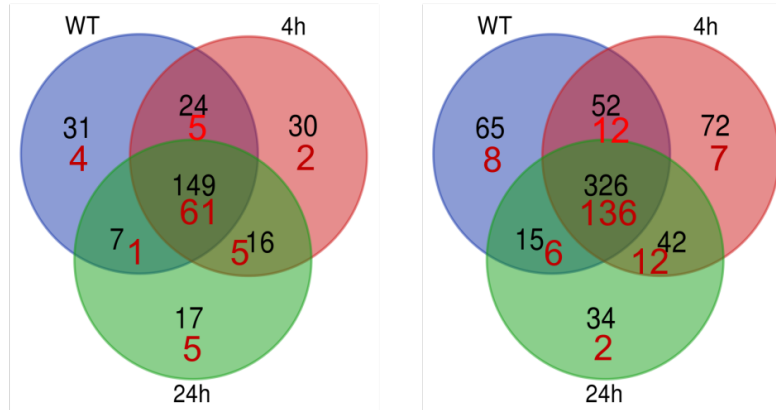


Figure 18 : Répartition des SNPs tombant dans des pics de chromatine ouverte selon les groupes. Nombres noirs : #SNPs total (considérant les SNPs en LD), nombres rouges : # signaux GWAS.

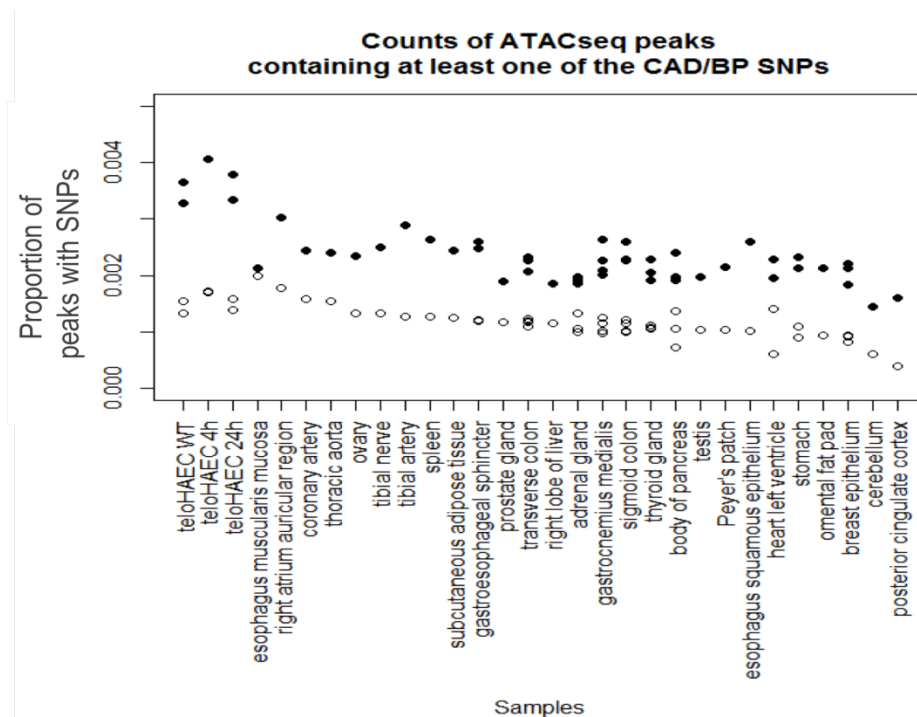


Figure 19 : Enrichissement de SNPs dans des pics de chromatine ouverte pour les teloHAEC et des tissus reliés aux fonctions endothéliales. Comparaison avec les données d'ENCODE dans 23 tissus. Chaque point représente un échantillon. Points blancs : SNPs pour les MCV, points noirs : SNPs pour l'HYP.

régions de chromatine ouverte bien qu'il ne soit pas un tissu lié aux MCV. Bien que ses fonctions et les types cellulaires le composant soient distincts, cela pourrait provenir de la morphologie comparable de ce tissu avec celui des artères puisqu'il s'agit d'un tube musculaire devant s'adapter à la circulation d'éléments, telles les artères s'adaptant au flux sanguin.

4.2.2 Débalancements alléliques dans l'ouverture de la chromatine

Le fait que des SNPs non codants tombent dans des régions de chromatine ouverte fait qu'ils sont à même de réguler l'expression génique. Cela peut se faire de plusieurs façons telles qu'en sollicitant ou réprimant le recrutement d'un activateur ou répresseur transcriptionnel. Cela pourrait entre autre venir d'une différence d'affinité entre la protéine et la séquence d'ADN comme d'une différence au niveau de l'ouverture de la chromatine qui est limitante pour l'accessibilité des protéines se liant sur l'ADN. Dans le cas de ce dernier scénario, il est possible de calculer un débalancement allélique, c'est-à-dire qu'un SNP à un locus donné sera davantage ouvert pour un allèle par rapport à un autre allèle. Pour la dysfonction endothéliale, seul deux SNPs présente un AI, soit rs35787595 et rs7667633 aux conditions de 4h ($p = 0,0005$) et 0h ($p = 0,0009$) respectivement (Tableau 15). À noter que ceux-ci ne sont pas significatifs lorsque les p-valeurs sont corrigées pour le nombre de tests.

Tableau 15 : Association de SNPs à un débalancement allélique dans l'ouverture de la chromatine.

SNPs dans un pic	WT		4h		24h		SNPs associés par GWAS	gène
	# séq.	p-valeur	# séq.	p-valeur	# séq.	p-valeur		
rs35787595	15	0,8	12	0,0005	11	0,2	rs3851738	CFDP1
rs7667633	22	0,0009	35	0,002	32	0,6	rs4593108	EDNRA

Le gène CFDP1 (craniofacial development protein 1) situé sur le chromosome 16 encode une protéine ubiquitaire (GTEx) qui a pour fonctions l'adhésion cellulaire, le développement de l'organisme multicellulaire ainsi que la régulation de l'apoptose des fibroblastes, de la prolifération et de la morphologie cellulaire (UniProt). Une des deux isoformes de cette protéine a d'ailleurs récemment été reportée pour son rôle dans l'organisation de la chromatine (110). Le gène EDNRA (Endothelin-1 receptor) localisé sur le chromosome 4 encode un récepteur à la membrane cellulaire exprimé dans plusieurs tissus dont principalement le cœur et les poumons (GTEx) impliqué entre autres dans le

développement du cœur, la régulation de la pression sanguine et la vasoconstriction (UniProt). Toutefois, considérant le peu de résultats et l'ordre de grandeur des p-valeurs, il est difficile d'en tirer des conclusions. Considérant le peu de données dû à l'expérience faite en *duplicata*, il faut cependant considérer qu'un faible signal sera difficilement mesurable.

4.2.3 Régions de chromatine différenciellement ouvertes avec un traitement au $TNF\alpha$

Avec les données de chromatine ouverte obtenues par ATAC-seq, il est possible de regarder si l'induction d'un dysfonctionnement endothélial entraîne une modification de l'ouverture de la chromatine. Dans un tel cas, il est attendu que ces loci soient d'intérêt dans le cadre des MCV. Pour ce faire, la couverture des pics ATAC-seq a été mesurée et une analyse a été faite pour déterminer une ouverture différentielle, signe qu'il y aurait un changement de conformation de la chromatine. Sur les 95 491 pics, 3133 sont différenciellement ouverts, soit près de 3% des pics. Ceux-ci sont répartis sur le génome (Annexe 3). Il a été démontré que les variants responsables d'un trait complexe sont distribués uniformément. Cela serait dû au fait que les gènes sont suffisamment inter-reliés pour former un réseau dans lequel même les gènes indirectement impliqués seront modulés. Toutefois, ils sont enrichis dans les éléments régulateurs de l'expression et les régions transcriptionnellement actives dans les tissus impliqués pour le trait complexe (106). Cette répartition est donc considérée normale et les pics différenciellement ouverts sont utilisés comme première information lors de la sélection des régions d'intérêt pour prioriser les SNPs d'intérêt dans la dysfonction endothéliale (Section 5.1.5).

4.2.4 Enrichissement de motifs liés par des facteurs de transcription

La majorité des SNPs étant dans des régions non-codantes de l'ADN, il est attendu que leur effet soit au niveau de la régulation de l'expression génique. Dans les faits, Il existe un enrichissement de variants non codants dans les éléments régulateurs en cis chez l'humain (111). Une possibilité de mécanisme serait que le variant change la séquence d'un motif de liaison de sorte à modifier la capacité des facteurs de transcription à le lier. D'ailleurs, il y aurait une corrélation entre l'ouverture de la chromatine et la liaison des TFs (112). Ces protéines sont des activateurs et répresseurs transcriptionnels qui vont agir au niveau de la régulation et l'initiation de la transcription. En effet, ils sont responsables en grande partie des patrons d'expression génique spécifique aux types cellulaires (113, 114). Ils ont d'ailleurs une part importante dans les maladies (115). Il est estimé que des motifs différenciellement liés par ces protéines lors d'une dysfonction endothéliale indiquent une voie biologique d'intérêt pour les MCV. Afin d'explorer cette avenue, les motifs liés ont été comparés entre

les conditions pour voir s'il y avait un enrichissement de motifs liés pour chaque facteur de transcription. Pour ce faire, l'algorithme CENTIPEDE basé sur un modèle probabiliste de mélange bayésien hiérarchique a été employé. Ce dernier identifie tous les motifs de liaison potentiels à partir de la base de données JASPAR (Version 2018) de motifs connus. Pour chaque motif, il y a une matrice (Position Weight Matrix (PWM)) avec l'importance de chaque nucléotide à chaque position qui prend en compte le fait que les motifs ne sont pas des séquences fixes. Il calcule une probabilité *a priori* d'être lié pour chaque motif trouvé selon des informations génomiques telles qu'un score global pour la séquence du motif, la conservation évolutive de séquence et la distance avec le site d'initiation de la transcription le plus proche. Les scores alléliques pour une matrice de fréquences alléliques à chaque position du motif se calculent comme suit :

$$W_{att.,obs.} = \log_2 \frac{p(att., obs.)}{p(att.)}$$

où $p(att.)$ est la probabilité de fréquence de chaque nucléotide estimée à 0,25 étant donné la distribution relativement uniforme attendue des 4 nucléotides possibles et $p(att., obs.)$ est la fréquence du nucléotide à une position déterminée dans le motif. Une seconde matrice avec les scores alléliques est ainsi générée. Un score global, correspondant à la somme des scores alléliques de l'allèle retrouvé à chaque position, est ensuite attribué à l'occurrence du motif. Une p-valeur de la probabilité qu'une séquence aléatoire ait un score global supérieur et une q-valeur du taux de faux positifs sont également calculées pour l'occurrence du motif. Par la suite, CENTIPEDE utilise des données expérimentales comme celles de Dnase-seq et histone ChIP-seq pour calculer une probabilité *a posteriori* que le motif soit lié. Dans le cas présent, ce sont les données d'ATAC-seq qui ont été utilisées. Cela permet de classer les motifs dans les catégories « liés » ou « non liés ». Ainsi, les motifs d'intérêts sont identifiés selon un seuil de probabilité de liaison, soit 99% pour cette analyse (Figure 20). Ces données sont ajoutées à la cartographie (Section 5.1.5).

Ensuite, pour analyser s'il y a une liaison différentielle des motifs, le nombre de motifs liés pour chaque facteur de transcription est compté puis normalisé selon le nombre de paires de bases ouvertes pour l'échantillon. Un test de t est ensuite effectué pour chaque TFs entre les conditions et une p-valeur est calculée. Tous facteurs de transcription ayant des motifs différentiellement liés avec des p-valeur de 0,05 ou inférieure entre au moins deux conditions sont considérés. Un total de 64 motifs sur les 719 testés ressortent positifs (Annexe 4). Une analyse de DAVID a été réalisée sur les TFs liant ces motifs afin de faire ressortir les voies biologiques impliquées. Des 44 voies impliquées significativement, il

faut noter la redondance des TFs tels que JUN et NFKB1 sans lesquels ce nombre serait drastiquement inférieur (Tableau 16). JUN encode le TF AP-1 régulant l'expression génique sous la voie de signalisation de l'AMPc (UniProt). NFKB1 (Nuclear factor NF-kappa-B p105 subunit) est un TF pléiotropique impliqué dans de nombreuses voies de signalisation et participant entre autres dans l'inflammation, la croissance cellulaire et l'apoptose (UniProt). Ainsi, en prenant en compte cette redondance, seule les premières lignes du tableau seraient à considérer. Toutefois, ces voies de différenciation des ostéoclastes, de mauvaise régulation de la transcription dans le cancer et voie de signalisation des MAPKs sont générales et vastes de sorte qu'il est hasardeux d'en déduire quelque chose.

MCV	# SNPs	HYP
5 274	total	12 443
100	dans motifs liés	184

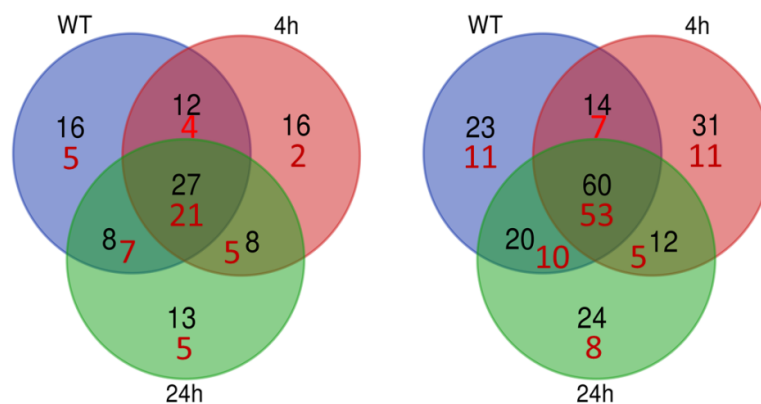


Figure 20 : Répartition des SNPs tombant dans des motifs liés par des facteurs de transcription selon les groupes. Nombres noirs : #SNPs total (considérant les SNPs en déséquilibre de liaison), nombres rouges : # signaux GWAS.

Tableau 16 : Analyse de voies biologiques de DAVID des TFs ayant des motifs différentiellement liés.

Termes	Gènes impliqués		p-valeur	Score Benjamini	
	(Compte absolu)	(%)			
Différenciation des ostéoclastes	JUN, FOSL2, NFKB2, NFKB1, IRF9, fosl1, junb, JUNB	8	13,8	1,89E-13	1,45E-11
Infection à HTLV-I	JUN, E2F1, EGR2, ATF1, ATF3, ELK4, NFKB2, NFKB1, fosl1	9	15,5	8,16E-13	3,14E-11
Mauvaise régulation de la transcription dans le cancer	ATF1, ELK4, CEBPE, NFKB1, ETV1, RXRA, TFE3, SIX1	8	13,8	8,64E-11	2,22E-09
Carcinogenèse virale	JUN, EGR2, NFKB2, NFKB1, IRF9, EGR3	6	10,3	1,20E-08	2,31E-07
Voie de signalisation des MAPKs	JUN, ELK4, NFKB2, NFKB1, junb	5	8,6	6,84E-07	1,05E-05
Hépatite B	JUN, E2F1, EGR2, NFKB1, EGR3	5	8,6	3,52E-06	4,52E-05
Tuberculose	VDR, NFYA, RFX5, NFKB1	4	6,9	6,21E-06	6,83E-05
Les voies du cancer	JUN, E2F1, NFKB2, NFKB1, RXRA	5	8,6	8,85E-06	8,51E-05
La légionellose	NFKB2, NFKB1, HSF1	3	5,2	8,70E-05	7,44E-04
Grippe A	JUN, NFKB1, IRF9	3	5,2	3,02E-04	0,0023
La coqueluche	JUN, NFKB1, IRF8	3	5,2	3,02E-04	0,0023
Voie de signalisation du récepteur Toll-like	JUN, NFKB1, IRF5	3	5,2	4,02E-04	0,0028
Cycle cellulaire	E2F1, TFD1, E2F4	3	5,2	5,16E-04	0,0033
Hépatite C	NFKB1, RXRA, IRF9	3	5,2	6,43E-04	0,0038
Maladie du foie gras non alcoolique (NAFLD)	JUN, NFKB1, RXRA	3	5,2	7,84E-04	0,0043
Voie de signalisation du TNF	JUN, NFKB1, JUNB	3	5,2	9,39E-04	0,0048
Cancer du poumon à petites cellules	E2F1, NFKB1, RXRA	3	5,2	9,39E-04	0,0048
Infection à herpès simplex	JUN, NFKB1, IRF9	3	5,2	9,39E-04	0,0048
Infection par le virus d'Epstein-Barr	JUN, NFKB2, NFKB1	3	5,2	9,39E-04	0,0048
Maladie inflammatoire de l'intestin (MICI)	JUN, TBX21, NFKB1	3	5,2	0,0011	0,0053
Signalisation des cellules épithéliales dans l'infection à Helicobacter pylori	JUN, NFKB1	2	3,4	0,0116	0,0515
Infection à Salmonella	JUN, NFKB1	2	3,4	0,0154	0,0644
Voie de signalisation NF-kappa B	NFKB2, NFKB1	2	3,4	0,0154	0,0644
Traitement et présentation de l'antigène	NFYA, RFX5	2	3,4	0,0154	0,0644
Maladie de Chagas (trypanosomiase américaine)	JUN, NFKB1	2	3,4	0,0193	0,0759
Voie de signalisation de l'adipocytokine	NFKB1, RXRA	2	3,4	0,0231	0,0860
La leishmaniose	JUN, NFKB1	2	3,4	0,0231	0,0860
Voie de signalisation du récepteur des cellules T	JUN, NFKB1	2	3,4	0,0269	0,0951
Voie de signalisation du récepteur des cellules B	JUN, NFKB1	2	3,4	0,0269	0,0951
Voie de signalisation des neurotrophines	JUN, NFKB1	2	3,4	0,0269	0,0951
MicroARN dans le cancer	E2F1, NFKB1	2	3,4	0,0344	0,1154
Synthèse et sécrétion d'aldostérone	ATF1, NR4A2	2	3,4	0,0344	0,1154
Dépendance à la cocaïne	JUN, NFKB1	2	3,4	0,0344	0,1154
Cancer du poumon non à petites cellules	E2F1, RXRA	2	3,4	0,0344	0,1154
Rougeole	NFKB1, IRF9	2	3,4	0,0344	0,1154
La leucémie myéloïde chronique	E2F1, NFKB1	2	3,4	0,0382	0,1222
Voie de signalisation du TGF-bêta	TFDP1, E2F4	2	3,4	0,0382	0,1222
Voie de signalisation de la prolactine	NFKB1, ESR2	2	3,4	0,0382	0,1222
Cancer du pancréas	E2F1, NFKB1	2	3,4	0,0382	0,1222
Voie de signalisation des œstrogènes	JUN, ESR2	2	3,4	0,0419	0,1284
voie de signalisation de l'AMPc	JUN, NFKB1	2	3,4	0,0419	0,1284
voie de signalisation des Wnt	JUN, fosl1	2	3,4	0,0494	0,1444

4.2.5 Cartographie de la conformation chromosomique et du transcriptome des *teloHAEC* et priorisation de loci d'intérêt pour les MCV.

Les données de RNAseq, de Hi-C et d'ATAC-seq ont été combinées par Simon Lalonde sur un navigateur en ligne (<http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=snHalk1oBv&statusId=19443707>) avec les SNPs et les TFs liés afin d'être visualisées. Un schéma explicatif de ce navigateur est présenté en annexe (Annexe 5). Concernant les données du RNAseq, une analyse d'expression différentielle de même que l'analyse des voies biologique avec l'outil en ligne DAVID ont été réalisées par Ken Sin Lo. Il en ressort que la principale voie impliquée est celle de la signalisation du TNF α . Cela valide la stimulation au TNF α pour enclencher la dysfonction endothéliale. Les données Hi-C permettent de regarder les interactions de la chromatine de manière à identifier des TADS, mais la technique est limitée par la résolution qui est délimitée par la fragmentation du génome en « boîte » de taille définie lors de l'analyse bioinformatique effectuée par Ken Sin Lo. Pour pallier cette contrainte de résolution, une expérience de promotor Hi-C permettrait de lier les séquences régulatrices avec des promoteurs de manière à restreindre la capture des interactions à celles pertinentes dans la régulation de l'expression génique et donc liant les SNPs aux gènes et pouvant jouer un rôle dans les MCV. Il serait également possible d'effectuer une expérience de ChIPseq pour déterminer la position des différentes marques d'histone et pouvoir ainsi catégoriser l'état activé ou réprimé de l'expression génique.

Lors de la priorisation des loci d'intérêts pour les MCV, les régions de chromatine différenciellement ouvertes sont utilisées comme première information. Parmi tous les SNPs associés aux MCV ou en LD, un total de 4 SNPs reliés aux MCV et 16 SNPs liés à l'HYP tombent dans 3 et 12 pics ATAC-seq différenciellement ouverts respectivement (Tableau 17). Pour l'ensemble des régions de chromatine différenciellement ouvertes contenant des SNPs, les données de RNAseq, de Hi-C et d'ATAC-seq ainsi que celles obtenues à partir de plusieurs logiciels d'analyse en ligne ont été utilisées pour déterminer les loci les plus pertinents à caractériser pour commencer (Tableau 17). HaploReg (version 4.1) permet d'obtenir de l'information sur les séquences non-codantes du génome tel que le chevauchement avec des promoteurs et amplificateurs (enhancers), les motifs liés par des protéines et l'effet des SNPs sur ceux-ci, les eQTLs et le gène le plus près (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) (116). RegulomeDB (version 1.1) donne un score aux SNPs représentatif du risque que le celui-ci a d'avoir un effet sur la liaison des protéines (<http://www.regulomedb.org/>) (117). Ce score est basé sur des données d'éléments régulateurs et fonctionnels du génome tels que des modifications

Tableau 17 : Résumé des informations recueillies sur les 20 SNPs dans les pics de chromatine différenciellement ouverts. Les pics en rouge sont ceux sélectionnés pour être caractérisés avec l'édition de génome. Le panneau du haut comprend les données générées par le laboratoire alors que celui du bas (page suivante) contient les données récoltées à partir des logiciels en lignes. Rouge = loci sélectionnés, vert = augmentation significative, jaune = diminution significative.

Maladie	SNP							Pic ATACseq			RNAseq			LogFC			Hi-C						ATACseq		LogFC						
	GWAS	LD	Chr	Pos	REF	ALT	Génotype	Chr	Début	Fin	Gène	p ajustée	WT 4h	WT 24h	4h 24h	Chr_A	Début_A	Fin_A	Chr_B	Début_B	Fin_B	p ajustée	WT 4h	WTt 24h	4h 24h						
MCV	rs2820315	rs59474075	1	201885704	C	N	0/0	1	201885494	201887294												3.18E-05	0.0401	-0.3863	-0.4264						
MCV	rs2820315	rs34091558	1	201886769	TA	N	0/0																								
MCV	rs67180937	rs17163363	1	222828704	C	N	0/0	1	222827855	222828970	AIDA	5.21E-19	0.5427	0.4969	-0.0458	Faible interaction avec le promoteur de AIDA						1.34E-04	0.7160	0.6198	-0.0962						
MCV	rs13003675	rs283484	2	233646505	A	N	0/0	2	233646104	233647118	NGEF	4.26E-29	-0.1188	-0.3390	-0.2202							1.95E-07	0.5926	0.5750	-0.0176						
											INPP5D	2.40E-85	-0.5154	-0.0530	0.4625																
HYP	rs12405515	rs10752944	1	172363524	C	T	0/1	1	172363033	172363916	TNFSF18	0.00E+00	1.1355	0.9522	-0.1833							2.39E-05	0.8320	0.4771	-0.3549						
											TNFSF4	1.32E-62	0.2357	0.3824	0.1466																
HYP	rs2761436	rs11335756	1	208005668	C	N	0/0	1	208004888	208005909	CD46P1,CR1	5.38E-63	0.4027	0.7610	0.3582	CTL : 1	208000000	208005000	1	208130000	208135000	4.96E-14	1.0792	0.8392	-0.2400						
											PLXNA2	1.90E-17	-0.3091	-0.0671	0.2419	TNF : 1	207990000	208000000	1	208130000	208140000										
HYP	rs7437940	rs6811851	4	7911771	A	C	1/1	4	7911739	7912587	HTRA3	8.49E-74	-0.2991	-0.5801	-0.2810							8.86E-05	0.6167	0.4354	-0.1813						
HYP	rs10057188	rs10050467	5	77846683	C	T	0/1	5	77846602	77847346											4.02E-05	0.7857	0.6990	-0.0867							
HYP	rs10057188	rs10050475	5	77846722	C	G	0/1																								
HYP	rs2188962	rs11741255	5	131811182	G	N	0/0																								
								5	131810268	131812028	CSF2	0.00E+00	1.8771	1.5789	-0.2983	Locus riche en interactions dont plusieurs partant de gènes DE															
										C5orf56	1.45E-65	0.5284	0.5052	-0.0233														2.37E-47	1.1369	0.5972	-0.5397
										AC116366.6	9.86E-71	0.2997	0.4325	0.1328																	
										IRF1	0.00E+00	1.2449	0.8564	-0.3885																	
HYP	rs1953126	rs6478486	9	123655329	T	C	1/1	9	123654540	123656317	AFAP1L1,GR	5.35E-18	-0.0821	0.2307	0.3128	Interaction avec le promoteur de TRAF1 et plusieurs autres interactions															
										TRAF1	0.00E+00	1.6731	1.4761	-0.1971														6.85E-07	0.4701	0.2613	-0.2088
											GSN	1.93E-197	-0.2007	-0.5443	-0.3435																
HYP	rs1953126	rs1860823	9	123659304	T	C	1/1	9	123659276	123659833												2.61E-06	0.9691	0.5667	-0.4023						
HYP	rs1953126	rs1860824	9	123659339	G	T	1/1																								
HYP	rs1953126	rs201993692	9	123694228	ATT	N	0/0	9	123693555	123694484												3.93E-04	0.3939	0.1693	-0.2246						
HYP	rs7977389	rs12310510	12	50066244	T	C	0/1				RND1	1.12E-98	1.4748	0.7513	-0.7235	Pic au center de la boîte HiCCUP Interaction avec le promoteur de FMNL3															
HYP	rs7977389	rs7954994	12	50068089	C	T	0/1	12	50066044	50068301	RHEBL1	2.57E-13	0.2681	0.3425	0.0744													3.00E-08	0.1864	-0.3412	-0.5276
											FMNL3	1.45E-106	0.3037	0.2338	-0.0699																
HYP	rs2034618	rs58006707	15	83869063	C	N	0/0	15	83868831	83869652												4.68E-04	0.7515	0.7677	0.0162						
HYP	rs6060114	rs6058197	20	30194899	A	N	0/0	20	30194739	30196046	ID1	1.18E-107	-0.6894	-0.4984	0.1909	CTL : 20	30150000	30155000	20	30195000	30200000										
HYP	rs6060114	rs6088736	20	30195166	C	N	0/0				CCM2L	1.99E-33	-0.2558	-0.4810	-0.2253	TNF : 20	30150000	30155000	20	30195000	30200000	3.49E-09	-0.3499	-0.3868	-0.0369						
HYP	rs11701033	rs7276540	21	33823727	C	N	0/0	21	33823012	33823984	IFNGR2	7.03E-237	0.4714	0.2112	-0.2602							6.07E-04	0.6416	0.3163	-0.3253						

Maladie	SNP							Pic ATACseq			HaploReg v4.1				RegulomeDB		GTEx		
	GWAS	LD	Chr	Pos	REF	ALT	Génotype	Chr	Début	Fin	Promoter histone marks	Enhancer histone marks	Proteins bound	Motifs changed	eQTL	Gène à proximité		Score	Signification
MCV	rs2820315	rs59474075	1	201885704	C	N	0/0	1	201885494	201887294	CRVX	ESDR, LNG, FAT, STRM, BRST, MUS, SKIN, VAS, GI, ADRL, PLCNT, OVRY, HRT, LIV, BRN, BONE	ZNF263	HDAC2,HMG-IY,MeF2,Ncx,Sox,Zfp105,p300		LMOD1	2b	Likely to affect binding	-
MCV	rs2820315	rs34091558	1	201886769	TA	N	0/0				FAT, STRM, MUS, SKIN, GI, LNG, CRVX	ESDR, LNG, FAT, BRST, SKIN, BRN, GI, ADRL, MUS, PLCNT, HRT, OVRY, VAS, BLD, BONE	CFOS,CJUN,GATA2,AP2ALPHA,CEBPB,ELK4,JUND,P300,STAT3	DMRT2,Foxf2,Foxo,GATA,GR		LMOD1	2b	Likely to affect binding	-
MCV	rs67180937	rs17163363	1	222828704	C	N	0/0	1	222827855	222828970	BLD	BLD, STRM, LNG, BONE		Pax-4	33	MIA3	4	Minimal binding evidence	RP11, MIA3, AIDA, FAM177B, BROX, TAF1A
MCV	rs13003675	rs283484	2	233646505	A	N	0/0	2	233646104	233647118		FAT, BRST, BLD, STRM, SKIN, GI, ADRL, HRT, PANC, LNG, CRVX, MUS	NFKB,BCL11A,STAT3,P300	Myb,Nkx6-1,Pou1f1,VDR	11	GIGYF2	2b	Likely to affect binding	C2orf82, KCNJ13, GIGYF2
HYP	rs12405515	rs10752944	1	172363524	C	T	0/1	1	172363033	172363916		ESDR, FAT, BRST, BLD, STRM, MUS, BRN, SKIN, GI, HRT, LNG, VAS, BONE		Foxa,Pou2f2	4	PIGC	2b	Likely to affect binding	C1orf105, SUCO, PIGC
HYP	rs2761436	rs113535756	1	208005668	C	N	0/0	1	208004888	208005909	BLD, SKIN	BRST, BLD, SKIN, BRN, GI, ADRL, PLCNT, THYM, HRT, LNG, PANC, MUS, SPLN, VAS		Foxk1,Foxo,Sox,TEF	1	C1orf132	4	Minimal binding evidence	-
HYP	rs7437940	rs6811851	4	7911771	A	C	1/1	4	7911739	7912587	FAT, STRM, MUS, GI, LNG, BRN, SKIN, BONE	ESC, ESDR, LNG, IPSC, BRST, BLD, BRN, SKIN, FAT, GI, HRT, MUS, PLCNT, OVRY, LIV, VAS		Foxi1,HMG-IY,Sox	8	AFAP1	5	Minimal binding evidence	AFAP1
HYP	rs10057188	rs10050467	5	77846683	C	T	0/1	5	77846602	77847346					2	LHFPL2	5	Minimal binding evidence	-
HYP	rs10057188	rs10050475	5	77846722	C	G	0/1	5	77846602	77847346					2	LHFPL2	4	Minimal binding evidence	-
HYP	rs2188962	rs11741255	5	131811182	G	N	0/0	5	131810268	131812028	BLD, LIV	ESC, ESDR, LNG, IPSC, FAT, STRM, BRST, BLD, SKIN, LIV, BRN, GI, ADRL, HRT, MUS, PLCNT, THYM, PANC, BONE		GR,MAZR,NRSF,Nanog,Pou2f2,Rad21,SMC3,TCF12,WT1,ZNF263	44	C5orf56	5	Minimal binding evidence	SLC22A5, P4HA2, PDUM4, ACSL6
HYP	rs1953126	rs6478486	9	123655329	T	C	1/1	9	123654540	123656317	BLD	LNG, FAT, BRST, BLD, MUS, SKIN, VAS, GI, ADRL, HRT, PLCNT, THYM, SPLN, BRN, BONE	POL2,SP1,YY1,NFKB,OCT2,POL24H8,POU2F2,POL2B	Hic1,PTF1-beta,Sin3AK-20,ZBTB7A	55	9.3kb 3' of TRAF1	3a	Less likely to affect binding	PSMD5, TRAF1, PHF19, FBXW2, C5, RAB14, MEGF9
HYP	rs1953126	rs1860823	9	123659304	T	C	1/1	9	123659276	123659833	BLD	ESC, IPSC, BLD, BRN, SKIN, FAT, GI, ADRL, HRT, LNG, MUS, PLCNT, THYM, SPLN	POL24H8,HDAC8	NF-kappaB	51	5.4kb 3' of TRAF1	4	Minimal binding evidence	PSMD5, TRAF1, PHF19, FBXW2, C5, MEGF9
HYP	rs1953126	rs1860824	9	123659339	G	T	1/1	9	123659276	123659833	BLD	ESC, IPSC, BLD, BRN, SKIN, FAT, GI, ADRL, HRT, LNG, MUS, PLCNT, THYM, SPLN	POL24H8,HDAC8	Dobox4	55	5.3kb 3' of TRAF1	4	Minimal binding evidence	PSMD5, TRAF1, PHF19, FBXW2, C5, RAB14, MEGF9
HYP	rs1953126	rs201993692	9	123694228	ATT	N	0/0	9	123693555	123694484	STRM, BLD, MUS, BONE	ESDR, LNG, FAT, BRST, BLD, STRM, BRN, SKIN, VAS, LIV, GI, ADRL, KID, MUS, PLCNT, THYM, HRT, SPLN	NFKB,BATF,EBF1,MEF2A,MEF2C,POL2,SP1,YY1,PAX5C20	Bcl6b,Evi-1,Foxo,Foxp1,HDAC2,Mef2,TATA,Zfp105	2.8kb 5' of TRAF1	2b	Likely to affect binding	-	
HYP	rs7977389	rs12310510	12	50066244	T	C	0/1	12	50066044	50068301	FAT, BLD, BRN, GI, MUS	ESC, ESDR, LNG, IPSC, FAT, STRM, BRST, BLD, MUS, BRN, SKIN, LIV, GI, ADRL, HRT, KID, PLCNT, THYM, OVRY, PANC, SPLN, VAS, BONE		Cdx2,Hoxb13,Hoxb9,Hoxd10,SRF,Sox	4	FMNL3	2b	Likely to affect binding	SPATS2, MCRS1, C1QL4
HYP	rs7977389	rs7954994	12	50068089	C	T	0/1				FAT, BLD, STRM, MUS, BRN, SKIN, GI, BONE	ESC, ESDR, LNG, IPSC, FAT, STRM, BRST, BLD, BRN, SKIN, VAS, GI, ADRL, HRT, KID, MUS, PLCNT, THYM, OVRY, PANC, SPLN, LIV	GATA1		4	FMNL3	4	Minimal binding evidence	SPATS2, MCRS1, C1QL4
HYP	rs2034618	rs58006707	15	83869063	C	N	0/0	15	83868831	83869652	BLD	BLD, MUS, SKIN	NFKB	AP-1,Mrg,Nrf-2,TCF11:-MaFG	5	RP11-382A20.3	2b	Likely to affect binding	HGDFRP3, HOMER2, RP11, C15orf40
HYP	rs6060114	rs6058197	20	30194899	A	N	0/0	20	30194739	30196046	ESC, ESDR, IPSC, FAT, STRM, BRST, SKIN, VAS, LIV, BRN, GI, PANC, LNG, MUS, OVRY, HRT, THYM, CRVX, BLD, BONE	ESDR, BLD, BRN, HRT, PLCNT, GI, SPLN	POL2,POL24H8,USF2,TCF4,TAF1,ZBTB33,HMGN3,POL2B	Ascl2,BCL2,ElA,Myf,NRSF,ZEB1,p300	16	89bp 5' of MIR3193	2b	Likely to affect binding	HM13, DEF8124, COX4I2, MYLK2, TPX2, FRG1B
HYP	rs6060114	rs6088736	20	30195166	C	N	0/0	20	30194739	30196046	ESC, ESDR, LNG, IPSC, FAT, STRM, SKIN, VAS, LIV, BRN, GI, ADRL, PANC, MUS, HRT, OVRY, CRVX, BRST, BLD, BONE	ESDR, IPSC, BRST, BLD, BRN, HRT, PLCNT, THYM, GI, SPLN, MUS	POL2,POL24H8,TAF7,POL2B,CEBPB,GT F2f1,TBP,CHD2,ELF1,FOSL2,FOXA1,FOXK2,HDAC2,HNFG4,JUND,P300,SIN3A,K20,SP1,TCF4,CMYC,HEY1,MAX,NX11,GATA1,GATA2,STAT3	HDAC2,HNFG4,FXRA,Znf143	13	122bp 3' of MIR3193	4	Minimal binding evidence	HM13, DEF8124, COX4I2, MYLK2, TPX2, FRG1B
HYP	rs11701033	rs7276540	21	33823727	C	N	0/0	21	33823012	33823984	BONE	LNG, IPSC, FAT, STRM, BRST, MUS, SKIN, BRN, GI, HRT, PLCNT, PANC, SPLN, VAS	STAT3	EWSR1-FL1		FAM176C	2a	Likely to affect binding	-

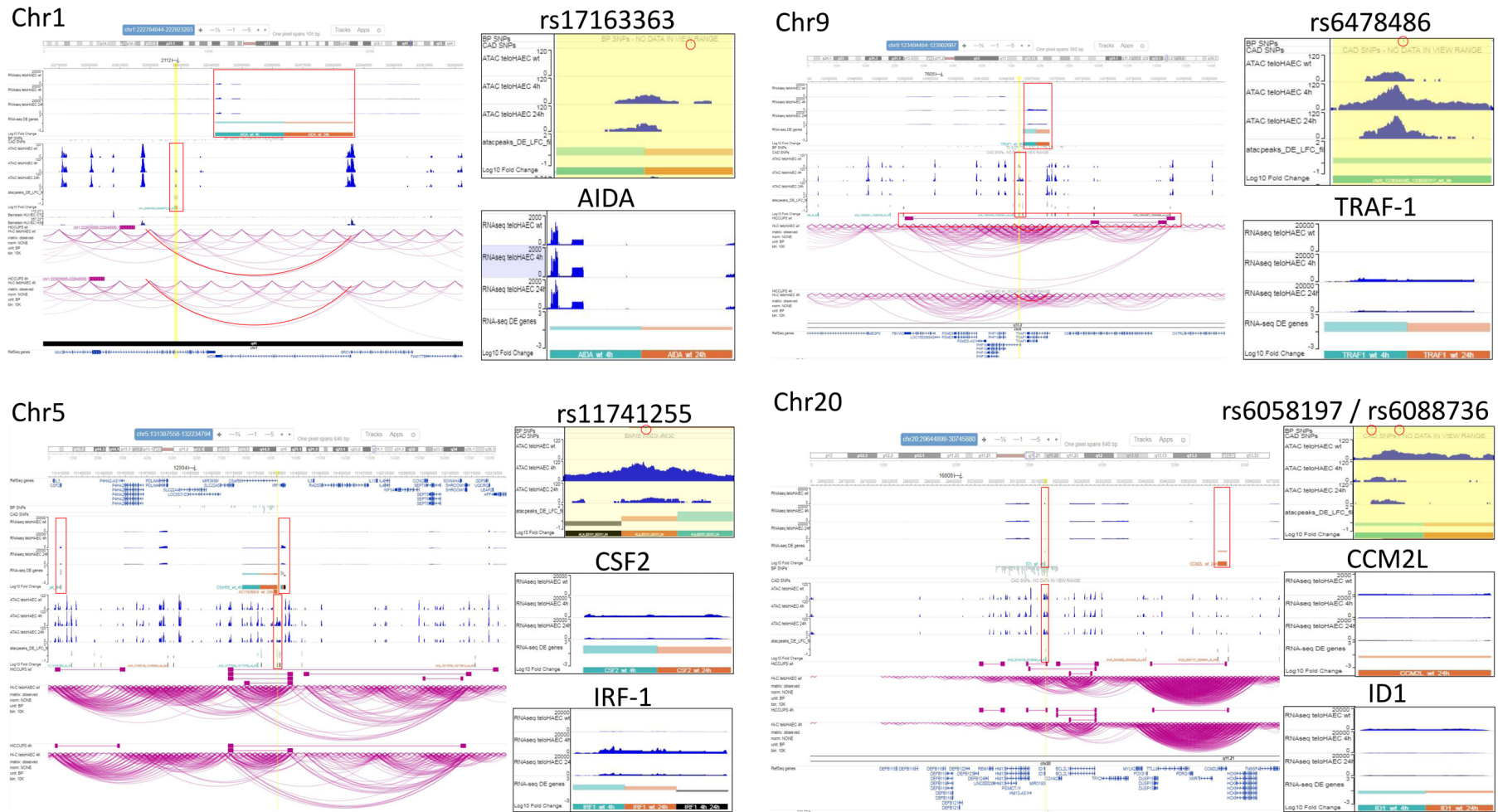


Figure 21 : Priorisation de 4 loci associés aux maladies coronariennes par GWAS. Les pics ATAC-seq différemment ouverts (ATAC-seq), comportant des SNPs (GWAS), à proximité de gènes différemment exprimés (RNAseq) et compris dans des régions d'interaction de la chromatine (Hi-C) sont délétés par CRISPR-Cas9. Les encadrés rouges représentent ce qui est significativement différent entre les groupes.

d’histones, la structure de la chromatine, les motifs de facteurs de transcription (117). GTEx, tel que mentionné précédemment, est une base de données transcriptomiques comprenant de multiples tissus et types cellulaires permettant d’identifier des eQTLs.

Au final, 4 de ces 20 régions sont sélectionnées (Figure 21) pour la suite des expériences en raison des gènes différentiellement exprimés à proximité (Tableau 18) et des différentes interactions chromosomiques. En effet, le premier pic de chromatine ouverte contenant le SNP rs17163363 interagit avec le promoteur du gène AIDA. Le second contient rs11741255, est dans une région riche en interactions chromosomiques et est collé sur le gène différentiellement exprimé IRF-1. Le troisième comporte le SNP rs6478486 et se trouve à un locus où il y a une perte d’interaction chromosomique avec la dysfonction endothéliale. Le quatrième cumule les deux SNPs rs6058197 et rs6088736 et est localisé dans une région riche en interactions chromosomiques.

Tableau 18 : Description des gènes DE dans les 4 régions sélectionnées (UniProt)

Chr	Gène DE	Rôle
1	AIDA (Axin interactor, dorsalization-associated protein)	Ventralisation lors de l’embryogénèse Inhibition de la voie de signalisation JNK
5	GM-CSF (Granulocyte-macrophage colony-stimulating factor)	Cytokine favorisant la croissance et différenciation cellulaire de granulocytes et macrophages
	IRF-1 (Interferon regulatory factor 1)	TF modulant l’expression génique de multiples processus cellulaires dont l’hématopoïèse, l’inflammation, la réponse immunitaire, la prolifération et la différenciation
9	TRAF1 (TNF receptor-associated factor)	Molécule adaptatrice dans l’activation de NF-kappa-B et JNK Permet l’ubiquitination et donc module l’apoptose
20	ID1 (DNA-binding protein inhibitor)	Régulateur transcriptionnel Inhibition du TF bHLH (basic helix-loop-helix) Régulation de multiples processus cellulaires dont l’apoptose, l’angiogénèse, la prolifération et la différenciation Inhibition de la différenciation des myocytes cardiaques
	CCM2L (Cerebral cavernous malformations 2 protein-like)	Régulation de l’adhésion intercellulaire, de la production de FGF (fibroblast growth factor) et de la morphogénèse cardiaque

4.2.6 Délétion de 4 loci d’environ 2kb par CRISPR-Cas9 avec double évènement de coupure Cas9

Afin d’étudier le rôle des 4 loci sélectionnés, une délétion d’environ 2kb a été réalisée en générant deux cassures double brins par le système d’endonucléase CRISPR pour retirer la séquence de la chromatine différentiellement ouverte. D’une part, cela complexifie l’expérience puisque deux évènements au sein

d'une cellule sont requis afin d'obtenir la délétion désirée, évènements dont l'efficacité varie grandement selon le modèle cellulaire, le loci ciblé et le mécanisme de réparation de l'ADN désiré (90). De plus, différents scénarios sont possibles, tels qu'aucune, une ou deux indels ou encore une recombinaison homologue avec le 2^e allèle pour réparer la séquence clivée, de sorte que la probabilité d'obtenir deux cassures causant le retrait de la séquence comprise entre celles-ci est moindre. Afin de valider la délétion, une PCR avec des amorces de part et d'autre de la délétion est réalisée (Annexe 6). Parce que l'efficacité de délétion effective est faible, des sous-populations ont été dérivées afin de pouvoir visualiser ce résultat positif. Les sous-populations positives sont présentement en train d'être dérivées afin d'obtenir des clones qui seront utilisés pour la caractérisation du rôle des loci délétés.

4- Discussion

Au niveau de la conformation chromosomique, il existe un enrichissement de SNPs associés aux MCV dans des régions de chromatine ouverte pour les EC (31). Les données des teloHAEC ont été comparées avec celles des divers tissus provenant de ENCODE pour tirer cette conclusion. Il faut cependant tenir compte des limitations de cette base de données qui ne contient pas d'expérience d'ATAC-seq dans des lignées cellulaires. Le fait que cela soit restreint à des tissus mixtes résulte en un signal qui est dilué par la diversité des cellules. De plus, les échantillons de ENCODE n'ont pas de phénotype de pathogénicité et ne sont pas soumis à des stimuli, de sorte qu'il n'est pas possible d'effectuer l'analyse dans un contexte de maladie.

Ensuite, il est considéré que les SNPs d'intérêt pour les MCV peuvent causer un AI dans des conditions spécifiques tel que la dysfonction endothéliale. Toutefois, l'analyse n'a permis d'identifier que deux SNPs marginalement significatif à une seule condition. Ces résultats étant non concluants, il faut considérer que ce n'est pas l'ensemble des SNPs associés par GWAS ou en LD qui sont hétérozygotes dans notre modèle cellulaire de teloHAEC dû au fait qu'il y a un génotype unique. Cela représente une limitation puisqu'il n'est pas possible de tester la totalité des SNPs. Pour pallier à cette lacune, il serait possible de faire l'expérience avec des lignées primaires de EC de différents génotypes. En plus de permettre d'observer l'ensemble des SNPs, les données seraient plus volumineuses de manière à augmenter la puissance de l'analyse et donc permettre d'identifier l'AI même s'il est minime. Plutôt que de regarder l'AI et pour poursuivre avec les données déjà générées, les pics différentiellement ouverts avec le traitement au TNF α ont été observés. Par la suite, il s'agissait de regarder les SNPs

tombant dans ces régions afin de déterminer lesquels sont impliqués dans les dysfonctions endothéliales et donc d'intérêt pour une investigation plus poussée.

Concernant la conceptualisation expérimentale des délétions réalisées par le système CRISPR-Cas9, la reconnaissance de deux séquences à proximité assure une spécificité accrue et permet une délétion de la région entière. D'autre part, la région délétée est large ce qui permet d'analyser le rôle de la région, mais pas nécessairement d'identifier les gènes causaux ou le variant responsable de la variation du risque. Ultérieurement, il serait possible d'explorer le rôle de ces loci de manière plus précise à l'aide de la nucléase Cas9 inactivée (deadCas9; dCas9) couplées avec des activateurs (VP64 : copies multiples du domaine trans-activateur de la protéine virale 16 de Herpes simplex (VP16)) ou répresseurs (KRAB : Krüppel-associated box domain of Kox1) transcriptionnels qui peuvent reproduire des gains ou pertes de fonction respectivement (118). La dCas9-VP64 peut activer modérément un gène par un facteur allant de 2 à 5 en moyenne et la dCas9-KRAB peut réprimer efficacement les gènes à près de 90-99% (119). Avec un unique sgRNA, le complexe protéique aura un effet jusqu'à 300bp du site d'initiation de la transcription (120), donc des régions moins larges comparativement aux délétions d'environ 2kb. Par la suite, il serait également possible d'utiliser des Cas9 ciblant un SNP de manière à créer de petits indels ou encore d'utiliser la Cas9 en combinaison avec un fragment d'ADN homologue portant l'allèle alternatif pour obtenir, via le mécanisme de réparation de l'ADN par recombinaison homologue, des cellules avec ce variant et ainsi directement caractériser l'effet de celui-ci, donc d'être encore plus précis (121).

L'ensemble de ces expériences ont permis de générer une cartographie détaillée à l'échelle génomique et de mettre en évidence quatre loci d'intérêt pour les MCV dans un contexte de dysfonction endothéliale. Le système CRISPR-Cas9 a été utilisé afin de retirer ces régions du génome et d'observer le phénotype résultant. Toutefois, considérant l'ensemble des SNPs associés aux MCV et à l'HYP ainsi que les SNPs en LD, il reste de nombreuses autres pistes à vérifier. Il serait possible de réaliser un screen CRISPR-dCas9 couplé avec le domaine activateur VP64 ou le domaine répresseur KRAB pour tous ces loci de manière à observer l'effet phénotypique des loci sur l'expression de marqueurs des fonctions endothéliales tels que les molécules d'adhésion et la production de NO. Les clones présentant une différence permettront de déterminer les SNPs impliqués dans les MCV.

5 - Conclusion

Les MCV ayant une forte répercussion sur l'économie et la santé globale, plusieurs GWAS ont permis d'associer de multiples SNPs à ces maladies. La causalité ne pouvant toutefois pas être inférée par ce type d'associations, mon projet de maîtrise c'est appliqué à investiguer l'implication de ceux-ci dans la modulation du risque de maladies sous un contexte de dysfonction endothéliale. rs9349379, un des SNPs associé au phénotype, dans le gène *PHACTRI* étant parmi les candidats les plus intéressants avec plusieurs éléments indiquant un effet causal, il fut le premier analysé. En raison de l'effet eQTL observé, soit une diminution de l'expression avec l'allèle alternatif G qui est associé à un plus haut risque de maladie, il semblerait que la réduction d'expression de *PHACTRI* affecte l'endothélium de manière à promouvoir l'athérosclérose. Toutefois, l'inactivation de *PHACTRI* n'altère pas de manière significative l'intégrité du cytosquelette d'actine et n'affecte guère l'expression génique au niveau d'une voie biologique de manière évidente ce qui réfuterait l'hypothèse initiale du rôle de *PACTRI* dans l'intégrité de l'endothélium vasculaire. Il faut cependant se rappeler que le mécanisme cellulaire liant ce SNP avec les MCV ne semble pas avoir un effet majeur de sorte qu'il est difficile de mettre en évidence son impact. Entre temps, il a été montré que le gène *EDNI* était également modulé par rs9349379 dans les SMC. Bien que cela n'empêche pas un rôle de *PHACTRI* dans les MCV en raison d'un rôle pléiotropique de rs9349379, ce fait combiné avec l'absence de résultats clairs de l'inactivation de *PHACTRI* enlève du poids à la pertinence de poursuivre l'investigation de ce gène. Parallèlement, la délétion de loci apparemment impliqués dans les MCV permettra d'investiguer sur le rôle de SNPs dans la modulation du risque de ces maladies. Cela permettra d'adresser la seconde hypothèse selon laquelle plusieurs des SNPs associés aux MCV influenceraient le risque de maladies via un rôle direct sur la fonction endothéliale.

Ainsi, mon projet de maîtrise a permis de caractériser le gène *PHACTRI* qui était alors très peu connu, de construire une cartographie exhaustive du transcriptome et de la conformation chromosomique à la grandeur du génome sous un phénotype de dysfonction endothéliale ainsi que de prioriser des loci modulant les fonctions endothéliales et donc potentiellement impliqués dans les MCV. Tel que mentionné, des expériences avec des cellules primaires, de ChIPseq et de promotor Hi-C ainsi qu'un screen CRISPR permettraient une meilleure investigation des SNPs associés aux MCV en permettant d'être plus représentatif de la réalité biologique, en allant chercher une information complémentaire pertinente sur la modulation génomique exercée par les SNPs et en appliquant l'analyse à l'ensemble

du génome respectivement. Ces expériences seront d'ailleurs incluses dans la suite des projets sur le long terme au laboratoire du Dr Lettre. L'arrêt de la sélection sur aussi peu que quatre loci permettait d'adresser rapidement la problématique pour que cela soit réaliste dans le cadre de ma maîtrise tout en donnant un aperçu de la caractérisation pouvant être menée. En ce sens, l'effet phénotypique des quatre délétions réalisées sera caractérisé sous stimulus inflammatoire pour mesurer l'effet sur l'adhésion cellulaire, la production d'oxyde nitrique et de radicaux libres, les propriétés anticoagulantes de même que la signalisation calcique.

Dans tous les cas, la compréhension génétique des traits complexes demeure un défi de taille de par la multitude de SNPs impliqués, de par l'effet individuel mineur de ces derniers et par l'impossibilité d'affirmer une causalité sans validation biologique. Il va sans dire que la compréhension des diverses voies impliquées constituerait un avancement considérable permettant d'exploiter celles-ci dans le développement d'un traitement alternatif dans les MCV et ainsi d'avoir un impact positif dans l'amélioration de la santé mondiale.

Annexe 1 - Observed *PHACTR1* transcripts

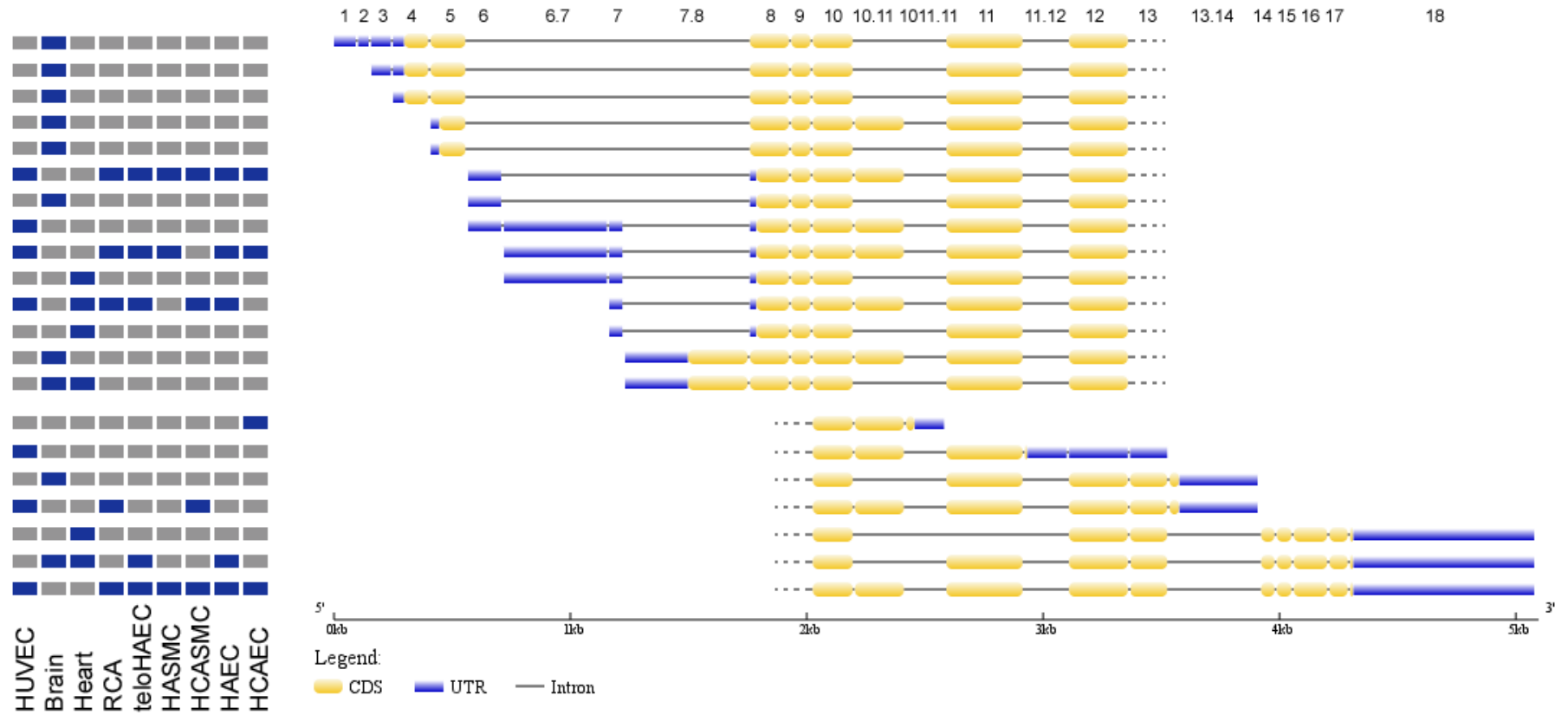
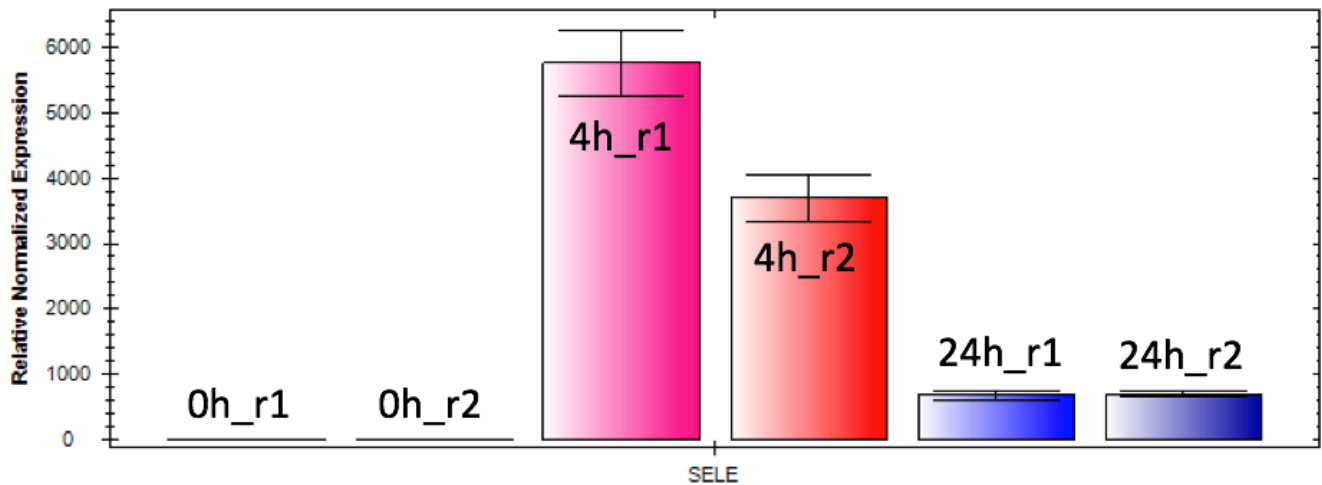


Figure S1. Summary of observed transcripts with their alternatively spliced 5' and 3' UTRs. The blue rectangles indicate transcripts that were seen in tissues samples whereas gray rectangles mean they were not observed. Introns aren't shown. Black lines indicate the absence of exons.

Annexe 2 - Traîtement au TNF α

A



B

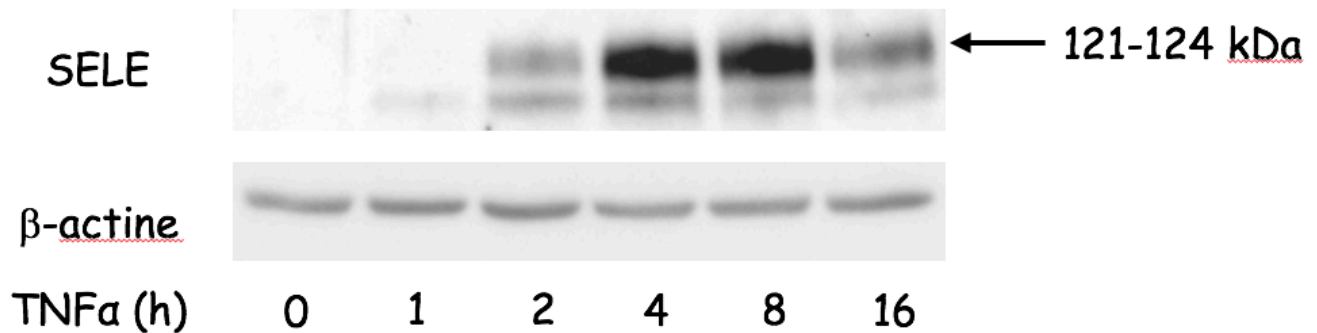


Figure 22 : Réponse des teloHAEC au stimulus de TNF α selon le temps d'exposition A) Expression relative des ARNm de *SELE*. Expérience en *duplicata*. Augmentation de l'expression après 4 heures et diminution à 24 heures. L'expression est normalisée avec les gènes TBP, HPRT et GAPDH. B) Expression relative des protéines de SELE. Augmentation puis diminution de l'expression avec le temps d'exposition pour une expression maximale entre 4 et 8 heures d'exposition. La β -actine sert de contrôle pour normaliser la quantité de protéines chargée sur gel.

Annexe 3 - Répartition uniforme de la chromatine ouverte

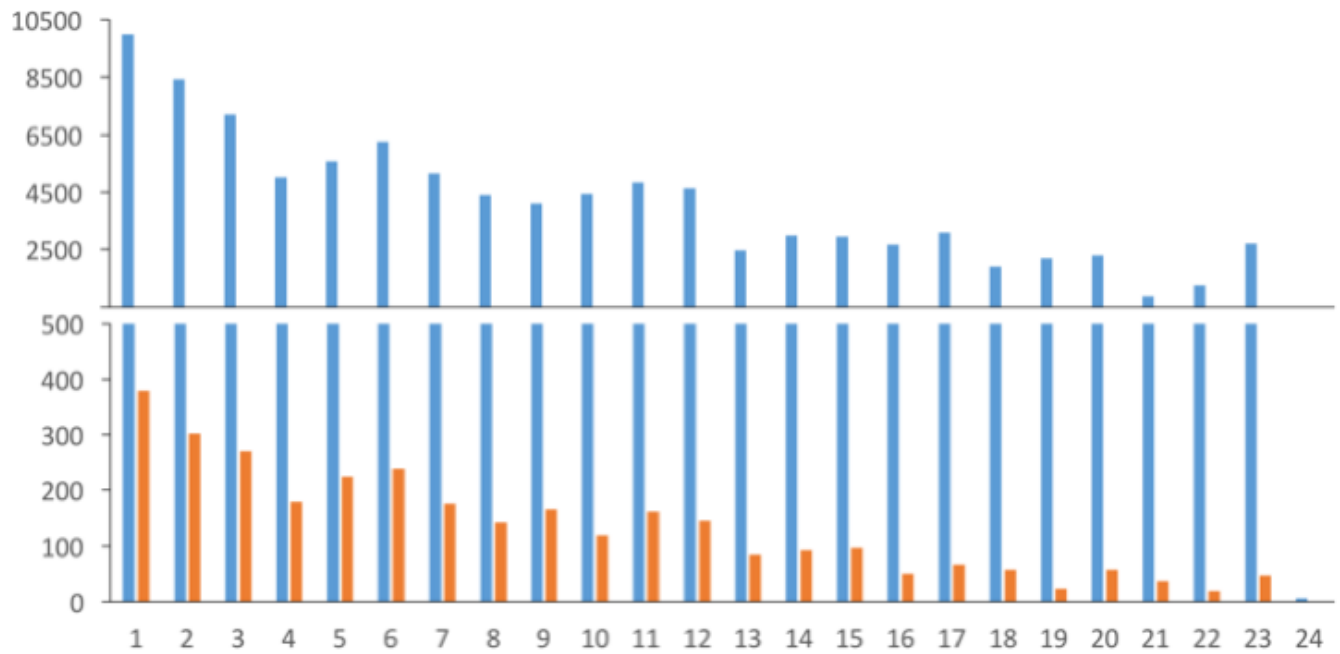


Figure 23 : Répartition des pics de chromatine ouverte selon les chromosomes. Bleu : pics ouverts, orange : pics différenciellement ouverts

Annexe 4 - Motifs différentiellement liés par des TFs.

Tableau 19 : 64 motifs différentiellement liés par des TFs déterminés par un test de t à partir de l'ensemble des motifs liés identifiés par CENTIPEDE. Jaune = liaison différentielle entre les deux conditions avec un $p \leq 0,05$.

Motifs	TFs	p-valeurs			Motifs	TFs	p-valeurs		
		wt_4h	wt_24h	4h_24h			wt_4h	wt_24h	4h_24h
MA0024.1	E2F1	0,9697	0,1100	0,0447	MA0736.1	GLIS2	0,6107	0,4739	0,0050
MA0024.2	E2F1	0,6196	0,4500	0,0139	MA0740.1	KLF14	0,7906	0,2708	0,0176
MA0060.1	NFYA	0,5286	0,2608	0,0109	MA0741.1	KLF16	0,6080	0,4614	0,0444
MA0060.2	NFYA	0,6795	0,1989	0,0440	MA0747.1	SP8	0,6958	0,5223	0,0116
MA0076.1	ELK4	0,4507	0,0300	0,0315	MA0749.1	ZBED1	0,2928	0,1258	0,0026
MA0105.4	NFKB1	0,2822	0,1731	0,0289	MA0761.1	ETV1	0,8602	0,1470	0,0475
MA0131.1	HINFP	0,5427	0,3141	0,0030	MA0770.1	HSF2	0,3554	0,6994	0,0325
MA0131.2	HINFP	0,8205	0,2480	0,0360	MA0778.1	NFKB2	0,2263	0,1636	0,0440
MA0258.2	ESR2	0,6106	0,3144	0,0374	MA0798.1	RFX3	0,3059	0,1252	0,0290
MA0470.1	E2F4	0,6260	0,3381	0,0104	MA0799.1	RFX4	0,2076	0,2620	0,0134
MA0472.2	EGR2	0,8609	0,4301	0,0244	MA0815.1	TFAP2C(var.3)	0,7766	0,3863	0,0205
MA0486.2	HSF1	0,3469	0,9908	0,0457	MA0831.1	TFE3	0,8273	0,2317	0,0427
MA0501.1	MAF::NFE2	0,0264	0,2157	0,6223	MA0831.2	TFE3	0,8790	0,2149	0,0140
MA0510.2	RFX5	0,1037	0,0141	0,0084	MA0834.1	ATF7	0,0884	0,0331	0,3347
MA0527.1	ZBTB33	0,8063	0,2662	0,0095	MA0837.1	CEBPE	0,1086	0,0470	0,4655
MA0591.1	Bach1::Mafk	0,0312	0,2196	0,5231	MA0872.1	TFAP2A(var.3)	0,8185	0,3713	0,0260
MA0600.1	RFX2	0,5106	0,2709	0,0065	MA0905.1	HOXC10	0,5789	0,2863	0,0338
MA0600.2	RFX2	0,0126	0,1478	0,0756	MA0911.1	Hoxa11	0,6205	0,2565	0,0499
MA0604.1	Atf1	0,4045	0,0045	0,2687	MA1118.1	SIX1	0,9036	0,3245	0,0323
MA0605.1	Atf3	0,0230	0,1640	0,4713	MA1122.1	TFDP1	0,6462	0,4580	0,0077
MA0611.1	Dux	0,5545	0,2231	0,0211	MA1127.1	FOSB::JUN	0,1114	0,0467	0,2932
MA0632.1	Tcf5	0,5374	0,3047	0,0123	MA1129.1	FOSL1::JUN(var.2)	0,0672	0,0468	0,0142
MA0652.1	IRF8	0,0434	0,0208	0,1032	MA1131.1	FOSL2::JUN(var.2)	0,0180	0,0205	0,2881
MA0653.1	IRF9	0,0430	0,0386	0,2143	MA1136.1	FOSB::JUNB(var.2)	0,0140	0,0180	0,0946
MA0656.1	JDP2(var.2)	0,1387	0,0920	0,0427	MA1139.1	FOSL2::JUNB(var.2)	0,0590	0,0262	0,4138
MA0690.1	TBX21	0,9953	0,4682	0,0343	MA1140.1	JUNB(var.2)	0,0293	0,0284	0,2337
MA0693.1	Ydr	0,9633	0,4452	0,0432	MA1143.1	FOSL1::JUND(var.2)	0,0397	0,0303	0,0127
MA0694.1	ZBTB7B	0,6051	0,8160	0,0160	MA1145.1	FOSL2::JUND(var.2)	0,0859	0,0402	0,1858
MA0695.1	ZBTB7C	0,6668	0,4692	0,0308	MA1146.1	NR1A4::RXRA	0,6395	0,5229	0,0123
MA0702.1	LMX1A	0,4938	0,9711	0,0079	MA1147.1	NR4A2::RXRA	0,6470	0,3696	0,0360
MA0703.1	LMX1B	0,4938	0,9711	0,0079	MA1419.1	IRF4	0,0597	0,0303	0,1904
MA0732.1	EGR3	0,6211	0,4397	0,0228	MA1420.1	IRF5	0,0130	0,0491	0,7918

Annexe 5 - Navigateur WashU Epigenome Browser

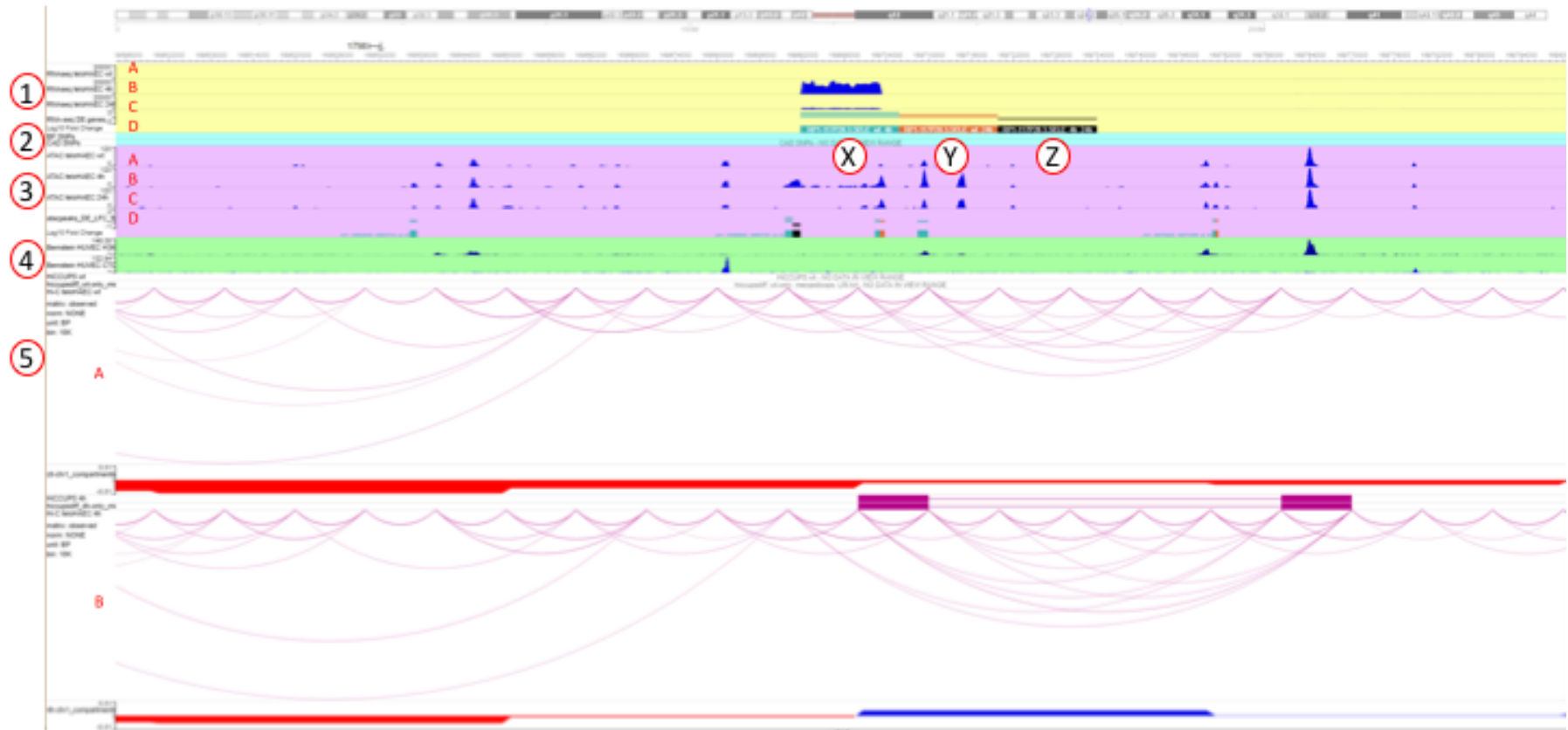


Figure 24 : Exemple du navigateur au loci de SELE présentant une expression différentielle du gène entre les trois conditions, trois pics de chromatine différemment ouverts et un gain d'interaction chromosomique sous stimuli. 1 (jaune): RNAseq, 2 (bleu): SNPs, 3 (mauve): ATACseq, 4 (vert): marques d'histone dans les HUVEC, 5 (blanc): Hi-C, A: WT, B: 4h, C: 24h, D: analyse différentielle, X: comparaison WT-4h, Y: comparaison, WT-24h, Z: comparaison 4h-24h

Annexe 6 - Validation de la délétion des 4 loci

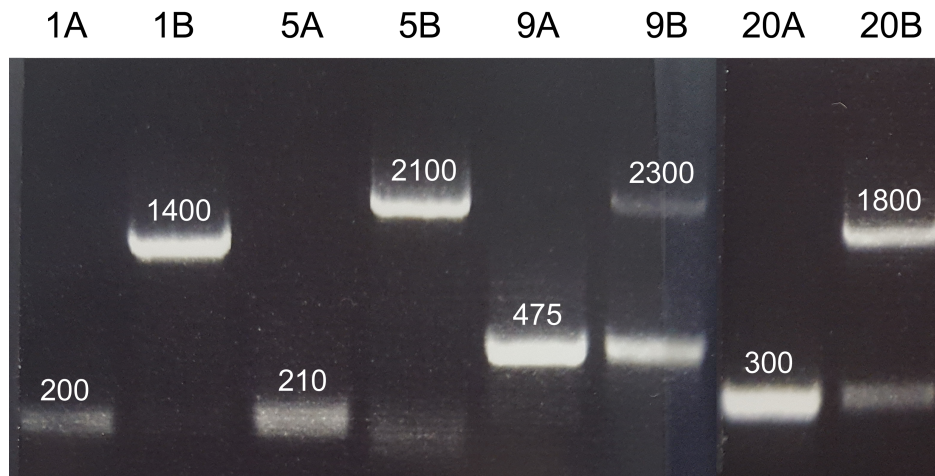


Figure 25 : Validation des délétions par PCR avec des amorces de part et d'autre des régions ciblées (amorces : Tableau 8). Chiffre = # chr, A = bande extraite non visible, purifiée et amplifiée une 2^e fois, B = première amplification à partir de l'ADNg.

6 - Références

1. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1545-602.
2. Suivi des maladies du cœur et des accidents vasculaires cérébraux au Canada. Agence de la santé publique du Canada; 2009.
3. Peden JF, Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum Mol Genet*. 2011;20(R2):R198-205.
4. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. *Nature*. 2011;473(7347):317-25.
5. Ozaki K, Tanaka T. Molecular genetics of coronary artery disease. *J Hum Genet*. 2016;61(1):71-7.
6. Brunner H, Cockcroft JR, Deanfield J, Donald A, Ferrannini E, Halcox J, et al. Endothelial function and dysfunction. Part II: Association with cardiovascular risk factors and diseases. A statement by the Working Group on Endothelins and Endothelial Factors of the European Society of Hypertension. *J Hypertens*. 2005;23(2):233-46.
7. Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis. *Cell*. 2011;145(3):341-55.
8. Lampugnani MG. Endothelial cell-to-cell junctions: adhesion and signaling in physiology and pathology. *Cold Spring Harb Perspect Med*. 2012;2(10).
9. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. *Nat Med*. 2011;17(11):1410-22.
10. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet*. 2004;364(9438):937-52.
11. Cahill PA, Redmond EM. Vascular endothelium - Gatekeeper of vessel health. *Atherosclerosis*. 2016;248:97-109.
12. Mayer B, Erdmann J, Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. *Clin Res Cardiol*. 2007;96(1):1-7.
13. Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*. 2010;19(3):227-9.

14. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol.* 2012;8(12):e1002822.
15. The International HapMap C. A haplotype map of the human genome. *Nature.* 2005;437:1299-320.
16. Ralston A. Gene interaction and disease: *Nature Education*; 2008 [1(1):16:[]
17. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-53.
18. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet.* 2014;94(4):559-73.
19. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):D1001-6.
20. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. RNA splicing is a primary link between genetic variation and disease. *Science.* 2016;352(6285):600-4.
21. Furlong LI. Human diseases through the lens of network biology. *Trends Genet.* 2013;29(3):150-9.
22. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res.* 2018;122(3):433-43.
23. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet.* 2017;49(9):1392-7.
24. Howson JMM, Zhao W, Barnes DR, Ho WK, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* 2017;49(7):1113-9.
25. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* 2017;49(9):1385-91.
26. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep.* 2017;7(1):2761.
27. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet.* 2017;49(1):54-64.
28. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Corrigendum: Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet.* 2017;49(10):1558.

29. Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet.* 2016;48(10):1162-70.
30. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47(10):1121-30.
31. Liu B, Pjanic M, Wang T, Nguyen T, Gloudemans M, Rao A, et al. Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. *bioRxiv.* 2018.
32. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet.* 2009;41(3):334-41.
33. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43(4):333-8.
34. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet.* 2012;44(8):890-4.
35. Winsvold BS, Bettella F, Witoelar A, Anttila V, Gormley P, Kurth T, et al. Shared genetic risk between migraine and coronary artery disease: A genome-wide analysis of common variants. *PLoS One.* 2017;12(9):e0185663.
36. Qi L, Parast L, Cai T, Powers C, Gervino EV, Hauser TH, et al. Genetic susceptibility to coronary heart disease in type 2 diabetes: 3 independent studies. *J Am Coll Cardiol.* 2011;58(25):2675-82.
37. Freilinger T, Anttila V, de Vries B, Malik R, Kallela M, Terwindt GM, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat Genet.* 2012;44(7):777-82.
38. Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, et al. Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat Genet.* 2013;45(8):912-7.
39. Debette S, Kamatani Y, Metso TM, Kloss M, Chauhan G, Engelter ST, et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet.* 2015;47(1):78-83.
40. Kiando SR, Tucker NR, Castro-Vega LJ, Katz A, D'Escamard V, Treard C, et al. PHACTR1 Is a Genetic Susceptibility Locus for Fibromuscular Dysplasia Supporting Its Complex Genetic Pattern of Inheritance. *PLoS Genet.* 2016;12(10):e1006367.
41. Fujimaki T, Oguri M, Horibe H, Kato K, Matsuoka R, Abe S, et al. Association of a transcription factor 21 gene polymorphism with hypertension. *Biomed Rep.* 2015;3(1):118-22.

42. Hager J, Kamatani Y, Cazier JB, Youhanna S, Ghassibe-Sabbagh M, Platt DE, et al. Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis. *PLoS One*. 2012;7(6):e38663.
43. O'Donnell CJ, Kavousi M, Smith AV, Kardina SL, Feitosa MF, Hwang SJ, et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation*. 2011;124(25):2855-64.
44. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*. 2012;43(12):3161-7.
45. Patel RS, Morris AA, Ahmed Y, Kavtaradze N, Sher S, Su S, et al. A genetic risk variant for myocardial infarction on chromosome 6p24 is associated with impaired central hemodynamic indexes. *Am J Hypertens*. 2012;25(7):797-803.
46. Paquette M, Dufour R, Baass A. PHACTR1 genotype predicts coronary artery disease in patients with familial hypercholesterolemia. *J Clin Lipidol*. 2018;12(4):966-71.
47. Bolton JL, Stewart MC, Wilson JF, Anderson N, Price JF. Improvement in prediction of coronary heart disease risk over conventional risk factors using SNPs identified in genome-wide association studies. *PLoS One*. 2013;8(2):e57310.
48. van Setten J, Isgum I, Smolonska J, Ripke S, de Jong PA, Oudkerk M, et al. Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction. *Atherosclerosis*. 2013;228(2):400-5.
49. Pechlivanis S, Muhleisen TW, Mohlenkamp S, Schadendorf D, Erbel R, Jockel KH, et al. Risk loci for coronary artery calcification replicated at 9p21 and 6q24 in the Heinz Nixdorf Recall Study. *BMC Med Genet*. 2013;14:23.
50. Wojczynski MK, Li M, Bielak LF, Kerr KF, Reiner AP, Wong ND, et al. Genetics of coronary artery calcification among African Americans, a meta-analysis. *BMC Med Genet*. 2013;14:75.
51. Coronary Artery Disease Genetics C. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43(4):339-44.
52. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. *Am J Hum Genet*. 2004;75(3):353-62.
53. Beaudoin M, Gupta RM, Won HH, Lo KS, Do R, Henderson CA, et al. Myocardial Infarction-Associated SNP at 6p24 Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries. *Arterioscler Thromb Vasc Biol*. 2015;35(6):1472-9.
54. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, et al. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell*. 2017;170(3):522-33 e15.

55. Zhang X, Johnson AD, Hendricks AE, Hwang SJ, Tanriverdi K, Ganesh SK, et al. Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. *Hum Mol Genet.* 2014;23(3):782-95.
56. Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. *Development.* 2007;134(23):4131-40.
57. Liu Y, Niu W, Wu Z, Su X, Chen Q, Lu L, et al. Variants in exon 11 of MEF2A gene and coronary artery disease: evidence from a case-control study, systematic review, and meta-analysis. *PLoS One.* 2012;7(2):e31406.
58. Weng L, Kavaslar N, Ustaszewska A, Doelle H, Schackwitz W, Hebert S, et al. Lack of MEF2A mutations in coronary artery disease. *J Clin Invest.* 2005;115(4):1016-20.
59. Wang L, Fan C, Topol SE, Topol EJ, Wang Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science.* 2003;302(5650):1578-81.
60. Allen PB, Greenfield AT, Svenningsson P, Haspeslagh DC, Greengard P. Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins. *Proc Natl Acad Sci U S A.* 2004;101(18):7187-92.
61. Reschen ME, Lin D, Chalisey A, Soilleux EJ, O'Callaghan CA. Genetic and environmental risk factors for atherosclerosis regulate transcription of phosphatase and actin regulating gene PHACTR1. *Atherosclerosis.* 2016;250:95-105.
62. Jarray R, Allain B, Borriello L, Biard D, Loukaci A, Larghero J, et al. Depletion of the novel protein PHACTR-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis. *Biochimie.* 2011;93(10):1668-75.
63. Allain B, Jarray R, Borriello L, Leforban B, Dufour S, Liu WQ, et al. Neuropilin-1 regulates a new VEGF-induced gene, Phactr-1, which controls tubulogenesis and modulates lamellipodial dynamics in human endothelial cells. *Cell Signal.* 2012;24(1):214-23.
64. Jarray R, Pavoni S, Borriello L, Allain B, Lopez N, Bianco S, et al. Disruption of phactr-1 pathway triggers pro-inflammatory and pro-atherogenic factors: New insights in atherosclerosis development. *Biochimie.* 2015;118:151-61.
65. Wiezlak M, Diring J, Abella J, Mouilleron S, Way M, McDonald NQ, et al. G-actin regulates the shuttling and PP1 binding of the RPEL protein Phactr1 to control actomyosin assembly. *J Cell Sci.* 2012;125(Pt 23):5860-72.
66. Mouilleron S, Wiezlak M, O'Reilly N, Treisman R, McDonald NQ. Structures of the Phactr1 RPEL domain and RPEL motif complexes with G-actin reveal the molecular basis for actin binding cooperativity. *Structure.* 2012;20(11):1960-70.
67. Mount PF, Kemp BE, Power DA. Regulation of endothelial and myocardial NO synthesis by multi-site eNOS phosphorylation. *J Mol Cell Cardiol.* 2007;42(2):271-9.

68. Lettre G. Investigation of the role of PHACTR1 in the vascular endothelium and its impact on myocardial infarction risk.
69. Dudzinski DM, Michel T. Life history of eNOS: partners and pathways. *Cardiovasc Res.* 2007;75(2):247-60.
70. Sessa WC, Garcia-Cardena G, Liu J, Keh A, Pollock JS, Bradley J, et al. The Golgi association of endothelial nitric oxide synthase is necessary for the efficient synthesis of nitric oxide. *J Biol Chem.* 1995;270(30):17641-4.
71. Pacher P, Beckman JS, Liaudet L. Nitric oxide and peroxynitrite in health and disease. *Physiol Rev.* 2007;87(1):315-424.
72. Ensembl [Internet]. 2015. Available from: http://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000112137;r=6:12716805-13290484.
73. Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet.* 2017;49(4):600-5.
74. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics.* 2016;32(20):3207-9.
75. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317-30.
76. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228-35.
77. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics.* 2015;31(16):2601-6.
78. Brodie A, Azaria JR, Ofran Y. How far from the SNP may the causative genes be? *Nucleic Acids Res.* 2016;44(13):6046-54.
79. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A.* 2017;114(9):2301-6.
80. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190-5.

81. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-60.
82. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science*. 2015;348(6235):666-9.
83. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348(6235):660-5.
84. Pazin MJ. Using the ENCODE Resource for Functional Annotation of Genetic Variants. *Cold Spring Harb Protoc*. 2015;2015(6):522-36.
85. Krause MD, Huang R-T, Wu D, Shentu T-P, Harrison DL, Whalen MB, et al. A genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *bioRxiv*. 2018.
86. Ruiz-Velasco M, Zaugg JB. Structure meets function: How chromatin organisation conveys functionality. *Current Opinion in Systems Biology*. 2017;1:129–36.
87. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell*. 2015;162(5):1051-65.
88. Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, et al. Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell*. 2015;162(5):1039-50.
89. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem*. 2016;85:227-64.
90. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-308.
91. Codina-Fauteux VA, Beaudoin M, Lalonde S, Lo KS, Lettre G. PHACTR1 splicing isoforms and eQTLs in atherosclerosis-relevant human cells. *BMC Med Genet*. 2018;19(1):97.
92. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55(4):611-22.
93. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3(7):RESEARCH0034.
94. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650-67.

95. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296-7.
96. Lalonde S, Stone OA, Lessard S, Lavertu A, Desjardins J, Beaudoin M, et al. Frameshift indels introduced by genome editing can lead to in-frame exon skipping. *PLoS One*. 2017;12(6):e0178700.
97. Aherrahrou R, Aherrahrou Z, Schunkert H, Erdmann J. Coronary artery disease associated gene *Phactr1* modulates severity of vascular calcification in vitro. *Biochem Biophys Res Commun*. 2017;491(2):396-402.
98. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. 2014;42(22):e168.
99. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3(6):1101-8.
100. Cooper JA. Effects of cytochalasin and phalloidin on actin. *J Cell Biol*. 1987;105(4):1473-8.
101. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
102. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13.
103. Knezevic II, Predescu SA, Neamu RF, Gorovoy MS, Knezevic NM, Easington C, et al. Tiam1 and Rac1 are required for platelet-activating factor-induced endothelial junctional disassembly and increase in vascular permeability. *J Biol Chem*. 2009;284(8):5381-94.
104. Patterson RL, van Rossum DB, Gill DL. Store-operated Ca²⁺ entry: evidence for a secretion-like coupling model. *Cell*. 1999;98(4):487-99.
105. dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, Berry DA, et al. Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev*. 2003;83(2):433-73.
106. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169(7):1177-86.
107. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*. 2017;123:56-65.
108. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 2015;4:1310.
109. Cairns J, Freire-Pritchett P, Wingett SW, Varnai C, Dimond A, Plagnol V, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol*. 2016;17(1):127.

110. Messina G, Atterrato MT, Prozzillo Y, Piacentini L, Losada A, Dimitri P. The human Cranio Facial Development Protein 1 (Cfdp1) gene encodes a protein required for the maintenance of higher-order chromatin organization. *Sci Rep.* 2017;7:45022.
111. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74.
112. 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;10(12):1213-8.
113. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol.* 2015;16(3):144-54.
114. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet.* 2011;12(4):283-93.
115. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet.* 2014;15(4):272-86.
116. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930-4.
117. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790-7.
118. Xu X, Qi LS. A CRISPR-dCas Toolbox for Genetic Engineering and Synthetic Biology. *J Mol Biol.* 2018.
119. La Russa MF, Qi LS. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol Cell Biol.* 2015;35(22):3800-9.
120. Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol.* 2016;17(1):5-15.
121. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell.* 2017;169(3):559.