

UNIVERSITÉ DE MONTRÉAL

**ÉTUDE GÉNÉTIQUE ET FONCTIONNELLE DE VARIANTES DE LA RÉGION
CHROMOSOMIQUE 3p21 ASSOCIÉE AUX MALADIES INFLAMMATOIRES
DE L'INTESTIN**

par

Marie-Pierre Lévesque

Département des Sciences Biomédicales

Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)
en sciences biomédicales
option générale

Avril 2011

© Marie-Pierre Lévesque, 2011.

UNIVERSITÉ DE MONTRÉAL
FACULTÉ DES ÉTUDES SUPÉRIEURES

Ce mémoire intitulé:

ÉTUDE GÉNÉTIQUE ET FONCTIONNELLE DE VARIANTES DE LA RÉGION
CHROMOSOMIQUE 3p21 ASSOCIÉE AUX MALADIES INFLAMMATOIRES DE
L'INTESTIN

présenté par : Lévesque Marie-Pierre

en vue de l'obtention du diplôme de : Maître ès Sciences

Département des Sciences Biomédicales

Faculté de Médecine

a été évalué par le jury d'examen constitué de :

AMRE Devendra, Ph.D., Président Rapporteur

RIOUX John David, Ph.D., Directeur Scientifique

KRAJINOVIC Maja, Ph.D., Membre du Jury

RÉSUMÉ

Des études de liaison et d'association génétiques ont permis d'identifier certains des facteurs de risque génétiques aux maladies inflammatoires de l'intestin (MII) dans la région chromosomique 3p21. Dans cette région, le polymorphisme nucléotidique simple (SNP) codant non-synonyme du gène *MST1*, rs3197999, encodant pour la mutation R689C, a été associé et répliqué à la fois à la colite ulcéreuse (CU) et à la maladie de Crohn (MC). Un autre SNP, corrélé à des SNP codants non-synonymes du gène *MST1R*, a également été associé à la MC. Afin de déterminer si d'autres variantes des gènes *MST1* et *MST1R* sont associés à la CU, nous avons testé pour association des SNP de ces gènes. Seul un proxy de R689C a montré un signal d'association significatif aux MII, ce qui suggère que R689C est la variante causale aux MII dans le gène *MST1*. En cherchant à déterminer si la région 3p21 contenait plusieurs signaux d'association mutuellement indépendants, trois SNP ont été identifiés comme possible facteurs de risque indépendants, et ont été génotypés dans des cas de CU et de MC et des témoins, puis nos résultats d'association ont été combinés à ceux provenant de trois autres cohortes indépendantes. Les trois SNP, R689C (*MST1*), rs6802890 et rs7629936 (*CDHR4*), sont associés aux MII, mais une étude d'association conditionnelle suggère qu'il existe en fait deux signaux d'association mutuellement indépendants dans la région 3p21. Le signal principal provient de R689C, une mutation de la protéine MSP. Cette protéine a un rôle dans l'inflammation chez les macrophages murins, et la migration, la cicatrisation et la survie chez les cellules épithéliales. Dans cette étude, le rôle de la MSP a été investigué dans des modèles de macrophages humains et de cellules épithéliales de côlon, et seule la phosphorylation d'AKT, un acteur dans la voie de signalisation de la survie cellulaire, a été modulée par la MSP dans nos modèles. Ce projet a donc permis d'apporter des connaissances sur les facteurs de risques génétiques aux MII dans la région 3p21, en identifiant 2 signaux d'association indépendants, et en nous informant sur le rôle de *MST1*, duquel provient le signal d'association principal, chez les cellules humaines.

Mots clés : 3p21, MST1, MST1R, association, méta-analyse, maladies inflammatoires de l'intestin, colites ulcéreuses, maladie de Crohn, phosphorylation d'AKT.

ABSTRACT

Linkage studies and association studies allowed the discovery of some of the genetic risk factors of inflammatory bowel disease (IBD) in the chromosomal region 3p21. In this region, the non-synonymous coding single nucleotide polymorphism (SNP) rs3197999, situated in the gene MST1 and encoding for the mutation R689C, has been associated to UC and CD multiple times, and another SNP, correlated to non-synonymous coding SNPs in the gene MST1R, has also been associated to CD. In order to verify if other variants of MST1 and MST1R are associated to UC, we tested the association of some of their SNPs. Apart from R689C, only its proxy showed a significant association signal to IBD. It suggests that R689C might be the causal variant of IBD in the region 3p21. In the aim to determine if the region 3p21 has multiple independent association signals, 3 SNPs have been identified, from the results of a published meta-analysis of UC genome-wide association studies, as being possibly independent risk factors for UC based on their correlation. Their association to IBD and their independence have been tested by genotyping them in a cohort composed of controls, and UC and CD cases. The results of the association tests have been combined, in a meta-analysis, to the results of 3 other independent association studies. The 3 SNPs, R689C (MST1), rs6802890 and rs7629936 (CDHR4) are associated to IBD, but the results of the subsequent conditional association tests suggest that there is only 2 independent association signals in the region 3p21. The main signal is arising from R689C, a mutation of the protein MSP. According to published studies, this protein has a function in the inflammation in murine macrophages, and also in the scattering, wound healing and survival of epithelial cells. In this thesis, we investigated the role of MSP in human macrophage models and in human colon epithelial cells, and it has been shown that MSP modulates the phosphorylation of AKT, an actor in the pathway of cellular survival. This project brought some knowledge about the IBD genetic risk factors in the region 3p21. We identified 2 independent association signals to IBD in this region, and the main signal is coming from a SNP in MST1, a gene which has a role, based on our results, in the survival in human colon epithelial cells.

Keywords : 3p21, MST1, MST1R, association, meta-analysis, inflammatory bowel diseases, ulcerative colitis, Crohn's disease, phosphorylation of AKT.

TABLE DES MATIÈRES

RÉSUMÉ.....	iv
ABSTRACT	vi
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	xi
LISTE DES FIGURES.....	xiii
LISTE DES ABBRÉVIATIONS	xv
REMERCIEMENTS.....	xix
CHAPITRE 1.....	1
1.1 LES MALADIES INFLAMMATOIRES DE L'INTESTIN.....	1
1.1.1 Généralités	1
1.1.2 La réponse inflammatoire.....	2
1.1.3 Les facteurs de risque environnementaux.....	5
1.1.4 L'épidémiologie génétique des Maladies Inflammatoires de l'Intestin	6
1.2 MII : LES ÉTUDES GÉNÉTIQUES.....	7
1.2.1 Les variations génétiques chez l'Homme	7
1.2.2 Les études de liaison.....	8
1.2.3 Les études d'association.....	9
1.2.4 Les voies biologiques impliquées dans la pathogénèse des MII.....	12
1.3 LA RÉGION CHROMOSOMIQUE 3p21 ASSOCIÉE AUX MII.....	14
1.3.1 Études de liaison et d'association dans la région 3p21	14
1.4 MSTI, MSTIR, R689C.....	17
1.4.1 MSP.....	17
1.4.2 RON.....	18
1.4.3 La voie de signalisation MSP/RON	19
1.5 HYPOTHÈSES ET OBJECTIFS	21
CHAPITRE 2.....	23
1.3 CULTURE CELLULAIRE.....	23
2.1.1 Lignées cellulaires et milieux	23
2.1.2 Purification des monocytes sanguins et culture pour différenciation en macrophages	24
2.2 TECHNIQUES DE BASE EN BIOLOGIE MOLÉCULAIRE.....	25
2.2.1 Extraction des protéines	25
2.2.2 Immunobuvardage de type Western.....	26
2.2.3 Réaction de polymérisation en chaîne quantitative.....	27
2.3 ÉVALUATION DE MODÈLES FONCTIONNELS	28
2.3.1 Expression de MSTIR dans les lignées cellulaires	28
2.3.2 Expression des gènes pro-inflammatoires	29

2.3.3	<i>Cicatrisation</i>	29
2.3.4	<i>Migration des cellules</i>	30
2.3.5	<i>Phosphorylation d'AKT</i>	31
2.4	EXPRESSION DE LA MSP RECOMBINANTE	32
2.4.1	<i>Production et validation du plasmide d'expression de MST1</i>	32
2.4.2	<i>Expression et purification de la MST1 recombinante</i>	34
2.5	GÉNOTYPAGE	36
2.5.1	<i>Génotypage sur la plateforme SEQUENOM</i>	36
2.5.1.1	<i>Conception des groupes d'amorces</i>	36
2.5.1.2	<i>Dilution des amorces</i>	37
2.5.1.3	<i>Amplification par PCR</i>	38
2.5.1.5	<i>Détermination des génotypes</i>	39
2.5.2	<i>Amplification spécifique de 3p21</i>	42
2.5.2.1	<i>Alignement de la région chromosomale 3p21</i>	42
2.5.2.2	<i>Conception des amorces spécifiques à la séquence de MST1</i>	42
2.5.2.3	<i>Optimisation des réactions de PCR</i>	43
2.5.3	<i>Échantillons d'ADN</i>	46
2.6	ANALYSES ET STATISTIQUES	47
2.6.1	<i>Essais Fonctionnels</i>	47
2.6.2	<i>Étude d'association des variantes de MST1 et MST1R</i>	48
2.6.3	<i>Méta-analyse des signaux de 3p21</i>	49
	CHAPITRE 3	51
3.1	ÉTUDES D'ASSOCIATION	51
3.1.1	<i>Recherche de nouvelles variantes codantes non-synonymes de MST1 et MST1R associées aux MII</i> 51	
3.1.1.1	<i>Étude de la structure génomique de 3p21</i>	52
3.1.1.1.1	<i>Alignement de la région 3p21</i>	52
3.1.1.1.2	<i>Amplification spécifique de la séquence de MST1</i>	53
3.1.1.2	<i>Test d'association des variantes de MST1 et MST1R</i>	57
3.1.2	<i>Test d'indépendance des signaux d'association de la région 3p21</i>	58
3.1.3	<i>Test d'association des trois SNP étant potentiellement des signaux d'association indépendants à la MII</i>	59
3.1.4	<i>Test d'association conditionnelle des trois SNP étant potentiellement des signaux d'association indépendants à la MII</i>	60
3.2	ÉTUDES FONCTIONNELLES	63
3.2.1	<i>Inhibition de l'expression de gènes pro-inflammatoires chez les macrophages prétraités à la MSP</i>	64
3.2.2	<i>Expression de MST1R dans différentes lignées cellulaires</i>	68
3.2.3	<i>Migration des cellules épithéliales des intestins traitées à la MSP</i>	69
3.2.4	<i>Induction par MSP de la phosphorylation de AKT chez nos modèles de cellules épithéliales du côlon</i>	72
3.3	PRODUCTION DE LA MSP RECOMBINANTE À PARTIR DE LA TRANSFECTION DE CELLULES DE MAMMIFÈRES	75
	CHAPITRE 4	79
4.1	<i>Deux signaux d'association indépendants sont impliqués dans la susceptibilité aux MII dans la région 3p21</i>	79
4.1.1	<i>Test d'association de variantes de MST1 et MST1R</i>	79

4.1.2	Test d'association conditionnel dans la région 3p21	82
4.2	<i>MSP, encodée par MST1, semble avoir un rôle régulateur de la phosphorylation de AKT, mais pas dans la modulation de la réponse inflammatoire</i>	85
4.2.1	Tester l'effet de MSP sur un modèle de réponse inflammatoire chez les macrophages	85
4.2.2	Tester l'effet de MSP sur des modèles de cellules épithéliales	87
4.3	<i>Limites de l'étude</i>	90
CONCLUSION		92
RÉFÉRENCES		94
ANNEXE A		101
ANNEXE B		102
ANNEXE C		131
ANNEXE D		132
ANNEXE E		140
ANNEXE F		141

LISTE DES TABLEAUX

Tableau I : Liste des études ayant publié un signal d'association dans la région 3p21.....	14
Tableau II: Recette de réactifs pour la fabrication des gels d'acrylamide 8%.....	26
Tableau III: Description et détails des manipulations pour chacun des anticorps utilisés dans les immunobuvardages.....	27
Tableau IV: Cycles de PCR quantitatif.....	28
Tableau V: Composition des réactions de digestion du plasmide HS-MST1-IM-1 par des enzymes de restriction.....	34
Tableau VI: Composition de la réaction de PCR (A) et des cycles de PCR (B) pour le génotypage.....	38
Tableau VII: Composition de la réaction de SAP.....	39
Tableau VIII: Composition de la réaction d'élongation.....	39
Tableau IX: Séquence et température de dissociation de chaque amorce spécifique au chromosome 3.....	43
Tableau X: Composition de chaque réaction d'amplification (A) spécifique de la séquence du chromosome 3, et cycles de PCR (B).....	44
Tableau XI: Les 14 variantes choisies pour tester les amorces spécifiques au chromosome 3 sur la plateforme de génotypage.....	45
Tableau XII: Cycle de PCR pour le génotypage avec les amorces spécifiques au chromosome 3.....	46

Tableau XIII: Autres variantes de MST1 et MST1R géotypés dans nos cohortes.	48
Tableau XIV: Résultats d'association de variantes de MST1 et MST1R dans notre cohorte italienne.....	58
Tableau XV: Résultats d'association des trois signaux indépendants dans la région 3p21 à partir des données des méta-analyses précédentes.....	59
Tableau XVI: Résultats de la méta-analyse des tests d'association par régression logistique et des analyses d'association conditionnelles pour MII, CU et MC.	61

LISTE DES FIGURES

Figure 1: Structure des tissus intestinaux.....	3
Figure 2: Voie de signalisation sommaire suite à la phosphorylation de RON après la stimulation au MSP	20
Figure 3: Détails du plasmide HS-MST1-IM-1	32
Figure 4: Détermination des génotypes sur la plateforme Sequenom	41
Figure 5: Plan des réactions de PCR de l'expérience d'optimisation de l'amplification spécifique du chromosome 3.....	44
Figure 6: Plan des réactions de PCR de l'expérience d'optimisation de l'amplification spécifique du chromosome 3.....	53
Figure 7: Électrophorèse de produits d'amplification de génomes humains ou hybrides rongeur-humain chr1 ou chr3.....	55
Figure 8: Régions en fort LD avec les trois signaux d'association de la région 3p21.....	63
Figure 9: Expression relative de marqueurs de différenciation chez les monocytes du sang et les macrophages qui en sont dérivés.....	65
Figure 10: Expression relative de TNF- α et IL12p40 chez les macrophages prétraités au MSP et stimulés au LPS	67
Figure 11: Expression relative de <i>MST1R</i> chez différentes lignées cellulaires.	68
Figure 12: Graphique à boîtes et moustaches présentant les résultats de cicatrisation des blessures de tissus épithéliaux intestinaux après différents traitements.....	71
Figure 13: Expression de AKT et Phospho-AKT chez des cellules HT-29, HCT-15 et HCT-116 traitées au FBS, HGF et MSP.....	74

Figure 14: Digestion du plasmide MST1	76
Figure 15: Immunobuvardage montrant les étapes de purification de la MSP recombinante.....	76
Figure 16 : Région définie par les SNP corrélés à nos signaux d'association indépendants ($r^2 \geq 0.5$ et 0.8).	84

LISTE DES ABBRÉVIATIONS

°C	degré Celsius
A	adénine
ADN	acide déoxyribonucléique
ANOVA	analyse de la variance
APC	cellule présentatrice d'antigène (<i>antigen presenting cell</i>)
Arg ou R	arginine
ARN	acide ribonucléique
ARNm	acide ribonucléique messenger
ATCC	<i>American Type Cell Collection</i>
BLASTN	<i>Nucleotide Basic Local Alignment Search Tool</i>
BSA	albumine de sérum bovin (<i>bovine serum albumin</i>)
C	cytosine
CEPH	Centre d'Étude du Polymorphisme Humain
CEU	résidents de l'Utah aux États-Unis d'origine européenne (CEPH)
ch	chacun
Chr	chromosome
CI	intervalle de confiance
cm	centimètre
CNV	nombre de copies variable (<i>copy number variation</i>)
ctl	contrôle
CU	colite ulcéreuse
Cys ou C	cystéine
DMEM	<i>Dulbecco's Modified Eagle medium</i>
dNTP	déoxyribonucléotide triphosphate (où N=A,C,G ou T)
EC ₅₀	concentration pour obtenir la moitié de l'effet maximal
ECL	chimiluminescence enchérie
EDTA	acide éthylènediaminetétraacétique
eQTL	étude d'association d'expression génétique
FBS	sérum fœtal bovin
Fig.	figure

g	accélération due à la gravité à la surface de la Terre
g	gramme
G	guanine
GWAS	étude d'association pangénomique (<i>genome-wide association study</i>)
h	heure
H ₂ O	eau
HEPES	acide 4-(2-hydroxyéthyle)-1-pipérazine-éthane-sulfonique
HGF	<i>hepatocyte growth factor</i>
His	histidine
HRP	peroxydase de raifort
IL	interleukine
iNOS	oxyde nitrique synthétase inductible
kb	kilo bases
kD	kilo Dalton
LB	<i>Lysogeny Broth</i>
LD	déséquilibre de liaison
LPS	lipopolysaccharide
M	molaire
mA	milliampère
MAF	fréquence de l'allèle mineur
Mb	mega bases
MC	maladie de Crohn
mg	milligramme
MgCl ₂	chlorure de magnésium
MII	maladies inflammatoires de l'intestin
min	minute
ml	millilitre
mM	millimolaire
MSP	<i>macrophage stimulating protein</i>
N ou Nbr.	nombre
Na ₃ VO ₄	orthovanadate de sodium
NaCl	chlorure de sodium
NaF	fluorure de sodium
NaH ₂ PO ₄	phosphate de sodium monobasique
NaOAc	acétate de sodium
NCBI	<i>National Center for Biotechnology Information</i>
ng	nanogramme
NIDDK	<i>National Institute of Diabetes and Digestive and Kidney diseases</i>
nM	nano molaire
nm	nanomètre
NO	oxyde nitrique
OMIM	<i>Online Mendelian Inheritance in Man</i>
OR	rapport de côte
P	phosphore

pb	paire de bases
PBS	solution saline-Phosphate
PCR	polymérisation en chaîne
pH	mesure de l'acidité
PMA	tétradécanoyl phorbol acétate
PMSF	Fluorure-methanosulfonylique-de-phenyl
PRR	récepteurs de reconnaissance de patron
Q1	premier quartile
Q3	troisième quartile
QC	contrôle de qualité
qPCR	PCR quantitatif
r^2	corrélation
RON	récepteur d'origine nantaise
rpm	révolution par minute
RPMI	<i>Roswell Park Memorial Institute medium</i>
SAP	phosphatase alcaline de crevette
SD	déviatiion standard
SDS	laurylsulfate de sodium
sec	seconde
SNP	polymorphisme nucléotidique simple
SOC	<i>super optimal broth/catabolite repression</i>
T	thymine
Tab.	tableau
TBS	solution saline-TRIS
TEMED	tétraméthyléthylènediamine
Th	lymphocytes T auxiliaire
Tm	température de dissociation
TRIS	trishydroxyméthylaminométhane
U	unité
ug	microgramme
ul	microlitre
uM	micro molaire
UTR	région non transcrite (<i>untranslated region</i>)
UV	ultra violet
V	Volt
WT	type sauvage
WTCCC	<i>Wellcome Trust Case Control Consortium</i>
x	fois

« Ce n'est pas parce qu'un problème n'a pas
été résolu qu'il est impossible à résoudre. »

Agatha Christie

REMERCIEMENTS

Je souhaite en tout premier lieu remercier le Dr. Rioux, mon directeur de recherche, pour m'avoir donné la chance d'évoluer dans son laboratoire et pour m'avoir transmis avec bienveillance toutes ces connaissances qui me seront indispensables dans ma carrière.

Je remercie également Philippe, mon superviseur, qui m'a aidé avec patience à diriger mes nombreuses expériences et que je ne remercierai jamais assez pour toutes les heures qu'il a prises de son temps pour corriger mes présentations, mes documents et mon mémoire.

Merci à vous aussi, Sylvain, Guy et Guillaume, pour vos précieux conseils sans lesquels mes expériences, mes analyses et mes présentations n'auraient tout simplement jamais abouti. Merci également à mes secrétaires préférées, Diana et Marie-Claude, pour l'organisation menée de main de maître de nos horaires chargés!

Un merci tout spécial à tous mes amis du labo : Catherine, Claudine, Geneviève D., Azadeh, Caroline, Gabrielle, Claire, Geneviève G., Mélissa, Ken, Fred, Amidou et sans oublier Christian...ouf! Je n'ai pas de mots pour dire à quel point vous m'avez apporté beaucoup dans ma vie. Ma maîtrise n'aurait vraiment pas été du gâteau sans vos conseils, vos rires et vos folies! Bref, je vous adore, vous allez terriblement me manquer, et j'aurai toujours une petite pensée pour vous lorsque je ferai des gaffes ou que je chanterai dans mon labo.

Je remercie aussi mes meilleures amies, Mélanie et Marie-Josée, pour les encouragements, leurs épaules réconfortantes, et tout simplement pour exister et être mes amies, même si je n'ai pas eu beaucoup de temps à leur consacrer ces dernières années...je vous adore tellement, les filles!

Je veux remercier mes parents, pour leur amour inconditionnel et pour m'avoir donné des valeurs humaines qui font, je crois, de moi une personne bien et heureuse. Je ne pourrai jamais vous remettre autant que ce que vous m'avez si généreusement donné, mais je vous promets de faire de mon mieux pour que vous soyez toujours fiers de moi.

Enfin, je ne peux pas clore mes remerciements sans louer mon conjoint, l'homme de ma vie, qui m'a supportée avec patience, qui a su me ramener à la réalité et éloigner mesangoisses par sa seule présence et son amour dévoué. Je t'aime tellement!

À tous, merci d'avoir été dans ma vie.

CHAPITRE 1

1. INTRODUCTION

1.1 LES MALADIES INFLAMMATOIRES DE L'INTESTIN

1.1.1 Généralités

Les maladies inflammatoires de l'intestin (MII) (OMIM #266600) rassemblent deux maladies chroniques et récurrentes se présentant sous forme d'inflammation du tube digestif; la maladie de Crohn (MC) et la colite ulcéreuse (CU). Elles affectent ensemble une personne sur 250 dans les pays industrialisés [1]. Les premiers symptômes, soit des crampes abdominales chroniques accompagnées d'une diarrhée importante et de saignements intestinaux, se développent surtout chez les jeunes adultes. Ces symptômes causent aussi une déficience de l'absorption des nutriments au niveau des intestins et mènent éventuellement à une perte de poids, de l'anémie et des carences alimentaires qui hypothèqueront considérablement la santé des patients [2]. De plus, 10 à 15% des patients requerront une ablation partielle des intestins à cause d'une inflammation trop persistante. Enfin, ces maladies peuvent également affecter les patients de manière indirecte puisque ceux-ci sont plus sujets à une détresse psychologique, telle que la dépression et l'anxiété, et plus à risque de développer un cancer du côlon. Les MII ont donc des répercussions importantes sur la vie des patients et leurs proches [3].

Bien qu'elles partagent de nombreux symptômes, la MC et la CU ont quelques particularités qui leur sont propre. La MC présente une inflammation transmurale, avec des ulcérations qui peuvent créer éventuellement des fistules. Ces dernières sont des ouvertures entre le lumen de l'intestin et d'autres organes, le péritoine ou l'extérieur du corps. Le développement de ces fistules peut mener à des infections péritonéales, de même que toucher des organes et affecter leur fonctionnement. De plus, la muqueuse intestinale est envahie de granulomes, des amas de macrophages et de cellules

inflammatoires. L'inflammation est discontinue, pouvant toucher tout le tube digestif, de la bouche à l'anus, mais affecte surtout l'intestin grêle, et plus particulièrement l'iléon [2, 3]. Pour la CU, l'inflammation est superficielle et se limite à la muqueuse intestinale. Cette dernière est couverte de petites ulcérations et est très fragile dû aux infiltrations de neutrophiles dans les tissus. De plus, l'inflammation progresse généralement de façon continue du rectum vers le côlon descendant. De nombreuses autres affections peuvent également se présenter chez les patients souffrants de la MC ou de la CU, comme l'arthrite, la cholangite sclérosante, et l'uvéite [3]. Malheureusement, bien que les traitements soient efficaces pour soulager les symptômes de ces maladies inflammatoires, ils ne les guérissent pas, les symptômes revenant souvent entre des phases de rémission. Les MII sont des maladies complexes et à ce jour, beaucoup d'efforts sont déployés en recherche afin de mieux identifier les facteurs génétiques et environnementaux qui mènent au développement de ces maladies complexes que sont les MII.

1.1.2 La réponse inflammatoire

Le système digestif est un organe complexe indispensable dans l'absorption des nutriments nécessaires au bon fonctionnement du corps. La figure 1 montre la structure des tissus intestinaux. Le tissu intestinal se compose de trois grandes couches : la séreuse qui est en contact avec l'intérieur du péritoine, la sous-muqueuse qui contient de nombreux muscles lisses, et la muqueuse, qui est en contact avec la lumière de l'intestin. Comme c'est la muqueuse qui est en contact avec les aliments et les micro-organismes, c'est également dans cette couche que l'on retrouve le plus de cellules ayant une fonction immunitaire ou un rôle dans l'absorption. Les entérocytes absorbent les nutriments, les ions et l'eau ingérée, et également sécrètent les immunoglobulines A qui leurs ont été transférées par les lymphocytes B des tissus lymphoïdes associés à l'intestin. Les cellules de Paneth sécrètent des lysozymes pour la défense de l'hôte. Les cellules de Goblet produisent des mucines, qui seront dissoutes dans l'eau sous forme de mucus intestinal. Ce mucus sert à limiter l'introduction des microorganismes dans la paroi intestinale, et aussi à lubrifier cette paroi. Le mucus contient

également des immunoglobulines et des protéines sécrétées par les entérocytes et les cellules de Paneth. Enfin, le tissu de la muqueuse est également infiltré de macrophages résidents ainsi que de neutrophiles qui ont pour rôle de limiter l'intrusion des microorganismes qui ont traversé la barrière épithéliale.

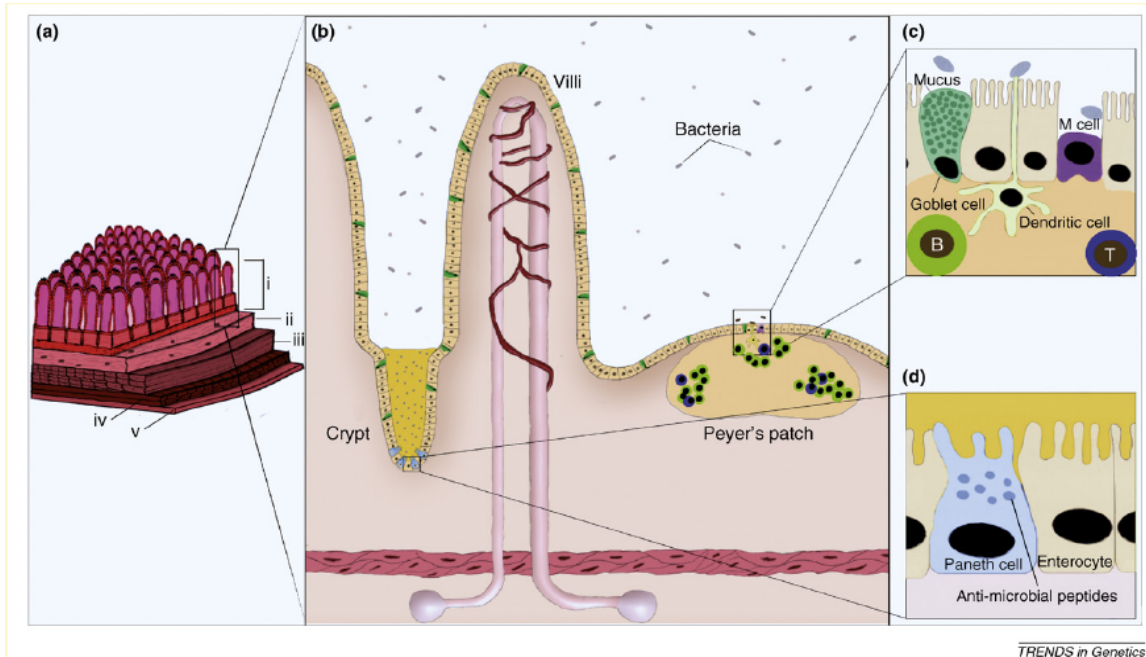


Figure 1: Structure des tissus intestinaux.

A) schéma des villi de la muqueuse (i) exposés au lumen de l'intestin, entouré de la sous-muqueuse (ii), de deux couches de muscles lisses (iii-iv) et de la séreuse (v). B) Détails de l'épithélium des tissus intestinaux. C) Schéma des cellules retrouvées à la surface des plaques de Peyer. D) Schéma des cellules de Paneth retrouvées dans les cryptes entre les villi. Modifié à partir de Budarf et al., 2009 [4].

Outre les cellules des tissus du système digestif, ce dernier comprend également un très grand nombre de microorganismes commensaux, qui contribuent également à fournir à notre corps des nutriments, à digérer certaines molécules complexes et à produire les vitamines que l'on ne peut pas produire [5]. Ces alliés, introduits dans notre organisme progressivement très tôt dans l'enfance, sont tolérés par notre système immunitaire lors

du développement de celui-ci. Leur présence dans la lumière de l'intestin ne déclenche en général aucune réaction inflammatoire.

L'intrusion de ces micro-organismes est limitée par des barrières physiques telles que le tissu épithélial, dont les cellules sont jointes entre elles par des jonctions serrées, et par le mucus. Cependant, il arrive parfois qu'un de ces micro-organismes commensaux, ou encore un pathogène, franchisse cette barrière. Une réponse immunitaire innée est immédiatement mise en branle afin d'éliminer cet organisme étranger. Les antigènes sont présentés par des cellules dendritiques aux macrophages résidents dans les tissus intestinaux et aux neutrophiles. De plus, les cellules M absorbent les antigènes dans le lumen et les transfèrent par transcytose aux lymphocytes T sur leur côté basolatéral. Les cellules immunitaires, qui reconnaissent les antigènes étrangers via des récepteurs spécialisés (pattern-recognition receptors, PRR) vont tenter d'éliminer les micro-organismes par phagocytose et par la libération d'agents oxydants et antimicrobiens. De plus, elles libèrent également des cytokines pro-inflammatoires telles que TNF- α , qui recrute au site d'inflammation des neutrophiles et des macrophages circulants, et IL-10, qui aide à la différenciation des lymphocytes T naïfs en lymphocytes régulateurs. Ces lymphocytes régulateurs inhibent la fonction pro-inflammatoire des lymphocytes effecteurs, et éventuellement permettent à l'inflammation de s'éteindre [4].

Chez les patients souffrant d'une MII, l'inflammation ne réussit pas à s'estomper et devient chronique, causant ainsi une production accrue de cytokines pro-inflammatoires, comme IL-23, qui différencie les lymphocytes T naïfs en lymphocytes de type Th1 et Th2. Ces lymphocytes pro-inflammatoires recrutent respectivement des lymphocytes T effecteur et des lymphocytes B différenciés. Éventuellement, la production de cytokines comme IL-6 et IL-12 par les cellules dendritiques et les macrophages mènent à la transformation des cellules T régulatrices en cellules pro-inflammatoire ayant un phénotype Th17 [6]. Les cytokines produites par les Th17 recruteront d'autres neutrophiles dans la muqueuse inflammée. Traditionnellement, on retrouve une réponse

de type Th1 et Th17 en général chez les patients atteints de MII, et les patients atteints de colite ulcéreuse expriment des cytokines typiques d'une réponse Th2 [7, 8]. En effet, on observe que la réponse inflammatoire chez les patients souffrant de MII mène à des infiltrations progressives de neutrophiles et de lymphocytes T et B effecteurs, qui détruisent peu à peu les tissus.

1.1.3 Les facteurs de risque environnementaux

On en sait encore très peu sur l'étiologie des MII, mais quelques facteurs environnementaux ont été mis en question dans les premières études épidémiologiques menées sur ces maladies. Entre autre, la hausse du nombre de cas dans les pays industrialisés lors de la deuxième moitié du 20^e siècle, et en particulier dans les milieux aisés des villes, là où l'eau potable est disponible [9], a soulevé l'hypothèse qu'une exposition en bas âge à un environnement trop hygiénique ne permettait pas à notre organisme de développer une tolérance aux micro-organismes qui se retrouveraient naturellement dans notre corps [10]. Ceci a été suggéré par la corrélation entre une diminution des lymphocytes T matures dans les plaques de Peyers chez les animaux élevés dans un environnement sans germes, et une déficience de la tolérance orale [11]. Cette hypothèse est entérinée par les tests microbiologiques, effectués sur les patients, qui n'ont pas permis de détecter chez eux de pathogène en particulier qui causerait la maladie. Des études ont également suggéré que le fait de ne pas avoir été allaité au sein, de consommer la pilule contraceptive, d'avoir des changements de diète ou d'environnement contribuait à augmenter les risques de développer une MII, mais aucune de ces études n'a été répliquée de façon significative [12]. Des études ont cependant démontré que la consommation de cigarettes augmentait significativement le risque de développer la maladie de Crohn, mais protégerait contre la colite ulcéreuse. Il a aussi été démontré que l'appendicectomie aurait un effet protecteur significatif contre le développement de la colite ulcéreuse [9].

1.1.4 L'épidémiologie génétique des Maladies Inflammatoires de l'Intestin

Plusieurs études épidémiologiques ont suggéré que le risque de développer des MII comporte une forte composante génétique. En fait, on s'attend à ce qu'une maladie complexe ayant des facteurs de risque génétiques soit retrouvée plus fréquemment, comparativement au reste de la population, chez différents membres d'une même famille puisque ces derniers partagent leur matériel génétique. On constate que 5 à 10% des patients atteints de MII ont un membre de leur famille au premier degré atteint de l'une ou l'autre des maladies inflammatoires de l'intestin. Comparé au reste de la population, le risque calculé pour un parent de développer la même MII qu'un patient est de 10-20 fois pour la CU et de 30-40 fois pour la MC[13]. Des études de jumeaux ont aussi contribué à déterminer la part du risque qui est génétique. Pour la MC, 50 à 58% des paires de jumeaux monozygotiques développeront tous les deux la maladie, contre 0 à 12 % chez les paires de jumeaux dizygotiques. Pour la CU, c'est de 6 à 14% des paires de jumeaux monozygotiques contre 0 à 5% chez les paires de jumeaux dizygotiques [13]. Ces études épidémiologiques démontrent que la part du risque qui est due à la génétique est plus importante pour MC que pour CU, mais également que ces maladies sont effectivement apparentées, et qu'elles ont probablement des facteurs de risque génétiques en commun.

Finalement, il a aussi été observé que des gens issus de certaines origines ethniques sont plus susceptibles que d'autres de développer une maladie. Pour les maladies inflammatoires de l'intestin, nous savons que ces maladies, bien qu'ayant une grande incidence dans les pays industrialisés, touchent plus les juifs ashkenazi que toute autre ethnie dans ces pays. De même, on observe une plus haute incidence chez les gens originaire du centre de l'Europe [12].

1.2 MII : LES ÉTUDES GÉNÉTIQUES

1.2.1 *Les variations génétiques chez l'Homme*

D'après les résultats du projet du génome humain, notre génome compterait environ 2.85 milliard de nucléotides [14]. Il a été suggéré qu'en moyenne, la séquence génomique entre deux individus diffère de 0,1%, ce qui représente environ une variation aux 1000 bases. Cette variation génétique est due soit à des insertions ou délétions de séquences nucléotidiques de longueur ou répétition différentes (régions au nombre de copies variable (CNV)), à des petits éléments répétitifs dont le nombre de répétitions varie d'une personne à l'autre (microsatellites), ou à des changements ponctuels de nucléotides, communément appelés *single nucleotide polymorphisms* (SNP). Il est estimé qu'environ 90% des variations chez l'homme sont des SNP communs [15]. Les SNP sont beaucoup plus utilisés dans les études génétiques récentes car le génotypage de ces variables bialléliques est moins coûteux, et les tests d'association sont plus faciles à analyser que pour les variantes polyalléliques comme les microsatellites et les CNV.

La transmission des allèles des variantes génétiques dans le génome suit les lois mendéliennes qui stipulent que les allèles sont transmis des parents aux descendants de façon aléatoire. Cependant, des allèles de variantes génétiques adjacentes sont souvent transmis conjointement étant liés physiquement sur une même molécule d'ADN. On dit que les allèles de deux variantes génétiques sont en liaison lorsqu'on observe une déviation de l'indépendance entre deux allèles transmis des parents aux enfants. La liaison donne lieu au déséquilibre de liaison (LD). On parle de déséquilibre de liaison lorsqu'on observe, dans une population, une sur- ou sous-représentation de combinaison d'allèles par rapport à ce qui est attendu par le hasard. La mesure du degré de LD entre deux variantes s'évalue de plusieurs façons, tel le coefficient de corrélation (r^2), qui mesure la corrélation entre les allèles de deux variantes. Le LD est un concept très important dans les études d'association d'une variante à une maladie puisque la variante

associée peut elle-même être en LD avec la variante causale de la maladie. Le LD entre les différentes variantes permet de tester indirectement l'association de plusieurs variantes liées entre elles en n'en génotypant qu'une seule [16].

La nature d'une variante dans le génome a aussi son importance. Des variantes codantes non-synonymes, c'est-à-dire induisant un changement d'acides aminés, ont été identifiées comme la cause de nombreuses maladies mendéliennes rares, de même que comme facteur de susceptibilité dans plusieurs maladies complexes [17]. Puisque qu'elles modifient la séquence de la protéine, elles peuvent potentiellement avoir un impact considérable sur la fonction de la protéine, en modifiant la structure tridimensionnelle, les sites d'attachement du ligand ou du récepteur, les fonctions catalytiques, et bien d'autres. Pour cette raison, ce type de variantes est très recherché dans une région associée à une maladie. Il a aussi été démontré que les variantes codantes synonymes ainsi que les variantes introniques peuvent avoir un effet sur la stabilité des ARNm et sur leur épissage [18]. Les variantes situées dans le promoteur d'un gène, dans les séquences non transcrites en amont ou en aval d'un gène, ou dans des régions reconnues par des facteurs de transcription peuvent également avoir un effet sur les niveaux d'expression d'un gène, et ultimement sur la production de la protéine. Enfin, beaucoup de pathologies ne se développent que par l'interaction de plusieurs variantes causales de même qu'avec des composantes environnementales [19, 20]. Ainsi, la composition des variantes présentes chez un individu contribue à augmenter ou à réduire sa susceptibilité à développer une maladie ou un phénotype en particulier, comparativement à un autre individu.

1.2.2 Les études de liaison

Les premiers efforts pour identifier les composants génétiques des maladies mendéliennes ont donné naissance à la cartographie de liaison (*linkage mapping*). Ce type d'analyses convient surtout aux maladies simples ayant des facteurs de risque génétiques rares qui ont un effet très fort qui augmente considérablement les chances de développer la maladie. Ce type d'analyse convient particulièrement pour la cartographie

de variantes familiales. La cartographie de liaison consiste à étudier la transmission conjointe de régions génomiques et d'une maladie dans une ou des familles en étudiant les événements de recombinaison. Dans ce type d'études, plus il y a de générations dans une famille, plus il y aura d'événements de recombinaison, et plus il sera facile de circonscrire la région liée à un phénotype. Depuis 1996, les premières cartographies de liaison appliquées à la maladie de Crohn et la colite ulcéreuse ont permis l'identification, d'après la base de données de gènes et de phénotype OMIM (*Online Mendelian Inheritance in Man*), d'au moins 16 loci de susceptibilité aux MII [21-28]. Les liaisons les plus répliquées sont pour les régions chromosomiques 3p, 3q, 5q31, 6p21, 14q, 16cen et 19p pour MC. Bien que la part génétique de la colite ulcéreuse soit moins importante que pour la MC, les loci 3p, 6p21 et 19p sont aussi liés à cette maladie. Malheureusement, certaines de ces régions sont longues de quelques dizaines de méga bases et contiennent jusqu'à des centaines de gènes. Il est donc difficile de déterminer quel sont les gènes conférant le risque aux MII dans ces régions.

1.2.3 Les études d'association

Les études d'association génétique tentent d'établir une corrélation entre les allèles d'une variante génétique, comme un SNP, et le risque de développer une maladie ayant une composante génétique. Dans ce but, la fréquence allélique de variantes génétiques ou la fréquence des génotypes est comparée entre les cas et les témoins provenant d'une même population. Ce type d'étude permet une cartographie plus fine des régions grâce au concept de LD, mais est sensible à la taille et à la stratification de la population. Plusieurs approches peuvent être utilisées pour ce type d'études, soit par gènes candidats, par régions candidates ou par une approche pangénomique.

Ainsi, des études d'association ont été utilisées afin de raffiner les régions de risque de la MC. Les régions identifiées par cartographie de liaison ont été utilisées comme loci de susceptibilité aux MII par ces études d'association dans des cohortes de quelques centaines d'individus. Par ce type de cartographie fine, le locus *IBD5* (5q31) a pu être

raffiné à une région de 250 kb, et les variantes R701W, G908R et 3020c du gène *CARD15* (16cen) ont pu être clairement identifiées comme facteurs de susceptibilité aux MII [29-31].

D'autres études d'association ont aussi été faites sur des gènes candidats dont la sélection était basée sur des prémisses, telles leur localisation dans des loci liés aux MII, ou leur fonction et leur profil d'expression démontrant un rôle potentiel du gène candidat dans des mécanismes pathophysiologiques des MII. Finalement, la sélection d'autres gènes candidats était aussi basée sur des phénotypes similaires aux MII causés par leur mutation dans des modèles animaux. Toutefois, les connaissances sur la fonction de nombreux gènes et sur la pathogenèse des maladies sont encore trop peu pour que l'approche par gène candidat soit efficace pour trouver la majorité des facteurs de risques génétiques d'une maladie.

Le développement de plateformes de génotypage à haut débit (Sequenom, Illumina, Affymetrix) ainsi que de meilleures connaissances des variantes et des haplotypes publiées par HapMap ont permis de tester l'association d'un grand nombre de SNP répartis dans tout le génome, donnant naissance aux études pangénomiques. Les SNP utilisés dans ces études pangénomiques ont été sélectionnés afin de capturer une part importante de la variabilité génétique commune retrouvée dans les populations humaines. Ces études non basées sur des prémisses fonctionnelles et indépendantes des modèles acceptés dans les MII ont ouvert de nouvelles portes pour identifier de nouveaux gènes candidats et processus pathophysiologiques. Ces études ont grandement contribué au développement des connaissances sur la génétique de la maladie de Crohn.

Cependant, malgré de nombreuses études testant jusqu'à des milliers d'individus, les dizaines de loci validés par ces études d'association par gène candidat ou pangénomiques n'expliquent qu'une fraction du risque génétique observé des MII. Cela est peut-être dû au fait que les MII sont des maladies complexes qui se déclarent suite à la présence de

nombreux facteurs de risque génétiques ayant chacun un petit effet individuel, et que les études pangénomiques manquaient jusqu'ici de puissance pour détecter ces derniers. Dans le but de détecter les variantes ayant un faible effet, des consortiums internationaux se sont formés pour augmenter la taille des cohortes afin d'augmenter la puissance des études.

En effet, depuis 2008, quatre méta-analyses combinant les résultats de plusieurs études d'association pangénomiques ont été publiées sur les MII, augmentant à 71 le nombre de loci confirmés pour MC, et à 49 pour CU [32-35]. Jusqu'à maintenant, l'héritabilité expliquée serait d'environ 16% pour CU et 23.2% pour MC [32, 34]. Ce taux, encore faible, pourrait être expliqué par le fait que pour la plupart de ces loci, les variantes causales ayant un effet direct sur le phénotype n'ont pas encore été identifiées. Heureusement, pour certains loci, des variantes codantes non-synonymes ont été identifiées clairement comme variantes causales par des essais fonctionnels; par exemple, un variant du gène d'autophagie *IRGM* qui est reconnu par un microARN [36]. Mais pour d'autres loci, qui ne contiennent qu'un gène, aucune variante codante non-synonyme n'a encore été identifiée. Certains loci ne contiennent aucun gène, alors que d'autres en contiennent des centaines. Les prochaines étapes importantes consisteront donc à identifier des gènes candidats dans ces loci, ainsi que des variantes candidates susceptibles d'être causales.

Pour ce faire, diverses techniques sont envisagées. Des études d'association d'expressions génétiques avec des marqueurs (eQTL) permettent d'identifier des variantes associées à l'expression d'un gène qui aurait un rôle dans le développement des MII. Aussi, le re-séquençage des loci associés pourrait permettre de trouver des variantes additionnelles autant communes que rares associées aux MII dans une population, et ainsi d'expliquer une partie de l'héritabilité manquante. Enfin, la plupart des études d'association pangénomiques rapportent le signal d'association le plus fort dans un locus, alors que certains contiennent plus d'une variante causale, par exemple *CARD15* et ses trois variantes causales étant toutes des facteurs de risque indépendants à la MC [37].

Une analyse du LD et des haplotypes dans ces régions et une analyse d'association conditionnelle par régression logistique de différents SNP peut permettre de confirmer l'indépendance de signaux situés dans un même locus. Il est également indispensable que plus d'essais fonctionnels viennent corroborer les résultats des études d'association afin de confirmer le rôle d'un gène, ainsi que l'effet des variantes candidates sur la fonction de ce gène.

1.2.4 Les voies biologiques impliquées dans la pathogénèse des MII

Des essais fonctionnels ont été menés pour tester le rôle des gènes associés jusqu'ici aux MII. Ainsi, certains mécanismes menant au développement des maladies inflammatoires de l'intestin ont été élucidés.

Entre autre, les gènes de l'autophagie *ATG16L1* et *IRGM* ont été associés aux MII. L'autophagie est une fonction des cellules permettant l'élimination de déchets ou de micro-organismes qui sont dans le cytoplasme des cellules. Il a été démontré que les cellules épithéliales intestinales et les cellules présentatrices d'antigènes (APC) des patients porteurs de l'allèle de risque de *ATG16L1* produisaient moins d'agents antimicrobiens [38], et qu'un *knockdown* de *IRGM* résulte en une diminution de l'autophagie de mycobactéries et de salmonelle chez les cellules épithéliales intestinales [39-41].

CARD15 est un autre exemple de gène associé à la MC dont la fonction a été étudiée. Ce gène encode pour le récepteur de reconnaissance de patron NOD2 exprimés par les cellules épithéliales des intestins. Des études ont démontrés que les patients porteurs des allèles de risque ont une sécrétion déficiente de la cytokine IL-1beta produite par les macrophages activés [42]. Or, cette cytokine aurait un rôle à jouer dans la perméabilité des jonctions serrées des cellules épithéliales [43]. L'introduction des allèles de risque de *CARD15* par transfection dans des cellules épithéliales intestinales a également provoqué

un défaut d'activation de la voie dépendante de NF- κ B lorsque les cellules étaient stimulées à l'aide de ligands bactériens [44, 45]. Ces allèles provoquent donc une sous-reconnaissance des antigènes, ce qui peut mener à un délai de la réponse inflammatoire [40, 46].

De nombreux gènes ayant un rôle dans la différenciation des lymphocytes T naïfs en lymphocytes Th1 et Th17, et dans la production de cytokines, ont été associés aux MII. Par exemple, *STAT3* et *JAK*, tous deux associés aux MII, sont des éléments clé dans la signalisation sous-jacente à l'activation du récepteur IL23R, aussi associé aux MII. Le ligand d'IL23R, IL23, est composé de deux sous-unités, dont l'une IL12p40, est encodée par *IL12B*, un autre gène associé. Les lymphocytes T naïfs stimulés avec IL-23 sont différenciés en lymphocytes Th17, qui produisent IL-22. Cette cytokine, également associée aux MII, active l'expression de peptides antimicrobiens par les cellules épithéliales des intestins, via une signalisation dépendante de STAT3. L'interféron-gamma, encodé par le gène *IFNG*, aussi associé, est responsable de la différenciation des lymphocytes T naïfs en lymphocytes Th1. Une accumulation de défauts dans l'activation de la réponse inflammatoire et dans la différenciation des cellules immunitaires pourrait expliquer la réponse inflammatoire exagérée observée chez les patients ayant une MII.

Des dizaines de gènes associés aux MII ont encore un rôle inconnu dans la pathogénèse des MII, et beaucoup de travail reste encore à faire pour les comprendre pleinement. Certains mécanismes insoupçonnés, comme c'était le cas pour l'autophagie, ne manqueront sans doute pas de nous surprendre.

1.3 LA RÉGION CHROMOSOMIQUE 3p21 ASSOCIÉE AUX MII

1.3.1 Études de liaison et d'association dans la région 3p21

Tableau I : Liste des études ayant publié un signal d'association dans la région 3p21. Les études sont classées par ordre chronologique de la date de publication.

Année de publication	Auteurs	Type d'étude	Cas / Témoins	Marqueur	Valeur-p / Score LOD	Locus	Référence
1996	Satsangi et al.	liaison	186 familles MII	D3S1076 D3S1573	0.0026 0.00021	3p21	28
2000	Rioux et al.	liaison pangénomique	158 trios MII	D3S1766 D3S1285	2.40	3p21	27
2001	Paavola et al.	liaison	93 familles MII	D3S2432	1.68	3p22	26
2001	Hampe et al.	liaison	268 trios MII	D3S1304 D3S1283	1.65 / 1.40	3p26 3p24	47
2002	Duerr et al.	liaison	178 trios MII	D3S1297 D3S3525	3.69 / 1.40	3p26	48
2007	WTCCC	GWAS	2000 MC/3000 ctl	rs9858542	3.58×10^{-8}	BSN	49
2007	Parkes et al.	association approche candidate	1182 MC/2024 ctl	rs9858542	4.9×10^{-8}	BSN	51
2007	Raelson et al.	GWAS	477 trios MC	rs9821675 rs1131095	1.99×10^{-6} 8.06×10^{-5}	CAMKV APEH	50
2008	Beckly et al.	association approche candidate	608 MC/670 CU/1592 ctl	rs1128535	0,002 (MC)	TRAIP	53
2008	Goyette et al.	association approche candidate	1155MII/754 ctl/1062 trios	rs9858213 rs3197999	1.48×10^{-6} 7.25×10^{-6}	RNF123	52
2008	Fisher et al.	GWAS	1841 CU/2935 ctl 1841 CU/4408 ctl	rs3197999 rs9858542	8.0×10^{-6} 1.3×10^{-6}	MST1 BSN	56
2008	Barrett et al.	méta-analyse GWAS	5555 MC/6638 ctl/1339 trios	rs3197999	1.15×10^{-12}	MST1	33
2008	Franke et al.	association approche candidate	1850 MC/1103 CU/1817 ctl	rs9858542	1.25×10^{-5} CU 5.06×10^{-7} MC	BSN	54
2009	Imielinski et al.	GWAS	2895MII/7854 ctl	rs3197999	2.41×10^{-11}	MST1	58
2009	Barrett et al.	GWAS	2361 CU/5417 ctl	rs9858542	7.0×10^{-9}	BSN	55
2010	Mc Govern et al.	méta-analyse GWAS	4702 CU/8371 ctl	rs3197999	3.8×10^{-9}	MST1	35
2010	Franke et al.	GWAS	1043 CU/1703 ctl	rs3197999	0.024	MST1	57
2010	Franke et al.	méta-analyse GWAS	6333 MC/15056 ctl	rs3197999	6.17×10^{-17}	MST1	34
2011	Anderson et al.	méta-analyse GWAS	6687 CU/19718 ctl	rs9822268	1.60×10^{-17}	APEH	32

La région 3p21 a été liée et associée plusieurs fois aux MII. Le tableau I montre une liste des études publiées qui ont détecté un signal d'association dans cette région.

Ainsi, la région 3p a été liée à la MC dès la première étude de cartographie de liaison pangénomique effectuée en 1996 [28]. Cette découverte a ensuite été répliquée par plusieurs autres cartographies de liaisons [26, 27, 47], et par des études d'association avec microsatellites tentant de définir la région exacte d'où provient ce signal [47, 48].

Par la suite, en 2007, deux premières études d'association pangénomique indépendantes [49, 50] ont détecté des signaux d'association à la MC pour des SNP de la région 3p21, rs1131095 et rs9858542, confirmant les résultats de liaison des études précédentes pour cette région. Le signal d'association au SNP rs9858542 a ensuite été répliqué dans une étude d'association par locus candidat [51]. Malheureusement, le SNP rs9858542, situé dans le gène *BSN*, induit un changement d'acide aminé synonyme dans une protéine n'ayant pas un rôle dans l'immunité ou la pathogenèse des MII. Cependant, plusieurs autres gènes dans ce locus sont intéressants.

Par la suite, dans le but d'identifier les véritables variantes causales dans ce locus, une première étude d'association par gènes candidats a été publiée par notre laboratoire [52]. En effet, afin de réduire la région étudiée, un premier raffinement de cette région avait été effectué en chevauchant les régions liées dans les études de cartographie précédentes, résultant ainsi en une région de 30 Mb contenant plus de 400 gènes. Ensuite, une sélection de 12 gènes candidats dont la fonction ou le profil d'expression dans la littérature démontrait le plus de liens avec l'immunité, l'inflammation, les intestins ou les MII, a été faite. Des 89 SNP couvrant la variabilité génétique commune de ces gènes candidats, 5 SNP montrant une association nominale ont été testés en réplification. Quatre SNP ont répliqué l'association et ont été associés à la fois à la CU et la MC, indiquant que cette région de susceptibilité est partagée pour ces deux maladies. Cette première étude a permis de déterminer que ces signaux d'association définissent ensemble une

grande région en LD de 1,2 Mb, et qu'il n'y a en fait qu'un seul signal dans cette région. Enfin, le SNP ayant obtenu le signal d'association combiné le plus fort, rs9858213, est en LD avec rs9858542, le signal répliqué dans ce locus. Une région définie par les SNP très corrélés avec ces deux SNP a permis de délimiter une région de 336kb dans laquelle a été identifié le SNP codant non-synonyme rs3197999, causant le changement R689C dans la protéine MSP (*macrophage stimulating protein*) encodée par le gène *MST1* (*macrophage stimulating 1*). Publiée au même moment, une étude d'association indépendante a permis d'associer à la MC un SNP corrélé à des variantes codantes non-synonymes du gène *MST1R* dans la région 3p21 [53] et qui n'est pas en LD avec les SNP associés ci-haut. Par la suite, une autre étude par approche candidate a répliqué le signal d'association à la fois à la MC et à la CU du SNP rs9858542 [54], confirmant ce locus comme facteur de risque génétique à la CU. Enfin, plusieurs études d'association pangénomiques aux IBD ont répliqués dans différentes populations caucasiennes les signaux d'association aux SNP rs3197999 et rs9858542 [55-58].

Enfin, depuis 2008, une vague de méta-analyses d'études d'association pangénomiques, qui réunissent chacune des milliers de cas et près du double de contrôles caucasiens, ont été effectuées afin de donner plus de puissance aux études pour détecter les variantes ayant un faible effet. Les valeurs-p d'association obtenues pour rs3197999 dépassent grandement le seuil significatif pour une étude d'association pangénomique de 5×10^{-8} , et vont jusqu'à 6.17×10^{-17} pour MC et 4.6×10^{-16} pour CU [32-35]. Plus particulièrement, des analyses eQTL dans une méta-analyse de CU récente publiée par Anderson et al. [32] pointent également vers *MST1* comme gène candidat dans la région 3p21. Ainsi, au moment de débiter le projet d'étude décrit dans ce mémoire, *MST1* était l'un des meilleurs gènes candidats dans ce locus, car la nature de sa variante R689C en fait un facteur potentiel de risque aux IBD. Enfin, le manque de LD entre ce signal d'association et celui détecté dans le gène *TRAIIP* suggère que la région 3p21 contient au moins deux signaux d'association indépendants. Ce deuxième signal d'association est en LD avec des

variantes codantes non-synonymes du gène *MST1R*, ayant un rôle dans la même voie de signalisation que *MST1* et donc ayant une fonction intéressante.

1.4 MST1, MST1R, R689C

1.4.1 MSP

Le gène *MST1* encode pour le précurseur de la MSP. Ce précurseur inactif de 78 kD, est exprimé par les hépatocytes et est en libre circulation dans le sang. Lorsqu'il arrive à un site d'une inflammation, le précurseur est clivé par l'un des facteurs de coagulation XIIa, XIa et la kallikréine, ou par la matriptase, une protéine exprimée à la surface des macrophages ayant démontré une haute efficacité à cliver la pro-MSP [59, 60]. Une fois clivée en deux sous-unités (alpha : 53 kD, et beta : 25 kD) qui demeurent liées ensemble par un lien disulfure entre les cystéines 468 et 588 [61], la MSP devient active et peut librement se lier à son récepteur, le Récepteur-d'origine-nantaise (RON).

La protéine MSP contient quatre domaines *Kringle* et un domaine sérine-protéase, ce qui l'inclus dans la famille du plasminogène, de la prothrombine, de l'urokinase et de l'*Hepatocyte growth factor/scatter factor* (HGF). Contrairement au reste de la famille des protéines *Kringle*, la MSP de même que le HGF, avec qui elle partage une homologie de 45% de sa structure, ont un site sérine-protéase inactif, dû à trois changements d'acides aminés dans le site actif de ce domaine [62]. Malgré que ce domaine soit inactif, il permet cependant à la sous-unité beta de MSP de se lier avec une grande affinité ($EC_{50}=0,25nM$) à RON, bien que ce récepteur ait également une certaine affinité pour l'unité alpha ($EC_{50}=16,9nM$) [63, 64]. La liaison entre l'unité beta et le récepteur RON s'effectue surtout via une pochette entourée de deux groupes d'acides aminés de charge positive, composés des arginines 683, 687, et 689, ainsi que 637, 639 et 641 [65]. Ce site

est extrêmement conservé chez les mammifères [66], et il a déjà été démontré par mutagenèse ponctuelle que R683 est essentiel à la liaison de la chaîne beta de MSP avec RON, ce qui suggère l'importance de ces groupements d'arginine autour du site de liaison de la MSP [63]. Enfin, bien que le site d'interaction entre l'unité alpha et le récepteur RON reste inconnu, il a été démontré que seule la liaison entre RON et la MSP sous forme d'hétérodimères (alpha/beta) permet une réponse biologique sous-jacente commençant par l'autophosphorylation du récepteur RON.

1.4.2 RON

RON est encodé, à partir du gène *MST1R*, sous forme d'un précurseur de 190kD. Ce précurseur est clivé par les enzymes du réticulum endoplasmique, puis RON est exprimé à la surface des cellules sous la forme d'un hétérodimère transmembranaire, où l'unité alpha (40 kD) et l'unité beta (150 kD) sont liées par un lien disulfure [67-69]. Ce récepteur est exprimé dans une grande variété de tissus et de cellules humaines, dont les poumons, la peau, le côlon, l'intestin grêle, les macrophages, le thymus, les oviductes et les muqueuses nasales [67, 69-71]. Dans les lignées cellulaires humaines immortalisées, RON est particulièrement exprimé dans les carcinomes mammaires (T47D) [67], les cancers du poumon (H249, H66, H596) [72] et dans les adénocarcinomes du côlon (HT-29, HCT-15) [73, 74]. La séquence d'acide-aminés de la protéine humaine RON est homologue à 73,6% de son équivalent chez la souris [68].

L'unité extracellulaire alpha contient une partie d'un domaine SEMA (sémaphorine), dont le rôle probable est d'entremettre les interactions entre protéines. La partie extracellulaire de la chaîne beta contient l'autre partie du domaine SEMA, mais aussi un domaine PSI (plexine, sémaphorine, intégrine), dont l'implication dans l'adhésion entre les récepteurs a été suggéré [75]. En effet, il a été démontré que non seulement RON se présentait à la surface des cellules sous forme de dimères, mais qu'il était également capable d'interagir avec d'autres types de récepteurs : EPO, MET, lequel est le récepteur de HGF, et IL-3R [76-78]. RON contient également dans sa partie extracellulaire un

domaine IPT (*Ig-like*, plexine, facteur de transcription), qui semble avoir un rôle dans l'immunité et le développement neuronal [75]. Le domaine intracellulaire tyrosine-kinase de ce récepteur présente de nombreux résidus tyrosine qui ont des rôles bien précis. Les tyrosines 1238 et 1239, à la suite d'une stimulation avec MSP, permettent l'autophosphorylation des dimères RON et l'activation de son site catalytique. Les tyrosines 1353 et 1360 sont le site d'arrimage multifonctionnel de ce récepteur, là où commence la voie de signalisation sous-jacente [67, 79-81].

1.4.3 La voie de signalisation MSP/RON

La voie de signalisation sous-jacente à l'autophosphorylation de RON, suite à une stimulation avec son ligand, active plusieurs processus très divers, dépendamment du type cellulaire étudié. Il a été démontré que le site d'arrimage multifonctionnel de RON peut interagir avec diverses tyrosines kinases et protéines adaptatrices, tels que PLC γ , p85, Shc, Grb2, PBK et Src [79] (fig. 2). Ce sont des intermédiaires de cascades de signalisation menant soit à l'inhibition de l'inflammation, à la survie cellulaire, à la mort cellulaire, à la motilité et l'adhésion des cellules, à des changements morphologiques ou à la résorption des os [82]. Il existe également d'autres effets de la stimulation de RON avec la MSP, mais on ne connaît pas les voies de signalisation empruntées. Globalement, cette voie de signalisation semble avoir beaucoup d'effet sur l'inflammation, l'agressivité des cancers et la cicatrisation.

En effet, il a été démontré que MSP activait le rôle régulateur des macrophages et inhibait l'inflammation. Les macrophages de souris stimulés avec la MSP phagocytent les

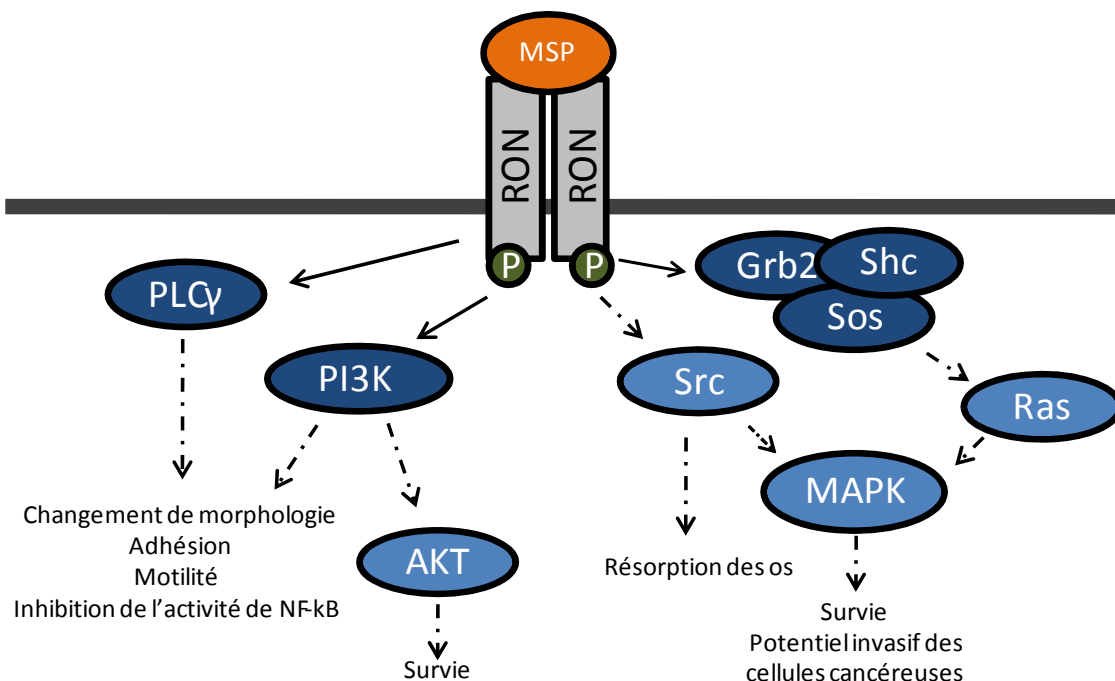


Figure 2: Voie de signalisation sommaire suite à la phosphorylation de RON après la stimulation au MSP
Les flèches pleines représentent des liaisons directes entre les protéines, alors que les flèches pointillées représentent des liaisons indirectes (Inspiré de Danilkovitch et Leonard, 2001 [83]).

érythrocytes entourés par le complément C3bi [84], il y a inhibition chez ces cellules de la production de iNOS et de NO [85-87], et transcription de IL-6 [88]. MSP peut aussi agir comme chimio attractant pour ces cellules. Également, un prétraitement à la MSP de macrophages murins stimulés avec un ligand bactérien comme LPS mène à une répression de l'activité de NF-κB dépendante du LPS, et il en découle une diminution de la production de molécules pro-inflammatoires comme Cox-2, IL12p40, IL12p35 et TNF-α [88, 89]. Une stimulation à la MSP augmente aussi la transcription de l'arginase, un compétiteur de l'iNOS pour la L-arginine, le substrat qui lui sert pour la production de NO [90]. Enfin, les macrophages sont pourvus à leur surface de la matriptase, une protéase capable de cliver et ainsi d'activer la MSP [91]. La présence de cette protéine

sur ce type de cellules suggère que la MSP pourrait avoir un rôle spécifique à accomplir sur la fonction des macrophages.

Aussi, une fois MSP lié à RON, qui se présente à la surface des cellules sous forme de dimères, ce dernier peut s'autophosphoryler et activer des voies propices à aider la cicatrisation des tissus, mais également à exacerber l'agressivité des cellules cancéreuses. Il a été démontré que MSP protège les kératinocytes humains et murins contre l'apoptose, mais aussi induit chez ces cellules des réponses chimiotactiques et de la prolifération [92, 93]. Ces effets sont sans doute bénéfiques à la cicatrisation, d'autant plus que la pro-MSP est clivée et activée par des facteurs de coagulation, donc sur des sites où la barrière épithéliale brisée nécessite la cicatrisation. MSP pourrait avoir un rôle dans la cicatrisation en attirant les cellules épithéliales, en modifiant leur motilité vers la brèche, et en favorisant leur survie et leur prolifération. MSP pourrait également attirer sur le site les macrophages et ainsi promouvoir la phagocytose des débris et des micro-organismes qui auraient pu pénétrer la barrière épithéliale. La surexpression de RON et de MSP dans des exsudats de cicatrice de tissus épithéliaux chez les souris corrobore cette idée [70]. À la lumière de ces rôles pour RON, il n'est donc pas étonnant que plusieurs études affirment trouver une surexpression de RON dans une multitude de lignées cancéreuses. La suppression de son effet anti-apoptotique et de croissance cellulaire lors de son *knockdown* vient appuyer l'importance de RON dans l'agressivité de ces cancers.

1.5 HYPOTHÈSES ET OBJECTIFS

Pour de nombreux loci associés aux MII, on ne connaît toujours pas les variantes qui sont les facteurs de risque pour ces maladies. Beaucoup de travail doit être fait afin d'identifier les gènes candidats, les variantes causales, mais aussi la fonction de ce gène et l'effet de la mutation causale sur cette fonction. Ce mémoire se concentre plus particulièrement sur les signaux d'association détectés dans la région chromosomique 3p21.

En effet, des signaux d'association aux MII dans la région 3p21 ont été détectés pour des variantes en LD avec les régions codantes des gènes *MST1* et *MST1R*, encodant respectivement MSP et RON. Le LD entre certains des signaux est très faible, ce qui soulève notre première hypothèse que la région 3p21 contiendrait plusieurs facteurs indépendants de risque aux MII provenant des gènes *MST1* et *MST1R*.

Chez les macrophages de souris, il a été démontré que MSP avait un effet anti-inflammatoire. Plusieurs études publiées montrent que MSP a aussi un effet sur la migration, la cicatrisation des tissus et la survie cellulaire chez les cellules épithéliales. Notre deuxième hypothèse est qu'un défaut de la voie de signalisation MSP/RON, nos deux candidats, pourrait perturber la réponse inflammatoire ou la survie des cellules intestinales chez l'humain et causer une MII.

De plus, la mutation R689C (rs3197999) est située dans le site de liaison de MSP avec son récepteur RON, alors qu'une mutation à un acide-aminé semblable a complètement inhibé cette liaison. Notre troisième hypothèse est que la mutation R689C modifie la fonction normale de la protéine MSP.

Ainsi, le travail décrit dans ce mémoire vise à la fois à identifier des variantes causales potentielles dans la région 3p21, mais aussi à vérifier le rôle d'un des gènes candidats et tester l'effet d'une de ces mutations par essai fonctionnel. Plus précisément, les objectifs sont :

1. Tester pour association d'autres variantes de *MST1* et *MST1R*, les deux gènes candidats dans la région 3p21.
2. Effectuer une méta-analyse d'études d'association et d'analyses conditionnelles logistiques de trois variantes associées nominalement et peu corrélées entre elles

qui ont été génotypées dans des études précédentes, afin de s'assurer de l'indépendance des signaux d'association dans la région 3p21.

3. Tester *in vitro* l'effet inhibiteur de MSP, encodée par MST1, sur la transcription de molécules pro-inflammatoire chez les macrophages humains dérivés des monocytes du sang, stimulés avec un ligand bactérien.
4. Tester *in vitro* l'effet d'un traitement à la MSP sur la cicatrisation, la dispersion et la phosphorylation d'AKT chez des cellules humaines de lignées immortalisées d'épithélium du côlon.
5. Produire une protéine recombinante MSP purifiée et activée à partir de la transfection d'un plasmide commercial dans des cellules de mammifère et tester l'effet de cette protéine sur la fonction normale de MSP sur un des modèles de cellules humaines cités plus haut.

CHAPITRE 2

MATÉRIEL ET MÉTHODES

1.3 CULTURE CELLULAIRE

2.1.1 Lignées cellulaires et milieux

Les lignées cellulaires humaines de monocytes THP-1 et HL-60, la lignée de cellules humaines embryonnaires de rein HEK-293, ainsi que les lignées de cellules épithéliales humaines du côlon HCT-15, HT-29, et HCT-116 proviennent de la collection *American Type Cell Collection* (ATCC) et ont été cultivées selon les méthodes et les milieux suggérés par le fournisseur. Les cellules ont été cultivées dans des milieux contenant 10%

de sérum bovin foetal (*fetal bovine serum*, FBS) et 1% de pénicilline/streptomycine. Exceptionnellement, les HL-60 ont été cultivées dans un milieu contenant 20% de FBS, et le milieu des THP-1 est additionné de 0,1% de β -Mercaptoéthanol. Les THP-1, les HL-60 et les HCT-15 ont été cultivées dans un milieu RPMI 1640 Glutamax, les HEK-293 dans du DMEM et les HCT-116 ainsi que les HT-29 ont été cultivées dans du milieu McCoy's. Toutes les cellules ont été cultivées en incubateur réglé à 37°C et 5% CO₂. Les monocytes du sang ont été cultivés à l'incubateur dans du milieu Iscove's + 10% FBS et 1% pénicilline/streptomycine.

2.1.2 *Purification des monocytes sanguins et culture pour différenciation en macrophages*

Pour l'extraction de monocytes du sang, 80 ml de sang ont été prélevés sur deux volontaires sains, par ponction veineuse, dans des tubes contenant comme anticoagulant de l'acide éthylènediaminetétraacétique (EDTA). Le sang a été dilué dans une solution saline-Phosphate (PBS) (GIBCO, Invitrogen) dans une proportion de 3:5, puis les différents éléments du sang ont été séparés par gradient de densité avec du Ficoll-Paque (GE Healthcare, 17-1440-02). Le sang dilué a donc été déposé en aliquots de 12ml à la surface de 3 ml Ficoll-Paque dans des tubes de 15 ml. Les tubes sont ensuite centrifugés à 2000 rpm pour 30 minutes sans freins. Les cellules les plus denses (érythrocytes, neutrophiles) se retrouvent au fond, et les globules blancs, moins denses, se retrouvent entre la couche de sérum et la couche de Ficoll, formant une couche communément appelée « *buffy coat* ». Les globules blancs ont été récupérés à l'aide d'une pipette pasteur jetable. Au total, 10 ml de cellules ont été diluées 1:5 avec du PBS dans deux tubes de 50 ml. Les globules blancs ont ensuite été centrifugés 10 minutes à 1500 rpm et les culots de cellules rincés 3 fois avec 25 ml de milieu complet chaud. Ensuite, les cellules ont été suspendues dans un total de 20 ml de milieu complet, distribuées dans 10 pétris Cellbind (Corning, Fisher) de 6 cm pour cellules adhérentes, et placées à l'incubateur pendant 2 h30 min. Finalement, les pétris ont été rincés doucement 4 fois au

PBS afin d'éliminer les cellules non-adhérentes, et 2 ml de milieu complet ont été ajoutés aux monocytes adhérents avant de les retourner à l'incubateur. Pour l'obtention de macrophages, les monocytes du sang ont été différenciés pendant 7 jours à l'incubateur avec des changements de milieu frais aux 2-3 jours. Les volontaires utilisés pour ce protocole ont tous signé un formulaire de consentement approuvé par le comité d'éthique de l'institut de Cardiologie de Montréal. Des macrophages humains ont aussi été obtenus par un protocole de différenciation des monocytes THP1. Brièvement, 8×10^5 cellules THP-1 ont été ensemencées par pétri de 6 cm dans 2 ml de milieu RPMI complet. Afin de les différencier, les cellules ont été traitées avec une concentration finale de 10ng/ml de phorbol myristate acétate (PMA) (Sigma, P1585) et laissées à l'incubateur 48 heures.

2.2 TECHNIQUES DE BASE EN BIOLOGIE MOLÉCULAIRE

2.2.1 *Extraction des protéines*

Toutes les extractions de protéine ont été faites selon le même protocole. Plus précisément, pour les cellules cultivées dans des plaques à 6 puits Cellbind (Corning, Fisher), les cellules ont été rincées 2 fois dans du PBS, 300 ul de tampon de lyse (5 mM HEPES, 1% Triton x-100, 4 mM EDTA, 100mM NaF, 10 mM Na_3VO_4 , 1% PMSF, 1% Cocktail d'Inhibiteurs de Protéases (Sigma, P8340)) ont été ajoutés à chacun des puits, puis les plaques de cellules ont été incubées sur glace pendant 30 minutes avant de collecter les lysats cellulaires. Pour les culots de cellules, les cellules ont été rincées 2 fois dans 1 ml de PBS, les culots ont été suspendus dans 300ul de tampon de lyse, et les tubes ont été incubés sur glace pendant 30 min. Lorsqu'une évaluation de la phosphorylation était prévue, le tampon de lyse a été additionné du Cocktail d'inhibiteurs de Phosphatases 1 dans une proportion de 1 :100 (Sigma, P2850). Les lysats cellulaires ont ensuite été centrifugés 30 minutes à 4°C à vitesse maximale et les culots de débris cellulaires ont été jetés. La concentration en protéines de ces extraits cellulaires a été déterminée à l'aide du kit BCA de quantification de protéine (Pierce, Thermo Fisher,

23227) d'après les indications du fabricant et à l'aide d'un lecteur de plaque à une absorbance de 562nm. Les échantillons de protéines ont été conservés à -80 °C.

2.2.2 Immunobuvardage de type Western

Les immunobuvardages de type western ont été réalisés dans le but d'évaluer la quantité de MSP recombinante produite ainsi que le niveau de phosphorylation de AKT induit par la stimulation des cellules épithéliales du côlon avec du MSP. Le protocole utilisé est tiré des méthodes décrites dans le livre de référence *Current Protocols in Molecular Biology* [94]. Pour chaque anticorps, les détails de l'immunobuvardage sont décrits dans le tableau III. Brièvement, les extraits de protéines ont été transférés dans les puits d'un gel SDS-Tris 8% acrylamide (voir tableau II), puis ont migré sous conditions dénaturante à 65mA pendant 1 heure dans un tampon SDS-TRIS 1X (0.025M Tris base + 0.192 M glycine + 0,1% SDS).

Tableau II: Recette de réactifs pour la fabrication des gels d'acrylamide 8%.

Réactif	Vol. Gel de séparation (Bas)	Vol. Gel de superposition (Haut)
Acrylamide 30%	4 ml	650 ul
tampon Tris-Cl/SDS pH 8.8 4x (0,5 M Tris-Cl + 0,4 % SDS)	3,75 ml	-
Tris-Cl/SDS pH 6.8 4x (1.5 M tris-Cl + 0,4 % SDS)	-	1,25 ml
H ₂ O	7,25 ml	3,05 ml
10% ammonium persulfate	50 ul	25 ul
TEMED 100%	10 ul	5 ul

Les protéines ont ensuite été électro-transférées sur une membrane de nitrocellulose pendant 12 heures à 25V. Pour la détection de protéine, les membranes ont été incubées 1 ou 2 heures dans un tampon de blocage 5% lait dans du tampon TTBS (TBS 1X + 0,1% Tween-20), ou 3% BSA dans du tampon TBS 1X (10 mM Tris-Cl pH 7.5 + 15 mM NaCl), suivi d'une incubation de 1 heure avec les anticorps primaires dirigés contre les protéines d'intérêt (Tableau III) dilué dans le tampon de blocage. Les membranes ont été

rincées 4 fois 10 minutes dans un tampon TTBS. Les membranes ont ensuite été incubées pendant 1 heure avec un anticorps secondaire anti-lapin ou anti-souris étiqueté avec la peroxydase du raifort (HRP, *horseradish peroxidase*) dilué dans un tampon de 5% lait dans du TTBS, rincées 4 fois 10 min dans un tampon TTBS et une fois 20 min dans un tampon TBS. Finalement les membranes ont été traitées à l'ECL (PerkinElmer, NEL 10300IEA) pour révéler les protéines par chimioluminescence.

Tableau III: Description et détails des manipulations pour chacun des anticorps utilisés dans les immunobuvardages.

Anticorps	Source	Tampon de blocage	Dilution	Temps d'incubation
Primaire, lapin IgG anti-huPhosphoAKT	R&DAF887	3% BSA/TBS	1/5000	2h
Primaire, lapin IgG anti-huAKT	R&DAF2055	5% lait écrémé/TTBS	1/5000	2h
Primaire, souris IgG anti-Penta-Histidines	Qiagen 34660	3% BSA/TBS	1/2000	1h
Secondaire, chèvre anti-lapin	Cedarlane CLCC43007	5% lait écrémé/TTBS	1/5000	1h
Secondaire, lapin anti-souris	Pierce 31457	5% lait écrémé/TTBS	1/5000	1h

2.2.3 Réaction de polymérisation en chaîne quantitative

Les cellules sont rincées doucement une fois au PBS, puis lysées dans 700 ul de tampon de lyse RLT (Qiagen, RNeasy Plus Mini Kit, 74134). Les acides ribonucléiques messagers (ARNm) ont été extraits à l'aide du kit RNeasy de Qiagen d'après la méthode suggérée par le fabricant, et quantifiés par la plateforme Bioanalyzer 2100 (Agilent) avec le kit RNA 6000 Nano Reagents (Agilent, 5067-1511). Pour chaque échantillon, 1 ug d'ARN a été utilisé afin de préparer une librairie d'ADN complémentaires à l'aide du kit High Capacity cDNA Reverse transcription (AB, 4368813). Afin de mesurer l'expression des gènes d'intérêts *TNF- α* , *MST1R*, *IL12B*, *CD-14* et *CD-206*, ainsi que du

gène de référence *HPRT1* (amorces : annexe C), une réaction de polymérisation en chaîne (PCR) quantitative a été fait sur un instrument de PCR MX3005P (Stratagene), en utilisant le kit Platinum SYBR Green qPCR Super-mix UDG (Invitrogen, 11733-046) d'après le protocole du fabricant. Les cycles de réaction de PCR sont décrits dans le tableau IV.

Tableau IV: Cycles de PCR quantitatif

Nombre de cycles	Température et durée
1x	95°C 10min
40x	1)95°C 30sec 2)60°C 1min
1x	1)55°C 30sec 2) 95°C 30sec

2.3 ÉVALUATION DE MODÈLES FONCTIONNELS

2.3.1 Expression de *MST1R* dans les lignées cellulaires

Afin de vérifier l'expression du gène *MST1R* dans les différentes lignées cellulaires utilisées lors de nos essais, l'expression de l'ARNm a été évaluée par PCR quantitatif. L'extraction des ARNm en triplicata d'un million de cellules HL-60, THP-1, HT-29, HCT-15, HCT-116 et HEK-293, la rétrotranscription et le PCR quantitatif ont été faits tels qu'expliqués précédemment. Puisque dans une expérience préalable nous avons déterminé l'expression basale de *MST1R* chez les THP-1 différenciés en macrophages, leur expression est utilisée comme base pour comparer les expressions de *MST1R* chez les autres lignées de cellules.

2.3.2 Expression des gènes pro-inflammatoires

Dans le but d'évaluer l'effet de la MSP recombinante sur une réponse inflammatoire induite par un ligand bactérien (LPS) chez les macrophages, 2 modèles différents ont été évalués. Les macrophages dérivés des monocytes circulants et de la différenciation des cellules THP-1 ont été prétraités pendant 4 heures avec du milieu complet contenant ou non 80ng/ml de MSP recombinante de type sauvage (R&D Systems, 352-MS/CF). Les macrophages ont ensuite été stimulés par l'addition de milieu complet additionné ou non de 10ng/ml de LPS (Sigma,L-2654). Après 3 heures de cette stimulation, l'expression des ARN des gènes *TNF- α* , *IL12p40*, *CD-14* et *CD-206* ont été mesurés afin de déterminer le niveau de différenciation des macrophages, et de mesurer la production de molécules pro-inflammatoires. Des monocytes ou cellules THP-1 non-différenciés ont également été lysés de cette manière pour être utilisés comme contrôle de différenciation.

2.3.3 Cicatrisation

Afin de mesurer l'effet de la MSP sur la motilité cellulaire, un modèle de cicatrisation d'un épithélium dérivé de lignées de cellules épithéliales du côlon a été utilisé. Brièvement, des cellules de lignée épithéliale du côlon HT-29 ont été distribuées dans des plaques à 6 puits, à grande confluence : 15 000 000 cellules/puits. Ces cellules ont ensuite été cultivées pendant deux jours supplémentaires afin de s'assurer d'obtenir une couche épithéliale confluente et différenciée bien régulière. Après deux lavages au PBS à température pièce, les cellules ont été cultivées dans des milieux sans sérum pour 12h. Puis trois blessures ont été faites dans l'épithélium de cellules à l'aide d'un embout de pipette P20 stérile, et trois égratignures perpendiculaires aux blessures ont été pratiquées à l'aide d'une lame de rasoir sous la plaque, créant ainsi 9 intersections facilement repérables pour nos observations ultérieures. Les cellules ont ensuite été lavées 2 fois au PBS à température pièce afin d'éliminer les cellules détachées par les blessures, et traitées avec soit 40ng/ml de HGF recombinant (R&D Systems, 294-HG/CF), soit 250ng/ml de MSP recombinante de type sauvage, ou portant la mutation C672A (R&D

Systems, 4306-MS/CF), soit 10ng/ml de PMA, ou laissées sans traitement. Les photos de six intersections différentes par puits ont été prises au temps zéro, 24h, 48 h et 72 h, à l'aide d'un microscope en contraste de phase à grossissement 40x et du logiciel ZEN LE 2009 (Carl Zeiss). Pendant l'expérience, les milieux sans FBS, avec ou sans MSP, HGF ou PMA sont changés aux 24 heures. La largeur de la blessure a été mesurée en pixels à l'aide du programme ImageJ (Wayne Rasband, NIH, v 1.43u) et la distance de migration des cellules à travers la blessure est exprimée comme la différence en pixels entre la mesure aux temps 72h et zéro, 48h et zéro ou 24h et zéro. Six mesures de la distance de migration ont été faites pour chaque point temporel et traitement, et sont résumées dans le graphique sous forme de boîtes et moustaches (*boxplot*) pour chaque traitement. L'effet de chaque traitement a été calculé par ANOVA avec le programme d'analyses statistiques R, et ensuite rapportés dans le graphique sous forme de valeur-p. La normalité des données a aussi été vérifiée par graphique quartile-quartile avec le même programme.

2.3.4 Migration des cellules

Pour déterminer l'effet de la stimulation de MSP sur la dispersion de cellules épithéliales du côlon, des cellules HCT-15, HCT116 et HT-29 ont été réparties à très faible confluence (HCT-15 : 100 000, HCT116 : 80 000, et HT-29 : 50 000 cellules/puits) dans des plaques de six puits et laissées à l'incubateur pour 24 h. Douze heures avant les traitements, les cellules ont été rincées 2 fois au PBS, et les milieux ont été changés pour du milieu sans FBS. Les cellules ont été ensuite soumises à différents traitements (PMA 10ng/ml, MSP recombinante C672A 250ng/ml, ou HGF 50ng/ml) ou laissées sans traitement. Pour chaque traitement, quatre colonies de cellules par puits ont été visualisées à grossissement 100x par microscope à contraste de phase, et photographiées aux temps 0, 24, 48 et 72h. Des photographies à grossissement 40X de la zone entourant les colonies étudiées ont été également prises aux temps zéro et 24h, afin de s'assurer de la conformité de ces colonies par rapport à l'ensemble des colonies du puits. La dispersion des cellules de chaque colonie, ainsi que leur morphologie générale, ont été comparés de façon qualitative entre les différents traitements à travers le temps.

2.3.5 *Phosphorylation d'AKT*

Les cellules HCT15, HCT116 et HT29 ont étéensemencées à confluence moyenne (HCT-15 : 250 000, HCT116 : 200 000, et HT-29 : 160 000 cellules/puits) dans des plaques à six puits. Après 24 heures d'incubation, les cellules ont été rincées deux fois au PBS et les milieux remplacés pour des milieux sans FBS. Douze heures plus tard, les cellules ont été traitées avec soit du HGF 50 ng/ml pour 10 minutes, du MSP recombinant de type sauvage 250ng/ml pour 5 minutes, du FBS (Invitrogen, 12483-012) 10% pour 30 minutes, ou laissées sans traitement. Sous les mêmes conditions expérimentales, les HCT-15 ont été aussi stimulées avec de la MSP recombinante mature produite au laboratoire par transfection (voir méthodes : production de la MSP recombinante). Les milieux ont ensuite été retirés, et les puits rincés deux fois au PBS froid. Les cellules ont été lysées pour une extraction des protéines totales (voir méthodes : extraction des protéines) et ces dernières ont été détectées par immunobuvardage à l'aide d'anticorps dirigés contre la protéine AKT phosphorylée et non phosphorylée.

2.4 EXPRESSION DE LA MSP RECOMBINANTE

2.4.1 Production et validation du plasmide d'expression de *MST1*

Nous avons acheté le plasmide d'expression de la *MST1* recombinante du kit d'expression Insect/Mammalia HS-*MST1*-IM-1 (Qiagen, EIM0003619) (fig. 3).

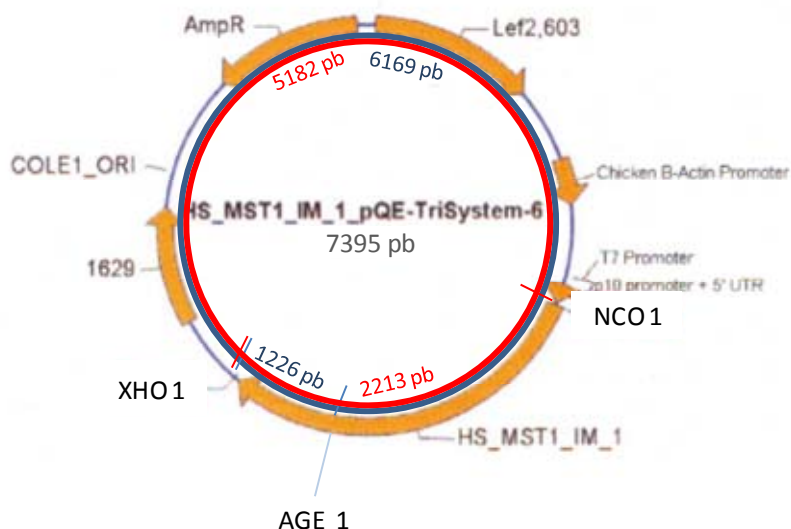


Figure 3: Détails du plasmide HS-*MST1*-IM-1 (Qiagen, EIM0003619) avec ses trois sites de restriction uniques ainsi que la longueur des fragments générés par les différentes digestions. (image modifiée d'après la carte du plasmide HS-*MST1*-IM-1 (Qiagen)).

Ce plasmide contient la séquence codante de *MST1* optimisée, c'est-à-dire contenant des substitutions de nucléotides permettant l'usage des codons plus fréquemment utilisés chez les eukaryotes, afin d'augmenter la stabilité et la production de protéines dans un système d'expression eukaryotique. Il contient aussi en 5' la séquence du peptide signal N-terminal pour la sécrétion de la protéine et en 3' la séquence d'une étiquette C-terminale de 10 histidines (10xHis). Cette étiquette permet de purifier la protéine MSP à l'aide de billes de nickel, mais aussi de détecter la protéine à l'aide d'un anticorps anti-Penta-Histidines. Afin de produire une grande quantité de ce plasmide pour nos

transfections, une suspension de *E. coli* compétentes NEB-alpha (NEB, C2988J), c'est-à-dire de bactéries traitées pour faciliter l'introduction de plasmides, a été transformée par la méthode de « *heat shock* ». La transformation a été effectuée d'après le protocole du fournisseur. Spécifiquement, 100 µl de suspension de bactéries ont été mélangés à 10 µg de plasmide, incubés sur glace 30 min, et placés à 42 °C 30 secondes (sec.). Le tube a ensuite été remis sur glace pour 5 min, et 950 µl de milieu SOC (Invitrogen 15544-034) ont été ajoutés. Le tube a été incubé sous agitation à 37 °C pendant 1h. Les bactéries ainsi transformées ont été ensemencées sur des pétris contenant du milieu LB-agar (10g/L tryptone, 5 g/L extrait de levures, 5g/L NaCl, 15g/L agar, 100µg/ml ampicilline), puis incubées à 37°C pour 24h. Le lendemain, 5 colonies ont été prélevées et mises en culture liquide dans 5 ml de milieu LB (10g/L tryptone, 5 g/L extrait de levures, 5g/L NaCl, 100µg/ml ampicilline) sous agitation à 37°C pour 24h. Ensuite, 200 µl de ces cultures ont été transférés dans 900 ml de milieu LB additionné d'ampicilline, puis incubés 24h à 37°C sous agitation. Le plasmide a été purifié à l'aide du kit Gen Elute HP plasmid Maxiprep. (Sigma, NOA0310-IKT). La concentration de la suspension de plasmide a été déterminée par absorbance (à 230nm) à l'aide d'un spectrophotomètre. Afin de confirmer l'identité du plasmide, 1 µg de plasmide a été digéré soit par une combinaison d'enzymes de restriction Xho1 (Invitrogen, 15231-020) et Nco1 (Invitrogen, 15421-019), soit par une combinaison de Xho1 et Age1 (NEB, R05525) à 37°C pour 1h30min (Tab. V). Le plasmide digéré a ensuite migré sur un gel d'agarose 0,5% (0,25g agar + 0.5 µg/ml bromure d'éthidium + 50 ml de TAE (40mM Tris-acétate, 2mM Na₂EDTA-2H₂O; pH=8.5)), à 70V pendant 45 minutes.

Tableau V: Composition des réactions de digestion du plasmide HS-MST1-IM-1 par des enzymes de restriction.

Réactifs	Vol (uL) XHO 1 - NCO 1	Vol (uL) XHO 1 – AGE 1
Plasmide (1 ug/ul)	2.08	2.08
Tampon NEBuffer 1 10x	1	1
BSA (2mg/ml)	0.5	0.5
Enzyme XHO 1 (10U/ul)	0.5	0.5
Enzyme AGE 1 (5U/ul)	-	0.5
Enzyme NCO 1 (10U/ul)	0.5	-
H ₂ O	5.42	5.42
Total	10	10

Les bandes ont été photographiées sous ultra-violet (UV) et comparées au patron de digestion attendu selon la carte du plasmide d'origine. La séquence codant *MST1* dans le plasmide a aussi été séquencée à l'aide de la technologie Sanger sur un Analyseur d'ADN 3730x1 d'Applied Biosystems. Les amorces antisens et le plasmide ont été fournis au service de séquençage de Génome Québec. Enfin, la séquence a été alignée à la séquence fournie par Qiagen. Puisque le plasmide sera utilisé ultérieurement pour des transfections dans des lignées cellulaires, il a été stérilisé par précipitation dans une solution 1/10 de NaOAc 3M pH 5.2 + 4 volumes d'éthanol 100%, puis centrifugé 30 minutes à vitesse maximale à 4°C. Le surnageant a été éliminé sous une hotte stérile. Le plasmide a ensuite été suspendu dans de l'eau sans Rnase/Dnase à une concentration de 1ug/ul.

2.4.2 Expression et purification de la *MST1* recombinante

Pour exprimer la protéine MSP recombinante, 500 000 cellules HEK293 dans 2 ml de milieu complet ont été distribuées par puits d'une plaque à 6 puits et mises à l'incubateur 24 heures. Pour chaque puits de transfection, 500 ul de milieu réduit en sérum Opti-MEM (GIBCO, Invitrogen, 31985), 4.3 ul de lipofectamine LTX (Invitrogen, 15338-100), et 2.5 ug de plasmide *MST1* stérile ont été combinés et incubés à température pièce pendant 30 minutes. Ce mélange a ensuite été ajouté dans chaque puits, suivi d'une incubation de 12 h. Comme contrôle négatif, des puits ont été transfectés selon le même protocole mais sans plasmide. Le lendemain, le milieu de chaque puits a été remplacé par du milieu complet et les cellules incubées pour une période de 6 heures. Puis les cellules ont été

rincées doucement 4 fois au PBS à température pièce, et incubées dans 2 ml de milieu sans sérum (DMEM 1% pénicilline/streptomycine (pen/strep)) pour 12 heures avant la récolte de protéines. Comme la MSP est une protéine sécrétée, le surnageant des cellules a été transféré dans un tube dans lequel sont ajoutés du 100uM de phényle méthyl sulfonyle fluoride (PMSF) (Sigma, P7626) et 1% de cocktail d'inhibiteurs de protéases de Sigma, et a été centrifugé à 4000 rpm 10 minutes à 4°C afin retirer les débris cellulaires. Les protéines intracellulaires ont également été récoltées en lysant les cellules (voir méthodes : extraction des protéines). Afin d'obtenir un échantillon de MSP avant les étapes de concentration et de purification, les protéines du surnageant des cellules ont été précipitées. L'équivalent du surnageant de 3 puits a été additionné de 4 volumes d'éthanol froid et centrifugé à vitesse maximale 30 min à 4°C. Les culots de protéines précipitées ont été séchés, puis suspendus dans 100 ul d'eau, et cet échantillon a été conservé pour visualisation par immunobuvardage à l'aide d'un anti-corps anti-penta-His. Afin de concentrer les protéines du surnageant, l'équivalent de 9 puits transfectés a été transféré dans des colonnes de concentration et exclusion par poids moléculaire AMICON Ultracel-30k de 15 ml (Millipore, UFC-903008). Ces colonnes ont été centrifugées pendant 20 minutes à 3200 x g à 4°C, puis rincées avec 4 ml de PBS et centrifugées de nouveau. Le retentât de la colonne (environ 150ul) a été recueilli, le filtre rincé avec 150 ul de tampon NPI-20-T (50 mM NaH₂PO₄ + 300 mM NaCl + 20 mM imidazole 0.05% Tween-20, pH 8.0), et les deux volumes combinés sur glace. À cette étape, 100 ul de cet échantillon a été conservé pour visualisation par immunobuvardage à l'aide d'un anticorps anti-penta-His ou anti alpha-MSP. Les 200 ul restant ont été utilisés pour la purification avec les billes de nickel. Pour ce faire, le 200 ul de protéines concentrées ont été ajouté dans un tube à 50 ul de billes d'agarose magnétiques Ni-NTA (Kit plasmide MST1, Qiagen) préalablement rincées dans le tampon NPI-10-Tlg (50 mM NaH₂PO₄ + 300 mM NaCl + 10 mM imidazole 0.05% Tween-20, 1% Igepal CA-630, pH 8.0). Le tube a été incubé 1 hr à 4°C sur un agitateur rotatif afin de permettre l'attachement de l'étiquette 10XHis aux ions de nickel. Les billes ont ensuite été lavées 4 fois dans 500ul de tampon NPI-20-T afin de retirer les protéines sans étiquette 10XHis.

Finally, the proteins with a 10XHis tag were eluted by suspension of beads in 50 µl of elution buffer NPI-500-T (50 mM NaH₂PO₄ + 300 mM NaCl + 500 mM imidazole 0.05% Tween-20, pH 8.0) followed by a 1 minute incubation at room temperature. This elution step was repeated a second time to optimize the recovery of the recombinant MSP. A portion of the purified MSP was incubated for 1 h at 37°C with 100ng of human kallikrein (Calbiochem, 420307) per µl of purified MSP to cleave and activate the latter. The extracts obtained at each step were kept for immunoblotting to visualize the proportion of MSP lost between the steps.

2.5 GÉNOTYPAGE

2.5.1 *Génotypage sur la plateforme SEQUENOM*

Our genotyping was performed on the Sequenom platform. This platform allows testing of up to 38 SNPs simultaneously on a single sample of DNA. The miniaturization of this technique also allows genotyping 384 samples on a single microchip using small amounts of samples and reagents. The experimental approach of the Sequenom platform consists in a multiplexed amplification of the regions surrounding each SNP of interest, followed by the determination of the genotype of each individual by extension of a single base corresponding to the alleles of the SNPs of interest and a separation of the extension products by mass spectrometry. The primers were designed so that they are compatible in a single reaction. Incompatible primers are simply separated in another group.

2.5.1.1 Conception des groupes d'amorces

The specialized software of the Sequenom genotyping platform allows grouping the largest number of SNPs in a limited number of

groupes, et de concevoir les différentes amorces d'amplification ou d'élongation propres à chaque SNP dans des réactions multiplexées. La première étape consiste à obtenir les séquences entourant les SNP (environ 200 paires de bases en aval et en amont du SNP) à partir de la base de données publique de variantes, dbSNP, et les concaténer dans un seul et unique fichier texte compatible aux logiciels de conception de Sequenom. Ce fichier est ensuite téléchargé dans la suite de logiciels en ligne RealSNP de SEQUENOM. En premier lieu, ProxSNP identifie dans les séquences fournies l'endroit précis où se situent les SNP adjacents à notre SNP d'intérêt, basée sur les données prélevées dans dbSNP, afin d'éviter le positionnement des amorces à ces endroits. Puis le fichier est analysé par le programme PreXTEND, qui aligne les séquences nucléotidiques avec le génome de référence dans NCBI. Il identifie les régions de la séquence qui ne sont rencontrées qu'une seule fois dans le génome et pour lesquelles il sera possible de faire des amorces, évitant ainsi d'obtenir des produits d'amplification non-spécifiques. Les SNP pour lesquels il est impossible de trouver une séquence unique sont rejetés. Finalement, ce fichier est analysé par le logiciel Assay Design (Sequenom, v 3.1 et 4.0), qui conçoit les amorces d'amplification ainsi que l'amorce d'élongation, et qui compose les groupes d'amorce compatibles entre elles dans une réaction. La composition des groupes, les séquences des amorces et les masses attendues pour l'amorce et ses produits d'élongation selon l'allèle sont tous inscrits dans un fichier Excel qui sera téléchargé dans la plateforme de génotypage afin de définir les paramètres de lecture.

2.5.1.2 Dilution des amorces

Les différentes amorces d'amplification et d'élongation sont commandées chez IDT (Integrated DNA technologies). Après réception, les amorces d'amplification sont diluées individuellement à 100 μ M, et celles d'élongation sont diluées à 1000 μ M. Les amorces d'amplification ont été groupées d'après le design du fichier de ASSAY DESIGN et diluées à une concentration finale de 500nM dans du tampon TE 0,1x. Les amorces d'élongation ont également été combinées selon les groupes conçus par ASSAY DESIGN, où chaque groupe multiplexés d'amorce a été divisé en trois sous-groupes,

selon leur masse pour être ensuite combinées à une concentration finale de à 7, 10 et 17 uM respectivement dans du tampon TE 0,1x. Une étape de vérification des amorces, consistant en l'évaluation de la quantité relative chaque amorce dans un groupe, est ensuite faite sur le spectromètre de masse MassArray de la plate-forme de génotypage et les concentrations des amorces sont alors ajustées selon la concentration suggérée par le logiciel d'analyse Typer.

2.5.1.3 Amplification par PCR

L'amplification des séquences entourant les variantes d'intérêt a été faite par PCR dans des plaques à 384 puits à l'aide du kit IPLEX GOLD REAGENT de SEQUENOM. (tab. VI). Chaque puits comprend 250ng d'ADN génomique d'individus de notre cohorte et chaque plaque contient également au moins 2 puits sans ADN comme contrôle négatif. Après l'amplification, les produits de PCR sont traités pendant 20 min à 37 °C à la *Shrimp Alkaline Phosphatase* (SAP; kit IPLEX de SEQUENOM) (tab. VII) afin de neutraliser les dNTPs libres avant l'étape d'élongation. (tab. VIII)

Tableau VI: Composition de la réaction de PCR (A) et des cycles de PCR (B) pour le génotypage.

Réactifs	Volume (ul) 1x réaction
Tampon 10x	3
25mM MgCl ₂	0.5
10mM ch dNTPmix	0.6
Hotstar Taq 5U/ul	0.2
H ₂ O	20.1
DNA (50ng/ul)	5
Mix d'amorces (10uM ch)	0.6
Total	30

A.

Nombre de cycles	Température et durée
1x	94°C 10min
45x	1)94°C 30sec 2)56°C 30sec 3)72°C 1min
1x	72°C 5min

B.

Tableau VII: Composition de la réaction de SAP.

Réactifs	Volume (ul) 1 réaction
H ₂ O	1,55
Tampon SAP 10X	0,17
Enzyme SAP	0,3
Total	2

2.5.1.4 Réaction d'élongation

Les réactifs d'élongation sont ensuite ajoutés à la plaque contenant les produits d'amplification (tab. VIII). Cette étape vise à permettre l'hybridation d'amorces sur la séquence adjacente aux variantes d'intérêt, et d'allonger cette amorce en ajoutant la base complémentaire aux allèles du SNP testé. Comme chaque base a un poids différent, la réaction d'élongation permet de produire des produits de masses différentes mais connues pour chaque allèle d'un SNP. Ces poids ont été prédéterminés par ASSAY DESIGN lors de la conception des amorces d'élongation, dépendamment de l'allèle attendue dans le produit d'élongation

Tableau VIII: Composition de la réaction d'élongation.

Réactifs	Volume (ul) 1 réaction
H ₂ O	0,76
Tampon IPLEX 10X	0,2
Mélange de terminaison	0,2
Mélange d'amorces d'élongation	0,8
Enzyme IPLEX	0,04
Total	2

2.5.1.5 Détermination des génotypes

Les réactions d'élongation sont ensuite diluées dans 16 ul d'H₂O et nettoyées à la résine afin de retirer les sels pouvant causer du bruit de fonds lors de la lecture. Un volume de 2 ul des produits d'élongation est ensuite transféré par robot nanodispenseur sur une micropuce Spectrochip II (SEQUENOM, 10117-2) contenant une matrice propriétaire

développée par Sequenom, pour être lues au spectrophotomètre de masse MALDI-TOF MS (*matrix-assisted laser desorption/ionization time of flight mass spectrometer*) MassArray (SEQUENOM). Les produits d'élongation sont d'abord vaporisés sur la puce par un laser. Ils sont ensuite séparés ou migrés sous vide selon leur charge dans un champs magnétique à haute intensité. Comme la charge d'une molécule d'ADN est directement corrélée à sa longueur et à sa séquence, et donc à sa masse, la vitesse de déplacement de chaque produit d'élongation dans le champs magnétique sera fonction directe de la masse de celui-ci. Ainsi, les produits d'élongation atteignent le détecteur et sont enregistrés à différents temps selon leur masse respective. Comme la masse dépend de la séquence de l'oligo, et que l'élongation à un certains SNP ne peut être que pour l'un ou l'autre des 2 allèles attendus du marqueur biallélique, on peut déterminer le génotype pour chaque individu (fig. 4) d'après cette masse. Les résultats de lecture pour tous les individus de chaque puce sont exportés dans un fichier dans lequel est affiché l'identité de l'échantillon et son génotype pour chaque SNP dans le groupe de SNP testé. Avant le contrôle de qualité et l'analyse d'association dans le logiciel Plink, les fichiers de format « Sequenom » sont convertis en fichiers de type PED et MAP utilisables dans PLINK.

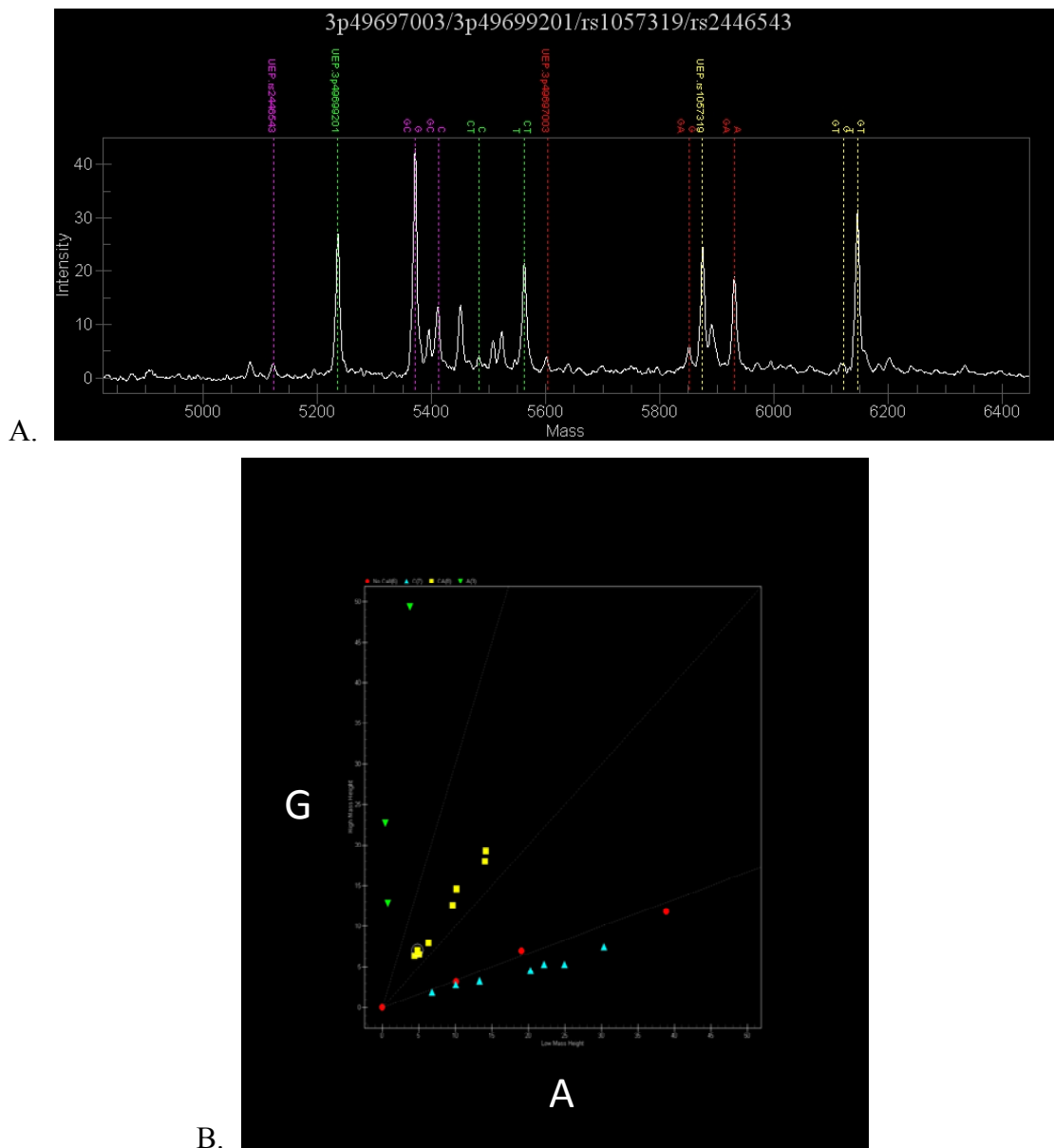


Figure 4: Détermination des génotypes sur la plateforme Sequenom

A : Mesure de la masse de chaque amplicon, de son intensité et identification des allèles. **B :** Graphique des génotypes pour un SNP. Chaque axe représente l'intensité pour un allèle, et chaque point représente le génotype d'un individu. Les points jaunes représentent les hétérozygotes, et les points rouges les individus dont le génotype n'a pu être déterminé.

2.5.2 Amplification spécifique de 3p21

2.5.2.1 Alignement de la région chromosomale 3p21

Les premiers génotypages que nous avons effectués de tous les SNP testés de *MST1* ont échoués. Nous avons donc vérifié si la séquence de la région 3p21 autour du gène *MST1* avait une particularité. Une séquence d'environ 44 000 paires de bases entourant le gène *MST1* a été tirée de la construction NCBI36.3 du génome humain, disponible sur le site du projet d'annotation génomique Ensembl, et a été alignée à l'ensemble du génome humain à l'aide du programme de recherche de séquence nucléotidique de la base de données NCBI, BLASTN 2.2.19 (Annexe B). Ceci a permis de valider la présence de deux duplications de la région 3p21 sur le chromosome 1, et aussi d'identifier les différences entre les duplications et la séquence d'origine afin de concevoir des amorces spécifiques pour l'amplification de la séquence du chromosome 3.

2.5.2.2 Conception des amorces spécifiques à la séquence de *MST1*

En utilisant l'alignement des séquences entourant la région du gène *MST1* sur les duplications identifiées sur le chromosome 1, des paires d'amorces spécifiques pour l'amplification de régions d'intérêt du gène *MST1* ont pu être conçues. Le positionnement de ces amorces a été conçu de sorte qu'elles soient positionnées si possible à l'extérieur de la région dupliquée, ou que les bases ultimes de la terminaison 3' s'aligne sur des différences ponctuelles entre la séquence du chromosome 3 et celles du chromosome 1. De cette façon, les extrémités 3' des amorces spécifiques ne s'hybrideront parfaitement qu'à la séquence du chromosome 3, et moins bien aux séquences du chromosome 1. De plus, les amorces situées dans la région dupliquée contiennent un changement additionnel de nucléotide à proximité de la terminaison 3', déstabilisant ainsi l'hybridation avec la séquence du chromosome 1. Enfin, les amorces ainsi construites ont été analysées par le

programme BioMATH (Promega, outils en ligne), qui calcule leur température de dissociation (T_m). Les amorces devaient avoir une T_m de 52 à 56 °C (tab. IX). Les amorces ont été commandées chez IDT. Lorsque reçues, les amorces ont été diluées par paire à une concentration de travail de 10uM chaque.

Tableau IX: Séquence et température de dissociation de chaque amorce spécifique au chromosome 3.

Nom	Séquence 5'→3'	Température de dissociation (°C)
Chr3vschr1_1_sens	aacatggctttatgtctgacaag	53,5
Chr3vschr1_1_anti	ggtcccagtagccaatatgg	55,0
Chr3vschr1_2_sens	gccagcctccagccttg	59,5
Chr3vschr1_2_anti	agggctccgtgtagggaaat	58,3
Chr3vschr1_3_sens	ggcacagggaacgccaca	60,7
Chr3vschr1_3_anti	cacctgtcactaagctgta	53,9

2.5.2.3 Optimisation des réactions de PCR

Afin de déterminer la température d'hybridation optimale pour obtenir un produit d'amplification spécifique de la région 3p21 et pour vérifier la compatibilité de ces amorces avec la méthode de génotypage de la plateforme Sequenom, les amorces ont été évaluées sur de l'ADN génomique humain, de même que sur de l'ADN issue de lignées cellulaires hybrides rongeur/humain contenant soit le chromosome 1 ou le chromosome 3 humain. L'ADN génomique humain provient de l'individu NA12707-A4 de la collection CEPH de HapMap. Les ADN de cellules hybrides rongeurs/humain contenant seulement un chromosome complet humain, soit le chromosome 3 (GM10253) ou le chromosome 1(GM13139), font partie du répertoire génétique de cellules humaines du *National Institute of General Medical Science* (NIGMS), échantillonnage 2, version 3. Ces génomes hybrides permettent de vérifier, en l'absence du chromosome 3, que les amorces spécifiques au chromosome 3 n'amplifient pas le chromosome 1. En plus des 3 paires d'amorces spécifiques au chromosome 3, une paire d'amorces spécifiques au chromosome 1 et une paire spécifique au chromosome 3 ont été utilisées comme contrôles à l'extérieur de la région dupliquée. Les 5 paires d'amorces utilisées sont

décrites au tableau IX, et les conditions d'amplification sont décrites à la figure 5 et au tableau X.

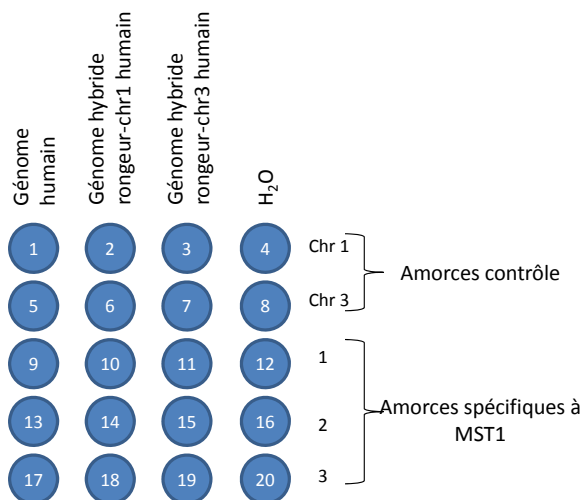


Figure 5: Plan des réactions de PCR de l'expérience d'optimisation de l'amplification spécifique du chromosome 3.

Tableau X: Composition de chaque réaction d'amplification (A) spécifique de la séquence du chromosome 3, et cycles de PCR (B).

Réactifs	Volume (ul) 1 réaction
H ₂ O	0,9
Tampon de PCR 10X	0,5
MgCl ₂ 25mM	0,4
dNTP 25mM ch	0,1
Mélange d'amorces 500nM ch	1
Hotstar Taq 5U/ul	0,1
ADN 2,5ng/ul	2
Total	5

A.

Nombre de cycles	Température et durée
1x	94°C 5min
35x	1)94°C 30sec 2)Gradient entre 50°C et 65°C 3)72°C 1min 30sec
1x	72°C 5min

B.

Dix ul de chaque produit d'amplification ont ensuite migrés sur un gel d'agarose 1,5 % (2,25g agarose + 147,75 ml TAE + 7,5 ul de bromure d'éthidium) pendant 30 minutes à 90V.

Afin de valider que ces amorces sont compatibles avec les méthodes de génotypage sur la plateforme Sequenom, une seconde expérience a été faite en multiplexant les trois paires d'amorces spécifiques au chromosome 3 aux températures optimales identifiées. Ces amorces ont ensuite été évaluées lors d'un test de génotypage comparatif avec les amorces d'amplification conçues de façon classique par les logiciels Sequenom. Les SNP génotypés sont de trois types; le type A sont des variantes qui ne représentent pas des différences entre les séquences du chromosome 3 et 1 après alignement, et qui sont fortement corrélés à des SNP sur le chromosome 3 à l'extérieur de la région dupliquée; les variantes de type B sont publiées dans les bases de données publiques mais qui représentent clairement des différences nucléotidiques entre les duplications et la séquence du chromosome 3 d'après l'alignement; finalement il y a enfin les variantes « test » identifiées lors d'expérience de séquençage du gène MST1 et dont on ne peut déterminer si ce sont de vraies variantes basé sur l'alignement (tab. XI). Un SNP situé en dehors de la région dupliquée est aussi utilisé comme contrôle positif de génotypage.

Tableau XI: Les 14 variantes choisies pour tester les amorces spécifiques au chromosome 3 sur la plateforme de génotypage.

A = vrai SNP B= différence entre chr1 et chr3.

No	Paire d'amorces	SNP	position	Type de variante
1	contrôle positif	rs9875617	49609700	A
2	Chr3vsChr1_1	rs3197999	49696536	A
3	Chr3vsChr1_1	rs1057319	49696565	B
4	Chr3vsChr1_1	3p49696610	49696610	TEST
5	Chr3vsChr1_1	3p49696626	49696626	TEST
6	Chr3vsChr1_1	rs2446543	49696759	TEST
7	Chr3vsChr1_1	rs13085791	49696785	A
8	Chr3vsChr1_1	3p49697003	49697003	B
9	Chr3vsChr1_2	rs6777426	49698885	B
10	Chr3vsChr1_2	3p49699143	49699143	B
11	Chr3vsChr1_2	rs41291698	49699176	TEST
12	Chr3vsChr1_2	3p49699201	49699201	TEST
13	Chr3vsChr1_3	3p49700308	49700308	B
14	Chr3vsChr1_3	3p49701074	49701074	TEST

Le génotypage a été fait sur 19 individus tirés de la collection CEPH (cohorte 4, annexe A), ainsi que sur les génomes provenant de lignées cellulaires hybrides rongeur/humain, ou de l'eau. Les amorces de PCR « classique » et les amorces d'élongation ont été conçues par le logiciel ASSAY DESIGN tel que décrit plus tôt. Les amorces classiques ont ensuite été diluées ensemble afin d'obtenir un mélange d'amorces avec une concentration de 500nM chacune. La même chose a été faite parallèlement pour les amorces spécifiques au chromosome 3. L'amplification par PCR a été faite avec les amorces spécifiques au chromosome 3 (voir tableau XII pour les cycles de PCR), et avec les amorces de PCR conçue de façon « classique » avec Assay Design. Le reste de l'expérience de génotypage a été fait comme décrit plus haut.

Tableau XII: Cycle de PCR pour le génotypage avec les amorces spécifiques au chromosome 3.

Nombre de cycles	Température et durée
1x	94°C 10min
45x	1)94°C 30sec 2)58°C 30sec 3)72°C 1min 30sec
1x	72°C 5min

Les graphiques de génotypage ont ensuite été extraits de la plateforme et ont été consultés visuellement afin de déterminer la spécificité des résultats de génotypage (Annexe D).

2.5.3 Échantillons d'ADN

Les ADN utilisés pour les tests d'association de variantes de la région chromosomique 3p21 proviennent de différentes cohortes (voir annexe A). Le diagnostic de la maladie de Crohn ou de la colite ulcéreuse a été basé sur des critères établis, suite aux analyses cliniques, radiologiques et endoscopiques, et de rapports histologiques. Une entrevue

avec le patient et une revue de son dossier médical ont été fait afin de compléter les données phénotypiques. Un consentement éclairé écrit de tous les participants, ainsi que l'approbation du comité d'éthique local ont été obtenus.

2.6 ANALYSES ET STATISTIQUES

2.6.1 Essais Fonctionnels

Lors de l'évaluation par PCR quantitatif de l'expression des gènes pro-inflammatoires *TNF- α* , *IL12p40*, des gènes codants pour des marqueurs de différenciation des monocytes en macrophages *CD-14* et *CD-206* et du gène codant pour le récepteur RON, *MST1R*, les différents traitements appliqués aux cellules ont été faits en triplicatas. De plus, chaque échantillon a été évalué en duplicata sur la plaque. À la suite d'une inspection des courbes d'amplification dans chaque puits à l'aide du logiciel MxPro (Stratagene 2007, Mx3005P v 4.10 build 389 schema 85), les puits montrant des courbes d'amplification irrégulières étaient éliminés. Le signal d'amplification de chaque puits a été corrigé sur le signal du marqueur de référence ROX (kit SYBR green, Invitrogen, 11733-046), et avec l'efficacité des courbes standards faites à partir d'une dilution sériée d'un des échantillons. L'expression des gènes d'intérêt dans chaque échantillon a été normalisée sur celle du gène de référence HPRT1 dans les mêmes échantillons, et l'expression a été calibrée par rapport à un échantillon sans traitement. Dû à la petite taille de l'échantillonnage, un test de Wilcoxon et un test de T de Welch ont été effectués à l'aide du programme Office Excel (Microsoft 2007, 12.0.6514.5000 SP2 MSO 12.0.6425.1000) afin de vérifier si les différences d'expression mesurées étaient significatives entre les différents traitements. Les barres d'erreur dans les graphiques représentent l'écart-type.

2.6.2 Étude d'association des variantes de *MST1* et *MST1R*

Nous avons testé pour association des SNP des gènes *MST1* et *MST1R* (tab. XIII) dans notre cohorte italienne de cas de CU et de témoins (cohorte 5, annexe A). Ces SNP proviennent de deux sources différentes :

- Des SNP identifiés par re-séquençage des gènes de *MST1* et *MST1R* chez des cas CU et des témoins. Les variantes ayant une fréquence de l'allèle mineur (MAF) >1%, un score élevé de vraisemblance tel que déterminé par le logarithme Syzygy 0.9 (Broad Institute 2009, Manuel A. Rivas et Mark J. Daily), et ne représentant pas des différences évidentes entre les séquences dupliquées sur le chromosome 1 et celle du chromosome 3 ont été testées.
- D'autres SNP provenant des bases de données publiques, tel que HapMap et dbSNP. Nous avons sélectionné les SNP ayant une MAF >1% chez les CEPH de la collection de génomes caucasiens du projet HapMap et ceux qui ne représentaient pas, après une inspection du fichier d'alignement, des différences entre les séquences dupliquées sur le chromosome 1 et celle du chromosome 3.

Tableau XIII: Autres variantes de *MST1* et *MST1R* génotypés dans nos cohortes.

No	Gene	SNP	Source
1	<i>MST1</i>	rs13085791	Banque de données dbSNP
2	<i>MST1</i>	rs1057319	Banque de données dbSNP
3	<i>MST1</i>	rs41291698	Banque de données dbSNP
4	<i>MST1</i>	3p49696610	Re-séquençage
5	<i>MST1R</i>	rs12489386	Banque de données dbSNP
6	<i>MST1R</i>	rs35986685	Banque de données dbSNP
7	<i>MST1R</i>	rs2230593	Banque de données dbSNP
8	<i>MST1R</i>	rs1062633	Banque de données dbSNP
9	<i>MST1R</i>	rs2230592	Banque de données dbSNP

En raison de leurs duplications et du fait que dbSNP contient de nombreux faux SNP (en vérité des différences entre la séquence du chromosome 3 et ses séquences dupliquées sur le chromosome 1), les séquences FASTA des SNP du gène MST1 n'ont pas été analysées par proxSNP ni par preXTEND, et ont été greffées au fichier de sortie final après l'analyse de preXTEND. Aussi, lors de l'amplification par PCR les températures d'hybridation ont été fixées à 58 °C plutôt qu'à 56 °C, et la durée de l'élongation a été fixée à 1 min 30 sec. Mis à part ces exceptions, le génotypage a été effectué comme décrit plus haut.

Seuls les individus ayant un taux de génotypage supérieur à 80% et les SNP ayant un taux de génotypage de 70% sont conservés pour l'analyse d'association subséquente. La différence de fréquence d'un allèle entre les cas et les témoins est testée par test de Khi-carré avec le logiciel d'analyse de données génétique Plink. L'équilibre de Hardy-Weinberg a été testé pour chaque SNP, et la fréquence de l'allèle mineure a également été calculée à l'aide de Plink.

2.6.3 Méta-analyse des signaux de 3p21

Afin de donner plus de puissance à notre étude pour détecter de faibles signaux d'association indépendants dans la région 3p21, nous avons fait la liste de tous les SNP ayant une valeur p supérieure à 1×10^{-4} dans l'étude d'association pangénomique de CU de McGovern. L'association au SNP rs3197999 a été considéré comme le signal principal venant de la région 3p21, et tous les SNP ayant une corrélation, à ce SNP, inférieure à 0.4 ont été gardé dans la liste. Le SNP le plus associé de cette liste est rs7629936 du gène CDHR4. Ensuite, les SNP ayant une corrélation inférieure à 0.4 à ce SNP ont été gardé : seul rs6802890, un autre SNP de CDHR4, est resté. Cela suggère qu'il y a potentiellement 3 signaux indépendants d'association dans la région 3p21. Ces SNP ont été génotypés chez tous les individus de la cohorte ITACC (cohorte 6, annexe A; 1110 témoins, 1647 cas de CU, 799 cas de MC) sur la plateforme Sequenom. Pour rs3197999, une partie des individus de la cohorte se superpose aux individus génotypés lors de l'identification de rs3197999 [66]. Les résultats d'association des trois SNP ont été

obtenus par régression logistique pour chaque sous-phénotype de la maladie (MC, CU et MII). Nous avons aussi obtenu les résultats post-contrôle de qualité (QC) d'association de 3 études d'association pangénomiques effectuées sur des cohortes indépendantes (voir cohortes 7, 8 et 9, annexe A). Pour ces études, seuls les individus ayant un taux de génotypage supérieur à 96% pour les SNP rs3197999, rs6802890 et rs7629936 ont été gardés pour les analyses subséquentes. Les résultats d'association des trois SNP provenant de nos analyses et de celles, de chaque étude d'association pangénomique (Sweden (341 témoins, 923 cas de CU), Cedars (2882 témoins, 925 cas de MC et 723 cas de CU), NIDDK (2415 témoins, 940 cas de MC et 1002 cas de CU)) ont été obtenus à l'aide du logiciel Plink pour chaque sous-phénotype de la maladie (MC, CU et MII). Des analyses d'association conditionnelle ont aussi été faites pour chaque SNP et chaque phénotype afin de tester l'indépendance des signaux d'association. Pour chaque cohorte, les composantes principales ont été utilisées comme covariable dans nos analyses afin de corriger les résultats sur une possible stratification de la population. Les résultats d'association et d'analyse conditionnelle des différentes cohortes ont tous été alignés afin de comparer le même brin d'ADN entre les études. Ensuite, pour chacun des SNP, une méta-analyse à effet fixe a été effectuée à l'aide du logiciel de méta-analyse de données génétiques Metal. La méta-analyse consiste à calculer la moyenne des effets de chaque SNP dans les études d'association et pour les analyses conditionnelles, pondérée sur l'inverse de la variance de chaque étude. Pour ce faire, un effet global β est calculé à

partir de la formule suivante : $\beta = \frac{\sum_i \beta_i \times w_i}{\sum_i w_i}$ où i représente les études d'association

pangénomiques, et $w_i = \frac{1}{\text{var}(\beta_i)}$. La variance est calculée à partir de l'écart-type (SE)

de l'effet : $\text{var}(\beta_i) = \text{SE}(\beta_i)^2$. Une cote Z est ensuite calculée à partir de l'effet

global β : $Z = \frac{\beta}{\text{SE}(\beta)}$, puis une valeur p est déduite à partir de la cote Z. Ce sont ces

valeurs p qui sont rapportées dans le tableau de résultat. Le rapport de cote (OR, odds ratio), aussi dans ce tableau, a été déduit à partir de la formule $OR = e^{\beta}$.

CHAPITRE 3

RÉSULTATS

3.1 ÉTUDES D'ASSOCIATION

3.1.1 Recherche de nouvelles variantes codantes non-synonymes de MST1 et MST1R associées aux MII

C'est par une approche par gène candidat que notre groupe a identifié dans la région 3p21 le SNP codant non-synonyme rs3197999 (R689C) du gène *MST1* associé aux MII [66]. Une autre étude a identifié aussi dans la même région un SNP corrélé à des variantes codantes non-synonymes du gène *MST1R* et associé à la MC, mais qui ne montre qu'une faible corrélation au SNP rs3197999 [53]. Fonctionnellement, *MST1* et *MST1R* semblent être de bons gènes candidats pour les MII. Nous avons émis l'hypothèse que d'autres variantes de ces gènes pourraient être associées aux MII. Nous avons donc testé l'association de variantes dans ces gènes, provenant soit des bases de données publiques, soit d'expérience de re-séquençage de ces gènes dans nos cohortes de patients MII et de témoins.

3.1.1.1 Étude de la structure génomique de 3p21

3.1.1.1.1 Alignement de la région 3p21

Lors des premiers génotypages que nous avons effectué dans un sous-groupe de nos cohortes (cohortes 1, 2 et 3, voir annexe A), le génotypage de tous les SNP testés de *MST1* a échoué contre seulement 17% des SNP localisé dans *MST1R*. Afin de vérifier si la séquence de la région 3p21 autour du gène *MST1* avait une particularité, nous avons aligné la région 3p21 complète contre le génome de référence.

Nos résultats d'alignement montrent que deux grandes régions localisées sur le chromosome 1 ont une très grande identité avec une partie de la séquence de la région 3p21 englobant le gène *MST1*. Cette observation confirme les résultats de 2 études antérieures qui démontraient l'existence de duplicata de la région chromosomique 3p21, entourant le gène *MST1*, dans la région 1p36 du chromosome 1 [95, 96]. La première région dupliquée couvre les bases de la position 49 696 425 à 49 709 548 (~13 Kb) (NCBI36) du chromosome 3 et montre une identité de 96% à la séquence nucléotidique d'origine, tandis que la seconde, plus courte, allant de la position 49 696 425 à 49 701 085 (~ 4.6 Kb), partage 97% d'identité avec cette dernière. Les séquences nucléotidiques alignées sont montrées dans l'annexe B. Presque la totalité de la séquence du gène *MST1* est située dans la région dupliquée et seules 15 paires de bases dans la région non traduite (UTR) 3' de *MST1* ne sont pas dupliquées. La figure 6 donne un aperçu de l'étendue des duplications, et montre que la plus grande de ces duplications englobe aussi près de la moitié de la séquence du gène voisin, *RNF-123*. Les promoteurs de *MST1* et *RNF-123* sont également inclus dans la plus grande des duplications. Les séquences des duplications divergent sensiblement de la séquence de la région 3p21 montrant des substitutions, délétions et insertions de un ou plusieurs nucléotides. La plus grande des 2 duplications contient aussi plusieurs délétions de plus de 5 nucléotides de long (Figure 6).

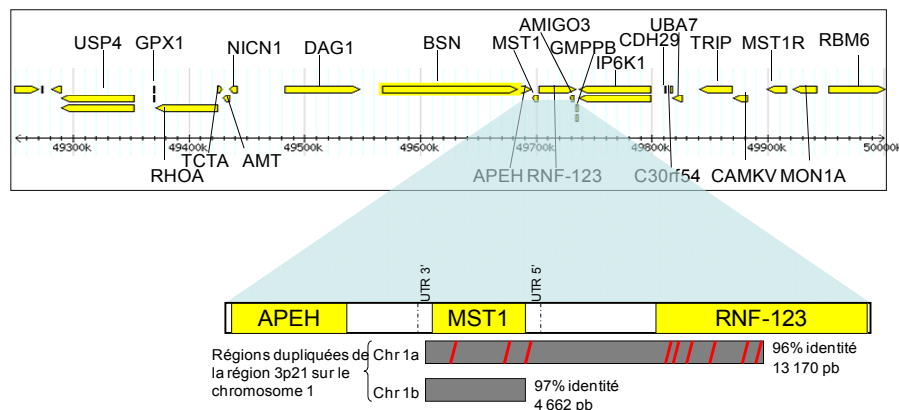


Figure 6: Plan des réactions de PCR de l'expérience d'optimisation de l'amplification spécifique du chromosome 3.

L'encadré montre le segment 49 250 kb-50 000 kb de la région 3p, tiré de HapMap d'après la construction NCBI36. Les blocs gris correspondent aux duplications sur le chromosome 1 ; Chr1a = position 16 925 192 - 16 965 362 et Chr1b = position 17 076 939 - 17 188 516, d'après la construction NCBI36. Les bandes rouges représentent les délétions > 5 pb dans les séquences des duplications.

3.1.1.1.2 Amplification spécifique de la séquence de *MST1*

Comme la plupart des variantes de *MST1* que nous voulons évaluer dans notre étude d'association sont situées dans cette région dupliquée et que les amorces conçues pour amplifier la séquence de *MST1* peuvent aussi amplifier la séquence des duplications sur le chromosome 1 il est clair que cette situation explique la faible qualité du génotypage que nous avons obtenu dans cette région. Plus précisément, lors de l'amplification et du génotypage des variantes dans la région, l'amplification en parallèle des séquences du chromosome 1 avec celle du chromosome 3 cause une sur-représentation de certains "allèles" retrouvés sur le chromosome 1 et un déplacement des groupes dans les graphiques d'agrégation rendant difficile la détermination des génotypes par les logiciels de la plate-forme de génotypage. De plus, plusieurs variantes publiées dans les bases des données publiques de SNP pour le gène *MST1* semblent en fait n'être que des différences nucléotidiques entre la séquence de *MST1* sur le chromosome 3 et celles des duplications

sur le chromosome 1. À l'aide de l'alignement des séquences, ces différences entre les chromosomes ont pu être discernées des véritables SNP de *MST1*. Ainsi, des 36 SNP de *MST1* disponibles dans la banque de données dbSNP (construction NCBI36) au moment de cette expérience, 6 étaient clairement des différences entre les duplications, et 21 étaient des différences entre le chromosome 1 et le chromosome 3. Aussi, 13 des variantes génotypées dans nos premiers essais étaient en fait des différences entre le chromosome 3 de référence et les duplications, tandis que 7 des 30 des variantes génotypés étaient en fait des variantes retrouvées sur seulement une des 2 duplications, donc des différences entre les duplications et la séquence du chromosome 3. Il a donc fallu concevoir des amorces spécifiques à la séquence de *MST1* sur le chromosome 3 afin de permettre le génotypage des SNP dans cette région. En prenant avantage des divergences entre les séquences dupliquées et d'origine, nous avons conçu 3 paires d'amorces spécifiques couvrant les régions contenant la plus grande densité de variantes de *MST1* à tester. Chacune des trois paires d'amorces spécifiques au chromosome 3 a été validée par PCR sur différents échantillons contrôle, soit de l'ADN génomique humain, soit de l'ADN provenant de génome hybride de rongeur et contenant soit le chromosome 1 ou 3 humain. De plus, des amorces spécifique à des séquences du chromosome 1 et 3 mais à l'extérieur des régions concernées par les duplications ont été utilisées comme contrôles positifs pour évaluer les amplifications spécifiques. Afin d'identifier la température d'hybridation qui permet une hybridation optimale de ces amorces à la séquence de 3p21 et qui limite l'hybridation sur les séquences dupliquées, chaque réaction a été évaluée dans une série de températures d'hybridation différentes lors du PCR (fig. 7). Les trois paires d'amorces du chromosome 3 semblent offrir une amplification spécifique à des températures entre 55-60°C puisque qu'on voit y voit une forte amplification en présence du chromosome 3 et qu'on ne voit que de faibles produits d'amplification non-spécifiques à des températures d'hybridation <55°C en présence du chromosome 1. De plus, aux mêmes températures d'hybridation, les amorces contrôles qui sont spécifiques aux chromosomes 1 et 3 semblent donner un seul produit d'amplification dans les lignées hybride contenant le chromosome 1 et 3 respectivement.

Puisque ces amorces semblent amplifier de façon spécifique les séquences du chromosome 3, elles ont été utilisées dans une expérience de génotypage comparative à petite échelle sur une sélection de variantes de *MST1*.

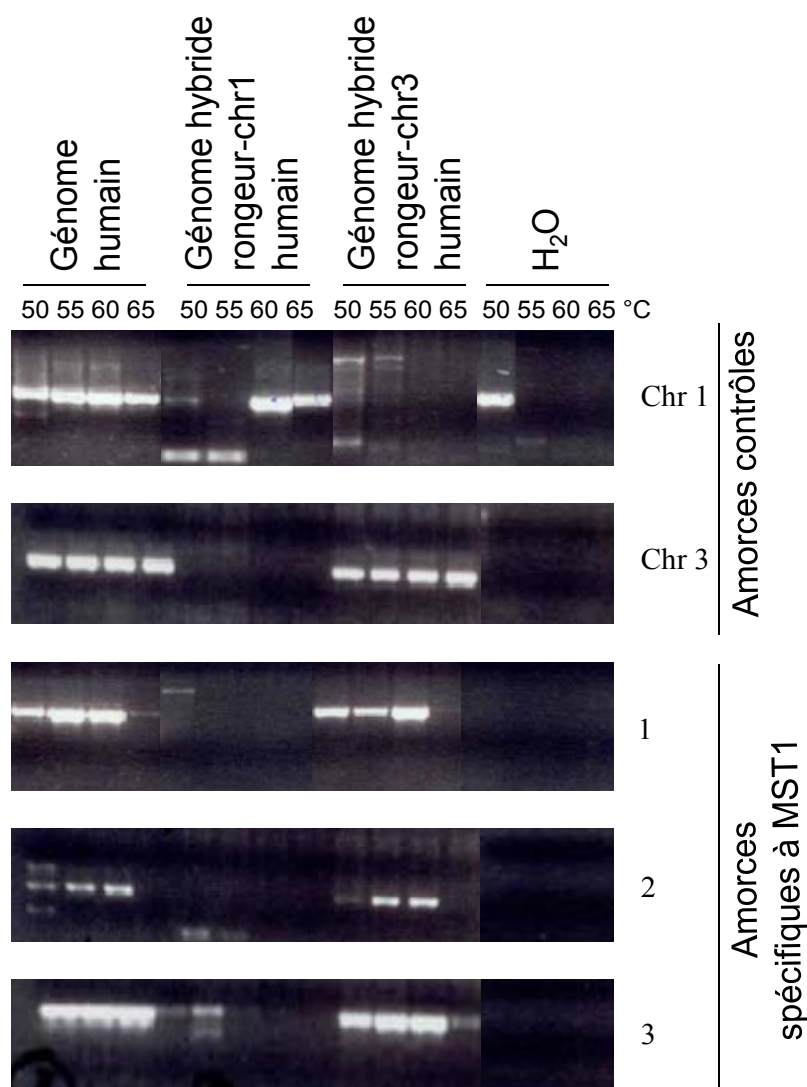


Figure 7: Électrophorèse de produits d'amplification de génomes humains ou hybrides rongeur-humain chr1 ou chr3. Les génomes ont été amplifiés à l'aide de paires d'amorces contrôles s'hybridant à l'extérieur des séquences dupliquées, ou avec les paires d'amorces conçues pour être spécifiques au chromosome 3. Les amplifications ont été faites à des températures d'hybridation de 50, 55, 60 ou 65°C.

Afin d'évaluer la spécificité de l'amplification de la région entourant *MSTI* lors des génotypages sur la plateforme Sequenom, des variantes étant clairement des différences nucléotidiques entre les séquences dupliquées ont été génotypées chez 19 individus de la cohorte de témoins sains de la collection CEPH (annexe A) avec soit les amorces spécifiques à la région de *MSTI*, soit des amorces conçues de façon classique avec le logiciel Assay Design. D'autres variantes ont également été testées sur ces mêmes échantillons avec ces amorces afin de déterminer si elles étaient véritablement des variantes de *MSTI*. Les graphiques d'agrégation de génotypes, présenté en annexe D, montre les résultats du génotypage pour chacune des 14 variantes testées chez les individus CEPH.

Le SNP contrôle rs9875617, située à l'extérieur des séquences dupliquées, montre que la réaction d'amplification et le génotypage avec les amorces standards ont bien fonctionnés pour cette région. Les variantes qui représentent des différences nucléotidiques entre les séquences dupliquées; 3p49700308, 3p49699143, rs6777426, 3p49697003 et rs1057319, montrent une tendance à être plus monomorphiques avec les amorces spécifiques qu'avec les amorces classiques, puisqu'en amplifiant plus spécifiquement la séquence d'intérêt, ces amorces permettent d'éliminer la sur-représentation de certains allèles causée par l'amplification des séquences des duplications. Pour les SNP reconnus, comme rs13085791 et rs3197999, on peut voir une nette amélioration de la détermination des génotypes avec les amorces spécifiques aux chromosomes 3 comparativement aux amorces classiques. Enfin, des 6 variantes dont le type restait à déterminer, 3 (3p49696626, 3p49699201 et 3p49696610) semblent monomorphiques chez nos 19 individus CEPH, alors que pour les 3 dernières (rs41291698, 3p49701074 et rs2446543), la mauvaise qualité de l'amplification n'a pas permis de déterminer leur génotype.

Puisque ces amorces permettent dans certains cas d'améliorer le génotypage des SNP situés dans la région dupliquée, elles ont été utilisées dans une expérience subséquente de génotypage de variantes de *MSTI* dans une cohorte de taille plus importante.

3.1.1.2 Test d'association des variantes de *MST1* et *MST1R*

Nous avons testé l'association, dans notre cohorte de CU, des variantes de *MST1*, ainsi que de variantes de *MST1R* à l'aide des amorces spécifiques au chromosome 3. Plus spécifiquement, les SNP sélectionnés à partir des bases de données publiques et d'expériences de séquençage de *MST1* et *MST1R*, et étant identifiés clairement dans notre alignement comme n'étant pas des différences entre les duplications et le chromosome 3 ont été sélectionnés (pour la liste des variantes génotypées, voir matériel et méthode, tableau XIII).

Des 9 SNP à tester, deux ont été rejetés lors de la conception des groupes d'amorces. Sept SNP ont été génotypés, et après le contrôle de qualité (SNP ayant un taux de génotypage supérieur à 80%, et les individus ayant un taux de génotypage de 90%), 5 SNP ont été conservés. Ces faibles taux (comparativement à des seuils de plus de 95% normalement rapportés dans les études d'association) ont été conservés afin de permettre l'analyse des variants dont le génotypage est compliqué par l'existence des duplications dans leur locus. L'analyse d'association a été faite avec 1310 cas de CU et 756 contrôles (cohorte 5, annexe A), avec un taux de génotypage général de 96.5%. Tous les SNP étaient en équilibre de Hardy-Weinberg. Les résultats du test d'association sont montrés dans le tableau XIV. Des SNP de la région 3p21 testés dans la cohorte ITACC, un SNP montre un signal d'association significatif (rs13085791, valeur-p de 0,0008223, la valeur-p de seuil avec correction de Bonferroni pour le nombre de tests = 0.00625). Ce SNP, situé dans le gène *MST1*, est parfaitement corrélé ($r^2=1$) avec R689C d'après les données de LD. Aucune des variantes de *MST1R* testés n'a donné de signal d'association significatif. Ceci suggère que le signal d'association dans le gène *MST1* provient de la variante codante non-synonyme R689C, et que nous n'avons pas pu répliquer le signal d'association corrélé à *MST1R*. Ceci peut être dû au fait que notre cohorte ne comporte que des cas de CU et que ce signal est exclusif à la MC, ou que notre cohorte n'a pas la puissance de détecter ce signal d'association pour la CU si ce dernier est dû à une variante ayant un faible effet.

Tableau XIV: Résultats d'association de variantes de MST1 et MST1R dans notre cohorte italienne.

(1647 cas, 1110 contrôles). La valeur p est bilatérale car l'effet des allèles des variantes testées est inconnu (protecteur ou à risque).

SNP	Position	Allèle mineur	Fréquence cas	Fréquence témoins	OR	95%CI	valeur-p	Gène
rs13085791	Chr3:49696802	A	0,319	0,269	1,272	1.104-1.464	0,0008223	MST1
rs1062633	Chr3:49899944	G	0,456	0,485	0,8888	0.781-1.011	0,07283	MST1R
rs12489386	Chr3:49902507	T	0,056	0,048	1,17	0.877-1.561	0,2847	MST1R
rs35986685	Chr3:49910530	G	0,013	0,010	1,347	0.714-2.54	0,3562	MST1R
rs2230593	Chr3:49915082	A	0,055	0,058	0,9353	0.712-1.23	0,6318	MST1R

3.1.2 Test d'indépendance des signaux d'association de la région 3p21

La région 3p21 a été plusieurs fois associée aux MII dans les études d'association précédentes, et nous avons pu répliquer le signal d'association à R689C dans notre test d'association précédent, mais pas celui corrélé à *MST1R*. Pour vérifier si la région 3p21 contient plus d'un signal d'association à la CU mais aussi à la MC, et provenant de variantes ayant un faible effet, nous avons dû sélectionner des variantes représentant potentiellement des facteurs de risque aux MII, et les génotyper dans une cohorte de grande taille comprenant des cas de MC et de CU. Les SNP génotypés dans une méta-analyse d'études d'association de CU [35], pour laquelle nous avons les résultats d'association et dont les valeurs-p d'association étaient sous le seuil de 1×10^{-4} , ont été classés en ordre décroissant de valeur-p. Les SNP corrélés à $r^2 > 0.4$ au SNP rs3197999, considérés comme faisant partie du même signal d'association, ont été retirés de ce groupe, ce qui a permis d'identifier rs7629936, le prochain SNP le plus associé et seulement faiblement corrélé au signal principal de rs3197999. Ensuite, les SNP corrélés à $r^2 > 0.4$ à rs7629963 ont été retirés de la liste, ce qui a permis d'identifier rs6802890. Cela suggère, basé sur les données de corrélation des individus caucasiens disponibles dans HapMap, que la région 3p21 contient 3 signaux d'association indépendants potentiels (tableau XV).

Tableau XV: Résultats d'association des trois signaux indépendants dans la région 3p21 à partir des données des méta-analyses précédentes. Les valeurs ont été tirées des résultats de McGovern et al. [35].

SNP	position	gène	valeur-p CU	valeur-p MC
rs3197999	chr3:49696536	MST1	1.36E-06	2.17E-07
rs6802890	chr3:49807265	intergénique, CDHR4	2.93E-06	2.90E-04
rs7629936	chr3:49811711	CDHR4	2.38E-06	5.84E-06

3.1.3 Test d'association des trois SNP étant potentiellement des signaux d'association indépendants à la MII

Afin de tester l'association des signaux observés dans la région 3p21, nous avons génotypé les SNP rs6802890, rs7629936 et rs3197999 dans différentes cohortes CU et MC (cohorte 6, annexe A), car ces signaux semblent être généralisé aux MII. Bien que nous avons déjà génotypé le SNP rs3197999 sur une partie des individus de la cohorte ITACC dans notre article original [66], nous avons génotypé les SNP rs319799, rs6802890 et rs7629936 pour cette étude dans notre cohorte ITACC entière composé des cas de CU et de MC. Un test d'association par régression a ensuite été fait à partir des résultats de génotypage après le contrôle de qualité (post-QC). De plus, en combinant ces résultats avec les résultats d'association de ces mêmes SNP dans une partie des cohortes retrouvées dans les méta-analyses de MC et de CU [33, 35], nous avons obtenu plus de puissance pour détecter une association. L'un de ces SNP, rs6802890, est très fortement corrélé (HapMap CEU $r^2 > 0.8$) au SNP rs1128535, corrélé lui-même à des variantes codantes non-synonymes de *MST1R* [53]. La fréquence de l'allèle mineur de chacun des SNP a également été calculée dans toutes nos cohortes (annexe E), et nous observons que les fréquences entre les cohortes sont semblables pour les trois SNP. Cependant, l'allèle mineur de rs3197999 (T), semble légèrement plus fréquent dans notre sous-groupe juif (MAF= 41.48%) que dans les autres groupes (MAF=28.49-34.73%). Aussi, la fréquence

de l'allèle mineur de rs7629936 (A) dans le sous-groupe juif (MAF= 18.42%) est légèrement inférieure à la fréquence des autres groupes (MAF=20.59-24.26%).

Les résultats dans le tableau XVI A montrent que les trois SNP sont associés aux MII en général, avec des valeurs-p dépassant de seuil significatif pour une étude d'association pangénomique. Les résultats d'association pour MC et CU montrent que ces signaux d'association semblent être généralisés aux MII. Le SNP rs397999 montre le signal d'association le plus fort, avec une valeur-p de 1.52×10^{-11} pour MII, 8.56×10^{-8} pour MC et 4.70×10^{-9} pour CU. L'effet de l'allèle A de rs7629936 semble être protecteur ($OR_{MII} = 0.841$ (0.789-0.895)), alors que les effets des allèles A de rs6802890 et T de rs3197999 semblent plutôt conférer un risque (OR_{MII} respectivement = 1.159 (1.100-1.221) et 1.208 (1.143-1.275)). Les effets de ces allèles sont également conservés dans les résultats d'analyses avec MC et CU.

3.1.4 Test d'association conditionnelle des trois SNP étant potentiellement des signaux d'association indépendants à la MII

Afin de tester l'indépendance de nos trois signaux d'association, nous avons ensuite fait un test d'association conditionnel à partir de nos résultats de génotypage. Nous avons combinés ces résultats dans une méta-analyse avec les résultats d'association conditionnels de ces SNP fait à partir des résultats des trois études d'association auxquelles nous avons accès. Les résultats d'analyse d'association conditionnelle (tab. XVI B) montrent que le signal d'association principal, à rs3197999, contribue significativement au signal d'association généralisé aux MII détecté pour rs6802890, et que les signaux d'association aux MII et à la CU des SNP rs7629936 et rs3197999 sont réciproquement indépendants. Nous pouvons donc conclure qu'il existe deux signaux d'association indépendants à la CU dans la région 3p21, le premier englobant les SNP rs3197999 et rs6802890, et le second le SNP rs7629936. Pour la MC, le signal semble provenir de rs3197999 seulement, mais comme la cohorte d'individus MC est beaucoup moins importante de celle des individus CU, il est possible que notre étude n'ait pas

obtenu assez de puissance pour détecter l'indépendance du signal d'association provenant de rs7629936.

Afin de vérifier l'hétérogénéité entre les cohortes, un test d'hétérogénéité par Khi-carré été fait. Aucune hétérogénéité n'a été détectée.

Tableau XVI: Résultats de la méta-analyse des tests d'association par régression logistique et des analyses d'association conditionnelles pour MII, CU et MC.
La méta-analyse réunit les résultats des études d'association pour les SNP rs3197999, rs7629936 et rs6802890, génotypés dans les cohortes 6, 7, 8 et 9 présentées dans l'annexe A. Les valeurs p sont bilatérales. Les valeurs p en gras sont significatives (valeur-p de seuil avec correction de Bonferroni = 0.0167).

	SNP	Allèle	OR	95%CI	valeur-p	HetKhi2	Hetval-p
MII	rs3197999	T	1,208	1,143-1,276	1,52E-11	3,531	0,3168
	rs6802890	A	1,159	1,100-1,221	2,78E-08	0,893	0,8271
	rs7629936	A	0,841	0,789-0,895	7,00E-08	1,121	0,772
CU	rs3197999	T	1,215	1,139-1,297	4,70E-09	2,174	0,5371
	rs6802890	A	1,170	1,100-1,246	7,48E-07	3,361	0,3392
	rs7629936	A	0,834	0,773-0,899	2,32E-06	0,796	0,8504
MC	rs3197999	T	1,215	1,132-1,305	8,56E-08	3,732	0,1547
	rs6802890	A	1,153	1,077-1,235	4,24E-05	0,735	0,6924
	rs7629936	A	0,852	0,783-0,927	1,96E-04	0,564	0,7543

A.

	SNP	Condition	OR	95%CI	valeur-p	HetKhi2	Hetval-p
MII	rs6802890	rs3197999	1,056	0.987-1.130	0,1134	5,121	0,1631
	rs7629936		0,898	0.839-0.961	0,001859	2,945	0,4002
	rs3197999	rs6802890	1,164	1.084-1.250	3,013E-05	8,089	0,04421
	rs7629936		0,899	0.832-0.971	0,006688	0,775	0,8555
	rs3197999	rs7629936	1,166	1,099-1,237	3,598E-07	5,883	0,1175
	rs6802890		1,100	1,032-1,172	0,003401	0,527	0,913
CU	rs6802890	rs3197999	1,069	0,986-1,158	0,1061	7,765	0,05113
	rs7629936		0,891	0,822-0,967	0,005429	1,682	0,641
	rs3197999	rs6802890	1,160	1,066-1,263	0,0005477	6,953	0,07342
	rs7629936		0,898	0,819-0,985	0,02265	0,739	0,8641
	rs3197999	rs7629936	1,171	1,092-1,256	9,777E-06	3,461	0,3258
	rs6802890		1,111	1,029-1,198	0,006752	3,232	0,3572
MC	rs6802890	rs3197999	1,041	0,951-1,138	0,3838	0,764	0,6826
	rs7629936		0,915	0,836-1,003	0,0575	2,161	0,3394
	rs3197999	rs6802890	1,187	1,081-1,304	0,0003313	3,525	0,1716
	rs7629936		0,912	0,822-1,011	0,07879	1,958	0,3757
	rs3197999	rs7629936	1,180	1,092-1,274	2,523E-05	5,753	0,05632
	rs6802890		1,102	1,013-1,197	0,02309	2,253	0,3242

B. Nous avons également calculé le déséquilibre de liaison entre les trois SNP pour chacune de nos cohortes ainsi que pour les individus caucasiens de HapMap (fig. 8 A). Ce LD est rapporté en r^2 . Nous avons aussi défini des régions de 3p21 comprenant des variantes fortement corrélées ($r^2 > 0.8$) à chacun des SNP indépendants (fig. 8B) à l'aide des données génétiques de la population caucasienne CEPH, tirées du projet HapMap. Les trois SNP sont plutôt en faible corrélation ($r^2 < 0.45$) entre eux. MST1R, le gène codant pour le récepteur de MSP, est situé dans la région en forte corrélation avec le SNP rs6802890.

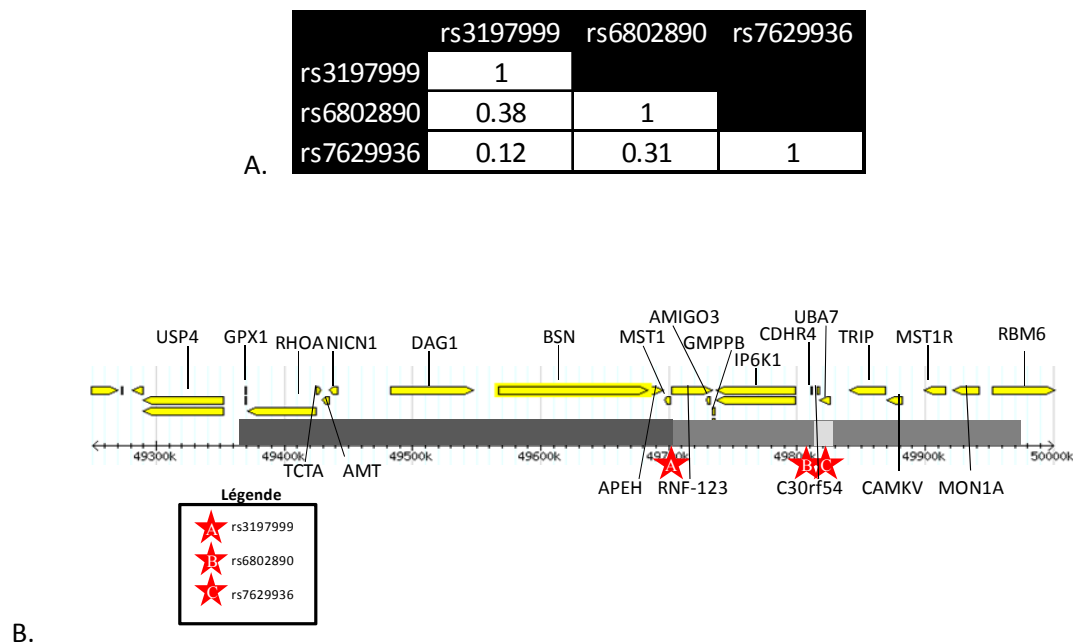


Figure 8: Régions en fort LD avec les trois signaux d'association de la région 3p21
La corrélation (r^2) entre les SNP, tirée des données des CEU de HapMap
(construction NCBI36). B : Régions en fort LD (HapMap CEU $r^2 > 0.8$) avec les trois
SNP. Les régions sont représentées par des blocs gris.

Bien que le LD entre les trois SNP testés soient faibles, les résultats de l'analyse d'association conditionnelle nous montre que le signal de rs6802890 serait dérivé des signaux d'association avec les SNP rs3197999 et rs7629936. Ainsi, la région 3p21 semble avoir 2 signaux d'association indépendants aux MII provenant de ces derniers.

3.2 ÉTUDES FONCTIONNELLES

Le signal d'association aux MII le plus fort, provenant de la variante codante non-synonyme R689C, est situé dans le gène *MST1*. Ce gène est un excellent candidat pouvant avoir un rôle à jouer dans l'étiologie des MII. En effet, plusieurs études ont démontré chez les macrophages murins que la MSP avait un rôle anti-inflammatoire.

Entre autre, il a été démontré que la production d'oxyde nitrique (NO) et de l'enzyme *induced nitric oxide synthase* (iNOS) induite par LPS est inhibée chez les macrophages de souris traités au MSP [85, 87], et cela en inhibant l'activité de NF-kB [86]. Un prétraitement minimum de 2 hrs avec la MSP supprimerait également l'expression de gènes de la réponse inflammatoire, tels que *IL12B*, induite chez les macrophages péritonéaux murins stimulés au LPS et IFN- γ [88]. Une étude a aussi démontré qu'un prétraitement avec MSP augmentait la production de O_2^- , normalement inhibée par NO, chez les macrophages humains de différents types. Dans cette même étude, le récepteur de MSP, RON, a également été faiblement détecté chez les macrophages humains dérivés de monocytes du sang [97]. Cependant, la mesure directe de la production de NO est difficile à faire chez les macrophages humains car leur production est beaucoup plus faible que chez les souris [98]. Nous avons donc décidé de tester l'effet d'un traitement à la MSP sur des modèles de macrophages et de cellules épithéliales humains en mesurant l'expression de gènes de la réponse inflammatoire, la migration des cellules lors de la cicatrisation et la phosphorylation de AKT dans des modèles expérimentaux en réponse à différents stimuli. Nous avons aussi décidé de tester l'effet de traitement à la MSP sur des cellules humaines, basé sur des modèles murins, afin de valider le rôle de la MSP chez l'Homme.

3.2.1 *Inhibition de l'expression de gènes pro-inflammatoires chez les macrophages prétraités à la MSP*

Nous avons isolé des monocytes du sang de volontaires sains et les avons différenciés en macrophages. En premier lieu, nous avons mesuré l'expression de marqueurs de la différenciation des monocytes en macrophages afin de valider la différenciation des cellules. Ensuite, une partie de ces cellules ont été soumise à un prétraitement avec MSP pendant 4 heures avant une stimulation de 3h au LPS. L'expression de gènes pro-inflammatoires comme *IL12B* et *TNF-a* a été mesurée puisque ces gènes possèdent des promoteurs reconnus par NF-kB, qui est activé par LPS.

La figure 9.A montre les résultats de l'expression des ARNm de *CD-14*, un marqueur que l'on retrouve principalement chez les monocytes non-différenciés [99]. Nos cellules différenciées pendant 7 jours en culture expriment significativement moins de *CD-14* ($p_{\text{Welch}}=0.000356$) que les monocytes, suggérant que ces cellules sont différenciées. La figure 9.B montre que l'expression de *CD-206*, un marqueur de la différenciation des macrophages, chez ces cellules est significativement plus élevée chez les cellules cultivées pendant 7 jours ($p_{\text{Welch}}=0.00140$) que chez les monocytes, ce qui confirme que ces cellules expriment un phénotype de monocyte différencié en macrophage.

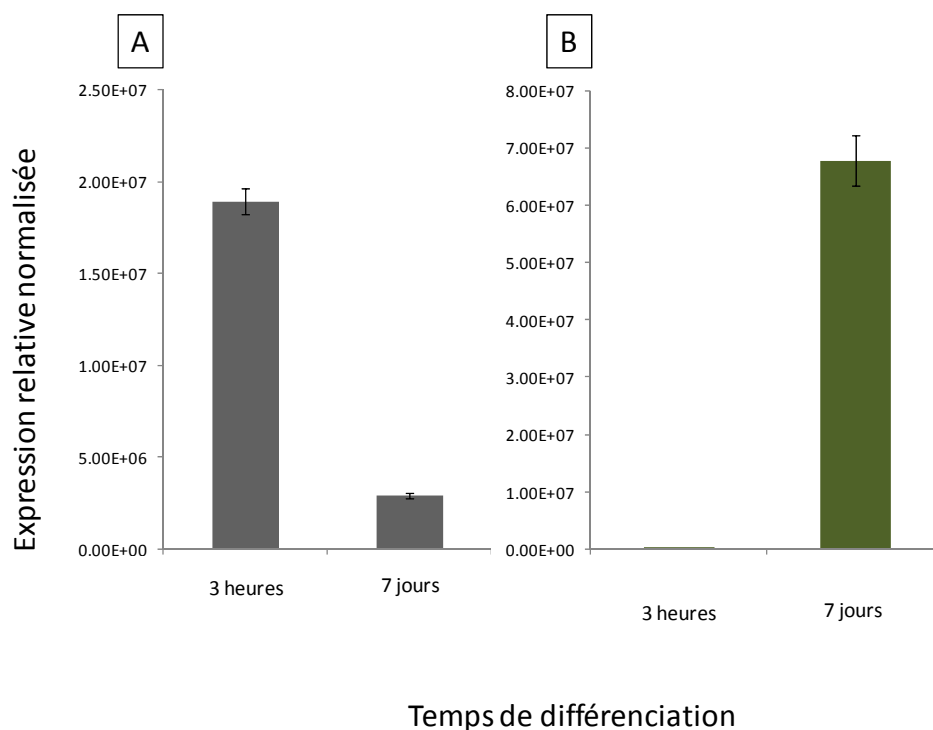


Figure 9: Expression relative de marqueurs de différenciation chez les monocytes du sang et les macrophages qui en sont dérivés.

A) Expression relative de *CD-14* chez les monocytes du sang et les macrophages qui en sont dérivés après 7 jours de culture. Barres = déviation standard . N=3. Le gène de référence est *HPRT1*. B) Expression relative de *CD-206* chez les monocytes du sang et les macrophages qui en sont dérivés après 7 jours de culture. Barres = déviation standard . N=3. Le gène de référence est *HPRT1*.

La figure 10 montre l'impact de la MSP sur l'expression de *TNF- α* et de *IL12B* en réponse à un stimulus pro-inflammatoire (A et B, respectivement). D'abord la différence d'expression entre les cellules stimulées au LPS et celles sans traitement est significative (*TNF- α* $p_{\text{Welch}}=0.00122$, *IL12B* $p_{\text{Welch}}=0.000156$). Le LPS induit donc une forte réponse inflammatoire chez nos macrophages dérivés des monocytes du sang. Cependant, la différence entre les macrophages stimulés au LPS prétraités ou non avec MSP n'est pas significative (*TNF- α* $p_{\text{Welch}}=0.427$, *IL12B* $p_{\text{Welch}}=0.128$). Ces résultats démontrent que MSP n'a pas un effet inhibiteur significatif sur l'expression de *TNF- α* ou *IL12B* induite par LPS dans ce modèle. Afin d'évaluer l'effet de MSP sur un modèle différent, nous avons ensuite évalué l'effet de MSP sur des macrophages dérivés de cellules et avons obtenu des résultats similaires à ceux obtenus avec les macrophages dérivés de monocytes du sang (résultats non présentés).

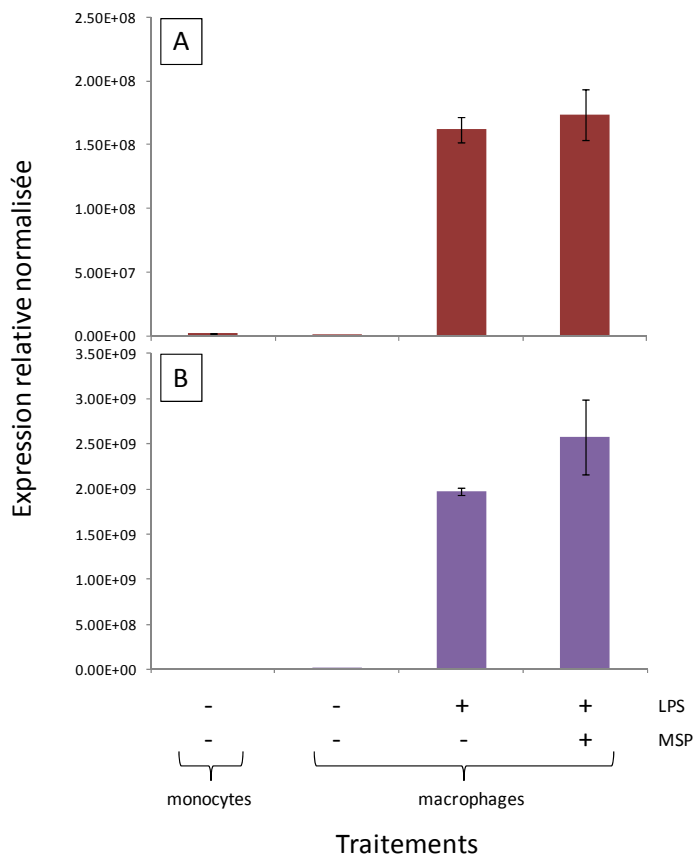


Figure 10: Expression relative de TNF-a et IL12p40 chez les macrophages prétraités au MSP et stimulés au LPS

A) Expression relative de *TNF-a* chez les macrophages prétraités au MSP et stimulés 3 heures au LPS. Barres = déviation standard. N=3. Le gène de référence est *HPRT1*. B) Expression relative de *IL12B* (IL12p40) chez les macrophages prétraités au MSP et stimulés 3 heures au LPS. Barres = déviation standard. N=3. Le gène de référence est *HPRT1*.

Devant ces résultats, nous avons décidé de vérifier l'expression des ARNm de *MSTIR* par PCR quantitatif dans ces deux modèles cellulaires de macrophages humains. Les tests préliminaires n'ont démontré aucune expression significative de *MSTIR* chez les macrophages dérivés des monocytes du sang ou dérivés des THP1 (résultats non présentés).

3.2.2 Expression de *MST1R* dans différentes lignées cellulaires.

Afin de trouver des lignées cellulaires exprimant le récepteur RON, encodé par *MST1R*, et donc susceptibles de répondre à un traitement avec la MSP, nous avons sélectionné, d'après les résultats des bases de données d'expression publiques, des lignées cellulaires qui expriment potentiellement RON. L'expression des ARNm de *MST1R* a été mesurée dans ces cellules par PCR quantitatif (fig. 11). Comparativement aux monocytes (ici représentés par les THP1 non différenciés), qui n'expriment pas l'ARNm de *MST1R* [97], les lignées cellulaires épithéliales de l'intestin HT-29, HCT-15 et HCT-116 expriment de très hauts niveaux de cet ARNm. Les HL-60 et les HEK293 sont utilisées aussi comme contrôles négatifs. Dans une expérience indépendante évaluant le niveau d'expression de *MST1R* chez les macrophages dérivés des monocytes sang les niveaux des ARNm de ce gène ont démontré un niveau d'expression similaire à celui des cellules THP-1 (résultats non présentés).

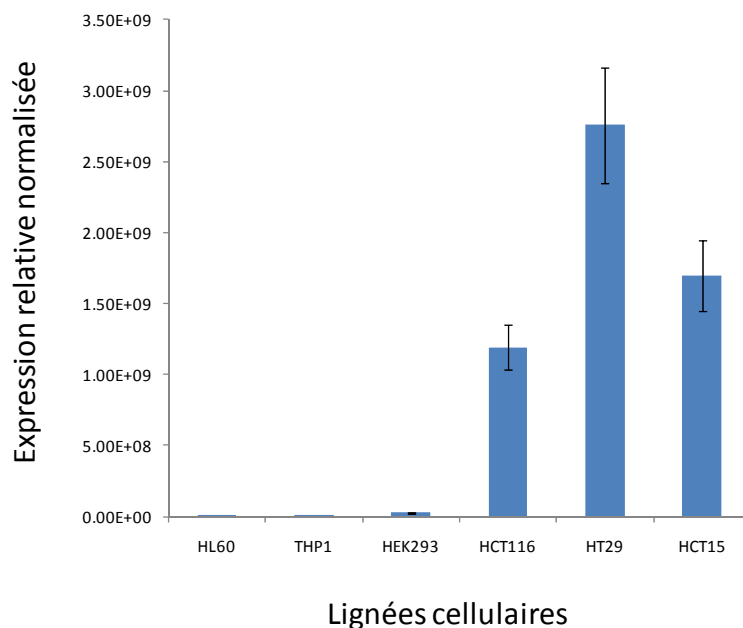
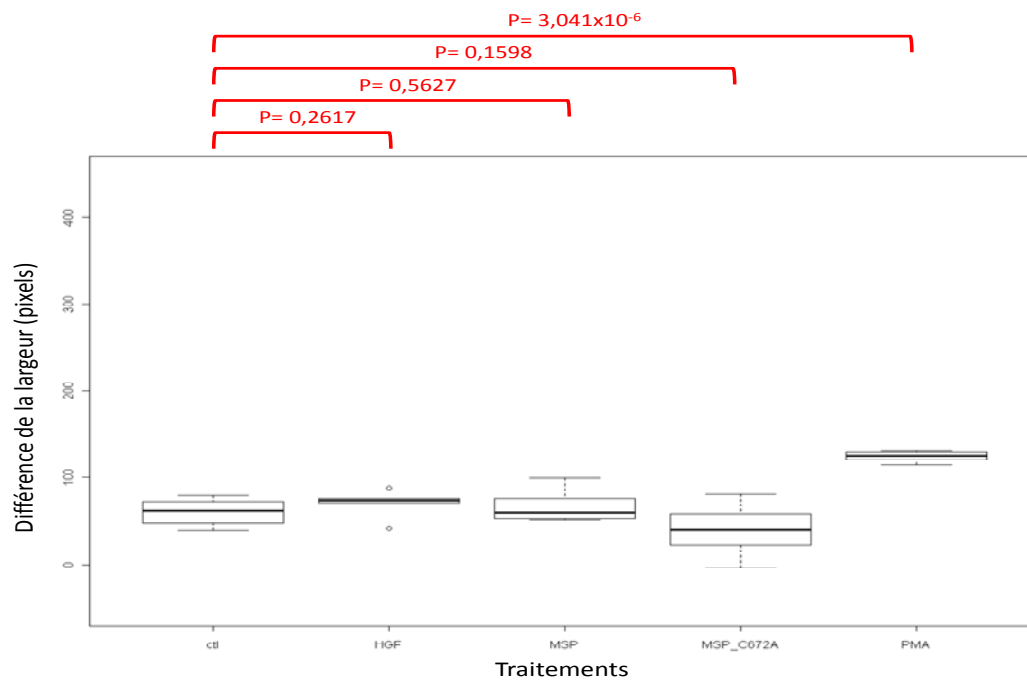


Figure 11: Expression relative de *MST1R* chez différentes lignées cellulaires. Les barres d'erreur représentent la déviation standard (SD). Le gène de référence est *HPRT1*. N=3.

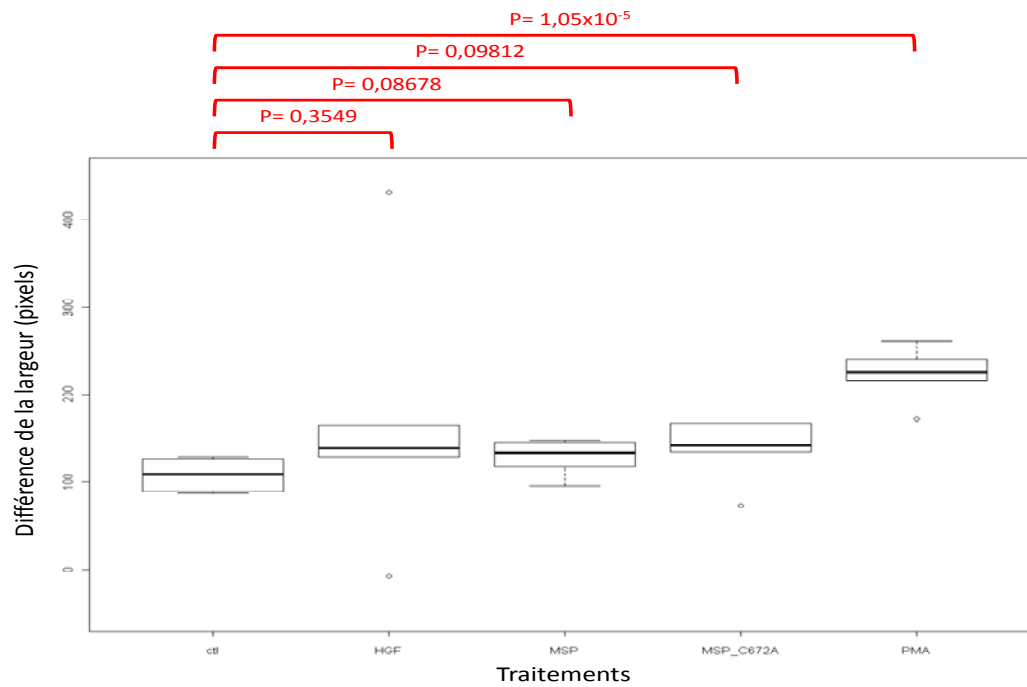
3.2.3 Migration des cellules épithéliales des intestins traitées à la MSP

Après avoir déterminé que les lignées cellulaires épithéliales de l'intestin expriment les ARNm *MST1R*, nous nous sommes basés sur la littérature afin d'appliquer des modèles fonctionnels décrits précédemment qui nous permettraient d'évaluer l'effet de MSP. Nous avons d'abord testé l'effet d'une stimulation à la MSP sur la cicatrisation d'une blessure infligée aux épithélia de cellules du côlon HCT-15, HCT-116 et HT-29. Comme il a été démontré que HGF et PMA activaient la migration des cellules HT-29 dans ce modèle, ces traitements ont été utilisés comme contrôles positifs dans nos expériences [100]. Afin de nous donner plus de chance d'observer une réponse à la stimulation avec MSP dans nos modèles, nous avons également vérifié l'effet de la MSP recombinante C672A, dont l'activité spécifique est de 10 à 20 fois plus grande que la MSP recombinante de type sauvage (WT) [101]. Les résultats de l'expérience de cicatrisation sur les HT-29 sont présentés dans la figure 12.

A 24h



B 48h



C 72h

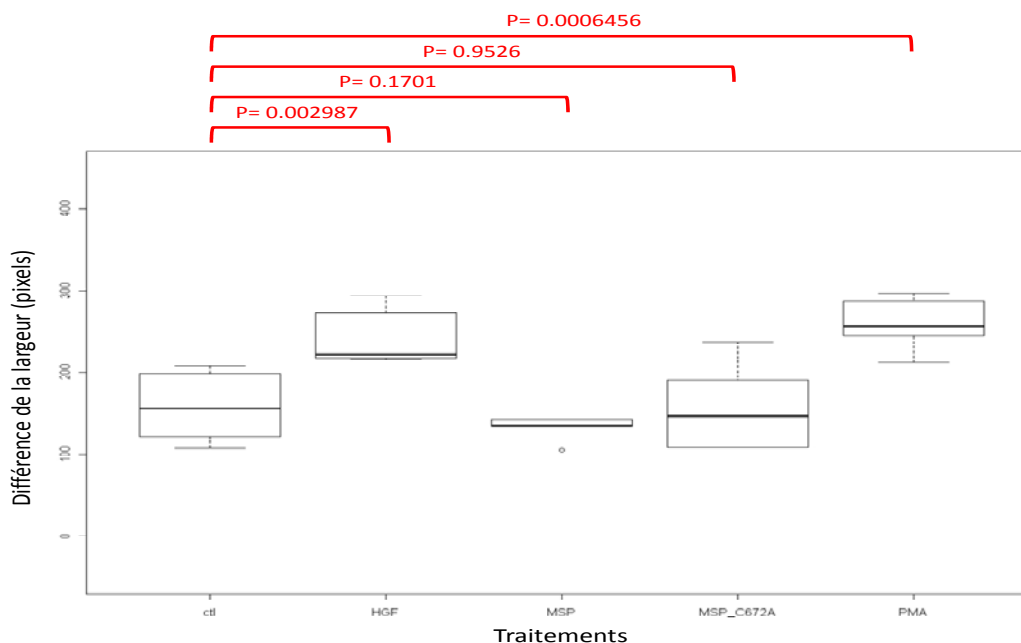


Figure 12: Graphique à boîtes et moustaches présentant les résultats de cicatrisation des blessures de tissus épithéliaux intestinaux après différents traitements. La cicatrisation est mesurée en différence de largeur de la blessure entre les temps de traitements 24 et zéro h en A, 48 et zéro h en B et 72 et zéro h en C. Les contours des boîtes représentent les premier (Q1) et troisième quartiles (Q3), alors que la ligne du centre représente la médiane. La longueur des barres, qui mesure l'indice de dispersion, est calculée par $\pm 1.50 \times (Q3-Q1)$. Les mesures qui tombent à l'extérieur de ces barres sont représentées par un cercle. N=6.

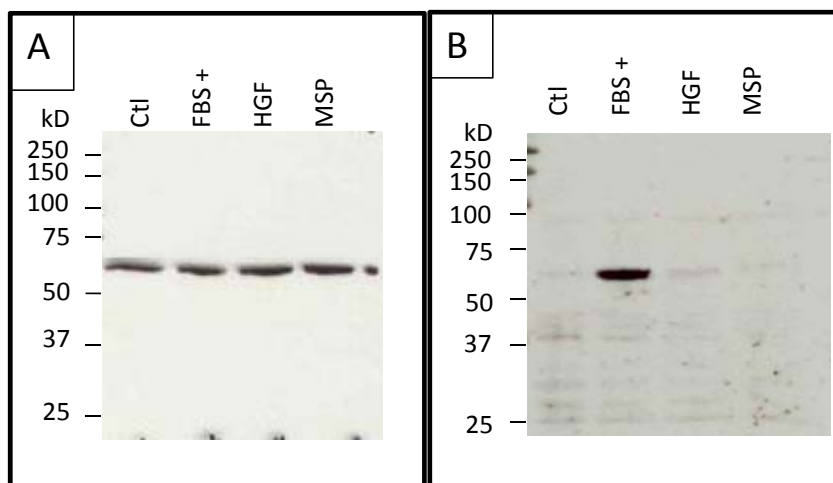
Après 24 h et 48 h de stimulation, nous observons que seul le traitement au PMA semble avoir eu un effet significatif (respectivement $p= 3.04 \times 10^{-6}$ et 1.05×10^{-5}) sur la cicatrisation des blessures des tissus épithéliaux de cellules HT-29, comparativement aux échantillons sans traitement. Après 72 h de stimulation, seuls les traitements au PMA et au HGF ont significativement augmenté la cicatrisation des tissus par rapport au contrôle sans traitement (respectivement $p= 0.000646$ et 0.00299)

Nous avons aussi évalué l'effet de MSP sur la migration des cellules épithéliales du côlon dans un modèle de dispersion des colonies de cellules. Après une distribution des cellules HCT-15, HT-29 et HCT-116 à faible densité dans des puits de culture cellulaire, nous avons traité celles-ci au PMA, HGF et MSP recombinante de type sauvage. Nous avons ensuite photographié des colonies de cellules à différents temps, et évalué de façon visuelle non-quantitative les changements de morphologie ou la migration des cellules composant les colonies. Les colonies présentées dans l'annexe F sont un exemple tiré du lot de 4 colonies échantillonnées par puits. D'après nos observations, le traitement au PMA augmente considérablement la dispersion des colonies pour les trois lignées cellulaires dès 24 h de traitement, tandis que les cellules traitées au HGF semblent légèrement plus allongées à 24 h que les cellules du contrôle sans traitement. Nous n'avons observé aucune différence morphologique des colonies ou des cellules avec les traitements au MSP comparativement aux contrôles négatifs. Les résultats dans ces modèles confirment ce qui est vu dans le modèle de cicatrisation précédent, c'est-à-dire que la MSP ne semble pas avoir un effet observable sur la migration des cellules épithéliales de côlon.

3.2.4 Induction par MSP de la phosphorylation de AKT chez nos modèles de cellules épithéliales du côlon

Afin de mesurer l'effet de MSP sur l'activation de la voie de signalisation sous-jacente à la phosphorylation de RON, nous avons évalué la phosphorylation d'AKT par immunobuvardage. En effet, il a été rapporté que AKT est activée à la suite d'un traitement au MSP dans des cellules épithéliales de l'intestin [82]. Afin d'évaluer le niveau le niveau de phosphorylation de AKT en comparaison à des niveaux minimums de base, nous avons privé les cellules HCT-15, HCT-116 et HT-29 de sérum pour 16 hrs avant le traitement avec la MSP recombinante de type sauvage. Normalement, chez les cellules privées de sérum, la phosphorylation d'AKT est réduite ou absente, puisque certains facteurs de croissance présents dans le sérum induisent cette phosphorylation.

Des traitements au FBS et au HGF ont été utilisés comme contrôles positifs car il a été démontré que ces traitements induisaient la phosphorylation de AKT [102, 103] dans ces cellules. Nous avons aussi utilisé un anticorps anti-AKT afin de vérifier la quantité relative d'AKT entre les différents échantillons. Chez les trois lignées de cellules épithéliales du côlon, les contrôles positifs de 30 minutes de stimulation au FBS et de 10 minutes de stimulation au HGF nous a permis de valider la phosphorylation de AKT (bandes à 60 kD) dans ce modèle. Le HGF semble induire plus faiblement la phosphorylation d'AKT chez les HT-29 et HCT-15 que chez les HCT-116. Notre traitement avec MSP, ne semble pas induire de façon significative la phosphorylation d'AKT dans les cellules HT-29, et seulement faiblement chez les HCT-116 comparativement à nos deux contrôles positifs. Nous observons cependant que chez les HCT-15, le traitement à la MSP induit clairement la phosphorylation d'AKT.



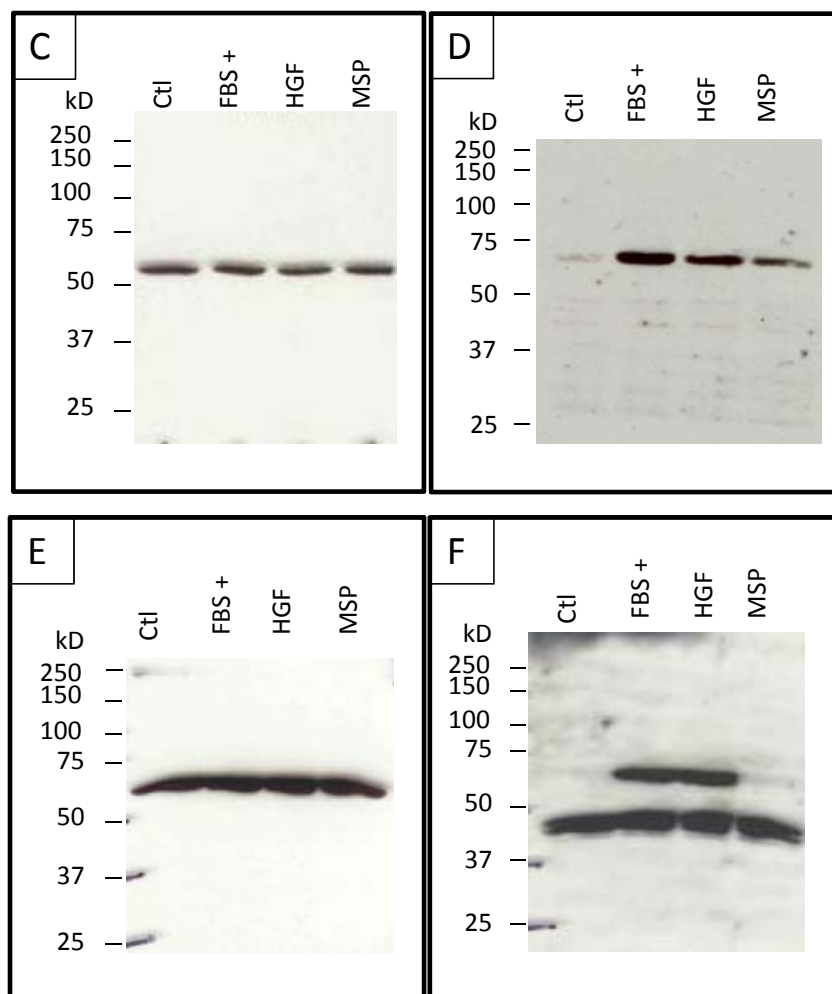


Figure 13: Expression de AKT et Phospho-AKT chez des cellules HT-29, HCT-15 et HCT-116 traitées au FBS, HGF et MSP.
Expression de AKT (A,C,E) et Phospho-AKT (B,D,F) chez des cellules HT-29 (A,B), HCT-15 (C,D) et HCT-116 (E,F) traitées au FBS, HGF et MSP. Vingt ug d'extrait de protéine HT-29 et HCT-15 ont été chargés par puits, et 30 ug pour HCT-116.

La stimulation avec MSP de certaines lignées humaines épithéliales du côlon induirait donc la phosphorylation d'AKT. Comme le seul récepteur connu de MSP est RON, cette phosphorylation est très probablement sous-jacente à l'activation de la phosphorylation de RON.

3.3 PRODUCTION DE LA MSP RECOMBINANTE À PARTIR DE LA TRANSFECTION DE CELLULES DE MAMMIFÈRES

Puisque nous avons trouvé un modèle de cellules humaines sur lequel un effet fonctionnel de la MSP est mesurable, nous avons tenté de produire une MSP recombinante par transfection d'un plasmide contenant la séquence humaine de MST1 dans des cellules de mammifères. Ce plasmide servira dans des expériences de mutagenèse futures visant à déterminer l'effet de la mutation R689C sur la fonction de MSP dans ce modèle. En effet, la mutation R689C (rs3197999), dû à sa position dans la pochette d'affinité de MSP pour RON, et à l'introduction d'une cystéine supplémentaire, pourrait potentiellement affecter la fonction ou la maturation de MSP.

Le plasmide d'expression Insect/Mammalia de Qiagen (voir matériel et méthode) contenant la séquence optimisée du gène MST1, incluant en position N-terminale le peptide signal pour la sécrétion extracellulaire de la protéine et une étiquette d'histidines en position C-terminale, a d'abord été introduit dans des bactéries E.coli afin d'obtenir une grande quantité de ce plasmide. Afin de vérifier l'identité du plasmide produit, ce dernier a été digéré par les enzymes de restriction XhoI, NcoI et AgeI. Les produits de la digestion ont été chargés sur un gel et visualisés par UV en présence de bromure d'éthidium (fig. 14). D'après la carte du plasmide (matériel et méthode, fig. 3), la digestion avec ces enzymes de restriction devrait nous permettre d'obtenir des fragments de 5182 pb et 2213 pb avec la combinaison XhoI et NcoI, et des fragments de 6169 pb et 1226 pb avec la combinaison XhoI et AgeI. D'après la figure 14, la digestion avec XhoI et NcoI nous permet de détecter des fragments d'un peu plus de 5 kb et 2.4 kb, et la digestion avec XhoI et AgeI donne des fragments de 6 kb et 1.3 kb. Ces fragments sont sensiblement plus longs que prévus, et d'autres analyses seront nécessaires afin de valider des résultats. Le séquençage du plasmide purifié ainsi que l'alignement de sa séquence avec la séquence fournie par Qiagen a aussi permis de valider l'identité de ce plasmide comme étant bien le plasmide MST1 Insect/Mammalia et l'intégrité de la séquence de MST1. Nous avons ensuite introduit ce plasmide d'expression dans des cellules de

mammifères HEK293, et la MSP recombinante sécrétée dans le milieu de culture a ensuite été purifiée tel que décrit dans la section 2.4.2 des Matériel et Méthodes. Nous avons ensuite digéré une partie de la pro-MSP purifiée avec la

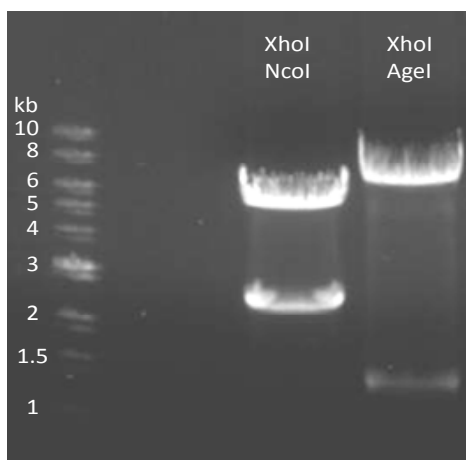


Figure 14: Digestion du plasmide MST1 Insect/Mammalia avec une combinaison d'enzymes de restriction XhoI et AgeI, et XhoI et NcoI.

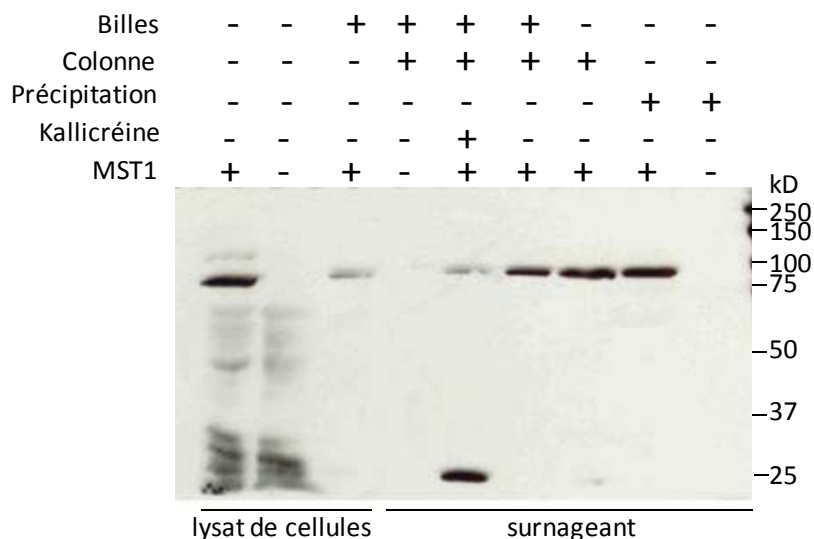


Figure 15: Immunobuvardage montrant les étapes de purification de la MSP recombinante.

Les cellules transfectées ont été lysées. Le surnageant a été ou précipité, ou concentré dans une colonne d'exclusion par poids moléculaire. Une partie de la MSP sécrétée et concentrée dans la colonne a été purifiée sur billes. Une partie des la MSP purifiée a ensuite été digérée par la kallibréine. 1 :10 des lysats et du concentré et précipité de protéines de surnageant ont été chargées par puits correspondant. 1 :3 du total de billes a été chargé par puits correspondant.

kallibréine afin de la convertir en MSP mature. À chaque étape de purification et de maturation, des échantillons ont été prélevés et chargés dans un gel, puis visualisés par immunobuvardage avec un anticorps dirigé contre les penta-histidines (fig. 15). Les lysats et le surnageant de cellules de HEK 293 non transfectés avec le plasmide ne montrent aucune bande détectable avec l'anticorps dirigé contre les penta-histidines, tandis que le lysat et le surnageant de cellules transfectées montre tel qu'attendu une protéine à 80kDa correspondant à la pro-MSP, suggérant une transfection et une expression efficace du plasmide, mais montrant aussi qu'une certaine quantité de MSP reste dans la cellule et n'est pas sécrété. L'échantillon pris à la suite de l'étape de concentration des surnageants par colonne nous démontre qu'il y a peu de perte de MSP dans cette étape. L'étape de purification par bille nous montre qu'il y a un peu de perte de MSP dans cette étape puisque le volume chargé équivaut au tiers du volume de surnageant, contre 1 :10 du volume de surnageant chargé pour la colonne. Enfin, la MSP produite et purifiée a été digérée par la kallibréine afin d'être activée. La kallibréine digère presque complètement la MSP produite puisque l'on distingue une bande à 27 kB qui correspond à l'unité beta de MSP qui porte l'étiquette de histidines.

Nous pouvons donc produire de la MSP recombinante par transfection, la purifier dans des quantités satisfaisantes et la digérer à l'aide de la kallibréine. L'optimisation de la digestion de MSP par la kallibréine a déterminé qu'une concentration de 4.5 ng de kallibréine par ul de MSP purifiée selon notre protocole, et une digestion de 1 h

permettait une digestion optimale de la MSP (résultats non montrés). Une quantification de la MSP produite a également été tentée, mais les résultats n'ont pas été concluants.

Également, la MSP ainsi produite et activée par la kallibréine a été utilisée dans une expérience de phosphorylation de AKT chez nos lignées épithéliales du côlon. Aucune phosphorylation de AKT n'a pu être détectée suite à la stimulation de ces cellules avec la MSP recombinante produite (résultats non montrés).

D'autres expériences seront nécessaires dans le futur afin de quantifier la MSP produite par transfection, et aussi afin de re-tester l'effet de cette MSP sur notre modèle chez des lignées de cellules humaines épithéliales du côlon.

CHAPITRE 4

DISCUSSION

4.1 *Deux signaux d'association indépendants sont impliqués dans la susceptibilité aux MII dans la région 3p21.*

4.1.1 Test d'association de variantes de MST1 et MST1R

Les expériences réalisées dans ce mémoire visent à améliorer notre compréhension des signaux d'association provenant de la région chromosomique 3p21 pour les maladies inflammatoires de l'intestin. En effet, de multiples études de liaison et d'association ont identifié cette région génomique comme facteur de risque aux MII. Le signal d'association le plus répliqué jusqu'ici provient du SNP codant non-synonyme rs3197999, situé dans le gène *MST1* [58, 66], qui encode la protéine MSP. Cette variante encode pour la mutation R689C de MSP, et pourrait, en introduisant une cystéine supplémentaire et en étant située dans le site de liaison de MSP avec son récepteur, être dommageable pour la fonction et la maturation de MSP. Le signal d'association dans 3p21 provient très probablement de la variante R689C. Cependant, une étude indépendante a aussi identifié un signal d'association à la maladie de Crohn corrélé à des SNP codants non-synonymes de *MST1R* [53]. Ce gène encode pour la protéine RON, le récepteur de MSP, et est situé à 250 kb du gène *MST1*. La corrélation entre ces deux signaux étant faible ($r^2 > 0.387$), il est possible que ces deux signaux d'association soient indépendants l'un de l'autre. De plus, *MST1* et *MST1R* sont tous les deux de bons gènes candidats dans l'optique des MII, puisque qu'ils ont un rôle lié à l'inflammation. Enfin, bien que l'existence de duplications de la séquence entourant le gène *MST1* puisse mettre en doute la validité du génotypage

des SNP situés au cœur de cette séquence, le LD entre ces variantes et celles situées à l'extérieur de la région dupliquée reste notre meilleur outil afin de confirmer la légitimité des signaux qui y sont détectés. Tel est entre autre le cas de R689C, fortement corrélé au SNP rs9858542 mainte fois associé aux MII et situé à l'extérieur de la région dupliquée. La conception d'amorces spécifiques à la séquence du chromosome 3 est également une étape déterminante afin d'obtenir des résultats valides. Le bon taux de génotypage de R689C dans nos essais, indiquant une amplification spécifique de la séquence sur le chromosome 3, est dû au positionnement des amorces de PCR et d'extension sur des différences nucléotidiques entre les duplications et la séquence de *MST1*. Ces observations suggèrent la validité des résultats de génotypage pour R689C.

Ainsi, comme les études d'association précédentes n'ont pas fait un génotypage exhaustif des gènes *MST1* et *MST1R*, nous avons testé l'association d'autres variantes de *MST1* et *MST1R*. Dans ce but, nous avons génotypé dans une cohorte de 1110 témoins et 1647 cas de CU, des SNP identifiés dans une expérience de re-séquençage de ces gènes ou dans la base de données dbSNP. Lors de l'étape de contrôle de qualité, les seuils de taux de génotypage des ADN et des SNP ont été gardés assez haut pour éliminer les génotypes contenant trop d'erreurs, mais aussi assez bas pour permettre l'analyse des variants dont le génotypage est compliqué par l'existence des duplications dans la région de *MST1*.

Les tests d'association subséquents nous montrent que autre que R689C, une seule variante codante synonyme de *MST1*, R641R (rs13085791), a un signal d'association significatif à la CU. R641R est parfaitement corrélée à R689C ($r^2=1$), notre signal d'association principal dans la région 3p21. Puisque les SNP codant non-synonymes sont plus aptes à induire un dysfonctionnement des protéines comparativement aux variantes codantes synonymes, le signal d'association à R641R est probablement dû à sa grande corrélation avec R689C. En effet, il a été suggéré que les arginines autour de ce site sont indispensables à cette liaison[63]. De plus, l'introduction d'une autre cystéine dans la séquence d'acides aminés de MSP pourrait permettre la formation d'un pont disulfure

supplémentaire qui pourrait changer le repliement de la protéine[66]. Une seule des variantes codante non-synonyme corrélée ($r^2 > 0.5$) à R689C est un SNP du gène *BSN*. Il introduit un codon stop dans la protéine, et pourrait avoir un effet fort sur la fonction de sa protéine, mais il n'a jamais été génotypé. Cependant, la protéine *basoon*, encodée par *BSN*, semble avoir un rôle connu que dans la formation de la matrice des terminaisons neuronales, et est exprimée surtout dans le cerveau, ce qui ne fait pas de *BSN* un bon gène candidat ayant un rôle à jouer dans la pathogénèse des MII. Nos résultats suggèrent donc que R689C est la variante causale d'où origine ce signal d'association. Pour *MSTIR*, aucune des variantes testées ne montre de signal d'association à la CU.

Cependant, on ne peut pas exclure la possibilité que d'autres variantes additionnelles de *MSTI* et *MSTIR* soient associées aux MII, pour diverses raisons. Entre autre, probablement que tous les SNP de ces gènes n'ont pas été génotypés. Aussi, une étude indépendante avait précédemment détecté un autre signal d'association à la MC en corrélation avec *MSTIR* [53]. Malgré qu'aucun signal d'association à la CU n'a été détecté pour *MSTIR* dans notre cohorte, il est aussi possible que ce signal soit plutôt un facteur de risque exclusif à la MC, ou que la taille de notre cohorte ne soit pas suffisante (1110 témoins et 1647 cas) pour détecter une association significative à la CU pour un variant ayant un faible effet.

Ainsi, bien que nos résultats suggèrent que R689C est une variante causale dans *MSTI*, et qu'aucune association n'est détectée pour les SNP de *MSTIR* qui ont été testés, il serait nécessaire de génotyper les variantes de ces gènes dans les cas de MC de notre cohorte, ainsi que d'augmenter de la taille des cohortes génotypées avant de conclure qu'il n'y a pas d'autres signaux d'association dans la région 3p21.

4.1.2 Test d'association conditionnel dans la région 3p21

Les résultats d'études association aux MII publiées nous suggèrent qu'il y a plus d'un signal d'association dans la région 3p21. Bien que nous ayons précédemment validé que l'association pour R689C du gène *MST1*, il est possible que la région 3p21 contienne d'autres facteurs de risque ayant un effet trop faible pour que nous les ayons détectés dans notre étude précédente. Dans le but de trouver des signaux d'association indépendants potentiels dans la région 3p21 dus à des facteurs de susceptibilité ayant un faible effet ou exclusifs à l'un ou l'autre des phénotypes de MII, nous avons voulu tester l'association de variantes de la région 3p21 dans une cohorte ayant une très grande taille, et comprenant des cas de CU et de MC. Une méta-analyse de plusieurs études d'association est une des méthodes éprouvées [104] qui permet d'atteindre ces objectifs lorsque l'on n'a pas accès aux données de génotypage bruts, en combinant les résultats d'association provenant de différentes études. Pour ce faire, nous avons dû en premier lieu limiter le nombre de variantes dont l'association conditionnelle est à tester à celles qui semblent strictement indépendantes. Nous avons donc identifiés, en consultant les résultats disponibles de la méta-analyse d'études d'association à la CU de McGovern (6791 témoins et 2693 cas de CU) [35], tous les SNP ayant un signal d'association inférieur à 1×10^{-4} , et qui sont faiblement corrélés entre eux ($r^2 < 0.4$). L'association à R689C a été identifiée comme le signal principal dans cette région, et deux autres variantes, l'une intronique (rs6802890) et l'autre codante synonyme A71A (rs7629936) du gène *CDHR4*, ont également montré des signaux suggestifs d'association indépendants. Le SNP intronique est un proxy ($r^2=1$) du signal corrélé à *MST1R*[53]. Il semble donc potentiellement y avoir trois signaux indépendants d'association aux MII dans la région 3p21. Nous avons ensuite génotypé ces trois SNP dans notre cohorte de 1110 témoins, 1647 cas de UC et 799 cas de MC, puis nous avons fait des tests d'association par régression logistique. Nous avons aussi demandé à des collaborateurs de nous fournir les résultats d'association venant de 3 études indépendantes d'association pangénomique aux MII pour ces 3 SNP. Ces résultats ainsi que ceux de notre étude ont

ensuite été combinés dans une méta-analyse à effet fixe. Ainsi, nous avons obtenus l'équivalent des résultats d'association de 6441 témoins, 4295 cas de CU et 2664 cas de MC pour ces 3 SNP.

Les résultats de cette méta-analyse nous montrent que les trois SNP sont associés significativement à la fois pour les MII, MC et pour CU, et que le signal le plus fort est détecté pour R689C (R689C : 1.52×10^{-11} , *CDHR4**intronique : 2.78×10^{-8} , *CDHR4**A71A : 7.0×10^{-8}), ce qui confirme qu'il y a potentiellement trois variantes causales dans la région 3p21. L'indépendance de ces trois signaux a ensuite été testée par régression logistique conditionnelle. Pour ce faire, ce type d'analyse a été faite sur nos données de génotypage pour chaque SNP et pour chaque phénotype. Nous avons également demandé à nos collaborateurs de faire la même analyse et de nous envoyer les résultats afin de les combiner dans une autre méta-analyse. D'après les résultats, le signal à R689C est indépendant des signaux des deux autres SNP pour les MII, MC et CU. Le signal d'association pour *CDHR4**A71A dans les MII et dans la CU est indépendant de celui de R689C. Cependant, d'après l'analyse faite avec les données de la MC, ce signal ne semble pas indépendant à celui de R689C. Le signal provenant du SNP intronique dans *CDHR4* ne semble pas non plus indépendants du signal de R689C dans les MII, la MC, et la CU.

Cela suggère donc que le principal signal d'association aux MII dans la région 3p21 provient bien de R689C. La région définie par les SNP corrélé à 0.5 à R689C est longue de 474 kb et chevauche 14 gènes, mais ne contient pas d'autres variantes codantes non-synonymes connues dans un gène intéressant (fig. 16). Le second signal d'association semble provenir de la variante *CDHR4**A71A et être exclusif à la CU. Cette variante est située dans le gène appelé *CDHR4*, pour *cadherin-related family member 4*. Le rôle et l'expression de ce gène sont encore inconnus, mais nous savons toutefois que la protéine encodée par ce gène partage des domaines semblables à ceux des cadhérines, des protéines membranaires ayant une fonction dans l'adhésion des cellules dépendante des

ions calcium. Aucun SNP n'est corrélé à *CDHR4**A71A à $r^2 > 0.8$. Cependant, la région définie par les SNP en corrélation de 0.5 avec *CDHR4**A71A, longue de 426 kb et englobant 16 gènes, semble chevaucher *MST1* mais pas *MST1R*, et il est donc possible que le deuxième signal provienne d'une autre variante de *MST1*. Il existe un autre gène dans la région corrélée, *RHOA*, qui pourrait avoir un rôle intéressant. En effet, *RHOA* encode une protéine du signal de transduction des changements de l'actine et qui est une cible de *Y. pseudotuberculosis*, qui cause des désordres digestifs. Cependant, il n'y a aucun SNP codant non-synonymes de ce gène corrélé à ce signal. Aussi, l'absence d'association de ce signal avec la MC pourrait représenter un manque de puissance dans notre cohorte de MC (n=2664 dans MC, n=4295 dans CU et n=6959 dans MII), et donc il faudra sans doute génotyper cette variante dans un plus grand échantillon de MC.

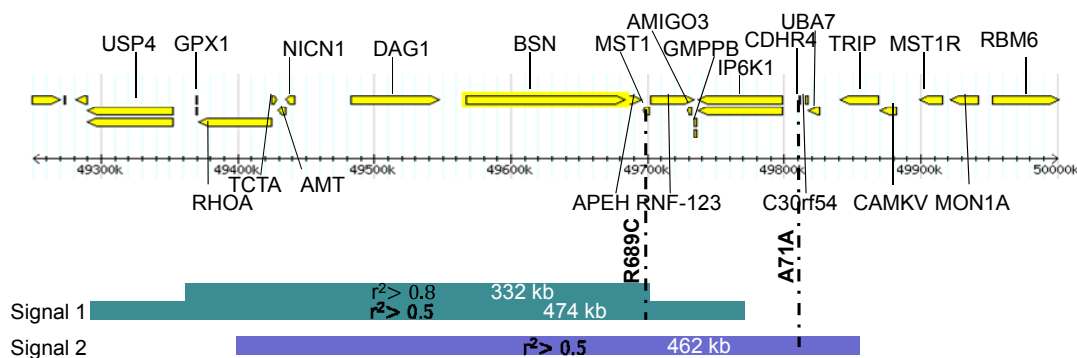


Figure 16 : Région définie par les SNP corrélés à nos signaux d'association indépendants ($r^2 > 0.5$ et 0.8).

En conclusion, il y aurait potentiellement deux signaux d'association aux MII dans la région 3p21. Puisque le SNP associé dans l'étude de Beckly et al. [53] est un proxy du SNP intronique de *CDHR4*, il est probable que son association aux MII ne soit pas indépendant du signal au SNP R689C. Également, d'après nos résultats de génotypage de variantes de *MST1* et *MST1R*, il ne semble pas y avoir d'autres variantes de ces gènes associées à la CU. Cependant, il reste encore beaucoup d'autres variantes pour lesquelles le test d'association n'a pu être fait. Il serait intéressant de re-séquencer des régions en

LD avec nos deux signaux, géotyper toutes ces variantes dans une cohorte de grande taille et enfin, faire une analyse logistique conditionnelle afin de tester l'indépendance des nouveaux signaux d'association que l'on pourrait trouver dans la région 3p21.

4.2 *MSP, encodée par MST1, semble avoir un rôle régulateur de la phosphorylation de AKT, mais pas dans la modulation de la réponse inflammatoire*

4.2.1 Tester l'effet de MSP sur un modèle de réponse inflammatoire chez les macrophages

Comme la variante R689C dans le gène *MST1* montre le signal d'association aux MII le plus fort dans la région 3p21, que le gène *MST1* semble être un candidat ayant un rôle dans l'inflammation, et que la variante R689C a de forte chance de modifier la fonction et le repliement normal de la protéine MSP, nous nous sommes penchés sur le rôle de cette protéine. Une fois la fonction de MSP chez les cellules humaines mise dans le contexte des MII, nous pourrions éventuellement vérifier l'effet de la mutation R689C sur nos modèles. Malheureusement, les connaissances sur la fonction de MSP proviennent presque toutes d'études menées sur des modèles chez la souris. Nous avons donc adapté des modèles avec des cellules humaines d'après les modèles chez la souris.

Tout d'abord, on sait que la pro-MSP est produite principalement par les hépatocytes sous forme inactive, et est mise en circulation dans le système sanguin. Sur les sites d'inflammation, la pro-MSP est clivée par des facteurs de coagulation, ainsi que par la matriptase, une enzyme à la surface des macrophages, ce qui permet à la MSP de s'activer et de se lier à son récepteur, RON. Une fois ce dernier auto-phosphorylé, il y a activation d'une variété de voies de signalisation qui vont de l'inhibition de la réponse

inflammatoire chez les macrophages, à la migration des cellules, à la cicatrisation des cellules épithéliales et à la survie cellulaire. Il a en effet été démontré qu'un prétraitement avec la MSP chez les macrophages de souris stimulés au LPS pouvait inhiber la production de NO, de iNOS et de IL12p40 via une inactivation de la voie de signalisation de NF-kB [85-88]. La seule étude faite chez les macrophages humains a démontré que MSP inhibait la production de molécules pro-inflammatoires en augmentant la production de O_2^- .

Nous avons donc évalué l'effet d'un prétraitement à la MSP suivi d'une stimulation au LPS sur des macrophages humains dérivés de monocytes circulants ou dérivés de la lignée cellulaire THP-1, en mesurant l'expression de *TNF- α* et *IL12B* (IL12p40), des gènes pro-inflammatoires transcrits par le facteur de transcription NF-kB. Nous avons également mesuré l'expression de marqueurs de différenciation des macrophages afin que nos modèles soient le plus comparables possible aux modèles de macrophages de souris de la littérature. Cependant, dans ces deux modèles, bien que les monocytes différenciés expriment des marqueurs de différenciation propres aux macrophages, aucun effet de la MSP n'a été observé sur l'expression des RNA de *TNF- α* et *IL12B*.

Les différences entre nos résultats et les résultats publiés auparavant dépendent probablement du modèle cellulaire utilisé, qui diffèrent à deux niveaux. Premièrement, nos modèles diffèrent au niveau de l'espèce. Ainsi, dans la majorité des expériences, les modèles utilisés étaient une lignée de macrophages de souris Raw264.7 transfectée avec un plasmide contenant la séquence de RON, ou des macrophages péritonéaux de souris, alors que nous utilisons des macrophages humains. Or, il est connu que les macrophages de souris ont une réponse inflammatoire un peu différente de celle des macrophages humains, puisque ces derniers ne produisent presque pas de NO en réponse au LPS, contrairement aux macrophages de souris [105]. Deuxièmement, notre modèle diffère de ceux dans la littérature par l'origine des macrophages. En effet, des macrophages d'exsudat péritonéaux ont été utilisés, alors que nous avons plutôt utilisé des

macrophages dérivés des monocytes du sang ou de la lignée THP-1. Or il a été démontré que RON n'était exprimé que par les cellules hautement différenciées dans la cavité péritonéale. De plus, il a été montré chez les rats que les monocytes du sang sont différenciés en macrophages hautement différenciés « résidents » après 6 jours de culture. Bien que nous ayons différencié nos monocytes en macrophages pendant 7 jours, peut-être les monocytes du sang nécessitent-ils une stimulation spécifique dans les tissus afin d'atteindre ce niveau de différenciation et produire RON. Enfin, dans la seule étude utilisant des macrophages humains dérivés des monocytes du sang, seule la production de O_2^- était mesurée suite à un prétraitement au MSP [97]. Il est possible que la production de O_2^- soit considérablement plus marquée chez ces cellules que l'expression des gènes pro-inflammatoires tels que *IL12B* et *TNF-a*.

Ainsi, nos résultats nous suggèrent que dans nos modèles de macrophages humains dérivés des monocytes du sang et des THP1, MSP n'a pas l'effet inhibiteur de la production de IL12p40 tel que vu chez les cellules de souris. Cependant cela n'exclut pas la possibilité pour MSP d'avoir un autre rôle en lien avec les MII dans les cellules humaines.

4.2.2 Tester l'effet de MSP sur des modèles de cellules épithéliales

En vue de valider d'autres fonctions potentielles de la MSP chez l'humain, nous avons criblé la littérature afin de trouver d'autres modèles à adapter aux cellules humaines. Il a été démontré que MSP induisait la migration ainsi que des changements de morphologie chez des cellules épithéliales de souris [106, 107]. Également, quelques études mesurant l'effet de MSP chez les cellules épithéliales humaines ont démontré que MSP induit la phosphorylation de AKT [82, 108], une sérine/thréonine protéine kinase en amont de la voie de signalisation de la survie cellulaire.

Nous avons donc stimulé des cellules humaines de lignées épithéliales de côlon (HCT-15, HCT-116 et HT-29) avec de la MSP et nous avons observé son effet sur la cicatrisation et la migration des cellules. L'expression de RON a été mesurée chez ces différentes lignées (voir résultats, figure 11) afin de s'assurer que ces cellules pouvaient être réceptives à la MSP. Nous avons également mesuré la phosphorylation d'AKT dans ces lignées lorsqu'elles étaient stimulées avec MSP.

Bien que les cellules ont bien répondu aux contrôles positifs de HGF et FBS, des facteurs de croissance, nous n'avons observé aucun effet de la MSP sur la morphologie et la migration des cellules, ni sur la cicatrisation des tissus de cellules épithéliales du côlon. Par contre, la phosphorylation d'AKT en réponse à une stimulation avec MSP a été observée que pour les cellules HCT-15, mais pas pour les HCT-116, bien qu'elles aient bien répondues aux contrôles positifs.

Les différences entre nos résultats et ceux des études précédentes peuvent être causés par différents facteurs. Entre autre, dans les expériences précédentes de migration des cellules [106], des pétris étaient recouverts de collagène, et ou de laminine-5, de fibronectine et de vitronectine ont été utilisés, alors que la surface de nos pétris ont simplement été traités pour l'attachement des cellules. Il a d'ailleurs été suggéré que la cicatrisation de tissus de kératinocytes était induite par l'interaction de RON avec l'intégrine $\alpha 6\beta 4$ à la surface de la cellule [106]. Aussi, les différences entre nos résultats et ceux des études précédentes résident peut-être dans la concentration de MSP utilisée lors des essais. En effet, dans la plupart des expériences précédentes, une dose de MSP presque deux fois plus élevée que celle dans nos expériences a été utilisée. La seule exception dans une étude [108] où une MSP humaine isolé à partir du plasma a été utilisée à une concentration moindre que dans nos expériences. Il est également possible que l'affinité de la MSP recombinante que nous avons acheté soit moins grande que la MSP provenant du plasma humain, puisqu'elle est produite par des cellules de rongeurs.

Néanmoins, notre expérience montre que chez les HCT-15, MSP induit significativement la phosphorylation d'AKT. Cela prouve d'abord que la MSP utilisée jusqu'ici a une affinité pour le récepteur RON dans ce modèle, mais aussi suggère que la MSP a probablement un rôle dans la survie cellulaire chez les cellules épithéliales humaines. En effet, l'activation d'AKT mène à une inhibition des voies apoptotiques, mais également à une augmentation de la migration des cellules ainsi qu'à une augmentation de la production de protéines[109]. Dans le cadre des MII, on peut aisément imaginer qu'un déficit de la survie cellulaire pourrait particulièrement avoir plusieurs effets néfastes sur la santé des tissus intestinaux. Nous savons entre autre que les cellules de l'épithélium de l'intestin se renouvellent à un rythme régulier de quelques jours. Alors que les cellules indifférenciées mûrissent en cellules épithéliales au fond des cryptes entre les villosités intestinales, les cellules plus vieilles sont déplacées vers l'extrémité des villosités, et au fur et à mesure qu'elles se détachent de la matrice extracellulaire, elles entrent en apoptose. Alors que la cellule apoptotique se desquame, il y a sécrétion d'une substance visqueuse dont la nature est encore inconnue, protégeant le trou laissé par la cellule morte, en attendant que les cellules voisines ne le combrent en se rapprochant. Ce processus limite l'intrusion des micro-organismes à travers la barrière épithéliale dans l'intestin normal. Or, il a été noté que chez les patients souffrant de CU et de MC, ce processus est compromis, puisque le taux de cellules apoptotiques est beaucoup plus élevé que chez les individus non affectés, et il y a de grandes régions d'épithélium intestinal non protégées. L'apoptose peut également être un mécanisme utilisé par des microorganismes pathogènes pour se libérer des cellules hôtes et ainsi se disséminer dans les cellules voisines. Également, les toxines produites par les bactéries peuvent induire l'apoptose des cellules épithéliales, et ainsi permettre l'intrusion bactérienne à travers la barrière épithéliale. Une mutation de la MSP qui limiterait son rôle anti-apoptotique pourrait favoriser l'invasion des microbes à travers la barrière épithéliale et la transmission des microbes intra-cytoplasmiques, ce qui pourrait mener à une inflammation continue comme vu chez les patients de MII.

Enfin, nous savons maintenant que la MSP ne semble pas inhiber l'expression des gènes de la réponse inflammatoire *IL12B* et *TNF- α* dans nos modèles de macrophages. Elle ne semble pas non plus avoir d'effet significatif sur la migration et la morphologie des cellules épithéliales du côlon, ni sur la cicatrisation de leur tissu. Cependant, notre modèle de la phosphorylation d'AKT induite par MSP chez les cellules épithéliales du côlon nous permettra de vérifier l'effet de la mutation R689C sur la fonction normale de MSP. Nous pourrions également, dans le même but, tester directement la phosphorylation de RON dans ce modèle puisque nous savons que ce dernier répond à la stimulation avec MSP. Il serait également intéressant, à la lumière de ces résultats, de mesurer directement la survie cellulaire des cellules épithéliales de l'intestin exposées à la MSP. Cela pourrait confirmer l'importance de la fonction de MSP sur l'intégrité des tissus intestinaux et expliquer son association aux MII.

4.3 Limites de l'étude

Comme dans toute étude, celle présentée dans ce mémoire a des limitations à la fois dans les essais génétiques et dans les essais fonctionnels.

Entre autres, les tests d'association des variants de la région 3p21 ont surtout ciblé des variants des gènes candidats *MST1* et *MST1R*. La sélection de ces candidats se base sur des prémisses telles que la fonction des gènes et la nature de leurs variantes ayant possiblement un rôle à jouer dans la pathogenèse des MII d'après la littérature. À part les variantes associées et indépendantes identifiées à l'aide des résultats de la méta-analyse de McGovern, cette étude ne teste pas les variantes situées dans d'autres gènes candidats, ni d'autres variantes intergéniques qui pourraient également être causales.

L'existence des duplications dans la région de *MST1* est également un facteur contraignant dans cette étude puisqu'elles diminuent la qualité du génotypage. Afin de

nous permettre d'analyser raisonnablement les variants dans cette région, nous avons diminué les seuils de contrôle de qualité du génotypage, ce qui augmente possiblement la quantité de données erronées lors de l'analyse. De plus, le nombre de variantes testées dans cette région est limité par le désign des amorces spécifique au chromosome 3. D'autres essais devront être faits afin d'améliorer le génotypage des variantes dans la région dupliquée.

Aussi, dans la méta-analyse présentée dans notre étude, le nombre plus restreint de cas de MC ne nous permet probablement pas d'aller chercher autant de signaux d'association comme pour la CU, ce qui pourrait expliquer que notre deuxième signal d'association ne semble pas indépendant du signal principal à R689C pour la MC.

Également, les résultats des essais fonctionnels présentés ici se limitent aussi à nos modèles cellulaires *in vitro*. La réponse des cellules en contact avec la matrice cellulaire et avec d'autres cellules est probablement légèrement différente de celle observée *in vitro*. Les prochains essais testant la fonction de MSP chez les cellules épithéliales ou les macrophages bénéficieraient possiblement du contact de ces cellules avec une matrice extracellulaire.

Le nombre d'observation par essai est limité : un plus grand nombre de duplicatats donnerait des résultats plus significatifs et probablement moins variables. Enfin, les mesures de l'expression de la MSP recombinante par WB ainsi que les mesures de la migration cellulaire chez les cellules épithéliales du colon sont purement qualitative. Afin de permettre de vérifier si les résultats sont significatifs, il faudrait trouver une méthode afin de mesurer ces observations dans les prochains essais semblables.

CONCLUSION

Les maladies inflammatoires de l'intestin, regroupant la colite ulcéreuse et la maladie de Crohn, touchent ensemble près d'une personne sur 250 dans les pays industrialisés, ce qui en fait des maladies communes. Malgré de nombreuses découvertes de gènes impliqués dans ces maladies, nous en savons encore bien peu sur leur étiologie. Il reste donc encore beaucoup de travail à accomplir afin d'identifier d'autres voies à cibler pour contrer le développement des MII.

Puisque pour la majorité des loci associés aux MII, les variantes causales sont encore inconnues, une partie des études à venir, comme celle présentée dans ce mémoire, consisteront à cribler ces loci afin d'identifier les facteurs de risques qu'ils contiennent. La région de 3p21 qui a été associée plusieurs fois aux MII englobe les gènes *MST1* et *MST1R*, des gènes ayant un rôle dans l'inflammation et encodant MSP et RON, respectivement. Dans le but d'identifier d'autres variantes de *MST1* et *MST1R* qui auraient pu être associées aux MII, nous avons génotypé pour ces gènes d'autres SNP dont l'association n'a pas été testée. Ainsi un seul des SNP, fortement corrélé à la variante R689C de MSP, a montré un signal d'association significatif dans notre cohorte de CU. Ces résultats suggèrent que le signal d'association qui englobe *MST1* provient de R689C.

Une partie de l'héritabilité inconnue des MII peut résider dans le fait qu'une région associée peut contenir plus d'une variante causale. Nous avons vérifié si c'était le cas pour la région 3p21 d'abord en identifiant, à l'aide des résultats d'une méta-analyse d'études d'association pangénomique à la CU, les SNP associés faiblement corrélés entre eux. Trois SNP ont été identifiés et génotypés dans notre cohorte composée de cas de CU et de MC et de témoins. Nos résultats de génotypage ont ensuite été combinés aux résultats de 3 autres études d'association pangénomique indépendantes ce qui nous a permis de détecter un signal d'association significatif pour ces trois SNP. Une analyse d'association conditionnelle nous a permis de tester l'indépendance de ces trois signaux,

et nous avons pu en conclure que la région 3p21 contient 2 signaux d'association mutuellement indépendants. Le signal d'association aux MII le plus fort provient de R689C, mais qu'il existe également un signal indépendant secondaire pour la CU corrélé à la variante *CDHR4**A71A.

Une des étapes déterminantes dans la recherche en génétique reste de tester le rôle fonctionnel des gènes et des variantes associés. Nous avons, en premier lieu, défini le rôle de MSP dans des modèles humains d'inflammation, de survie cellulaire, de migration cellulaire ou de cicatrisation tels que démontrés dans les modèles murins publiés précédemment. Ces expériences nous ont permis de conclure que dans nos modèles d'inflammation induite par LPS chez les macrophages dérivés de THP-1 et de monocytes du sang, MSP n'inhibe pas la réponse inflammatoire. Dans les modèles de cicatrisation des tissus épithéliaux du côlon et de migration des cellules épithéliales de l'intestin HT-29, HCT-15 et HCT-116, MSP n'a pas démontré non plus l'effet activateur vu dans la littérature. Cependant, MSP induit clairement chez les cellules HCT-15 la phosphorylation d'AKT, un acteur dans la voie de signalisation de la survie cellulaire. Or, un déséquilibre de l'apoptose des cellules épithéliales de l'intestin pourrait faciliter l'intrusion de bactéries à travers la barrière épithéliale et ainsi déclencher une réponse inflammatoire continue, tel que vu chez les patients de MII. Nous avons ensuite produit une MSP recombinante à partir de la transfection, dans des cellules de mammifères, d'un plasmide dans lequel la mutation R689C sera introduite. Il sera alors possible de vérifier l'effet de cette mutation sur la phosphorylation d'AKT induite par MSP. C'est ainsi qu'à partir des résultats d'association d'une variante, nous avons progressé jusqu'à l'identification d'un modèle fonctionnel dans lequel nous pourrions mesurer son effet, ce qui nous éclairera un peu plus sur la pathogenèse des MII.

RÉFÉRENCES

1. Consortium, U.I.G., *Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region*. Nat Genet, 2009. **41**(12): p. 1330-1334.
2. Podolsky, D.K., *Inflammatory Bowel Disease*. New England J. Med., 2002. **347**(6): p. 417-429.
3. Friedman, S.e.B., R.S., *Harrison's Principles of Internal Medicine* A. Fauci, Editor 2008, McGraw-Hill. p. 1886-1898.
4. Budarf, M.L., et al., *GWA studies: rewriting the story of IBD*. Trends in Genetics, 2009. **25**(3): p. 137-146.
5. Bäckhed, F., et al., *Host-Bacterial Mutualism in the Human Intestine*. Science, 2005. **307**(5717): p. 1915-1920.
6. Murai, M., et al., *Regulatory T-cell stability and plasticity in mucosal and systemic immune systems*. Mucosal Immunol, 2010. **3**(5): p. 443-449.
7. Heller, F., et al., *Interleukin-13 Is the Key Effector Th2 Cytokine in Ulcerative Colitis That Affects Epithelial Tight Junctions, Apoptosis, and Cell Restitution*. Gastroenterology, 2005. **129**(2): p. 550-564.
8. Brand, S., *Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease*. Gut, 2009. **58**(8): p. 1152-1167.
9. Loftus, E.V.J., *Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences*. Gastroenterology, 2004. **126**(6): p. 1504-1517.
10. Rook, G.A.W. and L.R. Brunet, *Microbes, immunoregulation, and the gut*. Gut, 2005. **54**(3): p. 317-320.
11. Maeda, Y., et al., *The Failure of Oral Tolerance Induction is Functionally Coupled to the Absence of T Cells in Peyer's Patches under Germfree Conditions*. Immunobiol., 2001. **204**(4): p. 442-457.
12. Karlinger, K., et al., *The epidemiology and the pathogenesis of inflammatory bowel disease*. European Journal of Radiology, 2000. **35**(3): p. 154-167.
13. Binder, V., *Genetic Epidemiology in Inflammatory Bowel Disease*. Digestive Diseases, 1998. **16**(6): p. 351-355.
14. Consortium, H.G.S.I., *Finishing the euchromatic sequence of the human genome*. Nature, 2004. **431**(7011): p. 931-945.
15. Consortium, T.I.H., *The International HapMap Project*. Nature, 2003. **426**(6968): p. 789-796.
16. Ziegler, A.e.K., I.R., *A statistical approach to genetic epidemiology*. 2ème édition ed2007: Wiley-Blackwell Inc.
17. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing*. PNAS, 2009. **106**(45): p. 19096-19101.

18. Chen, R., et al., *Non-Synonymous and Synonymous Coding SNPs Show Similar Likelihood and Effect Size of Human Disease Association*. PLoS ONE, 2010. **5**(10): p. e13574.
19. Boks, M.P.M., et al., *Investigating gene–environment interaction in complex diseases: increasing power by selective sampling for environmental exposure*. International Journal of Epidemiology, 2007. **36**(6): p. 1363-1369.
20. Sing, C.F., J.H. Stengard, and S.L.R. Kardia, *Genes, Environment, and Cardiovascular Disease*. Arterioscler Thromb Vasc Biol, 2003. **23**(7): p. 1190-1196.
21. Cho, J.H., et al., *Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: Evidence for epistasis between 1p and IBD1*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(13): p. 7502-7507.
22. Duerr, R.H., et al., *High-Density Genome Scan in Crohn Disease Shows Confirmed Linkage to Chromosome 14q11-12*. American journal of human genetics, 2000. **66**(6): p. 1857-1862.
23. Hampe, J., et al., *A Genomewide Analysis Provides Evidence for Novel Linkages in Inflammatory Bowel Disease in a Large European Cohort*. The American Journal of Human Genetics, 1999. **64**(3): p. 808-816.
24. Hugot, J.-P., et al., *Mapping of a susceptibility locus for Crohn's disease on chromosome 16*. Nature, 1996. **379**(6568): p. 821-823.
25. Ma, Y., et al., *A genome-wide search identifies potential new susceptibility loci for Crohn's disease*. Inflammatory Bowel Diseases, 1999. **5**(4): p. 271-278.
26. Paavola, P., et al., *Genetic analysis in Finnish families with inflammatory bowel disease supports linkage to chromosome 3p21*. Eur. J. Hum. Genet. , 2001. **9**: p. 328 - 334
27. Rioux, J.D., et al., *Genomewide Search in Canadian Families with Inflammatory Bowel Disease Reveals Two Novel Susceptibility Loci*. The American Journal of Human Genetics, 2000. **66**(6): p. 1863-1870.
28. Satsangi, J., et al., *Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12*. Nat Genet, 1996. **14**(2): p. 199-202.
29. Hugot, J.-P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 599-603.
30. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 603-606.
31. Rioux, J.D., et al., *Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease*. Nat Genet, 2001. **29**(2): p. 223-228.
32. Anderson, C.A., et al., *Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47*. Nat Genet, 2011. **43**(3): p. 246-252.
33. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease*. Nat Genet, 2008. **40**(8): p. 955-962.

34. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci*. Nat Genet, 2010. **42**(12): p. 1118-1125.
35. McGovern, D.P.B., et al., *Genome-wide association identifies multiple ulcerative colitis susceptibility loci*. Nat Genet, 2010. **42**(4): p. 332-337.
36. Brest, P., et al., *A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease*. Nat Genet, 2011. **43**(3): p. 242-245.
37. Croucher, P.J.P., et al., *Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations*. Eur J Hum Genet, 2003. **11**(1): p. 6-16.
38. Cadwell, K., et al., *A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells*. Nature, 2008. **456**(7219): p. 259-263.
39. McCarroll, S.A., et al., *Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease*. Nat Genet, 2008. **40**(9): p. 1107-1112.
40. Shih, D. and S. Targan, *Insights into IBD pathogenesis*. Current Gastroenterology Reports, 2009. **11**(6): p. 473-480.
41. Singh, S.B., et al., *Human IRGM Induces Autophagy to Eliminate Intracellular Mycobacteria*. Science, 2006. **313**(5792): p. 1438-1441.
42. van Heel, D.A., et al., *Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease*. The Lancet, 2005. **365**(9473): p. 1794-1796.
43. Al-Sadi, R.M. and T.Y. Ma, *IL-1 β Causes an Increase in Intestinal Epithelial Tight Junction Permeability*. The Journal of Immunology, 2007. **178**(7): p. 4641-4649.
44. Girardin, S.E., et al., *Nod2 Is a General Sensor of Peptidoglycan through Muramyl Dipeptide (MDP) Detection*. Journal of Biological Chemistry, 2003. **278**(11): p. 8869-8872.
45. Inohara, N., et al., *Host Recognition of Bacterial Muramyl Dipeptide Mediated through NOD2*. Journal of Biological Chemistry, 2003. **278**(8): p. 5509-5512.
46. Smith, A.M., et al., *Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease*. The Journal of Experimental Medicine, 2009. **206**(9): p. 1883-1897.
47. Hampe, J., et al., *Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease*. Gut, 2001. **48**(2): p. 191-197.
48. Duerr, R.H., et al., *Evidence for an inflammatory bowel disease locus on chromosome 3p26: linkage, transmission/disequilibrium and partitioning of linkage*. Human Molecular Genetics, 2002. **11**(21): p. 2599-2606.
49. Consortium, W.T.C.C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls*. Nature, 2007. **447**(7145): p. 661-678.
50. Raelson, J.V., et al., *Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci*. Proceedings of the National Academy of Sciences, 2007. **104**(37): p. 14747-14752.

51. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility*. Nat Genet, 2007. **39**: p. 830-832.
52. Goyette, P., et al., *Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis*. Mucosal Immunol, 2008. **1**: p. 131-138.
53. Beckly, J.B., et al., *Two-stage candidate gene study of chromosome 3p demonstrates an association between nonsynonymous variants in the MST1R gene and Crohn's disease*. Inflammatory Bowel Diseases, 2008. **14**(4): p. 500-507.
54. Franke, A., et al., *Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis*. Nat Genet, 2008. **40**: p. 713-715.
55. Barrett, J.C., *Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes*. Nat. Genet., 2009. **41**: p. 703-707.
56. Fisher, S.A., *Genetic determinants of ulcerative colitis include the ECMI locus and five loci implicated in Crohn's disease*. Nat. Genet., 2008. **40**: p. 710-712.
57. Franke, A., et al., *Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL)*. Nat Genet, 2010. **42**(4): p. 292-294.
58. Imielinski, M., et al., *Common variants at five new loci associated with early-onset inflammatory bowel disease*. Nat Genet, 2009. **41**(12): p. 1335-1340.
59. Wang, M.H., A. Skeel, and E.J. Leonard, *Proteolytic cleavage and activation of pro-macrophage-stimulating protein by resident peritoneal macrophage membrane proteases*. The Journal of Clinical Investigation, 1996. **97**(3): p. 720-727.
60. Wang, M.H., et al., *Proteolytic conversion of single chain precursor macrophage-stimulating protein to a biologically active heterodimer by contact enzymes of the coagulation cascade*. Journal of Biological Chemistry, 1994. **269**(5): p. 3436-3440.
61. Yoshimura, T., et al., *Cloning, sequencing, and expression of human macrophage stimulating protein (MSP, MST1) confirms MSP as a member of the family of kringle proteins and locates the MSP gene on chromosome 3*. Journal of Biological Chemistry, 1993. **268**(21): p. 15461-15468.
62. Donate, L.E., et al., *Molecular evolution and domain structure of plasminogen-related growth factors (HGF/SF and HGF1/MSP)*. Protein Science, 1994. **3**(12): p. 2378-2394.
63. Danilkovitch, A., M. Miller, and E.J. Leonard, *Interaction of Macrophage-stimulating Protein with Its Receptor*. Journal of Biological Chemistry, 1999. **274**(42): p. 29937-29943.
64. Danilkovitch, A., A. Skeel, and E.J. Leonard, *Macrophage Stimulating Protein-Induced Epithelial Cell Adhesion Is Mediated by a PI3-K-Dependent, but FAK-Independent Mechanism*. Experimental Cell Research, 1999. **248**(2): p. 575-582.
65. Carafoli, F., et al., *Crystal structure of the β -chain of human hepatocyte growth factor-like/macrophage stimulating protein*. FEBS Journal, 2005. **272**(22): p. 5799-5807.

66. Goyette, P., et al., *Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis*. *Mucosal Immunol*, 2008. **1**(2): p. 131-138.
67. Giovanni, G., et al., *RON is a heterodimeric tyrosine kinase receptor activated by the HGF homologue MSP*. *EMBO J.*, 1994. **13**: p. 3524-3532.
68. Iwama, A., et al., *Molecular cloning of a novel receptor tyrosine kinase gene, STK, derived from enriched hematopoietic stem cells*. *Blood*, 1994. **83**(11): p. 3160-3169.
69. Ronsin, C., et al., *A novel putative receptor protein tyrosine kinase of the met family*. *Oncogene*, 1993. **8**: p. 1195-1202.
70. Nanney, L.B., et al., *Proteolytic Cleavage and Activation of pro-Macrophage-Stimulating Protein and Upregulation of its Receptor in Tissue Injury*. 1998. **111**(4): p. 573-581.
71. Sakamoto, O., et al., *Role of macrophage-stimulating protein and its receptor, RON tyrosine kinase, in ciliary motility*. *The Journal of Clinical Investigation*, 1997. **99**(4): p. 701-709.
72. Willett, C.G., et al., *Differential screening of a human chromosome 3 library identifies hepatocyte growth factor-like/macrophage-stimulating protein and its receptor in injured lung. Possible implications for neuroendocrine cell survival*. *The Journal of Clinical Investigation*, 1997. **99**(12): p. 2979-2991.
73. Chen, Y.-Q., et al., *Overexpression and Activation of the RON Receptor Tyrosine Kinase in a Panel of Human Colorectal Carcinoma Cell Lines*. *Experimental Cell Research*, 2000. **261**(1): p. 229-238.
74. Wang, M.-H., A.L. Kurtz, and Y.-Q. Chen, *Identification of a novel splicing product of the RON receptor tyrosine kinase in human colorectal carcinoma cells*. *Carcinogenesis*, 2000. **21**(8): p. 1507-1512.
75. Bork, P., et al., *Domains in plexins: links to integrins and transcription factors*. *Trends in Biochemical Sciences*, 1999. **24**(7): p. 261-263.
76. A Follenzi, S.B., P Gual, M C Stella, P Longati and P M Comoglio, *Cross-talk between the proto-oncogenes Met and Ron*. *Oncogene*, 2000. **19**: p. 3041-3049.
77. Herley, M.T., A.D. D'Andrea, and P.A. Ney, *Physical and functional interactions between the erythropoietin receptor and a truncated form of the STK/RON receptor tyrosinekinase*. *Blood*, 1999. **94**(supplement 1): p. 652a.
78. Mera, A., et al., *Induction of Cell Shape Changes through Activation of the Interleukin-3 Common β Chain Receptor by the RON Receptor-type Tyrosine Kinase*. *Journal of Biological Chemistry*, 1999. **274**(22): p. 15766-15774.
79. Iwama, A.Y.N., Suda T, *STK/RON receptor tyrosine kinase mediates both apoptotic and growth signals via the multifunctional docking site conserved among the HGF receptor family*. *EMBO J.*, 1996. **15**: p. 5866-5875.
80. Longati, P., Bardelli, A., Ponzetto, C., Naldini, L., & Comoglio, P., *Tyrosines1234-1235 are critical for activation of the tyrosine kinase encoded by the MET proto-oncogene (HGFreceptor)*. *Oncogene*, 1994. **9**: p. 49-57.
81. Ponzetto, C., et al., *A multifunctional docking site mediates signaling and transformation by the hepatocyte growth factor/scatter factor receptor family*. *Cell*, 1994. **77**(2): p. 261-271.

82. Danilkovitch, A., et al., *Two Independent Signaling Pathways Mediate the Antiapoptotic Action of Macrophage-Stimulating Protein on Epithelial Cells*. *Mol. Cell. Biol.*, 2000. **20**(6): p. 2218-2227.
83. Danilkovitch, A. and E. Leonard, *Kinases involved in MSP/RON signaling*. *Journal of Leukocyte Biology*, 1999. **65**(3): p. 345-348.
84. Skeel A, Y.T., Showalter SD, Tanaka S, Appella E & Leonard EJ., *Macrophage stimulating protein: purification, partial amino acid sequence, and cellular activity*. *J. Exp. Med.*, 1991. **173**: p. 1227-1234.
85. Chen, Y.-Q., J.H. Fisher, and M.-H. Wang, *Activation of the RON Receptor Tyrosine Kinase Inhibits Inducible Nitric Oxide Synthase (iNOS) Expression by Murine Peritoneal Exudate Macrophages: Phosphatidylinositol-3 Kinase Is Required for RON-Mediated Inhibition of iNOS Expression*. *The Journal of Immunology*, 1998. **161**(9): p. 4950-4959.
86. Liu, Q.-P., et al., *Negative Regulation of Macrophage Activation in Response to IFN- γ and Lipopolysaccharide by the STK/RON Receptor Tyrosine Kinase*. *The Journal of Immunology*, 1999. **163**(12): p. 6606-6613.
87. Wang, M.H., et al., *Macrophage-stimulating protein inhibits induction of nitric oxide production by endotoxin- or cytokine-stimulated mouse macrophages*. *Journal of Biological Chemistry*, 1994. **269**(19): p. 14027-14031.
88. **Morrison, A.C.**, et al., *Macrophage-stimulating protein, the ligand for the stem cell-derived tyrosine kinase/RON receptor tyrosine kinase, inhibits IL-12 production by primary peritoneal macrophages stimulated with IFN- γ and lipopolysaccharide*. *J. Immunol*, 2004. **172**: p. 1825-1832.
89. Zhou, Y.-Q., et al., *Activation of the RON Receptor Tyrosine Kinase by Macrophage-stimulating Protein Inhibits Inducible Cyclooxygenase-2 Expression in Murine Macrophages*. *Journal of Biological Chemistry*, 2002. **277**(41): p. 38104-38110.
90. Morrison, A.C. and P.H. Correll, *Activation of the Stem Cell-Derived Tyrosine Kinase/RON Receptor Tyrosine Kinase by Macrophage-Stimulating Protein Results in the Induction of Arginase Activity in Murine Peritoneal Macrophages*. *The Journal of Immunology*, 2002. **168**(2): p. 853-860.
91. Bhatt, A.S., et al., *Coordinate expression and functional profiling identify an extracellular proteolytic signaling pathway*. *Proceedings of the National Academy of Sciences*, 2007. **104**(14): p. 5771-5776.
92. Wang, M.H., Montero-Julian, F.A., Dauny, I., Leonard, E.J., *Requirement of phosphatidylinositol-3 kinase for epithelial cell migration activated by human macrophage stimulating protein*. *Oncogene*, 1996. **13**: p. 2167-2175.
93. Wang, M.-H., et al., *Macrophage-Stimulating Protein Induces Proliferation and Migration of Murine Keratinocytes*. *Experimental Cell Research*, 1996. **226**(1): p. 39-46.
94. Gallagher, S.R., *One-Dimensional SDS Gel Electrophoresis of Proteins*, in *Current protocols in molecular biology*, R.B. Frederick M. Ausubel, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and K. Struhl, Editors. 1999, John Wiley & Sons Inc. p. 10.2A.1-10.2A.34.

95. Friezner Decen, S.J., et al., *Structure of the human DIF15S1A locus; a chromosome 1 locus with 97% identity to the chromosome 3 gene coding for hepatocyte growth factor-like protein. Mitochondrial DNA*, 1998. **8**(6): p. 409-413.
96. van der Drift, P., et al., *Multiple MSP Pseudogenes in a Local Repeat Cluster on 1p36.2: An Expanding Genomic Graveyard?* Genomics, 1999. **62**(1): p. 74-81.
97. Brunelleschi, S., et al., *Macrophage Stimulating Protein (MSP) evokes superoxide anion production by human macrophages of different origin. British Journal of Pharmacology*, 2001. **134**(6): p. 1285-1295.
98. Albina, J., *On the expression of nitric oxide synthase by human macrophages. Why no NO?* Journal of Leukocyte Biology, 1995. **58**(6): p. 643-649.
99. Gantner, F., et al., *In vitro differentiation of human monocytes to macrophages: change of PDE profile and its relationship to suppression of tumour necrosis factor- α release by PDE inhibitors. British Journal of Pharmacology*, 1997. **121**(2): p. 221-231.
100. Herrera, R., *Modulation of hepatocyte growth factor-induced scattering of HT29 colon carcinoma cells. Involvement of the MAPK pathway. J Cell Sci*, 1998. **111**(8): p. 1039-1049.
101. Wahl, R.C., et al., *Mutation of Cys672 Allows Recombinant Expression of Activatable Macrophage-stimulating Protein. Journal of Biological Chemistry*, 1997. **272**(24): p. 15053-15056.
102. Kanayama, M., et al., *Hepatocyte growth factor promotes colonic epithelial regeneration via Akt signaling. American Journal of Physiology - Gastrointestinal and Liver Physiology*, 2007. **293**(1): p. G230-G239.
103. Tari, A.M. and G. Lopez-Berestein, *Serum predominantly activates MAPK and Akt kinases in EGFR- and ErbB2-over-expressing cells, respectively. International Journal of Cancer*, 2000. **86**(2): p. 295-297.
104. Smith, N.L., et al., *Association of Genome-Wide Variation With the Risk of Incident Heart Failure in Adults of European and African Ancestry / Clinical Perspective. Circulation: Cardiovascular Genetics*, 2010. **3**(3): p. 256-266.
105. Schneemann, M. and G. Schoeden, *Macrophage biology and immunology: man is not a mouse. J. Leukoc. Biol.*, 2007. **81**: p. 579.
106. Santoro, M.M., G. Gaudino, and P.C. Marchisio, *The MSP Receptor Regulates [alpha]6[beta]4 and [alpha]3[beta]1 Integrins via 14-3-3 Proteins in Keratinocyte Migration. Developmental Cell*, 2003. **5**(2): p. 257-271.
107. Wang, D., et al., *Activation of the RON receptor tyrosine kinase attenuates transforming growth factor- β 1-induced apoptotic death and promotes phenotypic changes in mouse intestinal epithelial cells. Carcinogenesis*, 2005. **26**(1): p. 27-36.
108. Wang, M.H., et al., *Altered expression of the RON receptor tyrosine kinase in various epithelial cancers and its contribution to tumourigenic phenotypes in thyroid cancer cells. The Journal of Pathology*, 2007. **213**(4): p. 402-411.
109. Kennedy, S.G., et al., *The PI3-kinase/AKT signaling pathway delivers an anti-apoptotic signal. Genes Dev.*, 1997. **11**: p. 701-713.

ANNEXE A

Détails des cohortes utilisées lors des génotypages de nos variantes

	Nom	Ethnie	Centre	Cohortes					
				Nbr. témoins	Nbr. Cas	Nbr. CU	Nbr. MC	Trios	Nbr. Total
1	CEU	américaine/européenne du nord et de l'ouest	Centre d'Étude du Polymorphisme Humain (HapMap)	94	0	0	0	0	94
2	ITACC	italienne	Hôpital CSS San Giovanni Rotondo (Italie)	401	784	632	152	0	1185
3	FAM	italienne et canadienne	Multiplés centres au Québec, Hôpital CSS San Giovanni Rotondo (Italie)	0	0	0	0	21 (CU), 153 (MC)	174 trios
5	ITACC	italienne	Hôpital CSS San Giovanni Rotondo (Italie)	1110	1647	1647	0	0	2757
6	ITACC	italienne	Hôpital CSS San Giovanni Rotondo (Italie)	1110	2446	1647	799	0	3556
7	Cedars	américaine caucasienne	Medical Genetics Institute, Cedars-Sinai Medical Center	2882	1648	723	925	0	4530
8	NIDDK	américaine caucasienne	National Institute of diabetes and digestive, kidney diseases IBD Genetics Consortium	2415	1942	1002	940	0	4357
9	SWEDEN	suédoise	Swedish UC GWA study organisation for study of IBD, Swedish national program for IBD	341	923	923	0	0	1264

****Les témoins et les cas de colite ulcéreuse sont les même pour les cohortes 5 et 6. Seulement 27 individus de la cohorte 2 ne sont pas dans la cohorte 6.***

ANNEXE B

Fichier d'alignement de la region 3p21 avec le genome humain à l'aide du script BLASTN de NCBI. Query = 3p21. Sbjct (1 et 2) = duplications 1 et 2 sur le chromosome 1. Les séquences soulignées en noirs montrent la position des amorces spécifiques à la séquence de MST1. Les autres séquences soulignées représentent les différents variants venant de nos expériences de re-séquençage et de la base de données dbSNP. Les séquences en majuscules représentent les séquences codantes.

BLASTN 2.2.19+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

RID: REA570A1015

Database: human build 36.3 reference assembly genomic scaffolds
368 sequences; 2,870,843,926 total letters
Query= 3 dna:chromosome chromosome:NCBI36:3:49681393:49726110:1
Length=44718

	Score
E	
Sequences producing significant alignments:	(Bits)
Value	
ref NT_022517.17 Hs3_22673 Homo sapiens chromosome 3 genomic ...	8.064e+04
0.0	
ref NT_004873.16 Hs1_5030 Homo sapiens chromosome 1 genomic c...	2.171e+04
0.0	
 >ref NT_004873.16 Hs1_5030 Homo sapiens chromosome 1 genomic contig, reference assembly	
Length=3518496	
 Features flanking this part of subject sequence:	
45085 bp at 5' side: hypothetical protein LOC55672	
22727 bp at 3' side: hypothetical protein	
 Score = 2.171e+04 bits (24080), Expect = 0.0	
Identities = 12746/13224 (96%), Gaps = 153/13224 (1%)	
Strand=Plus/Minus	

UTR non dupliquée

Ttttttttttttttttttttttacaggcataaagaggaaacatggcctttatgtctgacaag...

↓debut region alignee chr3:49696423 ←vers fin du gene mst1

49696391(-)

query 15032 ...AAGTTTTGTCTCCCAAGGCATATGGCATCAAGGCTGGGCCTAACCCAGTCTCATGACC
15091

|||||
sbjct 3369609 AAGTTTTGTCTCCCAAGGCATATGGCGTCAAGGCTGGGCCTAACCCAGTCTCATGACC
3369550

sbjct 3476524 AAGTTTTGTCTCCCAAGGCATATGGCGTCAAGGCTGGGCCTAACCCAGTCTCATGACC
3476583

AR696H_L Codon stop rs3197999
49696452A R689C
T A

query 15092 TTGTGAATCCAGTCCACAAACACAGAGACACGCGTGAAGACAGCTGGCCAGCGGACCTT
15151

|||||
sbjct 3369549 TTGTGAATCCAGTCCACAAACACAGAGACACGCGTGAAGACAGCTGGCCAGCGGACCTT
3369490

sbjct 3476584 TTGTGAATCCAGTCCACAAACACAGAGACACGCGTGAAGACAGCTGGCCAGCGGACCTT
3476643

rs1057319 E676K
I679R 49696575L675Q T669S C667F
C T T C A

query 15152 GCGCATACTCGGTTGGGGATTATAATTCCTTTCAGGCCAGCAGTTGTGGTAAAGCAG
15211

|||||
sbjct 3369489 GCGCATACTCGGTTGGGGATTCTAATTCCTTTCAGGCCAGCAGTTGTGGTAAAGCAG
3369430

sbjct 3476644 GTGCATACTCGGTTGGGGATTCTAATTCCTTTCAGGCCAGCAGTTGTGGTAAAGCAG
3476703

↓début exon 18 mst1

P664L G659Q
A T

query 15212 GCAAGTGGGCCCCGTAGTCACCctggcaggtaggagaactgatgagggccccgggccac
15271

|||||
sbjct 3369429 GCAAGTGGGCCCCGTAGTCACCctggcaggtaggagaactgatgagggccccgggccac
3369370

sbjct 3476704 GCAAGTGGGCCCCGTAGTCACCctggcaggtaggagaactgatgagggccccgggccac
3476763

A

query 15272 agcaatgactagcctgcttcatgataaaatagttcatttctagcccccatacccttcca
15331

|||
sbjct 3369369 agccatgactagcctgcttcatgataaaatagttcatttctagcccccatacccttcca
3369310

sbjct 3476764 agcaatgactagcctgcttcatgataaaatagttcatttctagcccccatacccttcca
3476823

```

query 15332   gggctggcccagggcctgccaccaacCTCACAGGCCCCACAGGGGCCAACAGTCCCTC
15391
|||||
sbjct 3369309 gggctggcccagggcctgccaccaacCTCACAGGCCCCACAGGGGCCAACAGTCCCTC
3369250
sbjct 3476824 gggctggcccagggcctgccaccaacCTCACAGGCCCCACAGGGGCCAACAGTCCCTC
3476883
          rs13085791proxy de rs3197999   R637STOP
rs17854561
          R641R   A   S628S
query 15392   AGTGCACATCTCACTCTCCGCACACGTCTCGGTGCTTGATGTTACTCTCGTTGCA
15451
|||||
sbjct 3369249 AGTGCACATCTCGCTCTCCGCACATGTCTCGGTGCTTGATGTTACTCTCGTTGGA
3369190
sbjct 3476884 AGTGCACATCTCGCTCTCCGCACATGTCTCGGTGCTTGATGTTACTCTCGTTGGA
3476943

          début exon 17 mst1↓
query 15452   GATGACATTCAGCAAGGCCACATTTAGGACTGTGTCATTACCCGTACctgcagtgagggg
15511
|||||
sbjct 3369189 GATGACATTCAGCAAGGCCACATTTAGGACTGTGTCATTACCCGTACctgcagtgagggg
3369130
sbjct 3476944 GATGACATTCAGCAAGGCCACATTTAGGACTGTGTCATTACCCGTACctgcagtgagggg
3477003

query 15512   aatggggagagggagaggggtcctgcaggaagatccagggctgggcctcctggccaccagc
15571
|||||
sbjct 3369129 aatggggagaaggagacgggtcctggaggaagatccagggctgggcctcctggccaccagc
3369070
sbjct 3477004 aatggggagaaggagacgggtcctggaggaagatccagggctgggcctcctggccaccagc
3477063
          49696977   W607RA605VI604T
          A   G   A   G
query 15572   agtcctgtgcactatgctcttacCTTTGGTCTCACCCAGCCTGCAATCTCACACTTGGT
15631
|||||
sbjct 3369069 agtcctgtgcactatgctcttacCTTTGGTCTCACCCAGCCTGCAATCTCACACTTGGT
3369010
sbjct 3477064 agtcctgtgcactatgctcttacCTTTGGTCTCACCCAGCCTGCAATCTCACACTTGGT
3477123

          W593W   P590P   R583H
          T   C   T
query 15632   CCCTGGAGGCACCACATACCAATTCAGGGGGCAGGCAGATCAGGGCCACACGCTGGTTCAG
15691
|||||
sbjct 3369009 CCCTGGAGGCACCACATATCATTTCAGGGGGCAGGCAGATCAGGGCCACACGCTGGTTCAG
3368950
sbjct 3477124 CCCTGGAGGCACCACATACCAATTCAGGGGGCAGGCAGATCAGGGCCACACGCTGGTTCAG
3477183

          49697099
          début exon 16 mst1↓   A
query 15692   GGTACAGATctttagcaagaatgggggcactcagggctgaggccacaaggcccagccc
15751

```

|||
 sbjct 3368949 GGTACAGATcttta^gcaagaatgggggcaactcagggctctgaggccacaaggctcagccc
 3368890
 sbjct 3477184 GGTACAGATcttta^acaagaatgggggcaactcagggctctgaggccacaaggctcagccc
 3477243

 query 15752 cacctcacaccctcccaggttgtccacatacCTCTCCAGCTTGAGCAGGACAAGCTGGGA
 15811
 |||
 sbjct 3368889 cacctcacaccctcccaggttgtccacatacCTCTCCAGCTTGAGCAGGACAAGCTGAGA
 3368830
 sbjct 3477244 cacctcacatcctcccaggttgtccacatacCTCTCCAGCTTGAGCAGGACAAGCTGGGA
 3477303

 RS1057288 G551S
 V561L C
 query 15812 GCCTGAGGGCCCA^{AG}CATCTTGGCTACTGGGACCCGCTGTAG^{CC}CTGGCTCTCCATG
 15871
 |||
 sbjct 3368829 GCCTGAGGGCCCA^{CAC}CAGCATCTTGGCTACTGGGACCCGCTGTAGGCC^TGGCTCTCCATG
 3368770
 sbjct 3477304 GCCTGAGGGCCCA^{CAC}CAGCATCTTGGCTACTGGGACCCGCTGTAGGCC^TGGCTCTCCATG
 3477363

 RS6446287 G533C
 Q546Q T
 query 15872 TTGTGGGTTCTGGAACAGGGTGCCCAACCATACCTCATAG^{CC}CGTGAGAGGCATATGGct
 15931
 |||
 sbjct 3368769 TTGTGGGTTCTGGAACAGGGTGCCCAACCATACCTCATAG^{CC}CGTGAGAGGCATATGGct
 3368710
 sbjct 3477364 TTGTGGGTTCTGGAACAGGGTGCCCAACCATACCTCATAG^{CC}CGTGAGAGGCATATGGct
 3477423

 RS9837520
 query 15932 gggagagaagctctgctaggtcatttgtgactctcag^ttccattgccccaggctcacttg
 15991
 |||
 sbjct 3368709 gggagagaagctctgctaggtcatttgtgactctcag^tccgttgccccaggctcacttg
 3368650
 sbjct 3477424 gggagagaagctctgctaggtcatttgtgactctcag^tccgttgccccaggctcacttg
 3477483

 T T
 query 15992 ttagcttgcctgggaa^gggggaaggtgggatgagactgggtccccaacacaaggagg
 16051
 |||
 sbjct 3368649 ttagcttgcctgggaa^{ag}ggggaaggtgggatgagactgggtccccaacacaaggagg
 3368590
 sbjct 3477484 ttagcttgcctgggaa^{ag}ggggaaggtgggatgagactgggtccccaacacaaggagg
 3477543

 query 16052 ctcacCAGGAGGAGAAGCACTGCCGGGCAGTCAGTATCCACTGCTCCTTCACTAGAGACC
 16111
 |||
 sbjct 3368589 ctcacCAGGAGGAGAAGCACTGCCGGGCAGTCAGTATCCACTGCTCCTTCACTAGAGACC
 3368530
 sbjct 3477544 ctcacCAGGAGGAGAAGCACTGCCGGGCAGTCAGTATCCACTGCTCCTTCACTAGAGACC
 3477603

↓début exon 14 mst1

```

query 16112 CCCC GCAGAAATGCTGGCCCTGCctagaggagtggggaattaggacaggtaacagactcc
16171
|||||
sbjct 3368529 CCCC GCAGAAATGCTGGCCCTGCctagaggagtggggaattaggacagggaacagactcc
3368470
sbjct 3477604 CCGCGCAGAAATGCTGGCCCTGCctagaggagtggggaattaggacagggaacagactcc
3477663

query 16172 tgggacagatgctagacctgccatcttctggctaggacctctgggggcagggatagattc
16231
|||||
sbjct 3368469 tgggacagatgctagacctgccatcttctggctaggacctctgggggcagggatagattc
3368410
sbjct 3477664 tgggacagatgctagacctgccatcttctggctaggacctctgggggcagggatagattc
3477723

query 16232 ccagccccagtgggataaaccacagaggacacaacctcagctcctctctgtgggagacag
16291
|||||
sbjct 3368409 ccagccccagtgggataaaccacagaggacacaacctcagctcctctctgtgggagacag
3368350
sbjct 3477724 ccagccccagtgggataaaccacagaggacacaacctcagctcctctctgtgggagacag
3477783

query 16292 gcagttgtgcctcacCGATTCCGCAAGCTGACTGTCCAGGGTGAGTTGCCCGGATGGCCC
16351
|||||
sbjct 3368349 gcagttgtgcctcacCAATTCCGCAAGCTGACTGTCCAGGGTGAGTTGCCCGGATGGCCC
3368290
sbjct 3477784 gcagttgtgcctcacCGATTCCGCAAGCTGACTGTCCAGGGTGAGTTGCCCGGATGGCCC
3477843

query 16352 RS3882325 V485AV4841 R478R RS3882323
CCAGCCACGCGCAGCTTGGAAACGCGCTGATCCAGCCGATCCACCCTCTTGCCACACTTC
16411
|||||
sbjct 3368289 CCAGCCACGCGCAGCTTGGAAACAACGCTGATCCAGCCGATCCACCCTCTTGCCACACTTC
3368230
sbjct 3477844 CCAGCCACGCGCAGCTTGGAAACGCGCTGATCCAGCCGATCCACCCTCTTGCCACACTTC
3477903

query 16412 TCAAACTGCACCTGGTctgtaggatgggggtgggctggatgaaaccagactgtgtggatg
16471
|||||
sbjct 3368229 TCAAACTGCACCTGGTctgtaggatgggggtgggctggatgaaaccagactgtgtggatg
3368170
sbjct 3477904 TCAAACTGCACCTGGTctgtaggatgggggtgggctggatgaaaccagactgtgtggatg
3477963

query 16472 RS11393273 -458D
tcgtgggctaagggcctgaccataactggcccaactcctaacCTGGGGGGTCCAGGA
16530
|||||
sbjct 3368169 tcgtgggctaagggcctgaccataactggcccaactcctaacCTGGGGGGTCCAGGA
3368110

```

sbjct 3477964 tcgtgggctaaggcctgaccataactggcccaactcgtaacCTGGGGGGTCCAGGA
3478023

↓début exon 12 mst1

query 16531 tTGATGGCGGCTGGTCATCAGctgaaagacaaagttcactgggggtaaggagccagcct
16590

|||||
sbjct 3368109 TTGATGGCGGCTGGTCATCAGctgaaagacaaagttcactgcggttaaggagccagcct
3368050

sbjct 3478024 TTGATGGCGGCTGGTCATCAGctgaaagacaaagttcactgggggtaaggagccagcct
3478083

T

query 16591 ttgg--gtgggggctgaggcagggcatgggggaagcgtcactagtgtcacCGCAGCGT
16648

|||||
sbjct 3368049 ttggtggtgaggctgaggcagggcatggggcaagcgtcactagtgtcacCGCAGCGT
3367990

sbjct 3478084 ttggtggtgaggctgaggtaaggtcatggggcaagcgtcactagtgtcacCGCAGCGT
3478143

D455Y rs4525916

query 16649 CGCAGGGCACAGTAGTGAATGGGGTCCCTGGGTCCATCGTGTAGCACCAGGGCCCATGG
16708

|||||
sbjct 3367989 CGCAGGGCACAGTAGTGAATGGGGTCCCTGGGTCCATCGTGTAGCACCAGGGCCCATGG
3367930

sbjct 3478144 CGCAGGGCACAGTAGTGAATGGGGTCCCTGGGTCCATCGTGTAGCACCAGGGCCCATGG
3478203

RS11288337 RS11295598

P410P RS9713630

query 16709 CTATCCCCATCTGGT-CTGGCAGAAGTTCTCCTCCAGTTGTGCATGCGGTTTCGGAGGTA
16768

|||||
sbjct 3367929 CTATCCCCATCTGGGT-CTGGCAGAAGTTCTCCTCCAGTTGTGCATGCGGTTTCGGAGGTA
3367871

sbjct 3478204 CTATCCCCATCTGGGT-CTGGCAGAAGTTCTCCTCCAGTTGTGCATGCGGTTTCGGAGGTA
3478262

P405R rs12054095 ↓début exon 11 mst1

query 16769 AATGTGAACctaggcggagcgggagcaaaatcgtggcagggtagtctcaaccatttcca
16828

|||||
sbjct 3367870 AATGTGAACcagggcgggagcgggagcgaagtctggcagggtagtctcaaccatttcca
3367811

sbjct 3478263 AATGTGAACcagggcgggagcgggagcaaaatcgtggcagggtagtctcaaccatttcca
3478322

ggctctggtcccagacaccaaagcatgccgcccagggttagggccctggcggggccggg
16888

|||||
sbjct 3367810 ggctctggtcccagacatcaaagcatgccgcccagggttagggccctggcggggccggg
3367751

sbjct 3478323 ggctctggtcccagacatcaaagcatgccgcccagggttagggccctggcggggccggg
3478382

P402P

49698328

RS12054096

A

P400H

W394G RS2985131

A

query 16889 agcaccagggactca**g**TG**CGG**CTT**CG**CGCGTCTCAGCGGA**CA**GC**G**CTGGCACTGGAc
 16948
 |||||
 sbjct 3367750 agcaccagggattca**c**TG**CGG**CTT**GT**GCGGCGTCTCAGCGGACC**CGCG**CTGGCACTGGAC
 3367691
 sbjct 3478383 agcaccagggactca**c**TG**CA**GCTT**GT**GCGGCGTCTCAGCGGACC**AGCG**CTGGCACTGGAC
 3478442
 T385T début exon 10 mst1↓
 A R379R P375C A374A
 query 16949 ACCCTTGC**GGT**CTT**G**CTGACCGTGC**CG**GTACTGCT**CCCC**CGCGCCGTGGTAGCAGTc
 17008
 |||||
 sbjct 3367690 ACCCTTGC**GGT**CTT**G**CTGACCGTGC**CG**GTACTGCT**CCCC**CGCGCCGTGGTAGCAGTc
 3367631
 sbjct 3478443 ACCCTTGC**GGT**CTT**G**CTGACCGTGC**CG**GTACTGCT**CCCC**CGCGCCGTGGTAGCAGTc
 3478502
 RS4052581 RS4052582 RS4052583
 query 17009 t**g**gg**cg**gg**tg**cg**gg**cagccatcaggccgagacctcgccccggccctccggttccaggct
 17068
 |||||
 sbjct 3367630 t**g**gg**cg**gg**tg**cg**gg**cagccatcaggccgagac**ccc**gccccggccctccggttccaggct
 3367571
 sbjct 3478503 t**g**gg**cg**gg**tg**cg**gg**cagccatcaggccgagacctcgccccggccctccggttccaggct
 3478562
 query 17069 tccagccc-ggctctgtagcccccaagcttgggcctcac**CCTGGGGCCGCACGTCGTCTG**
 17127
 |||||
 sbjct 3367570 tccagccccggctctgtagcccccaagcttgggcctcac**CCTGGGGCCGCACGTCGTCTG**
 3367511
 sbjct 3478563 tccagccccggctctgtagcccccaagcttgggcctcac**CCTGGGGCCGCACGTCGTCTG**
 3478622
 RS2087732 rs16854791
 A353G A352V M350T
 CRS4052562
 query 17128 TACAACGCCGATCTGGTAGCAA**AGGC**CG**CG**CATGCCGGGCCGCAGGTGAAGCACC
 17187
 |||||
 sbjct 3367510 TACAACGCCGATCTGGTAGCAA**AGCC**ACGCG**CA**TGCCGGGCCGCAGGTGAAGCACC
 3367451
 sbjct 3478623 TACAACGCCGATCTGGTAGCAA**AGCC**ACGCG**CG**TGCCGGGCCGCAGGGTGAAGCACC
 3478682
 début exon 9 mst1↓
 début exon 9 mst1↓
 P335L P331R L327P
 A rs3887843 G
 query 17188 AGGGCGCCTCTGAGCCGTC**GGG**GTTCCGGCA**A**AGTTCTCCCGA**A**AGTCTctaagcaggc
 17247
 |||||
 sbjct 3367450 AGGGCGCCTCTGAGCCGTC**GGG**GTTCCGGCA**G**AAGTTCTCCCGA**A**AGTCTctaagcaggc
 3367391
 sbjct 3478683 AGGGCGCCTCTGAGCCGTC**G**AGGTTCCAGC**G**AAGTTCTCCCGA**A**AGTCTctaagcaggc
 3478742

03P49698643


```

sbjct 3367035 tgtacggtactccacgggatatgctctcaggtcacgcccagcccctctgacctccccggc
3366976
sbjct 3479101 tgtacggtactccacgggatatgctctcaggtcacgcccagcccctctgacctccccggc
3479160

query 17668 caagccacgcccctcccaaggttcccaggtaccctcccaggcctggtccccgccccta
17727
|||||
sbjct 3366975 caagccacgcccctcccaaggttcccaggtaccctcccaggcctggtccccgccccta
3366916
sbjct 3479161 caagccacgcccctcccaaggttcccaggtaccctcccaggcctggtccccgccccta
3479220

query 17728 cCGCAGCGGGGAGGTCACAGAACTCTCGCTCGATCTGCGGATCCGTAGTGTAGCACCAT
17787
|||||
sbjct 3366915 cCGCAGCGGGGAGGTCACAGAACTCTCGCTCGATCTGCGGATCCGTAGTGTAGCACCAT
3366856
sbjct 3479221 cCGCAGCGGGGAGGTCACAGAACTCTCGCTCGATCTGCGGATCCGTAGTGTAGCACCAT
3479280

query 17788 GGCCGCTCGGAGCCGTCAGGATTCGGCAATAGTTGTCGTCCAGACCTTGGTCGAGGAAC
17847
E247Q_D N242S N238S D232N
G G C C T
|||||
sbjct 3366855 GGCCGCTCGGAGCCGTCAGGATTCGGCAATAGTTGTCGTCCAGACCTTGGTCGAGGAAC
3366796
sbjct 3479281 GGCCGCTCGGAGCCGTCAGGATTCGGCAATAGTTGTCGTCCAGACCTTGGTCGAGGAAC
3479340

query 17848 ctgggggcgtaagggcggtgaacaagaccctgggactctggcttatctggccccgcc
17907
03p49699253
49699232 49699242 49699253
|||||
sbjct 3366795 ctgggggcgtaagggcggtgaagaagaccctgggactctggcttatctggccccgcc
3366736
sbjct 3479341 ctgggggcgtaagggcggtgaagaagaccctgggactctggcttatctggccccgcc
3479400

query 17908 aattgccctacacggagccctgcccctggagtctggaccttcctagcccggccccag
17967
|||||
sbjct 3366735 agttgccctacacggagccctgcccctggagtctggaccttcctagcccggccccag
3366676
sbjct 3479401 agttgccctacacggagccctgcccctggagtctggaccttcctagcccggccccag
3479460

query 17968 gacgccgataccgctacgcgtactTGCCCGCTCGAAGGGGTGCTGGTGGGGTGCTGA
18027
G228S
T

```

```
|||||
sbjct 3366675 ggcgccgataaccgcctacgcgtacTTGCCCGGCTCGAAGGGGTGCTGGTGC GGGTGCTGA
3366616
sbjct 3479461 ggcgccgataaccgcctacgcgtacTTGCCCGGCTCGAAGGGGTGCTGGTGC GGGTGCTGA
3479520

                                T206T                                G201D_R
                                |                                TG
query 18028 AGATCCCAGCGCTGGCACTCGCGCCCTGACTCGGTGCGGTCTACCGCGCCGCGGTATTCC
18087

|||||
sbjct 3366615 AGATCCCAGCGCTGGCACTCGCGCCCTGACTCGGTGCGGTCTACCGCGCCGCGGTATTCC
3366556
sbjct 3479521 AGATCCCAGCGCTGGCACTCGCGCCCTGACTCGGTGCGGTCTACCGCGCCGCGGTATTCC
3479580

                    ↓début exon 6 mst1
query 18088 TCGCCATTGCACCAGACACACGCGGctggagacaaagagccagtggttcctggatggac
18147

|||||
sbjct 3366555 TCGCCATTGCACCAGACACACGCGGctggagacaaagagccagtggttcctggatgggc
3366496
sbjct 3479581 TCGCCATTGCACCAGACACACGCGGctggagacaaagagccagtggttcctggatgggc
3479640

                                E188V
query 18148 gtgggcttgccctccactctcccagcttgaccggcgccgcttacCTCCGGCAGGAT
18207

|||||
sbjct 3366495 gtgggcttgccctccactctcccagcttgaccggcgccgcttacCCAACCGGCAGGAT
3366436
sbjct 3479641 gtgggcttgccctccactctcccagcttgaccggcgccgcttacCCAACCGGCAGGAT
3479700

                                Y170H
query 18208 TTGATGCCGAGCTCTGGAAGCGCACAGCAGGGTCTGTTGTCTAGCACCAAGGACCTCCG
18267

|||||
sbjct 3366435 TTGATGCCGAGCTCTGGAAGCGCACGGCAGGGTCTGTTGTGTGGCACCAAGGACCTCCG
3366376
sbjct 3479701 TTGATGCTGCAGCTCTGGAAGCGCACGGCAGGGTCTGTTGTGTGGCACCAAGGACCTCCG
3479760

                                T147T T145M
query 18268 GGGTCGCCATCAGGGTTACGGCAGAAGTTCTCTCCAGGCCATTCCGGAGAGGCGGTG
18327

|||||
sbjct 3366375 GGGTCGCCATCAGGGTTACGGCAGAAGTTCTCTCCAGGCCATTCCGGAGCGTGGGCATG
3366316
sbjct 3479761 GGGTCGCCATCAGGGTTACGGCAGAAGTTCTCTCCAGGCCATTCCGGAGCGTGGGCATG
3479820

                    ↓début exon 5 mst1
query 18328 TACctgagggcccagagcatcactatagtgtgtgctgggggaaggtcccaggccgggacg
18387

|||||
sbjct 3366315 TACctgagggcccagagcatcactatagtgtgtgctgggggaaggtcccaggccgggacg
3366256
sbjct 3479821 TACctgagggcccagagcatcactatagtgtgtgctgggggaaggtcccaggccgggacg
3479880

                                P139P
```

```

query 18388 gaggggaaggtgtttgtctcacTTGTGATCATTGAGGAACTTGTGGCTCCAAGCCTGGCA
18447
|||||
sbjct 3366255 gaggggaaggtgtttgtcccacTGGTGATCGTTCGGGAACTTGTGGCTCCAAGCCTGGCAG
3366196
sbjct 3479881 gaggggaaggtgtttgtctcacTTGTGATCGTTCGGGAACTTGTGGCTCCAAGCCTGGCAG
3479940

P130S R119Q
RS41291712 T
query 18448 GACAGGCCACCCACGGTCGTGGCCATGGTGCCCGGTACCCAACCCCATTTGTTTCATGATG
18507
|||||
sbjct 3366195 GACAGGCCACCCACGGTCGTGGCCATGGTGCCCGGTACCCAACCCCATTTGTTTCATGATG
3366136
sbjct 3479941 GACAGGCCACCCACGGTCGTGGCCATGGTGCCCGGTACCCAACCCCATTTGTTTCATGATG
3480000

TV107I ↓début exon 4 mst1
query 18508 CAGGTCCGTACGTAGTctgggagcaagagacagaagatcaacttgggctgaggtcccctg
18567
|||||
sbjct 3366135 CAGGTCCGTATGTAGTctgggagcaagagacagaagatcaacttgggctgaggtcccctg
3366076
sbjct 3480001 CAGGTCCGTATGTAGTctgggagcaagagacagaagatcaacttgggctgaggtcccctg
3480060

A
query 18568 tctcccaccctgcccctctccacccccacttgcCTTTCTTCTGGAAGAGGTCACAGCGCC
18627
|||||
sbjct 3366075 tctcccaccctgcccctctccacccccactcgcCTTTCTCTGGAAGAGGTCACAGCGCC
3366016
sbjct 3480061 tctcccaccctgcccctctccacccccactcgcCTTTCTCTGGAAGAGGTCACAGCGCC
3480120

query 18628 CAGAACGCCGCAGCCTCGTGTGGGGCGAGTGTTGAGTCCATGGCAGCAGTTGGCAACCAT
18687
|||||
sbjct 3366015 CAGAATGCCGCAGCCTCGTGTGGGGTGAAGTGTGAGTCCATGGCAGCAGTTGGCAACCAT
3365956
sbjct 3480121 CAGAATGCCACAGCCTTGAGTGGGGCGAGTGTTGAGTCCATGGCAGCAGTTGGCAACCAT
3480180

H76D C ↓début exon 3 mst1
query 18688 GGCTGCTCACGTTGTAGTGAAGGCCctggagagaagaaggcacaaggtaacgccacagc
18747
|||||
sbjct 3365955 GGCTGCTCACATTGTAGTGAACGCCctggagagaagaaggcacaaggtaacgccacggc
3365896
sbjct 3480181 GGCTGCTCACATTGTAGTGAACGCCctggagagaagaaggcacaaggtaacgccacggc
3480240

M64T G
query 18748 ccaggcttccctgccccagctttatctaggcccagtgggcactcacCGGCAGTCCATTA
18807
|||||
sbjct 3365895 ccaggctcccctgccccagctttatctatgcccagtgggcactcacCGGCAGTCCATTA
3365836

```

sbjct 3480241 ccaggctcccctgccccagtcttatctaggcccagtggccactcacCAGCAGTCC**ATTA**
 3480300

P62L R59H
 A T
 query 18808 **AGG**CCCACAG**CGC**ACCAGCACACTCTTCAGCATCTGCCACATCCTCCTGCCAAGGCCCGG
 18867

|||
 sbjct 3365835 **AGG**CCCACAG**CGC**ACCAGCACACTCTTCAGCATCTGCCACATCCTCCTGCCAAGGCCCGG
 3365776

sbjct 3480301 **AGG**CCCACAG**CGC**ACCAGCACACTCTTCAGCATCTGCCACATCCTCCTGCCAAGGCCCGG
 3480360

H39Q L37- Q27STOP
 T A
 query 18868 GCACCACCGC**ATG**TAG**TCG**GTGCTGTAGCTCTGTGCCCGGAGCAC**TTG**GAAGTCATTCA
 18927

|||
 sbjct 3365775 GCACCACCGC**T**TGTAGC-GGCGCTGTAGCTCTGTGCCCGGAGCAC**T**AGAAGTCATTCA
 3365717

sbjct 3480361 GCACCACCGC**A**TGTAGC**A**GGTGCTGTAGCTCTGTGCCCGGAGCAC**T**GGAAGTCATT**C**G
 3480420

|début exon 2 mst1
 query 18928 **ATGGCGAGCGCTGCC**Cctgcagagtaggcatgagtggtgaggtcaggtgggcatacatg
 18987

|||
 sbjct 3365716 **ATGGCGAGCGCTGCC**Cctgca-----gagtggtgaggtcaggtgggcatacatg
 3365667

sbjct 3480421 **ATGGCGAGCGCTGCC**Cctgcagagtggtgagtggtgaggtcaggtgggcatacatg
 3480480

query 18988 tcagtaatgtgtattggcatgtccacactttgttcattcaggggatcaaagctacaaggc
 19047

|||
 sbjct 3365666 tcagtaatgtgtattggcatgtccacagttgttcattcaggggatcaaagctacaaggc
 3365607

sbjct 3480481 tcagtaatgtgtattggcatgtccacactttgttcattcaggggatcaaagctacaaggc
 3480540

query 19048 ttctgggatggaccctgtatgcactttcaagggccagtctag-ccccctgcacagatact
 19106

|||
 sbjct 3365606 ttctgggatggaccctgtatgcactttcaagggccagtctagccccctgcacagatact
 3365547

sbjct 3480541 ttctgggatggaccctgtatgcactttcaagggccagtctagccccctgcacagatact
 3480600

query 19107 tgtgaaaaaatttcccctgggaagcaggcccagacttggtagttatcaccggtgcctct
 19166

|||
 sbjct 3365546 tgt-caaaaaatttcccctgggaagcaggcccagacttggtagttatcaccggtgcctct
 3365488

sbjct 3480601 tgtcaaaaaatttcccctgggaagcaggcccagacttggtagttatcaccggtgcctct
 3480660

A
 query 19167 gtt**tagtggccaggcaccg**ggctcagatctaacacatcagctctgtgagagcagtggt
 19226

```

||| |||
sbjct 3365487 |gtgtagtgccagccaccgggctcagatctaacacacacgctctgtgagagcagtgggt
3365428
sbjct 3480661 |gtgtagtgccagccactgggctcagatctaacacatacagctctgtgagagcagtgggt
3480720

query 19227 |gatgaagcttgcccccactcatctctcaaagtgagaatgctaaggctcagagccatcacat
19286
|||
sbjct 3365427 |gatggagcttgccccatctcatctctcaaagtgagaatgctaaggctcagagccatcacat
3365368
sbjct 3480721 |gatggagcttgccccatctcatctctcaaagtgagaatgctaaggctcagagccatcacat
3480780

query 19287 |taccagccaggggcctggctaggcattcagactccaaatctgggctctcacctgcaca
19346
|||
sbjct 3365367 |taccagccaggggcctggctaggcattcagactccaaagctgggctctcacctgcaca
3365308
sbjct 3480781 |taccagccaggggcctggctaggcattcagactccaaatctgggctctcacctgcaca
3480840

query 19347 |aaggcatatgctaggttagaggggtagatcaggctcaggaggggtcactgcctgctgtgt
19406
|||
sbjct 3365307 |aaggcatacgctaggttagaggggtagatcaggctcaggaggggtcactgcctgctgtgt
3365248
sbjct 3480841 |aaggcatacgctaggttagaggggtagatcaggctcaggaggggtcatttctgctgtgt
3480900

query 19407 |gcgtgcatctgtgtggtcctgacactgcttcagtgctagagcagacgtgctaatagaggc
19466
|||
sbjct 3365247 |gcgtgcatctgtgtggtcctaactgcttcagtgctagagcagacgtgctaatagaggc
3365188
sbjct 3480901 |gcgtgcatctgtgtggtcctaactgcttcagtgctagagcagacgtgctaatagaggc
3480960

query 19467 |ctaagtgggcaatgtctatgtgttcctagggttcccagctgtgctcaagaggcaaggt
19526
|||
sbjct 3365187 |ctaagtgggccatgtctatgtgttcctagggttcccagctgtgctcaagaggcaaggt
3365128
sbjct 3480961 |ctaagtgggccgtgtctgtgtgttcctagggttcccagctgtgctcaagaggcaaggt
3481020

query 19527 |cactgccccatgccactgagcctctggctccccgacttttttctcatcctagaatagga
19586
|||
sbjct 3365127 |cactgccccatgccactgagcctctggctccccgacttttttctcatcccagaatagga
3365068
sbjct 3481021 |cactgccccatgccactgagcctctggctccccgacttttttctcatcccagaatagga
3481080
```

```

query 19587  gaatggggccaaccccctcctgaaggcagatggggatcagggttgggggcacttacCAGG
19646
      |||||||
sbjct 3365067  gaatggggccaagcccctcctgaaggcagatggggatcaaggttgggggcactcacCAGG
3365008
sbjct 3481081  gaatagggccaaccccctcctgaaggcagatggggatcagggttgggggcactcacCAGG
3481140
                                     début du gene mst1(-) 49701099↓
                                     P5S
                                     A
                                     codon start
query 19647  GACCCCTAAGCATTGAGTCAGAAGCAGCAGGAGTGGGAGCCACCCCATCCTTCTGGCTGG
19706
      |||||||
sbjct 3365007  GACCCCTAAGCATTGAGTCAGAAGCAGCAGGAGTGGGAGCCACCCCATCCTTCTGGCTGG
3364948
sbjct 3481141  GACCCCTAAGTATTGAGTCAGAAGCAGCAGGAGTGGGAGCCACTC 3481186
                                     ↓début exon 1 mst1
query 19707  AGGCTGCACTGTgaccaccacagcccatccgggaagtgtgaaacctgtccctacggg
19766
      |||||||
sbjct 3364947  AGGCTGCACTGTgaccaccacagcccatccgggaagtgtgaaacctgtccctacggg
3364888

query 19767  attgggtggctctggctccacacgtcagctcaggcctgctggaccctgacctgagacct
19826
      |||||||
sbjct 3364887  attgggtggctctggctccgcacgtcagctcaggcctgctggaccctgacctgagacct
3364828

query 19827  ggtgacaggagccatgaggggcccaggcctcaggtcccacaggtcagttgcaagggcctag
19886
      |||||||
sbjct 3364827  ggtgacaggagccatgaggggcccaggcctcaggtcccacaggtcagttgcaagggcctag
3364768

query 19887  tacag-----cttagtgacagggtgtaggaaggttgggtg
19922
      |||
sbjct 3364767  cacagctagccccctgagaggcctccctacttagtggtcaggtgtaggaaggttgggtg
3364708

query 19923  gggacacttgaggtgccctgggggtgggggtgaaaccctctgcagcctagtacagccaag
19982
      |||
sbjct 3364707  gggccacttgaggtgccctgggggtgggggtgaaaccctctgcagcctagtacagccaag
3364648

query 19983  ggcattgtgaaagtgagagctgccagaggtctgggctccagcccctgggtgacattaac
20042
      |||||||
sbjct 3364647  ggcattgtgaaagtgacagctgccagaggtctgggctccagcccctgggtgacattaac
3364588

query 20043  ttttctgaatttcggtttttccctgtaaattggggaaaattacttggcccagcttcaaca
20102

```

```

|||||
sbjct 3364587 ttttctgaatttcggtttttccctgtaaattggggaaaattacttgccccagcttcaaca
3364528

query 20103 ctgcctcccctccacccttttgctgctgccttatcaggccccagctgtaggcaggtcagc
20162

sbjct 3364527 ctgcctcccctccacccttttgctgctgccttatcaggccccagctgtaggcaggtcagc
3364468

query 20163 ccacacccggcggcagagcccagagaggctcgctcatctcagccgccagaagtgtcatcc
20222

sbjct 3364467 ccacgcccggcggcagagcccagagaggctcgctcatctcagccgccagaagtgtcatcc
3364408

query 20223 caggggaaggtgagcagggctgtccactaggactaaataggggtggaggtaacaggtagca
20282

sbjct 3364407 caggggaaggtgagcgaagctgtccactaggactaaataggggtggaggtaacaggtagca
3364348

query 20283 gggactctggcaggaacggcctttcagacgttttaaggaccgtctgagcggggcctctaa
20342

sbjct 3364347 gggactctggcaggaaccgcctttcagacgttttcaggaccgtctgagcggggcctccaa
3364288

query 20343 gtatgggtaggagacagggcggcctagggattccggttagcatgtcgcgttctgccgcagc
20402

sbjct 3364287 gtatgggtaggagacagggcggcctagggattccggttagcacgtcgcgttctgccgcagc
3364228

query 20403 agggctgggagcatccccagagcagggcggaggtcggaggtggtgaggcctttcggcgaag
20462

sbjct 3364227 agggctgggagcatccccggggcagggcggaggtcggaggtggtgaggcctttcggcgaag
3364168

query 20463 ctgaggcctggaacaacccgggtgggaagccatggaggggggtcccctagcggaggctgggc
20522

sbjct 3364167 ctgaggcctggaacaacccgggtgggaagccacggaggggggtcccctagcggaggctgggc
3364108

query 20523 gcggggcagtgcgctgcgcggtggaggctcagcgcgcaagcgtgcgctgatttgcc
20582

sbjct 3364107 gcggggcgggtgcgctgcgcggtggaggctcagcgcgagcgtgcgctgatttgcc
3364048

query 20583 ctgaacggacgccgtagccaagaggttgggcggatgtgtgagccgggtcgctcgggccg
20642
|||||

```

↓début du gène rnf123 49701994(+)

```

sbjct 3364047 ctgaacggacgccgtagccgagaggttgggcggatgttgtgaaccgggtcgcgccggccg
3363988

query 20643 aggctccgggtgagtgaggggcccagggctctgagcctggcgggttccggctattccct
20702

sbjct 3363987 aggctcggggtgagtgaggggcccagggctctgagcctggcgggttccggctattccct
3363928

query 20703 tcgcggtccgggtagtcacgtgagggcacctgcagcgcagggatctgggtgcgaaac
20762

sbjct 3363927 tcgcggtccagtttagtccactgagggcacctgcagcgcagggatctgagtcgaaac
3363868

query 20763 tccgagcgaagctggtctggcaggtcctgtggagcttggggccaggcggggcgacttg
20822

sbjct 3363867 tccgagcgaagctggtctggcaggtcctgtggagcttggggccaggcggggcgacttg
3363808

query 20823 ggggagccatgcgcccccctcgccgtccttgggtcccagacggagtaggttcagttctg
20882

sbjct 3363807 ggggagcca-----tccttgggtcccagacggagtaggttcagttctg
3363764

query 20883 agtccggataccccggcggggttgacgggatagccgtgacagggagtccccggcgggg
20942

sbjct 3363763 agccccgataccccggcaggttgacgggatagccgtgacagggagtccccggcgggg
3363704

query 20943 acaaggtccgcctctcacctgggtcctcagcaactgactcccggagcccactccgggtgt
21002

sbjct 3363703 acaaggtccgcctctcacctgggtcctcagcaactgacacccggag-----tgggtgt
3363651

query 21003 ctggaaggtctgctcatcaaatgaatgactgagtttaatcagatcttcaactactctggt
21062

sbjct 3363650 ctggaaggtctgctcatcaaatgaatgactgcatttaatcagatcttcaactactctggt
3363591

query 21063 agtgaaggggaccgatcatggagcccaccgagggcccaggtcgggaaaatcaccttc
21122

sbjct 3363590 agtgaaggggaccgatcatggagggcccaccgagggcccaggtcgggaaaatcaccttc
3363531

query 21123 ccgaggagcaggcgtggtgcagacccacaagggttggcgggagtagagggtctgttt
21182

sbjct 3363530 ccgaggagcaggcgtggtgcagacccacaggggttggcgggagtagagggtctgttt
3363471

```



```

query 21183   ggggtgagataagagactgaactgcagagtagggagtgctggctaaggcagtgctgcaga
21242
          ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363470 ggggcgagataagagactgaacggcagagtagggagtgctggctaaggcagtgctgcaga
3363411
          ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21243   gcgaaactctagacctgcgctgtctaataagaactttatatgatggggaacgttctacc
21302
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363410 gcgaaactctagacctgcgctgtctaataagaactttatatgatggggaacgttctacc
3363351
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21303   tttgtgctgtccaacagagtaaccattagccacatgtggtcattgaactcttgaatgtg
21362
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363350 tttgtgctgtccaatagagtaaccgtagccacatgtggtcattgaactcttgaatgtg
3363291
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21363   gctaaggcaactgagaaatcgaattgttaatttttaaaactttaaatthaagtaaacctca
21422
          ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363290 gctgaggcaactgagaaatcgaattgttaatttttaaaactttaaatthaagtaaacctca
3363231
          ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21423   tgtggatgacggctgctgtgttggacagcaccgctacacatcagtgaggccagcaccagg
21482
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363230 cgtggatgacggctgctgtgttggacagcaccgctacacatcagtgaggccagcaccagg
3363171
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21483   gactcagtgaggttctggagcagctgcagcagcaacatgatgtgagggatagatgcact
21542
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363170 gactcagtgaggttttggagcagctgcagcagcaacatgatgtgagggatagatgcact
3363111
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21543   aggaggaccaatggagtttgccatgagggtgggcaagggtggtgcttggatgaggtggg
21602
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363110 aggaggaccaatggagtttgccatgagggtgggcaagggtggtgcttggatgaggtggg
3363051
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21603   aggccacatgtgtgagtgggaatggtgaaactatcaggaccatggataaaagaaggagg
21662
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3363050 aggccacatgtgtgagtgggaacagtgaaactatcaggaccatggataaaagaaggagg
3362991
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

query 21663   aaccagggtgaggaaagaggctccagccttgtggggagtaagtgggacttatggctag
21722
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
sbjct 3362990 aaccagggtgaggaaagaggctccagccttgtggggagtaagtgggacttatggctag
3362931
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

query 21723 agatgaccagaccagtgggtggtctgcacttgggggaatgtgtaggattgtcgcttctggt
21782
      |||
sbjct 3362930 agatgaccagaccagtgggtggtctgcacttgggggaatgtgtaggattgtcgcttctggt
3362871

query 21783 tgtgagctggactcagtatggcttacaactgagatggatacaggggctctgccctagga
21842
      |||
sbjct 3362870 tgtgagctggactcagtatggcttacaaccgagatgaatacaggggctctgccctagga
3362811

query 21843 cacctcactttggacattttggctctgatttccacagagctgagacattttgttatggttc
21902
      |||
sbjct 3362810 cacctcactttggacattttggctctgatttccacagagctgagacattttgttatggttt
3362751

query 21903 ggcaaaaatataatctagtagctctctgccagccctgtgctgggacagggacactg
21962
      |||
sbjct 3362750 ggcaaaaatataatctagtagctctctgccagccctgtgctgggacagggacactg
3362691

query 21963 gagaacaagaagtggccctgttacatgggcttagaccctggctaccctccctgtgtagg
22022
      |||
sbjct 3362690 gagaacaagaagtggccctgttacatgggcttagaccctggctaccctgcctgtgtagg
3362631

query 22023 aagctgggagcccctgcattcccctgcctatgtgttagtgtgcatgcccctggcctcac
22082
      |||
sbjct 3362630 aagctgggagcccctgcattcccctgcctatgtgttagtgtgcatgcccctggcctcac
3362571

query 22083 ctgctccaagccttctgcatggcaggactgaggtcaggtgtgggggcccctggctgg
22142
      |||
sbjct 3362570 ctgctccaagacttctgcatggcaggactgaggtcaggtgtgggggcccctggctgg
3362511

query 22143 ccctgagtagacaggctctgtgtctggctcagccccaggaccactggctgcccatgaga
22202
      |||
sbjct 3362510 ccctgagtagacaggctctgtgtctgcctcagcctccaggaccactggctgcccatgaga
3362451

query 22203 gatgaaggatggcatccaagggggccggcatgtctttctcccgaagagctataggctga
22262
      |||
sbjct 3362450 gacgaaggatggcatccaagggggccggcgtgtctttctcccgaagagctgtaggctga
3362391

query 22263 cctcagatgctgagaaatccaggggtcacaggttaagagctggcaggggaaggaaggaggctc
22322

```

```
|||||
sbjct 3362390 cctcagatgctgagaaatccagggtcacaggttaagggctggcagagaaggaaggaggctc
3362331

query 22323 ctcttggtgggaggagaaggaagggacaaaggcctcaggccttgctgcacctgtggctgg
22382

sbjct 3362330 ctcttggtgggaggagaaggaagggacaaaggcctcaggccttgctgcacctgtggctgg
3362271

query 22383 cctaggaccagaagtctctggggccagcagggccaggagctgcactgatcatgtggccc
22442

sbjct 3362270 cctaggaccagaagtctctggggccagcagggccaggagctgcactcatcgtgtggccc
3362211

query 22443 ttttctcccttctgacttggtggctcaggcattgtgcaggagaagctgctgaatgactacc
22502

sbjct 3362210 ttttctcccttctgacttggtggctcaggcattgtgcaggagaagctgctgaatgactacc
3362151

query 22503 tgaaccgcatcttttctcttctgaacatgcaccccagcagccaccagcaggtatggct
22562

sbjct 3362150 tgaaccgcatc-tttcctcttctgaccatgcaccccagcagccaccggcaggtatggct
3362092

query 22563 ggggtgggtggcccctcctcactctgctggggccagaccagaataaggaggctgcagttgta
22622

sbjct 3362091 ggggtgggcgcccctcctcactctgctggggccagagcagaataaggaggctgcagttgta
3362032

query 22623 gggaaaggagcctgggctctggacagataggcttaggctctgttctcactgtgctcctg
22682

sbjct 3362031 gggaaaggagcctgggctctggacagataggcttaggctctgttctcactgtgctcctg
3361972

query 22683 gctgcgagatctctggccagagagtctgccttccgtactctgcttcttctctgtctccagg
22742

sbjct 3361971 gctgcgagatctctggccagagagtctgccttccgtacgctgcttcttctctgtctccagg
3361912

query 22743 cagtgctagttccttctgagagt cagggagcagaggtcagggagcgggtgcacattcatg
22802

sbjct 3361911 cagtgctagttccttctgagagt cagggagcagaggtcagggagcgggtgcacattcatg
3361852

query 22803 cccagctagcacctggcacctactccatacccagcatctgccctgcatgtacctgcct
22862

|||||
```

sbjct 3361851 3361792 cccagctagcacctggcacctactccacaccagcatctgccctgcatgtacctgcct

query 22863 22922 caccttggccccagcctggccccttgtagcttctgactgaccttgtaagaacaaactcc

sbjct 3361791 3361732 caccttggccccagcctggccccttgtagcttctgactgaccttgtaagaacaaactcc

query 22923 22982 tccagggccttttctttacaagtttaataacttttaagtgtagtcagtataaaagttacaa

sbjct 3361731 3361672 tccagggccttttctttacaagtttaataacttttgagtgtagtcagtgtaaaagttacaa

query 22983 23042 gtacattttagcagaaacttagagggccaggcacagtggttcacacctgtaatcccagca

sbjct 3361671 3361612 gtacattttagcagaaacttagagggccaggcacagtggttcacacctgtaatcgcagca

query 23043 23102 ctttgggaggccaaggtgggtagattgcttgagtccaggagttcaagaccagcctgagca

sbjct 3361611 3361552 ctttgggaggccaaggtgggtagattgcttgagtccaggagttcaagaccagcctgagca

query 23103 23162 acatggcaagaccccatctctacaataaaatacagaaattacctgggcatggtggtgtgc

sbjct 3361551 3361492 acatggccagaccccatctctacaataaaatacagaaattacctgggcatggtggtgtgc

query 23163 23222 acctgtagtcccagctgttagggaggctgatgtgggaggatcacttgagcctaggaggtt

sbjct 3361491 3361432 acctgtagtcccagctgttagggaggctgatgtgggaggatcacttgagcctaggaggtt

query 23223 23282 gaggctgcagtgagttggcctctgcatgccactgcactgtagcctgggtgacagagcaag

sbjct 3361431 3361372 gaggctgcagtgagctggcctctgcatgccactgcactgtagcctgggtgacagagcaag

query 23283 23342 accctgtctctaaaaaaagaaggaaaagcacctgctatcttgttattttaacatgactat

sbjct 3361371 3361312 accctgtctctaaaaaaagaaggaaaagcacctgctatcttgttattttaacatgactat

query 23343 23401 tcaaatgtgattcttttctg-tttctttctttgtttacatttttaattttaattttg

sbjct 3361311 3361252 tcaaatgcgattcttttctgttttctttctttgtttacatttttaattttaattttg

```
query 23402 tgggtatgtagtaggtatatatatatttatagggtacatgagatactttgatacaggcatgc
23461
|||||
sbjct 3361251 tgggtatgtagcaggtatatgtatttatagggtatatgagatactttgacacaggcatgc
3361192
|||||

query 23462 aatgcgtaataatcacatcatggtaaatggggtatccatcccctcaagcatttatctttt
23521
|||||
sbjct 3361191 aatgcgtaataatcacatcatggtaaatggggtatccatcccctcaagcatttatctttt
3361132
|||||

query 23522 gtcttataaacaatctagttataactcttttagttatttttaaatgtacaattattattta
23581
||
sbjct 3361131 gtgttataaacaatctagttataactcttttagttatttttaaatgtacaattattattta
3361072
|||||

query 23582 ctatagtcactctgttctgctaccagataggctctcattaattcaattttttgtacgcat
23641
|||
sbjct 3361071 ctacagtcactctgttctgctaccagacaggctcttattaattcgattttttgtacgcat
3361012
|||||

query 23642 tagccatccccatctccccacaagccctatactacccttccaagcctctggtaaccatcc
23701
|||||
sbjct 3361011 tagccatccccatctccccacaagccctatactaccgttccaagcctctggtaaccatcc
3360952
|||||

query 23702 agtactttgcctctatgagttcaattgttttaatttttagctgtcacaataagtataa
23761
|||||
sbjct 3360951 agtactttgcctctatgagttcaattgttttaatttttagctgtcacaataagtataa
3360892
|||||

query 23762 catgcaatgttgtcttttctgcctggcttatttcatttaacataatgacctccagttc
23821
|||||
sbjct 3360891 catgcagtgttgtcttttctgcctggctcattcacttaacataatgacctccagttc
3360832
|||||

query 23822 catccatgttgttgcatattacaggatctcattctttttatggccgaatagtactcogt
23881
|||||
sbjct 3360831 catccatgttgttgcaattacaggatctcattctttttatggctgaatagtactctgt
3360772
|||||

query 23882 tgtgtatatgtaccatgttttctttatttattcatctgttgatggacacttagattgctt
23941
|||||
sbjct 3360771 tgtgtatctgtaccgtgttttctttatttattcatctgttgatggacacttagattgctt
3360712
|||||
```

query 23942 ccaaactctggctattgtgaacagtgctgcaacaaacatgggaatgcag-----
23990
|||||
sbjct 3360711 ccaaactctggctattgtgaacagtgctgcaacaaacatgggaatgcagatatctcttcg
3360652

query 23991 -----atatctctttgatataactgatttcccttcttt
24022
|||||
sbjct 3360651 atatactgatttcccttctttgaggtatataatctcttcgatataactgatttcccttcttt
3360592

query 24023 caggtatataaccagcagtgagggttgctggatcatatggtagctcaatTTTTtagtttagt
24082
|||||
sbjct 3360591 caggtatataaccagcagtgaggattgctggatcatatggtagctcaatTTTTtagtttagt
3360532

query 24083 tttttttttttgaaaaa--ttttttttttgagacggagtcttgctctggtgccaggc
24140
|||||
sbjct 3360531 ttttctttttgaaaaaattttttttttttgagacagagtcttgctctggtgcacaggc
3360472

query 24141 tggagtgcaggggcacaatctcagctcactgcaagctccgcctcccagggtcatgccatt
24200
|||||
sbjct 3360471 tggagcgcaggggcgcaatctcagctcactgcaagctccgcctcccagggtcacgcaatt
3360412

query 24201 cttctgcctcagcctcccaagtagctgggactgcaggtgcccgccaccatgccagctaa
24260
|||
sbjct 3360411 cttgggcctcagcctcccaagtagctgggactacaggcggcccgccaccagggtccagctaa
3360352

query 24261 ttttttgcatttttagtagagacgggggtttcaccgtgttagcaggatggtcttgatctc
24320
|||||
sbjct 3360351 ttttttgtcttttagtagagacgggggtttcaccgtgttagcaaggatggtcttgatctc
3360292

query 24321 ctgagctcgtgatccgcccacctcggcctcccaaagtgctgggattacaggcgtgagcca
24380
|||||
sbjct 3360291 ctgagctcgtgatccaccgcctcggcctcccaaagtgctgggattacaggcgtgagcca
3360232

query 24381 gtgtgcctggcc--tttttttttttttttttttttttttttttttgagatggagttttgctcttg
24438
|||||
sbjct 3360231 gtgtgcctggccttttttttttttttttttttttttttttttttggagaaggagttttgctcttg
3360172

query 24439 ttgcccagcctggagtgcagtgccatgatcttggtcactgcaacctctgcctcccaagt
24498

||||| |
sbjct 3360171 ttgccaggctggagtgcagtggcatgatcttggctcactgcaacctctgcctcccagat
3360112

query 24499 tctagtaattctcctgcctcagcctcccagatagctgggattacaggcaciaaccaccac
24558

||||| |
sbjct 3360111 tctagcaattctcctgcctcagcctcccgtgtagctgggattataggcacacaccaccac
3360052

query 24559 acccagttaattttggatatttttagtagagatggggtttcgccatggtggccaggctggt
24618

||||| |
sbjct 3360051 gccagttaattttggatatttttagtacagatggggtttcgctgtgttggccaggctggt
3359992

query 24619 ctcgaaactcctagcctcaggtgatcctcccacctcgacgtcccaaagtgtgagattaca
24678

||||| |
sbjct 3359991 ctcgaaactcctagcctcaggtcatcctcccacctcgacgtcccaaagtgtgagattaca
3359932

query 24679 ggc atgagccaccgcacctgaccaatttttagtttttgaggaacctccaaactgttctc
24738

||||| |
sbjct 3359931 ggc atgagccaccacacctgatcaatttttagtttttcaggaacctccaaactgttctc
3359872

query 24739 gatagtgattgcactattataatttacattcccactaacagcatacaggggttcacacac
24798

||||| |
sbjct 3359871 gatagtgattgcactattataatttacattcccactaacagcgtacaggggttcacacac
3359812

query 24799 ctcgccagcatttgttattgcctgtcttttggataaaagccatttttagctggtgtgagat
24858

||||| |
sbjct 3359811 ctcgccagcatttgttattgcctgtcttttggataaaagccattttaactgggtgtgagat
3359752

query 24859 gat atctc---gtagt tttgatttgcatttctctgatgaccagtgacgttgaatgccttt
24915

||||| |
sbjct 3359751 gat atctcattgtagt tttgatttgcatttctctgatgaccagtgacgttgaataccttt
3359692

query 24916 ccatatgcctg tttaccattcgtatgtcttcttttgagaaatgtctgttcgatcttttg
24975

||||| |
sbjct 3359691 ccatatgcctg tttgctatttgcatttctcttttgagaaatgtctgttcgatcttttg
3359632

query 24976 cccatcttttgatggaattattagtttttttccatagaggttgtttgcgcttctttgtata
25035

||||| |

sbjct 3359631 cccatcttttgatggaactattagatatttttcctatatagttggttgcggtctttatata
3359572

query 25036 ttctggttattaatcccttgtcagatgggtagttacaaatattttctccatttggtggg
25095
|||||

sbjct 3359571 ttctggttattaatcccttgtcagatgggtagttacaaatattttctccatttggtggg
3359512

query 25096 ctgtctcttcactttggtgatcggttactttgctatgcagaagcttttaacttcatgtc
25155
|||||

sbjct 3359511 ctgtctcttcactttggtgatcggttactttgctatgcagaagcttttaacttcatgtc
3359452

query 25156 tgtttatcttataatgtatgtagttcagattatgctgtatagaaaactttaaaaaattt
25215
|||||

sbjct 3359451 tgtttatcttataatgtatgtagttcagattatgctgtatagaaaactttaaaaaattt
3359392

query 25216 atcaccttcagttcttttcttaaaacgttttattgagctatgattaatatacaataaac
25275
|||||

sbjct 3359391 gtcaccttcagttcttttcttaaaacatttttattgagctatgattaatatacaataaac
3359332

query 25276 tacacatacttaaagtgtacaatttgataagtgttaacatatataaaaaccgtgaaac
25335
|||||

sbjct 3359331 tacacacacttaaagtgtacaatttgataagtgttaacatatagataaaaaccgtgaaac
3359272

query 25336 catcactaccatcaaggtaatggacatgtccctcacttccaaaagtctatagaagtttgc
25395
|||||

sbjct 3359271 catcagtagcatcaaggtaatggacatgtccctcacttcc-----agaagtttgc
3359222

query 25396 atcctatttttatcatcaaagtgtatcccgtgagtatttttcccattatcaatttatct
25455
|||||

sbjct 3359221 atcctatttttattatcaaagtgtatcccgtgagtatttttccacattatcaatgtatct
3359162

query 25456 tactgagcatctttagg[RS9858213]ccagggggtctgctgtggcccataccagtgccctagtttggg
25515
|||||

sbjct 3359161 tactgagcatctttagg[RS9858213]ccagggggtctgctgtggcccatacaagtgccctagtttggg
3359102

query 25516 aggacttttgcttttgtaacaatgctgtgaggatgccagggctggagtcgtgtggcta
25575
|||||

sbjct 3359101 aggacttttgcttttgtaacaatgctgtgaggatgccagggctggagtcgtgtggcta
3359042


```
query 25576 gagtgtttgaactgcagtcacttttaagactggtgccagattacttcttcaaaggatcaa
25635
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
sbjct 3359041 gagtgtttgaactgcagtcacttttaagactggtgccagattacttcttcaaaggatcaa
3358982

query 25636 gtaggttctccctccctgcagtgaggcacggtgtccctgtctctcacccctcatcagc
25695
      ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
sbjct 3358981 gtaggttctccgtccctgcagtgaggcacggtgtccctgtctctcaccccttatcagc
3358922

query 25696 cttagctgttctgtcttttacgacttctgcccttaggcaggtccaaaaggcagcaacgt
25755
      ||||||||||||||||||||||||||||||||||| ||||||||||||||||||| |
sbjct 3358921 cttagctgttctgtcttttacgacttctgctcttaggcaggtccaaaaggcagcatcat
3358862

query 25756 tacctgttctatttaagcaactgtgggtgggtcagaaagtgcctgcttgagagagttcctg
25815
      ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
sbjct 3358861 tacctgttctatttaagcaattgtgggtgggtcagaaagtgcctgcttgagagagttcctg
3358802

query 25816 cccacccccacccatactcagtcctggcctgttctcagagccatctgatccctacctca
25875
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
sbjct 3358801 cccacccccacccatactcagtcctggcctgttctcagagccatctgatccctacctca
3358742

query 25876 tgtgataccagagatggtcctactgctgctgctgctgctgctgtggccgccaggagagacagc
25935
      ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
sbjct 3358741 tgtgataccagagatggtcctgctgctgctgctgctgctgctgtggccgccaggagagacagc
3358682

query 25936 aggtggtagggctgtcagtcagtggttgtgggtattcagggccctgaactgggacctagg
25995
      ||||||||||| ||||||||||||||||||| |||||||||||||||||||
sbjct 3358681 aggtggtagggccgtcagtcagtggttgtggggattcagggccctgaactgggacctagg
3358622

query 25996 agatccaaaagggtcaggggaagttgctggctctggccaaatgggcagtgctgtaggcact
26055
      |||||||||||||||||||||||||||||||||||||||||||||||||||
sbjct 3358621 agatccaaaagggtcaggggaagttgctggctctggccaaatgggcagtgctgtaggcact
3358562

query 26056 cagtggtggctgtctcaggccgactgcccacagatatgtgcagtcctgctctcaggatcag
26115
      |||||||||||||||||||||||||||||||||||||||||||||||||||
sbjct 3358561 cagtggtggctgtctcaggccgactgcccacagatatgtgcagtcctgctctcaggatcag
3358502
```

```
query 26116 gcacacgacttttggggtcaggtacaagagtagctgctccaagtgtgggcttgtaagaa
26175
      |||
sbjct 3358501 gcacacgactttgggggtcaggtacaagagtagccactccaagtgtgggcttgtaagaa
3358442
      |||

query 26176 ctgttttcttttgagacaggggtctcgctatgtcacagtcatactgcaacctcaa
26235
      |||
sbjct 3358441 ctgttttctttgaggcaggggtctcgctatgtcccagtcatactgcaacctcaa
3358382
      |||

query 26236 gtcccaggctcaaggaatcctcccacctcagcctcccacatagctgggaccacaagcaca
26295
      ||
sbjct 3358381 gttccaggctcaaggaatcctcccacctcagcctcccacatagctgggaccacaagcaca
3358322
      ||

query 26296 tgccacagcaccagctaa----tttttttttttaagagacaggggtcttactgtgtcac
26351
      |||
sbjct 3358321 tgccacagcaccagctaaattttttttttttaagaggcaggggtcttactgtggcac
3358262
      |||

query 26352 ccaggctggtctcaaaccctggggtcaagtgattctcccacctcagcctcccaaagtgc
26411
      |||
sbjct 3358261 ccaggctggtctcaaaccctggggtcaagtgattctcccacctcagcctcccaaagtgc
3358202
      |||

query 26412 tgagactacaggtgtgagccaccatgcttggcctgtaactgcttttgaggaaggctcctt
26471
      |||
sbjct 3358201 tgagactacaggtgtgagccaccatgcttggcctgtaactgcttttgagaaaggctcctt
3358142
      |||

query 26472 tatccttgggcccactggggtattcagagatggggtggctttagccttctatgctacta
26531
      |||
sbjct 3358141 tatcctt-ggtccactggggtattcagagatggggtggctttagccttctatgctacta
3358083
      |||

query 26532 gctgagcaccgactgcatagccaccacctggccttggcctgggacagagcagcagtgg
26591
      |||
sbjct 3358082 gctgagcaccgactgcatagccaccacctggccttggcctgggacagagcagcagtgg
3358023
      |||

query 26592 ggaggggcctttgtgcatctctgctcctacagctcaggggtgggagtggcttacct-gg
26650
      |||
sbjct 3358022 ggaggggcctttgtgcatctctgctcctacagctcaggggtgggagtggcttacctggg
3357963
      |||

query 26651 ggaatatgtgatccctgagcctgggtagtggctactttgtcaggggtgc-ttgtccaacct
26709
```

sbjct	3357962	 ggaatatgtgatccctgagcctgggtagtggctactttgtcagggtgctttgtgccacct
	3357903	
query	26710	ttgtccccttccagagctttctgtctgtgtggatgagtggggaagaactggcattaag
	26769	
sbjct	3357902	 tcgtccccttccagagctttctgtctgtgtggatgagtaggtaggaactggcatgaaag
	3357843	
query	26770	gtggaagagcaggtggtggcaccctgtgggcagccaggtgtgtaatggcctctgttctg
	26829	
sbjct	3357842	 gtggaagagcaggtggtggcagcctgtgggcagccaggtgtgtaatggcctctgttctg
	3357783	
query	26830	cctgagtcctgctggttaccaccacctcactccctgctattcttagcagctgagctcagg
	26889	
sbjct	3357782	 cctgagtcctgctggttaccaccacctcactccctgctattcttagcagctgagctcagg
	3357723	
query	26890	acagcactgactggcattagacagtaccctttccctctaggcctgtgtgttgagaggat
	26949	
sbjct	3357722	 acagcactgactggcattagacagtaccctttccctctaggcctgtgtgttgagaggat
	3357663	
query	26950	gagggggccaccagacagatttttcccaactcctttggctcgggtaggtgggtcatgtg
	27009	
sbjct	3357662	 gagggggccaccagacagatttttcccaactcctttggctcgggtaggtgggtcatgtg
	3357603	
query	27010	ctccgctgaccatttatggactggcgggtacaggacagagccatccagcagctcagaggg
	27069	
sbjct	3357602	 ctctgctgaccatttatggactggcgggtacaggacagagccatccagcagctcagaggg
	3357543	
query	27070	ctgagcctttcatacttccagaacatgtccattcaacacagactcccctggctgggtgg
	27129	
sbjct	3357542	 ctgagcctttcatacttccaggacatggccattcaacacagactcccctggctgggtgg
	3357483	
query	27130	tggcgagactggaatctgccccaaaggaaccacaggcctaggggcaagggttgactagg
	27189	
sbjct	3357482	 tggcgagactggaatctgccccaaaggaaccacaggcctaggggcaagggttgactagg
	3357423	
query	27190	tggccttggtaactgaccacttgggaaccagcatttctcccagccttcttctccagaca
	27249	

sbjct 3357422 tggccttgggtactgaccactttgggaaccagcatttctcccagccttcttcttccagaca
3357363

query 27250 tcatcagcacagccacctttgagagcggacaggttatgtgctatgcagacagcacttggt
27309

sbjct 3357362 tcatcagcccagccacctttgagagcggacaggttatgtggatgcagacagcacttggt
3357303

query 27310 attccctgggctcttcacagcatcttaacataagtcaactattctgtttttacagatgaa
27369

sbjct 3357302 attccctgggctcttcacagcatcttaacataagtcaactattctgtttttacagatgaa
3357243

query 27370 ggaattaagacttagagagggctac-----ttttttttttttttttttgagacagagtct
27423

sbjct 3357242 ggaattaagacttagagagggctactttttttttttttttttttgagacggagtct
3357183

query 27424 tgctctgtcaccagactagagtgcagtgccacgatcatggctcactgcaacctccgcct
27483

sbjct 3357182 tgctctgtcaccagactagagtgcagtggtgcgatcatggctcactgcaacctccgcct
3357123

query 27484 cccaggttcaagcaattctcctgcctccgcctccaagtagctggtattacaggcacatg
27543

sbjct 3357122 cccaggttcaagcaattctcctgcctccgcctccaagtagctggtattacaggcacatg
3357063

query 27544 ccaccgcgccagc-----atthtttagtagagtcgcaatthccacatgthggcca
27593

sbjct 3357062 ccaccgcgccagctactthtttgatthtttagtagagtcgcaatthccacatgthggcca
3357003

query 27594 ggctggtcttgaactactgacctcgtgatctgcccgcctcagcctcccaaagtgctggga
27653

sbjct 3357002 ggctggtcttgaactactgacctcatgatctgcccgcctcagcctcccaaagtgctggga
3356943

query 27654 ttacaggcatgagccaccgcacctggccacatac--ttttttttttttttttttgagac
27711

sbjct 3356942 ttacaggcgtgagccaccgcacctggccacatactttttttttttttttttttgagac
3356883

query 27712 agagtcttactcttgttgcccgggctgaagtacaatgacacgatcttggttactacaac
27771

sbjct 3356882 agagtcttactcttgttgcccgggctgaagtacaatgacacgatctcggttactacaac
3356823

```

query 27772 ctctgccccaccaaggttcaagctattctcctgcctctgtctcccgagtagctgggatta
27831
      |||
sbjct 3356822 ctctgccccaccaaggttcaagcgattctcctgcctctgcctcccgagtagcggggatta
3356763
      |||

query 27832 tacgcatgcccaccacactcagctaatttttgtatttttagtagggatggggtttcacc
27891
      |||
sbjct 3356762 tacgaatgtgccaccacgcccagctaatttttgtgttttagtagggatggggtttcacc
3356703
      |||

query 27892 atgtggccaggctggtctcgaactcctgacctcaggtgatctgcccgcctcggcctcca
27951
      |||
sbjct 3356702 atgtggccaggctcgtcccgaactcctgacctcaggtgatctgcccgcctcggcctcca
3356643
      |||

query 27952 aagtgctgggattacagggatggggccaccacacccggcctggactgggtataacttttgag
28011
      |||
sbjct 3356642 aagtgctgggattacagggacggggccaccacacccagtctggactgggtataacttttgag
3356583
      |||

query 28012 ttagtgagagaattcggccttctgacctcctgcatgccagagtctaggactaggcacagc
28071
      |||
sbjct 3356582 ttagtgagagaattcagccttctgacctcctgcatgccagagtctaggactaggcacagc
3356523
      |||

query 28072 ccttttctggagtctaggcagccccctgtgagctcaggggaaaggatgggatttagtg
28131
      |
sbjct 3356522 ccttttctggagtctaggcagccccctgtgagctcaggggaatggatgggatttagtg
3356463
      |
vers fin du gene rnf123 49733965(+)->      ↓fin de region alignee chr3:49709548
query 28132 tcactctggctgatttcaaggactt 28155
      |||
sbjct 3356462 tcactctggctgatatacaaggactt 3356439
      |||

```

ANNEXE C

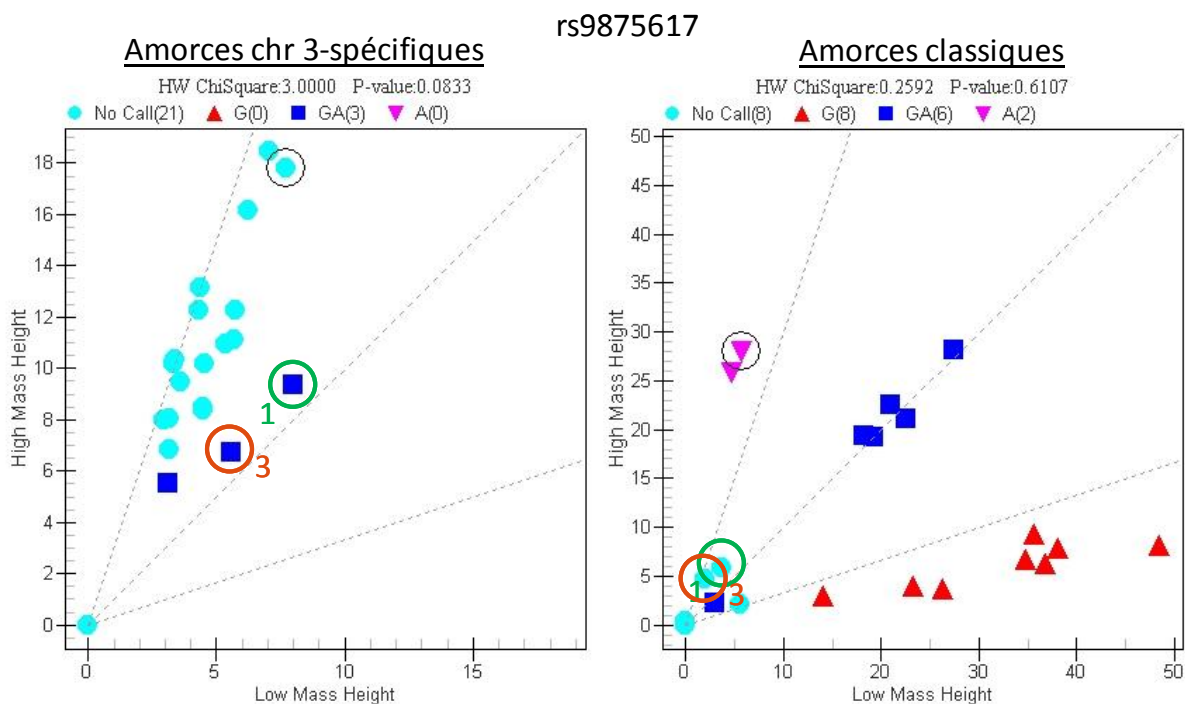
Détail des amorces utilisées pour les qPCR.

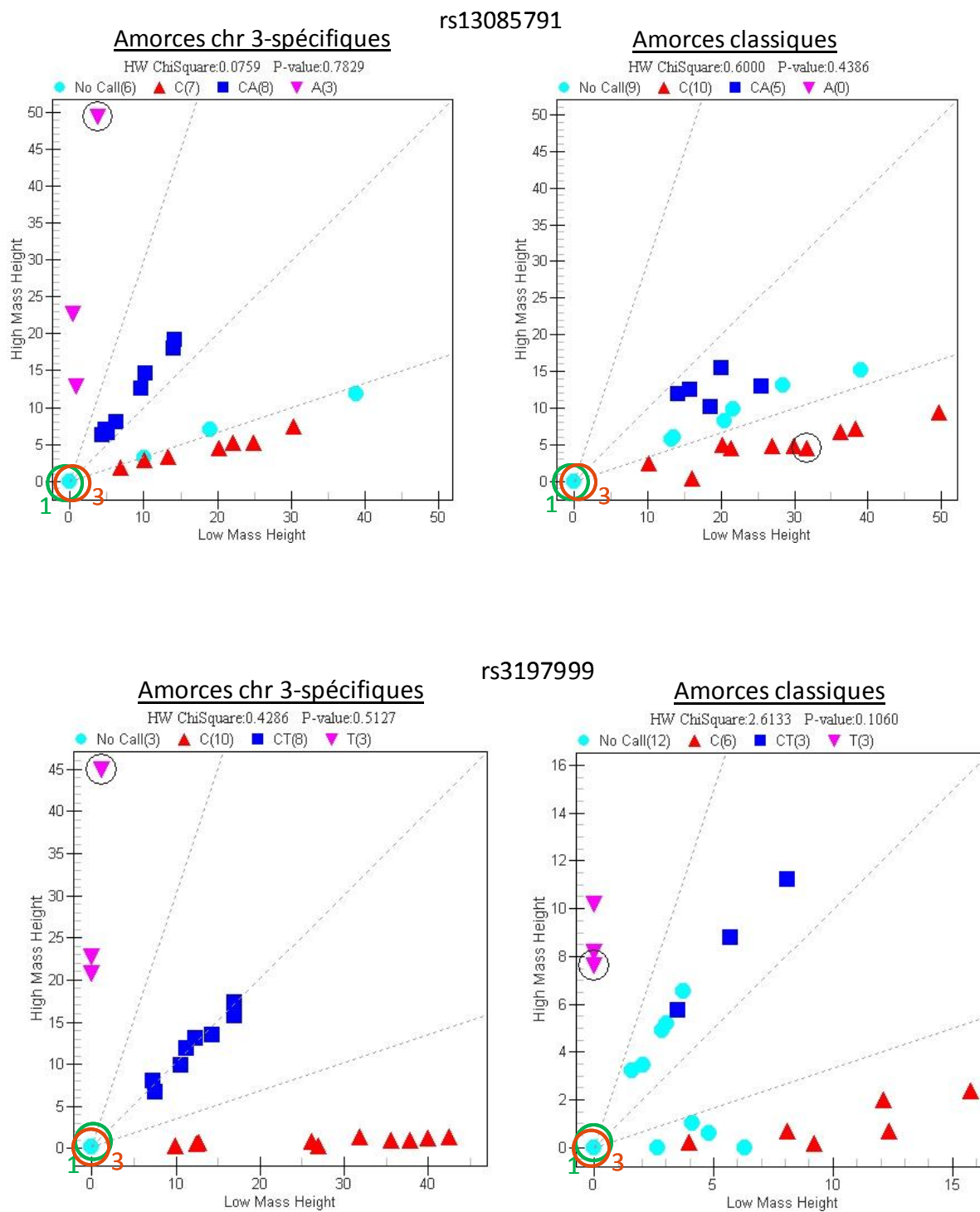
Gène	Amorces	Séquence
CD-14	sens anti	GAGGTTCGGAAGACTTATCG ACGGCTCTAGGTTGAGAC
CD-206	sens anti	GACTGTGTTGTTATTATTGGAG TCGTTGCTGGAGGATTAG
HPRT1	sens anti	TGGCGTCGTGATTAGTGATG CAGAGGGCTACAATGTGATGG
IL12B	sens anti	CTGAGGAGAGTCTGCCATTG CTTGCCCTGGACCTGAACG
MST1R	sens anti	ATATGTGCCACGGTGACC CCTTGACTGTGAATGACTCG
TNF- α	sens anti	GACCTCTCTCTAATCAGC CTGGTTATCTCTCAGCTC

ANNEXE D

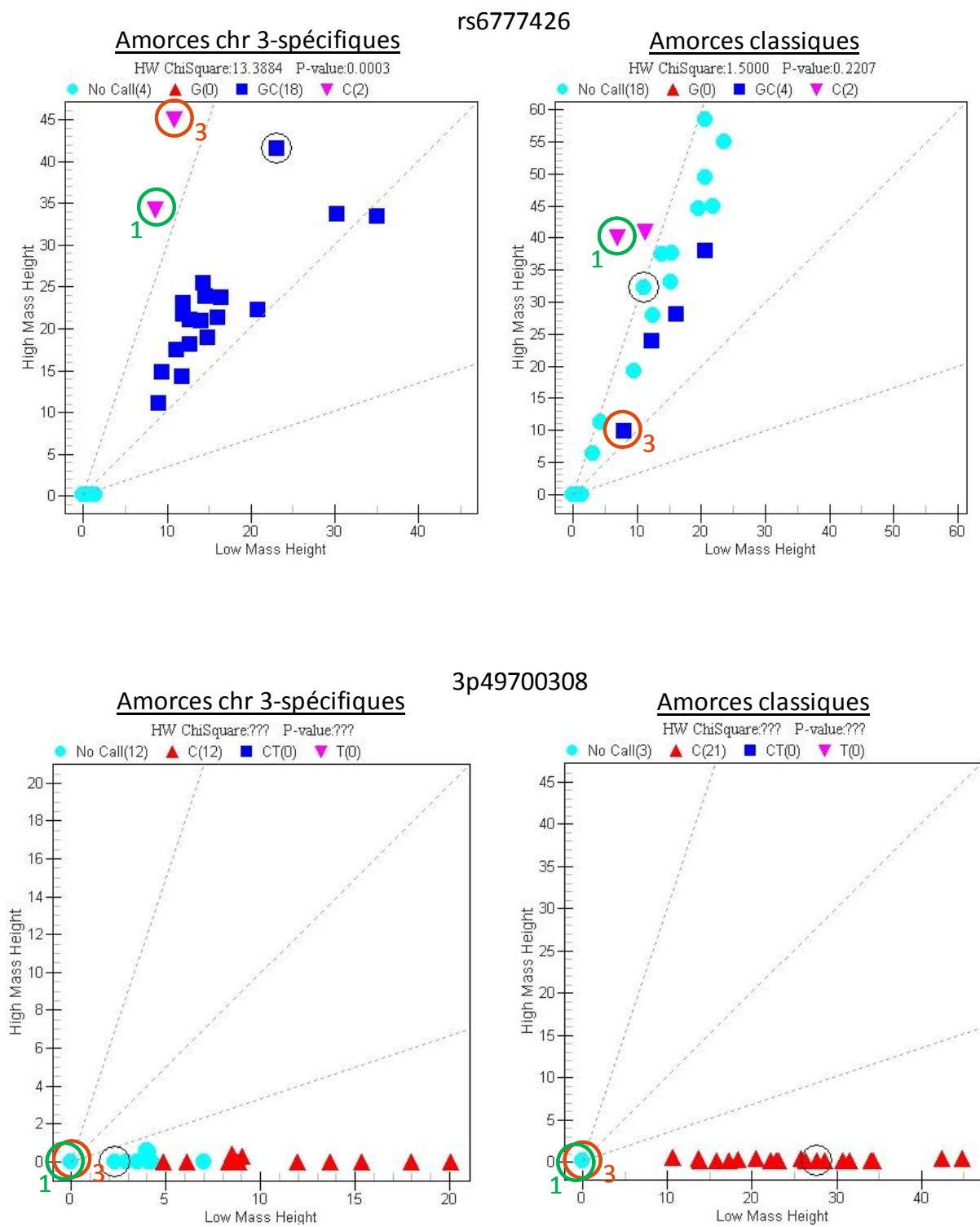
Graphiques d'aggrégations des tests de génotypage. Les graphiques sont groupés par 2 pour chaque SNP, un pour les amorces classiques, l'autre pour les amorces spécifiques à la séquence de MST1. Les cercles aqua représentent les individus dont le génotype n'a pu être déterminé. Les carrés bleu désignent les individus ayant été classifié comme hétérozygotes, et les triangles rouges ou rose représentent les individus ayant été classés monozygotes pour l'un ou pour l'autre des allèles.

A) Variantes de la région 3p21



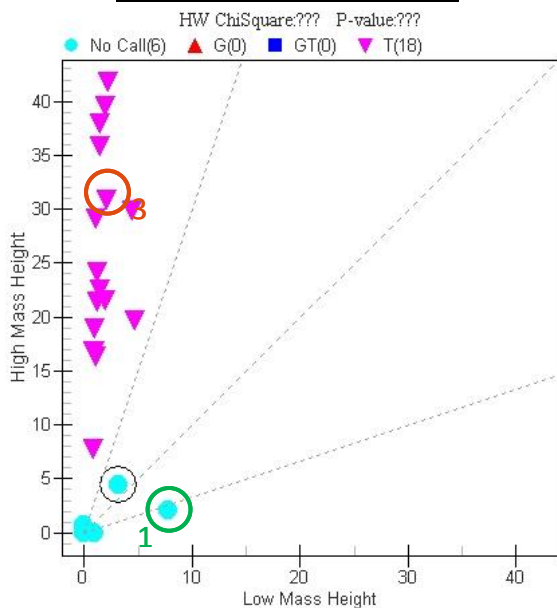


B) Différences entre les duplications

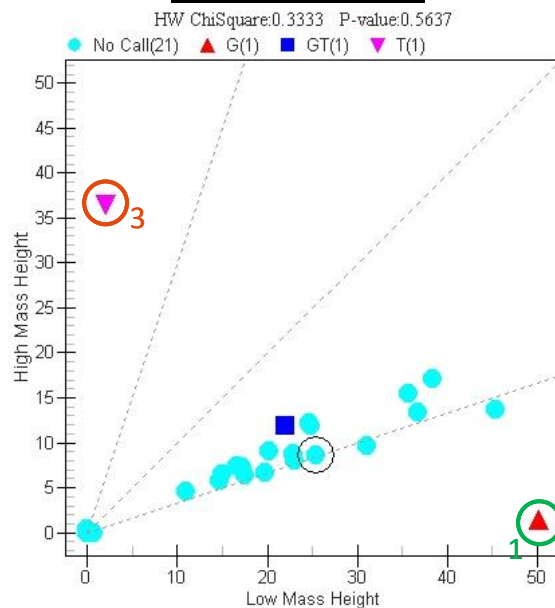


rs1057319

Amorces chr 3-spécifiques

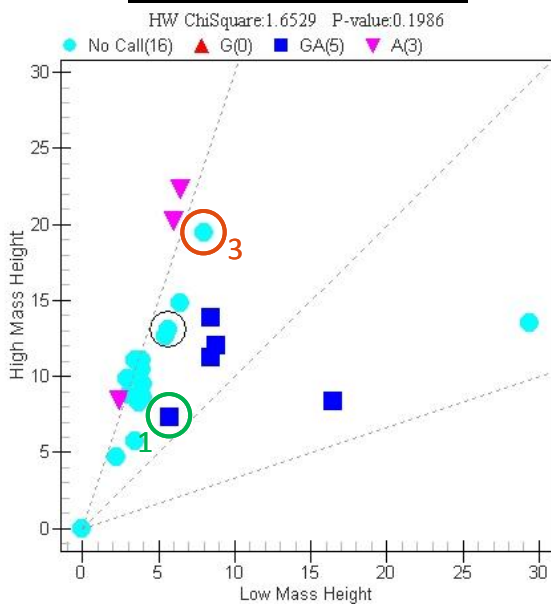


Amorces classiques

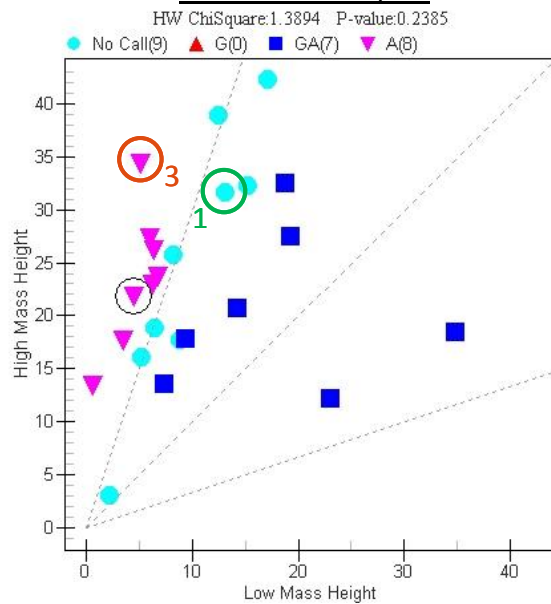


3p49697003

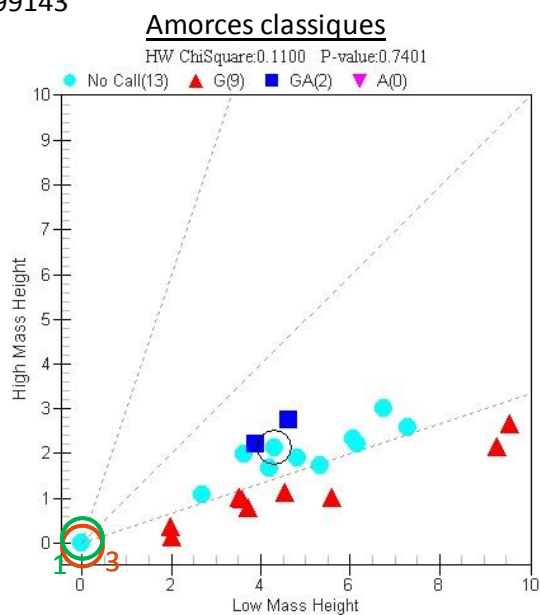
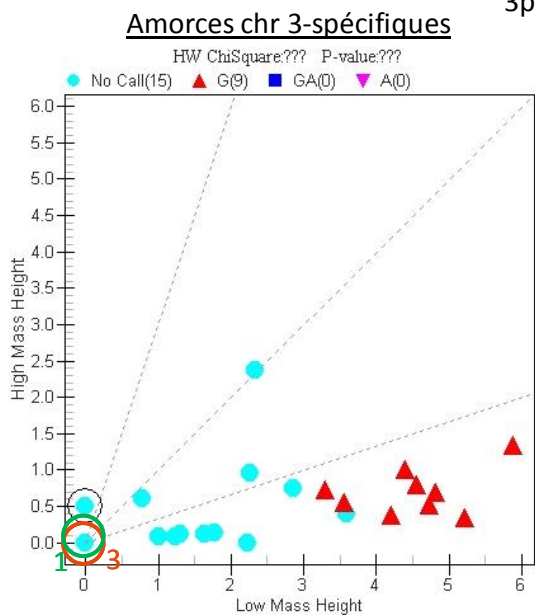
Amorces chr 3-spécifiques



Amorces classiques

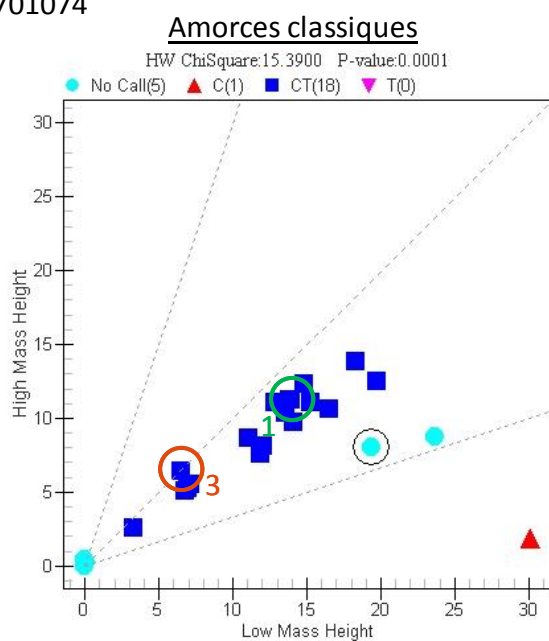
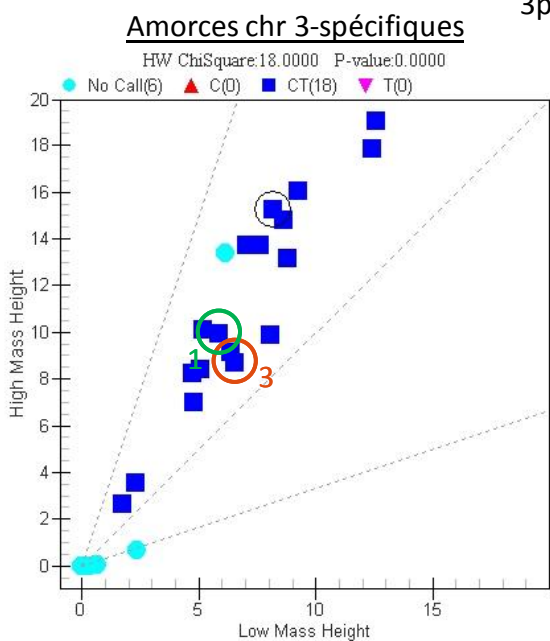


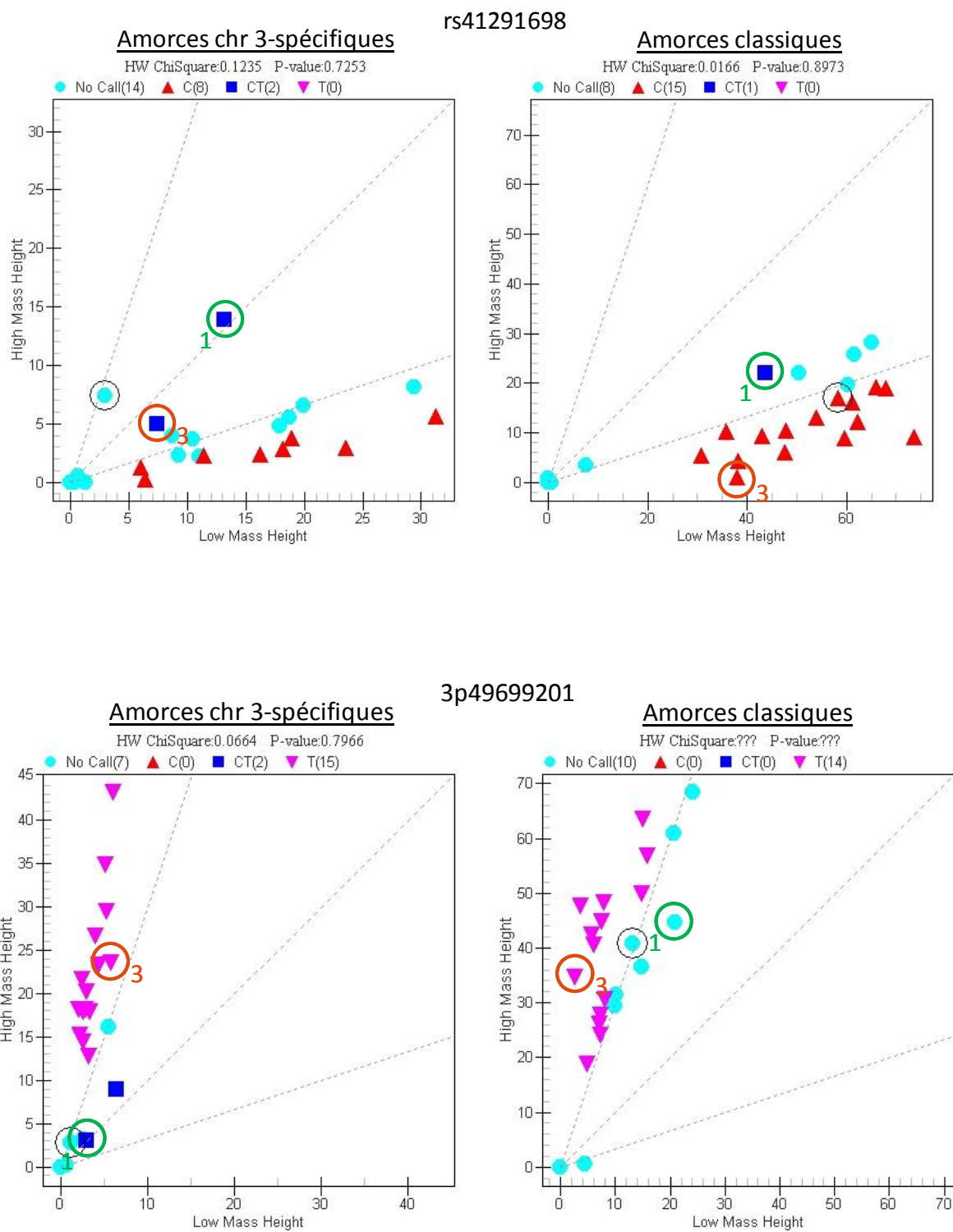
3p49699143



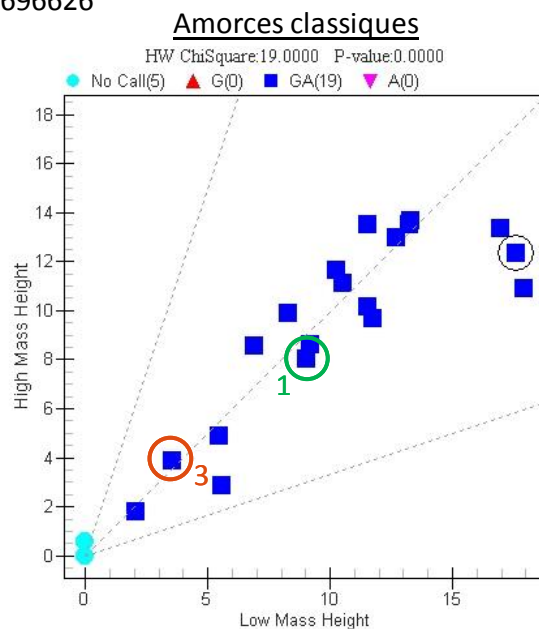
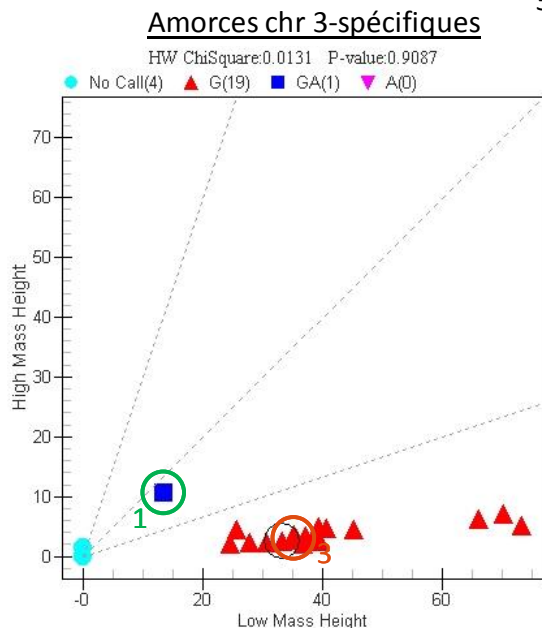
C) Variantes à tester

3p49701074

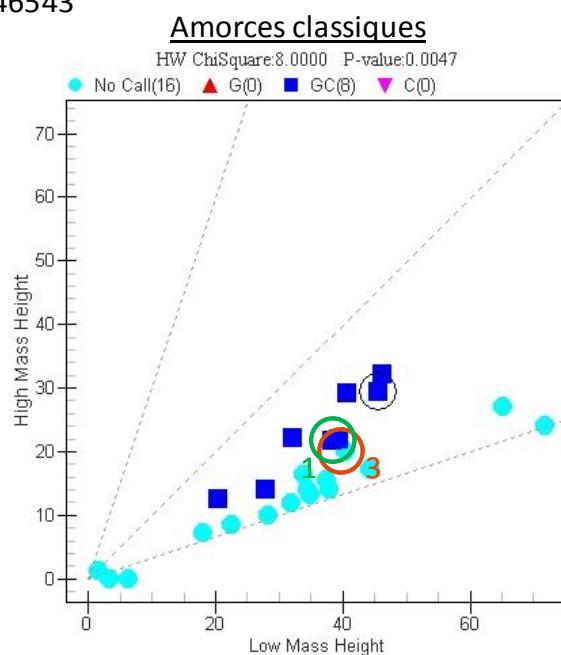
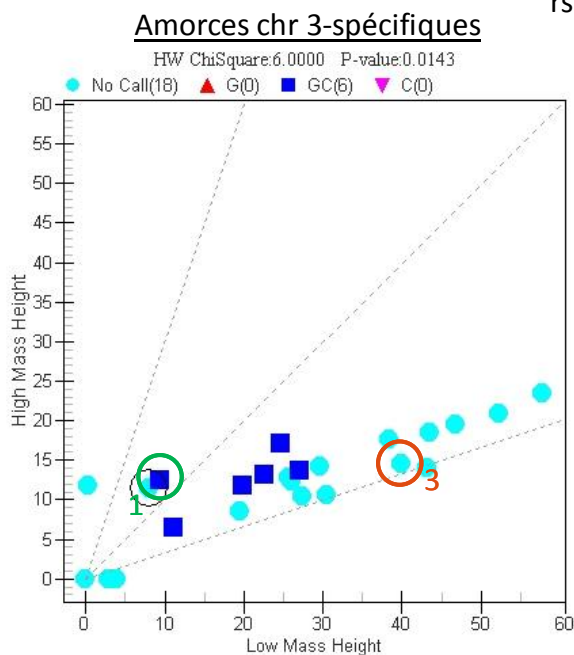




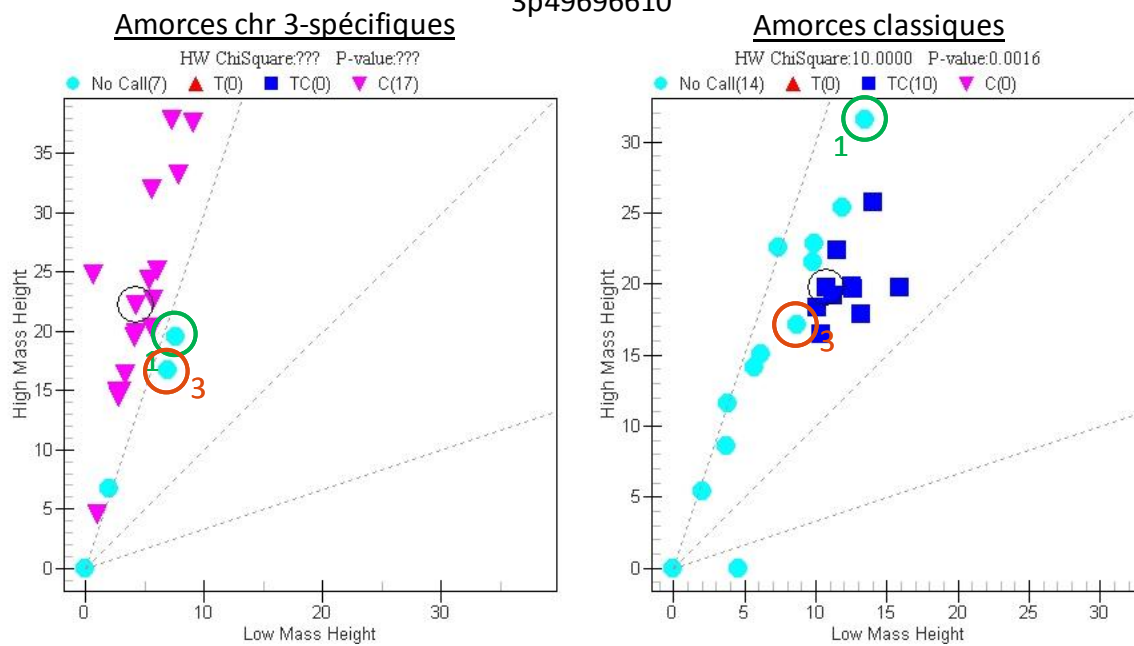
3p49696626



rs2446543



3p49696610



ANNEXE E

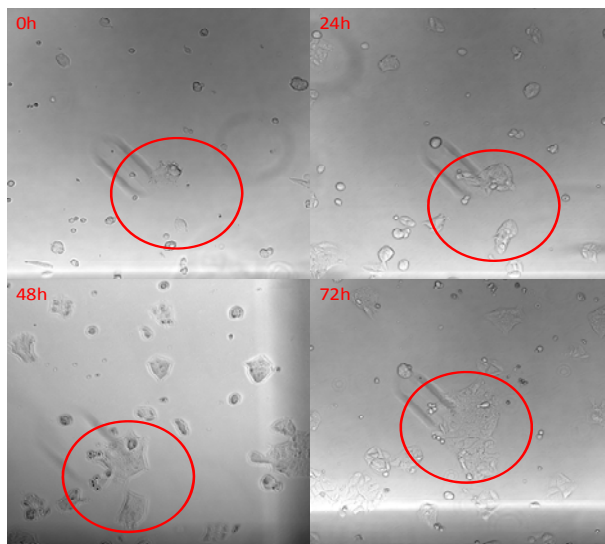
MAF des SNP pour chaque cohorte utilisée dans la méta-analyse.

SNP	Cohorte		MAF
rs3197999	ITACC		0.3069
	Cedars	Total	0.2994
		Juifs	0.4148
		non-Juifs	0.2849
	Sweden		0.3473
NIDDK		0.3431	
rs6802890	ITACC		0.4549
	Cedars	Total	0.4793
		Juifs	0.4056
		non-Juifs	0.4886
	Sweden		0.4448
NIDDK		0.4522	
rs7629936	ITACC		0.2059
	Cedars	Total	0.2361
		Juifs	0.1842
		non-Juifs	0.2426
	Sweden		0.2195
NIDDK		0.2176	

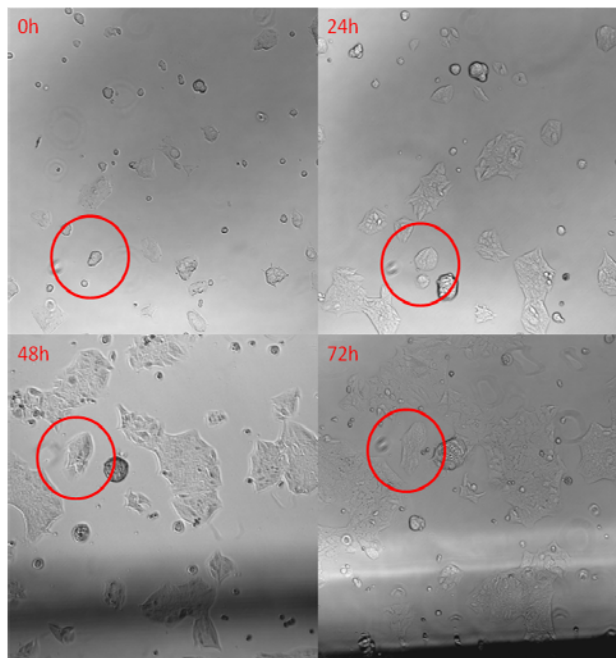
ANNEXE F

Photos de l'expérience de migration des cellules épithéliales du côlon traitées soit au PMA, MSP ou HGF aux temps 0h, 24h, 48h et 72h pour les cellules HT-29, HCT-15 et HCT-116. Les photos proviennent d'un duplicata (1/4).

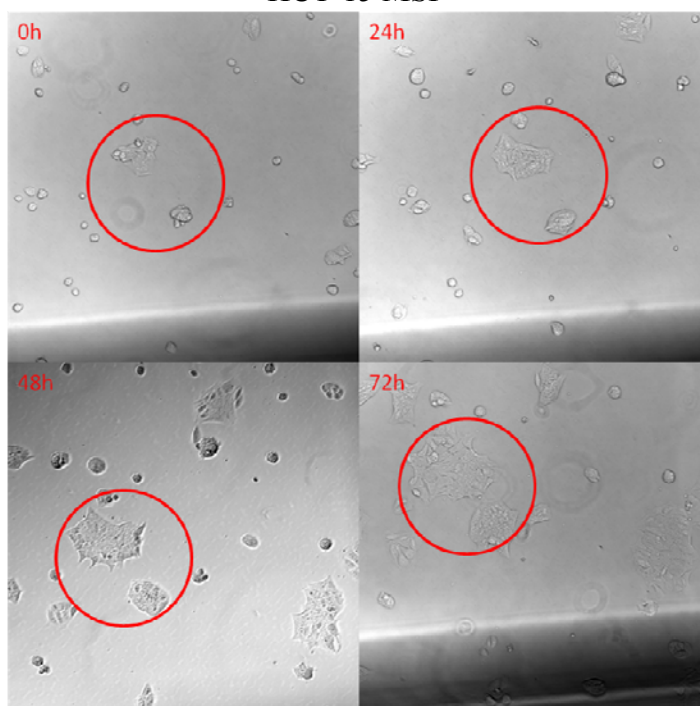
HCT-15 ctl



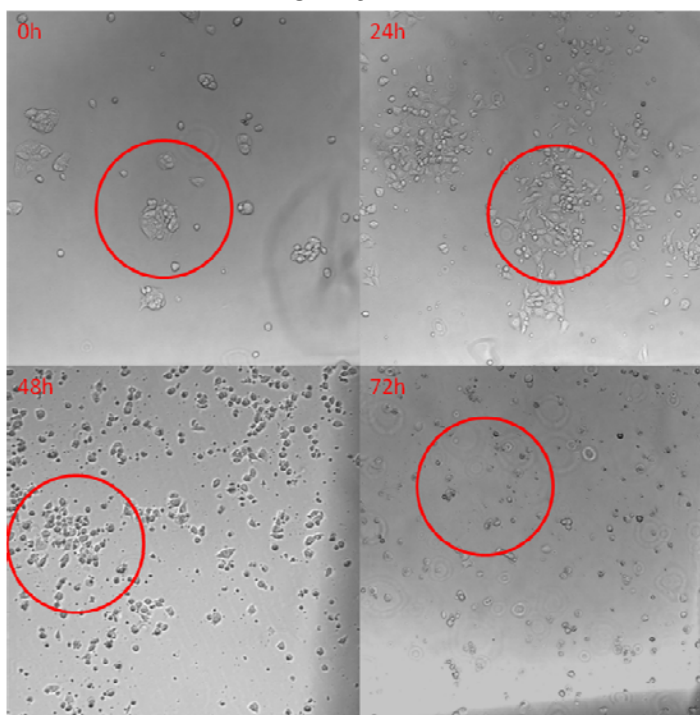
HCT-15 HGF



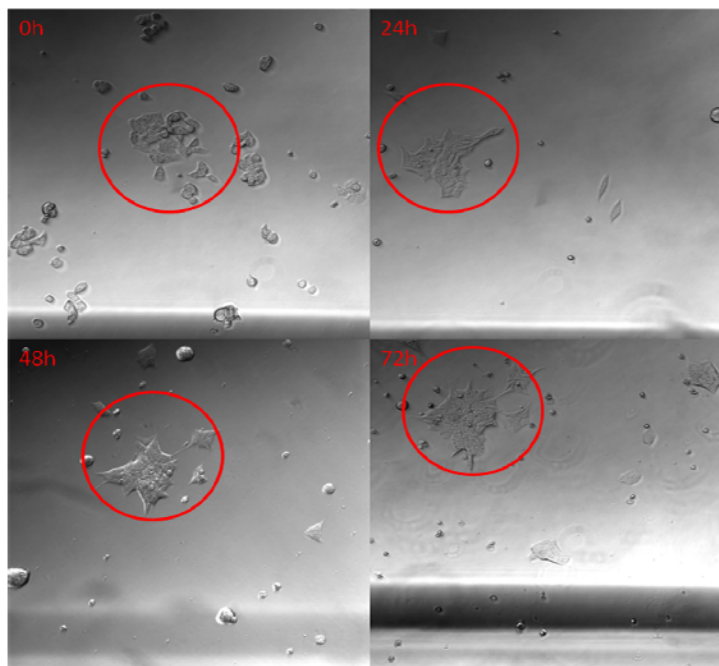
HCT-15 MSP



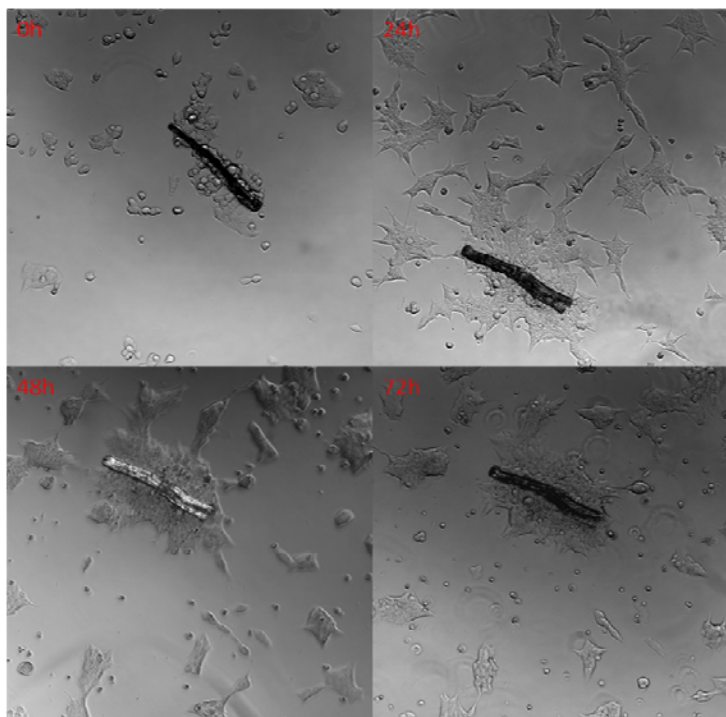
HCT-15 PMA



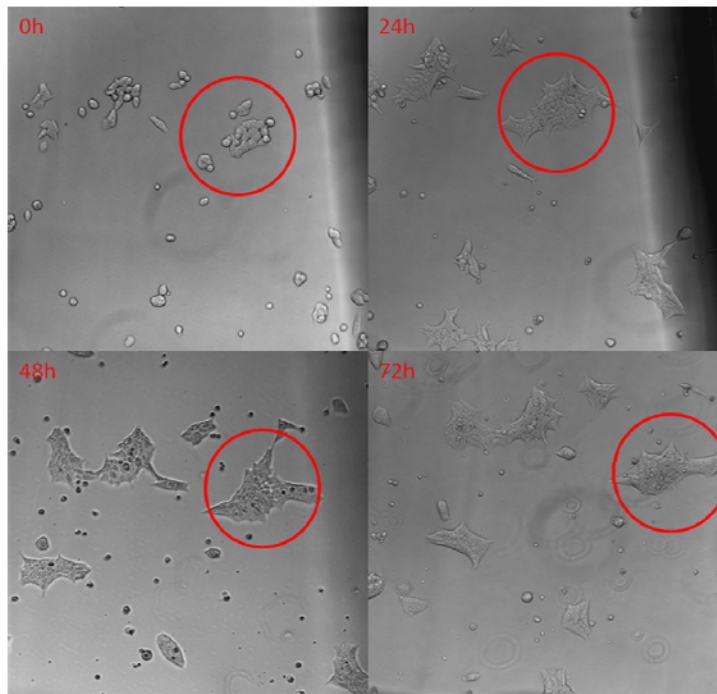
HCT-116 ctl



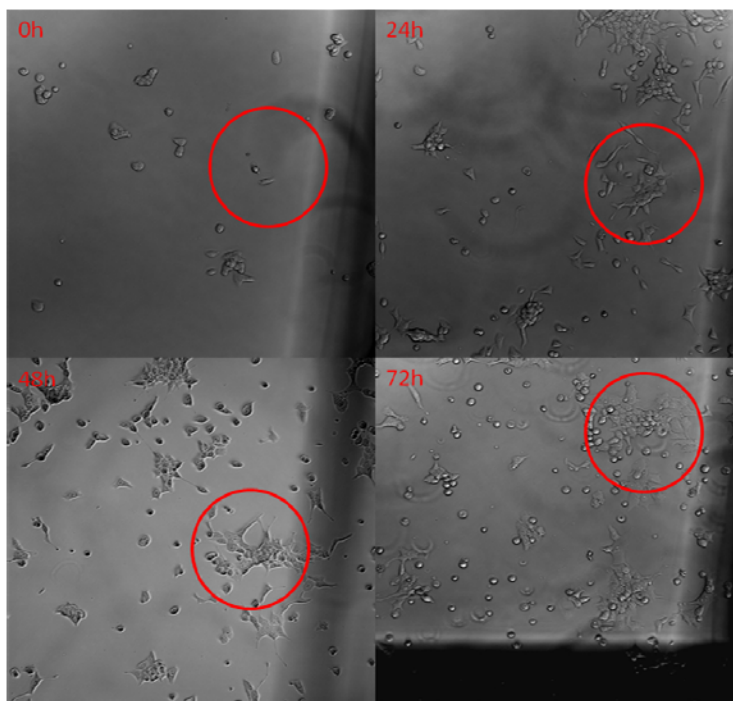
HCT-116 HGF



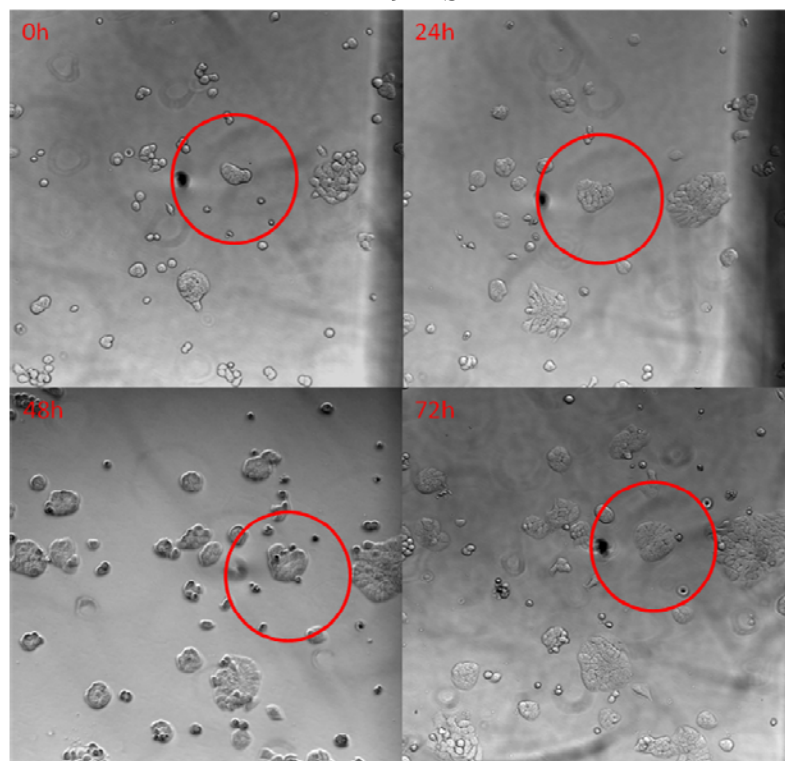
HCT-116 MSP



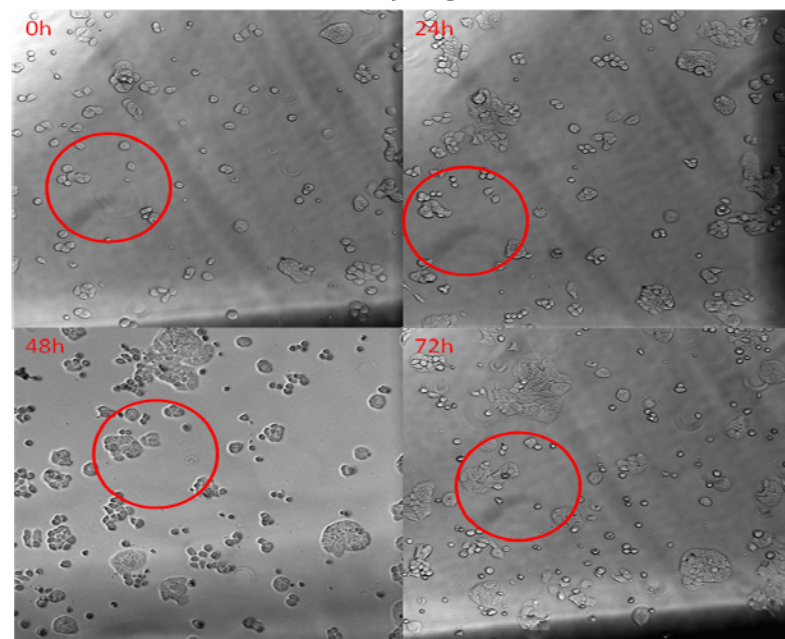
HCT-116 PMA



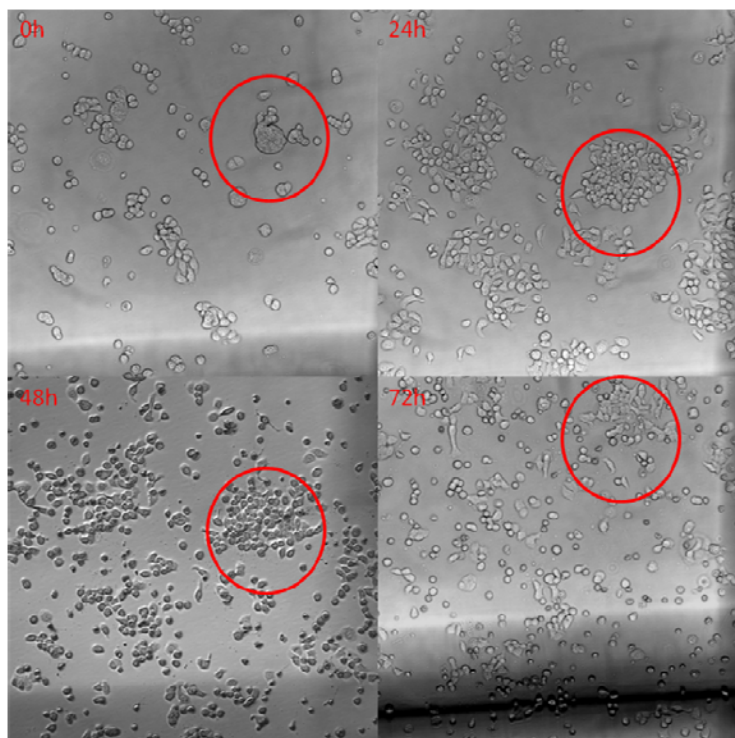
HT-29 MSP



HT-29 HGF



HT-29 PMA



HT-29 ctl

