

Université de Montréal

**MAST3 : facteur de risque génétique aux maladies
inflammatoires de l'intestin et modulateur
d'inflammation**

par

Catherine Labbé

Département de sciences biomédicales

Faculté de médecine

Thèse présentée à la Faculté de médecine
en vue de l'obtention du grade de doctorat
en sciences biomédicales

5 août, 2011

© Catherine Labbé, 2011

Université de Montréal
Faculté de médecine

Cette thèse intitulée :

MAST3 : facteur de risque génétique aux maladies inflammatoires de l'intestin
et modulateur d'inflammation

Présentée par :
Catherine Labbé

a été évaluée par un jury composé des personnes suivantes :

Daniel Sinnett, président-rapporteur
John D. Rioux, directeur de recherche
Zoha Kibar, membre du jury
Yohan Bossé, examinateur externe
Gaëtan Mayer, représentant du doyen de la FES

Résumé

La maladie de Crohn (MC) et la colite ulcéreuse (CU) sont des maladies inflammatoires chroniques du tube digestif qu'on regroupe sous le terme maladies inflammatoires de l'intestin (MII). Les mécanismes moléculaires menant au développement des MII ne sont pas entièrement connus, mais des études génétiques et fonctionnelles ont permis de mettre en évidence des interactions entre des prédispositions génétiques et des facteurs environnementaux - notamment la flore intestinale – qui contribuent au développement d'une dérégulation de la réponse immunitaire menant à l'inflammation de la muqueuse intestinale. Des études d'association pangénomiques et ciblées ont permis d'identifier plusieurs gènes de susceptibilité aux MII mais les estimations de la contribution de ces gènes à l'héritabilité suggèrent que plusieurs gènes restent à découvrir. Certains d'entre eux peuvent se trouver dans les régions identifiées par des études de liaison génétique.

L'objectif de mon projet de doctorat était d'identifier un ou des facteurs de risque génétique dans la région chromosomale 19p (identifiée comme région de liaison IBD6) et de le/les caractériser au niveau fonctionnel. Nous avons d'abord entrepris une cartographie d'association de la région 19p. À la suite du génotypage successif de deux cohortes indépendantes, nous avons identifié un SNP intronique et quatre SNP codants dont un non-synonyme, rs8108738, tous localisés dans le gène *microtubule associated serine threonine kinase gene-3 (MAST3)* et associés aux MII. Peu d'information fonctionnelle sur MAST3 était disponible. Par contre MAST2, une protéine encodée par un gène de la même famille, régule l'activité du facteur de transcription inflammatoire NF- κ B. Nous avons confirmé l'implication de MAST3 dans l'activité de NF- κ B via un knockdown de *MAST3* et des essais gène-rapporteur. Pour poursuivre la caractérisation fonctionnelle de MAST3, nous avons choisi une approche non ciblée pour étudier les effets de la variation des niveaux d'expression de MAST3 sur la cellule. C'est-à-dire que nous avons créé un 1^{er} modèle cellulaire de surexpression du gène *MAST3* dans les

cellules HEK293 et analysé l'expression pangénomique endogène. La validation de l'expression génique dans un 2^e modèle cellulaire de knockdown et de type cellulaire différent (THP1), nous a permis d'identifier et de contrer les effets non-spécifiques dus aux niveaux non-physiologiques. Notre étude d'expression a mené à l'identification d'un groupe de gènes dont l'expression est régulée par MAST3. Ces gènes sont majoritairement impliqués dans des fonctions immunitaires (cytokines pro-inflammatoires, régulateurs de NF- κ B, migration cellulaire, etc.) et une forte proportion est régulée par NF- κ B. Nous avons évalué l'importance du groupe de gènes régulés par MAST3 dans la présentation clinique des MII à travers des études d'expression dans des biopsies intestinales de patients atteints de CU. Nous avons constaté que l'expression de ces gènes est significativement supérieure dans les régions enflammées par rapport aux régions saines de la muqueuse intestinale des patients atteints de CU. Globalement, les résultats de nos études suggèrent que le facteur de risque aux MII *MAST3* agit via la voie du facteur de transcription NF- κ B pour influencer l'expression d'un groupe de gènes impliqués dans l'inflammation intestinale typique des MII.

Chaque étude génétique sur les MII a le potentiel d'orienter les recherches fonctionnelles vers de nouvelles voies biologiques causales. Le dévoilement des mécanismes moléculaires sous-jacents à ces voies permet d'augmenter les connaissances sur le développement de ces maladies vers une compréhension plus complète de la pathogenèse qui permettra d'optimiser le diagnostic et le traitement de ces maladies.

Mots-clés : Maladies inflammatoires de l'intestin, susceptibilité génétique, chromosome 19p, MAST3, NF- κ B, expression pangénomique, inflammation

Abstract

Inflammatory bowel diseases (IBD) refer to different chronic inflammatory diseases of the digestive tract mainly Crohn's disease (CD) and ulcerative colitis (UC). Mechanisms leading to the pathogenesis of IBD are not completely understood, but genetic and functional studies have highlighted interactions between genetic predispositions and environmental factors, such as the intestinal microbiota, as contributors to the deregulation of the immune response that leads to inflammation of the intestinal mucosa. Genome-wide and targeted association studies have identified several IBD susceptibility genes. However, estimations of the contribution of these genes to heritability of the disease suggest that more are to be discovered. Some of these genes may be in previously identified IBD linkage regions. The objective of my doctoral project was first, to identify risk factors in the 19p chromosomal region (IBD6), previously identified following a genome-wide linkage study, then to characterise them functionally.

We first performed a comprehensive association mapping study of the 19p region. Our two-stage genotyping strategy led to the identification of one intronic SNP and four coding SNP –including one non-synonymous SNP, rs8108738 – all located in the microtubule associated serine threonine kinase gene 3 (*MAST3*) and associated to IBD. Very limited functional information on *MAST3* was available at that time. However *MAST2* (a gene in the same family as *MAST3*), is involved in the regulation of inflammation master switch, transcription factor NF- κ B. We confirmed the involvement of *MAST3* in the modulation of NF- κ B via a knockdown of *MAST3* and gene reporter assays. In order to further characterize the function of *MAST3*, we chose a non-targeted approach to study the effects of the modulation of *MAST3* levels on the cell. More specifically, we created a 1st cell model of *MAST3* overexpression in HEK293 cells and analysed the resulting genome-wide endogenous gene expression. Validation in a 2nd

cell model consisting of a knockdown of *MAST3* in THP1 cells, allowed to identify non-specific gene expression due to non-physiological *MAST3* levels. Our expression study led to the identification of a group of genes whose expression is modulated by *MAST3*. These genes are mainly involved in immune functions (pro-inflammatory cytokines, NF- κ B regulation, cellular migration, etc) and a majority is regulated by transcription factor NF- κ B. We evaluated the importance of this *MAST3*-regulated gene set in the clinical manifestation of IBD through an expression study on biopsies of UC patients. We found that the expression the *MAST3*-regulated gene set was significantly enriched in inflamed region of the intestinal mucosa of UC patients compared to healthy region. Taken together, the results of our study suggest that IBD risk factor *MAST3* acts on the NF- κ B pathway to influence the expression of a group of genes involved in intestinal inflammation typical of IBD.

Every genetic study on IBD has the potential to lead functional research towards new causal biological pathways. The unravelling of the molecular mechanisms underlying these pathways aims to improve the comprehension of the pathogenesis of IBD and hopefully will allow for optimization of diagnostic and treatment of these diseases.

Keywords : Inflammatory bowel diseases, genetic susceptibility, chromosome 19p, *MAST3*, NF- κ B, genome-wide expression, inflammation

Table des matières

Résumé.....	i
<i>Abstract</i>	iii
Table des matières.....	v
Liste des tableaux.....	ix
Liste des figures.....	xi
Liste des abréviations.....	xiii
Dédicace.....	xv
Remerciements.....	xvi
Avant-Propos.....	xviii
INTRODUCTION.....	1
Chapitre 1 Le système digestif dans la santé et les maladies inflammatoires.....	2
1.1 Biologie du système digestif en santé : inflammation contrôlée.....	4
1.1.1 Microbiome bactérien.....	4
1.1.2 Les mécanismes physiques de défense.....	7
1.1.2.1 Le péristaltisme.....	7
1.1.2.2 La paroi intestinale.....	9
1.1.2.3 Le mucus.....	9
1.1.3 Les mécanismes biologiques de défense.....	10
1.1.3.1 Manifestations circulatoires.....	10
1.1.3.2 Manifestations humorales non-spécifiques.....	10
1.1.3.3 Manifestations immunitaires.....	11
1.1.4 La tolérance immunologique.....	17
1.2 Les maladies inflammatoires de l'intestin.....	19
1.2.1 Épidémiologie des maladies inflammatoires de l'intestin.....	19
1.2.2 Manifestations cliniques.....	20
1.2.3 Traitement.....	26
1.2.3.1 L'acide 5-aminosalicylique.....	26

1.2.3.2 Les corticostéroïdes	26
1.2.3.3 Les immunomodulateurs.....	27
1.2.3.4 Les antibiotiques	27
1.2.3.5 Les traitements biologiques	28
1.2.3.6 La chirurgie.....	29
1.2.4 Études des causes immunobiologiques.....	30
1.2.5 Facteurs de risque environnemental.....	33
1.2.5.1 La cigarette.....	33
1.2.5.2 L'appendicectomie.....	35
1.2.5.3 Autres facteurs	36
Chapitre 2 Les maladies inflammatoires de l'intestin: des maladies génétiques..	37
2.1 Une histoire familiale.....	37
2.2 Structure du génome et variations.....	38
2.2.1 Contexte qui a favorisé les découvertes.....	38
2.2.2 Variations et marqueurs génétiques.....	39
2.2.2.1 Les répétitions en tandem	39
2.2.2.2 Les SNP	40
2.3 Stratégie d'identification de facteurs de risque génétique sur les MII.....	44
2.3.1 Études de liaison: les premières découvertes génétiques des MII	44
2.3.1.1 <i>NOD2</i>	48
2.3.1.2 <i>IBD5</i>	49
2.3.1.3 Le complexe majeur d'histocompatibilité	49
2.3.2 Études d'association: la révolution pangénomique.....	50
2.3.2.1 Les études d'association par gènes candidats	51
2.3.2.2 Les études d'association pangénomiques	51
2.3.3 L'héritabilité manquante: stratégie d'identification des autres facteurs de risque génétique	54
2.4 Des découvertes génétiques qui ont mené à une meilleure compréhension des mécanismes sous-jacents au développement des MII	59

2.4.1 La voie du récepteur de l'interleukine 23 (IL23R)	59
2.4.2 L'autophagie et les MII.....	62
Chapitre 3 La région candidate 19p	65
3.1 Identification du locus.....	65
3.2 Études par gènes candidats sur le chromosome 19	68
3.3 Hypothèses et objectifs du projet	70
Bibliographie.....	72
RÉSULTATS	95
Chapitre 4 MAST3: facteur de risque aux maladies inflammatoires de l'intestin	96
ARTICLE 1	98
Contribution des auteurs	99
Abstract	100
Introduction.....	101
Results	103
Discussion	108
Materials and methods	111
Acknowledgments.....	116
Tables and Figures	117
Supplementary Data	128
References.....	131
Chapitre 5 MAST3: modulateur d'inflammation.....	135
ARTICLE 2.....	137
Contribution des auteurs	138
Abstract	139
Introduction.....	140
Results	142
Discussion	147
Materials and Methods.....	151
Acknowledgements.....	155

Tables and Figures	156
Supplementary Data.....	165
References.....	178
DISCUSSION.....	182
Chapitre 6 Discussion	183
6.1 Résumé du projet et mise en contexte.....	183
6.1.1 MAST3 : facteur de risque génétique aux MII	183
6.1.2 Premier lien fonctionnel immunitaire: MAST3 et NF- κ B.....	186
6.1.3 MAST3 influence l'expression de gènes pro-inflammatoires et anti-inflammatoires	188
6.1.4 Certains gènes dont l'expression est modulée par MAST3 sont aussi des facteurs de risque aux MII connus.....	190
6.1.5 Les gènes dont l'expression est modulée par MAST3 sont surexprimés dans les tissus enflammés typiques des MII	191
6.2 Rôle de la kinase MAST3 dans l'inflammation.....	193
6.2.1 Modèle fonctionnel	193
6.2.2 Effets biologiques	196
6.3 Perspectives.....	199
6.3.1 Limitations de l'étude d'association	199
6.3.2 Limitations des études fonctionnelles.....	202
6.3.3 Impact de mon étude sur la recherche sur les MII.....	203
6.3.4 De la génétique au fonctionnel	204
Conclusion	206
Bibliographie.....	208
ANNEXES	216

Liste des tableaux

INTRODUCTION

Tableau 1. Comparaison de la MC et la CU	22
Tableau 2. Premiers loci MII identifiés par liaison.....	47
Tableau 3. Résumé des résultats de liaison en fonction du phénotype	67

RÉSULTATS

Article 1

Table 1. Study cohorts	117
Table 2. Top association	118
Table 3. Association results for <i>MAST3</i> coding SNPs.....	120
Supplementary Table 1: Quality control steps and results from the screening stage...	129
Supplementary Table 2: Quality control steps and results from the replication stage.	129

Article 2

Table 1. Differentially expressed genes (Fold increase>2) in the <i>MAST3</i> overexpression model.....	156
Table 2. Effect of a <i>MAST3</i> stable knockdown (KD) on expression of genes in stimulated THP1 cells confirms results from the overexpression model.....	158
Table 3. <i>MAST3</i> stable knockdown (KD) influences the expression of immune genes in stimulated THP1 cells.	159
Table 4. <i>MAST3</i> -regulated gene set enrichment tests in patient biopsies.....	160
Table 5. Comparison of the expression of the <i>MAST3</i> -regulated genes between BPN and BPI.....	161
Supplementary Table 1. All gene significantly differentially expressed ($P < 0.05$) in cells overexpressing the <i>MAST3</i> gene.....	165
Supplementary Table 2. Major Gene Ontology biological processes involving genes upregulated in our <i>MAST3</i> overexpression model.	172

Supplementary Table 3. Comparison of the expression of gene from the MAST3 regulated gene set between BC and BPN.....	173
Supplementary Table 4. Sequences of the RT qPCR primers.....	174

ANNEXE I

Tableau I1. Loci MII confirmés par les études d'association pangénomiques.....	218
Tableau I2. Gènes-candidats correspondant aux loci MC.....	222
Tableau I3. Gènes-candidats correspondant aux loci CU.....	225

Liste des figures

INTRODUCTION

Figure 1. Le système digestif	3
Figure 2. Inflammation contrôlée.....	6
Figure 3. Structure et fonction immunitaire de la paroi intestinale	8
Figure 4. La signalisation via les TLR.....	14
Figure 5. Photos de sections d'intestin grêle sain (gauche) et avec maladie de Crohn (droite).....	21
Figure 6. Photos de sections de côlon sain (gauche) et avec colite ulcéreuse (droite) ...	23
Figure 7. Manifestations extra-intestinales les plus communes affectant les patients souffrant de MII	25
Figure 8. Déséquilibre de liaison	42
Figure 9. Représentation classique d'étude de liaison et d'association	45
Figure 10. Loci de susceptibilité à la MC connus.....	53
Figure 11. Tailles requises des cohortes pour des études pangénomiques	56
Figure 12. La voie IL23R et l'inflammation	61
Figure 13. Une étude de liaison pangénomique identifie un pic de liaison sur le chromosome 19p.....	66

RÉSULTATS

Article 1

Figure 1. Association and LD analyses in the <i>MAST3</i> region.....	121
Figure 2. <i>MAST3</i> correlation neighbourhood.....	123
Figure 3. Schematic representation of the <i>MAST3</i> protein	124
Figure 4. <i>MAST3</i> mRNA expression in human cell lines and primary immune cells .	125
Figure 5. <i>MAST3</i> knockdown attenuates LPS-mediated NF- κ B activation.....	126
Supplementary Figure 1. <i>MAST3</i> expression in primary mouse immune cells..	128

Article 2

Figure 1. MAST3-regulated genes are involved in critical immune function.162
Figure 2. MAST3 regulates gene expression through NF- κ B1.....163
Supplementary Figure 1. Effect of *MAST3* overexpression on MAST3 protein levels at different times post-transfection.....175
Supplementary Figure 2. RT qPCR validation of expression chip results.....176
Supplementary Figure 3. Effect of a stable knockdown of the *MAST3* gene on MAST3 protein levels.....177

DISCUSSION

Figure 1. Cibles potentielles de MAST3 dans la voie NF- κ B.195

ANNEXE II

Figure II1. Expression of *MAST3* mRNA post-transfection in *MAST3* overexpressed cells.228
Figure II2. Effect of *MAST3* transfection on MAST3 protein levels at different times post-transfection.229
Figure II3. Unsupervised hierarchical clustering of normalized expression data.230
Figure II4. Effect of a KD of *MAST3* on *MAST3* mRNA levels.....231
Figure II5. Effect of a stable *MAST3* knockdown on MAST3 protein levels at different times post-transfection.232
Figure II6. *MAST3* overexpression data normalization.....233
Figure II7. Biopsy expression data normalization234

Liste des abréviations

λ_s , risque relative associé à la fratrie

ADN, acide désoxyribonucléique

ARN (RNA), acide ribonucléique (*ribonucleic acid*)

ARNm (mRNA), ARN messenger (messenger RNA)

CNV, variation en nombre de copies

CD#, *cluster of differentiation #*

CTD, cellule dendritique

CU, colite ulcéreuse (*ulcerative colitis*)

DL (LD), déséquilibre de liaison (*linkage disequilibrium*)

DRO, dérivés réactifs de l'oxygène

DSS, *dextran sulfate sodium*

EAP (GWAS), étude d'association pangénomique
(genome-wide association study)

HEK, cellules embryonnaires de reins (*human embryonic kidney cells*)

HLA, antigènes des leucocytes humains (*human leukocyte antigen*)

IC, intervalle de confiance

Ig, immunoglobuline

IL, interleukine

Kb, kilobase

LOD, *logarithm of the odds*

LPS, lipopolysaccharide

MAST3, *microtubule associated serine/threonine kinase 3*

MDP, *muramyl dipeptide*

MC (CD), maladie de Crohn (Crohn's disease)

Mb, megabase

MII (IBD), maladies inflammatoires de l'intestin (*inflammatory bowel diseases*)

NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases

NK, cellules tueuses naturelles

NF- κ B, *nuclear factor-kappa B*

NOD2, *nucleotide-binding oligomerization domain containing 2*

OR, rapport de cote (*odds ratio*)

PAMP, *pathogen-associated molecular pattern*

Pb (bp), pair de base (base pair)

PCR, *polymerase chain reaction*

PGH, projet génome humain

PMA, *phorbol 12-myristate 13-acetate*

PP, plaque de Payer

PRR, *pattern recognition receptors*

qPCR, PCR quantitatif en temps réel (*real time quantitative PCR*)

QC, contrôle de qualité (*quality control*)

RT, répétition en tandem

RT qPCR, *reverse transcriptase real time quantitative PCR*

shRNA, *small hairpin RNA*

SNP, single nucleotide polymorphism

TLR, *Toll-like receptor*

T_H, lymphocytes T auxiliaires (T helper cells)

Tregs, lymphocytes T régulateurs

TS, type sauvage

*Pour ma sœur, unique et originale,
Mon poème*

Remerciements

Six ans c'est à la fois court et long. Court parce que c'est déjà fini. Long parce que c'est enfin fini. À l'heure des bilans, je peux admettre, somme toute, que ce fut un parcours généralement heureux avec quelques moments plus difficiles que je chéris parce qu'ils m'ont permis de grandir. Ces six dernières années m'ont fait progresser de façon exponentielle tant au plan scientifique qu'au plan humain. J'ai presque l'impression d'avoir enfin atteint l'âge adulte. Mais, je ne me suis pas rendue là toute seule, il y a beaucoup de gens, qui sur la route, m'ont tenu la main.

Ce n'est pas un secret, il y a beaucoup de roulement dans le personnel d'un laboratoire et en six ans, j'ai vu des amis partir, mais j'ai eu la chance d'en rencontrer beaucoup. Je dois d'abord remercier John Rioux, mon directeur de thèse, d'avoir rendu possibles toutes ces rencontres en prenant le pari de m'accepter dans son laboratoire. Je réalise la chance que j'ai eu de travailler avec un chef de file en génétique, et je pense qu'avec le recul et plus loin dans ma carrière, je vais prendre encore plus conscience de la dimension de ce que j'ai appris grâce à vous. Merci infiniment.

Nancy, merci de m'avoir accueillie dans le labo, merci de m'avoir rassurée et épaulée dans les débuts;

Céline, rythme cardiaque au repos 42, énergie pure, mais éthique de travail incroyable, merci de ta gentillesse et merci de m'avoir fait rire;

Phil, parce que tu es resté le plus longtemps, j'ai l'impression d'avoir eu un frère au labo, on a eu quelques frictions, mais elles ont toujours été passagères, merci (aussi t'as des mollets de fous!);

Gab, en si peu de temps tu t'es rendue essentielle. Tu es charmante, mais dans le plus noble des sens;

Sylvain, Guy, merci pour les précieux conseils, sollicités ou non ;-)

Tout le monde du labo de Guillaume, vous avez su apporter un regain de dynamisme et d'énergie;

Catherine, Jasmine, mes compagnes dans l'aventure du Ph. D., merci de vous en être sortie et de m'aider à en faire autant!

Marie-Pierre, Ken, Fred, Gen, Clo, mes amis. Merci de la collaboration, de l'appui et de la confiance que vous m'accordez, et on continue à se voir, parce qu'à chaque fois c'est mémorable!

Marcia, I am sorry I am such a lousy French teacher. I wish I could have done more because I feel like you have given me so much. You are one of the greatest teachers of Life I have ever had. Thank you.

Et parce que c'est probablement mon dernier diplôme (ou en tout cas le plus haut grade que j'attendrai jamais), j'ai envie de remercier tous les professeurs que j'ai eus, de ceux qui m'ont appris à écrire et compter, à ceux qui m'ont enseigné Darwin et Mendel, merci pour l'inspiration et le désir d'aller plus loin.

Merci à ma sœur et mon frère, Stéphanie et Gaël et à ma merveilleuse coloc et amie Susie pour l'appui constant et le sentiment de se savoir aimé. Merci à mes premiers professeurs, ma mère et mon père, qui sont aussi fiers aujourd'hui qu'il y a 25 ans dans la cour de l'école...

Merci.

Avant-Propos

Généralement quand je présente ce que je fais dans la vie, j'aborde le sujet des maladies inflammatoires de l'intestin (MII). Les personnes de professions non-médicales se demandent généralement de quoi il s'agit et puis je mentionne la maladie de Crohn et la colite ulcéreuse et soudain ils connaissent quelqu'un qui en est atteint. Les MII sont des maladies communes. Ce constat se révèle d'autant plus vrai au Canada où malheureusement, on rapporte des taux de prévalence et d'incidence de maladies de Crohn parmi les plus élevés au monde¹. Concrètement, plus de 200 000 Canadiens sont atteints de MII. De plus, ces maladies sont décrites, reconnues et étudiées depuis des décennies: 1859 pour la colite ulcéreuse et 1932 pour la maladie de Crohn^{2;3}. Malgré le fardeau humain et financier (dans un contexte de système de santé publique) que représentent les MII, et malgré le fait qu'elles soient reconnues depuis longtemps, elles représentent toujours un défi au niveau du diagnostic et du traitement. Les traitements typiques sont majoritairement basés sur l'apaisement des symptômes dont souffre le patient plutôt que l'élimination des causes de ces maladies. Les raisons? Les mécanismes moléculaires et cellulaires précis à l'origine de ces maladies ont été obscurs pendant très longtemps et le sont toujours jusqu'à un certain point, toutefois cet état est appelé à changer...

La révolution génétique propulsée par les avancées scientifiques des dernières années, notamment la publication de la séquence complète du génome humain et les technologies qui en découlent, a permis d'attaquer la recherche sur les causes des maladies d'un nouvel angle, celui de la génétique. Cette révolution est toujours en cours avec des innovations continues que ce soit au niveau des techniques de génotypage et de séquençage qu'au niveau des marqueurs génétiques utilisés pour étudier la variabilité (des microsatellites, des SNP, des variantes en nombre de copies). On peut maintenant

relier des portions de la séquence du génome, des loci, à des maladies. Les personnes portant une certaine combinaison de variantes associées localisées dans ces loci sont susceptibles de développer la maladie. Les MII sont parmi les maladies communes, les plus étudiées au niveau génétique. Ce travail a porté fruit avec l'identification et la validation d'une centaine de facteurs de risque aux MII. On sait maintenant que les MII se déclenchent suite à des interactions entre des prédispositions génétiques et des facteurs environnementaux qui contribuent à la dérégulation de la réponse immunitaire aux antigènes de la flore intestinale. Les mécanismes moléculaires précis ne sont pas complètement connus, mais les études d'association pangénomiques ont permis d'identifier des dizaines de loci et de gènes de susceptibilité aux MII. Ces gènes font partie de voies biologiques dont le dévoilement permettra d'identifier des biomarqueurs pour faciliter le diagnostic et des cibles pour orienter la recherche thérapeutique. L'émergence de traitements biologiques est un témoignage des progrès qu'a amené l'étude des voies biologiques.

Chaque étude génétique a le potentiel d'orienter les recherches fonctionnelles sur de nouvelles voies biologiques ou d'améliorer les connaissances sur des voies connues. La suite logique et essentielle de ces études génétiques passe par la caractérisation fonctionnelle des gènes associés qui vise à expliquer les mécanismes moléculaires impliqués dans la pathogenèse des maladies. Dans le cadre de mes études doctorales, j'ai contribué à clarifier une partie du puzzle moléculaire que sont les MII. Je présente dans cet ouvrage, mon étude génétique qui a mené à l'identification d'un gène de susceptibilité aux MII appelé MAST3 et mes études fonctionnelles qui ont permis d'impliquer MAST3 dans la régulation du facteur de transcription NF- κ B et dans la régulation de l'expression de gènes immunitaires qui représentent une signature biologique de l'inflammation typique des MII.

Bonne lecture.

INTRODUCTION

Chapitre 1

Le système digestif dans la santé et les maladies inflammatoires

L'absorption par le système digestif est un processus d'une efficacité remarquable: moins de 5% des glucides, des acides gras et des protéines ingérés sont excrétés par les adultes ayant une alimentation normale⁴. Les types de nourritures et les apports nutritionnels à travers le monde présentent des différences considérables, mais l'absorption intestinale s'adapte et demeure efficace. La digestion est déclenchée par une phase cérébrale c'est-à-dire que la vue, l'odeur ou même l'idée de la nourriture amorce la digestion. Cet ouvrage ne traitera toutefois pas de ce processus fascinant, pour une bonne revue sur la digestion et l'absorption des nutriments et des vitamines, consultez Farrell et al.⁴

Dès la quatrième semaine de développement embryonnaire, les grandes lignes du système digestif sont mises en place (Figure 1). Au troisième trimestre de gestation le fœtus peut même commencer à avaler du liquide amniotique et à en y absorber de petites quantités de sucres⁵. Cette activité sert à préparer le système digestif à fonctionner à plein régime. Cependant, c'est seulement après la naissance, lorsque l'enfant commence à absorber de la nourriture provenant de l'environnement extérieur, que le système digestif se confronte au défi majeur qu'il aura à gérer tout au long de la vie soit extraire et absorber les nutriments essentiels tout en protégeant l'organisme des attaques de l'extérieur.

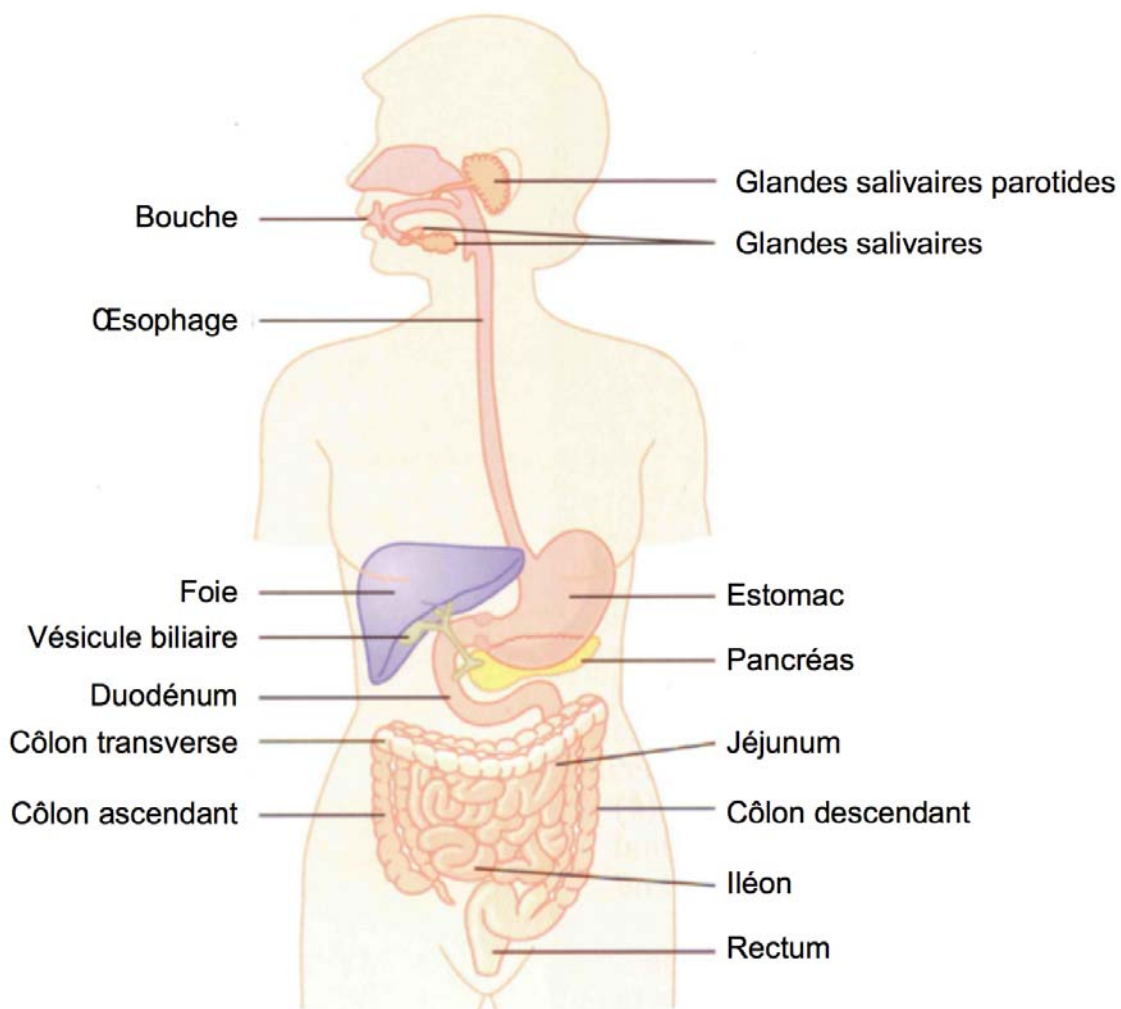


Figure 1. Le système digestif

Le système digestif humain est un tube d'une longueur approximative de 9 m. Son rôle premier est la digestion des aliments et l'assimilation des nutriments. Le processus de digestion est amorcé dans la bouche par la salive produite par les glandes salivaires, riches en enzymes digestives. Le bol alimentaire se rend à l'estomac via l'œsophage. L'estomac le digère de façon mécanique (brassage) et chimique (sucs gastriques). Le résultat est une pâte appelée chyme qui passe dans l'intestin grêle (composé du duodénum, du jéjunum et de l'iléon) où la majorité de l'absorption se produit. Le matériel non absorbé passe dans le côlon qui se charge de récupérer l'eau et certaines vitamines. Le matériel restant se mêle au mucus et aux bactéries de la flore intestinale pour devenir matière fécale qui est stockée temporairement dans le rectum avant d'être éliminée par l'anus. Figure adaptée de Guyton et Hall⁶.

1.1 Biologie du système digestif en santé : inflammation contrôlée

1.1.1 Microbiome bactérien

Le système gastro-intestinal a pour mission première d'absorber les nutriments de la nourriture digérée et d'éliminer les déchets du processus de digestion. Le processus d'absorption se déroule dans la couche supérieure du tube digestif qu'on appelle l'épithélium gastro-intestinal. Cette couche mince (d'une seule cellule d'épaisseur) et perméable est texturée de villosités qui en décuplent la surface. Ainsi, la surface totale de l'intestin grêle est d'environ 300m² chez l'humain⁷.

Toute cette surface grouille de vie puisque la lumière intestinale abrite un milieu riche en microorganismes. On estime que la densité des microorganismes dans le côlon est de 10¹¹-10¹² cellules/ml de contenu intestinal. Cette concentration se traduit par 1.5 kg de bactéries dans le système digestif. Dans l'ensemble du corps humain, on estime que les bactéries sont au moins 10 fois plus nombreuses que les cellules humaines. Les différentes espèces bactériennes du corps vivent en symbiose⁸. Uniquement dans le tube digestif, cette communauté de micro-organismes est composée d'au moins 1000 espèces bactériennes distinctes⁹. Des analyses génétiques récentes ont permis de faire la lumière sur la nature de la flore microbienne du système digestif¹⁰. Les résultats de ces analyses montrent une grande diversité entre les individus au niveau des espèces bactériennes avec lesquelles ils vivent. Cependant, certains embranchements bactériens comme les *firmicutes* et les *bacteroidetes* sont prédominants dans les intestins de tous les vertébrés¹⁰.

Certaines de ces bactéries assistent leur hôte dans des fonctions essentielles comme la digestion des sucres complexes et la production de vitamines et d'autres molécules. Toutefois, la quantité astronomique de bactéries menace sans cesse la santé de

l'organisme. La fonction du système immunitaire est de maintenir l'homéostasie avec le microbiome intestinal en empêchant les bactéries pathogènes, mais aussi les bactéries commensales de traverser la barrière intestinale. Dans le cas d'une invasion, le système immunitaire déclenche une inflammation contrôlée.

L'inflammation est un processus normal et essentiel qui permet au corps de se protéger des invasions pathogènes. La paroi intestinale a un rôle clé dans cette protection. En plus de son rôle dans l'absorption des nutriments, la paroi intestinale est un organe immunitaire. En effet, elle est constituée de cellules immunitaires dont la diversité et le nombre est sans égale dans les autres organes du corps¹¹. De plus, l'inflammation implique un influx de cellules immunitaires vers les régions touchées. Conséquemment, on parle d'inflammation contrôlée pour exprimer l'idée que ces cellules immunitaires présentes en grands nombres, permettent d'assurer une protection essentielle au système digestif et indirectement à tout le corps. Les mêmes processus sont impliqués dans l'inflammation pathologique et dans l'inflammation contrôlée. Toutefois dans ce derniers cas, l'organisme limite les dommages causée via de l'autorégulation.

Dans la lumière intestinale, le système immunitaire est confronté à des antigènes de plusieurs sources (alimentaires, du soi, de la flore intestinale, de microbes pathogènes). Le système immunitaire gère ces antigènes via un processus d'inflammation physiologique et contrôlée (Figure 2). L'organigramme de la figure 2 présente une classification synthétique des différents niveaux de défense impliqués dans l'inflammation contrôlée pour mieux comprendre l'importance et le rôle de chacune d'eux. Au départ, on peut séparer le contrôle des stimuli antigéniques en deux types principaux : défense physique et défense biologique.

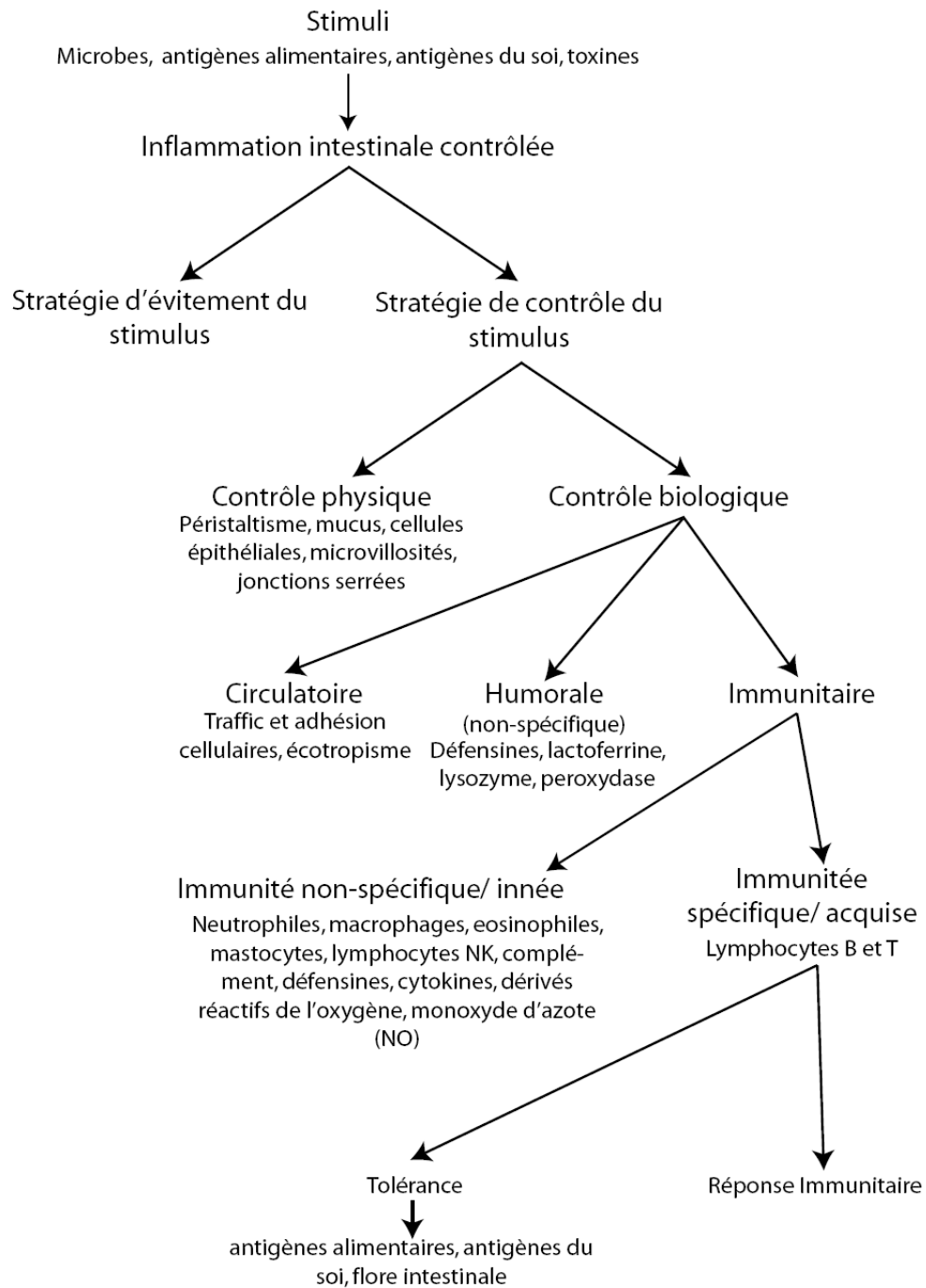


Figure 2. Inflammation contrôlée

Représentation schématique des différentes facettes de l'inflammation physiologique contrôlée. Figure adaptée de ¹¹

1.1.2 Les mécanismes physiques de défense

Les mécanismes physiques comprennent le péristaltisme, la paroi intestinale et le mucus.

1.1.2.1 Le péristaltisme

Le péristaltisme intestinal se caractérise par l'action coordonnée des muscles intestinaux pour propulser le bol alimentaire ou chyme en aval dans le système. Il vise à prévenir l'accumulation de matière alimentaire, de mucus et de bactéries et ce faisant réduit la possibilité d'absorber un surplus d'antigènes qui causeraient de l'inflammation¹². Des études ont montré une augmentation de la quantité de microorganismes intestinaux chez les rats dont le péristaltisme est bloqué chimiquement¹³.

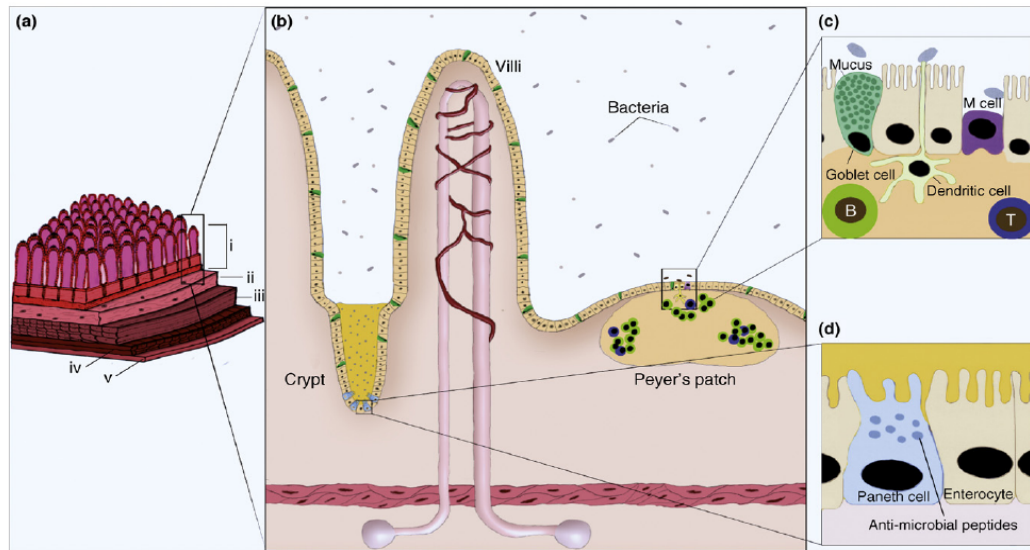


Figure 3. Structure et fonction immunitaire de la paroi intestinale

La paroi intestinale a un rôle actif dans la protection contre les microorganismes présents dans le système digestif. (a) La paroi intestinale est composée de différentes couches de tissus: la muqueuse est directement en contact avec la lumière (i), la sous-muqueuse (ii) entoure et supporte la muqueuse tout en la reliant aux couches musculaires sous-jacentes (iii-iv), finalement la séreuse entoure les couches musculaires. (b) L'épithélium intestinal est texturé de cryptes et de villosités et est composé de différents types cellulaires. (c) Les cellules en gobelet sécrètent les mucines qui forment un film protecteur à la surface de l'épithélium. (d) Les cellules de Paneth produisent différentes classes de peptides antimicrobiens comme les défensines et les lysozymes. Les lymphocytes représentant les principaux acteurs de la réponse immunitaire spécifique sont organisés en tissus lymphoïdes comme les plaques de Payer (PP) (b) et les follicules lymphoïdes isolés. Ces lymphocytes sont principalement des lymphocytes T effecteurs et des lymphocytes B différenciés (c). Souvent localisées près des PP, les cellules M et les cellules dendritiques échantillonnent l'environnement et en présentent un aperçu aux lymphocytes agissant donc comme cellules présentatrices d'antigènes ¹⁴.

1.1.2.2 La paroi intestinale

La paroi intestinale agit comme barrière mécanique contre les invasions. Elle est composée de plusieurs couches de tissus (Figure 3). La muqueuse est directement en contact avec la lumière intestinale. Les jonctions serrées entre les cellules de la muqueuse empêchent les microbes de pénétrer l'épithélium. Immédiatement sous la muqueuse se trouvent la sous-muqueuse, puis deux couches musculaires et enfin la séreuse qui recouvre les couches superficielles (Figure 3A). Un élément clé du maintien de l'homéostasie avec le microbiome intestinal est de minimiser le contact entre les microorganismes et l'épithélium par le biais de substances biochimiques qui renforcent la barrière épithéliale. Ces substances sont sécrétées par à peu près toutes les cellules spécialisées de l'épithélium (Figure 3B). Les cellules en gobelet sécrètent les glycoprotéines mucines qui forment le mucus (section 1.1.2.3). Les entérocytes et les cellules de Paneth produisent de façon constitutive, les défensines- α . Les défensines sont des protéines qui favorisent l'élimination des bactéries par les cellules phagocytaires en attaquant l'intégrité de la membrane bactérienne par le biais de la formation de pores¹⁵.

1.1.2.3 Le mucus

Le mucus, quant à lui, forme une couche protectrice qui recouvre l'épithélium sur une épaisseur d'environ $150\mu\text{M}$ ¹⁶. Les cellules en gobelet qui font partie des cellules épithéliales intestinales (Figure 3) sécrètent les glycoprotéines mucines qui s'assemblent pour former la structure gélatineuse qui caractérise le mucus. Les gènes qui encodent les différentes protéines mucines s'expriment de façon différentielle selon la localisation intestinale ce qui suggère que les différentes mucines ont des rôles distincts dans la protection de la muqueuse¹⁷. Le mucus prévient le contact direct entre la muqueuse intestinale et les antigènes, forme un piège où peuvent s'enliser les microbes, mais aussi favorise la rétention à la surface épithéliale des molécules de défenses antibactériennes¹⁸.

1.1.3 Les mécanismes biologiques de défense

Les mécanismes biologiques se caractérisent par des manifestations circulatoires, humorales non-spécifiques et immunitaires.

1.1.3.1 Manifestations circulatoires

On parle de manifestations **circulatoires** pour décrire les phénomènes de migrations et d'écotropisme (*homing*) des cellules immunitaires. Dans des circonstances normales, les leucocytes circulent constamment dans tout le corps et leur mouvement est étroitement régulé et directionnel. Par exemple, les lymphocytes patrouillent l'organisme à la rencontre des antigènes en circulant dans les vaisseaux sanguins puis dans les tissus et dans le système lymphatique. Ils acquièrent une préférence pour un tissu particulier à travers leur première rencontre avec un antigène et ils peuvent revenir vers ce lieux¹⁹.

On appelle ce phénomène *écotropisme*. Les macrophages et les granulocytes n'ont pas de mémoire antigénique, mais ils migrent à partir des vaisseaux sanguins vers les tissus en réponse à des signaux chimiques produits consécutivement à un traumatisme ou une infection. Par exemple, des études chez la souris ont montré que la *macrophage inflammatory protein 2* (MIP-2) aussi appelée *chemokine (CXC motif) ligand 2* (CXCL2) exprimée par les cellules épithéliales intestinales attire les macrophages²⁰.

Dans le cas des lymphocytes comme des macrophages cette migration dépend des interactions avec les molécules d'adhésion telles que *inter-cellular adhesion molecule 1* (ICAM-1)²¹.

1.1.3.2 Manifestations humorales non-spécifiques

L'organisme exerce également un contrôle des antigènes via certaines substances possédant des propriétés antimicrobiennes qui font partie de la circulation **humorale non-spécifique** c'est-à-dire qu'elles circulent dans tout le corps via les liquides corporels. Les défensines (section 1.1.2.2) en sont un exemple. La lactoferrine est une

autre de ces substances, il s'agit d'une glycoprotéine qui peut se lier au fer, un élément utilisé dans le métabolisme de nombreux organismes. La lactoferrine est synthétisée non seulement par les cellules épithéliales, mais elle est aussi présente dans les granules des neutrophiles²². La lactoferrine augmente également la phagocytose et l'adhérence cellulaire et influence la sécrétion de cytokines proinflammatoires^{23; 24}. Le lysozyme et les peroxydases sont d'autres exemples d'enzymes qui possèdent des propriétés antimicrobiennes²⁵.

1.1.3.3 Manifestations immunitaires

Au cœur du système de protection du corps, se place la défense **immunitaire**. Deux principaux mécanismes de défense immunitaire œuvrent dans le but de protéger l'organisme : la réponse immunitaire innée et non-spécifique et la réponse immunitaire spécifique. La réponse innée est une réponse générale, rapide et efficace, mais qui ne confère pas d'immunité à long terme. La réponse immunitaire spécifique est plus tardive et comme son nom l'indique, spécifique. Elle implique une plus forte réponse immunitaire et une mémoire de la signature des pathogènes qui ont été une menace. Ces deux mécanismes travaillent conjointement dans le microcosme intestinal afin de préserver l'homéostasie.

Réponse immunitaire innée

La réponse immunitaire innée est orchestrée par divers mécanismes qui incluent le système du complément, des peptides antimicrobiens, et différents types de leucocytes. L'élément clef de ces mécanismes est qu'ils sont activés immédiatement pour rapidement contrôler et contenir l'élément pathogène. Les leucocytes impliqués dans la réponse immunitaire innée comprennent :

- Les lymphocytes NK (pour *Natural Killer*) représentent un sous-type de lymphocytes qui ont pour fonction de détruire les cellules anormales (cancéreuses, infectées par un virus, etc) en provoquant l'apoptose²⁶. Ils

produisent également plusieurs cytokines comme l'interféron γ (INF- γ) et le *Tumor Necrosis Factor α* (TNF α), suite à l'activation de leurs récepteurs membranaires²⁷. Ces cytokines attirent et assistent les cellules phagocytaires présentatrices d'antigènes.

- Les granulocytes neutrophiles et éosinophiles sont des leucocytes qui sont caractérisés par des granules intra-cytoplasmiques. Ces granules contiennent des agents antimicrobiens (comme la lactoferrine) qui sont relâchés lors de l'ingestion des microorganismes. Les granulocytes sécrètent également des cytokines comme TNF α et l'interleukine 8 (IL-8)²⁸.
- Les monocytes et leur forme différenciée, les macrophages, sont des cellules phagocytaires présentatrices d'antigènes qui sécrètent également des cytokines comme TNF α et interleukine 10 (IL-10)²⁹.
- Les cellules dendritiques (CDT) relient les réponses immunitaires innées et spécifiques. Leur principale fonction est d'interagir avec les pathogènes et d'en présenter les antigènes aux cellules du système immunitaire spécifique. Cependant, les CDT ont aussi un rôle dans l'immunité innée puisqu'elles ont également la capacité de sécréter des cytokines dont la nature dépend de leur localisation dans l'organisme et leur niveau de différenciation. Les CDT du plasma sont potentiellement impliquées dans la lutte contre les virus via la sécrétion de la cytokine interleukine 1 (IL-1) et les CDT myéloïdes sécrètent la cytokine interleukine 12 (IL-12) en abondance^{30; 31}.

Dans le système immunitaire inné, la reconnaissance des molécules pathogènes est directement inscrite dans le génome. C'est-à-dire que les cellules d'un même type possèdent les mêmes récepteurs qui reconnaissent les mêmes motifs moléculaires bactériens. Ce n'est pas le cas chez les cellules impliquées dans la réponse spécifique (section 1.1.3.3 p.15). Les récepteurs impliqués dans la reconnaissance des motifs cellulaires bactériens sont appelés PRR (*pattern recognition receptors*). Les PRR sont activés par différents types de motifs bactériens appelés PAMP (*pathogen-associated*

molecular pattern). Les *Toll-like receptors* (TLR) représentent un type de PRR. Ces récepteurs ont pour fonction de cribler l'environnement cellulaire pour détecter la présence de menaces bactériennes. Ils sont présents en abondance chez les monocytes, les macrophages et les CTD, mais aussi chez d'autres types cellulaires comme les lymphocytes B et les cellules de l'épithélium intestinal. À l'heure actuelle, on distingue dix TLR chez l'humain³²⁻³⁷. Les TLR1, TLR2, TLR4, TLR5, TLR6, TLR11 sont localisés à la surface cellulaire. Ils répondent à des ligands spécifiques qui, pour la plupart, composent la membrane bactérienne, tels que des lipides, lipoprotéines ou des protéines. Par exemple, les TLR4 conjugués à leur corécepteur *Myeloid Differentiation Factor 2* (MD-2) reconnaissent les lipopolysaccharides (LPS), une famille de molécules constituante de la membrane externe des bactéries à Gram négatif³⁸. Les TLR3, TLR7, TLR8 et TLR9 sont exprimés dans les vésicules intracellulaires comme le réticulum endoplasmique, les endosomes, les lysosomes et les endolysosomes. Ces TLR sont activés par des acides nucléiques bactériens³⁹. Suite à l'activation du récepteur par son ligand, une cascade de signalisation est enclenchée (Figure 4). Celle-ci mène le plus souvent à l'activation du facteur de transcription NF- κ B qui à son tour active un programme de transcription de gènes (*TNF*, *IL8*, etc) ayant pour but de déclencher des mécanismes d'élimination de la menace bactérienne à un plus grande échelle⁴⁰. Ces mécanismes incluent la synthèse de dérivés réactifs de l'oxygène, l'activation du complément, la sécrétion de chémokines et de cytokines impliquées dans la chimiotaxie des macrophages phagocytaires et la sécrétion de protéines antimicrobiennes comme les défensines.

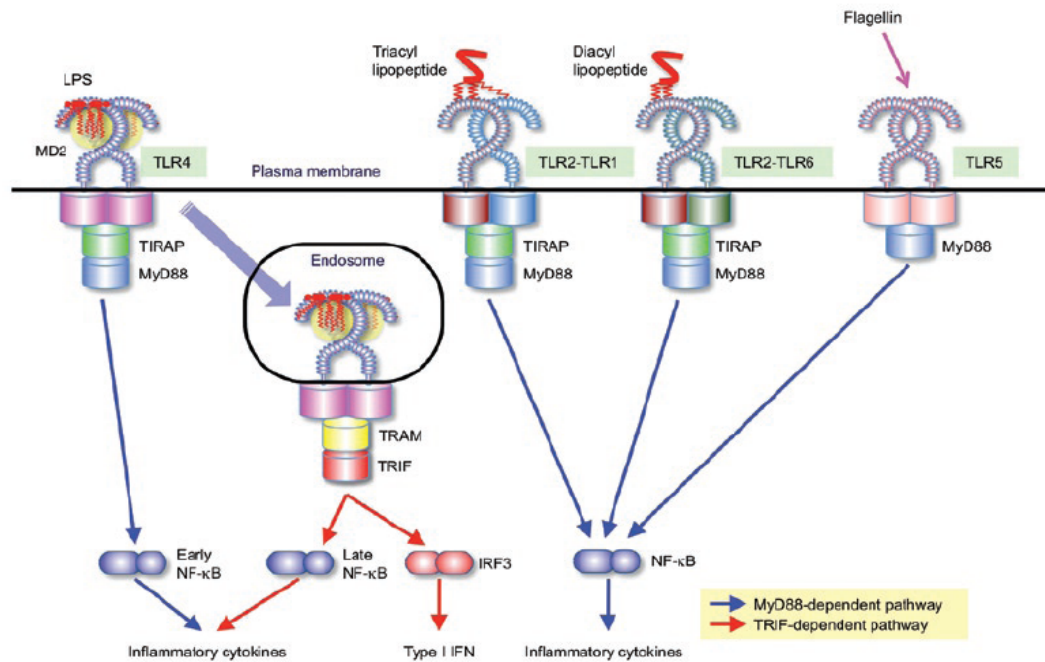


Figure 4. La signalisation via les TLR

Différents types de récepteurs TLR sont activés par des ligands distincts. Les TLR4 activés par les LPS enclenchent une cascade de signalisation qui passe par l'activation de différents partenaires dont TRIF *Related Adaptor Molecule* (TRAM) et *TIR domain-containing adaptor molecule 1* (TRIF). L'aboutissement de cette cascade est l'activation du facteur de transcription NF-κB³⁹.

D'autres types de récepteurs sensibles aux motifs moléculaires bactériens existent. Spécifiques aux vertébrés supérieurs, les récepteurs *nucleotide-binding oligomerization domain containing 1* et *2* (NOD1 et NOD2) sont exprimés par plusieurs types cellulaires comme les macrophages, les cellules dendritiques, les cellules de Paneth, les kératinocytes, les cellules épithéliales de l'intestin, des poumons et de la cavité buccale⁴¹. Principalement identifiés comme récepteurs cytosoliques, mais des études récentes révèlent leur présence à la membrane plasmique^{42; 43}. Le ligand de NOD1 est le *D-glutamyl-meso-diaminopimelic acid* (iE-Dap) et celui de NOD2 est le *muramyl dipeptide* (MDP), tous deux des constituants d'un polymère appelé peptidoglycane qui compose la membrane bactérienne externe. La stimulation de NOD1 ou NOD2 par leur ligand entraîne l'activation de NF-κB et de la voie des *mitogen-activated protein kinase* (MAPK) ce qui conduit, comme avec les récepteurs TLR, à la transcription d'un programme de gènes ayant comme fonction d'éliminer la menace antigénique.

La réponse immunitaire spécifique

Les récepteurs des lymphocytes B et T sont des membres de la famille des immunoglobulines qui sont impliqués dans la **réponse immunitaire spécifique**. Les immunoglobulines ne sont pas encodées telles quelles dans le génome et ne sont pas prédestinés à reconnaître un antigène en particulier. La stratégie du système immunitaire est de générer un très grand éventail de récepteurs au hasard par réarrangement génique, de sélectionner les lymphocytes naïfs qui possèdent les récepteurs utiles par la présentation d'antigènes pour lesquels ils sont spécifiques et de leur faire subir une expansion clonale⁴⁴. Ce processus de sélection nécessite du temps. Il faut en moyenne 4 à 5 jours avant que l'expansion clonale soit complète et que les lymphocytes « naïfs » se soient différenciés en lymphocytes effecteurs. Pour cette raison, la réponse immunitaire spécifique est plus tardive que la réponse immunitaire non-spécifique⁴⁵.

Les lymphocytes B, sécréteurs d'immunoglobulines (Ig), jouent un rôle clé dans le contrôle de l'inflammation intestinale. Plus de 70% de ces lymphocytes se retrouvent

dans la muqueuse intestinale. De plus, il y a plus d'anticorps IgA sécrétés dans l'organisme que de tous les autres isotypes confondus⁴⁶. La production des anticorps est stimulée par les CDT qui échantillonnent l'environnement bactérien à la surface de l'épithélium intestinal. À la suite de quoi, les CDT présentent leurs antigènes bactériens aux lymphocytes B immatures, qui se différencient en plasmocytes et produisent des IgA spécifiques⁹. Les anticorps IgA traversent l'épithélium vers la lumière intestinale au moyen du transport intracellulaire via les récepteurs pIgR exprimés à la surface des cellules épithéliales⁴⁶. Ces IgA se lient aux bactéries et ainsi limitent leur association avec l'épithélium intestinal et empêchent leur pénétration dans les tissus. Les détails de la fonction des IgA dans ces mécanismes sont toujours méconnus, mais ils pourraient impliquer l'enlèvement des bactéries dans la couche de mucus et la promotion de la phagocytose⁴⁷. Les lymphocytes B agissent également comme cellules présentatrices d'antigènes. Une étude récente a montré que la coopération entre les cellules dendritiques et les lymphocytes comme cellules présentatrices d'antigènes augmentait la différenciation des lymphocytes T CD4+ en lymphocyte T régulateurs (Tregs)⁴⁸.

La muqueuse de l'intestin grêle et du côlon contient une population hétérogène de lymphocytes T dont environ 60-70% sont des lymphocytes CD4+ (*cluster of differentiation 4*) de type T auxiliaire (T_H pour *Helper*) et 20-30% sont des lymphocytes CD8+ de type T cytotoxique⁴⁹. La majorité sont oligoclonaux ayant été activés par les cellules présentatrices d'antigènes dans les tissus lymphatiques comme les plaques de Payer. Ils présentent un profil d'expression suggérant leur activation – expression de l'interleukine 2 (*IL2*), *interleukine receptor α (IL2RA)* et *interleukine receptor β (IL2RB)*⁵⁰. Les lymphocytes T auxiliaires (T_H) ont des fonctions essentielles dans la réponse immunitaire spécifique. Entre autres, ils assistent les lymphocytes B dans la production d'anticorps et la commutation isotopique, ils recrutent et activent les lymphocytes T cytotoxiques, les macrophages et les granulocytes⁵¹. Différents sous-types de T_H se distinguent par les cytokines qu'ils sécrètent et leur localisation tissulaire.

Les T_{H1} produisent de l'INF γ et sont impliqués dans l'activation des macrophages et l'élimination des pathogènes intracellulaires et les T_{H2} produisent les interleukines 4, 5, 10 et 13 (IL-4, IL-5, IL-10 et IL-13) et sont impliqués dans la production des anticorps de type IgE, le recrutement des éosinophiles et l'élimination des pathogènes extracellulaires⁵²⁻⁵⁴. Les T_{H17} produisent les interleukines 17, 21 et 22 et sont importants dans l'élimination des bactéries extracellulaires et des mycètes^{55;56}. Les lymphocytes T régulateurs représentent également un sous-type de lymphocytes T auxiliaires. Les Tregs contrôlent l'activation, la différenciation et la prolifération des autres lymphocytes T et répriment l'inflammation à la suite d'infection microbienne. Ils sécrètent les cytokines IL-10 et *transforming growth factor- β* (TGF β) qui sont toutes deux essentielles pour accomplir leurs fonctions anti-inflammatoires⁵⁷. De leur côté, les lymphocytes T CD8⁺ cytotoxiques contribuent à freiner la propagation de pathogènes en attaquant directement les cellules infectées et en provoquant leur apoptose⁵⁸.

1.1.4 La tolérance immunologique

Le tube digestif étant exposé à des antigènes provenant de la nourriture et des bactéries de la flore intestinale qui stimulent constamment la réponse immunitaire, un important constituant de l'inflammation contrôlée est la tolérance immunologique. Ce phénomène prévient les réactions immunitaires délétères causées par des protéines qui doivent être absorbées pour combler les besoins nutritionnels et par les antigènes de la flore intestinale commensale qui participent au métabolisme énergétique, à l'absorption des nutriments et à la santé en général⁵⁹. La tolérance immunologique implique la distinction des antigènes du Soi et du non-Soi. La distinction des antigènes se fait à deux niveaux. La **tolérance centrale** se produit durant la maturation des lymphocytes dans le thymus durant le développement fœtal. Les antigènes qui reconnaissent le Soi sont détruits⁶⁰. La **tolérance périphérique** est développée plus tard contre des antigènes (provenant de la nourriture, de la flore intestinale) qui sont rencontrés par les lymphocytes T mature. Différents mécanismes ont évolué pour gérer la tolérance

périphérique tels que l'anergie, la délétion et la suppression active. L'anergie est causée par la reconnaissance de l'antigène dans des conditions sous-optimales de costimulation. Sous ces conditions les lymphocytes T deviennent incapables de reconnaître l'antigène s'il leur est présenté plus tard par des cellules présentatrices d'antigènes dans des conditions adéquates⁶¹. La délétion se caractérise par l'élimination par apoptose des cellules ayant été exposées plusieurs fois au même antigène –une caractéristique importante de l'immunité intestinale. Cette élimination est contrôlée par le récepteur de mort cellulaire *TNF receptor superfamily, member 6* (Fas) qui est exprimé à de hauts niveaux sur les lymphocytes T intestinaux^{62; 63}. La suppression active est gérée par les lymphocytes T régulateurs qui sécrètent des cytokines anti-inflammatoires comme IL-10 et TGF β et ce faisant freinent l'activation et la fonction des autres cellules T⁵⁷. La tolérance immunologique est essentielle au contrôle de l'inflammation intestinale et la perte cette tolérance est un acteur majeur dans l'inflammation intestinale pathologique.

1.2 Les maladies inflammatoires de l'intestin

Dans un système digestif en santé, des réactions immunitaires contrôlées régulent les interactions entre les bactéries et le corps pour maintenir la tolérance immunologique et l'homéostasie intestinale. Dans certains cas, l'inflammation devient chronique, on parle alors de maladies inflammatoires de l'intestin (MII). Ces maladies regroupent deux types principaux soit la maladie de Crohn (MC) et la colite ulcéreuse (CU) qui diffèrent dans leurs manifestations cliniques (voir section 1.2.2).

1.2.1 Épidémiologie des maladies inflammatoires de l'intestin

Les taux d'incidence les plus élevés de MC et CU se trouvent dans les populations d'Amérique du Nord, du Royaume-Uni et d'Europe. En Amérique du Nord, les taux d'incidence de la MC sont estimés à 3,1-14,6 par 100 000 et ceux de la CU à 2,2-14,3 par 100 000⁶⁴. Dans les régions industrialisées ci-haut mentionnées, les taux se stabilisent alors qu'ils continuent d'augmenter dans les populations du sud de l'Europe, de l'Asie ainsi que dans la plupart des pays en voie de développement⁶⁵. Les taux de prévalence varient beaucoup en fonction des origines ethniques. En Amérique de Nord par exemple, les taux de prévalence les plus élevés se retrouvent chez les Juifs Ashkénazes⁶⁶. Dans cette région, les taux de prévalence pour la MC sont plus faibles chez les gens d'origine hispanique (4,1 par 100 000) ou asiatique (5,6 par 100 000) que pour les caucasiens (43,6 par 100 000) ou les Afro-Américains (29,8 par 100 000)⁶⁷. La MC est extrêmement rare en Amérique du sud et en Afrique, sauf en Afrique du sud où l'incidence dans la population caucasienne est de 2.6 par 100 000⁶⁴. La tendance est la même pour la CU. Des taux d'incidence plus élevés sont observés chez les Américains caucasiens que chez les Afro-Américains et les taux d'incidence sont généralement plus bas en Asie que dans les populations occidentales⁶⁴.

Même si le diagnostic de MII peut se produire tout au cours de la vie, celui de MC se produit le plus fréquemment de 15 à 30 ans avec un second pic d'incidence moins important entre 60 et 70 ans^{68; 69}. L'âge moyen de diagnostic de la MC est de 29.5 ans⁷⁰. Le pic d'incidence de diagnostic de la CU se produit entre 20 et 29 ans⁶⁹. Différentes études ont montré que les MII affectaient légèrement plus les femmes que les hommes. Cette prédominance se remarque plus particulièrement chez les femmes en fin d'adolescence et au début de l'âge adulte ce qui suggère que les hormones pourraient jouer un rôle dans l'expression de la maladie⁶⁴.

1.2.2 Manifestations cliniques

La MC a été identifiée pour la première fois par un chirurgien allemand en 1623, mais c'est le médecin américain Burril B. Crohn qui l'a décrite officiellement et lui a donné son nom en 1932². La CU a été décrite pour la première fois en 1859 par le médecin britannique Sir Samuel Wilks³. Dans les deux cas, il s'agit de maladies récidivantes. Les patients se présentent avec entre autres des diarrhées excessives, du sang, du pus et/ou du mucus dans les selles, des douleurs abdominales intenses et quelquefois de la fièvre⁷¹. Ces manifestations sont dues à de l'inflammation sévère du tube digestif. La MC et la CU diffèrent dans leurs manifestations cliniques (voir Tableau 1).

La maladie de Crohn est caractérisée par de l'inflammation qui peut s'étendre dans tout le tube digestif de la bouche à l'anus (Figure 5). L'inflammation est profonde et affecte souvent toute les couches de tissus qui composent l'intestin. La maladie se manifeste par des régions d'ulcération profonde qui peuvent occasionner des zones localisées d'amas de leucocytes (granulomes) ou connexions tubulaires entre les intestins et les organes avoisinants (fistules)⁷². Quelquefois la maladie produit des blocages de l'intestin dus à des formations cicatricielles consécutives à de l'inflammation (strictions). La MC se manifeste le plus souvent dans la région distale de l'intestin grêle et la région proximale du côlon. Près de 50% des patients ont des manifestations d'inflammation dans l'iléon et le côlon. Pour le tiers des patients, l'inflammation est confinée à l'intestin grêle, avec

l'iléon terminal le plus souvent affecté. Enfin, pour 20 à 25% des patients, le côlon seul est affecté par la maladie⁷². Dans la MC, l'inflammation est discontinue, c'est-à-dire que des zones d'inflammation sont entrecoupées de zones de tissus d'apparence saine. Cette nature discontinue occasionne des présentations cliniques très variables d'un patient à l'autre.



Figure 5. Photos de sections d'intestin grêle sain (gauche) et avec maladie de Crohn (droite)

La muqueuse intestinale atteinte par la MC est enflée, enflammée, et présente des ulcérations. Dans la MC, l'inflammation est souvent discontinue avec des zones d'apparence saine qui entrecoupent des zones enflammées.⁷³

Tableau 1. Comparaison de la MC et la CU

	Maladie de Crohn (MC)	Colite ulcéreuse (CU)
Incidence	3,1-14,6 par 100 000	2,2-14,3 par 100 000
Présentation clinique de la maladie	Tout le système digestif de la bouche à l'anus peut être atteint. Une majorité des cas sont atteints au niveau de l'intestin grêle distal et du côlon proximal; la moitié des patients ont l'iléon et le côlon atteints; le tiers est atteint à l'intestin grêle seulement (surtout l'iléon); de 20 à 25% des patients ont seulement le côlon d'atteint.	Se développe dans le rectum et progresse de façon proximale.
	Inflammation transmurale de la muqueuse et la sous-muqueuse.	Inflammation continue.
Distribution sociogéographique	Taux plus élevés dans les pays de l'hémisphère nord et dans les populations industrialisées plus riches et urbaines.	Taux plus élevés dans les pays de l'hémisphère nord et dans les populations industrialisées plus riches et urbaines.
Facteurs de risque environnemental	La cigarette expose à la MC.	La cigarette et l'appendicite sont des facteurs protecteurs.

Tableau modifié d'après ¹⁴

Pour la CU, l'inflammation est restreinte à la muqueuse et la sous-muqueuse et s'étend de façon proximale à partir de l'anus pour inclure le côlon en partie ou en totalité (Figure 6). Les fistules et les granulomes ne font pas partie des manifestations courantes de la maladie. De plus, l'inflammation est continue sans entrecouplement avec des tissus sains. Dans 5% des cas de maladies inflammatoires de l'intestin, il est impossible d'établir un diagnostic définitif de MC ou CU à cause des similitudes et du chevauchement entre les pathologies⁷⁴ (voir le tableau 1 pour une comparaison de la MC et la CU).

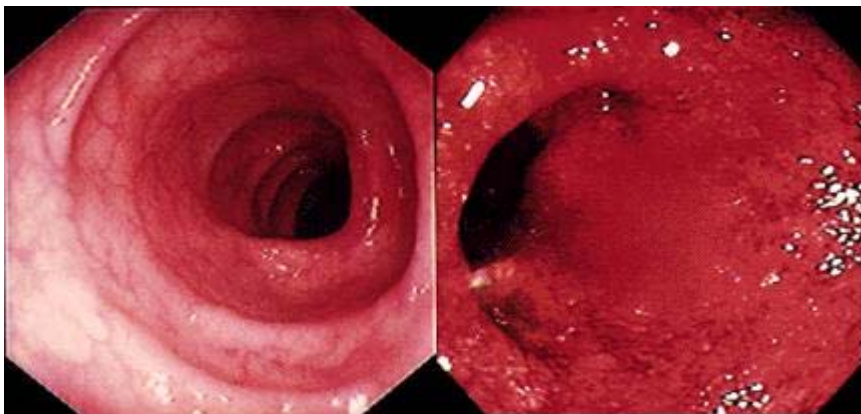


Figure 6. Photos de sections de côlon sain (gauche) et avec colite ulcéreuse (droite)

Le côlon présenté dans la photo de droite est en phase active, mais bénigne de colite ulcéreuse. La rougeur marquée est caractéristique de l'état d'inflammation (<http://www.virtualmedicalcentre.com/diseases.asp?did=173>).

Aux complications caractéristiques des MII dues par exemple à la cicatrisation s'ajoutent des complications extra-intestinales qui se produisent chez 25% des patients⁷¹. Elles peuvent affecter tous les systèmes, mais les plus communes visent la peau, les yeux, la bouche, les articulations et le foie (voir Figure 7). L'arthrite et l'érythème noueux sont, respectivement, des manifestations rhumatoïde et dermatologique courantes. L'anémie est aussi fréquente⁷⁵. De plus, les MII sont parmi les facteurs de risque les plus importants de développer un cancer intestinal. La prévalence de cancer colorectal (CCR) est de 0.3% dans la population générale alors qu'elle est de 3.7% chez les patients atteints de CU ce qui représente un risque 12 fois plus grand pour ces derniers^{76; 77}. Une récente méta-analyse constituée de 20 analyses indépendantes et plus de 40 000 patients a révélé que les risques de CCR et de cancer de l'intestin grêle étaient respectivement de 2 à 3 fois et de 18.75 fois plus élevé chez les patients souffrant de MC que pour la population générale⁷⁸. Les facteurs de risque pour les CCR reliés aux MII incluent l'étendue, la sévérité et la durée de la MII, les MII à début précoce, les manifestations extra-intestinales telles que la cholangite sclérosante primaire et l'histoire familiale de cancer. Chez un patient atteint d'une MII sévère depuis plus de 30 ans, les risques de développer un CCR sont de 18%⁷⁹. Les patients atteints de MC sont également plus à risque de développer d'autres cancers qui affectent la portion supérieure du système digestif, les poumons, la vessie et la peau. Les patients atteints CU sont plus à risque de développer des cancers du foie et de la vésicule biliaire et des leucémies⁸⁰.

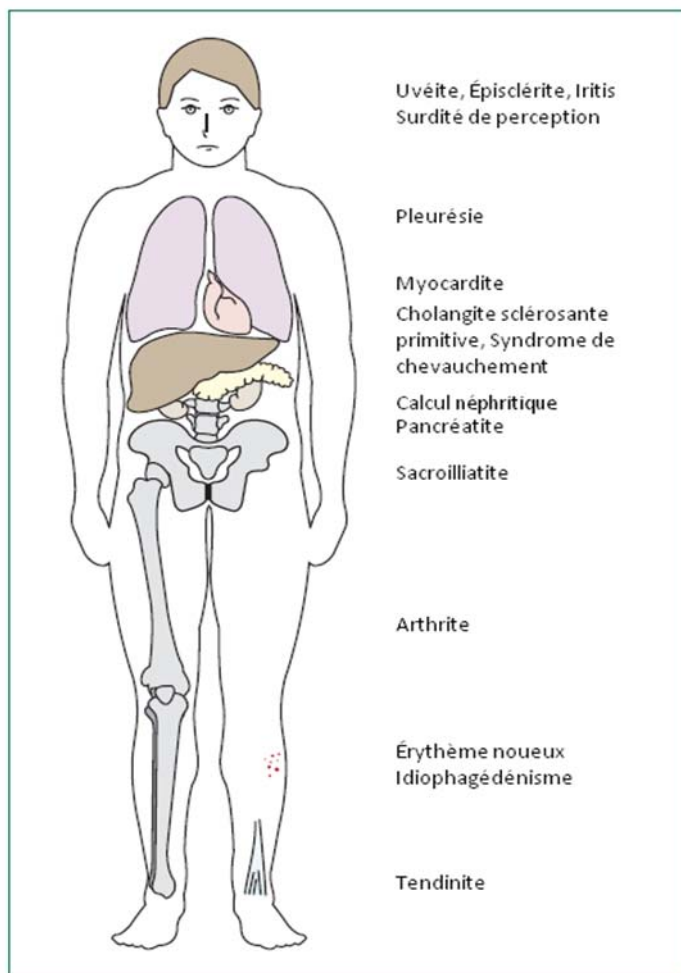


Figure 7. Manifestations extra-intestinales les plus communes affectant les patients souffrant de MII

25% des patients atteints de MII souffrent de manifestations extra-intestinales qui peuvent affecter tous les systèmes mais les plus communes impliquent la peau, les yeux, la bouche, les jointures et le foie. Figure adaptée de⁷¹

1.2.3 Traitement

La thérapie à long terme pour soigner les patients atteints de MII est généralement multidimensionnelle et dictée par le type de maladie ainsi que les sites atteints. Les médicaments pour gérer les symptômes comme les antidiarrhéiques sont souvent utilisés.

1.2.3.1 L'acide 5-aminosalicylique

Le premier médicament utilisé pour soulager l'inflammation est généralement un dérivé de l'acide 5-aminosalicylique (5-ASA) tel que le sulfalazine^{72; 81}. On suppose que ce composant agit en bloquant la production de prostaglandines et de leukotriènes ce qui en retour inhiberait la chimiotaxie des neutrophiles, désactiverait les dérivés réactifs de l'oxygène et inhiberait l'activation de NF-κB. Il existe une grande variété de médicaments de types 5-ASA qui doivent être sélectionnés en fonction de la localisation de la maladie. Par exemple, pour des patients souffrant d'inflammation colique distale, un suppositoire sera approprié, des comprimés à oraux à diffusion lente ou couplés à une matrice pH-dépendante sont également disponibles pour les gens souffrant d'inflammation à l'intestin grêle, etc. Le maintien de la rémission par traitement aux 5-ASA est possible pour la CU mais plus difficile pour la MC⁸².

1.2.3.2 Les corticostéroïdes

Les corticostéroïdes sont habituellement prescrits lorsque les 5-ASA n'ont pas d'effet. La prednisone entre autres est efficace à des doses de 40 à 60 mg/jours pour les patients atteints de MC ou de CU de niveau modéré à sévère⁷¹. Les corticoïdes sont surtout prescrits dans le cas de crise et leurs effets à long terme n'ont pas été démontrés. De plus, ils sont souvent accompagnés d'effets secondaires graves qui peuvent toucher de multiples systèmes (hypertension, hyperglycémie, ostéoporose, insomnie, glaucome,

acné, etc) et occasionnent une dépendance à long terme. Toutefois, ces effets peuvent être contrôlés par des suppléments alimentaires ou d'autres médicaments⁸¹.

1.2.3.3 Les immunomodulateurs

Pour maintenir la rémission, des cas modérés à sévères de MII, les médicaments immunosuppresseifs ou immunomodulateurs sont souvent utilisés. Les plus utilisés sont l'azathioprine et son métabolite actif la mercaptopurine. Les risques de la prise de ces médicaments incluent une augmentation des risques de lymphomes et les patients doivent être suivis pour vérifier leurs niveaux de leukocytes^{72; 81}. Certaines personnes sont porteuses d'allèles spécifiques du gène *TMPT* (*thiopurine s-methyl transferase*) qui les rendent susceptibles de développer des réactions de toxicité suite à l'administration de ce type de médicaments⁸². En effet, le gène *TMPT* code pour une enzyme qui est impliquée dans le métabolisme de ces drogues et qui a comme fonction d'en éliminer les dérivés toxiques. Les patients devraient donc idéalement être génotypés avant l'administration de ces médicaments. Les mécanismes d'action responsables des effets thérapeutiques de ce type de médicaments observés sur les patients atteints de maladies inflammatoires demeurent obscurs, mais ils pourraient inclure la suppression d'un sous-groupe de lymphocytes T⁸².

1.2.3.4 Les antibiotiques

Des études cliniques ont montré l'efficacité des antibiotiques pour traiter un sous-groupe de patients atteints de MC mais son effet n'a pas été démontré pour la CU. Le metronidazole par exemple, est efficace pour les patients ayant des fistules péri-anales⁸³. L'effet des antibiotiques sur la maladie de Crohn témoigne de l'importance de la flore intestinale dans le développement de la maladie.

1.2.3.5 Les traitements biologiques

Les traitements biologiques représentent la nouvelle vague de traitement contre les MII. Un anticorps monoclonal anti-TNF α , l'infliximab, commercialisé sous le nom de Remicade, est utilisé en clinique depuis 1997⁸⁴. Son efficacité a été démontrée chez les patients atteints de CU et de MC et il est maintenant utilisé en thérapeutique d'induction et d'entretien. Cet anticorps se lie au TNF α -une cytokine pro-inflammatoire sécrétée par les macrophages activés - soluble et lié à la membrane cellulaire. Il favorise l'élimination de TNF α soluble et membranaire et la lyse des macrophages et des lymphocytes T activés et induit l'apoptose de ces derniers. Ce faisant, il contrôlerait la réponse immunitaire à la muqueuse intestinale⁸⁵. Des études cliniques ont démontré l'efficacité de l'infliximab pour induire et maintenir la rémission de l'inflammation intestinale, prévenir les fistules, prévenir l'utilisation de stéroïdes, réduire les besoins d'hospitalisation et de chirurgies et améliorer la qualité de vie^{84; 86-88}. L'infliximab est approuvé pour le traitement de la MC et la CU qui ne répondent pas aux thérapies conventionnelles mentionnées précédemment⁸⁵. Il s'agit d'un traitement efficace, mais coûteux et il peut être dangereux, certains cas de mortalité due à la tuberculose ayant été enregistrés⁸⁹.

D'autres anticorps ciblant TNF α et également d'autres cytokines font partie d'essais en phase clinique. Le Fontoluzimab, un anticorps humanisé anti-interféron γ a été étudié en phase II et semble avoir fait diminuer les taux détectés du biomarqueur inflammatoire, protéine C-réactive, chez les patients atteints de MC modérée à sévère⁹⁰. Un anticorps dirigé contre la sous-unité p40 des interleukines 12 et 23 (IL12Bp40) semblent également avoir des effets positifs⁹¹. Certains nouveaux traitements sont axés sur le blocage des lymphocytes T. Un anticorps dirigé contre le corécepteur des lymphocytes T, CD3, a montré des effets chez des patients atteints de CU sévère et réfractaire aux corticostéroïdes dans une étude en phase I⁹². Une étude de phase II a cependant identifié un taux élevé de toxicité et d'effets secondaires⁹³. Deux autres anticorps le basiliximab et le daclizumab dirigés contre la chaîne α du récepteur à l'interleukine 2 (CD25) et

approuvés comme médicaments antirejet lors de greffes ont généré des résultats préliminaires mitigés dans le traitement des MII^{94; 95}. D'autres traitements biologiques ciblent l'inhibition de la migration et de l'adhésion cellulaire. Des inhibiteurs de molécules d'adhésion comme les intégrines bloquent la migration des leucocytes vers les sites d'inflammation. Deux études en phase II suggèrent que le natalizumab un anticorps monoclonal contre l'intégrine $\alpha 4$ induit la rémission des patients souffrant de MC active^{96; 97}.

1.2.3.6 La chirurgie

Dans certains cas de maladies réfractaires à la médication, la chirurgie est indiquée. En effet, l'ablation du côlon et du rectum, même si elle n'est pas souhaitable, guérit la CU. Toutefois, moins de 10% des patients ont recours à ce type d'opération⁸¹. Pour la maladie de Crohn, la chirurgie fait partie intégrante du traitement, à la fois pour contrôler les symptômes et pour traiter les complications. Par exemple, la chirurgie est indiquée dans les cas de perforation de l'intestin grêle ou du côlon, mais aussi dans les cas d'inflammation sévère réfractaire aux autres traitements, d'hémorragie sévère et de néoplasies⁹⁸. Dans les 20 ans suivant le diagnostic de MC, trois patients sur quatre auront eu recours à la chirurgie. Le taux de chirurgie dans les trois ans suivant le diagnostic oscille entre 25% et 45% dépendamment des pays⁷². La chirurgie chez les patients atteints de MII a comme ligne directrice de conserver au maximum les fonctions intestinales et sexuelles. Mais très souvent, la chirurgie a des conséquences fonctionnelles importantes comme la déficience en vitamine B12 ou les diarrhées d'acides biliaires et également, dans le cas de la maladie de Crohn, un taux élevé de rechute⁹⁹. Plusieurs types de chirurgies peuvent être envisagés selon le type et l'étendue de l'inflammation^{72; 81}.

Traditionnellement, les traitements ciblaient les symptômes et non les causes des MII. Récemment, une meilleure compréhension des causes immunobiologiques (section 1.2.4) a permis le développement des médicaments biologiques. À moyen terme,

l'accroissement des connaissances sur la susceptibilité génétique (chapitre 2) pourra sans doute contribuer à des traitements personnalisés et donc plus efficaces.

1.2.4 Études des causes immunobiologiques

Dans le système digestif en santé, les mécanismes de régulation anti-inflammatoires font contrepoids aux signaux inflammatoires pour maintenir l'équilibre immunologique. Cet équilibre peut être brisé pour différentes raisons notamment lorsque l'intégrité ou la perméabilité de la barrière intestinale est compromise ou lorsque les réponses immunitaires innées ou spécifiques sont disproportionnées. Les modèles murins sont très utiles pour étudier le développement de l'inflammation intestinale typique des MII. Même s'il n'existe pas de modèle qui présente exactement les mêmes manifestations physiologiques dont souffrent les patients atteints de MII (qui peuvent se présenter de façon très différente chez deux patients différents), les nombreux modèles murins permettent d'étudier différents aspects de l'inflammation.

La **perméabilité de la barrière intestinale** a été proposée comme facteur étiologique puisque qu'elle est compromise chez les patients atteints de MII. De plus, leurs parents sains du 1er degré montrent également des signes anormaux de perméabilité intestinale^{100; 101}. Le modèle de souris dextran sulfate sodium (DSS) permet d'étudier la perméabilité cellulaire. Il est généré en ajoutant le polymère DSS dans l'eau de la souris pendant quelques jours. Le résultat est une colite caractérisée par des diarrhées, du sang dans les selles, de l'ulcération et des infiltrations de granulocytes^{102; 103}. Comme la colite se produit quand même chez les animaux souffrant d'immunodéficience sévère combinée (*Severe combined immunodeficiency* ou SCID) ou qui sont déficients par délétion (knockout) pour les gènes *recombinant activating gene 1 or 2 (rag1, rag2)*, en d'autres termes chez les animaux ne possédant pas de lymphocytes T ou B, la réponse immunitaire spécifique ne joue pas un rôle important dans ce modèle¹⁰⁴. Deux études récentes se sont penchées sur le rôle des récepteurs TLR4 dans l'intégrité épithéliale intestinale^{105; 106}. Ces deux études faites chez des souris knockout pour les gènes *TLR4*

ou *myeloid differentiation primary response gene 88 (MYD88)* – encodant une protéine impliquée dans la signalisation via les TLRs - suggèrent que la signalisation via les TLRs est importante pour le maintien de la barrière épithéliale. Dans les deux études, les souris TLR4 *-/-* ont moins souffert de l'inflammation que les souris de type sauvage (TS). Étrangement les deux groupes ont noté une plus grande perte de poids, de saignements intestinaux et un plus haut taux de mortalité chez les souris TLR4 *-/-* soumises au DSS que les souris TS soumises au DSS. Cette susceptibilité au DSS a été attribuée à des problèmes de prolifération épithéliale et une augmentation de la translocation des bactéries commensales aux nodules lymphatiques¹⁰⁵.

Les macrophages et les CDT sont des acteurs importants de la **réponse immunitaire innée**. Dans les tissus enflammés typiques des MIIs, les macrophages sont activés et produisent différentes cytokines pro-inflammatoires telles que IL-1 α , IL-1 β et TNF α ¹⁰⁷; ¹⁰⁸. Les patients atteints de MC présentent un plus grand nombre de macrophages dans la muqueuse intestinale que les témoins en santé et ces macrophages sécrètent davantage des cytokines IL-23 et TNF α et contribuent à la production de INF γ par les lymphocytes T¹⁰⁹. Les CDT intestinales expriment davantage les récepteurs TLR2 et TLR4 chez les patients atteints de MII que les témoins en santé et chez les patients atteint de MC les CDT du côlon produisent plus de cytokines IL-12 et IL-6¹¹⁰. Les modèles murins permettent d'expliquer de façon moléculaire, les causes de ces différences dans l'expression de cytokines. Par exemple, les modèles de souris knockout pour le gène *NOD2* présentent des anomalies dans la production de cytokines¹¹¹. De plus, ces modèles murins montrent entre autre que les polymorphismes de *NOD2* associés à la MC (section 2.3.1.1) causent une perte des effets régulateurs médiés par l'activation et la signalisation de *NOD2*¹¹².

Au niveau de la **réponse immunitaire spécifique**, de nombreuses altérations immunobiologiques sont notables chez les patients atteints de MII. Les lymphocytes B sécrètent davantage d'IgA¹¹³. Les patients atteints de CU sécrètent davantage d'IgG1 et

les patients atteints de MC sécrètent davantage d'IgG toutes classes confondues¹¹⁴. Les sous-types de lymphocytes T prédominants présentent également des différences. Depuis l'identification des sous-types de lymphocytes T CD4+ T_H1 et T_H2 qui se caractérisent par des profils d'expression de cytokines distincts, chez la souris⁵² puis chez l'humain¹¹⁵, l'étude des lymphocytes T CD4+ a considérablement progressé. On a identifié d'autres sous-types, dont les Tregs et les T_H17^{55; 56; 116}. Plusieurs observations témoignent de l'importance des sous-types T_H1 et T_H17 dans l'étiologie de la MC. Par exemple, les taux élevés de production des cytokines INF- γ et IL-12 par les cellules mononuclées de la muqueuse intestinale^{117; 118}. De plus, la MC est associée à des taux élevés de IL-17 et d'INF γ par les lymphocytes T_H17 et aussi d'IL-21 qui régule la production de IL-17^{119; 120; 121}. La CU est considérée comme une maladie de type T_H2 atypique puisqu'on observe un sous-ensemble des cytokines T_H2 classiques. On observe notamment des taux élevés de production d'IL-5 et IL-13 par les lymphocytes T_H et d'IL-13 par les lymphocytes NK dans les tissus enflammés de patients atteints de CU¹²². Cependant, les lymphocytes T_H17 jouent également un rôle dans la pathologie¹²⁰. Les constats sont moins clairs pour les Tregs mais certaines observations suggèrent que le nombre et la fonction des Tregs pourraient être modifiés chez les patients souffrant de MII ce qui contribuerait au maintien de l'inflammation^{123; 124}. De plus, les modèles murins montrent que les Tregs sont essentiels pour bloquer la colite induite expérimentalement¹²⁵ et que les cytokines anti-inflammatoires TGF β et IL-10 produites par les Tregs semblent être responsables de ce blocage¹²⁶.

Tous ces bouleversements cellulaires s'ils se produisent dans le tube digestif de façon chronique provoquent des manifestations anatomiques qui conduisent aux maladies inflammatoires de l'intestin.

1.2.5 Facteurs de risque environnemental

L'hypothèse la plus communément acceptée sur la cause des MII suggère que la combinaison de facteurs environnementaux et de prédispositions génétiques (chapitre 2) provoque une dérégulation du système immunitaire gastro-intestinal face au microbiome bactérien. À titre d'exemple pour appuyer l'importance de la génétique dans le développement de la maladie, les études de jumeaux monozygotes sont souvent citées. La plupart présente des taux de concordance d'environ 50-60% pour la MC^{127; 128}. Ce qui veut dire que 40 à 50% des gens ayant une constitution génétique identique sont discordants en ce qui a trait à la MC. Les influences environnementales jouent donc un rôle tout aussi important que la génétique dans le développement des MII. On les classe donc dans les maladies complexes puisque l'interaction de plusieurs facteurs environnementaux et génétiques contribue à leur cause. Le principal facteur environnemental qui influence l'inflammation et la réponse immunitaire est le microbiome bactérien (chapitre 1). Classiquement, deux autres facteurs ont été étudiés abondamment: la cigarette et l'appendicectomie.

1.2.5.1 La cigarette

Étrangement, la cigarette est considérée comme un facteur protecteur pour la CU et facteur de risque pour la MC. En effet, plusieurs études ont confirmé cette affirmation dont une métaanalyse publiée en 1989 par Calkins qui conclue que les fumeurs ont considérablement moins de chance de développer une CU que les gens qui n'ont jamais fumé, c'est-à-dire que le risque relatif approché (*odds ratios*, OR) est de 0,4 (95% Intervalle de confiance (IC) 0,34-0,48)¹²⁹. Mais cette « protection » n'est conférée qu'aux fumeurs actifs, un calcul du risque des anciens fumeurs donne un OR de 1,64 (95% IC 1,35-1,98). La cigarette semble influencer la présentation clinique de la maladie. Dans une autre étude, on note que les fumeurs ont moitié moins de chance d'être hospitalisés que les non-fumeurs¹³⁰. Les timbres de nicotine semblent avoir quelques effets bénéfiques sur le taux de rémission de la maladie¹³¹. Par contre,

plusieurs études ont montré que la cigarette est un facteur de risque pour la MC. La métaanalyse de Calkins suggère que les fumeurs sont plus à risque pour cette maladie avec un OR de 2.0 (95% IC 1,65-2,47)¹²⁹. Les anciens fumeurs, avec un OR de 1,80 (95% IC 1,33-2,51) seraient aussi à risque, mais moins que les fumeurs actifs donc arrêter de fumer serait bénéfique pour les patients souffrant de MC. Entre autres, la cigarette aggrave les manifestations cliniques de la MC, en favorisant la formation de fistules ou de strictions¹³².

On ignore les mécanismes exacts régissant l'interaction entre la cigarette et le développement des MII. Certaines observations permettent cependant d'avancer des hypothèses. Par exemple, les macrophages exposés au monoxyde de carbone sécrètent moins de la cytokine pro-inflammatoire TNF- α et plus de la cytokine anti-inflammatoire d'IL-10 sous stimulation LPS que les macrophages non exposés¹³³. La cigarette a des effets délétères sur les phagocytes qui passent par la diminution de leur pouvoir bactéricide et bactériostatique¹³⁴. De plus, la cigarette et la nicotine diminuent les taux de cytokines dans la muqueuse du côlon des fumeurs souffrant de MII, spécifiquement de IL-1 β et d'IL-8 pour la CU et de IL-8 pour la MC¹³⁵. D'autres observations suggèrent qu'une augmentation de la production de mucus chez les fumeurs serait bénéfique¹³⁶. Les effets néfastes seraient entre autres modulés par les dérivés réactifs de l'oxygène (DRO). Les fumeurs génèrent plus de DRO et leur pouvoir antioxydant est plus faible¹³⁷. De plus, on a observé que l'exposition à la fumée de cigarette augmente la production de DRO et favorise l'infiltration des neutrophiles dans le côlon de rat atteint de colite induite chimiquement¹³⁸. Toutes ces observations n'expliquent pas les différences observées entre les effets de la cigarette sur la MC et la CU cependant les différences de susceptibilité génétique pourraient être en cause. Par exemple, le gène de susceptibilité *NOD2* est spécifique à la MC (voir chapitre suivant)¹³⁹. Conséquemment, certaines voies biologiques sont spécifiquement impliquées dans l'une ou l'autre de ces maladies et un effet de synergie de ces voies avec la cigarette pourrait expliquer les différences.

1.2.5.2 L'appendicectomie

Plusieurs études ont associé l'appendicectomie à des risques réduits de développer la CU. Une métaanalyse de 17 études populationnelles réalisées dans les années 90 impliquant 3600 cas et plus de 4600 contrôles a montré que l'appendicectomie était associée à un OR de 0,312 (95% IC 0,261-0,373) des risques subséquents de CU¹⁴⁰. Une étude suédoise publiée en 2001 a également observé un effet protecteur à la CU d'une appendicectomie suivant une appendicite avec perforation (OR 0,58 95% IC 0,38-0,87), d'une appendicite sans perforation (OR 0,76 95% IC 0,65-0,90) et d'une adénopathie mésentérique (OR 0,57 95% IC 0,36-0,89)¹⁴¹. De plus, la protection contre la CU a été observée uniquement chez les patients ayant subi leur appendicectomie avant l'âge de 20 ans. Par ailleurs, cette étude n'a pas identifié de lien entre l'appendicectomie liée à des douleurs abdominales non spécifiques et la CU (OR 1,06 95% IC 0,74-1,52). Cette observation a poussé ce même groupe à investiguer davantage et dans une étude subséquente, il conclut que l'appendicectomie en elle-même ne protège pas contre la CU mais l'inflammation de l'appendice sous-jacente est un facteur protecteur¹⁴².

Le groupe suggère plusieurs théories pour expliquer le mécanisme derrière cette protection. Entre autres, la théorie de l'immunité constitutionnelle selon laquelle les individus développeraient une réponse immunitaire distincte, souvent simplifiée par la dichotomie T_H1/T_H2, suite à des facteurs génétiques et environnementaux. Le profil de réponse immunologique des gens ayant eu une appendicite dans l'enfance serait de type T_H1 alors que la CU est de type T_H2. Cette différence préviendrait le développement de la CU. Pour appuyer cette hypothèse, une étude a montré que les patients souffrant de CU ayant eu une appendicectomie (sans doute suite à une appendicite) présentent généralement des symptômes plus bénins que les patients n'ayant pas eu d'appendicectomie¹⁴³. Il ne semble pas y avoir de lien clair entre l'appendicectomie et la MC. Une méta-analyse a présenté l'appendicectomie comme un facteur de risque pour la MC mais la majeure partie du risque a été observée dans la première année et après 5 ans le risque n'était plus significatif¹⁴⁴. Les auteurs suggèrent que ces observations

traduisent des problèmes de diagnostic, c'est-à-dire que les patients souffrant de MC non-diagnostiquée subissent une appendicectomie injustifiée. En somme, les différences observées entre les effets des facteurs de risque environnemental comme la cigarette et l'appendicite sur le développement de la MC et la CU sont multifactorielles et semblent reliées au contexte génétique.

1.2.5.3 Autres facteurs

D'autres facteurs de risque environnemental qui contribueraient à provoquer des réactions immunitaires dérégulées typiques des MII ont été proposés quoique leurs liens directs avec les MII ont été moins systématiquement confirmés. Une bonne proportion de ces facteurs de risque tourne autour de l'hypothèse de l'hygiène avancée par Strachan¹⁴⁵. L'hypothèse s'appuie sur le fait qu'au cours du 20^e siècle, les pays industrialisés ont bénéficié d'améliorations considérables dans les niveaux d'hygiène due entre autre à l'accès à de l'eau potable, à de l'eau chaude, à la réfrigération des aliments ainsi que la réduction de la taille des familles. Parallèlement, les incidences d'allergies et de maladies inflammatoires et auto-immunes ont augmenté, par exemple entre 1968 et 1983, l'incidence de MC chez les enfants écossais a triplé¹⁴⁶. Selon l'hypothèse de l'hygiène un enfant du modernisme pourrait être surprotégé contre les infections communes. Lorsqu'il entre en contact avec une infection plus tard durant sa vie, des réactions inflammatoires disproportionnées se déclenchent menant au développement de maladies comme les MII¹⁴⁷. Des facteurs spécifiques, indicateurs du niveau d'hygiène, tels que les infections au *Helicobacter pylori*, la taille de la fratrie, l'ordre de naissance dans la fratrie, expositions aux animaux et grandir en ville ou en campagne ont été considérés mais les études ont donné des résultats mitigés¹⁴⁷⁻¹⁴⁹. Toutefois, la localisation géographique pourrait influencer le développement des MII puisque les études épidémiologiques montrent des disparités dans la distribution de la prévalence des MIIs en fonction des pays.

Chapitre 2

Les maladies inflammatoires de l'intestin: des maladies génétiques

2.1 Une histoire familiale

Plusieurs observations permettent d'avancer que les MII sont au moins en partie des maladies génétiques. En effet, plusieurs études montrent qu'un individu souffrant de MII est plus susceptible d'avoir un membre de sa famille qui en souffre aussi. Des études populationnelles concluent que de 5 à 10% des patients ont un membre de leur famille rapprochée qui souffre aussi de MII¹⁵⁰. Le risque relatif à la fratrie est estimé à 30 à 40 fois pour la MC et 10 à 20 fois pour la CU. Les études de jumeaux supportent la théorie d'une contribution génétique, les taux de concordance étant plus élevés pour les monozygotes que pour les dizygotes surtout pour la MC (27-58% vs 0-12%) mais aussi pour la CU (6-15% vs 0-6%)^{150; 151}. De plus, il a été observé que l'âge de diagnostic et le phénotype de la maladie sont aussi héréditaires dans la MC¹⁵². Finalement, des études ont montré que les membres de la famille rapprochée de patients souffrant de MC ou CU sont plus à risque de développer la CU ou la MC, respectivement, c'est-à-dire que certains facteurs de risques génétiques sont communs à ces deux maladies^{150; 153; 154}. Mais les MII sont des maladies complexes, d'un point de vue génétique. Une architecture complexe de gènes, d'interactions géniques, d'environnement, d'interactions gènes-environnement est à l'origine de la pleine manifestation du phénotype de la maladie¹⁵⁵. De plus, chaque gène n'a souvent qu'un effet modeste sur le risque de développer la maladie.

2.2 Structure du génome et variations

2.2.1 Contexte qui a favorisé les découvertes

Durant la dernière décennie, le domaine de la génétique humaine a été le théâtre de plusieurs découvertes et avancées révolutionnaires. La plus grande percée a sans doute été réalisée avec la complétion de la séquence du génome humain^{156; 157}. Plusieurs autres développements ont contribué à augmenter les connaissances sur les maladies complexes: l'identification d'un grand nombre de marqueurs génétiques ponctuels (entre autres des *Single Nucleotide Polymorphisms* (SNP) et des variantes en nombre de copies (*copy number variants* ou CNV)); des avancées dans la compréhension de la corrélation et du déséquilibre de liaison entre les SNP; le développement de plusieurs méthodes de génotypage à haut débit; le séquençage du génome d'autres espèces comme la souris et l'émergence de nouvelles disciplines telle que la bio-informatique. Ces nouveaux outils et techniques d'analyse ont ouvert la porte à des études d'association à grande échelle utilisant des marqueurs génétiques pour examiner des gènes ou des régions du génome dans le but de trouver des variantes responsables de la susceptibilité aux maladies ou tout autre trait complexe.

Les prémisses du Projet Génome Humain (PGH) ont vu le jour au milieu des années 80, quand quelques visionnaires ont décidé de briser les rangs et l'idéologie généralement acceptée que les recherches en biologie doivent toujours être menées par une hypothèse¹⁵⁸. La justification majeure du PGH était que d'acquérir de l'information fondamentale sur notre matériel génétique allait augmenter la compréhension de la génétique humaine et des rôles des gènes dans la santé et la maladie. De plus, dès les débuts, un des principes du PGH était que les données et ressources devaient être disponibles rapidement à toute la communauté scientifique avant même publication¹⁵⁸. Les bénéfices de ce projet sont donc énormes. Ils incluent l'accès public à la structure et la séquence de chaque gène humain qui permet d'étudier rapidement et relativement

facilement la fonction et la régulation d'un gène. Ce type d'information facilite énormément l'étude des effets de variantes de séquence.

2.2.2 Variations et marqueurs génétiques

Le génome humain comprend environ 3,3 milliards de paires de bases d'ADN^{156; 157}. On a d'abord surestimé le nombre total de gènes du génome humain, mais les études phylogéniques récentes comptabilisent environ 20 500 gènes codant pour des protéines¹⁵⁹. De nombreuses études ont montré qu'il existe plusieurs variations ou différences pouvant être identifiées par la comparaison du génome de deux individus, on les appelle des variantes génétiques. Ces variantes expliquent les différences phénotypiques entre les individus, que ce soit au niveau de l'apparence, de la susceptibilité aux maladies ou de la réponse thérapeutique différentielle. Il existe plusieurs types de variantes génétiques.

2.2.2.1 Les répétitions en tandem

Les répétitions en tandem (RT) et les SNP, entre autres, ont été abondamment utilisés comme marqueurs dans les études génétiques. Les RT sont définies par des blocs de séquences répétées. Ces répétitions peuvent être localisées sur un ou plusieurs chromosomes différents. On leur donne des noms spécifiques en fonction de la taille moyenne de leur bloc répété, des plus petites répétitions (les microsatellites) aux répétitions plus longues les CNV. Les répétitions composent environ 50% du génome^{156; 157}. Les définitions varient, mais on nomme habituellement les répétitions d'unités composées de 9 nucléotides ou moins, des **microsatellites**, et les répétitions d'unités de 10 nucléotides ou plus, des **minisatellites**¹⁶⁰. Les RT sont instables, leur taux de mutation variant souvent entre 10^{-3} et 10^{-7} par division cellulaire est de 10 à 100 000 fois plus élevé que dans le reste du génome¹⁶¹. La plupart de ces mutations causent des polymorphismes de répétition, c'est-à-dire que des unités complètes de répétition sont ajoutées ou retranchées. Une bonne partie des RT est située dans les gènes ou les régions

régulatrices. En effet, environ 17% des gènes contiennent des répétitions dans leur séquence¹⁶². Les répétitions de 3 et de 6 nucléotides sont de loin les plus fréquentes, possiblement parce qu'il y a de la sélection contre les variations qui changent le cadre de lecture¹⁶³. Les RT ne se trouvent pas dans tous les types de gènes, on les trouve davantage dans les gènes qui encodent des protéines régulatrices comme les facteurs de transcription¹⁶⁴. Les polymorphismes de RT peuvent être liés à des variations de fonction qui provoquent des maladies dont les plus connues sont le syndrome de l'X fragile¹⁶⁵, la maladie de Kennedy¹⁶⁶ et la maladie de Huntington¹⁶⁷. Typiquement, la sévérité de la maladie est plus importante et l'apparition des symptômes est plus hâtive chez les patients portant un plus grand nombre de RT. Il existe beaucoup de variations dans la longueur des RT (c'est-à-dire beaucoup d'allèles) chez les non affectés, mais ces dernières ne dépassent pas le seuil pathogénique. Comme plusieurs allèles sont souvent présents à un même locus, les génotypes familiaux sont souvent très informatifs. En conséquence, les microsatellites ont beaucoup été utilisés dans les études de génétique des populations et de cartographie génétique par approche de liaison (section 2.3.1).

2.2.2.2 Les SNP

La forme la plus commune de variantes génétiques est le polymorphisme nucléotidique d'une seule base (SNP). À l'heure actuelle, il y a environ 20 000 000 de SNP entrés dans la base de données dbSNP (Build 132, rs # validés, <http://www.ncbi.nlm.nih.gov/projects/SNP/>). Dans deux génomes humains tirés au hasard, 99,9% de la séquence d'ADN est identique. Les 0,1% restants contiennent des variations de séquence en majorité des SNP¹⁶⁸. La plupart des SNP sont communs et bialléliques. Deux SNP sur trois impliquent le remplacement d'une cytosine par une thymine. Les SNP sont stables, très abondants et distribués dans tout le génome. On estime qu'un SNP est présent toutes les mille paires de base environ¹⁶⁹.

L'étude systématique sur des SNP communs a permis de découvrir que les individus qui portent l'allèle d'un SNP à un site précis portent souvent des allèles spécifiques à des sites voisins. C'est-à-dire que les variations génétiques qui sont proches l'une de l'autre tendent à être héritées en bloc haplotypique, de dix à quelques centaines de kilobases, en déséquilibre de liaison (DL) (Figure 8). Une fois que ces blocs sont identifiés, ils peuvent être traités comme des allèles^{170; 171}. Un ou plusieurs allèles sur le bloc prédisent l'identification des autres allèles. Ce DL existe à cause de l'ascendance partagée des chromosomes contemporains. Quand une nouvelle variante causale apparaît à la suite d'une mutation – que ce soit un SNP, une insertion/délétion, ou autre – cette variation est liée au chromosome unique où elle est apparue et elle est marquée par une combinaison distincte de variantes¹⁷². Cette constatation a servi de base à la création du projet HapMap mené par un consortium international qui en Octobre 2002¹⁷³ s'est donné pour objectif de déterminer les patrons communs de variations dans la séquence du génome humain et de mettre ces informations à la disponibilité de tous. Le consortium voulait développer une carte de ces patrons génétiques à travers le génotypage d'un million de SNP et la détermination de leur fréquence et du degré de corrélation entre eux dans des échantillons d'ADN de populations d'ascendance africaine, asiatique et européenne¹⁷³. La base de données HapMap compte aujourd'hui plus de 4 millions de SNP dans la population CEPH d'ascendance européenne (*HapMap data release HapMap data release #28, August 2010, on NCBI B36 assembly, dbSNP b126*). Concrètement, cette base de données permet de sélectionner un nombre restreint de SNP pour l'étude de la variation sur un chromosome entier voire sur tout le génome. Cette innovation permet de réduire les coûts de génotypage et donc de faciliter les projets à grande échelle. La multiplication des études d'association pangénomiques depuis cinq ans est le témoignage de la réussite du projet HapMap.

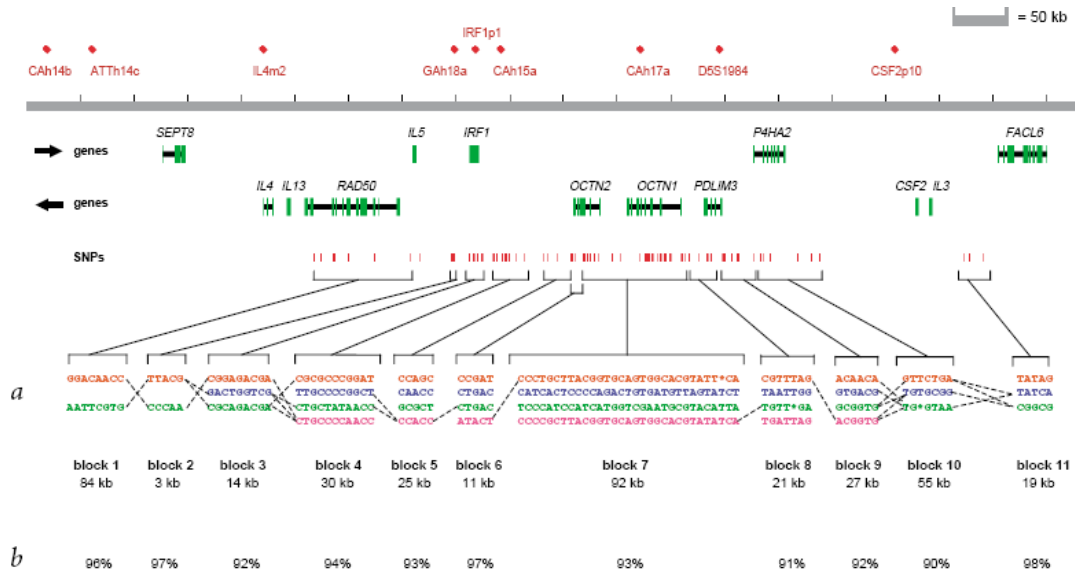


Figure 8. Déséquilibre de liaison

Cette figure est une représentation de la distribution des variantes de séquence de la région chromosomique 5q31 (IBD5) dans une cohorte d'ascendance européenne. Ces variantes sont arrangées en blocs haplotypiques communs plus ou moins longs (séquences de couleur) qui sont transmis tel quel et caractérisés par une faible variabilité génétique (a). Ce phénomène d'association préférentielle des variantes est appelé déséquilibre de liaison. Des épisodes de recombinaisons séparent les blocs entre eux. Les lignes pointillées représentent les associations les plus communes entre les blocs. Enfin en b, les pourcentages représentent la proportion des chromosomes observés qui correspondent aux haplotypes communs représentés ici¹⁷⁰.

Selon la nature et la localisation des variantes génétiques, elles peuvent être silencieuses ou influencer la voie biologique dans laquelle la protéine encodée par le gène joue un rôle. Par exemple, certaines variations génétiques mènent à des différences dans la structure de la protéine et modifient sa fonction, alors que d'autres variantes peuvent modifier le niveau auquel le gène ou la protéine est exprimé. Quelques unes de ces variations qui ont des conséquences fonctionnelles peuvent influencer les phénotypes normaux (ex. taille, couleur des yeux, etc.) alors que d'autres peuvent influencer des phénotypes cliniques et donc être considérés comme facteurs de risque génétique. Ce sont ces derniers qui sont les sujets d'études génétiques visant à identifier les gènes de susceptibilité à des maladies.

2.3 Stratégie d'identification de facteurs de risque génétique sur les MII

2.3.1 Études de liaison: les premières découvertes génétiques des MII

L'identification du gène et de la variante causale d'une maladie est un défi important. Dans le cas des maladies monogéniques, il y a un locus à identifier. Pour les maladies complexes, comme les MII, il y en a plusieurs et la plupart des gènes ont un effet modeste sur le phénotype. La cartographie génétique est le terme utilisé pour décrire la localisation de gènes impliqués dans des phénotypes sur la base de la corrélation avec des variantes génétiques sans qu'il n'y ait d'a priori sur leurs fonctions¹⁶⁸. Deux stratégies de cartographie génétique ont été particulièrement utilisées : les études de liaison et les études d'association (Figure 9).

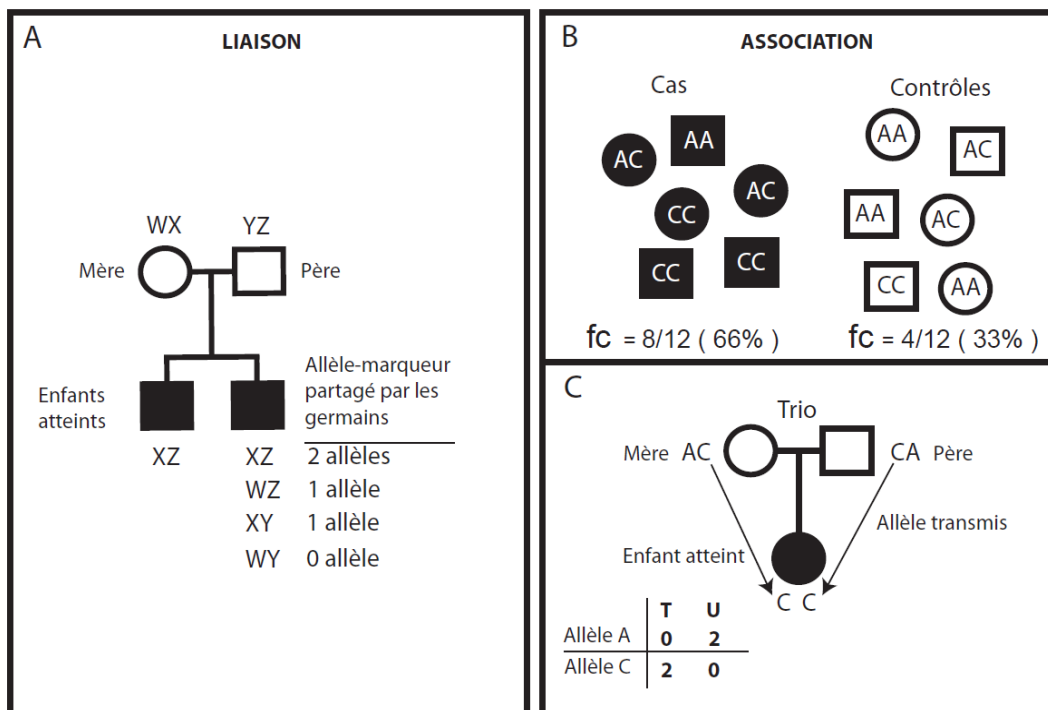


Figure 9. Représentation classique d'étude de liaison et d'association

Les études de liaison ont pour but de localiser les régions chromosomiques qui contiennent des gènes de susceptibilité en employant des tests statistiques indirects qui examinent la cotransmission des marqueurs génétiques et de la maladie (ou du trait complexe étudié) au sein de familles. Les études de liaison sont efficaces pour étudier les facteurs génétiques hautement pénétrants qui sont rares dans une population. Typiquement, plusieurs familles sont incluses dans ces études et le partage d'allèles entre les membres d'une famille qui sont atteints par la maladie est analysé. Par exemple, en **A**, il y a deux parents non atteints et deux enfants atteints. On s'attend à ce que les enfants atteints partagent 0, 1 ou 2 allèles dans une proportion de 0,25/0,50/0,25 respectivement. Si les deux enfants d'une même famille partagent des allèles d'un marqueur spécifique plus souvent que dicté par le hasard, il est possible que le marqueur soit lié à la maladie. Même si les allèles partagés entre deux enfants d'une même famille doivent être les mêmes, les allèles partagés entre les familles peuvent être différents et toujours être liés à la maladie. D'autres parts, les études d'association testent des allèles spécifiques à travers tous les individus d'une population étudiée. Il existe deux types communs de designs expérimentaux pour les études d'association : les études populationnelles de type cas/témoins (ou contrôles) (**B**) et les trios qui comprennent deux parents et leur enfant atteint (**C**). Les études d'association de cas/témoins testent si un allèle est corrélé à une maladie en déterminant si les fréquences alléliques diffèrent entre les cohortes de cas et de témoins non-apparentés. Des centaines voire des milliers d'échantillons sont testés dans les études cas/témoins. À des fins de simplification, on illustre ici (**B**) une différence évidente entre les fréquences de l'allèle C dans une étude de 6 cas (66%) et 6 témoins (33%) ce qui « suggère » une association de cet allèle avec le phénotype de la maladie. Dans les trios, l'allèle transmis (T) est l'équivalent du cas et l'allèle non transmis (U pour *untransmitted*), du témoin. Le principe est démontré (**C**) en utilisant un seul trio, mais typiquement des centaines de trios sont testés. Dans cet exemple, l'allèle C est transmis deux fois (100%) à l'enfant atteint « suggérant » une association. Les études de cas/témoins sont plus efficaces que les études basées sur des familles parce qu'elles requièrent moins d'échantillons pour procurer la même information (1 cas et 1 contrôle au lieu des 3 membres du trio). Cependant, dans les études populationnelles, il y a quelquefois des différences de fréquences alléliques non reliées au trait étudié. Ce phénomène est appelé stratification. Les trios ne sont pas susceptibles à la stratification puisque le contrôle, constitué du chromosome non transmis, est parfaitement apparié. Les études de liaison et d'association peuvent être pangénomiques ou ciblées à une région chromosomique précise selon le choix des marqueurs. (Modifiée d'après ¹⁷⁴).

Les études de liaison (Figure 9A.) mettent à profit la structure familiale pour suivre les patrons de transmission des allèles-marqueurs et des phénotypes étudiés, une région chromosomique est identifiée comme contenant un locus potentiel quand la maladie et l'allèle-marqueur sont cotransmis plus souvent qu'on s'y attendrait. Cette stratégie d'analyse génétique a été conçue en 1913 par Sturtevant pour ses études sur les drosophiles¹⁷⁵. Mais les études de liaison sur les humains ont longtemps été impossibles à cause de la petite taille des familles (comparée aux études animales), les croisements aléatoires et le petit nombre de marqueurs génétiques connus. La première étude de liaison réussie chez les humains a identifié le gène responsable de la maladie de Huntington en 1983¹⁷⁶. De façon générale, les microsatellites ont été utilisés comme marqueurs de prédilection dans les études de liaison. Dès 1987, une carte pangénomique de quelques 400 marqueurs microsatellites était disponible¹⁷⁷. En 1996, la carte comportait plus de 5000 microsatellites¹⁷⁸. Ces outils génétiques ont facilité l'identification de gènes liés à des maladies.

Quoique très utiles pour identifier les gènes de maladies mendéliennes monogéniques classiques, les études de liaison ont été moins puissantes à identifier les gènes de maladies complexes¹⁶⁸. La plupart des maladies communes et des traits quantitatifs ont des architectures complexes par lesquelles le phénotype est déterminé par la combinaison des interactions entre plusieurs gènes et l'environnement¹⁷⁹. L'effet individuel de chaque facteur de risque pour la maladie sera donc modeste. De plus, les variantes correspondant aux facteurs de risque ont généralement des fréquences élevées et une pénétrance faible. Ces constatations combinées aux tailles relativement petites des cohortes utilisées classiquement dans les études de liaison expliquent la puissance statistique faible et le peu de succès qu'ont eu ces dernières avec les maladies complexes.

Des études de liaison utilisant un ensemble important de marqueurs combinés à une cohorte très large composée de grandes familles peuvent obtenir quelque succès dans

l'étude des maladies complexes. Cependant, les loci de liaison qui font souvent plus de 10 mégabases doivent être ensuite criblés lors d'études d'association pour pouvoir préciser le gène ou la variante causals¹⁸⁰. Tout de même, les études de liaison ont contribué à identifier quelques-unes des régions chromosomiques contenant des gènes de susceptibilité aux MII (Tableau 2)¹⁸¹. L'importance de la liaison est exprimée le plus souvent en terme de *logarithm*₍₁₀₎ of the odds (LOD score). Le LOD score donne une estimation du niveau de liaison entre deux loci, par exemple le locus marqueur et le locus maladie. Un LOD score de plus de 3,6 est considéré comme significatif dans une étude de liaison pangénomique¹⁸². Ce seuil a été le sujet de nombreux débats, mais concrètement, il décrit le LOD score minimal pour lequel on peut affirmer avec moins de 5% d'erreur (P<0,05) que les deux marqueurs sont liés¹⁸².

Les loci IBD1 à IBD6 ont été identifiés comme significatifs dans des études de liaison pangénomique alors que les loci IBD7 et IBD8 suggèrent des liaisons aux MII (voir références tableau 2).

Tableau 2. Premiers loci MII identifiés par liaison

Locus	Région chr.	Commentaires	Référence
IBD1	16q12	Spécifique à la MC, gène <i>NOD2</i>	183
IBD2	12p	Possiblement spécifique à la CU	184
IBD3	6p	locus MII, région HLA	185
IBD4	14q11-12	Possiblement spécifique à la MC Spécifique à la MC, agrégat de gènes codant pour des cytokines	186
IBD5	5q31		181
IBD6	19p	locus MII	181
IBD7	1p35	locus MII	187
IBD8	16p12	Possiblement spécifique à la MC	188

Les premières tentatives d'identification de facteurs de risque génétique aux MII ont été réalisées par des approches de liaison dans des familles multigénérationnelles. Plusieurs régions génomiques ont été identifiées (Tableau 2), mais elles étaient souvent très grandes (plusieurs mégabases) et elles contenaient de multiples gènes (quelquefois des centaines). Pour identifier les gènes causals dans ces régions, les chercheurs se sont tournés vers les études d'association qui sont basées sur le calcul des différences entre les fréquences alléliques dans les cas et dans les contrôles (voir section 2.3.2).

2.3.1.1 *NOD2*

En 2001, le premier locus porteur de risque pour la MC et l'un de ceux qui sont systématiquement répliqués, *NOD2*, a été identifié par Hugot *et al* dans la région de liaison IBD1 sur le chromosome 16, une région qui avait été identifiée précédemment par le même groupe^{139; 183}. Les résultats démontraient un signal d'association à la MC centré sur le gène *NOD2* qui s'amenuisait de part et d'autre du gène. Les auteurs ont donc conclu que *NOD2* était le facteur de risque. Parallèlement, deux groupes indépendants ont publié l'association du gène avec la MC par le biais d'études d'association par gènes candidats^{189; 190}. Trois variantes codantes communes ont été identifiées via ces études (Leu1007insC, Arg702Trp, Gly908Arg). Elles sont toutes localisées dans le domaine LRR (pour *leucine-rich repeat*) impliqué dans la reconnaissance au ligand. Plusieurs autres variantes rares sont soupçonnées d'avoir un impact significatif sur le risque de développer la MC mais leur rareté complique l'évaluation de leur impact véritable¹³⁹. Tel que mentionné dans le chapitre 1 de cet ouvrage (section 1.1.3.3), les récepteurs NOD2 font partie des PRR et reconnaissent des motifs bactériens spécifiques. Ils ont un rôle important dans la réponse immunitaire innée puisqu'ils activent la voie de signalisation de NF- κ B suivant la liaison avec leur ligand MDP¹⁹¹.

2.3.1.2 IBD5

Toujours en 2001, un autre locus IBD qu'on a appelé IBD5 a été identifié dans la région chromosomique 5q31 par cartographie de liaison¹⁹². Parce que cette région renferme un fort déséquilibre de liaison, il est difficile d'identifier une seule variante causale. La région de liaison mesure 243kb et contient cinq gènes : *interferon regulatory factor 1 (IRF1)*, *solute carrier family 22 member 4 et 5 (SLC22A4 et SLC22A5)*, *PDZ and LIM domain protein 4 (PDLIM4)* et *prolyl 4-hydroxylase (P4HA2)*. Ces gènes se révèlent tous être des candidats intéressants du point de vue fonctionnel et les études génétiques entreprises jusqu'à maintenant n'ont pas réussi à identifier le ou les gène(s) causal(s). Un peu plus d'attention a été donnée à *SLC22A4*, qui encode un transporteur organique de cation de type sodium indépendant, et *SLC22A5*, qui encode un transporteur de carnitine à haute affinité de type sodium-dépendant. Ces protéines sont soupçonnées de jouer un rôle important au niveau de la perméabilité de la barrière épithéliale intestinale, plus spécifiquement au niveau de l'élimination des petits cations endogènes, des drogues et des toxines environnementales¹⁹³. Une étude a proposé l'hypothèse que des variantes causales sont situées dans *SLC22A4* et *SLC22A5* et forment un haplotype, TC, porteur de risque de développer la MC¹⁹⁴. D'autres études ont montré que l'effet de ces variantes ne pouvait être séparé de l'effet total d'IBD5^{195; 196}.

2.3.1.3 Le complexe majeur d'histocompatibilité

Le locus IBD3 qui est constitué du complexe majeur d'histocompatibilité (CMH) a été lié aux MII formellement à la suite d'une étude ciblée du chromosome 6p¹⁹⁷. De plus, il est le seul locus qui ait passé le seuil de signification pangénomique dans les méta-analyses de liaison MII ($P < 0,0001$)¹⁹⁷⁻¹⁹⁹. Ce locus hautement polymorphique contient les antigènes des leukocytes humains (HLA) et plus de 200 autres gènes impliqués dans l'immunité²⁰⁰. La densité de gènes fonctionnellement intéressants dans la région et le très haut déséquilibre de liaison rendent difficile l'identification formelle des facteurs de

risque génétique. Plusieurs articles de revue témoignent de l'importance du CMH dans la susceptibilité aux MII^{201; 202}.

2.3.2 Études d'association: la révolution pangénomique

Une des hypothèses concernant l'origine génétique des maladies communes est qu'elles sont causées par des variantes communes. Les études d'association, au cours desquelles la fréquence des allèles à des sites polymorphiques est comparée entre des populations d'affectés et de témoins, sont considérées comme étant des outils plus efficaces pour étudier les traits complexes parce qu'elles ont un plus grand pouvoir statistique pour détecter des variantes communes ayant de faibles effets sur le risque de développer la maladie²⁰³. Il existe deux designs principaux d'études d'association: les études populationnelles de type cas/témoins et les trios qui comprennent deux parents et leur enfant atteint (Figure 9B et 9C). Les études d'association exploitent le déséquilibre de liaison généré par les recombinaisons à travers toute l'histoire ancestrale pour identifier des variantes. À la différence des études de liaison qui tiennent seulement compte des méioses récentes dans chaque famille étudiée. Une approche d'association est capable de détecter des effets génétiques substantiels (i.e., $OR > 2$) sur un phénotype dans des cohortes relativement petites (environ $n=200$) et a un pouvoir élevé de détection de variante de faible effet ($OR < 2$) mais nécessite de larges cohortes ($n > 1000$)²⁰⁴. Jusqu'à récemment, le petit nombre de SNP découverts et l'absence de méthode efficace de génotypage à haut débit empêchaient de pleinement exploiter cette méthode.

Aujourd'hui avec les quelques 20 millions de SNP de la base de données dbSNP (NCBI dbSNP Build 132) et le développement de la technologie des micropuces, les études d'association à grande échelle deviennent possibles. Les études d'association ont permis l'identification de 99 gènes ou loci de susceptibilité aux MII^{205; 206}. Deux stratégies ont été utilisées dans les études d'association: les études par gènes candidats et les études de type pangénomique.

2.3.2.1 Les études d'association par gènes candidats

Historiquement, on a commencé par les études par gènes candidats parce les moyens nécessaires pour les réaliser étaient plus simples que études pangénomiques (moins de marqueurs génétiques, moins de génotypage). Dans ce type d'étude, des gènes sont sélectionnés pour l'étude en fonction de leur localisation génomique, par exemple dans un locus identifié par liaison, ou de leurs fonctions et de leurs implications présumées dans l'étiologie de la maladie²⁰⁷. Ces études sont limitées par les connaissances fonctionnelles sur les gènes et les hypothèses postulées. En effet, même si ces hypothèses sont larges et impliquent, par exemple, tous les gènes d'une voie métabolique, elles vont identifier seulement une fraction des facteurs de risque génétique même pour les maladies pour lesquelles la pathophysiologie est relativement connue. Si les défaillances physiologiques fondamentales d'une maladie sont inconnues, l'approche par gènes candidats sera clairement inadéquate pour expliquer complètement les bases génétiques de la maladie²⁰⁸. De plus, les premières études d'association par gènes candidats avaient généralement de faibles puissances statistiques de détection de variantes associées et étaient plus susceptibles de générer des résultats faux-positifs à cause des petites tailles de cohortes et des seuils de signification statistiques très permissifs²⁰⁹.

2.3.2.2 Les études d'association pangénomiques

Dans les années 90, l'approche par criblage pangénomique a été proposée pour les études d'association²⁰³. Il fallait d'abord développer un catalogue de variantes génétiques humaines communes, des techniques de génotypage à grande échelle puis des outils d'analyses pour toute l'information générée. La publication de la séquence du génome humain^{157; 182}, le développement du projet HapMap¹⁷³ et l'avènement des techniques de génotypage par micropuces ont permis les études d'association pangénomiques (EAP) (voir section 2.2.1). Parce qu'aucune supposition n'est faite sur la localisation génomique des variantes associées, cette approche exploite les forces des études d'association sans avoir à deviner l'identité des gènes²⁰⁸. Les EAP représentent

donc une méthode non biaisée qui peut être tentée même en l'absence d'information concernant la fonction ou la localisation des gènes causals. Elles fournissent des informations non biaisées et détaillées sur l'héritabilité des maladies à un niveau qui n'était pas envisageable il y a quelques années seulement. Le nombre important de publications de ce type d'étude dans les dernières années témoignent de leurs popularités.

Une longue période sans découverte majeure de nouveaux loci de susceptibilité aux MII a suivi l'identification de NOD2 et du locus IBD5. Ceci s'explique par le fait que la plupart des études d'association tentées durant cette période étaient restreintes au génotypage dans de petites cohortes, d'un petit nombre de variantes génétiques localisées dans des gènes sélectionnés pour leur implication potentielle dans la pathogenèse des MII (étude d'association par gènes candidats). Grâce aux études pangénomiques, le nombre de loci connus associés aux MII a augmenté de façon exponentielle. La figure 10 présente un graphique du nombre de loci connus associés à la MC en fonction de l'année et révèle l'impact des études d'association pangénomiques (la première dans le domaine a été publiée en 2005²¹⁰) dans l'identification des loci de susceptibilité à cette maladie.

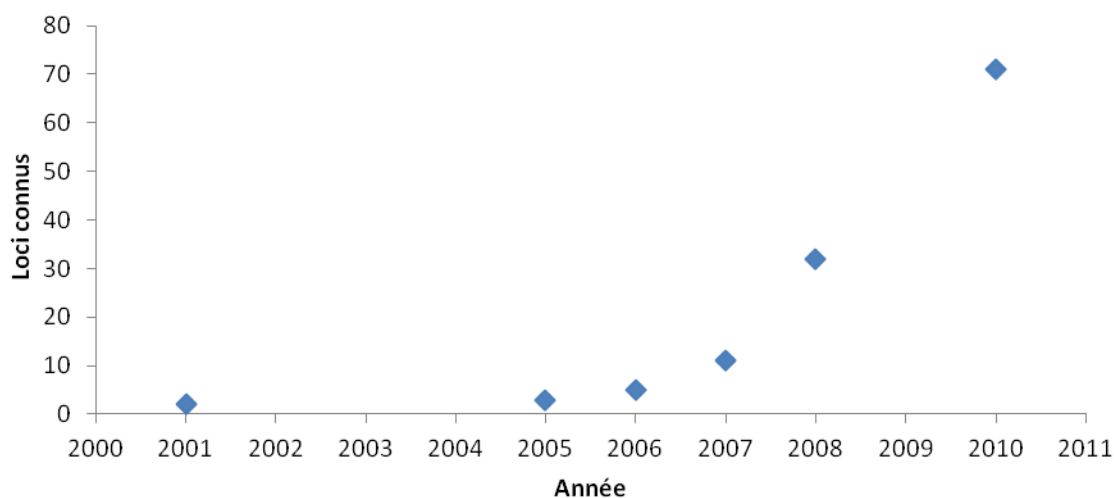


Figure 10. Loci de susceptibilité à la MC connus

Le graphique montre l'augmentation exponentielle du nombre de loci de susceptibilité à la MC connus dans les quatre dernières années. Cette augmentation coïncide avec le développement des études d'association pangénomiques.

La plus récente métaanalyse porte le nombre de loci connus de la MC, confirmés par des études pangénomiques, à 71²⁰⁶. La dernière métaanalyse sur la CU recense 47 loci²⁰⁵.

Les tableaux I1 et I3 (en annexe) présentent la liste des loci identifiés jusqu'à maintenant par les études d'association pangénomiques. En considérant le chevauchement entre les loci de susceptibilité aux deux maladies, il y a à ce jour 99 loci de susceptibilité aux MII. La principale caractéristique de ces derniers est leur effet faible qui se traduit par un OR entre 1 et 1,2 pour la plupart des loci (Tableau I1). Plusieurs de ces loci contiennent plusieurs gènes et certains n'en contiennent pas du tout. Les tableaux I2 et I3 (en annexe) présentent les gènes prédits dans chaque locus associé. Le facteur de risque génétique (identifié) le plus important pour la MC mais aussi pour la CU selon les dernières métaanalyses semble être *IL23R* avec un OR de 2,67 et 1,72, pour la MC et la CU respectivement. Même si un faible pourcentage des loci impliqués dans la MC, la CU et d'autres maladies inflammatoires a été identifié jusqu'à maintenant, on observe que certains gènes ont été identifiés seulement dans la MC (comme *NOD2*), dans la CU (comme *HNF4α*), alors que d'autres sont spécifiques

aux MII au sens large (*MST1*) et d'autres encore sont partagés avec d'autres maladies inflammatoires (comme *IL23R*). Les études bio-informatiques d'enrichissement génique réalisées sur les gènes candidats dans les loci identifiés révèlent de façon générale, une surreprésentation de molécules impliquées dans les processus immunitaires telles que des cytokines, des récepteurs de cytokines et des régulateurs de la réponse immunitaire innée et acquise^{205; 206}. Les gènes identifiés impliquent spécifiquement divers phénomènes tels que l'autophagie (ex. :*ATG16L1*, *IRGM*, *DAP*), les processus moléculaires liés à la barrière épithéliale (ex. :*GNAI2*, *HNF4α*, *CDH1*, *LAMB1*), le développement des lymphocytes B (*BLIMP-1*) et la différenciation des lymphocytes T (*IL23R*, *IL10*, *IL7R*, *IL21*, *IFNG*) dans le développement des MII.

2.3.3 L'héritabilité manquante: stratégie d'identification des autres facteurs de risque génétique

Le *Genome-Wide Association Studies catalog* recense 867 publications d'études d'association pangénomiques (www.genome.gov/gwastudies, accédé le 5 mai 2011). Toutes ces publications ont considérablement augmenté le nombre de gènes de susceptibilité aux maladies complexes connus. Toutefois, pour la plupart des maladies complexes comme les MII, un pourcentage relativement faible de l'héritabilité est expliqué par les variantes découvertes jusqu'à maintenant. Par exemple, la dernière métaanalyse de six études pangénomiques sur la CU qui totalisent 6782 cas et 20099 contrôles a identifié 31 nouveaux loci pour amener le nombre total de loci connus à 47. Malgré tout, les auteurs de l'étude estiment que seulement 16% de l'héritabilité est expliquée par ces loci²⁰⁵. Parallèlement, une première métaanalyse comportant 3,230 cas et 4,829 contrôles a porté le nombre de loci de la MC à 32²¹¹. Cette étude expliquait environ 20% de l'héritabilité. La plus récente métaanalyse du même groupe totalisant six études pangénomiques individuelles soit 6333 cas et 15056 contrôles a porté le nombre de loci de MC à 71, et malgré tout les auteurs calculent que seulement 23,2% de l'héritabilité est maintenant expliquée²⁰⁶. Une faible augmentation de 3% alors que le

nombre de loci connus a plus que doublé! Ce constat révèle plusieurs réalités dont l'effet individuel modeste de chaque gène et l'imperfection des SNP-marqueurs à capturer complètement l'effet des variantes causales. Conséquemment, plusieurs variantes et plusieurs gènes porteurs de risque demeurent jusqu'ici inconnus.

La puissance statistique des études d'association pangénomiques actuelles est l'une des raisons qui explique cette situation. Altshuler *et al.* ont calculé la taille des cohortes pour s'assurer un pouvoir de détection de 90%, 50% et 10% en fonction du risque relatif approché (OR) (figure 11)¹⁶⁸. Étant donné que la majorité des OR pour les gènes de susceptibilité aux MII se situent entre 1 et 1.5, les tailles de cohortes requises pour avoir 90% de pouvoir statistique de détecter une association significative avec une variante dont la fréquence allélique est >10% sont de l'ordre de dizaines de milliers d'échantillons (figure 11). La taille des cohortes nécessaires est augmentée considérablement quand la fréquence est plus petite. Pour une variante rare dont la fréquence est <1% et le OR est de l'ordre de 1.2, la taille requise pour avoir 90% de chance d'identifier une association significative monte au-delà de 100 000 échantillons.

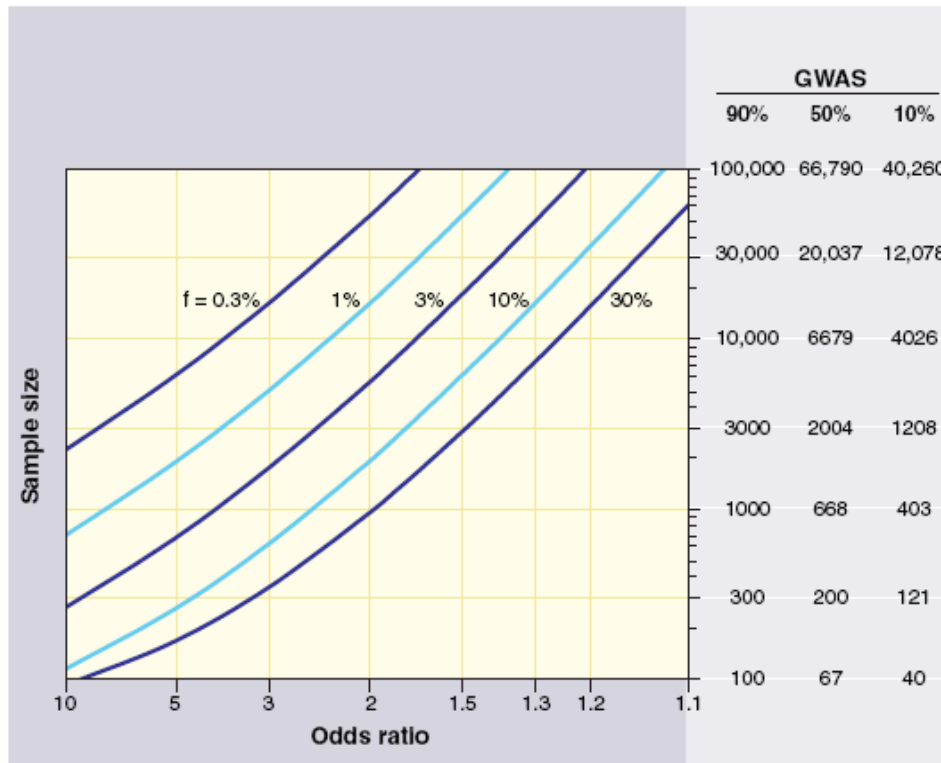


Figure 11. Tailles requises des cohortes pour des études d'association pangénomiques

Ce graphique présente le nombre total d'échantillon requis N (où le nombre de contrôles ou de cas est égale $N/2$) pour cartographier une variante génétique en fonction de l'effet de cette variante (Odds Ratio, OR) et de sa fréquence (f). Dans les trois premières colonnes de droite, sont alignés les N calculés pour les études d'association pangénomiques (GWAS) utilisant des variantes communes. Les tailles des cohortes requises pour obtenir un résultat significatif avec une valeur $P < 10^{-8}$ à différents niveaux de pouvoir statistiques (90%, 50%, 10%) sont précisées. Figure modifiée d'après ¹⁶⁸

En partie à cause des imperfections dans les études pangénomiques actuelles, une partie du risque génétique des maladies inflammatoires de l'intestin est toujours inexplicée. Dans le jargon génétique, on parle d'héritabilité manquante (ou *missing heritability*). On estime qu'au moins une partie de cette héritabilité se trouve dans les variantes rares qui ne sont pas cartographiées adéquatement. Une des stratégies envisagées est de re-séquencer tous les loci identifiés pour identifier des variantes rares qui seraient

potentiellement causales. Des projets de re-séquençage à grande échelle se sont mis en branle. Aidés des nouvelles technologies de séquençage à haut débit, ces projets visent à augmenter et intégrer les connaissances sur la variation génétique humaine et son rôle dans les maladies. Le consortium HapMap 3 s'est donné pour but de séquencer des régions spécifiques du génome chez 692 individus provenant de dix populations mondiales et d'évaluer les niveaux de corrélation entre les SNP communs (<5%) et les variantes rares²¹². Les auteurs concluent que les variantes moins fréquentes sont moins partagées à travers les populations et que ces variantes sont moins bien imputées par les variantes communes. Ils soulignent l'importance de considérer des échantillons de populations provenant de différents continents, d'étudier cette variation, de mieux l'intégrer dans les études d'association et ce faisant, d'améliorer les chances d'identifier des loci.

Le projet 1000genomes²¹³ vise quant à lui à établir un catalogue plus complet de la variation du génome humain en séquençant au moins 1000 personnes provenant de différentes populations mondiales (incluant les échantillons HapMap). Par une couverture génomique de 4x par génome, ce projet vise à identifier tous les SNP avec une fréquence supérieure à 1%. Cette nouvelle carte permettra de génotyper ces SNP qui individuellement sont rares mais qui collectivement sont fréquents et qui pourraient expliquer une partie de l'héritabilité des maladies dites communes.

Des projets individuels de re-séquençage ayant pour but d'identifier les gènes causals dans les loci associés aux maladies à la suite d'études pangénomiques commencent également à être publiés. Les loci associés aux maladies complexes peuvent contenir plusieurs gènes (voir tableaux I2 et I3 en annexe). Le génotypage de variantes rares dans les exons ou les régions ayant des fonctions connues peut permettre de préciser le gène causal et l'impact des variantes sur la fonction (s'ils tombent dans des domaines fonctionnels connus). Par exemple, Nejentsev *et al* ont resequencé les exons de 10 gènes candidats pour le diabète de type 1 (DT1) dont 6 sont localisés dans des loci de

susceptibilité préalablement identifiés²¹⁴. Quatre variantes rares du gène *interferon induced with helicase C domain 1* (IFIH1) conférant un risque au DT1 indépendamment les unes des autres ont été identifiées. *IFIH1* est localisé dans le locus 2q24 qui est associé au DT1 (mais n'est pas unique dans ce locus)²¹⁵. Les variantes rares associées affectent l'épissage de l'ARNm et provoquent une troncature de la protéine. *IFIH1* encode une protéine impliquée dans l'activation de la réaction immunitaire suivant une infection aux picornavirus²¹⁶. L'infection par ce type de virus est plus commune chez les patients venant d'être diagnostiqués comme souffrant de DT1 que chez la population normale et précède l'apparition des auto-anticorps caractéristiques de la maladie²¹⁷. Cette étude de reséquençage a donc permis d'orienter les recherches futures vers une voie biologique qui est impliquée dans la manifestation de la maladie.

2.4 Des découvertes génétiques qui ont mené à une meilleure compréhension des mécanismes sous-jacents au développement des MII

2.4.1 La voie du récepteur de l'interleukine 23 (IL23R)

La découverte du gène de susceptibilité aux MII *NOD2* a depuis longtemps permis de faire le rapprochement entre la dérégulation de la réponse immunitaire innée et la pathogenèse de ces maladies. Dès les débuts des études génétiques sur les MII, la réponse immunitaire acquise était également soupçonnée de jouer un rôle dans la pathogenèse sans qu'on puisse vraiment le prouver. L'identification du facteur de risque *IL23R* codant pour une sous-unité du récepteur de l'interleukine 23 au cours de la première étude d'association pangénomique sur les MII a permis d'impliquer la réponse immunitaire acquise de façon définitive et ouvert une nouvelle avenue pour la recherche thérapeutique sur les maladies inflammatoires²¹⁸.

IL23R représente une des deux parties du récepteur hétérodimère de IL-23, l'autre sous-unité est encodée par le gène *IL12RB1* qui fait également partie du récepteur de l'interleukine 12 (IL-12). IL-23 est elle-même composée de deux sous-unités assemblées en hétérodimère, la sous-unité p19 (encodée par le gène *IL23A*) et la sous-unité p40 (encodée par le gène *IL12B*), cette dernière étant partagée avec IL-12. Le gène *IL12B* a depuis été associé à la maladie de Crohn²¹¹. L'association de ces deux gènes (*IL12B* et *IL23R*) avec les MII suggère que l'interaction entre le récepteur IL-23 et sa cytokine déclenchent une cascade de signalisation cruciale aux processus inflammatoires menant aux MII. De plus, pour étayer le rôle de la voie IL-23 dans le développement des MII, d'autres gènes encodant des molécules dont l'impact se fait sentir en aval du récepteur dans la voie ont récemment été identifiés comme facteurs de risque^{211; 219}. Par exemple, les gènes *janus kinase 2 (JAK2)* et *tyrosine kinase 2 (TYK2)* encodent des protéines qui s'associent aux récepteurs de cytokines dont celui d'IL-23

puis phosphorylent les molécules de signalisation comme *Signal Transducer and Activator of Transcription 3* (STAT3) (Figure 12). Toutefois, il faut souligner que ces molécules (JAK2, TYK2 et STAT3) ne sont pas spécifiques à la voie de IL-23, ils interviennent notamment dans la voie de IL-10 (qui a également été associée au MII²¹⁹). La variante d'*IL23R* la plus fortement associée aux MII, Arg381Gln, est située dans le domaine de liaison à la kinase JAK2, ce qui suggère que la variante module la cascade de signalisation via son interaction avec JAK2.

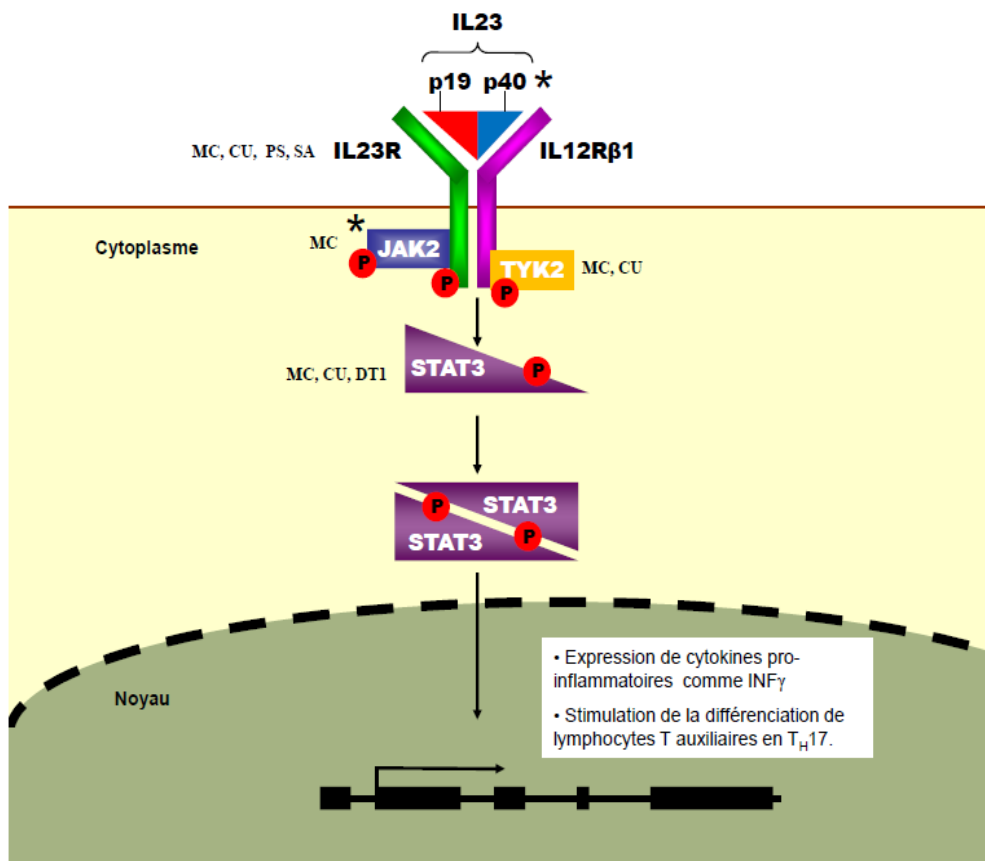


Figure 12. La voie IL23R et l'inflammation

La cytokine hétérodimère IL-23 se lie à son récepteur et ce faisant déclenche une cascade de signalisation, incluant la phosphorylation de STAT3 via JAK2 et TYK2, menant à l'expression de cytokines pro-inflammatoires et la stimulation de la différenciation des lymphocytes T auxiliaires en T_H17 . L'intérêt de la voie IL-23 est d'autant plus grand qu'en plus d'être impliquée dans les MII (maladie de Crohn MC, colite ulcéreuse CU), elle a été associée à plusieurs autres maladies inflammatoires telles que le psoriasis (PS), la spondylarthrite ankylosante (SA), et le diabète de type 1 (DT1). De plus, des thérapies biologiques (section 1.2.3.5) ciblant certaines molécules de la voie (ces molécules sont identifiées par un *) sont présentement étudiées. Figure modifiée d'après¹⁴.

Suite à sa phosphorylation, STAT3 s'homodimérise, se déplace dans le noyau et induit la transcription de cytokines telles que IL-17, IL-17F, IL-22 et IFN- γ . Les cytokines IL-17 et IL-17F sont majoritairement produites par les T_H17. Les T_H17 expriment de hauts niveaux du récepteur IL23R. Les cytokines IL-17 et IL-17F ont un rôle important dans la réponse immunitaire en générale et dans les maladies inflammatoires en particulier. Elles sont impliquées dans le recrutement et l'activation des neutrophiles durant la réponse immunitaire contre les envahisseurs pathogènes tels que les bactéries et champignons²²⁰. Chez les patients souffrant de MC en phase active, les niveaux de IL-17 et IL-23 sont augmentés comparés au niveau d'un groupe témoin²²¹. De plus, il a été démontré qu'IL-17 augmente l'expression de la cytokine IL-8 et du facteur de nécrose tumorale alpha (TNF- α) dans les cellules épithéliales traitées au LPS. Combinés aux associations de plusieurs gènes de la voie IL-23 et de gènes reliés à la différenciation de T_H17, ces résultats indiquent que l'activation de l'axe IL-23/IL17 est reliée de façon fondamentale à l'étiologie des MII et pourrait expliquer la nature récidivante de ces maladies par le biais de l'augmentation de la sensibilité de l'épithélium aux molécules bactériennes telles que le LPS²²¹.

La voie IL-23 a été associée à plusieurs autres maladies inflammatoires (voir figure 12). Ces observations révélatrices ont orienté les recherches thérapeutiques vers les molécules de la voie IL-23. Plusieurs essais cliniques sont en cours. En effet, les informations générées par les études génétiques et les études fonctionnelles et cliniques qui s'ensuivent pourront potentiellement permettre de mieux apparier les patients avec les traitements adaptés à leur type de maladie.

2.4.2 L'autophagie et les MII

Une découverte excitante et inopinée des études d'association pangénomiques sur la maladie de Crohn est l'importance de l'autophagie dans la pathogenèse de la maladie. L'identification du gène *ATG16L1* comme facteur de risque à la MC lors de deux études indépendantes a souligné l'implication de ce processus biologique dans

l'inflammation^{222; 223}. Toutefois, l'aspect marquant de cette découverte est venu par la suite lorsqu'on a réalisé que *NOD2* le premier facteur de risque aux MII identifié et *ATG16L1* agissaient dans la même voie biologique. En effet, les récepteurs NOD1 et NOD2 dirigent l'autophagie via le recrutement de la protéine ATG16L1 à la membrane plasmique au point de pénétration bactérienne²²⁴. Cette observation souligne l'intérêt des études fonctionnelles pour acquérir une meilleure compréhension du contexte biologique entourant les facteurs de risque génétique et ce faisant, une meilleure compréhension de la pathogenèse des maladies. D'autres parts, cela suggère que les nombreux facteurs de risque aux MII pourraient agir de concert dans un nombre limité de voies biologiques. L'identification de ces voies permettrait de comprendre enfin les mécanismes régissant la dérégulation de l'inflammation et ce faisant de cibler des traitements adéquats aux patients.

L'autophagie est un mécanisme ancien et conservé chez toutes les cellules eucaryotes. Dans sa forme la plus basique, il s'agit d'un mécanisme cellulaire dont le but est d'autodigérer les constituants intracellulaires lors de famine, de recycler les organelles superflues ou endommagées et de dégrader certaines protéines²²⁵. Lorsque le processus implique un pathogène intracellulaire, on parle de macroautophagie. Lors de la macroautophagie, une structure appelée membrane d'isolation se forme autour du pathogène. Ce dernier est encapsulé davantage dans une organelle formée d'une double membrane, appelée l'autophagosome, et transporté au lysosome pour être dégradé²²⁶. ATG16L1 fait partie des protéines clefs qui s'associent aux lipides membranaires et contrôle l'élongation et la stabilisation de l'autophagosome²²⁷.

ATG16L1 joue un rôle majeur dans la régulation négative des cytokines pro-inflammatoires. La délétion du gène *ATG16L1* entraîne la surrexpression des cytokines IL-1 β et IL-18 dans des macrophages de souris²²⁸. De plus, la fonction d'ATG16L1 serait intimement reliée à celle du récepteur NOD2. Le ciblage des bactéries par les autophagosomes est augmenté suite à une stimulation de NOD2 par MDP, et cette

augmentation ne se produit pas dans les cellules exprimant le variant associé aux MII, *ATG16L1* T300A^{224; 229}. NOD2 et ATG16L1 interagissent physiquement à la membrane plasmique et cette interaction pourrait permettre à NOD2 de recruter la machinerie reliée à l'autophagie lorsque des bactéries entrent dans la cellule²²⁴.

À la suite d'*ATG16L1*, d'autres gènes impliqués dans le processus de l'autophagie ont été associés aux MII. Le gène *IRGM* a été associé à la MC²³⁰ et le gène *LRRK2* est situé dans un locus associé à la MC²⁰⁶. Le gène *DAP* qui encode un régulateur négatif de l'autophagie a récemment été associé avec la CU²⁰⁵. Peu à peu, on commence à relier les différents facteurs de risque génétique dans des voies biologiques. Il est possible que les maladies complexes, comme les MII, prennent leur origine dans une confluence d'un nombre limité de voies biologiques interrompues par des protéines encodées par les facteurs de risque génétique. Un individu pourrait n'avoir besoin que d'un certain nombre de ces facteurs pour développer la maladie, la clef serait d'avoir la bonne (ou la mauvaise!) combinaison pour la développer²³¹.

Chapitre 3

La région candidate 19p

3.1 Identification du locus

Le point central de cet ouvrage est la région 19p qui a été identifiée pour la première fois en 2000 comme une région qui inclut un ou des facteurs de risque génétique pour les MII. L'étude de liaison pangénomique sur 158 familles de paires de germains canadiens atteints de MII¹⁸¹ s'est penchée sur la ségrégation de 312 marqueurs microsatellites. Le seuil accepté de signification pangénomique dans les études de liaison est un LOD score de 3,6¹⁸² et la région du chromosome 19p qu'on a appelé par la suite IBD6 le dépasse largement avec un LOD score de 4,6 (Figure 13). Selon cette étude le locus semble être partagé par les deux MII et il ne s'agit pas d'un locus lié au diagnostic précoce de la MC (voir tableau 3). Le facteur relatif à la fratrie λ_s - c'est-à-dire le risque relatif que le frère ou la sœur d'un atteint soit également atteint, comparé au risque de la population générale – est de 2,2. Les résultats d'une métaanalyse de 10 études de liaison pangénomiques indépendantes (dont l'étude originale¹⁸¹) totalisant 1952 paires de germains atteints de MII suggèrent également la présence d'un locus de liaison à la MC à la région 19p¹⁹⁸.

Cette région de liaison sur le chromosome 19p a été nommée IBD6 (pour *Inflammatory Bowel disease Locus 6*). La région qui s'étend sur 28,5Mb comprend tout le petit bras du chromosome 19.

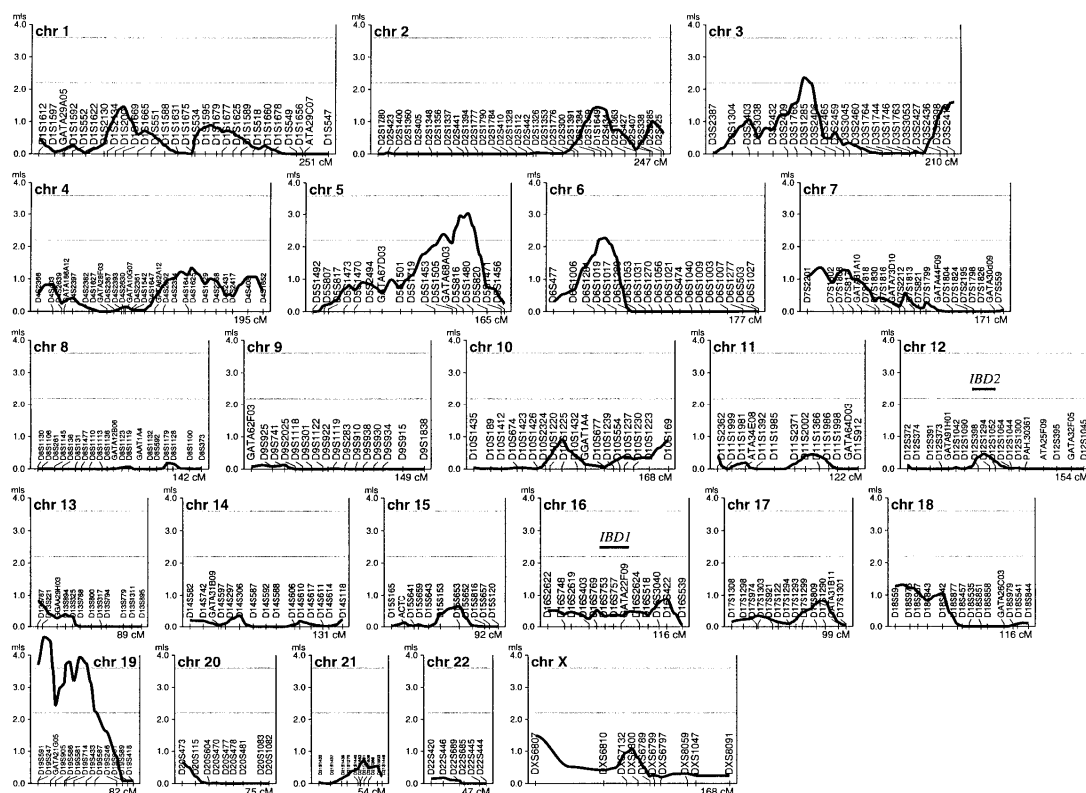


Figure 13. Une étude de liaison pangénomique identifie un pic de liaison sur le chromosome 19p

Résultats des analyses multipoints du criblage pangénomique. Les axes verticaux indiquent les LOD scores en fonction de la localisation chromosomique. Les traits verticaux sur les axes horizontaux indiquent la position des marqueurs microsatellites. Les deux lignes horizontales à l'intérieur de chaque graphique représentent les seuils acceptés dans les études pangénomiques pour une liaison suggérée (LOD score 2,2) et une liaison significative (LOD score 3,6). Les régions IBD1 et IBD2 sur les chromosomes 16 et 12, respectivement, représentent les loci de susceptibilité rapportés dans la littérature ^{183; 184} 181

Tableau 3. Résumé des résultats de liaison en fonction du phénotype

Catégorie Phénotypique	No.de familles	LOD score
MII	158	4.6
CU	24	2.9
MC	110	3
MC16	46	1.5

MII= maladies inflammatoires de l'intestin, CU=colite ulcéreuse, MC= maladie de Crohn, MC16= maladie de Crohn à diagnostic précoce (avant 16 ans).

3.2 Études par gènes candidats sur le chromosome 19

Le chromosome 19 fait partie du groupe des petits chromosomes métacentriques. Il se compose de plus d'éléments répétitifs (55%) que la moyenne génomique (44.8%). Son contenu en GC de 48% est également plus élevé que la moyenne génomique de 41%²³². Il est le plus dense au niveau génique. Le chromosome 19p contient 638 gènes connus²³². Plus de 25% des gènes du chromosome 19 sont membres d'une famille de gènes arrangés en tandem. Le plus grand groupe de gènes sur le chromosome 19 encode des facteurs de transcription de type C2H2 doigt de zinc (ZNF). 266 de ces gènes sont groupés en 11 agrégats sur le chromosome 19. Un de ces agrégats localisé dans la région péri-centromérique du chromosome 19p est exclusif à la lignée des primates²³³. La famille CYP4F est une famille de gènes d'un intérêt fonctionnel, localisés sur le chromosome 19p. Membres de la superfamille des cytochromes P450, les CYP4F codent pour des enzymes impliqués dans l'inflammation et le métabolisme des drogues²³⁴.

La région 19p présente donc un vaste éventail de gènes candidats ayant des fonctions immunitaires et inflammatoires et potentiellement impliqués dans les maladies inflammatoires de l'intestin. La suite logique de l'étude de Rioux *et al.*¹⁸¹ à l'époque était de tenter d'identifier le ou les gènes causaux par une étude d'association. C'est ce que Tello-Ruiz et collaborateurs ont fait avec leur analyse de 56 gènes candidats potentiels localisés dans la région 19p chez 180 trios et 343 patients MII et 207 contrôles indépendants²³⁵. Les gènes candidats ont été retenus pour leur intérêt fonctionnel basé sur des recherches de la littérature. Parmi les gènes sélectionnés se trouvent des gènes impliqués dans la migration cellulaire, des régulateurs immunitaires, des récepteurs de surface et des cytokines²³⁵. De plus, les auteurs ont tenté de répliquer les associations à des SNP codants non synonymes dans les gènes *C3* et *ICAMI* publiées précédemment dans la littérature²³⁶⁻²³⁸. Aucune association n'a pu être observée dans ces gènes, ni dans les 56 candidats.

Un autre groupe a identifié, dans le locus 19p13, un facteur de risque pour la maladie de cœliaque, le gène *MYO9B*²³⁹. La maladie de cœliaque est une maladie intestinale inflammatoire déclenchée par l'ingestion de gluten chez les individus qui possèdent une susceptibilité génétique²⁴⁰. *MYO9B* est exprimé dans les cellules intestinales humaines et sa surexpression dans les cellules épithéliales de rat occasionne des modifications morphologiques liées aux filaments d'actine^{241; 242}. Pour vérifier si les variantes de *MYO9B* entraînaient aussi la susceptibilité aux MII classiques, Van Bodegraven et al. ont génotypé 8 SNP, qui couvrent toute la variabilité du gène *MYO9B* ($r^2 > 0.8$), chez 2717 patients avec une MII et 4440 contrôles de descendance européenne²⁴³. Le SNP non-synonyme rs1545620 est celui à qui a été attribué le signal d'association le plus fort ($P = 1.9 \times 10^{-6}$) dans la cohorte MII. Toutefois, les auteurs déclarent que l'OR (1,2) et les fréquences alléliques (fréquence d'allèle mineur 43%) suggèrent une faible contribution au risque relatif à la fratrie ($\lambda_s < 1.02$) et donc que ce gène ne contribue pas de façon significative au pic de liaison observé par Rioux *et al.* dans le locus IBD6. Ce qui nous amène au but de la présente étude qui est d'identifier d'autres facteurs de risque aux MII dans le locus 19p.

3.3 Hypothèses et objectifs du projet

Les conclusions des études mentionnées précédemment dans les sections 3.1 et 3.2, nous amènent à postuler certaines hypothèses sur la région chromosomique 19p.

- Il existe une ou plusieurs variantes porteuses de risque génétique aux MII dans la région de liaison 19p.
- Ces variantes associées sont situées dans des gènes ou dans des régions non-codantes régulatrices qui modifient l'expression de gènes spécifiques associés aux MII.
- Ces gènes associés encodent des protéines ayant des fonctions immunitaires.
- Des modifications dans ces fonctions immunitaires ont un impact sur le phénotype des MII.

Les objectifs de ce projet de doctorat sont regroupés en deux volets. Le premier volet concerne la génétique alors que le second relève davantage de la biologie fonctionnelle.

- Dans le premier volet, les objectifs étaient de cribler de façon systématique la variation génétique dans la région du chromosome 19 appelé IBD6 pour identifier des variantes associées aux MII dans une première cohorte, et de répliquer les signaux d'association dans une cohorte indépendante. Ensuite, on voulait relier ce ou ces variantes génétiques à un (ou des gènes) précis afin d'identifier le (ou les) candidat(s) pour des études fonctionnelles.
- Dans le second volet, l'objectif était d'initier la caractérisation fonctionnelle du facteur de risque génétique, *MAST3*, identifié à la suite de l'étude génétique. C'est-à-dire, évaluer par des études d'expression l'étendue de son impact sur les tissus d'intérêt (immunitaires et intestinaux), ensuite décrire son rôle dans la

cellule (par le biais de surexpression et knockdown) pour pouvoir comprendre son impact sur l'inflammation et les maladies inflammatoires de l'intestin.

Bibliographie

1. Bernstein CN, Wajda A, Svenson LW, MacKenzie A, Koehoorn M, Jackson M *et al.* The epidemiology of inflammatory bowel disease in Canada: a population-based study. *Am J Gastroenterol* 2006; **101**(7): 1559-1568.
2. Crohn B, Ginzburg L & Oppenheimer EH. Regional ileitis, pathologic and clinical entity. *JAMA* 1932; **99**: 1323-1329.
3. De Dombal FT. Ulcerative colitis: definition, historical background, aetiology, diagnosis, naturel history and local complications. *Postgrad Med J* 1968; **44**(515): 684-692.
4. Farrell J. Digestion and Absorption of Nutrients and Vitamins *Feldman: Sleisenger and Fordtran's Gastrointestinal and Liver Disease, 9th ed.*, 2006, pp 2147-2197.
5. Mulvihill SJ, Stone MM, Fonkalsrud EW & Debas HT. Trophic effect of amniotic fluid on fetal gastrointestinal development. *J Surg Res* 1986; **40**(4): 291-296.
6. Guyton A & Hall J. Gastrointestinal Physiology *Medical Physiology*, 11th Edition edn. Elsevier Saunders Philadelphia, Pennsylvania, 2006, p 772.
7. Vander A, Sherman J & Luciano D. The digestion and absorption of food *Human Physiology*, 3rd edn. McGraw-Hill, 1980.
8. Savage DC. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 1977; **31**: 107-133.
9. Hooper LV & Macpherson AJ. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* 2010; **10**(3): 159-169.
10. Ley RE, Lozupone CA, Hamady M, Knight R & Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 2008; **6**(10): 776-788.
11. Fiocchi C. The normal intestinal mucosa: a state of "controlled inflammation" *Inflammatory bowel disease: from bench to bedside*, 2nd edn. Kluwer Academic Publishers, 2005, pp 101-120.

12. Mayer EA, Raybould H & Koelbel C. Neuropeptides, inflammation, and motility. *Dig Dis Sci* 1988; **33**(3 Suppl): 71S-77S.
13. Scott LD & Cahall DL. Influence of the interdigestive myoelectric complex on enteric flora in the rat. *Gastroenterology* 1982; **82**(4): 737-745.
14. Budarf ML, Labbe C, David G & Rioux JD. GWA studies: rewriting the story of IBD. *Trends Genet* 2009; **25**(3): 137-146.
15. Lehrer RI, Barton A, Daher KA, Harwig SS, Ganz T & Selsted ME. Interaction of human defensins with *Escherichia coli*. Mechanism of bactericidal activity. *J Clin Invest* 1989; **84**(2): 553-561.
16. Pullan RD, Thomas GA, Rhodes M, Newcombe RG, Williams GT, Allen A *et al*. Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis. *Gut* 1994; **35**(3): 353-359.
17. Chang SK, Dohrman AF, Basbaum CB, Ho SB, Tsuda T, Toribara NW *et al*. Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology* 1994; **107**(1): 28-36.
18. Hamer HM, Jonkers DM, Loof A, Vanhoutvin SA, Troost FJ, Venema K *et al*. Analyses of human colonic mucus obtained by an in vivo sampling technique. *Dig Liver Dis* 2009; **41**(8): 559-564.
19. Cahill RN, Poskitt DC, Hay JB, Heron I & Trnka Z. The migration of lymphocytes in the fetal lamb. *Eur J Immunol* 1979; **9**(3): 251-253.
20. Ohtsuka Y, Lee J, Stamm DS & Sanderson IR. MIP-2 secreted by epithelial cells increases neutrophil and lymphocyte recruitment in the mouse intestine. *Gut* 2001; **49**(4): 526-533.
21. Salmi M & Jalkanen S. Human leukocyte subpopulations from inflamed gut bind to joint vasculature using distinct sets of adhesion molecules. *J Immunol* 2001; **166**(7): 4650-4657.
22. Baggiolini M, De Duve C, Masson PL & Heremans JF. Association of lactoferrin with specific granules in rabbit heterophil leukocytes. *J Exp Med* 1970; **131**(3): 559-570.

23. Miyauchi H, Hashimoto S, Nakajima M, Shinoda I, Fukuwatari Y & Hayasawa H. Bovine lactoferrin stimulates the phagocytic activity of human neutrophils: identification of its active domain. *Cell Immunol* 1998; **187**(1): 34-37.
24. Palumbo D, Iannaccone M, Porta A & Capparelli R. Experimental antibacterial therapy with puroindolines, lactoferrin and lysozyme in *Listeria monocytogenes*-infected mice. *Microbes Infect*; **12**(7): 538-545.
25. Lee JY, Kim YY, Chang JY, Park MS & Kho HS. The effects of peroxidase on the enzymatic and candidacidal activities of lysozyme. *Arch Oral Biol*; **55**(9): 607-612.
26. Harnack U, Johnen H & Pecher G. Natural Killer Cell Line YT Exerts Cytotoxicity Against CD86+ Myeloma Cells. *Anticancer Res* 2011; **31**(2): 475-479.
27. Fauriat C, Long EO, Ljunggren HG & Bryceson YT. Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood* 2010; **115**(11): 2167-2176.
28. Ear T & McDonald PP. Cytokine generation, promoter activation, and oxidant-independent NF-kappaB activation in a transfectable human neutrophilic cellular model. *BMC Immunol* 2008; **9**: 14.
29. Foey AD, Feldmann M & Brennan FM. CD40 ligation induces macrophage IL-10 and TNF-alpha production: differential use of the PI3K and p42/44 MAPK-pathways. *Cytokine* 2001; **16**(4): 131-142.
30. Zuniga EI, McGavern DB, Pruneda-Paz JL, Teng C & Oldstone MB. Bone marrow plasmacytoid dendritic cells can differentiate into myeloid dendritic cells upon virus infection. *Nat Immunol* 2004; **5**(12): 1227-1234.
31. Kalinski P, Hilkens CM, Snijders A, Snijdewint FG & Kapsenberg ML. IL-12-deficient dendritic cells, generated in the presence of prostaglandin E2, promote type 2 cytokine production in maturing human naive T helper cells. *J Immunol* 1997; **159**(1): 28-35.
32. Chuang T & Ulevitch RJ. Identification of hTLR10: a novel human Toll-like receptor preferentially expressed in immune cells. *Biochimica et biophysica acta* 2001; **1518**(1-2): 157-161.

33. Chuang TH & Ulevitch RJ. Cloning and characterization of a sub-family of human toll-like receptors: hTLR7, hTLR8 and hTLR9. *Eur Cytokine Netw* 2000; **11**(3): 372-378.
34. Du X, Poltorak A, Wei Y & Beutler B. Three novel mammalian toll-like receptors: gene structure, expression, and evolution. *Eur Cytokine Netw* 2000; **11**(3): 362-371.
35. Medzhitov R, Preston-Hurlburt P & Janeway CA, Jr. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* 1997; **388**(6640): 394-397.
36. Rock FL, Hardiman G, Timans JC, Kastelein RA & Bazan JF. A family of human receptors structurally related to Drosophila Toll. *Proc Natl Acad Sci U S A* 1998; **95**(2): 588-593.
37. Takeuchi O, Kawai T, Sanjo H, Copeland NG, Gilbert DJ, Jenkins NA *et al.* TLR6: A novel member of an expanding toll-like receptor family. *Gene* 1999; **231**(1-2): 59-65.
38. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J Immunol* 1999; **162**(7): 3749-3752.
39. Kawai T & Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* 2010; **11**(5): 373-384.
40. Medzhitov R. Origin and physiological roles of inflammation. *Nature* 2008; **454**(7203): 428-435.
41. Franchi L, Warner N, Viani K & Nunez G. Function of Nod-like receptors in microbial recognition and host defense. *Immunol Rev* 2009; **227**(1): 106-128.
42. Barnich N, Aguirre JE, Reinecker HC, Xavier R & Podolsky DK. Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor- κ B activation in muramyl dipeptide recognition. *J Cell Biol* 2005; **170**(1): 21-26.
43. Kufer TA, Kremmer E, Adam AC, Philpott DJ & Sansonetti PJ. The pattern-recognition molecule Nod1 is localized at the plasma membrane at sites of bacterial interaction. *Cell Microbiol* 2008; **10**(2): 477-486.

44. Hozumi N & Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc Natl Acad Sci U S A* 1976; **73**(10): 3628-3632.
45. Janeway CA, Jr., Travers P, Walport M & Shlomchik MJ. *Immunobiology*, 5th Edition edn. Garland Publishing: New York, NY, 2001.
46. Macpherson AJ, McCoy KD, Johansen FE & Brandtzaeg P. The immune geography of IgA induction and function. *Mucosal Immunol* 2008; **1**(1): 11-22.
47. Fagarasan S & Honjo T. Intestinal IgA synthesis: regulation of front-line body defences. *Nat Rev Immunol* 2003; **3**(1): 63-72.
48. Moore C, Sauma D, Reyes PA, Morales J, Roseblatt M, Bono MR *et al.* Dendritic cells and B cells cooperate in the generation of CD4(+)CD25(+)FOXP3(+) allogeneic T cells. *Transplant Proc*; **42**(1): 371-375.
49. James SP, Fiocchi C, Graeff AS & Strober W. Phenotypic analysis of lamina propria lymphocytes. Predominance of helper-inducer and cytolytic T-cell phenotypes and deficiency of suppressor-inducer phenotypes in Crohn's disease and control patients. *Gastroenterology* 1986; **91**(6): 1483-1489.
50. Matsuura T, West GA, Youngman KR, Klein JS & Fiocchi C. Immune activation genes in inflammatory bowel disease. *Gastroenterology* 1993; **104**(2): 448-458.
51. Zhu J & Paul WE. CD4 T cells: fates, functions, and faults. *Blood* 2008; **112**(5): 1557-1569.
52. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA & Coffman RL. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol* 1986; **136**(7): 2348-2357.
53. Ansel KM, Djuretic I, Tanasa B & Rao A. Regulation of Th2 differentiation and Il4 locus accessibility. *Annu Rev Immunol* 2006; **24**: 607-656.
54. Szabo SJ, Sullivan BM, Peng SL & Glimcher LH. Molecular mechanisms regulating Th1 immune responses. *Annu Rev Immunol* 2003; **21**: 713-758.
55. Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, Murphy KM *et al.* Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* 2005; **6**(11): 1123-1132.

56. Park H, Li Z, Yang XO, Chang SH, Nurieva R, Wang YH *et al.* A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* 2005; **6**(11): 1133-1141.
57. Chen Y, Inobe J, Kuchroo VK, Baron JL, Janeway CA, Jr. & Weiner HL. Oral tolerance in myelin basic protein T-cell receptor transgenic mice: suppression of autoimmune encephalomyelitis and dose-dependent induction of regulatory cells. *Proc Natl Acad Sci U S A* 1996; **93**(1): 388-391.
58. Masopust D, Jiang J, Shen H & Lefrancois L. Direct analysis of the dynamics of the intestinal mucosa CD8 T cell response to systemic virus infection. *J Immunol* 2001; **166**(4): 2348-2356.
59. Strobel S & Mowat AM. Immune responses to dietary antigens: oral tolerance. *Immunol Today* 1998; **19**(4): 173-181.
60. Griesemer AD, Sorenson EC & Hardy MA. The role of the thymus in tolerance. *Transplantation* 2010; **90**(5): 465-474.
61. Quill H & Schwartz RH. Stimulation of normal inducer T cell clones with antigen presented by purified Ia molecules in planar lipid membranes: specific induction of a long-lived state of proliferative nonresponsiveness. *J Immunol* 1987; **138**(11): 3704-3712.
62. Alderson MR, Tough TW, Davis-Smith T, Braddy S, Falk B, Schooley KA *et al.* Fas ligand mediates activation-induced cell death in human T lymphocytes. *J Exp Med* 1995; **181**(1): 71-77.
63. De Maria R, Boirivant M, Cifone MG, Roncaioli P, Hahne M, Tschopp J *et al.* Functional expression of Fas and Fas ligand on human gut lamina propria T lymphocytes. A potential role for the acidic sphingomyelinase pathway in normal immunoregulation. *J Clin Invest* 1996; **97**(2): 316-322.
64. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004; **126**(6): 1504-1517.
65. Ahuja V & Tandon RK. Inflammatory bowel disease in the Asia-Pacific area: a comparison with developed countries and regional differences. *J Dig Dis*; **11**(3): 134-147.

66. Yang H, Taylor KD & Rotter JI. Inflammatory bowel disease. I. Genetic epidemiology. *Mol Genet Metab* 2001; **74**(1-2): 1-21.
67. Kurata JH, Kantor-Fish S, Frankl H, Godby P & Vadheim CM. Crohn's disease among ethnic groups in a large health maintenance organization. *Gastroenterology* 1992; **102**(6): 1940-1948.
68. Sonnenberg A. Age distribution of IBD hospitalization. *Inflamm Bowel Dis*; **16**(3): 452-457.
69. Johnston RD & Logan RF. What is the peak age for onset of IBD? *Inflamm Bowel Dis* 2008; **14 Suppl 2**: S4-5.
70. Loftus EV, Jr., Silverstein MD, Sandborn WJ, Tremaine WJ, Harmsen WS & Zinsmeister AR. Crohn's disease in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival. *Gastroenterology* 1998; **114**(6): 1161-1168.
71. Baumgart DC & Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007; **369**(9573): 1641-1657.
72. Sands BE. CHAPTER 108 – Crohn's Disease *Feldman: Sleisenger & Fordtran's Gastrointestinal and Liver Disease, 8th ed.* Elsevier Health Sciences, 2006.
73. Cheifetz AD, Moss AC & Peppercorn MA (2008). Crohn's Disease. In: *Knol*. <http://knol.google.com/k/crohn-s-disease#>.
74. Friedman S & Blumberg R (2008). Chapter 289. Inflammatory Bowel Disease. In: *Harrison's Principles of Internal Medicine*.
75. Larsen S, Bendtzen K & Nielsen OH. Extraintestinal manifestations of inflammatory bowel disease: epidemiology, diagnosis, and management. *Ann Med* 2010; **42**(2): 97-114.
76. Eaden JA, Abrams KR & Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001; **48**(4): 526-535.
77. Heitman SJ, Ronksley PE, Hilsden RJ, Manns BJ, Rostom A & Hemmelgarn BR. Prevalence of adenomas and colorectal cancer in average risk individuals: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2009; **7**(12): 1272-1278.

78. Laukoetter MG, Mennigen R, Hannig CM, Osada N, Rijcken E, Vowinkel T *et al.* Intestinal cancer risk in Crohn's disease: a meta-analysis. *J Gastrointest Surg* 2011; **15**(4): 576-583.
79. Kraus S & Arber N. Inflammation and colorectal cancer. *Curr Opin Pharmacol* 2009; **9**(4): 405-410.
80. Pedersen N, Duricova D, Elkjaer M, Gamborg M, Munkholm P & Jess T. Risk of extra-intestinal cancer in inflammatory bowel disease: meta-analysis of population-based cohort studies. *Am J Gastroenterol* 2010; **105**(7): 1480-1487.
81. Su C & Lichtenstein GR. CHAPTER 109 – Ulcerative Colitis *Feldman: Sleisenger & Fordtran's Gastrointestinal and Liver Disease, 8th ed.* Elsevier Health Sciences, 2006.
82. Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002; **347**(6): 417-429.
83. Jakobovits J & Schuster MM. Metronidazole therapy for Crohn's disease and associated fistulae. *Am J Gastroenterol* 1984; **79**(7): 533-540.
84. Targan SR, Hanauer SB, van Deventer SJ, Mayer L, Present DH, Braakman T *et al.* A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. *N Engl J Med* 1997; **337**(15): 1029-1035.
85. Clark M, Colombel JF, Feagan BC, Fedorak RN, Hanauer SB, Kamm MA *et al.* American gastroenterological association consensus development conference on the use of biologics in the treatment of inflammatory bowel disease, June 21-23, 2006. *Gastroenterology* 2007; **133**(1): 312-339.
86. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 2002; **359**(9317): 1541-1549.
87. Sands BE, Anderson FH, Bernstein CN, Chey WY, Feagan BG, Fedorak RN *et al.* Infliximab maintenance therapy for fistulizing Crohn's disease. *N Engl J Med* 2004; **350**(9): 876-885.
88. Lichtenstein GR, Bala M, Han C, DeWoody K & Schaible T. Infliximab improves quality of life in patients with Crohn's disease. *Inflamm Bowel Dis* 2002; **8**(4): 237-243.

89. Keane J, Gershon S, Wise RP, Mirabile-Levens E, Kasznica J, Schwiertman WD *et al.* Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med* 2001; **345**(15): 1098-1104.
90. Reinisch W, de Villiers W, Bene L, Simon L, Racz I, Katz S *et al.* Fontolizumab in moderate to severe Crohn's disease: a phase 2, randomized, double-blind, placebo-controlled, multiple-dose study. *Inflamm Bowel Dis* 2010; **16**(2): 233-242.
91. Mannon PJ, Fuss IJ, Mayer L, Elson CO, Sandborn WJ, Present D *et al.* Anti-interleukin-12 antibody for active Crohn's disease. *N Engl J Med* 2004; **351**(20): 2069-2079.
92. Plevy S, Salzberg B, Van Assche G, Regueiro M, Hommes D, Sandborn W *et al.* A phase I study of visilizumab, a humanized anti-CD3 monoclonal antibody, in severe steroid-refractory ulcerative colitis. *Gastroenterology* 2007; **133**(5): 1414-1422.
93. Baumgart DC, Targan SR, Dignass AU, Mayer L, van Assche G, Hommes DW *et al.* Prospective randomized open-label multicenter phase I/II dose escalation trial of visilizumab (HuM291) in severe steroid-refractory ulcerative colitis. *Inflamm Bowel Dis* 2010; **16**(4): 620-629.
94. Creed TJ, Norman MR, Probert CS, Harvey RF, Shaw IS, Smithson J *et al.* Basiliximab (anti-CD25) in combination with steroids may be an effective new treatment for steroid-resistant ulcerative colitis. *Aliment Pharmacol Ther* 2003; **18**(1): 65-75.
95. Van Assche G, Sandborn WJ, Feagan BG, Salzberg BA, Silvers D, Monroe PS *et al.* Daclizumab, a humanised monoclonal antibody to the interleukin 2 receptor (CD25), for the treatment of moderately to severely active ulcerative colitis: a randomised, double blind, placebo controlled, dose ranging trial. *Gut* 2006; **55**(11): 1568-1574.
96. Ghosh S, Goldin E, Gordon FH, Malchow HA, Rask-Madsen J, Rutgeerts P *et al.* Natalizumab for active Crohn's disease. *N Engl J Med* 2003; **348**(1): 24-32.
97. Gordon FH, Lai CW, Hamilton MI, Allison MC, Srivastava ED, Fouweather MG *et al.* A randomized placebo-controlled trial of a humanized monoclonal antibody to alpha4 integrin in active Crohn's disease. *Gastroenterology* 2001; **121**(2): 268-274.

98. Strong SA, Koltun WA, Hyman NH & Buie WD. Practice parameters for the surgical management of Crohn's disease. *Dis Colon Rectum* 2007; **50**(11): 1735-1746.
99. Rutgeerts P, Geboes K, Vantrappen G, Beyls J, Kerremans R & Hiele M. Predictability of the postoperative course of Crohn's disease. *Gastroenterology* 1990; **99**(4): 956-963.
100. Munkholm P, Langholz E, Hollander D, Thornberg K, Orholm M, Katz KD *et al.* Intestinal permeability in patients with Crohn's disease and ulcerative colitis and their first degree relatives. *Gut* 1994; **35**(1): 68-72.
101. Soderholm JD, Olaison G, Peterson KH, Franzen LE, Lindmark T, Wiren M *et al.* Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease. *Gut* 2002; **50**(3): 307-313.
102. Okayasu I, Hatakeyama S, Yamada M, Ohkusa T, Inagaki Y & Nakaya R. A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* 1990; **98**(3): 694-702.
103. Mahler M, Bristol IJ, Leiter EH, Workman AE, Birkenmeier EH, Elson CO *et al.* Differential susceptibility of inbred mouse strains to dextran sulfate sodium-induced colitis. *Am J Physiol* 1998; **274**(3 Pt 1): G544-551.
104. Dieleman LA, Ridwan BU, Tennyson GS, Beagley KW, Bucy RP & Elson CO. Dextran sulfate sodium-induced colitis occurs in severe combined immunodeficient mice. *Gastroenterology* 1994; **107**(6): 1643-1652.
105. Fukata M, Michelsen KS, Eri R, Thomas LS, Hu B, Lukasek K *et al.* Toll-like receptor-4 is required for intestinal response to epithelial injury and limiting bacterial translocation in a murine model of acute colitis. *Am J Physiol Gastrointest Liver Physiol* 2005; **288**(5): G1055-1065.
106. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S & Medzhitov R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 2004; **118**(2): 229-241.
107. Rugtveit J, Brandtzaeg P, Halstensen TS, Fausa O & Scott H. Increased macrophage subset in inflammatory bowel disease: apparent recruitment from peripheral blood monocytes. *Gut* 1994; **35**(5): 669-674.

108. Rugtveit J, Nilsen EM, Bakka A, Carlsen H, Brandtzaeg P & Scott H. Cytokine profiles differ in newly recruited and resident subsets of mucosal macrophages from inflammatory bowel disease. *Gastroenterology* 1997; **112**(5): 1493-1505.
109. Kamada N, Hisamatsu T, Okamoto S, Chinen H, Kobayashi T, Sato T *et al.* Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-gamma axis. *J Clin Invest* 2008; **118**(6): 2269-2280.
110. Hart AL, Al-Hassi HO, Rigby RJ, Bell SJ, Emmanuel AV, Knight SC *et al.* Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology* 2005; **129**(1): 50-65.
111. Watanabe T, Kitani A, Murray PJ & Strober W. NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol* 2004; **5**(8): 800-808.
112. Watanabe T, Asano N, Murray PJ, Ozato K, Tailor P, Fuss IJ *et al.* Muramyl dipeptide activation of nucleotide-binding oligomerization domain 2 protects mice from experimental colitis. *J Clin Invest* 2008; **118**(2): 545-559.
113. MacDermott RP, Nash GS, Bertovich MJ, Mohrman RF, Kodner IJ, Delacroix DL *et al.* Altered patterns of secretion of monomeric IgA and IgA subclass 1 by intestinal mononuclear cells in inflammatory bowel disease. *Gastroenterology* 1986; **91**(2): 379-385.
114. Scott MG, Nahm MH, Macke K, Nash GS, Bertovich MJ & MacDermott RP. Spontaneous secretion of IgG subclasses by intestinal mononuclear cells: differences between ulcerative colitis, Crohn's disease, and controls. *Clin Exp Immunol* 1986; **66**(1): 209-215.
115. Rotteveel FT, Kokkelink I, van Lier RA, Kuenen B, Meager A, Miedema F *et al.* Clonal analysis of functionally distinct human CD4+ T cell subsets. *J Exp Med* 1988; **168**(5): 1659-1673.
116. Al-Chalabi A, Andersen PM, Chioza B, Shaw C, Sham PC, Robberecht W *et al.* Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. *Hum Mol Genet* 1998; **7**(13): 2045-2050.
117. Parronchi P, Romagnani P, Annunziato F, Sampognaro S, Becchio A, Giannarini L *et al.* Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn's disease. *Am J Pathol* 1997; **150**(3): 823-832.

118. Monteleone G, Biancone L, Marasco R, Morrone G, Marasco O, Lizza F *et al.* Interleukin 12 is expressed and actively released by Crohn's disease intestinal lamina propria mononuclear cells. *Gastroenterology* 1997; **112**(4): 1169-1178.
119. Annunziato F, Cosmi L, Santarlasci V, Maggi L, Liotta F, Mazzinghi B *et al.* Phenotypic and functional features of human Th17 cells. *J Exp Med* 2007; **204**(8): 1849-1861.
120. Fujino S, Andoh A, Bamba S, Ogawa A, Hata K, Araki Y *et al.* Increased expression of interleukin 17 in inflammatory bowel disease. *Gut* 2003; **52**(1): 65-70.
121. Monteleone G, Monteleone I, Fina D, Vavassori P, Del Vecchio Blanco G, Caruso R *et al.* Interleukin-21 enhances T-helper cell type I signaling and interferon-gamma production in Crohn's disease. *Gastroenterology* 2005; **128**(3): 687-694.
122. Fuss IJ, Heller F, Boirivant M, Leon F, Yoshida M, Fichtner-Feigl S *et al.* Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J Clin Invest* 2004; **113**(10): 1490-1497.
123. Makita S, Kanai T, Oshima S, Uraushihara K, Totsuka T, Sawada T *et al.* CD4+CD25bright T cells in human intestinal lamina propria as regulatory cells. *J Immunol* 2004; **173**(5): 3119-3130.
124. Maul J, Loddenkemper C, Mundt P, Berg E, Giese T, Stallmach A *et al.* Peripheral and intestinal regulatory CD4+ CD25(high) T cells in inflammatory bowel disease. *Gastroenterology* 2005; **128**(7): 1868-1878.
125. Read S, Malmstrom V & Powrie F. Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation. *J Exp Med* 2000; **192**(2): 295-302.
126. Asseman C, Mauze S, Leach MW, Coffman RL & Powrie F. An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. *J Exp Med* 1999; **190**(7): 995-1004.
127. Orholm M, Binder V, Sorensen TI, Rasmussen LP & Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 2000; **35**(10): 1075-1081.

128. Halfvarson J, Bodin L, Tysk C, Lindberg E & Jarnerot G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003; **124**(7): 1767-1773.
129. Calkins BM. A meta-analysis of the role of smoking in inflammatory bowel disease. *Dig Dis Sci* 1989; **34**(12): 1841-1854.
130. Boyko EJ, Perera DR, Koepsell TD, Keane EM & Inui TS. Effects of cigarette smoking on the clinical course of ulcerative colitis. *Scand J Gastroenterol* 1988; **23**(9): 1147-1152.
131. Pullan RD, Rhodes J, Ganesh S, Mani V, Morris JS, Williams GT *et al*. Transdermal nicotine for active ulcerative colitis. *N Engl J Med* 1994; **330**(12): 811-815.
132. Cosnes J. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. *Best Pract Res Clin Gastroenterol* 2004; **18**(3): 481-496.
133. Otterbein LE, Bach FH, Alam J, Soares M, Tao Lu H, Wysk M *et al*. Carbon monoxide has anti-inflammatory effects involving the mitogen-activated protein kinase pathway. *Nat Med* 2000; **6**(4): 422-428.
134. King TE, Jr., Savici D & Campbell PA. Phagocytosis and killing of *Listeria monocytogenes* by alveolar macrophages: smokers versus nonsmokers. *J Infect Dis* 1988; **158**(6): 1309-1316.
135. Sher ME, Bank S, Greenberg R, Sardinha TC, Weissman S, Bailey B *et al*. The influence of cigarette smoking on cytokine levels in patients with inflammatory bowel disease. *Inflamm Bowel Dis* 1999; **5**(2): 73-78.
136. Zijlstra FJ, Srivastava ED, Rhodes M, van Dijk AP, Fogg F, Samson HJ *et al*. Effect of nicotine on rectal mucus and mucosal eicosanoids. *Gut* 1994; **35**(2): 247-251.
137. Kalra J, Chaudhary AK & Prasad K. Increased production of oxygen free radicals in cigarette smokers. *Int J Exp Pathol* 1991; **72**(1): 1-7.
138. Guo X, Wang WP, Ko JK & Cho CH. Involvement of neutrophils and free radicals in the potentiating effects of passive cigarette smoking on inflammatory bowel disease in rats. *Gastroenterology* 1999; **117**(4): 884-892.

139. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**(6837): 599-603.
140. Koutroubakis IE, Vlachonikolis IG & Kouroumalis EA. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: a critical review. *Inflamm Bowel Dis* 2002; **8**(4): 277-286.
141. Andersson RE, Olaison G, Tysk C & Ekblom A. Appendectomy and protection against ulcerative colitis. *N Engl J Med* 2001; **344**(11): 808-814.
142. Frisch M, Pedersen BV & Andersson RE. Appendicitis, mesenteric lymphadenitis, and subsequent risk of ulcerative colitis: cohort studies in Sweden and Denmark. *BMJ* 2009; **338**: b716.
143. Cosnes J, Carbonnel F, Beaugerie L, Blain A, Reijasse D & Gendre JP. Effects of appendectomy on the course of ulcerative colitis. *Gut* 2002; **51**(6): 803-807.
144. Kaplan GG, Jackson T, Sands BE, Frisch M, Andersson RE & Korzenik J. The risk of developing Crohn's disease after an appendectomy: a meta-analysis. *Am J Gastroenterol* 2008; **103**(11): 2925-2931.
145. Strachan DP. Hay fever, hygiene, and household size. *BMJ* 1989; **299**(6710): 1259-1260.
146. Barton JR, Gillon S & Ferguson A. Incidence of inflammatory bowel disease in Scottish children between 1968 and 1983; marginal fall in ulcerative colitis, three-fold rise in Crohn's disease. *Gut* 1989; **30**(5): 618-622.
147. Koloski NA, Bret L & Radford-Smith G. Hygiene hypothesis in inflammatory bowel disease: a critical review of the literature. *World J Gastroenterol* 2008; **14**(2): 165-173.
148. Bernstein CN, Rawsthorne P, Cheang M & Blanchard JF. A population-based case control study of potential risk factors for IBD. *Am J Gastroenterol* 2006; **101**(5): 993-1002.
149. Amre DK, Lambrette P, Law L, Krupoves A, Chotard V, Costea F *et al.* Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: a case-control study. *Am J Gastroenterol* 2006; **101**(5): 1005-1011.
150. Binder V. Genetic epidemiology in inflammatory bowel disease. *Dig Dis* 1998; **16**(6): 351-355.

151. Halfvarson J. Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm Bowel Dis* 2011; **17**(1): 6-12.
152. Bayless TM, Tokayer AZ, Polito JM, 2nd, Quaskey SA, Mellits ED & Harris ML. Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences. *Gastroenterology* 1996; **111**(3): 573-579.
153. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen TI & Binder V. Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991; **324**(2): 84-88.
154. Probert CS, Jayanthi V, Hughes AO, Thompson JR, Wicks AC & Mayberry JF. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* 1993; **34**(11): 1547-1551.
155. Taramelli R & Acquati F. The human genome project and the discovery of genetic determinants of cancer susceptibility. *Eur J Cancer* 2004; **40**(17): 2537-2543.
156. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
157. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG *et al.* The sequence of the human genome. *Science* 2001; **291**(5507): 1304-1351.
158. Collins FS, Morgan M & Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science (New York, NY)* 2003; **300**(5617): 286-290.
159. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* 2007; **104**(49): 19428-19433.
160. Richard GF, Kerrest A & Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 2008; **72**(4): 686-727.
161. Brinkmann B, Klintschar M, Neuhuber F, Huhne J & Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American journal of human genetics* 1998; **62**(6): 1408-1415.

162. Li YC, Korol AB, Fahima T, Beiles A & Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 2002; **11**(12): 2453-2465.
163. Legendre M, Pochet N, Pak T & Verstrepen KJ. Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome research* 2007; **17**(12): 1787-1796.
164. Verstrepen KJ, Jansen A, Lewitter F & Fink GR. Intragenic tandem repeats generate functional variability. *Nature genetics* 2005; **37**(9): 986-990.
165. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 1991; **65**(5): 905-914.
166. La Spada AR, Wilson EM, Lubahn DB, Harding AE & Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991; **352**(6330): 77-79.
167. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* 1993; **72**(6): 971-983.
168. Altshuler D, Daly MJ & Lander ES. Genetic mapping in human disease. *Science* 2008; **322**(5903): 881-888.
169. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**(6822): 928-933.
170. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ & Lander ES. High-resolution haplotype structure in the human genome. *Nature genetics* 2001; **29**(2): 229-232.
171. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; **296**(5576): 2225-2229.
172. HapMap. A haplotype map of the human genome. *Nature* 2005; **437**(7063): 1299-1320.
173. HapMap. The International HapMap Project. *Nature* 2003; **426**(6968): 789-796.

174. Goyette P, Labbe C, Trinh TT, Xavier RJ & Rioux JD. Molecular pathogenesis of inflammatory bowel disease: genotypes, phenotypes and personalized medicine. *Ann Med* 2007; **39**(3): 177-199.
175. Sturtevant AH. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* 1913; **14**(1): 43-59.
176. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983; **306**(5940): 234-238.
177. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K *et al.* A genetic linkage map of the human genome. *Cell* 1987; **51**(2): 319-337.
178. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; **380**(6570): 152-154.
179. Wang WY, Barratt BJ, Clayton DG & Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nature reviews* 2005; **6**(2): 109-118.
180. Suh Y & Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. *Mutation research* 2005; **573**(1-2): 41-53.
181. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *American journal of human genetics* 2000; **66**(6): 1863-1870.
182. Lander E & Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature genetics* 1995; **11**(3): 241-247.
183. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996; **379**(6568): 821-823.
184. Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K *et al.* Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nature genetics* 1996; **14**(2): 199-202.

185. Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, Macpherson AJ *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *American journal of human genetics* 1999; **64**(3): 808-816.
186. Ma Y, Ohmen JD, Li Z, Bentley LG, McElree C, Pressman S *et al.* A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999; **5**(4): 271-278.
187. Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci U S A* 1998; **95**(13): 7502-7507.
188. Hampe J, Frenzel H, Mirza MM, Croucher PJ, Cuthbert A, Mascheretti S *et al.* Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc Natl Acad Sci U S A* 2002; **99**(1): 321-326.
189. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; **357**(9272): 1925-1928.
190. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**(6837): 603-606.
191. Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF *et al.* Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* 2005; **307**(5710): 734-738.
192. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature genetics* 2001; **29**(2): 223-228.
193. Tein I. Carnitine transport: pathophysiology and metabolism of known molecular defects. *J Inherit Metab Dis* 2003; **26**(2-3): 147-169.
194. Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nature genetics* 2004; **36**(5): 471-475.

195. Noble CL, Nimmo ER, Drummond H, Ho GT, Tenesa A, Smith L *et al.* The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 2005; **129**(6): 1854-1864.
196. Silverberg MS, Duerr RH, Brant SR, Bromfield G, Datta LW, Jani N *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur J Hum Genet* 2007; **15**(3): 328-335.
197. Hampe J, Shaw SH, Saiz R, Leysens N, Lantermann A, Mascheretti S *et al.* Linkage of inflammatory bowel disease to human chromosome 6p. *American journal of human genetics* 1999; **65**(6): 1647-1655.
198. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD & Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 2004; **13**(7): 763-770.
199. Williams CN, Kocher K, Lander ES, Daly MJ & Rioux JD. Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. *Inflamm Bowel Dis* 2002; **8**(6): 375-381.
200. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 1999; **401**(6756): 921-923.
201. Ahmad T, Marshall SE & Jewell D. Genetics of inflammatory bowel disease: the role of the HLA complex. *World J Gastroenterol* 2006; **12**(23): 3628-3635.
202. Cho JH & Brant SR. Recent insights into the genetics of inflammatory bowel disease. *Gastroenterology* 2011; **140**(6): 1704-1712.
203. Risch N & Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**(5281): 1516-1517.
204. Whitcomb DC, Aoun E, Vodovotz Y, Clermont G & Barmada MM. Evaluating disorders with a complex genetics basis. the future roles of meta-analysis and systems biology. *Digestive diseases and sciences* 2005; **50**(12): 2195-2202.
205. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics* 2011; **43**(3): 246-252.
206. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics* 2011; **42**(12): 1118-1125.

207. Cardon LR & Bell JI. Association study designs for complex diseases. *Nature reviews* 2001; **2**(2): 91-99.
208. Hirschhorn JN & Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature reviews* 2005; **6**(2): 95-108.
209. Lohmueller KE, Pearce CL, Pike M, Lander ES & Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature genetics* 2003; **33**(2): 177-182.
210. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet* 2005; **14**(22): 3499-3506.
211. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics* 2008; **40**(8): 955-962.
212. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**(7311): 52-58.
213. The_1000genomes_Project_Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-1073.
214. Nejentsev S, Walker N, Riches D, Egholm M & Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009; **324**(5925): 387-389.
215. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nature genetics* 2006; **38**(6): 617-619.
216. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 2006; **441**(7089): 101-105.
217. Hyoty H & Taylor KW. The role of viruses in human diabetes. *Diabetologia* 2002; **45**(10): 1353-1361.

218. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**(5804): 1461-1463.
219. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nature genetics* 2008; **40**(11): 1319-1323.
220. Maloy KJ. The Interleukin-23 / Interleukin-17 axis in intestinal inflammation. *J Intern Med* 2008; **263**(6): 584-590.
221. Holtta V, Klemetti P, Sipponen T, Westerholm-Ormio M, Kociubinski G, Salo H *et al.* IL-23/IL-17 immunity as a hallmark of Crohn's disease. *Inflamm Bowel Dis* 2008; **14**(9): 1175-1184.
222. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature genetics* 2007; **39**(2): 207-211.
223. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics* 2007; **39**(5): 596-604.
224. Travassos LH, Carneiro LA, Ramjeet M, Hussey S, Kim YG, Magalhaes JG *et al.* Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nat Immunol* 2009; **11**(1): 55-62.
225. Rabinowitz JD & White E. Autophagy and metabolism. *Science* 2010; **330**(6009): 1344-1348.
226. Mizushima N, Yamamoto A, Matsui M, Yoshimori T & Ohsumi Y. In vivo analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol Biol Cell* 2004; **15**(3): 1101-1111.
227. Mizushima N, Kuma A, Kobayashi Y, Yamamoto A, Matsubae M, Takao T *et al.* Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *Journal of cell science* 2003; **116**(Pt 9): 1679-1688.

228. Saitoh T, Fujita N, Jang MH, Uematsu S, Yang BG, Satoh T *et al.* Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1beta production. *Nature* 2008.
229. Cooney R, Baker J, Brain O, Danis B, Pichulik T, Allan P *et al.* NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med*; **16**(1): 90-97.
230. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature genetics* 2007; **39**(7): 830-832.
231. Cadwell K. Crohn's Disease Susceptibility Gene Interactions, a NOD to the Newcomer ATG16L1. *Gastroenterology* 2010; **139**(5): 1448-1450.
232. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J *et al.* The DNA sequence and biology of human chromosome 19. *Nature* 2004; **428**(6982): 529-535.
233. Eichler EE, Hoffman SM, Adamson AA, Gordon LA, McCready P, Lamerdin JE *et al.* Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome research* 1998; **8**(8): 791-808.
234. Kalsotra A & Strobel HW. Cytochrome P450 4F subfamily: at the crossroads of eicosanoid and drug metabolism. *Pharmacol Ther* 2006; **112**(3): 589-611.
235. Tello-Ruiz MK, Curley C, DelMonte T, Giallourakis C, Kirby A, Miller K *et al.* Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19. *Eur J Hum Genet* 2006; **14**(6): 780-790.
236. Braun C, Zahn R, Martin K, Albert E & Folwaczny C. Polymorphisms of the ICAM-1 gene are associated with inflammatory bowel disease, regardless of the p-ANCA status. *Clin Immunol* 2001; **101**(3): 357-360.
237. Elmgreen J, Sorensen H & Berkowicz A. Polymorphism of complement C3 in chronic inflammatory bowel disease. Predominance of the C3F gene in Crohn's disease. *Acta Med Scand* 1984; **215**(4): 375-378.
238. Low JH, Williams FA, Yang X, Cullen S, Colley J, Ling KL *et al.* Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. *Inflamm Bowel Dis* 2004; **10**(3): 173-181.

239. Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E *et al.* Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nature genetics* 2005; **37**(12): 1341-1344.
240. Lionetti E & Catassi C. New clues in celiac disease epidemiology, pathogenesis, clinical manifestations, and treatment. *Int Rev Immunol* 2011; **30**(4): 219-231.
241. Muller RT, Honnert U, Reinhard J & Bahler M. The rat myosin myr 5 is a GTPase-activating protein for Rho in vivo: essential role of arginine 1695. *Mol Biol Cell* 1997; **8**(10): 2039-2053.
242. Wirth JA, Jensen KA, Post PL, Bement WM & Mooseker MS. Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail. *Journal of cell science* 1996; **109** (Pt 3): 653-661.
243. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM *et al.* Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 2006; **131**(6): 1768-1774.

RÉSULTATS

Chapitre 4

MAST3: facteur de risque aux maladies inflammatoires de l'intestin

Dans la première phase du projet, le but principal était d'identifier un ou des facteurs de risque aux maladies inflammatoires de l'intestin dans la région IBD6 ou 19p. Depuis l'identification de la région en 2000¹, d'autres groupes ont répliqué la liaison avec les MII ce qui suggère que la région contient un ou des facteurs de risque généraux à ces maladies. Les études par gènes candidats n'ont pas identifié de variante avec un effet assez important pour être responsable du pic de liaison identifié dans l'étude originale.

L'article suivant décrit notre étude de cartographie d'association de la région IBD6 qui a mené à l'identification du gène *MAST3* comme gène de susceptibilité aux MII. En résumé notre étude démontre:

- Un criblage de la région IBD6 suivi d'une étape de répllication
- L'identification d'un SNP associé localisé dans le second intron du gène *MAST3*
- L'association subséquente de quatre SNP codant incluant un SNP non-synonyme (Gly861Ser)
- Des analyses de déséquilibre de liaison et de corrélation entre les SNP qui excluent clairement les autres gènes situés dans cette région

De plus, nous avons étudié le contexte biologique et la fonction du gène *MAST3* et démontré que :

- *MAST3* était exprimé dans plusieurs types de cellules immunitaires (humaines et murines, lignées cellulaires et cellules primaires)
- *MAST3* est impliqué dans l'activation de NF- κ B dépendante des TLR4 dans des cellules stimulées au LPS.

Cet article démontre qu'une étude d'association ciblée à une région confirmée de liaison représente une stratégie gagnante pour identifier des gènes associés. De plus, cet article

place le gène causal dans un contexte biologique qui supporte les découvertes récentes des autres gènes de susceptibilité aux MII et qui implique la réponse immunitaire innée à des produits bactériens tel que le LPS dans le développement de ces maladies.

ARTICLE 1

MAST3: a Novel IBD Risk Factor that Modulates TLR4 Signaling

Catherine Labbé^{1,2}, Philippe Goyette¹, Céline Lefebvre¹, Christine Stevens³, Todd Green³, Marcela K. Tello-Ruiz³, Zhifang Cao⁴, Aimee L. Landry⁴, Joanne Stempak⁵, Vito Annese⁶, Anna Latiano⁶, Steven R. Brant⁷, Richard H. Duerr⁸, Kent D. Taylor⁹, Judy H. Cho¹⁰, A. Hillary Steinhart⁵, Mark J. Daly^{3,4}, Mark S. Silverberg⁵, Ramnik J. Xavier⁴, John D. Rioux^{1,2,3}.

- 1) Institut de Cardiologie de Montréal, Montréal, Qc, Canada;
- 2) Département de médecine, Université de Montréal, Montréal, Qc, Canada;
- 3) The Broad Institute of MIT and Harvard, Cambridge, MA, USA;
- 4) Center for Computational and Integrative biology and Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA;
- 5) Mount Sinai Hospital IBD Center, University of Toronto, Toronto, On, Canada;
- 6) Gastrointestinal & Endoscopy Units, I.R.C.C.S. Hospital, San Giovanni Rotondo, Italy;
- 7) Meyerhoff Inflammatory Bowel Disease Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA;
- 8) School of Medicine Division of Gastroenterology, Hepatology and Nutrition, University of Pittsburgh, Pittsburgh, PA, USA;
- 9) Medical Genetics Institute and IBD Center, Cedars-Sinai Medical Center, Los Angeles, CA, USA;
- 10) IBD Center, Division of Gastroenterology, Department of Medicine, Yale University, New Haven, CT, USA.

Running Head: Association Study Identifies MAST3 in IBD

Keywords: MAST3, Inflammatory Bowel Disease, TLR4, NF-κB

Cet article est d'abord paru dans *Genes and Immunity* : Labbé C, Goyette P, Lefebvre C, Stevens C, Green T, Tello-Ruiz MK, Cao Z, Landry AL, Stempak J, Annese V, Latiano A, Brant SR, Duerr RH, Taylor KD, Cho JH, Steinhart AH, Daly MJ, Silverberg MS, Xavier RJ, Rioux JD. *Genes and Immunity* (2008) **9**, 602–612; doi:10.1038/gene.2008.57; published online 24 July 2008; © Macmillan Publishers Ltd. L'article est inclus dans ma thèse avec la permission de l'éditeur et de tous les auteurs.

Contribution des auteurs

Pour cet article, j'ai effectué 80% des travaux incluant la sélection et le génotypage des SNP de l'étape de réplication, l'analyse des résultats de génotypage des étapes de criblage et de réplication et la rédaction de l'article. Annese, Latiano, Brant, Duerr, Taylor, Cho, Steinhart et Silverberg sont des chercheurs collaborateurs qui ont fourni des échantillons d'ADN. Stempak est assistante de recherche chez Silverberg et elle a participé à la collecte d'échantillon. Daly, Green, Stevens et Tello-Ruiz ont participé à la sélection et au génotypage des SNP de l'étape de criblage sur la plate-forme Illumina. Landry et Cao ont effectué les essais fonctionnels dans le laboratoire de Xavier. Goyette et Lefebvre ont participé au génotypage sur la plateforme Sequenom. Rioux est l'investigateur principal, il a participé à la conception de l'étude, a fourni des échantillons d'ADN et a participé aux analyses et à la rédaction de l'article.

Abstract

Inflammatory bowel disease (IBD) is a chronic disorder caused by multiple factors in a genetically susceptible host. Significant advances in the study of genetic susceptibility have highlighted the importance of the innate immune system in this disease. We previously completed a genomewide linkage study and found a significant locus (IBD6) on chromosome 19p. We were interested in identifying the causal variant in IBD6. We performed a two-stage association mapping study. In stage one, 1530 SNPs were selected from the HapMap database and genotyped in 761 patients with IBD. Among the SNPs that passed the threshold for replication, 26 were successfully genotyped in 754 additional patients (stage two). One intronic variant, rs273506 located in the *MAST3* gene was found to be associated in both stages (pooled $P=2 \times 10^{-4}$). We identified four *MAST3* coding variants, including a non-synonymous SNP rs8108738, correlated to rs273506 and associated to IBD. To test whether *MAST3* was expressed in cells of interest, we performed expression assays which showed abundant expression of *MAST3* in antigen presenting cells and in lymphocytes. The knockdown of *MAST3* specifically decreased TLR4 dependent NF- κ B activity. Our findings are additional proof of the pivotal role played by modulators of NF- κ B activity in IBD pathogenesis.

Introduction

Inflammatory bowel disease (IBD) is a term describing two complex diseases: Crohn's disease (CD) and ulcerative colitis (UC). Both CD and UC are common chronic autoimmune diseases of the gastrointestinal tract but they differ in terms of localization and extent of the inflammation. IBD primarily affects people living in industrialized countries. A recent epidemiological survey of IBD in the US reports prevalence rates of 201 per 100000 and 238 per 100000 for CD and UC, respectively². The precise causes are still unknown but different environmental factors, ranging from geographic location to lifestyle, hygiene and gut flora, have been proposed³. In addition, familial clustering and twin studies have clearly shown that inherited factors contribute to disease predisposition. The first gene discovered to be unequivocally associated with CD was *CARD15*^{4;5}. Shortly after, the IBD5 locus was also associated⁶. In the last year, more than ten new loci have been associated with susceptibility to or protection from IBD⁷⁻¹³. However, the contribution of each of these loci is predicted to be small and thus, all the genes discovered to date do not entirely explain the genetic risk for IBD.

To elucidate the genetic basis for CD and UC, we previously performed a genomewide linkage study that identified a significant linkage region (LOD score 4.6) on chromosome 19p, subsequently designated as the IBD6 locus¹. Stratifying the families based on phenotype (i.e., CD vs UC vs combined IBD) and age of diagnosis (early vs late), we found evidence of linkage in all subgroups; strongly suggesting that the 19p locus is a general risk factor for IBD. Supporting evidence for this locus was also found in two independent genomewide linkage studies^{14;15}. The IBD6 locus extends 28.5Mb and contains 638 known genes, many of which are involved in immunity and inflammation such as *ICAM1*, *ICAM3* and *IL12RB1*¹⁶. Our previous attempts, using candidates gene approaches^{17;18}, to identify the causal gene within this locus resulted in the identification of the variant A1011S in *MYO9B* as a risk factor for UC. However, this A1011S variant could only account for a relative risk to siblings (λ_s) of 1.02, and

unlike the original findings, was subphenotype specific, and therefore could not explain entirely the observed linkage signal.

To further refine the region containing the major IBD risk factor within the IBD6 locus, we used a systematic high density association mapping approach to analyze the variation in this region. To optimize the statistical power, our experiment was designed in two stages. Specifically, in the first stage, we typed 1530 SNPs selected to tag the common variants in the 28.5 Mb region identified in CEU HapMap dataset. Approximately, ~1800 samples (trios and cases/controls) were genotyped in this screening stage. In the second stage we replicated the most associated SNPs from the initial screen in ~2000 independent samples (trios and cases/controls). These two stages led to the identification of a single significantly replicated variant: rs273506 (combined $P= 1.8 \times 10^{-4}$). This variant is located in the intronic region of the microtubule-associated serine/threonine-protein kinase gene 3 (*MAST3*, NM_015016). We subsequently typed the known coding variants within *MAST3* and identified one missense, S861G, and three synonymous mutations that are significantly associated in IBD ($P < 5 \times 10^{-3}$). Combining these association results with the linkage disequilibrium (LD) and correlation patterns in this region indicate that the association signal is limited to a ~60kb region immediately surrounding the *MAST3* gene, suggesting a role for *MAST3* in IBD susceptibility.

Results

Two-stage association study identifies rs273506

Our goal was to identify the gene(s) responsible for the IBD6 linkage peak discovered in previous studies. Since candidate gene studies failed to do so, we embarked upon a comprehensive unbiased two-staged association mapping study to search for the specific gene that is responsible for the IBD risk factor in this region. In the first stage, a modest nominal P-value of 0.02 was used as threshold for significance in order to optimize the chance of identifying the association signal. In a second stage, the SNPs that passed this threshold were genotyped in an independent cohort to separate the true and false positives.

In order to tag the common variation in IBD6, we first selected 1530 SNPs that we genotyped in 761 IBD patients and parental or matched controls (see Table 1). A systematic quality control (QC) analysis of the data was performed resulting in a final dataset of 371 trios and 271 independent cases/203 matched controls. A total of 1238 SNPs with an average genotyping call rate of 98.5% passed the QC thresholds (see Supplementary Table 1). Primary testing was done using the IBD phenotype (CD and UC) because our previous linkage study indicated that the IBD6 locus contains a general IBD risk factor¹. In a secondary analysis, we tested association with the CD phenotype since this subgroup was sufficiently large (see Table 1). We found 26 SNPs associated with IBD and 33 with CD, for a total 37 non-redundant SNPs significantly identified through association testing (see Table 2 and Supplementary Table 3).

We then performed a replication study using 401 trios as well as 353 independent cases and 344 controls (see Table 1). We successfully typed 26 of the top SNPs (see Table 2 and Supplementary Table 2). A single marker, rs273506 had significant association values in both the IBD screening (stage 1) ($P=0.005$) and replication (stage 2) ($P=0.02$) cohorts. To better evaluate the significance and strength of the signal, we merged the

screen and replication cohorts (total number of post-QC samples: 682 trios, 582 independent cases and 505 controls) and obtained a combined P-value of 1.8×10^{-4} and an odds ratio (OR) of 1.23 [95% Confidence Interval (CI): 1.10, 1.38]. The SNP rs273506 is a G/A variant where the A allele has a frequency of 47% (in the stage 1 cohort) and is associated with the IBD phenotype. The rs273506 SNP is located in the second intron of the *MAST3* gene (Figure 1).

MAST3 gene as an IBD susceptibility gene

Given that the association identified in the screen and replicated in an independent cohort was located within *MAST3*, we were interested in testing the known coding variants within this gene. Specifically, we identified all five *MAST3* coding variants found within the CEU HapMap dataset: four synonymous and one non-synonymous. Four of these SNPs are on the same haplotype ($r^2=1$) and are correlated to the associated SNP rs273506 ($r^2>0.7$) in the HapMap dataset. One SNP, rs541225, is not correlated to rs273506 or to the other *MAST3* coding SNP ($r^2<0.1$) and thus was not typed. The four remaining *MAST3* coding SNPs were genotyped in our cohorts (total number of post-QC samples: 610 trios, 1105 independent cases and 713 controls) and the correlation is consistent with the HapMap data. Of the SNPs genotyped, there are three synonymous variants, rs740691 (H174H), rs2270623 (I536I) and rs2072490 (G1045G) and one non-synonymous, rs8108738 (S861G). All coding SNPs are significantly associated ($P<0.005$ see Table 3). Being on the same haplotype, the IBD risk conferred by each of these coding variants, as measured by the OR, is approximately the same (1.18 (95% CI [1.04, 1.34])).

Although this association mapping identified coding variants in *MAST3* that are associated to IBD, it is formally possible that they are simply in linkage disequilibrium (LD) with causal variation outside of the *MAST3* region. Therefore, we examined the LD and correlation patterns in this region (Figure 1 and 2). *MAST3* falls within a region of tight LD containing no other gene (LD block #4 on Figure 1). The significant association results of our screening stage are all located in a 60kb region surrounding rs273506 and implicate *MAST3* and not the surrounding genes (Figure 1 upper panel).

Furthermore, the region defined by all the HapMap SNPs correlated to rs273506 by an r^2 of 0.5 or greater, contains *MAST3* and the *PIK3R2* gene (see the correlation neighbourhood depicted in Figure 2). However, the region representing the SNPs correlated by an r^2 of 0.7 or higher is restricted to the *MAST3* gene, and thus favours the *MAST3* gene over the *PIK3R2* gene, which encodes a regulatory subunit of the phosphatidylinositol 3-kinase (PI3K). In our screening phase we typed SNPs tagging the variation in the *PIK3R2* region and these perfectly tagged ($r^2=1$) two of the three *PIK3R2* SNPs present in the HapMap database, including the only known non-synonymous coding SNP (rs1011320), and none were significant. Taken together, these results indicate that a variant in the *MAST3* gene is responsible for the association signal detected from SNP rs273506 and that the S861G variant is the most likely candidate.

IL12RB1, located 10kb upstream of *MAST3*, encodes one of the subunit of the receptor for IL12 and IL23. *IL23R* has been associated to ileal CD ⁷ and is a subunit of the receptor of a major cytokine involved in the differentiation of T-helper cells into T_H17 cells ¹⁹. We were particularly interested in examining the relationship of *IL12RB1* to the associated variants in *MAST3*. In the screening stage of the present study, we typed two SNPs located in *IL12RB1* but none of them had significant association results.

Altogether the two typed SNPs correlate to 76% of the HapMap intronic and exonic SNPs within *IL12RB1* with an $r^2 > 0.7$ (average $r^2 = 0.975$), suggesting absence of association between *IL12RB1* and IBD. Furthermore, our associated SNP, rs273506 is very weakly correlated to SNPs in this gene. In fact, the highest r^2 does not reach 0.4 (Figure 2). All these findings and the *MAST3* association results point to *MAST3* as the susceptibility gene.

MAST3 belongs to the MAST kinase family. The *MAST3* gene is 53.9 kb long and is encoded by 27 exons. *MAST3* encodes an open reading frame of 1309 amino acids. Alignments of the *MAST3* human amino acid sequence with the mouse ortholog sequence demonstrate 91% identity ²⁰. The *MAST3* protein is a four alpha-helix bundle composed of different domains including a protein kinase domain and a PDZ domain (a

common structural domain of signaling proteins designated by an acronym from the first letter of three proteins: PSD95, DlgA and zo-1) that mediates binding to PTEN (Figure 3)²¹. Four other members of this family are known: MAST1, MAST2, MAST4 and a mast-like protein, MASTL²²⁻²⁴. MAST1, 2, 3 and 4 have similar structure and share the PDZ and kinase domains whereas MASTL is a shorter protein lacking the PDZ domain. The four other members of the MAST family have amino acid sequence identity to MAST3 ranging from 45% for MASTL to 73% for MAST2. However, specific domains have higher identity. The MAST3 protein kinase domain has 89%, 89%, 90% identity with the corresponding MAST1, MAST2 and MAST4 protein kinase domain, respectively. The MAST3 PDZ domain identity is 73%, 78%, 76% with MAST1, MAST2 and MAST4, respectively. The associated synonymous SNP I536I is located within the kinase domain. The non-synonymous SNP S861G is not located within any UniProt predicted domains, but might be located in a yet unknown domain or interfere with protein folding (Figure 3).

MAST3 regulates TLR4-mediated NF- κ B pathway specifically

We next determined the expression of MAST3 in a panel of epithelial and immune cell lines by quantitative real time PCR (qPCR). As shown in Figure 4A, MAST3 is abundantly expressed in human embryonic kidney 293 cells (HEK293) and monocyte-like THP-1 cells. To further investigate the role of MAST3 in immunity, its expression was assessed in different human immune cells via qPCR. Here, it was found that MAST3 is constitutively expressed in CD4⁺ and CD8⁺ T cells and CD19⁺ B cells (Figure 4B). Specifically, CD4⁺ cells were found to have 35 fold higher amounts of MAST3 than placental control and CD8⁺ cells were found to have 65 fold higher amounts. We observed the same patterns of expression in primary murine immune cells (see Supplementary Figure 1).

Previous knockdown studies have shown that MAST2 is involved in the TLR4-mediated NF- κ B pathway^{25;26}. The expression of *MAST3* in HEK293 cells and availability of TLR reporter cells permitted us to examine if MAST3 also contributed to innate immune signaling pathways. To examine the potential role of MAST3, stable TLR4 and

TLR5 HEK293 cell lines were used to assess NF- κ B activity after alteration of MAST3 expression. The knockdown of MAST3 did not affect TLR5-mediated NF- κ B activity upon stimulation with flagellin, but decreased the TLR4 NF- κ B activation significantly following stimulation with lipopolysaccharides (LPS) (Figure 5). These results indicate that MAST3 regulates the NF- κ B activity specifically through the TLR4 pathway. Recent studies have highlighted that impairment of the regulation of host-organism interactions may contribute to the development of Crohn's disease and ulcerative colitis. Here we provide genetic and functional evidence that MAST3 is a genetic susceptibility factor that may contribute to the pathogenesis of IBD.

Discussion

The IBD6 locus on chromosome 19p was identified via genomewide linkage as a susceptibility locus common to the different form of IBD. In this study we aimed to identify specific risk factors responsible for the linkage signal in this region. As discussed in a recent meta-analysis on Crohn's disease²⁷, the high linkage in regions such as 5q and 6p are result of the combined effect of several genes. The identification of one modest impact gene in this specific study combined to the previous identification of *MYO9B* led us to believe that the 19p region also contains several susceptibility genes for IBD such as *MAST3*.

We report here the results of our two-stage association mapping study performed on a combined cohort of more than 1500 IBD cases and their respective parental or matched controls. A single SNP, rs273506, located in the second intron of the *MAST3* gene, was found to be associated in the screening and replication IBD cohorts with a combined P-value of 1.8×10^{-4} . Through LD and correlation patterns, we confirmed that the association signal was coming from *MAST3* and not one of the flanking genes such as *IL12RB1*, a gene involved in the same immune pathway as *IL23R* which as been associated to IBD⁷.

Further analysis of the genetic variation in *MAST3* identified four associated coding SNPs including one non-synonymous variant, S861G, and three synonymous coding SNPs (H174H, I536I, G1045G). The strength of the risk conferred by each of these coding variants of the *MAST3* gene (OR of ~ 1.18 with 95% CI [1.04, 1.34]) is in the same range as the OR of other common SNPs that have been recently associated to IBD^{6; 8; 9; 11; 12}. Consistent with the fact that they are on the same haplotype, all *MAST3* coding SNPs typed have the same OR. The SNP S861G, being non synonymous, seems the most likely of the variants tested in *MAST3* to have a functional consequence although regulatory SNPs in non-coding region have been shown to influence gene

expression²⁸. Thus we can not formally exclude the possibility that another variant than S861G is the actual causal variant.

Sequence identity analyses as well as functional assays have shown that MAST3 is a kinase²¹. Kinases are implicated in numerous cell processes several of which are important in regulation of immune responses. The MAST family is a relatively unknown branch of the kinases. The expression of the *MAST* family members is diverse; *MAST2* and *MAST4* are almost ubiquitous whereas *MAST1* is most expressed in the brain and *MASTL*, in B lymphocytes²⁹. Here we show that *MAST3* is mostly expressed in antigen presenting cells and lymphocytes. The MAST proteins contain a serine/threonine kinase domain and a PDZ protein interaction domain, not found in MASTL, which modulates interaction with different substrates. Not much is known about the protein interactions of MAST3, but PTEN, a regulator of cell growth and apoptosis, has been shown to be a substrate²¹. Valiente *et al.* showed via transfection and co-immunoprecipitation that MAST3 binds PTEN through its PDZ domain and by doing so increases PTEN stability, which in turn facilitates its phosphorylation by MAST3 and other kinases²¹.

Although little is known about MAST3 function, studies of the other family members can provide some insight. For example, the MAST2 protein has been shown to be involved in pathways important for inflammation as it regulates IL-12 p40 synthesis and NF- κ B activation^{25;26}. NF- κ B is a transcription factor known to activate the expression of many genes, including cytokines and cytokine subunits such as IL-12 p40, involved in the development of immune cells and the activation of inflammatory processes. IL-12 p40 is a subunit of IL-12, a cytokine involved in the differentiation of T_H1 cells and the production of other inflammatory cytokines such as INF γ and TNF- α . IL-12 p40 is also a subunit of IL23, a cytokine involved in the proliferation of Th₁₇ cells which in turn activates myeloid cells and NK cells to produce inflammatory cytokines, including IL6, TNF α and IL17, that drive intestinal inflammation³⁰. A genomewide association study on IBD identified a strong association to the *IL23R* gene⁷ highlighting the role of the IL23 axis in immune diseases. Given the high sequence identity within functional domains of the MAST family members and the role of MAST2 in inflammation,

MAST3 is likely to be involved in similar pathways. Previous studies have reported that *MAST2* is expressed in epithelial cells and antigen presenting cells²⁵. Consequently, we examined the expression of *MAST3* in a panel of cell lines and found the highest expression of *MAST3* in HEK293 cells and THP-1 monocytes. Further expression analysis revealed that *MAST3* is also expressed in primary murine and human immune cells (CD4+ and CD8+ T cells, CD19+ B cells). We next tested the influence of MAST3 on the specific activation of NF- κ B via TLR4 and TLR5 signaling and found that MAST3 is involved in the regulation of NF- κ B activation via TLR4 but not TLR5 signaling pathways. Members of the Toll family of transmembrane receptors (TLRs) recognize diverse structures associated with pathogenic organisms and can trigger an immune response. Deregulation of the immune response to commensal microbial gut flora is one of the potential causes of IBD. TLR4 responds to LPS, a major component of gram-negative bacteria cell wall. The interaction of LPS with TLR4 results in the recruitment of adaptor protein MyD88 and phosphorylation of IRAK and TRAF6 which leads to the phosphorylation of IKKs and the release of NF- κ B from its inhibitor, I κ B³¹. A variant of *TLR4*, 299G, was recently found to be associated to IBD and CD³². Decreases in airway responsiveness to inhaled LPS, NF- κ B activation and *IL1* expression are among the biological consequences of this variant³³. The MAST2 protein has been shown to play a role in the NF- κ B pathway through the binding and stabilization of TRAF6²⁵. We hypothesize that the MAST3 kinase is similarly involved in the stabilization and/or phosphorylation of one the components of the TLR4 cascade downstream of LPS binding to TLR4 receptor and upstream of NF- κ B activation and thus regulating NF- κ B activity. It is likely that MAST3 functions at a crossroad defining the nature of the mucosal immune system's encounters with luminal bacteria. Additional functional analyses will be needed to identify MAST3 targets in this cascade. Elucidation of MAST3's role in regulating protein interactions in intestinal inflammation will help to define the complex molecular events that contribute to IBD. Our findings add to the growing list of genes (e.g. *NOD2*, *MST1*, etc.) that indicate a pivotal role for modulators of NF- κ B activity, and more broadly genes of the innate immune response to microflora (e.g. *ATG16L1*, *IRGM* etc.), in IBD pathogenesis.

Materials and methods

DNA samples

Samples used here have been the subject of multiple other studies^{1; 7; 12; 18}. The screening cohort consists of 1863 samples (see Table 1) collected in Canada and Italy. The Canadian samples were collected by the Quebec IBD Genetics Consortium and at the Mount Sinai Hospital IBD Center (University of Toronto). Italian samples come from a population enrolled at San Giovanni Rotondo ‘CSS’ Hospital. The replication cohort consists of 1900 samples from the NIDDK IBD Genetics Consortium (IBDGC). The additional *MAST3* SNPs were genotyped on the replication cohort, part of the screen cohort (Italian and Quebec samples) and on additional Italian samples (total samples before QC: 3894). IBDGC cohorts were ascertained through Baltimore, Chicago, Los Angeles, Montreal, Pittsburgh, and Toronto Genetics Research Centers. In all populations considered, the diagnosis of IBD and classification as CD, UC or indeterminate colitis (part of the IBD cohort but not studied independently as a subphenotype) was confirmed by established criteria of clinical, radiological and endoscopic analysis, and from histology reports. A review of the patient’s chart, as well as an interview with the patient, was done to complete the phenotypic data. Written informed consent was obtained from all participants and ethics approval was granted in each of the participating institutions.

SNP selection

In order to select an appropriate set of SNPs to capture the common variation within the IBD6 region, SNP genotype information was downloaded from HapMap data phase I CEU (release 16a from March 1, 2005). To maximize the power, only SNPs with minor allele frequency greater than 10% were included in the study. A total of 1530 tag SNPs were selected using Tagger (ver.0.9.5) to be correlated with non-genotyped SNPs with a minimal r^2 of 0.7 (aggressive tagging). Additionally, three IBD5 SNPs (rs1007602, rs11739135, rs17622208), two *NOD2* (rs2066844, rs2066845) (used as positive controls) and one chromosome X (rs2106416 used as a gender control) were also

selected for inclusion in the genotyping. Three positive control SNPs were successfully genotyped and all were significantly associated (see supplementary Table 3). In line with other IBD association studies the NOD2 SNP, rs2066844, was the most associated SNP in our CD cohort.

Genotyping methods

For the screening stage, samples were genotyped on the Illumina BeadLab system (GoldenGate assay) at The Broad Institute Center for Genotyping and Analysis using a method previously described³⁴. The genotyping for the replication stage and *MAST3* coding SNPs was performed using primer extension chemistry and mass spectrometric analysis (iPlexGold assay, Sequenom) in the Laboratory of Genetics and Genomic Medicine of Inflammation in Montreal (www.inflammgen.org).

Quality control

Quality Control (QC) was performed on both screening and replication data (Supplementary Tables 1 and 2). Looking at the data distribution, we selected parameters thresholds that would optimize the quality of the data while keeping as many SNPs and samples as possible. During the QC of the screening data, SNPs with less than 75% of genotype calls were removed first to prevent weak assays from influencing the other QC parameters. At this point, one monomorphic SNP was also removed. We removed samples and SNPs with less than 90% and 80% of genotype calls, respectively. Finally, we evaluated the relationship between all samples using the Identity by State tests implemented in Plink (version v0.99q) and removed individuals that did not correspond to their pedigree data suggesting sample mix up or contamination. These analyses resulted in a final high quality dataset, used for subsequent association analysis, consisting of 1238 SNPs, 371 trios, 271 cases and 203 controls. Average call rate per individual and per SNP is 98.5%.

For the replication stage, we eliminated the SNPs with less than 85% of genotype calls and sample with less than 90% genotype calls. Then, we removed all the samples that were common with the screening samples to obtain a true independent replication

dataset consisting of 311 trios, 311 cases and 302 controls and genotyping results for 26 SNPs. Average call rate per individual and per SNP for this dataset is 98.9% and 99.0%, respectively.

Association testing

Association in trios was assessed by the transmission disequilibrium test (TDT)³⁵ and in cases/controls by a standard chi-squared (χ^2) test carried out on a 2x2 contingency table. To combine data from trios and cases/controls, the number of risk alleles transmitted in trios or found in cases was reformatted as a mean or expected count (ET and EC), observed count (OT and OC), and variance (varT and varC). A combined Z-score was then calculated by summation as $Z = [(OT+OC)-(ET+EC)]/\sqrt{(varT+varC)}$ ¹⁷. For the association result of the screening stage, the nominal two-tailed P-values were derived as the P-value corresponding to the standard normal cumulative distribution of a Z-score. Nominal associations ($P < 0.02$) in the IBD and CD cohorts were pursued in a replication phase to ensure the elimination of false positives. For the replication results a one-tailed test was used. The association scores of the additional coding and intronic SNPs of *MAST3* are the results of a two-tailed test.

Cell cultures

Cell lines were maintained under normal conditions (37°C, 5% CO₂) in standard culture media (DMEM containing 10% FCS+Fe⁺² and 50 µg/mL gentamicin for adherent cell lines; IMDM containing 10% FCS+Fe⁺², 100 µM β-mercaptoethanol, and 50 µg/mL gentamicin for suspension cell lines).

MAST3 expression in human cells

RNA from cell lines was extracted using RNeasy columns (Qiagen, Valencia, CA), according to the manufacturer's directions. 0.5 µg mRNA was used to prepare first strand cDNA using Bio-Rad's iScript enzyme (Hercules, CA). Real-time Reverse Transcriptase (RT) PCR was performed using the IQ SYBR Green Supermix (Bio-Rad) according to the recommended protocol, with the following primers: *MAST3* Forward: 5'-CTGTCGCCATTGTCGGTCC-3', *MAST3* Reverse: 5'-

ATAGCCGGAAGATGGGAGAGA-3', *GAPDH* Forward: 5'-GGAGCCAAACGGGTCATCATCTC-3', and *GAPDH* reverse: 5'-GAGGGGCCATCCACAGTCTTCT-3'. Amplicons were visualized on a 2% agarose gel to confirm correct band sizes. Samples were run in duplicate. All data were normalized to *MAST3* expression in SW480 cells and to *GAPDH* expression. In order to determine *MAST3* expression in human immune cells, a commercially available cDNA panel (Human Blood Fractions) was obtained from Clontech (Mountain View, CA). Quantitative PCR for *MAST3* was performed as above. All data were normalized to *MAST3* expression in the placenta control and to *GAPDH* expression.

MAST3 expression in mouse immune cells

Purified immune cell populations were isolated from the spleens of six week old C57BL/6 mice using magnetic beads (Miltenyi, Auburn, CA) (one liver for each cell type) according to the manufacturer's instructions; RNA from lung was also prepared as a reference control. Purity of the isolated cells was confirmed by FACS analysis (see supplemental data) and RNA was extracted as described above. Murine-specific *Mast3* primers are as follows: *Mast3 Mm* Forward: 5'-AGGCTGCATCTATCAGAGCG -3', *Mast3 Mm* Reverse: 5'- AGGCTCCTCATCGAAGCTCA-3'. Amplicons were visualized on a 2% agarose gel to confirm correct band sizes. Samples were run in duplicate. All data were normalized to cDNA from mouse lung and *Gapdh* expression.

Transfection and reporter assays

Three separate siRNAs specific for *MAST3*, along with corresponding negative controls, were obtained from Invitrogen (Carlsbad, CA). One day before transfection, 4×10^5 HEK293 cells in 1 ml of DMEM medium were plated per well in 12-well plates. The cells were transiently transfected with 1 uL of 20 nM siRNA against *MAST3* or corresponding siRNA controls using Transfectin according to the manufacturer's instructions. Three days later, the cells were rinsed in ice-cold PBS and harvested for reverse transcriptase qPCR as described above. All data was normalized to cDNA from cells transfected with control siRNA oligos and *GAPDH* expression. Of the three *MAST3*-specific siRNAs, one was chosen for further experiments (siMAST3#3).

TLR4 and TLR5 HEK293 stable cell lines were kindly provided by Dr. Douglas Golenbock (University of Massachusetts Medical School, MA). Those cells were maintained at 37°C and 5 % CO₂ in DMEM supplemented with 10 % fetal calf serum and 50 µg/mL gentamicin. One day before transfection, 4 x 10⁵ cells in 1 ml of DMEM medium were plated per well in 12-well plates. The cells were transiently transfected with 1 µL of 20 nM siRNA against *MAST3* as selected previously (forward sequence: GCUGAGGAUGAUACCAGCUACUUUG, or medium GC negative control siRNA) with 25 ng of NF-κB-luc and 0.1 ng of renilla-luc by using Transfectin according to the manufacturer's instructions. Three days later, cells were rinsed in ice-cold PBS and harvested for qPCR and reporter assays. Stimulation was done using 100 ng/ml LPS or 0.1 ng/ml Flagellin for another 6 hours. Reporter luciferase activity was measured by a standard protocol as supplied by the manufacturer (Dual Luciferase Reporter Assay System, Promega, Madison, WI) from three biological samples (i.e. three individual cell pellets). qPCR was performed to assess *MAST3* expression as described above. Samples were run in duplicate. All data was normalized to cDNA from cells transfected with control siRNA oligos and to GAPDH expression.

Acknowledgments

This project was funded by grants from the National Institute of Diabetes and Digestive and Kidney Diseases DK062432 to JDR and AI062773 to RJX.

The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research Resources.

C. Labbé is the recipient of a Fond de Recherche en Santé du Québec studentship award.

The authors would like to thank Marcia Budarf, Marie-Pierre Lévesque and Jing Lian for their help in the review of the manuscript.

Competing interests

The authors have declared that no competing interest exists.

Supplementary Information is available at the *Genes & Immunity*'s website

Tables and Figures

Table 1. Study cohorts

		Trios	Independent cases	Controls	Total samples
Screen	CD	366	241	236	1575
	UC	45	87	236	458
	IBD*	433	328	236	1863
Replication	CD	269	176	344	1327
	UC	132	177	344	917
	IBD	401	353	344	1900

* The IBD screen cohort also contains indeterminate colitis (see methods for details).

Table 2. Top association results

Marker	Position	Screening		Replication	Combined ³
		cohort ¹	P-value ²	P-value ²	P-value ²
rs11880212	334057	IBD	0.012	0.1065	0.0039
rs12980274	360977	CD	0.013	0.1172	0.4424
rs9304882	1140482	CD	0.002	N.T.	N.T.
rs8111699	1160714	CD	0.005	N.T.	N.T.
rs6510605	1325641	CD	0.016	0.4700	0.4544
rs10404242	2010585	IBD	0.009	N.T.	N.T.
rs611481	3659636	CD	0.005	0.4131	0.0395
rs9973235	4798713	CD	0.019	N.T.	N.T.
rs1564759	5406263	CD	0.01	0.1034	0.4594
rs551244	5912883	IBD	0.011	0.2218	0.1036
rs1982082	6273442	IBD	0.018	N.T.	N.T.
rs11881691	8656274	IBD	0.005	0.4017	0.0318
rs8101955	13300130	IBD	0.003	N.T.	N.T.
rs11879128	13442369	CD	0.019	N.T.	N.T.
rs892143	15506620	IBD	0.003	N.T.	N.T.
rs390017	16267034	IBD	0.014	0.1570	0.0070
rs2227356	16864405	IBD	0.001	0.1118	0.1193
rs8170	17250704	CD	0.012	0.3974	0.0757
rs3826700	17287501	IBD	0.015	0.2322	0.0165
rs7255307	17461244	CD	0.005	0.3314	0.1546
rs1078000	17482508	CD	0.007	0.2466	0.1865
rs7254755	17577582	CD	0.016	N.T.	N.T.
rs36692	17767839	CD	0.0008	0.4530	0.0619
rs36690	17771483	CD	0.008	N.T.	N.T.
rs273506	18082647	IBD	0.005	0.0237	0.0002
rs7258722	18669915	IBD	0.011	0.4370	0.0153
rs258575	19973264	CD	0.002	0.3153	0.0648
rs918442	20385941	IBD	0.007	0.4830	0.0312
rs11085397	20841228	CD	0.014	0.0368	0.0110
rs11666981	22517408	CD	0.016	0.2885	0.0947
rs431270	23260806	CD	0.017	N.T.	N.T.
rs167153	23269553	IBD	0.003	0.1125	0.0013

rs1020075	23332057	IBD	0.011	0.3326	0.0124
rs638080	23384048	IBD	0.011	0.2512	0.0725
rs629359	23414347	IBD	0.007	0.2117	0.0717
rs584460	23438251	IBD	0.013	0.3294	0.0607
rs2640201	23505118	CD	0.014	0.1888	0.0425

¹Cohort in which the SNP is most associated in screen; replication and combined values correspond to cohorts of the same phenotype

²P-value of the Z score.

³The total number of post QC samples for the combined cohort is 682 trios, 585 independent cases and 505 controls.

N.T.= Not tested due to assay or quality control failure.

Table 3. Association results for *MAST3* coding SNPs¹

Marker	MAF ²	Position	Location	AA ³ variant	Z score	P-value	OR [95% CI]
rs740691	0.475	18095441	exon 7	H174H	2.99	0.003	1.18 [1.04,1.34]
rs2270623	0.469	18106512	exon 15	I536I	3.03	0.002	1.17 [1.06,1.30]
rs8108738	0.468	18116359	exon 22	S861G	3.13	0.002	1.19 [1.05,1.34]
rs2072490	0.468	18118750	exon 25	G1045G	3.27	0.001	1.18 [1.04,1.34]

¹The total number of post QC samples for the *MAST3* coding SNPs is 610 trios, 1105 independent cases and 713 controls.

²MAF= minor allele frequency in founders

³AA= amino acid

Figure 1. Association and LD analyses in the *MAST3* region.

The upper panel summarizes the association results from the screen and replication studies. The most significantly associated SNP (screen and replication) is rs273506, located in the second intron of the *MAST3* gene as indicated in the middle panel by the yellow triangle and the red arrow. In an attempt to identify the causal variant, we then genotyped coding SNPs in *MAST3* and found that one missense and three synonymous SNPs were also significantly associated (as indicated by X symbols in the upper panel). In the lower panel, it can be seen that these associated SNPs are all found in a region of strong LD (depicted as D') and that the association does not extend beyond this region (as depicted by haplotype block 4 delimited by rs273504 and rs3736328).

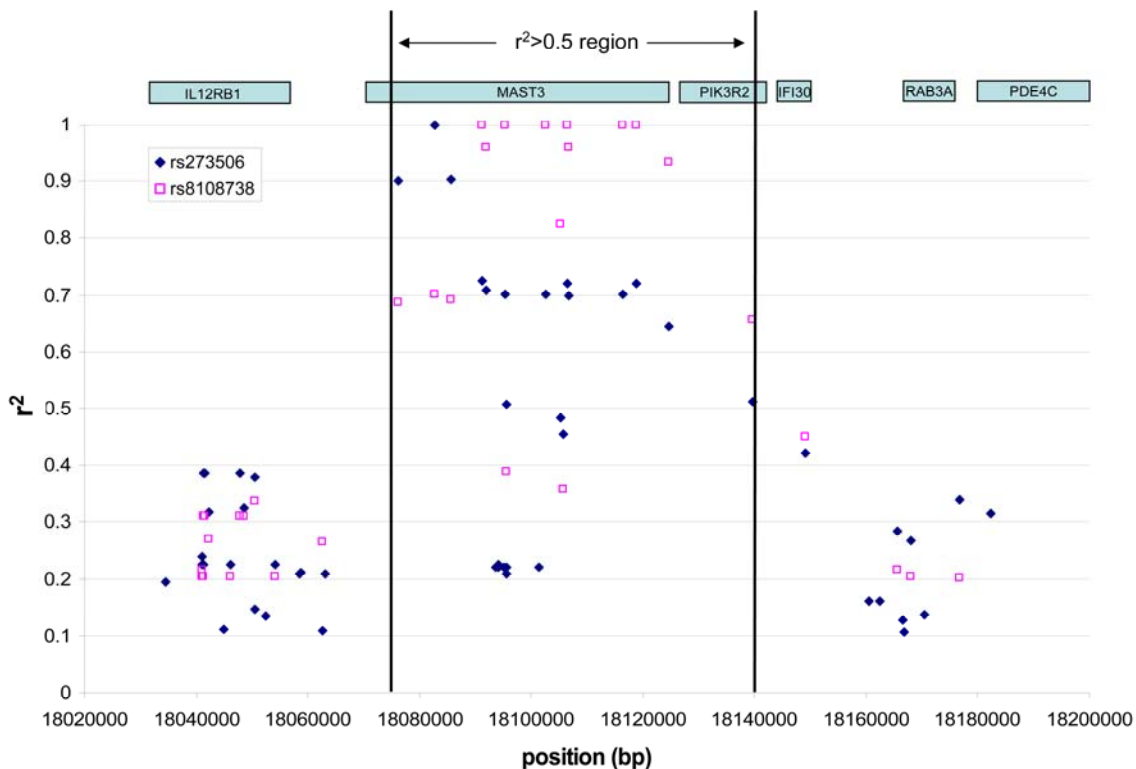


Figure 2. *MAST3* correlation neighbourhood.

Since correlation to a given allele can extend beyond haplotype block boundaries defined by D' , we examined the correlation neighbourhood surrounding the top associated SNPs. Specifically, we determined the correlation (r^2) between the significantly associated SNPs and all of the SNPs in the HapMap CEU population. All SNPs correlated at an $r^2 \geq 0.1$ with rs273506 or rs8108738 are plotted, with similar results obtained for rs740691, rs2270623 and rs2072490 (data not shown). The region defined by the SNPs that are correlated with the associated SNPs at an $r^2 \geq 0.5$ is emphasized by the black vertical lines (delimiting SNPs are rs273504 and rs2267922). This illustrates that not only are the associated SNPs all located within the *MAST3* gene but that all other known but untyped SNPs correlated with these define a correlation neighbourhood limited to *MAST3* as well as *PIK3R2*. These results along with the association results (Table 2, 3 and Figure 1) implicate *MAST3* rather than the flanking genes.

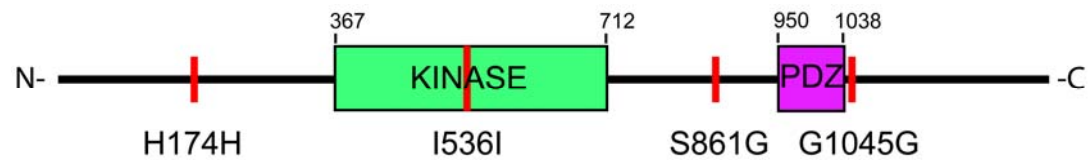


Figure 3. Schematic representation of the MAST3 protein.

The 1309 amino acids MAST3 protein contains two main functional domains: the kinase domain (in green) and the PDZ domain (in purple). The associated variants are represented by vertical red lines. The numbers indicate the amino acid delimitations as predicted by UniProt (<http://www.ebi.ac.uk/uniprot/>).

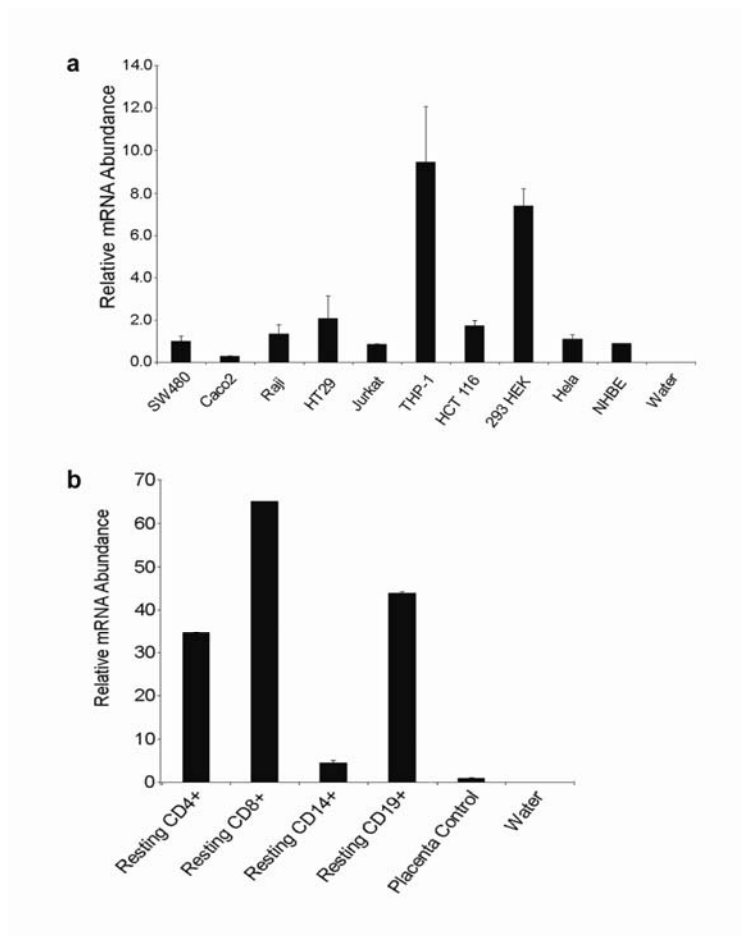


Figure 4. *MAST3* mRNA expression in human cell lines and primary immune cells.

A. Real time RT-PCR was used to assess the expression of endogenous *MAST3* gene expression in human cell lines. THP-1, a monocyte cell line, and HEK293, a kidney epithelial cell line, both showed high levels of *MAST3* gene expression (7-9 folds higher than SW480 reference cells). B. Real time RT-PCR was used to assess the expression of endogenous *MAST3* gene in human immune cell populations. *MAST3* expression was highest in CD4+, CD8+ T cells and CD19+ B cells. Error bars represent the standard deviation between triplicates.

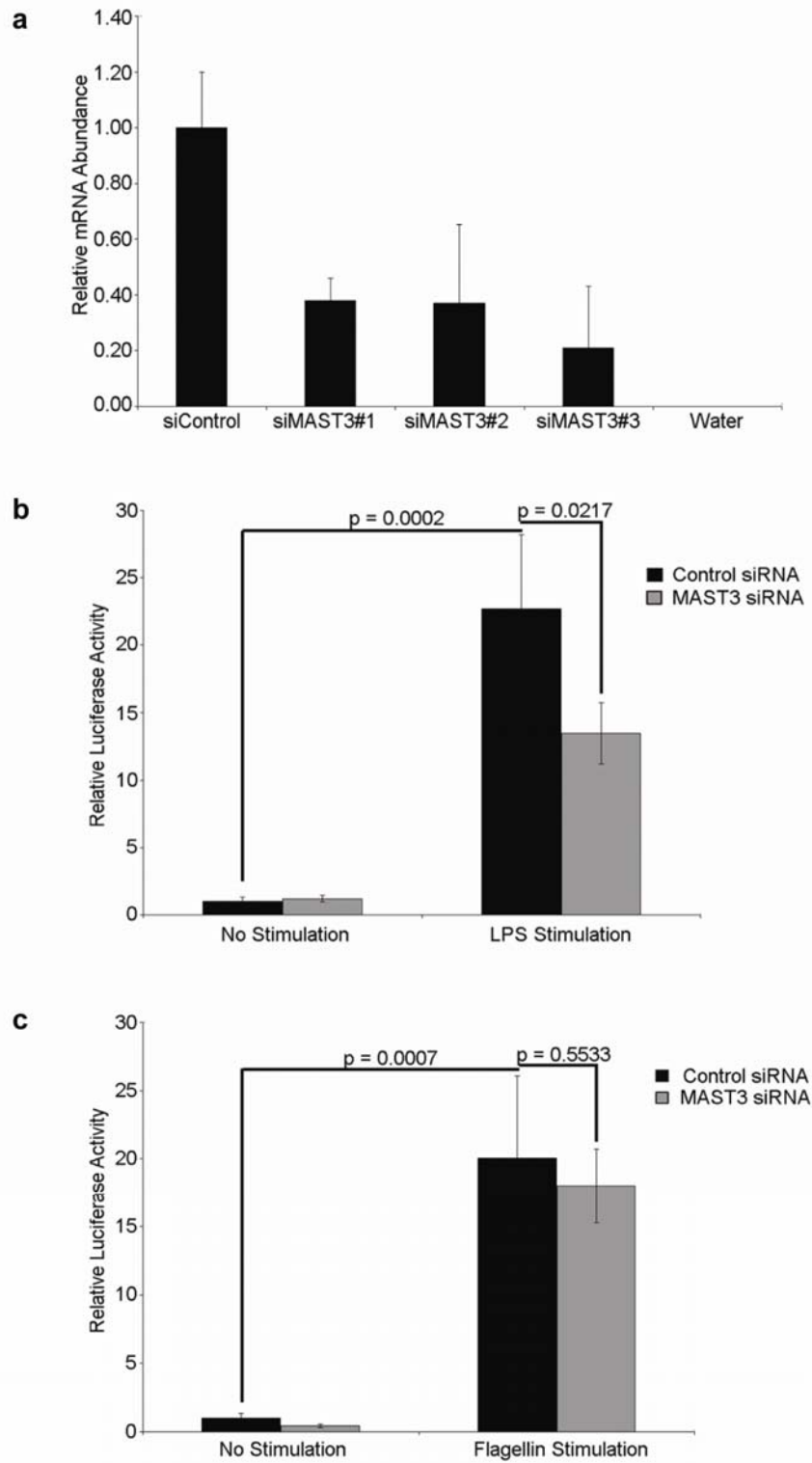
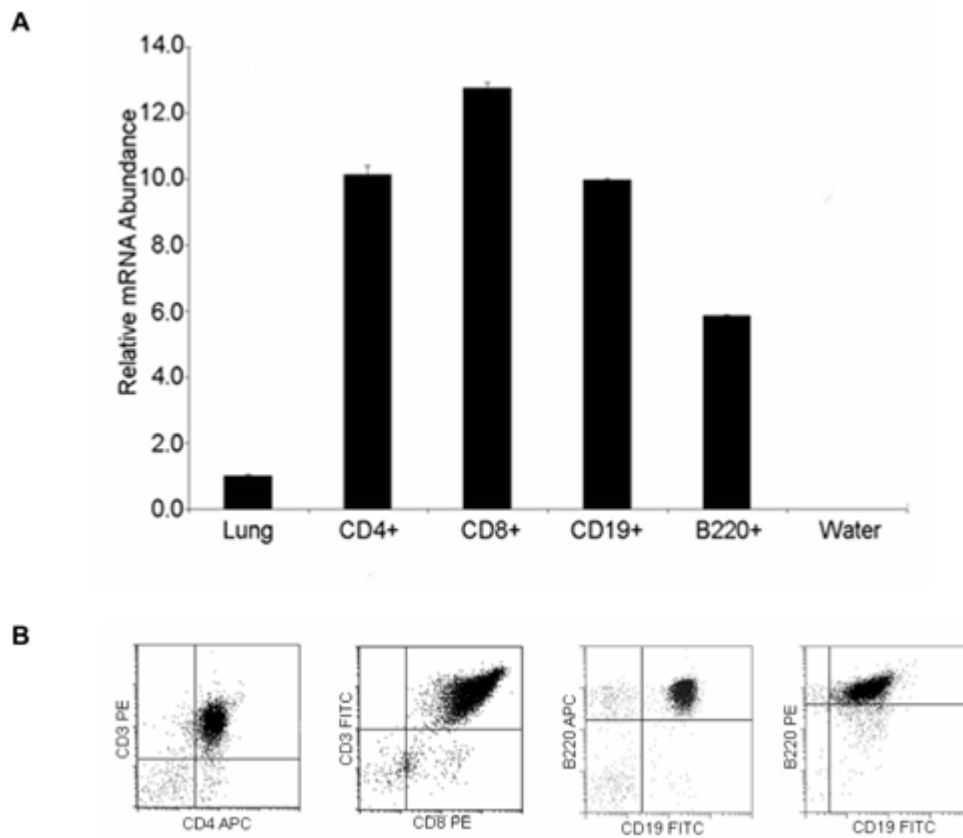


Figure 5. *MAST3* knockdown attenuates LPS-mediated NF- κ B activation.

A. *MAST3* siRNA knockdown efficiency in HEK293 cells was assessed using three *MAST3*-specific oligos and corresponding negative control. Of the three *MAST3* siRNAs, siMAST3#3 was selected for further experiments. B. *MAST3* expression was knocked down in TLR4 HEK293 cells, LPS-mediated NF- κ B activation was significantly reduced when compared to LPS stimulated cells transfected with a non-specific siRNA (P-value = 0.0217). C. Conversely, when *MAST3* expression was knocked down in TLR5 HEK293 cells, flagellin-mediated NF- κ B activation was comparable to NF- κ B activity in flagellin stimulated cells transfected with a non-specific siRNA (P-value = 0.553). Error bars represent the standard deviation between three samples taken from three individual cell pellets.

Supplementary Data



Supplementary Figure 1. *MAST3* expression in primary mouse immune cells.

A. Purified murine immune cell populations were obtained as described in Materials and Methods, and real time RT-PCR was used to assess endogenous *MAST3* gene expression. CD4+, CD8+, CD19+, and B220+ cells all showed strong *MAST3* expression (6 to 14 fold higher than reference RNA). B. FACS plots show overall purity of isolated murine immune cell fractions. Errors bars represent the standard deviation of two separate RT experiments using aliquots of the same samples.

Supplementary Table 1: Quality control steps and results from the screening stage

Steps	Details	SNPs	DNA	Trios	ME/ trio/ SNP	Cases	Controls
0		1536	1893	433	0,011	328	236
1	no call on Illumina (failed assays)	1259	1893	433	0,014	328	236
2	Removed SNPs with less than 75% genotyping	1245	1893	433	0,014	328	236
3	Removed monomorphic SNP (MAF<1%)	1244	1893	433	0,014	328	236
4	Removed families with more than 0.005 ME / trios /SNP	1244	1803	403	0,005	328	236
5	Removed individuals with low genotyping (<90%)	1244	1691	373	0,004	283	206
6	Removed SNPs with low genotyping (<80%)	1238	1691	373	0,004	283	206
7	Removed individuals that failed IBS tests	1238	1667	371	0,004	271	203

Supplementary Table 2: Quality control steps and results from the replication stage

Steps	Details	SNPs	DNA	Trios	ME/Trio/SNP	Cases	Controls
0		37	1900	401	0,00506	353	344
1	No call on Sequenom (failed assays)	31	1900	401	0,00539	353	344
2	Removed families with an excess of ME (>0.1/SNP)	31	1888	397	0,00333	353	344
3	Removed SNPs with low % genotyping (<85%)	26	1888	397	0,00116	353	344
4	Removed the individual with low % genotyping (<90%)	26	1825	362	0,00074	339	330
5	Removed samples common to step 1 samples	26	1608	311	0,00074	311	302

ME: Mendelian Errors

Supplementary Table 3. All association results for the screening stage

(Available online at:

<http://www.nature.com/gene/journal/v9/n7/supinfo/gene200857s1.html?url=/gene/journal/v9/n7/full/gene200857a.html>)

References

1. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000; **66**(6): 1863-1870.
2. Kappelman MD, Rifas-Shiman SL, Kleinman K, Ollendorf D, Bousvaros A, Grand RJ *et al.* The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. *Clin Gastroenterol Hepatol* 2007; **5**(12): 1424-1429.
3. Baumgart DC & Carding SR. Inflammatory bowel disease: cause and immunobiology. *Lancet* 2007; **369**(9573): 1627-1640.
4. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**(6837): 599-603.
5. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**(6837): 603-606.
6. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature genetics* 2001; **29**(2): 223-228.
7. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**(5804): 1461-1463.
8. Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Hasler R *et al.* Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS ONE* 2007; **2**(1): e691.
9. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007; **39**(2): 207-211.
10. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 2007; **3**(4): e58.

11. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007; **39**(7): 830-832.
12. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007; **39**(5): 596-604.
13. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**(7145): 661-678.
14. Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci U S A* 1998; **95**(13): 7502-7507.
15. Ma Y, Ohmen JD, Li Z, Bentley LG, McElree C, Pressman S *et al.* A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999; **5**(4): 271-278.
16. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J *et al.* The DNA sequence and biology of human chromosome 19. *Nature* 2004; **428**(6982): 529-535.
17. Tello-Ruiz MK, Curley C, DelMonte T, Giallourakis C, Kirby A, Miller K *et al.* Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19. *Eur J Hum Genet* 2006; **14**(6): 780-790.
18. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM *et al.* Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 2006; **131**(6): 1768-1774.
19. Steinman L. A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage. *Nat Med* 2007; **13**(2): 139-145.
20. NCBI (2008). HomoloGene: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene>. In.
21. Valiente M, Andres-Pons A, Gomar B, Torres J, Gil A, Tapparel C *et al.* Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. *J Biol Chem* 2005; **280**(32): 28936-28943.

22. Lumeng C, Phelps S, Crawford GE, Walden PD, Barald K & Chamberlain JS. Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases. *Nat Neurosci* 1999; **2**(7): 611-617.
23. Sun L, Gu S, Li X, Sun Y, Zheng D, Yu K *et al.* [Identification of a novel human MAST4 gene, a new member of the microtubule associated serine-threonine kinase family]. *Mol Biol (Mosk)* 2006; **40**(5): 808-815.
24. Walden PD & Cowan NJ. A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with microtubules of the spermatid manchette. *Mol Cell Biol* 1993; **13**(12): 7625-7635.
25. Xiong H, Li H, Chen Y, Zhao J & Unkeless JC. Interaction of TRAF6 with MAST205 regulates NF-kappaB activation and MAST205 stability. *J Biol Chem* 2004; **279**(42): 43675-43683.
26. Zhou H, Xiong H, Li H, Plevy SE, Walden PD, Sassaroli M *et al.* Microtubule-associated serine/threonine kinase-205 kDa and Fc gamma receptor control IL-12 p40 synthesis and NF-kappa B activation. *J Immunol* 2004; **172**(4): 2559-2568.
27. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD *et al.* Genome-wide association defines more than thirty distinct susceptibility loci for Crohn's disease. *Nature Genetics* 2008; **in press**.
28. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC *et al.* A genome-wide association study of global gene expression. *Nature genetics* 2007; **39**(10): 1202-1207.
29. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* 2002; **99**(7): 4465-4470.
30. McGovern D & Powrie F. The IL23 axis plays a key role in the pathogenesis of IBD. *Gut* 2007; **56**(10): 1333-1336.
31. Gay NJ & Gangloff M. Structure and function of Toll receptors and their ligands. *Annu Rev Biochem* 2007; **76**: 141-165.
32. De Jager PL, Franchimont D, Waliszewska A, Bitton A, Cohen A, Langelier D *et al.* The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun* 2007; **8**(5): 387-397.
33. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M *et al.* TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 2000; **25**(2): 187-191.

34. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D *et al.* High-throughput SNP genotyping on universal bead arrays. *Mutat Res* 2005; **573**(1-2): 70-82.
35. Spielman RS, McGinnis RE & Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**(3): 506-516.

Chapitre 5

MAST3: modulateur d'inflammation

Dans le deuxième volet du projet, le but était d'initier la caractérisation du gène de susceptibilité aux MII *MAST3*. La stratégie employée pour y arriver s'appuie sur deux idées :

1. Le peu d'informations disponibles dans la littérature sur le gène ou la protéine *MAST3* qui prévient l'élaboration d'hypothèses précises et supporte une stratégie d'étude non restreinte par des hypothèses.
2. Notre première démonstration qui implique *MAST3* dans la réponse immunitaire via la voie de NF- κ B¹.

L'article qui suit décrit la suite de nos études fonctionnelles sur *MAST3* qui se traduit par une étude approfondie sur l'effet de la modulation de l'expression du gène *MAST3* sur l'expression pangénomique endogène. En résumé notre étude démontre :

- L'effet d'une surrexpression du gène *MAST3* sur l'expression génique endogène dans un premier modèle *ex vivo* (cellules HEK293).
- La validation et la confirmation de ces résultats d'expression génique dans un second modèle cellulaire composé de cellules THP1 de type macrophage qu'on a soumises à un knockdown du gène *MAST3*.
- L'identification d'un groupe de gènes dont l'expression est régulée par *MAST3* et qui est impliqué dans des fonctions immunitaires.
- Une confirmation additionnelle du rôle de *MAST3* dans la modulation de l'activité de NF- κ B.
- La corrélation temporelle entre l'activité de NF- κ B et l'expression génique régulée par *MAST3* qui suggère que *MAST3* exerce son influence sur cette dernière via la voie de NF- κ B.

- L'observation que l'expression des gènes régulés par *MAST3* est significativement supérieure dans les zones enflammées de la muqueuse du côlon de patients atteints de colite ulcéreuse comparé à des zones saines.

Cet article présente une caractérisation fonctionnelle du gène de susceptibilité aux MII *MAST3*, réaffirme l'importance de la voie NF- κ B dans la pathogenèse des MII et relie l'expression d'un groupe de gènes dont l'expression est modulée par *MAST3* aux manifestations cliniques des MII.

ARTICLE 2

Genome-wide Expression Profiling Implicates a *MAST3*-Regulated Gene Set in Colonic Mucosal Inflammation of Ulcerative Colitis Patients

Catherine Labbé M.Sc^{1,2}, Gabrielle Boucher M.Sc¹, Sylvain Foisy Ph.D¹, Azadeh Alikashani M.Sc¹, Herbert Nkwimi M.Sc¹, Geneviève David B.Sc.^{1,2}, Mélissa Beaudoin M.Sc¹, Philippe Goyette Ph.D¹, Guy Charron Ph.D¹, Ramnik J. Xavier MD^{3,4} and John D. Rioux Ph.D^{1,2}

- 1) Institut de Cardiologie de Montréal, Montréal, Qc, Canada;
- 2) Département de Médecine, Université de Montréal, Montréal, Qc, Canada;
- 3) Center for Computational and Integrative Biology and Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.
- 4) Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, MA, USA.

Running Head: IBD risk factor *MAST3* regulates the expression of inflammatory cascade

Keywords: *MAST3*, Inflammatory Bowel Disease, NF- κ B

Cet article a été publié en ligne dans le journal *Inflammatory Bowel Disease* le 12 octobre 2011. Il est inclus dans cette thèse avec l'accord des éditeurs et de tous les co-auteurs.

Contribution des auteurs

Pour cet article, j'ai effectué 80% des travaux incluant la planification expérimentale, la création et la mise au point des modèles cellulaires, les essais fonctionnels, l'analyse des résultats et la rédaction de l'article. Boucher et Nkwimi sont des biostatisticiens qui ont participé à l'analyse de données d'expression sur les biopsies. Foisy est bioinformaticien et a participé à l'analyse des données des puces d'expression. Alikashani est assistante de recherche et a participé aux essais fonctionnels (Western blots). David est étudiante à la maîtrise et a été impliquée dans le recrutement des patients pour les biopsies et l'extraction d'ARN provenant des biopsies. Beaudoin est assistante de recherche et a participé aux études d'expression (manipulation des robots). Goyette et Charron sont associés de recherche et ont fournis des conseils pour le design expérimental et la rédaction de l'article. Xavier est gastroentérologue et a participé à la rédaction de l'article. Rioux est l'investigateur principal, il a participé à la conception de l'étude, a fourni les échantillons, a participé aux analyses et à la rédaction de l'article.

Abstract

Crohn's disease (CD) and ulcerative colitis (UC) are inflammatory bowel diseases (IBD) presumably caused by dysregulated immune responses to the gut microbiota. Genetic association studies have implicated dozens of chromosomal regions or loci in IBD susceptibility. The next challenge is to explain the individual role of each of these modest effect loci in disease state. We have previously identified *MAST3* as an IBD susceptibility gene through genetic fine-mapping of the 19p linkage region. Testing *MAST3* in a reporter assay provided preliminary evidence that *MAST3* modulates the activity of inflammation-related transcription factor nuclear factor kappa B (NF- κ B). Here, we further characterized the function of *MAST3* through an examination of the influence of the modulation of *MAST3* expression on endogenous genome-wide expression patterns. More specifically, we looked at differential gene expression resulting from overexpression and knockdown of the *MAST3* gene in epithelial and macrophage cell lines. We highlight a group of genes whose expression is modulated by *MAST3* and correlate their expression with NF- κ B activity. Their expression was found to be enriched in inflamed mucosal tissue of UC patients, confirming the importance of these genes in IBD. These *MAST3*-regulated genes are central to mucosal immune responses. Among them are pro-inflammatory cytokines (e.g. *CCL20*, *IL8*), regulators of NF- κ B (e.g. *TNFAIP3*, *LY96*, *NFKBIA*), genes involved in interferon-induced defense against pathogen invasion (e.g. *IFIT1*, *ISG15*) and genes involved in cell adhesion and/or migration (e.g. *CD44*, *TMOD1*). Taken together, these results confirm *MAST3* as a modulator of the inflammatory response through regulation of immune gene expression in the gut of IBD patients.

Introduction

Inflammatory bowel diseases (IBD) refer to Crohn's disease (CD) and ulcerative colitis (UC), two common inflammatory diseases of the gastrointestinal tract most prevalent in urban areas of North America and Europe. Current consensus assigns the causes of IBD to environmental factors - such as cigarette smoking, appendicitis, hygiene, and gut microbiota - combined to genetic predisposition that contribute to the development of a dysregulated immune response to the gut microbiota leading to mucosal inflammation². Genome-wide genetic association studies have identified 99 IBD susceptibility regions or loci and some of them contain multiple genes^{3;4}. Targeted association studies, such as ours¹, have also identified several susceptibility genes. However, estimation of heritability explained by these loci suggests that many more remain to be discovered. The important challenge is to identify individual role of the IBD genes to better understand the molecular mechanisms underlying disease onset and chronic inflammation.

In a previous study, we identified the microtubule associated serine/threonine protein kinase-3 (*MAST3*) gene as a genetic risk factor for inflammatory bowel disease (IBD) through association fine-mapping of the chromosome 19p region (also known as linkage region IBD6)¹. *MAST3* is one of the least studied members of the *MAST* kinase family. Most members of the *MAST* family are ubiquitously expressed but their expression is highest in the brain⁵. *MAST2*, the most studied member of this family, has been involved in immune reactions, more specifically in the regulation of NF- κ B⁶. Based on this knowledge, we performed a knockdown of *MAST3* in HEK293 cells and tested transcription factor NF- κ B activity through reporter assays. We showed that a knockdown of *MAST3* decreased significantly and specifically receptor TLR4 stimulated activity of NF- κ B¹.

In the current study, we aimed to further characterize the role of *MAST3* by modulating its expression, via overexpression and knockdown (KD) of the gene, in *ex vivo* cell

culture systems and determining the impact on genome-wide endogenous gene expression. Specifically, we first compared transcriptional profiles of mock-transfected and *MAST3*-overexpressing HEK293 cells using genome-wide microarray analysis. Our genome-wide analysis highlighted a group of 28 *MAST3*-regulated genes whose expression levels were increased by ≥ 2 -fold. In order to assess specificity of these results, we knocked down *MAST3* in immune cells (THP1) and determined the expression levels of a subset of the genes identified in the overexpression experiment. We observed that for 7/9 genes tested, the expression was significantly reduced and thus the results of our KD confirmed our previous overexpression model results. To assess the importance of the *MAST3*-regulated genes in the clinical presentation of IBD, we compared gene expression in colonic mucosal tissue of healthy controls and of UC patient non-inflamed and inflamed regions. We found a significant enrichment of the expression of the *MAST3* gene set in inflamed vs non-inflamed tissues ($P=0.005$).

The *MAST3*-regulated genes coordinate immune responses in intestinal inflammation. They include pro-inflammatory cytokines, regulators of NF- κ B, interferon induced genes, and genes involved in adhesion and cell migration and a majority of these genes are regulated by NF- κ B which suggests that *MAST3* controls their expression via the NF- κ B pathway.

Furthermore, the results of our expression study in patients biopsies suggest that *MAST3* can regulate the expression of a set of genes that discriminates inflamed from non-inflamed tissues and that the overexpression of these genes is part of the clinical manifestation of the disease.

Results

Overexpression model identifies a MAST3 regulated gene set involved in immune response

In order to gain insight into the function of the IBD susceptibility gene *MAST3*, we overexpressed *MAST3* cDNA in HEK293 cells (chosen because these cells are easily and efficiently transfected) and analysed the resulting *MAST3* transcriptional signature using genome-wide cDNA microarray. The endogenous expression of *MAST3* is relatively low in these cells, and was measured by reverse transcription quantitative polymerase chain reaction (RT qPCR) to be several orders of magnitude lower than in cells overexpressing *MAST3* (Figure II1 Annexe II) *MAST3* protein levels at 24h post-transfection were increased by 60-fold in these cells (Supplementary Figure 1 and Figure II2 Annexe II). Unsupervised hierarchical clustering of expression data clearly discriminates between cells overexpressing the *MAST3* gene and mock-transfected cells (Figure II3, Annexe II). In total, 195 probes showed significant ($P_{\text{corrected}} < 0.05$) differential expression (Supplementary Table 1). This represents 0.4% of all interrogated probes. The differentially expressed probes represent 146 different validated protein coding genes, 24 genes encoding small nuclear RNAs and 15 encoding hypothetical proteins. Of the 185 differentially expressed genes, 113 showed higher expression and 72 showed lower expression in cells overexpressing the *MAST3* gene compared to mock-transfected cells. Underexpressed genes (in cells overexpressing *MAST3*) consisted mostly of small nuclear RNAs and hypothetical proteins, so we decided to focus on the overexpressed genes. We were most interested in the top 28 genes with 2-fold or greater increases in expression (Table 1). We validated the microarray expression data by testing the expression of 5 genes randomly chosen among the top 28 using RT qPCR (Supplementary Figure 2).

This set of 28 genes (Table 1) whose expression appears to be regulated by *MAST3* shows a significant enrichment ($P < 0.01$) of functional involvement in defense response

(10-fold), inflammatory response (16-fold), response to wounding (11-fold), regulation of apoptosis (7-fold), and regulation of programmed cell death (7-fold) (Supplementary Table 2). The importance of this group of genes in the immune response can be further recognized in their known function based on the literature: they encode pro-inflammatory cytokines, regulators of the activity of NF- κ B, proteins involved in adhesion and cell migration, and proteins induced by interferon in response to pathogen invasion or tumor progression. Additionally, 61% of these genes are regulated by NF- κ B via functionally validated binding sites in their promoter (Figure 1 and Table 1) suggesting that MAST3 may act through this pathway.

The median fold-increase in gene expression in the MAST3 regulated gene set (top 28 genes) in *MAST3*-overexpressing cells compared to mock-transfected cells is 3. However, two genes are overexpressed by greater than 10-fold, *CCL20* ($P=1.31 \times 10^{-10}$, 38-fold) and *IL8* ($P=6.09 \times 10^{-6}$, 14-fold), suggesting that MAST3 levels have a greater influence on the expression of these genes than the other upregulated genes. The *CCL20* chemokine is the only known ligand of the CCR6 receptor and CCR6 signaling is involved in the recruitment of dendritic cells, and other antigen presenting , as well as CD4+T cells to the sites of epithelial inflammation⁷. The chemokine IL-8 is produced predominantly in the lamina propria of the colon and its level correlates with mucosal inflammation⁸.

Knockdown of MAST3 expression in immune cells confirms gene expression profile

In order to further study the effect of MAST3, we created a *MAST3* knockdown (KD). We were interested in confirming that the observed gene expression patterns from our overexpression model were reversed in the knockdown. Additionally, we wanted to study immune gene expression in an immune cell context so we selected monocyte-like THP1 cells. Our KD was created by transducing THP1 cells with *MAST3* small hairpin RNA (shRNA).

Interestingly, the KD has different effects on MAST3 levels at different stages of cell differentiation and stimulation (Supplementary Figure 3, Figures II4-II5 Annexe II). The

KD has the greatest effect on *MAST3* levels in cells differentiated by phorbol 12-myristate 13-acetate (PMA) and stimulated by lipopolysaccharide (LPS) for 24h (a decrease of 75% of mRNA levels and of 92% of *MAST3* protein levels). In order, to maximize the effect of the *MAST3* KD on endogenous gene expression, we studied THP1 cells treated with PMA and LPS for 24h. We tested a subset of the *MAST3* regulated gene set. Of the 9 genes tested, 7 are significantly underexpressed ($P < 0.05$) in our KD model compared to control and 2 show no significant difference (Table 2). Overall, this downregulation of gene expression in *MAST3* KD cells is consistent with the upregulation observed in the cells that overexpress the *MAST3* gene. Notably, *IL8* is highly expressed in PMA and LPS treated THP1 (4800-fold increase from unstimulated THP1 stage) and its expression level is reduced to 55% of the control expression in the KD ($P = 0.01$).

Our expression study in cells overexpressing the *MAST3* gene highlighted a group of upregulated genes that are regulated by transcription factor NF- κ B. We were interested in exploring the extent of the effect of a *MAST3* KD on the NF- κ B dependent endogenous gene expression but were somewhat limited by the HEK293 cell model (fibroblast like human embryonic kidney cells). The THP1 cell model allowed us to study the expression of additional NF- κ B regulated genes not necessarily expressed in HEK293 cells. We monitored the expression of several interleukins, TGF β 1 and members of the NF- κ B pathway (Table 3). As expected, based on our previous results, the KD of *MAST3* decreases the expression of a majority (4/6) of the pro-inflammatory genes tested. The KD also upregulates the expression of anti-inflammatory cytokine gene *IL10* to 210% of non target KD control ($P = 0.003$). These additional results in our THP1 *MAST3* KD model support the previous results from the overexpression model and suggest a prominent pro-inflammatory role for *MAST3*.

*NF- κ B activity correlates to *MAST3* regulated gene expression*

We were interested in studying the relationship between NF- κ B activity and the *MAST3* regulated gene set expression. We have previously shown that a KD of *MAST3*, in LPS stimulated HEK293 cells that constitutively express the TLR4 receptor, decreases the

activity of transcription factor NF- κ B¹. Thus, we first wished to confirm these findings in our two new cell models of overexpression in HEK cells and KD in THP1 cells. Then, we used the *IL8* and *CCL20* genes as proxies for the MAST3 regulated gene set to study the temporal correlation between NF- κ B activity and MAST3 gene set expression.

We tested the activity of NF- κ B at different times post-transfection of *MAST3* (Figure 2A). We observed a significant increase in activity specific to MAST3 beginning 12 hours post-transfection. The activity of NF- κ B in cells overexpressing the *MAST3* gene, at 12 hours and 24 hours post-transfection is 159% (P=0.01) and 354% (P=0.02) that of mock-transfected cells, respectively. This increase in NF- κ B activity is correlated with significant MAST3 mRNA (Figure II1 Annexe II) and protein level increases (Supplementary Figure 1).

The activity of NF- κ B in our THP1 knockdown confirms the results of our previous study in HEK293 cells. A significant decrease in NF- κ B activity (70% of control, P=0.02) is detected in THP1 cells differentiated with PMA and treated with LPS for 6h (Figure 2B). Importantly, these results correlate with decreased MAST3 mRNA (67.2% decrease) (Figure II4 Annexe II) and protein (86% decrease) levels (Supplementary Figure 3 and Figure II5 Annexe II).

The expression levels of MAST3 regulated genes *CCL20* and *IL8* genes are induced early and significant differences between *MAST3*-overexpressed cells and control can be observed as soon as 8h post-transfection (Figure 2C and 2D). Yet, the most significant fold differences in expression between cells overexpressing the MAST3 gene and mock-transfected cells are observed at 12 and 24h (10 and 32-fold increases respectively), and thus correlate temporally with the increases in NF- κ B activity.

MAST3 regulated gene set expression is enriched in inflamed tissues of UC patients

To establish the importance of the MAST3- regulated gene set in the clinical presentation of the disease, we tested its expression in human intestinal mucosa tissues from patient biopsies. In order to maximize our statistical power to detect differential

expression by limiting noise, we defined the MAST3-regulated gene set as the group of genes that had been overexpressed by 3-fold or greater in our cell model of MAST3 overexpression (Table 1, Top 14 genes). We were mostly interested in answering two questions: 1. Are there differences in gene expression between intestinal tissue from healthy people and normal looking intestinal tissue from patients affected with UC? 2. Are there differences between normal looking intestinal tissues and inflamed tissue from UC patients? In order to answer these questions, we analyzed the differential gene expression profiles between biopsies of controls (BC) and patients' non-inflamed (BPN) and inflamed (BPI) regions of the colon and rectum. Specifically, we were interested in discovering whether there was an enrichment of our MAST3-regulated genes as part of the inflammatory process in IBD patients.

There were no detectable differences between BC (n=16) and BPN (n=15) (Table 4 and Supplementary Table 3). In contrast, we observed significant differences between BPN (n=19) and BPI (n=15). Of all the probes of the array, 34.5% (8498 out of 24607) showed differential expression ($P < 0.05$). In our gene set, 71.4% of genes (10 out of 14) showed differential expression ($P < 0.05$), which represent a significant 2-fold enrichment ($P = 0.005$) (Table 4). After correction for multiple testing, we examined the expression patterns for individual genes. In general, expression of genes from the MAST3-regulated gene set was increased in the BPI compared to BPN. However, this trend did not extend to *MAST3* expression for which the difference between BPI and BPN did not reach statistical significance. For 5 genes the individual increase in expression was significant, *CD44* ($P = 0.001$), *TNC* ($P = 0.017$), *ISG15* ($P = 0.02$), *NFKBIZ* ($P = 0.02$) and *IL8* ($P = 0.031$) (Table 5).

Discussion

Genome-wide and targeted association studies have already identified numerous modest effect genes and estimations of cumulative heritability predict that they will continue to identify new genes for some time in the future. In order to use these genetic findings to benefit patients, functional studies are essential. The challenge is to explain the individual role of each of these genes in the pathogenesis of complex diseases.

Following the identification of *MAST3* as an IBD gene ¹, we were interested in studying its functional effect in the context of the immune system. Since *MAST3* had not been studied extensively, we decided to use a non-targeted approach to examine the effect of perturbing the expression of *MAST3* (overexpression) in a well-characterized cell system. The hypothesis being that overexpression of *MAST3* will have an influence on expression of genes in related biological pathways. We report here our results that highlight a group of *MAST3*-regulated genes heavily involved in immune responses. These genes are functionally linked to NF- κ B as a majority of them are regulated and/or regulators of the NF- κ B pathway. We show that the activity of NF- κ B is correlated temporally with *MAST3*-regulated gene expression implying that *MAST3* acts on the NF- κ B pathway to trigger immune reactions through changes in gene expression. The *MAST3* gene set is upregulated in colonic biopsies of inflamed regions from patients with UC as compared to non-inflamed regions. This suggests that *MAST3* contributes to the dysregulation of the immune response to microbiota underlying IBD through modulation of gene expression.

To examine the *MAST3*-regulated gene set in a clinically relevant context, we studied genome-wide expression in colonic biopsies from control and UC patients. We observed a significant increase in the expression of the *MAST3*-regulated genes in inflamed vs non-inflamed tissues. Since 20% of the genes in this set are encoding proinflammatory cytokines, including the key chemokines CCL20 and IL8, this emphasizes the role of the

inflammatory process in the initiation and/or maintenance of the disease state.

Interestingly, *IL8* is a chemokine that is important in innate immunity and highly regulated by *MAST3*. *IL8* expression has been shown to be increased in both UC and CD patients and *IL8* mRNA and protein level correlates to inflammation severity⁸. Mutations in *NOD2*, the first gene associated to CD⁹, were shown to regulate the expression of *IL8*¹⁰.

In addition, a few of the *MAST3*-regulated genes are involved in the regulation of the NF- κ B pathway which seems to also implicate *MAST3* in a feedback mechanism to prevent uncontrolled inflammation. *NFKBIA* is among the NF- κ B regulators within this gene set. *NFKBIA* encodes proteins I κ B α which sequesters the NF- κ B complex in the cytoplasm therefore inhibiting its nuclear localization and transcription factor functions¹¹. Another of these regulators is *TNFAIP3* encoding protein A20 which acts on the NF- κ B pathway through interaction with TRAF6¹². Overexpression of A20 has been shown to block TLR4 activation of NF- κ B¹³. MD-2, encoded by gene *LY96*, is the co-receptor to TLR4. The binding of LPS to TLR4 and subsequent NF- κ B activation is dependent on the formation of the TLR4-MD-2 complex¹⁴. *IER3* has been shown to interact with NF- κ B and inhibits its activation, thus interfering with downstream gene expression¹⁵.

Several IBD genes influence immune responses through modulation of the NF- κ B pathway. Notably, a recent study into the effect of the overexpression of IBD associated variant *NOD2*^{L1007finsC} on global gene expression underscores some NF- κ B regulated gene found in the *MAST3*-regulated gene set¹⁶. *NOD2*^{L1007finsC} was found to upregulate the expression of pro-inflammatory cytokine gene *IL8* and *CXCL2* and of NF- κ B regulator genes *NFKBIA*, *IER3* and *TNFAIP3*. Our combined findings give weight to the hypothesis that the dysregulation of a limited number of biological pathways is involved in the pathogenesis of IBD. *NOD2* is the first identified and among the most systematically replicated CD genes. As opposed to *MAST3*, it is also among the most studied functionally. *NOD2* has been involved in the regulation of immune gene expression and epithelial barrier integrity¹⁷. Here we show that another IBD risk factor,

MAST3, is modulating the expression of genes and regulating the activity of NF- κ B in a similar way and thus that its variants could translate into similar cellular consequences. It is possible that there is an inflammation gene set that is regulated by several IBD genes such as *NOD2* and *MAST3*. The upregulation of the expression of this inflammation gene set in mucosal gut tissues could trigger the conditions for the development of IBD. In this context of interconnectivity between IBD genes, functional studies that aim to decipher disease-causing pathways are becoming essential companions to genetic studies in the quest to improve our understanding of the molecular mechanisms underlying IBD.

Among *MAST3*-regulated genes, a few - *TNF*, *TNFRSF9* and *CCL20/CCR6* - are located in IBD associated loci which also underlines the interconnectivity of the IBD genes. The *TNF* gene is on chromosome 6p21 in the major histocompatibility complex (CMH) region and has been associated to CD following a metanalysis of genome-wide associated studies¹⁸. The *TNF* gene encodes tumor necrosis factor α , a pro-inflammatory cytokine that induces cell proliferation and differentiation and that is found in increased amounts in patients suffering from IBD^{19;20}. Biological therapies that target TNF have been used for several years to treat refractory IBD. Notably, the anti-TNF humanized antibody Infliximab has shown efficacy in reducing intestinal inflammation and maintaining remission, preventing fistulae, reducing the need for steroids treatment, reducing hospitalization and need for surgery, and improving quality of life²¹⁻²⁴. The *TNFRSF9* gene is located in a locus recently associated to UC³. *TNFRSF9* encodes a cell surface transmembrane protein that is involved in cell differentiation and activation^{25;26}. *TNFRSF9* has been shown to activate NF- κ B in a TRAF-dependent manner²⁷. *CCL20*, the most upregulated gene in our overexpression model, has not been associated to IBD. However, it encodes the sole ligand of receptor CCR6 which is encoded by a CD associated gene¹⁸. *CCL20* and *CCR6* are involved in the regulation of mucosal immunity through the recruitment of antigen presenting cells to inflammation sites and homing of dendritic cells and CD4+ T cells to the lymphoid tissue of the gut⁷. There is much interest in *CCR6* because it is the only receptor

expressed on every T_H17 cell subsets²⁸ and T_H17 driven immune responses are hypothesized to be responsible for CD pathogenesis²⁹.

Our genome-wide expression profiling is a first step into the investigation of the role of MAST3 in the immune response. Several signal transducers acting between TLR4 and NF- κ B are phosphorylated on serine and/or threonine. It is possible that the MAST3 kinase modulates the activity of NF- κ B through the phosphorylation and activation of one of these molecules. Also, MAST3 might physically interact and modulate the pathway through stabilization and chaperoning of a target. MAST2, a protein from the same family, has been shown to regulate NF- κ B through the formation of a complex with TRAF6⁶. Additional functional studies will be needed to pinpoint the precise target of MAST3 in the NF- κ B pathway. In the meantime, this study reinforces the importance of this pathway in IBD and identifies an IBD risk factor, *MAST3*, as a regulator of NF- κ B activity and immune gene expression.

Materials and Methods

Cell culture

THP1 cells (ATCC TIB-202) were cultivated in RPMI medium supplemented with 10% fetal bovine serum, 2-mercaptoethanol (final concentration 0.05 mM), and penicillin (100 units/mL)-streptomycin (100 ug/mL). TLR4 293-hTLR4A-MD2-CD14 HEK cells (InvivoGen #293-hTlr4md2cd14) were cultivated in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum, penicillin (100 units/mL)-streptomycin (100 ug/mL), blasticidin (10 ug/mL) and HygroGold (50 ug/mL, InvivoGen) at 37°C in a 5% CO₂ atmosphere.

MAST3 overexpression model

At day 0, twelve aliquots of 1×10^6 TLR4 293-hTLR4A-MD2-CD14 HEK cells (InvivoGen, San Diego, CA, USA) were diluted in 2ml of complete medium (without antibiotics) and plated in individual wells of two 6-well plates. At day 1, cells were transiently transfected using 6ug of pCMV-XL6 vector containing *MAST3* cDNA (SC316595 Origene, Rockville, MD, USA) or 6ug of pCMV-AC-GFP vector (PS100010 Origene, Rockville, MD, USA) and 15 uL of lipofectamin 2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Cells were collected 24h later (at day 2) and frozen. For the time course assay, cells were similarly transfected, and collected and frozen at 0h, 8h, 12h and 24h after transfection and used for protein or RNA extraction.

MAST3 knockdown model

Small hairpin RNA (shRNA) was introduced into THP1 cells by spinoculation with Mission lentiviral particles (Sigma, St-Louis, MO, USA) according to manufacturer's instructions. At day 1, aliquots of 50,000 cells were diluted in 100 uL of complete medium. Cells were transduced with lentiviral transduction particules carrying a vector with shRNA targeting *MAST3* and a puromycin resistance gene (clone ID

TRCCN0000200003 from Sigma, St-Louis, MO, USA, CCGGCGAGCCTTTCTGCCGACACAGCTCGAGCTGTGTCGGCAGAAAGGCTC GTTTTTTGT) or control transduction particules carrying a vector with non-target shRNA and a puromycin resistance gene(SHC002V Sigma, St-Louis, MO, USA) at MOI =10. Three independent aliquots of cells were transduced with *MAST3* shRNA and three others with non-target control shRNA. Cells with virus-containing medium were centrifuged at 800 g for 30 minutes at 32 °C. Virus-containing medium was aspirated and cell pellets were resuspended in 100 ul of complete media by gently pipetting the pellets up and down, and each resuspended pellet was transferred to its own well in a 96-well tissue culture plate. The plate was returned to the tissue culture incubator at 37°C in a 5% CO₂ atmosphere. At day 4, puromycin (final concentration 0.5mg/mL) was added to the medium. At day 12, all cells in the non-transduced control wells were dead. Cells were further propagated in medium containing puromycin and pools of transduced THP1 clones were used for differentiation. The 6 pools of transduced THP1 cells (3 *MAST3* knockdowns, 3 controls) were each separated in aliquots. One was kept untreated, one was treated with phorbol myristate acetate (PMA; P1585 Sigma, St-Louis, MO, USA) (10 ng/mL) for 48h and the other ones were treated with PMA (10 ng/mL) for 48h and with lipopolysaccharide (LPS; L2654 Sigma, St-Louis, MO, USA) (10 ng/mL) for 3h, 6h or 24h . Cells were then collected for proteins and RNA extraction.

RNA extraction, quantification and reverse-transcription (RT) in cell model

RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. RNA was quantified using the RNA 6000 Nano Kit (Agilent) and a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). All samples had RIN above 9.5. RNA samples (2 ug per rxn) were reverse-transcribed (random hexamers) using the High Capacity cDNA Reverse Transcription kit (ABI, Carlsbad, CA, USA).

Proteins and Western Blots

Proteins were extracted from cell lysate using RIPA Buffer (899100 Thermo, Waltham, MA, USA) according to the manufacturer's instructions. Proteins were quantified using Pierce BCA Protein Assay Kit (23225 Thermo, Waltham, MA, USA) according to the manufacturer's instructions. The absorbance was read on a Synergy 2 plate reader (Biotek, Winoosky, VT, USA). Western blots were performed according to standard methods, antibodies used for immunoblotting were anti-MAST3 (ab64952, Abcam, Cambridge, MA, USA), anti-actin (ab3280, Abcam, Cambridge, MA, USA), and goat anti-rabbit-HRP (CLCC43007, Cederlane, Burlington, NC, USA).

qPCR

Gene expression was monitored with 10 ng of total cDNA and SybrGreen (Invitrogen Carlsbad, CA, USA) based qPCR using the Mx3005P Real-Time PCR system (Agilent, Santa Clara, CA, USA). Cycling conditions were set at 95°C for 10 minutes, then 40 cycles of 95°C for 30 seconds and 60°C for 1 minute and 1 cycle of 55°C for 30 seconds and 95°C for 30 seconds. Samples were run in duplicates. Gene expression levels were normalized to the *HPRT1* gene expression levels. Human gene specific primer sequences used were: *MAST3* forward (5'GCAGCGAAGTGGACTATGG3'), *MAST3* reverse (5'GATGGTATTCAGGAGAGATGGG3'), *HPRT1* forward (5'TGGCGTCGTGATTAGTGATG3'), *HPRT1* reverse (5'CAGAGGGCTACAATGTGATGG3') (Other sequences in Supplementary Table 4).

MAST3 model of overexpression: Whole-genome gene expression analysis

Twelve RNA samples from TLR4 293-hTLR4A-MD2-CD14 HEK cells (6 *MAST3* overexpressed samples and 6 controls) were sent to Genome Quebec Innovation Center (Montreal, QC, Canada) (GQ) to be analyzed using the HumanHT-12 v4 Expression BeadChip (Illumina, San Diego, CA, USA). Preliminary data was analyzed at GQ using the Genome Studio Software (Illumina, San Diego, CA, USA). We performed further quality control assessment and data normalization using the Lumi library included in the Bioconductor package^{30,31}. First, data were adjusted by log₂ transformation and global mean normalization. Then, differentially expressed genes between cells overexpressing

the *MAST3* gene and mock-transfected cells were determined using a one-tailed Welch's t-test calculated using the MeV software³². P values were corrected for multiple testing using the adjusted Bonferroni correction. Genes with significant expression change ($P_{\text{adjusted}} < 0.05$) were further studied.

A query of the Database for Annotation, Visualization and Integrated Discovery (DAVID)³³ was used to identify functional pathways based on Gene Ontology biological processes.

NF- κ B activity Assays

NF- κ B activity was monitored using ELISA strip plates from the Transcription Factor kit for NF- κ B p65 (89858 Thermo, Waltham, MA, USA) according to the manufacturer's instructions. For THP1 cells and TLR4 HEK cells, 6 and 20 μ g of total protein lysate was used per well, respectively. For each of the three biological replicates at each point, luminescence from two independent wells was read on a Synergy 2 plate reader (Biotek, Winoosky, VT, USA).

Whole-Genome Expression in Patient Biopsies

Biopsies were collected from patients suffering from ulcerative colitis (UC) and healthy controls at Royal Victoria and Maisonneuve-Rosemont Hospitals in Montreal (QC, Canada). Patients had to be older than 18 year old and diagnosed with UC for at least 5 years. Patients under biological treatment (i.e.: Infliximab) were excluded from the study. Biopsies were collected during routine exams (patients) and during cancer screening exams (controls). Written informed consent was obtained from all participants and ethics approval was granted in each of the participating institutions. For patients, biopsies was taken from inflamed (on the margin of ulcerations) and noninflamed region. Biopsies were taken from the rectum (controls and most patients) and colon (some patients) and stored in RNAlater (Qiagen, Valencia, CA, USA) until RNA extraction. RNA was extracted in the following 24h using the RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer instructions. RNA was quantified using the RNA 6000 Nano Kit (Agilent) and a Bioanalyzer 2100 (Agilent). RIN was above 8. One μ g of RNA was sent to GQ and whole-genome expression was

evaluated using the Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, USA).

Statistical analyses for the enrichment analyses

Expression data from the biopsies was normalized using quantile normalization with the RMA algorithm within each analyzed pair of classes (control+ noninflamed and noninflamed+inflamed)^{34, 35}. Differential expression between each class was tested using Welch T tests. Data structure (inflamed and non-inflamed samples that came from the same individual and batches of Affymetrix (Santa Clara, CA, USA) chips hybridized on 2 separate occasions) was corrected using permutations (1000 permutations). Enrichment in the MAST3-regulated gene set was tested using Fisher's exact test. For assessment of the differential expression of individual genes, P values were further adjusted for FWER using a Bonferonni-type correction that incorporates correlation within the data.

Acknowledgements

The authors would like to thank Guillaume Lettre and Marcia Budarf for their careful review of the manuscript. This work was funded by grants from the U.S. National Institute of Diabetes and Digestive and Kidney Diseases (DK062432) and the Crohn's and Colitis Foundation of America (SRA512) awarded to JDR.

CL holds a doctoral training award from the Fond de Recherche en Santé du Québec.

Tables and Figures

Table 1. Differentially expressed genes (Fold increase>2) in the *MAST3* overexpression model

Gene	chr	P value	Fold increase	Function
<i>MAST3</i>	19	3.48E-04	42.00	
<i>CCL20</i>	2	1.31E-10	38.13	C, NP
<i>IL8</i>	4	6.09E-06	13.89	C, NP
<i>CD44</i>	11	1.13E-02	6.41	AM, NP
<i>IFIT2</i>	10	1.20E-03	6.23	I
<i>IER3</i>	6	4.05E-04	5.28	RN, NP
<i>NFKBIA</i>	14	4.41E-02	4.98	RN, NP
<i>IL32</i>	16	8.70E-04	4.60	C, NP
<i>ISG15</i>	1	6.82E-04	3.95	I
<i>TNFAIP3</i>	6	5.81E-04	3.83	RN, NP
<i>NFKBIZ</i>	3	8.69E-05	3.74	RN, NP
<i>IFIT1</i>	10	8.40E-06	3.66	I
<i>CD70</i>	19	5.98E-04	3.55	C, RN
<i>TNC</i>	9	3.97E-02	3.49	AM, NP
<i>SLC25A24</i>	1	3.82E-05	3.07	
<i>CXCL2</i>	4	5.62E-03	2.92	C, NP
<i>TNFRSF9</i>	1	1.40E-03	2.81	RN, NP
<i>IFIT3</i>	10	1.78E-04	2.80	I
<i>SGPP2</i>	2	3.89E-03	2.61	NP
<i>TNF</i>	6	2.77E-04	2.55	C, NP
<i>GAL</i>	11	2.53E-04	2.44	NP
<i>CD83</i>	6	5.69E-07	2.41	NP

<i>TNFRSF12A</i>	16	3.84E-03	2.35	RN, AM
<i>PMAIP1</i>	18	1.17E-03	2.28	I
<i>NFKB1</i>	4	2.87E-04	2.26	RN, NP
<i>LY96</i>	8	1.27E-02	2.22	RN
<i>CDK6</i>	7	5.01E-04	2.18	NP
<i>CALCB</i>	11	4.38E-03	2.17	
<i>TMOD1</i>	9	7.18E-05	2.00	AM

chr: chromosome. C: pro-inflammatory cytokine. I: induced by interferons. NP: regulated by NF- κ B. RN: regulator of NF- κ B. AM: involved in cell adhesion and/or migration.

Table 2. Effect of a *MAST3* stable knockdown (KD) on expression of genes in stimulated THP1 cells confirms results from the overexpression model.

Gene	Effect of <i>MAST3</i> KD (% of expression of non target KD control)*	
		P
<i>IL8</i>	55%	0.01
<i>IER3</i>	41%	0.02
<i>CD70</i>	76%	0.03
<i>CCL20</i>	no detected effect	0.80
<i>ISG15</i>	75%	0.02
<i>LY96</i>	40%	0.01
<i>NFKBIA</i>	92%	0.02
<i>TNF</i>	no detected effect	0.47
<i>TNFAIP3</i>	75%	0.02

-: no significant effect of the KD of *MAST3* on expression.

*: THP1 cells were differentiated using PMA and treated with LPS for 3h (*IER3*, *TNF*) or 24h (other genes).

Table 3. MAST3 stable knockdown (KD) influences the expression of immune genes in stimulated THP1 cells.

Gene	Effect of MAST3 KD (% of expression of non target KD control)*	P
<i>IL10</i>	210%	0.003
<i>IL12B</i>	200%	0.03
<i>IL18</i>	220%	0.04
<i>IL1A</i>	60%	0.01
<i>IL1B</i>	70%	0.03
<i>TGFB1</i>	65%	0.03
<i>TRAF6</i>	65%	0.002

*: THP1 cells were differentiated using PMA and treated with LPS for 6h (*IL10*, *IL1B*, *TGFB1*) or 24h (*IL12B*, *IL18*, *IL1A*, *TRAF6*).

Table 4. MAST3-regulated gene set enrichment tests in patient biopsies

	BC vs BPN		BPN vs BPI	
	MAST3 controlled gene set	All genes* on microarray	MAST3 controlled gene set	All genes* on microarray
P* < 0.05	3	7045	10	8498
Total	14	24066	14	24607
%	21.4	29.2	71.4	34.5
Enrichment P	0.825		0.005	

* Probe sets on the Affymetrix chip are used in lieu of genes. Some genes are represented by more than one probe set. P* = P is the empirical P after 1000 permutations. BC: biopsies from healthy control. BPN: biopsies from patients taken from non-inflamed region. BPI: biopsies from patients taken from inflamed region.

Table 5. Comparison of the expression of the MAST3-regulated genes between BPN and BPI

	Exp. BPN	Exp. BPI	P
<i>CD44</i>	8.01	8.98	0.001
<i>TNC</i>	6.90	8.31	0.017
<i>ISG15</i>	9.02	9.96	0.02
<i>NFKBIZ</i>	10.56	11.74	0.02
<i>IL8</i>	5.89	7.66	0.027
<i>NFKBIA</i>	10.57	11.11	0.061
<i>TNFAIP3</i>	9.47	9.96	0.07
<i>IER3</i>	10.22	10.89	0.137
<i>CCL20</i>	10.33	11.57	0.27
<i>IFIT2</i>	8.12	8.56	0.505
<i>SLC25A24</i>	11.61	11.44	0.609
<i>CD70</i>	6.72	6.82	1
<i>IFIT1</i>	7.81	7.80	1
<i>IL32</i>	9.05	9.09	1

Exp.: mean expression. BPN: biopsies from patients taken from non-inflamed region. BPI: biopsies from patients taken from inflamed region. P: Empirical P corrected for multiple testing.

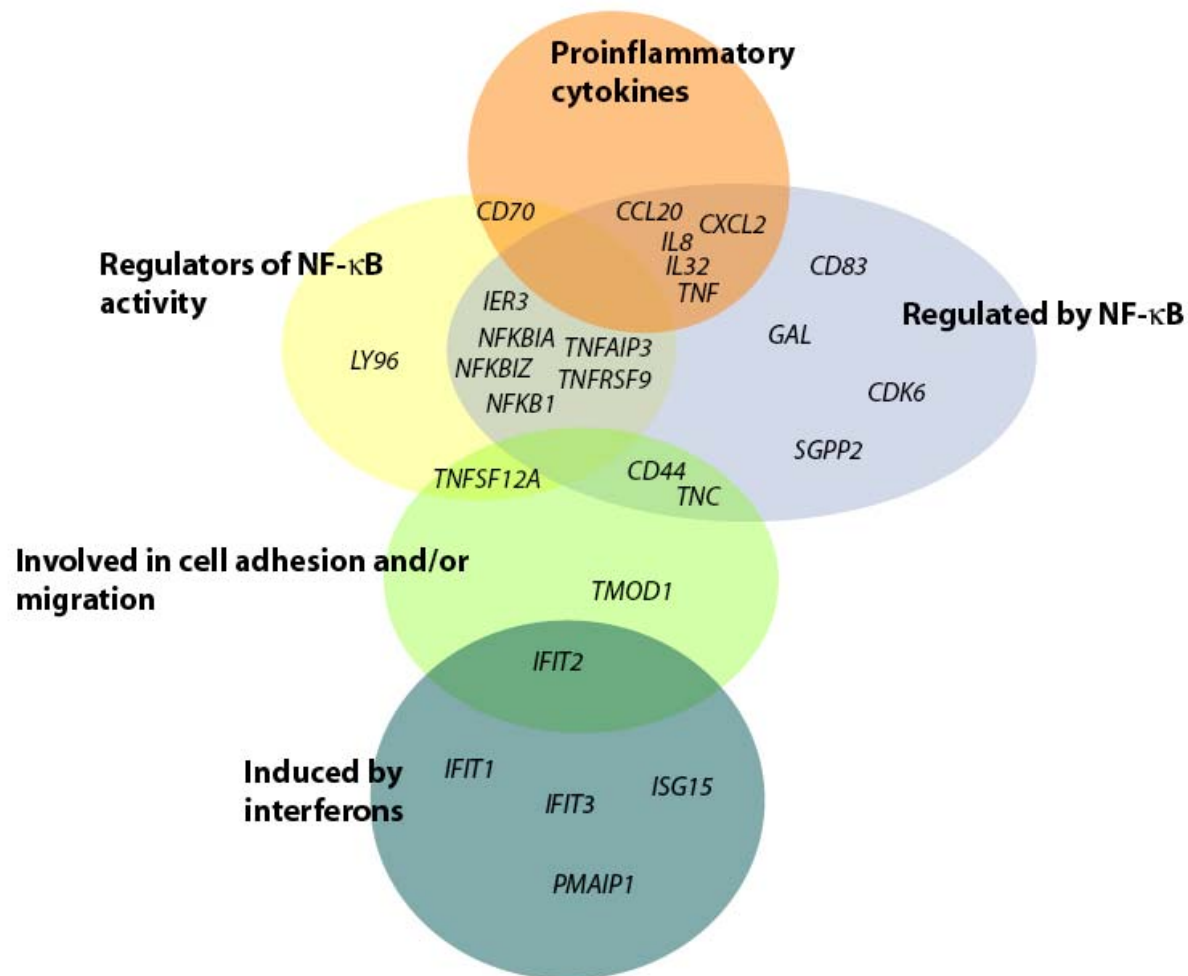
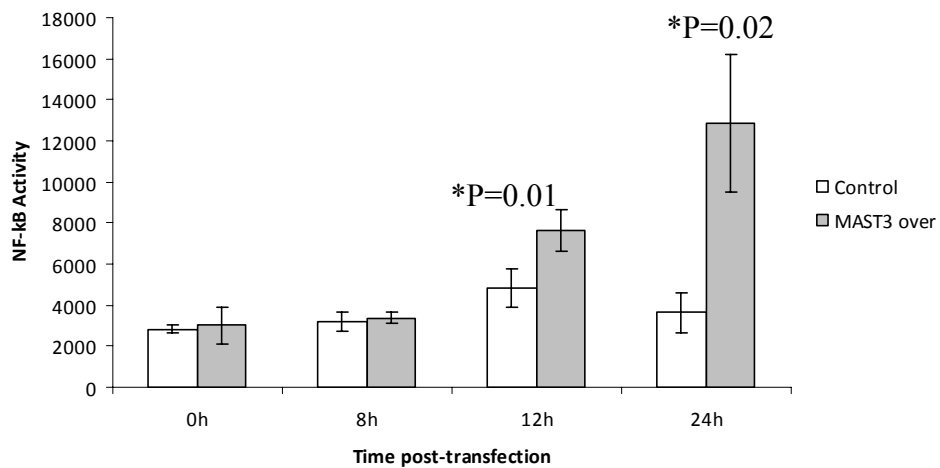


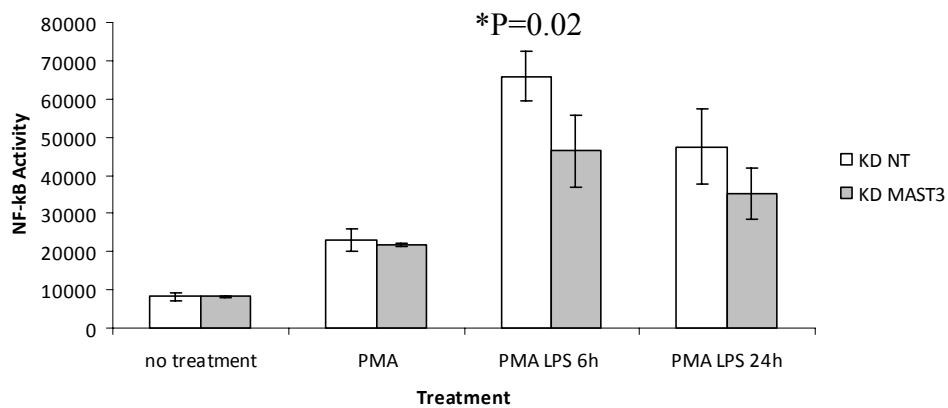
Figure 1. MAST3-regulated genes are involved in critical immune functions.

Genes in the MAST3-regulated gene set (genes upregulated by 2-fold or more in cells that overexpress the *MAST3* gene) were classified according to their function recovered from literature searches. The functions of the genes mainly cluster around the activity of NF-κB. Sixty-one percent of the 28 genes have experimentally confirmed NF-κB binding site in their promoter and are regulated by NF-κB. Thirty-two percent regulates the activity of NF-κB. Twenty-one percent are pro-inflammatory cytokines. Other functions include cell adhesion and migration and interferon mediated defense against pathogen invasion and tumor progression. Two genes out of the 28, *CALCB* and *SLC25A24* do not seem to be directly involved in immune functions and were not included in the diagram.

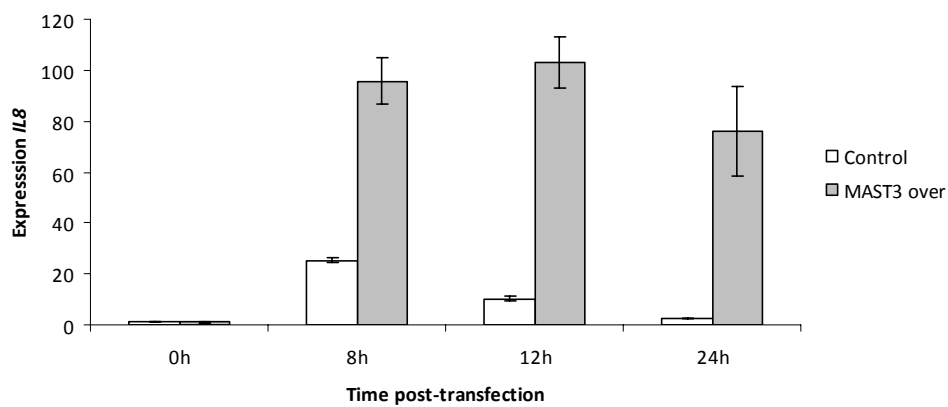
A.



B.



C.



D.

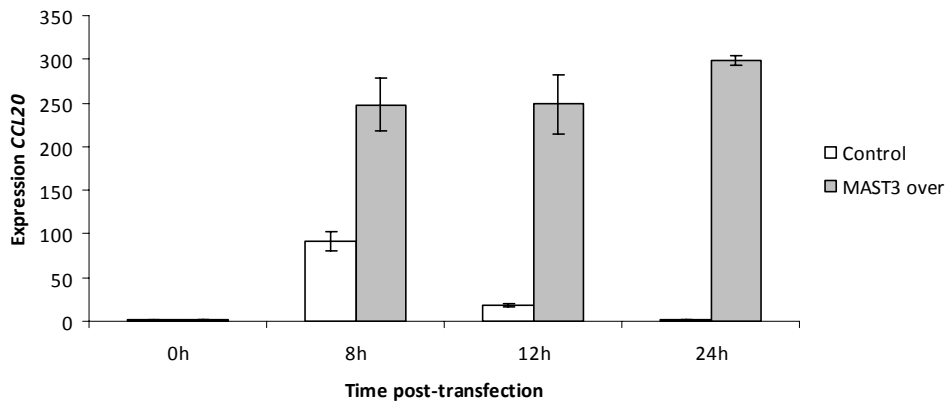


Figure 2. *MAST3* regulates gene expression through NF- κ B.

A. The overexpression of *MAST3* increases the activity of NF- κ B. The graph shows the activity of NF- κ B tested in HEK293 cells that overexpress the *MAST3* gene at different times post-transfection (*MAST3* over). The increase in activity is significantly detected after 12h (* $P=0.01$). B. The knockdown of *MAST3* decreases the activity of NF- κ B. The graph shows the activity of NF- κ B tested in THP1 cells at different stages of differentiation and stimulation following different treatment. The decrease in activity is significantly detected in phorbol 12-myristate 13-acetate (PMA) differentiated cells after 6h of lipopolysaccharide (LPS) stimulation (* $P=0.02$). C. and D. Expression of *MAST3*-regulated genes *IL8* (C) and *CCL20* (D) post-transfection in cells overexpressing the *MAST3* gene. Expression is normalized to *HPRT1* expression. Error bar represent standard deviation of three independent biological replicates. KD *MAST3*: knockdown *MAST3*, KD NT: non target knockdown control.

Supplementary Data

Supplementary Table 1. All gene significantly differentially expressed ($P < 0.05$) in cells overexpressing the *MAST3* gene.

RefSeq_Accession	Symbol	Expression x(OE/CTRL)	Adj P value
NM_015016	<i>MAST3</i>	42.00	3.48E-04
NM_004591	<i>CCL20</i>	38.13	1.31E-10
NM_000584	<i>IL8</i>	14.07	6.09E-06
NM_000584	<i>IL8</i>	13.89	3.97E-06
NM_001001391	<i>CD44</i>	6.41	0.011278061
NM_001547	<i>IFIT2</i>	6.23	0.001204265
NM_003897	<i>IER3</i>	5.28	4.05E-04
NM_020529	<i>NFKBIA</i>	4.98	0.044087354
NM_001012633	<i>IL32</i>	4.60	8.70E-04
NM_005101	<i>ISG15</i>	3.95	6.82E-04
NM_006290	<i>TNFAIP3</i>	3.83	5.81E-04
NM_001005474	<i>NFKBIZ</i>	3.74	8.69E-05
NM_001548	<i>IFIT1</i>	3.66	8.40E-06
NM_001252	<i>CD70</i>	3.55	5.98E-04
NM_002160	<i>TNC</i>	3.49	0.03971354
NM_013386	<i>SLC25A24</i>	3.07	3.82E-05
XM_941665	<i>LOC387763</i>	2.98	1.96E-05
NM_002089	<i>CXCL2</i>	2.92	0.005624934
NM_001561	<i>TNFRSF9</i>	2.81	0.001399281
NM_001031683	<i>IFIT3</i>	2.80	1.78E-04
XM_938742	<i>SGPP2</i>	2.61	0.003888478
NM_000594	<i>TNF</i>	2.55	2.77E-04
NM_015973	<i>GAL</i>	2.44	2.53E-04
NM_001040280	<i>CD83</i>	2.41	5.69E-07
NM_016639	<i>TNFRSF12A</i>	2.35	0.003839672
NM_002160	<i>TNC</i>	2.29	0.03776952

NM_021127	<i>PMAIP1</i>	2.28	0.001173748
NM_003998	<i>NFKB1</i>	2.26	2.87E-04
NM_004233	<i>CD83</i>	2.23	1.28E-06
NM_015364	<i>LY96</i>	2.22	0.012744945
NM_001259	<i>CDK6</i>	2.18	5.01E-04
NM_000728	<i>CALCB</i>	2.17	0.00437663
NM_203418	<i>RCAN1</i>	2.12	4.74E-04
NM_003275	<i>TMOD1</i>	2.00	7.18E-05
NM_203417	<i>RCAN1</i>	1.99	0.001139019
NM_014365	<i>HSPB8</i>	1.97	0.028745715
NM_022913	<i>GPBP1</i>	1.93	0.028743487
NR_029858	<i>MIR302C</i>	1.91	0.001760545
NM_001099678	<i>LRRC58</i>	1.91	0.014489579
NM_021127	<i>PMAIP1</i>	1.90	0.018972693
NM_003764	<i>STX11</i>	1.90	4.92E-04
NM_001554	<i>CYR61</i>	1.89	1.63E-04
NM_001012636	<i>IL32</i>	1.89	0.007492953
NM_006509	<i>RELB</i>	1.87	0.001447181
NM_052871	<i>MGC4677</i>	1.87	0.003494747
NM_014904	<i>RAB11FIP2</i>	1.86	0.002218478
NM_002982	<i>CCL2</i>	1.84	1.00E-05
NM_001040619	<i>ATF3</i>	1.83	0.008344874
NM_031412	<i>GABARAPL1</i>	1.83	1.63E-04
AJ420516		1.83	0.036260877
NM_004048	<i>B2M</i>	1.82	0.009514864
NM_033027	<i>AXUD1</i>	1.82	0.003138402
NM_005946	<i>MT1A</i>	1.82	0.018883374
NM_005903	<i>SMAD5</i>	1.81	0.003822554
NM_025079	<i>ZC3H12A</i>	1.81	0.009804687
NM_012323	<i>MAFF</i>	1.77	0.02726992
NM_025047	<i>ARL14</i>	1.76	0.010440555
NM_152858	<i>WTAP</i>	1.72	0.024817396

NM_003190	<i>TAPBP</i>	1.72	0.011685941
NM_016823	<i>CRK</i>	1.70	0.003074495
NM_001002857	<i>ANXA2</i>	1.69	0.001582759
NM_001425	<i>EMP3</i>	1.69	0.004254707
NM_001042369	<i>TROVE2</i>	1.66	0.01619699
CR622439		1.66	0.046200126
NM_015892	<i>CHST15</i>	1.66	0.003091501
NM_017736	<i>THUMPD1</i>	1.64	0.025530258
NM_003663	<i>CGGBP1</i>	1.63	0.014235752
NM_020466	<i>LYRM2</i>	1.62	0.015284507
NM_005239	<i>ETS2</i>	1.62	0.014084131
NM_014330	<i>PPP1R15A</i>	1.61	0.04305981
NM_002575	<i>SERPINB2</i>	1.60	0.027823232
NR_002323	<i>TUG1</i>	1.59	0.04239271
NM_199229	<i>RPE</i>	1.58	0.015281103
NM_006287	<i>TFPI</i>	1.57	0.019160675
NM_004251	<i>RAB9A</i>	1.57	1.16E-04
NM_080655	<i>C9orf30</i>	1.57	0.010326821
NM_138571	<i>HINT3</i>	1.56	0.003569022
NM_016823	<i>CRK</i>	1.56	0.019103818
NM_004728	<i>DDX21</i>	1.56	0.008450904
NR_003098	<i>SNHG1</i>	1.56	0.045457955
XM_927451	<i>LOC653226</i>	1.56	0.001124901
NM_002575	<i>SERPINB2</i>	1.55	0.030038996
NM_001660	<i>ARF4</i>	1.55	0.01218624
NM_001018069	<i>SERBP1</i>	1.54	0.024489302
NM_004776	<i>B4GALT5</i>	1.54	0.029393006
NM_201559	<i>FOXO3</i>	1.54	0.045295794
NM_001037174	<i>ARL5A</i>	1.54	0.004307957
NM_004417	<i>DUSP1</i>	1.53	0.006008911
NM_005139	<i>ANXA3</i>	1.53	0.012961811
NM_175624	<i>RAB3IP</i>	1.53	0.001175251

NM_002796	<i>PSMB4</i>	1.52	0.037042618
NM_031465	<i>C12orf32</i>	1.52	0.03357662
NM_153712	<i>TTL</i>	1.52	0.024349106
XM_927609	<i>LOC221710</i>	1.51	0.013918469
NM_006925	<i>SFRS5</i>	1.51	9.75E-04
NM_005444	<i>RQCD1</i>	1.50	3.66E-04
NM_001012756	<i>ZNF260</i>	1.50	0.048565622
NM_003877	<i>SOCS2</i>	1.49	0.009984685
NM_020803	<i>KLHL8</i>	1.48	2.32E-04
NM_174908	<i>CCDC50</i>	1.48	0.005193724
NM_032802	<i>SPPL2A</i>	1.48	4.61E-04
NM_014043	<i>CHMP2B</i>	1.47	0.01147545
BX538337		1.47	0.008317753
NM_017759	<i>INO80D</i>	1.47	0.003657858
NM_201557	<i>FHL2</i>	1.47	0.003084054
NM_001007067	<i>SDCBP</i>	1.46	0.03833343
NM_178517	<i>PIGW</i>	1.45	0.006815474
NR_001561	<i>CYCSL1</i>	1.45	0.012866
NM_181656	<i>C17orf58</i>	1.45	0.001354858
XM_937367	<i>IL7R</i>	1.45	0.045610905
NM_020412	<i>CHMP1B</i>	1.44	0.044499524
NM_025147	<i>COQ10B</i>	1.44	0.011245502
NM_005859	<i>PURA</i>	1.44	0.002112913
NM_022770	<i>GINS3</i>	1.44	0.003319947
NM_181656	<i>C17orf58</i>	1.44	0.001692695
NM_016078	<i>FAM18B</i>	1.43	0.001785247
NM_172177	<i>MRPL42</i>	1.43	0.024210248
NM_002908	<i>REL</i>	1.43	0.04924718
NM_021809	<i>TGIF2</i>	1.42	0.011688842
NM_001300	<i>KLF6</i>	1.41	0.017037055
NM_000617	<i>SLC11A2</i>	1.40	0.024676215
NM_004821	<i>HAND1</i>	0.71	0.019484956

NM_004463	<i>FGD1</i>	0.71	0.008387891
NM_005001	<i>NDUFA7</i>	0.70	0.044010684
NM_000820	<i>GAS6</i>	0.69	0.01588695
NM_144641	<i>PPM1M</i>	0.69	0.019506399
NM_054012	<i>ASS1</i>	0.69	0.027822183
NM_014214	<i>IMPA2</i>	0.68	0.012939415
NM_001013653	<i>LRRC26</i>	0.68	0.035538845
NM_005733	<i>KIF20A</i>	0.68	0.03582707
NM_014909	<i>VASH1</i>	0.68	0.004221957
NM_006383	<i>CIB2</i>	0.68	0.026606318
NM_032118	<i>WDR54</i>	0.67	0.003487259
NM_006855	<i>KDEL3</i>	0.67	0.024521606
NM_016609	<i>SLC22A17</i>	0.66	0.042047378
NM_002084	<i>GPX3</i>	0.66	0.012220169
NM_000599	<i>IGFBP5</i>	0.66	0.028279658
NM_016286	<i>DCXR</i>	0.66	0.027213806
NM_152739	<i>HOXA9</i>	0.66	0.013479603
NM_001024594	<i>C1orf53</i>	0.66	0.045547493
NM_004474	<i>FOXD2</i>	0.65	0.01620012
NM_000820	<i>GAS6</i>	0.65	0.001691935
NM_052839	<i>PANX2</i>	0.65	0.038621623
NM_002579	<i>PALM</i>	0.65	0.04225034
NM_001007256	<i>KLHDC9</i>	0.65	0.01139942
NM_019106	<i>SEPT3</i>	0.64	0.03553287
NM_000597	<i>IGFBP2</i>	0.64	0.027618175
NM_024040	<i>CUEDC2</i>	0.64	0.037346654
XR_018125	<i>LOC441154</i>	0.63	0.040506568
NM_080760	<i>DACH1</i>	0.62	0.004544208
NM_021004	<i>DHRS4</i>	0.62	0.019803086
NM_002048	<i>GAS1</i>	0.62	0.030644927
NM_004390	<i>CTSH</i>	0.62	0.044033036
NM_002528	<i>NTHL1</i>	0.62	0.025702296

NM_007129	<i>ZIC2</i>	0.62	0.016270809
NM_005896	<i>IDH1</i>	0.61	0.004873429
NR_000024	<i>SNORD46</i>	0.60	0.00787179
NM_002168	<i>IDH2</i>	0.59	8.11E-04
NR_024103	<i>C17orf93</i>	0.59	0.006295137
NM_021020	<i>LZTS1</i>	0.59	0.003652763
NM_003088	<i>FSCN1</i>	0.58	0.003710913
NM_001002018	<i>HCFC1R1</i>	0.56	0.01102881
NM_018390	<i>PLCXD1</i>	0.55	0.001204323
NM_005077	<i>TLE1</i>	0.55	0.001058865
NM_000477	<i>ALB</i>	0.54	0.001407475
NM_002775	<i>HTRA1</i>	0.52	0.046296883
NM_019058	<i>DDIT4</i>	0.52	0.004551901
NM_032772	<i>ZNF503</i>	0.50	0.007911245
NM_000848	<i>GSTM2</i>	0.48	0.0389311
NM_000561	<i>GSTM1</i>	0.45	0.004690235
NM_003970	<i>MYOM2</i>	0.44	0.024349527
NR_028372	<i>RNU6-15</i>	0.41	0.006706168
XM_210876	<i>C10orf41</i>	0.40	0.001170356
NR_004394	<i>RNU6-1</i>	0.40	0.023693664
NR_003051	<i>RMRP</i>	0.31	0.001427169
U62823		0.29	0.013889274
NR_023344	<i>RNU6ATAC</i>	0.28	0.001093215
NR_006881	<i>SNORD3C</i>	0.27	0.001543414
NR_004402	<i>RNU1F1</i>	0.27	5.44E-04
NR_002312	<i>RPPH1</i>	0.25	1.10E-04
NR_004430	<i>RNU1A3</i>	0.25	2.34E-04
NR_006880	<i>SNORD3A</i>	0.22	4.84E-04
NR_006882	<i>SNORD3D</i>	0.21	2.62E-04
XM_942124	<i>KIAA1666</i>	0.20	0.005818738
NR_001571	<i>RNY5</i>	0.19	0.002492848
NR_004400	<i>RNU1-5</i>	0.19	2.30E-04

NR_004426	<i>RNU1G2</i>	0.17	0.001273187
NR_004408	<i>RNU1-3</i>	0.17	4.20E-04
XM_001713808	<i>LOC100132564</i>	0.12	2.59E-07
XM_001713810	<i>LOC100134364</i>	0.08	2.10E-08
XM_001713809	<i>LOC100132394</i>	0.08	2.79E-08
NR_003287	<i>LOC100008589</i>	0.07	2.98E-08
NR_023371	<i>RN5S9</i>	0.05	2.10E-08
NR_001445	<i>RN7SK</i>	0.05	2.05E-05
NR_001445	<i>RN7SK</i>	0.04	5.43E-05

Supplementary Table 2. Major Gene Ontology biological processes involving genes upregulated in our MAST3 overexpression model.

Term	Coun t	%	PValue	Genes	List	Pop Hits	Pop Total	Fold Enrich.	Bonferroni	Benjamini	FDR
defense response	11	39.29	2.35E-08	<i>CD83, NFKBIZ, TNF, IL8, CD44, CCL20, LY96, CXCL2, IL32, NFKB1, GAL</i>	24	615	13528	10.08	1.57E-05	1.57E-05	3.52E-05
inflammatory response	9	32.14	3.65E-08	<i>NFKBIZ, TNF, IL8, CD44, CCL20, LY96, CXCL2, NFKB1, GAL</i>	24	325	13528	15.61	2.44E-05	1.22E-05	5.46E-05
response to wounding	10	35.71	1.02E-07	<i>NFKBIZ, TNF, IL8, CD44, CCL20, LY96, TNC, CXCL2, NFKB1, GAL</i>	24	530	13528	10.64	6.79E-05	2.26E-05	1.52E-04
regulation of apoptosis	10	35.71	3.39E-06	<i>TNFRSF9, IER3, TNF, CD44, NFKBIA, CD70, NFKB1, PMAIP1, TNFAIP3, GAL</i>	24	804	13528	7.01	0.0022645	5.67E-04	0.005086
regulation of programmed cell death	10	35.71	3.68E-06	<i>TNFRSF9, IER3, TNF, CD44, NFKBIA, CD70, NFKB1, PMAIP1, TNFAIP3, GAL</i>	24	812	13528	6.94	0.00245714	4.92E-04	0.005519
regulation of cell death	10	35.71	3.80E-06	<i>TNFRSF9, IER3, TNF, CD44, NFKBIA, CD70, NFKB1, PMAIP1, TNFAIP3, GAL</i>	24	815	13528	6.92	0.0025329	4.23E-04	0.00569

Supplementary Table 3. Comparison of the expression of gene from the MAST3 regulated gene set between BC and BPN

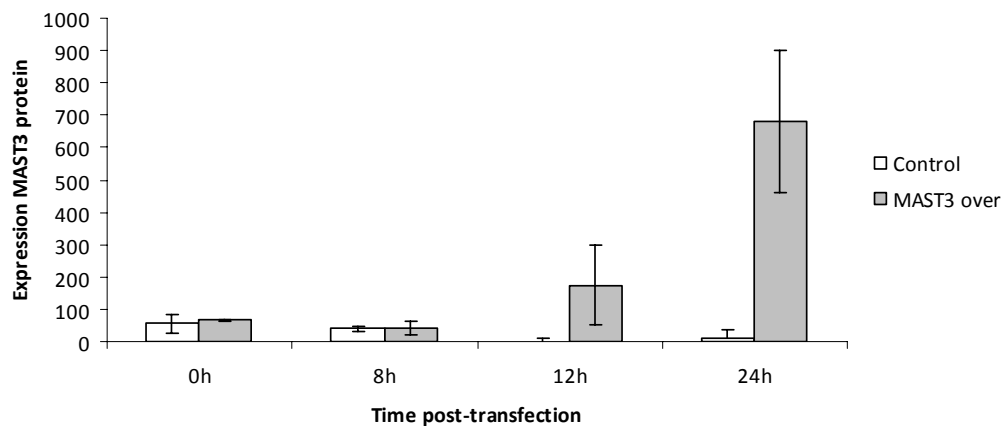
	Exp. BC	Exp. BPN	P
<i>IL8</i>	4.63	5.25	0.308
<i>TNFAIP3</i>	9.19	9.51	0.344
<i>NFKBIZ</i>	8.19	8.63	0.533
<i>MAST3</i>	7.77	7.48	0.668
<i>IER3</i>	10.51	10.19	0.925
<i>CCL20</i>	9.90	10.31	0.997
<i>NFKBIA</i>	10.72	10.61	0.998
<i>CD44</i>	7.46	7.55	1
<i>CD70</i>	6.75	6.83	1
<i>IFIT1</i>	7.62	7.83	1
<i>IFIT2</i>	8.00	8.17	1
<i>IL32</i>	8.85	8.97	1
<i>ISG15</i>	9.02	9.20	1
<i>SLC25A24</i>	11.69	11.64	1
<i>TNC</i>	6.78	6.89	1

BC: biopsies from healthy control. BPN: biopsies from patients taken from non-inflamed region.

Supplementary Table 4. Sequences of the RT qPCR primers.

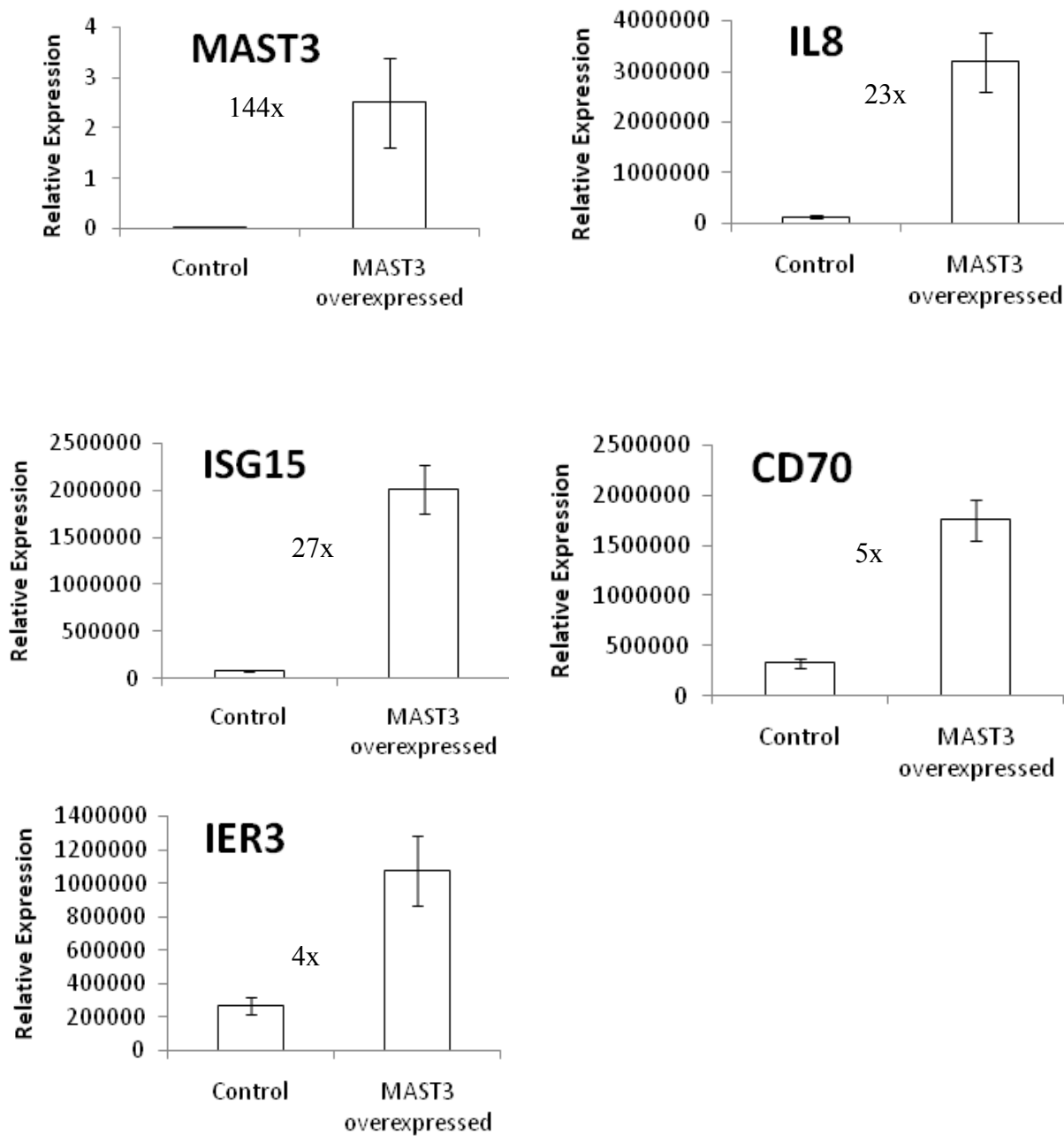
Gene	Orientation	Sequence
CCL20	F	CTCCACCTCTGCGGCGAATC
CCL20	R	AGATGATAGCATTGATGTCACAGC
CD44	F	CAACAACCTTCTACTCTGACATC
CD44	R	TGATAAGGAACGATTGACATTAG
CD70	F	ATCCAGCGCTTCGCACAGGC
CD70	R	CATGTAGATGCCATCACGAT
IER3	F	ACCATCGTCTTCTGCCAGA
IER3	R	AGTCACAGTTAGAAGGCGG
IFIT1	F	AGGATGAAGGACAGGAAG
IFIT1	R	GCAGTAAGACAGAAGTGG
IFIT2	F	ACCTCTGGACTGGCAATAG
IFIT2	R	CCTCTTCACCTTCTTCACG
IL32	F	CCCGAAGGTCTCTCTGATG
IL32	R	ATAAGCCGCCACTGTCTCC
ISG15	F	AGCCCACAGCCCACAGCCATG
ISG15	R	GTGGACAGCCAGACGCTGCT
NFKBIA	F	CTCCACTCCATCCTGAAG
NFKBIA	R	CTCCTGAGCATTGACATC
NFKBIZ	F	CAGCGGTGGAAGCGAAGG
NFKBIZ	R	TCATCAACAGGCGGACAGC
SLC25A24	F	CATTGCTGGTGTCTGTCTC
SLC25A24	R	TTGTCCTTCTTCAGTAAGTAAC
TNC	F	ACCTGAAGCCTGACACTG
TNC	R	CTTGCCATTCTCCATTCC
TNFAIP3	F	GGCGTTCAGGACACAGAC
TNFAIP3	R	TTCCAGTTCCGAGTATCATAGC
IL8	F	CGGAAGGAACCATCTCAC
IL8	R	CCACTCTCAATCACTCTCAG

F: forward, R: reverse



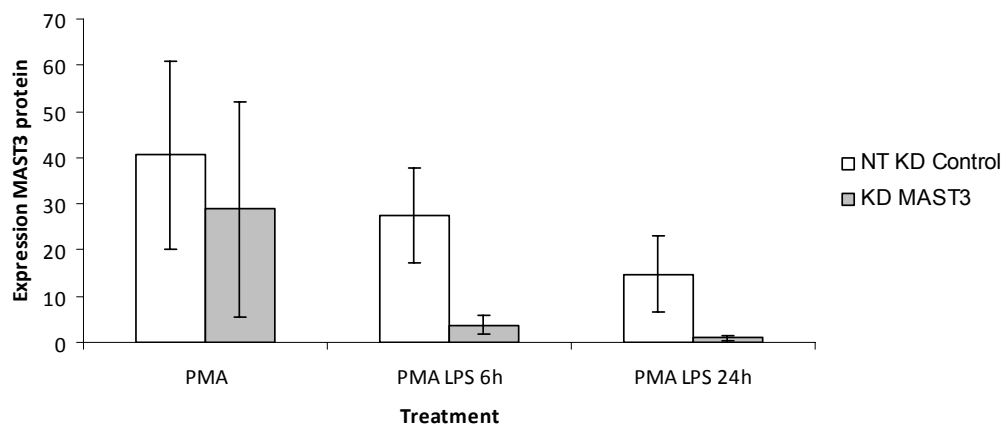
Supplementary Figure 1. Effect of *MAST3* overexpression on MAST3 protein levels at different times post-transfection.

Protein levels are relative to actin. Error bars represent the standard deviation between 3 independent biological replicates.



Supplementary Figure 2. RT qPCR validation of expression chip results.

The same RNA samples used on the chip were tested by qPCR. All gene tested were validated. The fold increase in expression is indicated on each bar chart. Expression is normalized to *HPRT1* expression. Error bars represent the standard deviation between 6 independent biological replicates.



Supplementary Figure 3. Effect of a stable knockdown of the *MAST3* gene on MAST3 protein levels.

Protein levels are relative to actin. There were no detectable MAST3 proteins in THP1 non stimulated cells. Error bars represent the standard deviation between 3 independent biological replicates. NT KD Control: non target knockdown control, KD MAST3: knockdown MAST3

References

1. Labbe C, Goyette P, Lefebvre C, Stevens C, Green T, Tello-Ruiz MK *et al.* MAST3: a novel IBD risk factor that modulates TLR4 signaling. *Genes Immun* 2008; **9**(7): 602-612.
2. Khor B, Gardet A & Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011; **474**(7351): 307-317.
3. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics* 2011; **43**(3): 246-252.
4. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics* 2011; **42**(12): 1118-1125.
5. Garland P, Quraishe S, French P & O'Connor V. Expression of the MAST family of serine/threonine kinases. *Brain Res* 2007.
6. Xiong H, Li H, Chen Y, Zhao J & Unkeless JC. Interaction of TRAF6 with MAST205 regulates NF-kappaB activation and MAST205 stability. *The Journal of biological chemistry* 2004; **279**(42): 43675-43683.
7. Ito T, Carson WFt, Cavassani KA, Connett JM & Kunkel SL. CCR6 as a mediator of immunity in the lung and gut. *Exp Cell Res* 2011; **317**(5): 613-619.
8. Daig R, Andus T, Aschenbrenner E, Falk W, Scholmerich J & Gross V. Increased interleukin 8 expression in the colon mucosa of patients with inflammatory bowel disease. *Gut* 1996; **38**(2): 216-222.
9. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**(6837): 599-603.
10. Li J, Moran T, Swanson E, Julian C, Harris J, Bonen DK *et al.* Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum Mol Genet* 2004; **13**(16): 1715-1725.
11. Ferreiro DU & Komives EA. Molecular mechanisms of system control of NF-kappaB signaling by IkappaBalpha. *Biochemistry* 2010; **49**(8): 1560-1567.

12. Heyninck K & Beyaert R. The cytokine-inducible zinc finger protein A20 inhibits IL-1-induced NF-kappaB activation at the level of TRAF6. *FEBS letters* 1999; **442**(2-3): 147-150.
13. O'Reilly SM & Moynagh PN. Regulation of Toll-like receptor 4 signalling by A20 zinc finger protein. *Biochem Biophys Res Commun* 2003; **303**(2): 586-593.
14. Park BS, Song DH, Kim HM, Choi BS, Lee H & Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 2009; **458**(7242): 1191-1195.
15. Arlt A, Rosenstiel P, Kruse ML, Grohmann F, Minkenberg J, Perkins ND *et al.* IEX-1 directly interferes with RelA/p65 dependent transactivation and regulation of apoptosis. *Biochimica et biophysica acta* 2008; **1783**(5): 941-952.
16. Billmann-Born S, Till A, Arlt A, Lipinski S, Sina C, Latiano A *et al.* Genome-wide expression profiling identifies an impairment of negative feedback signals in the Crohn's disease-associated NOD2 variant L1007fsinsC. *J Immunol* 2011; **186**(7): 4027-4038.
17. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 2005; **307**(5710): 731-734.
18. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics* 2008; **40**(8): 955-962.
19. Braegger CP, Nicholls S, Murch SH, Stephens S & MacDonald TT. Tumour necrosis factor alpha in stool as a marker of intestinal inflammation. *Lancet* 1992; **339**(8785): 89-91.
20. Reinecker HC, Steffen M, Witthoef T, Pflueger I, Schreiber S, MacDermott RP *et al.* Enhanced secretion of tumour necrosis factor-alpha, IL-6, and IL-1 beta by isolated lamina propria mononuclear cells from patients with ulcerative colitis and Crohn's disease. *Clin Exp Immunol* 1993; **94**(1): 174-181.
21. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet* 2002; **359**(9317): 1541-1549.
22. Sands BE, Anderson FH, Bernstein CN, Chey WY, Feagan BG, Fedorak RN *et al.* Infliximab maintenance therapy for fistulizing Crohn's disease. *N Engl J Med* 2004; **350**(9): 876-885.

23. Lichtenstein GR, Bala M, Han C, DeWoody K & Schaible T. Infliximab improves quality of life in patients with Crohn's disease. *Inflamm Bowel Dis* 2002; **8**(4): 237-243.
24. Targan SR, Hanauer SB, van Deventer SJ, Mayer L, Present DH, Braakman T *et al.* A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. *N Engl J Med* 1997; **337**(15): 1029-1035.
25. Jiang D, Chen Y & Schwarz H. CD137 induces proliferation of murine hematopoietic progenitor cells and differentiation to macrophages. *J Immunol* 2008; **181**(6): 3923-3932.
26. Halstead ES, Mueller YM, Altman JD & Katsikis PD. In vivo stimulation of CD137 broadens primary antiviral CD8+ T cell responses. *Nat Immunol* 2002; **3**(6): 536-541.
27. Arch RH & Thompson CB. 4-1BB and Ox40 are members of a tumor necrosis factor (TNF)-nerve growth factor receptor subfamily that bind TNF receptor-associated factors and activate nuclear factor kappaB. *Mol Cell Biol* 1998; **18**(1): 558-565.
28. Lim HW, Lee J, Hillsamer P & Kim CH. Human Th17 cells share major trafficking receptors with both polarized effector T cells and FOXP3+ regulatory T cells. *J Immunol* 2008; **180**(1): 122-129.
29. 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): 1152-1167.
30. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**(10): R80.
31. Du P, Kibbe WA & Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008; **24**(13): 1547-1548.
32. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA *et al.* TM4 microarray software suite. *Methods Enzymol* 2006; **411**: 134-193.
33. Huang da W, Sherman BT & Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**(1): 44-57.

34. Bolstad BM, Irizarry RA, Astrand M & Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; **19**(2): 185-193.
35. Gautier L, Cope L, Bolstad BM & Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004; **20**(3): 307-315.

DISCUSSION

Chapitre 6

Discussion

6.1 Résumé du projet et mise en contexte

6.1.1 MAST3 : facteur de risque génétique aux MII

Tout projet de recherche subit l'influence de son contexte historique qui inspire les méthodes, le design expérimental et même les hypothèses de départ. Ce phénomène est particulièrement marqué en génétique, un domaine qui, ces dernières années, a connu des avancées considérables à la fois au niveau des connaissances et des technologies. Conséquemment, les études génétiques d'aujourd'hui ne ressemblent pas à celles d'il y a 10 ans. Le volet génétique de mon projet de doctorat s'inscrit dans un contexte historique qui suit la publication de la séquence du génome humain^{1;2} et le développement du projet HapMap³ mais qui précède les études d'association pangénomiques (EAP). L'objectif initial était de cribler la région de liaison aux MII, IBD6 pour identifier le ou les facteurs de risque génétique. Dans l'étude originale de Rioux *et al.* publiée en 2000⁴, le pic de liaison calculé représentait un impressionnant OR de 4,6 pour cette région. Nous avons choisi d'utiliser la plateforme Illumina GoldenGate Assay qui donnait la flexibilité de pouvoir choisir 1536 SNP à génotyper simultanément. En sélectionnant des SNP marqueurs de la base de données HapMap, il était possible de couvrir la majorité ($r^2 > 0,7$) des SNP communs (fréquence d'allèle mineur > 10%) dans la région IBD6. À la suite du génotypage en deux étapes (une de criblage, une de réplication), nous avons identifié un SNP intronique commun (fréquence d'allèle mineure 47% chez les contrôles) localisé dans le gène *MAST3*. Ce SNP s'est avéré corrélé ($r^2 > 0,7$) à quatre SNP codants dont un non-synonyme et donc avec potentiellement une influence fonctionnelle sur la protéine. Toutes les variantes de *MAST3* associés ont un OR d'environ 1,2⁵.

L'impact génétique du facteur de risque *MAST3* sur le pic de liaison initial de la région IBD6 est difficile à évaluer puisqu'il impliquerait un génotypage des échantillons originaux qui ne sont plus disponibles. Toutefois, il est probable que le OR de 4,6 de l'étude originale soit surévalué dû à un biais dans la distribution des variantes génétiques. Cette occurrence se nomme *winner's curse* dans le jargon génétique et a été discutée abondamment dans le cadre d'étude de liaison et d'association^{6;7}. On suggère différentes méthodes pour contrer cet effet dont le calcul des paramètres, par exemple le OR, dans un ensemble de données indépendantes de l'étude originale⁶. En 2004, une étude anglaise indépendante, a répliqué la liaison au locus IBD6, avec toutefois un OR plus modeste de 1,59⁸. Une confirmation additionnelle s'est concrétisée par la métaanalyse de *Van Heel et al.*⁹ qui a identifié une région de liaison sur le chromosome 19p à partir des données de 10 études indépendantes (dont l'étude originale de *Rioux et al.*). De plus, le locus 19p est une vaste région, riche en gènes, qui n'a pas d'intérêt que pour les MII. En effet, des études de liaison l'ont reliée à différents traits complexes tels que l'autisme (OR=2,6)¹⁰, les niveaux totaux de cholestérol, (OR=3,89)¹¹, le syndrome des jambes sans repos (OR=2,6)¹² et le psoriasis (OR=4,06)¹³. Ces observations suggèrent deux types de constatations. D'une part, elles témoignent de la densité génique de la région et donc de la possibilité qu'elle arbore de multiples facteurs de risque aux maladies. D'autres parts, elles peuvent suggérer que la région contient des gènes cruciaux pour l'équilibre métabolique encodant des protéines qui évoluent dans des voies biologiques qui sont communes à plusieurs maladies complexes.

Les différentes études réalisées jusqu'à maintenant semblent appuyer l'hypothèse qu'il y ait plus d'un facteur de risque génétique indépendant influençant la susceptibilité aux MII localisé sur le chromosome 19p. D'ailleurs, la région est localisée sur le chromosome arborant la plus grande densité génique et une grande proportion de ces gènes ont des fonctions immunitaires¹⁴. Différentes études d'association par gènes candidats ont associé aux MII des variantes dans le gène *intercellular adhesion molecule 1 (ICAMI)* encodant une glycoprotéine impliquée dans l'adhésion des cellules

endothéliales et immunitaires^{8; 15}. Le gène encodant la protéine myosin IXB (*MYO9B*) impliquée dans le mouvement intracellulaire a également été associé via des études par gène candidat dans des populations néerlandaise¹⁶, espagnole¹⁷, italienne¹⁸, britannique¹⁹ et canadienne²⁰. Les EAP ont à ce jour identifié deux loci associés à la MC dans la région 19p²¹. Le premier locus est associé aux gènes *tyrosine kinase 2* (*TYK2*) et *intercellular adhesion molecule* (*ICAM3*) alors que le second tombe dans la région des gènes *glutathione peroxidase 4* (*GPX4*) et *strawberry notch homolog 2* (*SBNO2*). L'effet faible des variantes associées (OR généralement autour de 1,1-1,2) est un point commun à tous ces gènes. À ce titre, l'effet des variantes de *MAST3* se situe dans la moyenne (OR=1,2). Il est donc possible d'imaginer que la combinaison de tous ces facteurs de risques avec effets modestes (et d'autres facteurs de risque encore à découvrir) explique la force de la liaison au locus IBD6.

Dans le cadre de mon projet, nous n'avons, cependant, identifié qu'un seul facteur de risque dans IBD6 : *MAST3*. Cette situation s'explique par le fait que la puissance statistique de détecter des variantes associées inhérentes à chaque étude d'association est habituellement faible²². Cette puissance peut être calculée à l'aide des paramètres suivant : les fréquences alléliques, les OR et les tailles d'échantillons. Selon l'outil informatique *Genetic Power Calculator*²³, la puissance de mon étude de détecter un variant associé à $P < 0,05$, tel que rs273506 (le variant intronique associé de *MAST3*), avec un OR de 1,2, une fréquence de 47% dans un échantillon de 1500 cas, se situe autour de 20%. Conséquemment, on manque une bonne proportion des signaux d'association. Une solution pour pallier ce déficit serait d'augmenter les tailles de cohortes et c'est justement un des buts des initiatives de regroupements de différentes études pangénomiques en métaanalyses d'association.

La puissance de détection des EAP actuelles sur les MII reste faible²². La toute dernière métaanalyse d'EAP sur la MC a porté le nombre de loci associés à 71²¹. Cette étude comportait environ 6300 cas atteints de MC. Selon Altshuler *et al.* cette taille d'échantillon permet 90% de pouvoir statistique de détecter un variant à $P < 10^{-8}$ (le seuil

de signification pangénomique) avec fréquence allélique d'au moins 10% et OR de plus de 1,4²². Pour la MC, la majorité des variantes sont associées à OR plus faible de l'ordre de 1,1-1,2 et pour avoir 90% de puissance de détection de ce type de variantes, il faut une cohorte de plus de 15 000 cas. C'est sans doute ce qui explique que les gènes identifiés dans le locus IBD6 par les études d'association par gènes candidats ou région candidate (dans mon cas) n'aient pas été nécessairement identifiés dans les EAP actuelles. D'ailleurs, on estime que les 71 loci identifiés par les EAP n'expliquent que 23,2% de l'héritabilité²¹ de la MC. Le même scénario se produit pour la CU où les 47 loci identifiés par les EAP n'expliquent que 16% de l'héritabilité²⁴. Il est donc probable que dans l'avenir, avec des cohortes indépendantes ou encore des métaanalyses impliquant de plus grandes cohortes, plusieurs autres gènes de susceptibilité soient identifiés par les EAP.

6.1.2 Premier lien fonctionnel immunitaire: MAST3 et NF-κB

Le volet fonctionnel de mon projet s'est d'abord déroulé de façon exploratoire, puisque très peu de choses étaient connues sur la protéine MAST3. En effet, l'étendue des connaissances sur MAST3 nous a appris qu'il s'agit d'une kinase qui peut se lier à la phosphatase *phosphatase and tensin homolog* (PTEN) pour la stabiliser et la phosphoryler²⁵ et que MAST3 est exprimé abondamment dans différentes parties du cerveau murin²⁶. Nous avons d'abord évalué l'expression de MAST3 dans des cellules d'intérêt dans l'étude des MII et trouvé une expression importante dans les lymphocytes et les cellules présentatrices d'antigènes. Par la suite, d'autres kinases de la même famille qui ont été plus étudiées nous ont permis de développer des hypothèses fonctionnelles pour MAST3. Par exemple, la protéine MAST2 régule l'activité du facteur de transcription NF-κB via la molécule *TNF receptor associated factor 6* (TRAF6)²⁷. Nous avons donc mesuré l'activité de NF-κB due à MAST3 via un knockdown stable du gène *MAST3* combiné à des essais gène rapporteur (luciférase) dans des cellules embryonnaires de rein (HEK293). Nous avons constaté que MAST3 régule l'activité de NF-κB spécifique aux récepteurs TLR4⁵. Ensuite, deux nouveaux

modèles cellulaires, un de surexpression dans des cellules HEK293 et un de knockdown dans des cellules THP1 ont confirmé la modulation de NF- κ B par MAST3. Cette découverte s'est révélée être un premier lien fonctionnel entre MAST3 et la réponse immunitaire sur lequel on s'est appuyé pour proposer un modèle du rôle de MAST3 (section 6.2).

La modulation de l'activité du facteur de transcription inflammatoire NF- κ B est un thème commun à plusieurs protéines encodées par des gènes MII. Par exemple, *NOD2*, le premier gène MII encode un récepteur intracellulaire qui influence la susceptibilité à l'inflammation causée par la flore intestinale via la modulation de l'activité de NF- κ B²⁸. L'action de NOD2 sur la voie NF- κ B se fait via le recrutement de protéines de signalisation intracellulaire comme cIAP1 and cIAP2²⁹. Cette activation de la voie NF- κ B mène à l'expression de chemokines, de cytokines et de défensines qui à leur tour modulent la réponse antimicrobienne³⁰. La protéine *macrophage stimulating protein* (MSP) encodée par le gène de susceptibilité aux MII *macrophage stimulating 1* (*MST1*) bloque la dégradation de la protéine inhibitrice de NF- κ B, I κ B et la liaison de NF- κ B à l'ADN³¹. De plus, les derniers loci identifiés par métaanalyse de EAP contiennent plusieurs cytokines régulées et régulatrices de NF- κ B^{21; 24}. Toutes ces indications confirment la voie de NF- κ B comme ayant un rôle majeur dans le développement de l'inflammation typique des MII.

Le facteur de transcription NF- κ B a lui-même directement été impliqué dans les MII. Les modèles murins ont d'abord montré que la suractivation du facteur de transcription était liée au développement de colites³². Chez les patients atteints de MII, le niveau d'activité de NF- κ B est corrélé au niveau d'inflammation³³. Cette activité de NF- κ B provoque des hausses d'expression de cytokines pro-inflammatoires telles que TNF, IL-1 et IL-6³⁴. Ces cytokines sont impliquées dans les dommages à la matrice extracellulaire qui causent la dégradation de la muqueuse intestinale³⁵. Paradoxalement, NF- κ B est également impliqué dans le maintien de l'homéostasie intestinale. En effet,

des modèles murins pour lesquels la voie NF- κ B est bloquée, via knockout, développent de l'inflammation intestinale³⁶. Le double rôle de NF- κ B révèle un modèle complexe qui témoigne des interactions constantes du système immunitaire avec la flore intestinale et du besoin de réagir aux attaques, mais aussi de développer un certain niveau de tolérance pour maintenir l'homéostasie à la muqueuse intestinale.

6.1.3 MAST3 influence l'expression de gènes pro-inflammatoires et anti-inflammatoires

La génomique fonctionnelle et particulièrement les études d'expression pangénomiques ont été utilisées abondamment avec l'objectif d'identifier une signature typique des tissus cancéreux. L'idée derrière cette stratégie est que les changements d'expression géniques sont accompagnés de différences dans la biologie cellulaire qui occasionnent des différences dans la présentation clinique³⁷. En identifiant ces gènes qui correspondent à des cancers à progressions différentes, on peut d'abord mieux diagnostiquer la maladie et ensuite apporter des traitements plus adaptés et plus personnalisés³⁸. D'autres domaines de la biologie ont également su exploiter la mine d'information provenant des études d'expression pangénomiques. Par exemple, les études d'expression pangénomiques sont utilisées pour étudier la diversité génomique des bactéries, l'effet des microARN et l'envergure de la réponse aux interférons³⁹⁻⁴¹.

Dans le cadre de mon projet de doctorat, nous avons choisi d'utiliser les études d'expression pangénomiques pour étudier de façon non biaisée les effets fonctionnels de MAST3 dans la cellule. Étant donné, le peu de connaissances fonctionnelles sur MAST3, nous avons choisi une stratégie qui nous permettait de produire et d'identifier des effets cellulaires. Nous nous sommes appuyés sur deux modèles cellulaires pour amplifier l'effet de MAST3 (surexpression) ou le supprimer (knockdown) et ainsi d'avoir un maximum de chances de détecter des effets sur l'expression. Cette stratégie n'est pas sans conséquence. D'abord, les effets détectés ne sont pas d'ordre

physiologique et ensuite le type cellulaire influence l'expression génique. Nous avons pris le pari d'utiliser cette stratégie pour explorer l'effet général d'une surexpression de MAST3 dans les HEK293 sachant que nous allions confirmer l'effet inverse dans le knockdown d'un type cellulaire différent (THP1). De plus, l'utilisation des THP1 permettait d'étudier l'expression génique dans un contexte de cellules immunitaires (macrophages).

Notre étude a mené à l'identification d'un groupe de gènes dont l'expression est modulée par la présence de MAST3. Une bonne partie de ces gènes est impliquée dans la réponse immunitaire pro-inflammatoire, ils encodent des cytokines (*IL8*, *CCL20*, *CXCL2*), des protéines impliquées dans la migration cellulaire (*CD44*, *TNC*) et des protéines induites par les interférons en réponse à une invasion virale ou de cellules cancéreuses (*IFIT1*, *ISG15*). Paradoxalement, une autre partie de ces gènes est impliquée dans la régulation de l'inflammation à travers la régulation de NF- κ B (*IER3*, *LY96*, *TNFAIP3*). Cette dichotomie fonctionnelle semble suivre le double rôle de NF- κ B dans l'inflammation. D'ailleurs, plus de 60% des gènes modulés par MAST3 ont au moins un site de liaison fonctionnel et confirmé expérimentalement dans leur promoteur. C'est-à-dire que leur expression est régulée par NF- κ B. Cette proportion apparaît comme un enrichissement par rapport à la proportion de gènes contrôlés par NF- κ B dans le génome. En effet, des études se sont penchées sur l'identification de gènes cibles de NF- κ B via des immunoprécipitations de la chromatine et des puces d'ADN (*ChIP on chip*). Une étude ciblée dans des cellules HeLa (fibroblastes humains col de l'utérus) montre que 15.5% des gènes sur le chromosome 22 sont activés par NF- κ B⁴². Cependant, les auteurs estiment que cette proportion est surévaluée puisqu'elle prend en compte les pseudogènes. Une autre étude faite sur des cellules U937 (monocytes humains) identifie 348 gènes activés par NF- κ B sur 9496 testés ce qui représente environ 3.6%⁴³. Ces résultats semblent loin du 60% observé dans le groupe régulé par MAST3. Nous avons également démontré que l'expression des gènes régulés par MAST3 corrélait avec l'activité de NF- κ B dans nos modèles cellulaires de knockdown

et de surexpression. Ces résultats semblent confirmer qu'une bonne partie de l'effet de MAST3 sur l'expression génique se produit via la voie NF- κ B.

Les effets de MAST3 sur l'expression de gènes immunitaires sont semblables à ceux d'un autre gène MII : *NOD2*. Dans une étude de la réponse transcriptionnelle inhérente à *NOD2* et *NOD2* L1006fsinsC (une des variantes associées aux MII), Billmann *et al.* ont montré un programme d'activation génique composé à la fois de gènes pro-inflammatoires et de régulateurs de NF- κ B⁴⁴. Tout comme MAST3, *NOD2* module l'expression de cytokines comme *IL8* et *CXCL2*, et de régulateurs de la voie NF- κ B comme *TNFAIP3*, *IER3*, *NFKBIA*, *NFKBIZ*, *TNFRSF9*. Globalement, ces résultats semblent suggérer que la voie NF- κ B est une voie centrale régulée par plusieurs protéines encodées par des gènes de susceptibilité aux MII. L'activation de cette voie influence l'expression d'un groupe de gènes qui est impliqué dans l'inflammation.

6.1.4 Certains gènes dont l'expression est modulée par MAST3 sont aussi des facteurs de risque aux MII connus

Les facteurs de risque aux MII encodent des protéines qui interagissent au niveau des mêmes voies biologiques. Nous avons constaté cette observation dans le cadre de nos études d'expression. C'est-à-dire qu'il s'est avéré que MAST3 influence l'expression de gènes qui sont eux aussi des facteurs de risque aux MII (*TNF*, *TNFRSF9*, *CCL20/CCR6*). Par exemple, le gène *TNF* localisé sur le chromosome 6p21 dans la région HLA a été associé à la MC suite à la métaanalyse de Barrett *et al.*⁴⁵. Le gène *TNFRSF9* est localisé dans le locus 1p36 associé à la CU²⁴. Le gène qui est le plus surexprimé lors d'une surexpression de *MAST3*, *CCL20* n'a pas été directement associé aux MII. Par contre, il encode l'unique ligand du récepteur CCR6 encodé par un gène associé à la MC⁴⁵. Ces observations suggèrent de l'interaction fonctionnelle entre les gènes MII et appuient l'hypothèse d'un nombre restreint de voies biologiques impliquées dans la susceptibilité aux MII.

6.1.5 Les gènes dont l'expression est modulée par MAST3 sont surexprimés dans les tissus enflammés typiques des MII

Pour évaluer l'implication, dans les MII, des gènes identifiés à la suite des profils d'expression reliés à MAST3, nous avons étudié des biopsies de muqueuse du côlon de patients atteints de CU. L'objectif était de vérifier les observations de nos modèles *ex vivo* dans un contexte physiologique. Nos études d'expression ont montré un enrichissement de l'expression de ces gènes dans les tissus enflammés de la muqueuse intestinale des patients par rapport aux tissus sains de patients. Ce constat n'est pas surprenant puisqu'une majorité de ces gènes est impliquée dans des fonctions immunitaires. D'autres études ont souligné l'importance de l'un ou l'autre de ces gènes dans les MII. Ces études étaient pour la plupart des études ponctuelles, dont des études de gènes candidats choisis par rapport à leur fonction immunitaire. Par contraste, nous avons défini notre groupe de gènes à la suite d'une étude d'expression pangénomique dans des cellules surexprimant *MAST3* sans biais pour une fonction quelconque. Tout de même, notre étude réplique plusieurs de ces observations. Entre autres, Tsukada *et al.* ont démontré la surexpression de *IL8* dans la muqueuse colique de patients atteints de UC⁴⁶. Arijis *et al.* ont rapporté des niveaux élevés d'expression de gènes impliqués dans l'adhésion cellulaire et des gènes encodant des cytokines, notamment *CCL20*, *CXCL2* et *IL8*, dans la muqueuse intestinale de patients atteints de MII⁴⁷. Matzuda *et al.* ont montré une forte expression de *TNF* et de *IL8* dans les tissus enflammés de patients atteints de CU⁴⁸.

Le groupe de gènes modulés par MAST3 représente une signature d'inflammation qui semble distinguer les tissus sains des tissus enflammés chez les patients atteints de MII. Nous ne prétendons pas que cette signature d'inflammation soit uniquement due à l'action de MAST3. Toutefois, nous avons prouvé que MAST3 peut activer l'activité de NF- κ B et l'expression de ce groupe de gènes et donc contribuer avec d'autres facteurs

de risque encodant des régulateurs de la voie NF- κ B, tel que NOD2, à créer un environnement cellulaire propice à l'inflammation.

6.2 Rôle de la kinase MAST3 dans l'inflammation

6.2.1 Modèle fonctionnel

Les études fonctionnelles sur MAST3 réalisées au cours de mon projet de doctorat représentent un premier pas dans la compréhension du rôle MAST3 dans l'inflammation. Mes résultats permettent tout de même de spéculer sur des partenaires potentiels et des molécules cibles sur lesquelles sont dirigés les effets de MAST3. Deux types d'effets ont été envisagés : la phosphorylation de molécule de signalisation et la stabilisation de TRAF6 (Figure 1).

Le programme d'activation génique influencé par MAST3 passe par la voie NF- κ B. Dans cette voie, d'abord activée par stimulation des récepteurs cellulaires comme TLR4, la transmission de la signalisation passe majoritairement par l'activation de molécules via phosphorylation. Plusieurs molécules de cette voie sont phosphorylées en serine ou thréonine : TAK1, IKK α , IKB α , p38, etc (Figure 1). MAST3 est une kinase qui peut phosphoryler d'autres molécules en serine et thréonine, notamment PTEN²⁵. Il est possible que l'action de MAST3 sur la voie NF- κ B se produise à travers la phosphorylation de l'une de ces molécules.

D'autre part, Xiong *et al.* ont montré que MAST2, une protéine de la même famille de kinases que MAST3, contrôlait la voie NF- κ B à travers la formation d'un complexe avec TRAF6. MAST2 agit comme chaperon en stabilisant TRAF6 et en favorisant sa transubiquitination²⁷. Les protéines MAST3 et MAST2 partagent une grande identité de séquence : 73% de la séquence totale d'acides aminés et 78% de la séquence du domaine PDZ impliqué des les interactions protéine-protéine⁵. Il est donc possible que MAST2 et MAST3 aient des rôles redondants sur la voie NF- κ B via leur action sur TRAF6.

La première étape dans la compréhension du rôle de MAST3 a été franchie avec les connaissances acquises à travers l'étude de nos modèles cellulaires. L'effet fonctionnel directement relié aux variantes associées de MAST3 sur la voie NF- κ B reste à étudier. Puisque que ces variantes ont une prévalence plus importante chez les patients atteints de MII, il semble probable qu'elles aient un impact pro-inflammatoire d'un autre ordre, mais similaire au modèle de surexpression de MAST3 en terme de direction de l'effet: c'est-à-dire l'activation de la voie NF- κ B. Des essais fonctionnels additionnels seront cependant nécessaires afin de confirmer cette hypothèse.

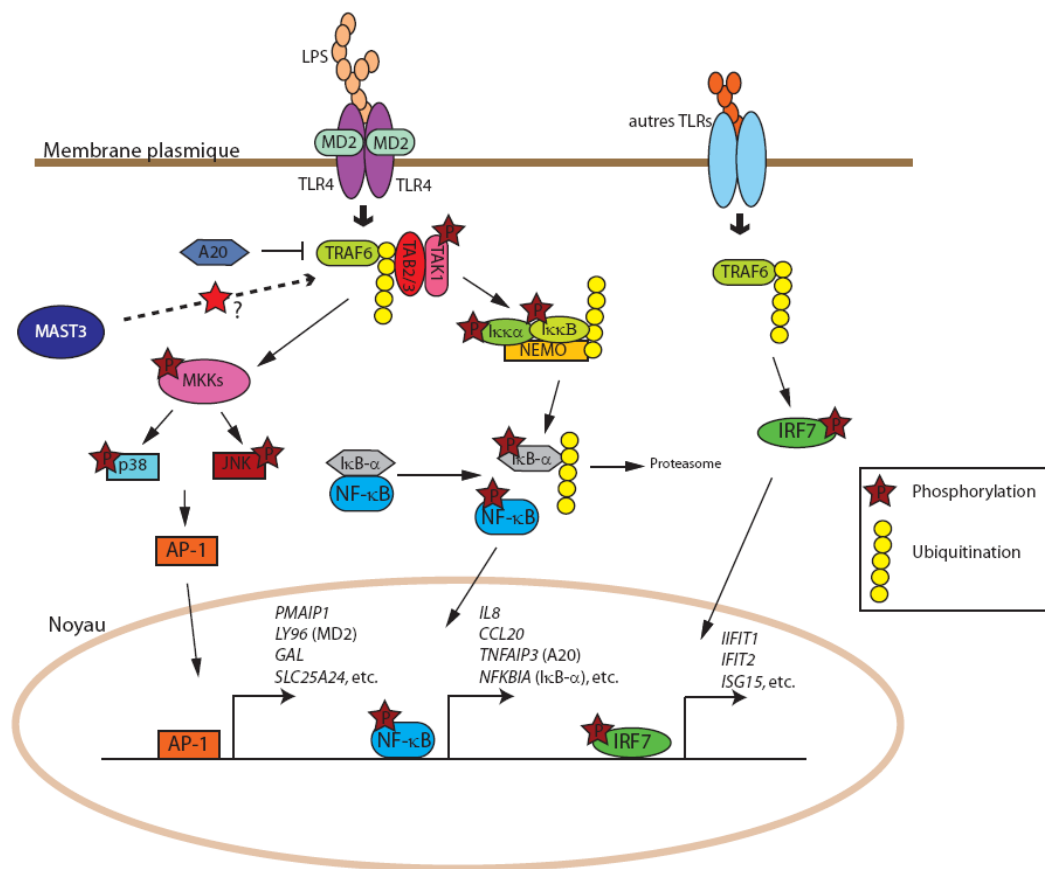


Figure 1. Cibles potentielles de MAST3 dans la voie NF-κB.

La liaison des sous-unités TLR4 avec le ligand LPS permet la dimérisation du récepteur TLR4 et son activation. S'en suit une cascade de signalisation culminant en l'activation de facteurs de transcription tels que NF-κB, AP-1 et IRF7 qui entraîne un programme d'expression de gènes impliqués dans des fonctions immunitaires. Des modifications post-transcriptionnelles comme de la phosphorylation et de l'ubiquitination se produisent pour activer les différentes molécules de la cascade. Certaines de ces molécules sont phosphorylées en serine/thréonine, potentiellement par MAST3 (étoile marron). MAST3 pourrait également agir sur la cascade à la stabilisation de TRAF6 (étoile rouge).

6.2.2 Effets biologiques

L'identification d'un groupe de gènes influencé par la modulation de l'expression de MAST3 dans la cellule permet de spéculer sur les effets biologiques de MAST3. De plus, l'identification de l'enrichissement de ce groupe de gènes dans les tissus enflammés de patients par rapport aux tissus sains de patients appuie l'hypothèse que le programme d'activation de gènes influencés par MAST3 a des conséquences sur l'état de la muqueuse intestinale. En clair, l'action de MAST3 en amont a des conséquences cellulaires observables qui se traduisent en manifestations cliniques des MII.

MAST3 module l'expression de **cytokines pro-inflammatoires** qui attirent les cellules immunitaires vers des régions qui deviennent des zones enflammées. Par exemple, IL-8 provoque la chimiotaxie et l'accumulation de neutrophiles et la formation d'œdèmes⁴⁹. L'IL-8 est essentiel au mouvement des neutrophiles, appelé diapédèse, à travers les cellules endothéliales formant les vaisseaux sanguins⁵⁰ et est impliqué dans la dégranulation de ces neutrophiles⁵¹. De plus, IL-8 est impliqué dans la croissance tumorale au niveau de l'angiogenèse⁵². IL-8 est également impliqué dans la production de métalloprotéases qui provoquent la dégradation de la matrice des cellules de la muqueuse intestinale^{53; 54}. La chemokine CCL20 est le seul ligand du récepteur CCR6 dont le gène est associé aux MII⁴⁵. CCR6 est exprimé dans plusieurs cellules immunitaires, dont les cellules dendritiques. Il a été démontré que CCR6 est impliqué dans l'induction de la migration de ces cellules vers l'épithélium intestinal aux sites de contact avec les pathogènes et autres antigènes⁵⁵. Par la suite, les cellules dendritiques migrent vers les nodules lymphatiques pour agir comme cellules présentatrices d'antigènes⁵⁵. CCR6 est également le seul récepteur à être exprimé sur tous les sous-types de cellules T_H17 et le complexe CCL20/CCR6 guide la migration des T_H17 vers les zones d'inflammation⁵⁶. Dans les plus récents modèles de développement de la MC, la réponse immunitaire due aux T_H17 se situe au premier plan⁵⁷.

MAST3 module l'expression de gènes encodant des **molécules d'adhésion** comme CD44. La protéine CD44 est impliquée dans l'adhésion intercellulaire et entre les cellules et la matrice extracellulaire. La présence de cytokines et chémokines dans le milieu cellulaire favorise l'expression de CD44 sur les lymphocytes T dont dépendent les interactions avec l'endothélium lors du roulement des lymphocytes⁵⁸. CD44 est également un site d'ancrage pour les molécules médiatrices d'inflammation comme MMP9⁵⁹. La liaison à CD44 augmente la phagocytose des neutrophiles apoptotiques par les macrophages⁶⁰. De plus, l'activation de CD44 influence l'expression de cytokines pro-inflammatoires⁶¹. L'expression du gène encodant la glycoprotéine tenascine C, *TNC*, est également modulée par MAST3. Cette protéine fait partie de la matrice extracellulaire et est impliquée dans l'adhésion cellulaire via des interactions avec la molécule d'adhésion fibronectine⁶². La tenascine est exprimée fortement dans les zones d'inflammation dues à l'acné⁶³ et au psoriasis⁶⁴. De plus, une étude récente a montré que cette protéine activait les récepteurs TLR4 et stimulait l'expression de cytokines pro-inflammatoire dans le contexte de l'arthrite rhumatoïde⁶⁵.

MAST3 module l'expression des gènes *IFIT1*, *IFIT2*, *IFIT3* qui encodent des protéines induites par introduction de virus à ARN ou ADN. Plus spécifiquement, ces protéines sont induites par les **interférons** sécrétés suite à une invasion virale. Elles ont été impliquées dans diverses fonctions cellulaires ayant pour but de stopper l'infection virale⁶⁶. La fonction la mieux caractérisée de ces protéines est l'inhibition de l'initiation de la traduction via la liaison au facteur eIF3⁶⁷. Les IFITs ont également été impliquées dans l'inhibition de la migration cellulaire⁶⁸ et de la prolifération cellulaire⁶⁹.

MAST3 influence l'expression de gènes impliqués dans un programme de freinage de l'inflammation via le **contrôle de la voie NF-κB**. La protéine MD-2, encodée par le gène *LY96* agit comme corécepteur de TLR4. La liaison du LPS au récepteur TLR4 et l'activation de NF-κB dépendent de la formation du complexe TLR4-MD-2⁷⁰. La protéine A20 encodée par le gène *TNFAIP3* coupe la transmission du signal dans la voie en désubiquitinant la molécule TRAF6⁷¹. Le gène *NFKBIA* encode la protéine IκBα qui

séquestre NF- κ B dans le cytoplasme et ce faisant inhibe sa translocalisation au noyau et son rôle de facteur de transcription⁷².

6.3 Perspectives

6.3.1 Limitations de l'étude d'association

Chacune de mes deux études présente des caractéristiques qui ont influencé la quantité et le type de résultats obtenus. Au niveau de mon étude d'association génétique, ces caractéristiques se traduisent par la taille des cohortes et la couverture de la variation génétique dans la région. Ces caractéristiques influencent la puissance statistique de détecter des signaux d'association. La puissance statistique est selon moi, une des principales limitations de l'étude. Tel que mentionné dans la section 6.1.1, la puissance de détecter une variante associée avec $P < 0,05$ ayant les caractéristiques de rs273506 (la variante intronique associée de *MAST3* - un OR de 1,2 et une fréquence de 47%) dans un échantillon d'environ 1500 cas, se situe autour de 20%. Ce qui implique qu'on manque une majorité des signaux d'association et qu'on s'expose à des résultats faux-positifs. Par exemple, sur les 37 SNP sélectionnés dans la première étape, un seul a répliqué dans la deuxième étape, rs273506 dans *MAST3*. Pour augmenter cette puissance statistique, on pourrait augmenter la taille des cohortes puisque cette caractéristique est directement reliée à la puissance de détection (Introduction Figure 11, page 56)²². La fréquence des variantes est une caractéristique qui entre aussi dans cette équation. Une augmentation de la taille des cohortes permettrait de capter plus facilement des signaux provenant de variantes de fréquences plus faibles. Nos cohortes totalisent environ 4000 individus (trios et cas-témoins pour les deux stades criblage et réplification) ce qui représentait un nombre impressionnant, il y a quelques années. Aujourd'hui, dans le domaine des études génétique sur les MII, la norme impose des tailles plus importantes. Les métaanalyses de CU et MC par exemple, comportent 6687 et 6333 cas respectivement^{21; 24}. Toutefois, même ces études ne détectent qu'une fraction des variantes responsable de l'héritabilité ce qui semble suggérer que des cohortes encore plus grandes soient nécessaires. Ce constat ne vient pas invalider le bien-fondé des études avec moins de 1000 cas et témoins bien appariés, il suggère cependant que les métaanalyses (comportant de plus en plus de cas) de toutes ces études pourront détecter davantage de variantes plus rares et à effet modeste.

La couverture de la variabilité procurée par mon étude influence également la puissance de détection. Une couverture plus grande n'était pas nécessairement possible à l'époque (la sélection des SNP est antérieure à HapMap) et considérant la taille des cohortes qui limitait de toute façon l'association avec des variantes rares, nous avons misé sur les SNP communs avec une fréquence allélique supérieure à 10% et nos tagSNP captent ces derniers avec un r^2 minimum de 0,7. Une partie de la variabilité (variantes plus rares et variantes communes non capturées par les tagSNP) nous échappe. L'amélioration de la couverture du génome via l'identification de nombreuses nouvelles variantes (SNP, CNV, etc) s'est produite de pair avec des innovations importantes dans les technologies de génotypage. Ces dernières permettraient d'augmenter de façon importante notre couverture du chromosome 19p et conjuguées à des cohortes plus grandes amélioreraient notre puissance de détection.

Nous nous sommes concentrés sur les variantes communes pour des questions de logistiques, et aussi à cause du paradigme sur les maladies complexes qui stipulent que les maladies communes sont causées par des variantes communes. Toutefois, de plus en plus, l'attention est tournée vers les variantes rares. Lors de notre étude d'association, nous avons identifié plusieurs variantes communes associées (fréquence d'allèle mineur autour de 46%) qui se situaient sur un même haplotype. Il a été impossible d'isoler une variante causale, mais le SNP non-synonyme rs8108738 est un candidat potentiel. Nous ne pouvons toutefois pas exclure que les variantes conférant le risque génétique soient rares et en déséquilibre de liaison avec les variantes communes identifiées. Le ré-séquençage pourrait permettre d'identifier des variantes rares causales chez les patients souffrants de MII. Dans le cadre d'un projet pilote qui visait à tester la faisabilité d'un projet de séquençage à grande échelle, les exons de *MAST3* et quelques autres gènes ont été séquencés dans des groupes de patients atteints de MII (les données ne sont pas dans la thèse). Trois nouveaux SNP ont été identifiés, mais leurs faibles fréquences alléliques (<0,01%) rendent plus difficiles les études fonctionnelles. Ces SNP devront être génotypés dans des cohortes de grandes tailles pour vérifier leur association. L'efficacité

de la stratégie de re-séquençage pour identifier des variantes rares associées aux MII et potentiellement causales vient d'être prouvée. Le projet de séquençage à grande échelle est paru dans *Nature Genetics* en octobre 2011⁷³. Dans le cadre de cette étude, 56 gènes associés à la MC ont été re-séquencés chez 350 patients atteints de MII. Plusieurs nouvelles variantes rares, notamment de *NOD2*, *IL23R* et *CARD9*, ont été identifiées. Certaines de ces variantes ont des effets fonctionnellement intéressants (par exemple, sur l'épissage de l'ARN messenger). Les auteurs estiment que ces nouvelles variantes rares expliquent de 1 à 2% de l'héritabilité. L'accès aux bases de données de séquençage publiques, comme celle du projet *1000Genomes*⁷⁴ permet également d'identifier des variantes plus rares potentiellement causales sur la base de leurs effets prédits par les outils de bioinformatiques. Par exemple, pour *MAST3* plusieurs variantes rares ont été identifiées dont certains SNP non-synonymes qui pourraient être dommageables pour la structure et la fonction de la protéine. Des études d'association centrées sur ces SNP combinées à des essais fonctionnels pourraient aider à comprendre leurs impacts sur les MII.

L'étude des variantes introniques et intergéniques comme variantes causales est souvent négligée dans les études d'association. Notamment dans notre étude, après l'identification initiale de la variante intronique associée, nous nous sommes tournés vers les variantes codantes pour identifier les variantes causales. Nous voulions passer aux études fonctionnelles rapidement et les variantes codantes, spécifiquement les non-synonymes, en raison de leur effet sur la séquence protéinique semblaient plus faciles à étudier. Cependant, les régions non-codantes renferment elles-aussi des variations qui peuvent avoir un impact sur les fonctions cellulaires. Ces variantes peuvent, par exemple, se trouver dans des régions qui affectent l'épissage de l'ARNm⁷⁵. Elles peuvent alors influencer la stabilité du messenger ou la séquence de la protéine. Les variantes peuvent également se trouver dans des régions régulatrices. Elles peuvent alors influencer l'expression du gène et l'abondance d'une protéine⁷⁶. Le SNP intronique associé de *MAST3* n'est pas localisé dans une région conservée et donc à moins de probabilité de se trouver dans une région régulatrice, mais il est possible que des

variantes rares ou communes corrélées à ce dernier le soient et modifient l'expression de *MAST3*. Cette avenue mériterait d'être explorée dans des études futures.

6.3.2 Limitations des études fonctionnelles

Nos études fonctionnelles représentaient un défi assez important puisque peu d'informations étaient disponibles quant au rôle de *MAST3* dans la cellule. Nous avons décidé de procéder avec des expériences exploratoires sur des modèles cellulaires de knockdown et de surexpression qui, via des études d'expression pangénomiques, nous permettaient d'obtenir une image générale des effets de *MAST3*. La limitation principale de ce design provient des modèles de surexpression et de knockdown. La quantité intracellulaire de *MAST3* dans ces modèles atteint des niveaux non-physiologiques. Ces niveaux nous permettent d'une part de détecter des effets plus facilement, d'autres parts nous exposent à des effets de toxicité non-reliés au rôle de *MAST3*. Toutefois, nous avons pallié en partie cette source d'erreurs en confirmant les résultats d'expression du modèle de surexpression dans un modèle de knockdown. Les gènes identifiés dans le groupe régulé par *MAST3* étaient surexprimés dans le modèle de surexpression et sous-exprimés dans le modèle de knockdown (par rapport à leurs contrôles respectifs) ce qui suggère l'authenticité de la modulation par *MAST3*.

La prochaine étape est de comprendre les effets spécifiques des variantes associées de *MAST3* dans l'inflammation. Plusieurs expériences peuvent être envisagées. Nous avons amorcé des expériences exploratoires sur l'effet du génotype de rs8108738 (comme proxy pour l'haplotype associé de *MAST3*) sur l'expression des gènes du groupe régulé par *MAST3* dans nos biopsies de patients et également dans des lignées de lymphocytes B immortalisés. Les tailles modestes des cohortes disponibles n'ont pas permis la détection de différences statistiquement significatives (ces résultats ne sont pas dans la thèse). Cependant, ce genre d'étude dans des cohortes de tailles plus importantes pourrait mettre la lumière sur les effets des variantes de *MAST3* sur l'expression de

gènes immunitaires (direction, intensité). En parallèle, la comparaison de modèles cellulaires *ex vivo* de surexpression des différents allèles associés, permettrait également d'étudier les effets des variantes sur l'expression de gènes, mais aussi rendrait possible beaucoup d'essais fonctionnels.

En termes d'essais fonctionnels, la phosphorylation est un des aspects que nous n'avons pas encore exploré. Nous avons découvert que MAST3 modulait l'activité de NF- κ B, toutefois, on ignore toujours le rôle précis de MAST3 dans cette voie. Une de mes hypothèses est que MAST3 phosphoryle une ou plusieurs molécules dans la voie de NF- κ B. La quantification (par exemple via Western blots) de la phosphorylation de cibles potentielles de MAST3 dans la voie (Figure 1) dans nos différents modèles cellulaires (surexpression, knockdown, et éventuellement les modèles exprimant les différentes variantes associées de *MAST3*) pourrait permettre de décortiquer la voie NF- κ B et préciser le rôle de MAST3.

6.3.3 Impact de mon étude sur la recherche sur les MII

J'ai identifié un facteur de risque aux MII. Malgré l'impact modeste des variantes associées, il s'agit tout de même d'un facteur de risque connu de plus. L'objectif ultime des études génétiques est de comprendre la totalité de l'héritabilité des maladies et à ce jour moins de 25% de l'héritabilité est expliquée. Un facteur de risque équivaut à un morceau du casse-tête dont le dévoilement total permettra un jour de comprendre entièrement l'étiologie des MII. De façon pratique, l'ensemble des variantes associées permettra d'une part, d'améliorer le diagnostic des MII via l'élaboration de classification moléculaire de maladies auxquelles on pourra prédire plus clairement des manifestations cliniques. Déjà, des études se sont penchées sur l'impact des variantes de *NOD2* sur le développement des manifestations cliniques de la MC⁷⁷. La considération d'un groupe plus important de gènes de susceptibilité à la MC permettra sans doute de les prédire encore mieux. D'autre part, la connaissance des facteurs de risque génétique

pourrait permettre de prédire la réponse thérapeutique et conséquemment de mieux adapter les protocoles de traitement. À ce propos, Arijis *et al.* ont montré qu'il était possible de distinguer la réponse thérapeutique à l'infliximab en fonction de l'expression d'un groupe de gènes⁷⁸.

Dans un autre ordre d'idée, mes résultats fonctionnels sur MAST3 ont permis de redécouvrir l'importance de la voie NF- κ B dans les mécanismes sous-jacents ces maladies. Nous avons donné un contexte biologique à un gène dont la fonction est méconnue. Nous avons relié MAST3 à un programme d'activation de gènes immunitaires. Nous avons aussi montré l'importance clinique de ce programme d'activation génique dans les biopsies de muqueuse intestinale provenant de zones enflammées de patients atteints de CU. De plus, nous avons découvert que l'activation de NF- κ B reliée à MAST3 est similaire à celle reliée à NOD2. Le programme d'activation génique de NOD2 et celui de MAST3 sont également semblables, ce qui souligne la redondance fonctionnelle des facteurs de risque aux MII. Cette redondance commence d'ailleurs à être évidente avec notamment plusieurs gènes MII impliqués dans la voie IL23R (*IL23R*, *JAK2*, *STAT3*, *TYK2*, *PTPN2*, *IL12B*)^{45; 79; 80} et plusieurs autres impliqués dans l'autophagie (*ATG16L1*, *NOD2*, *IRGM*, *LRKK2*, *DAP*)^{21; 24; 81-83}.

6.3.4 De la génétique au fonctionnel

Typiquement, la séquence génétique d'une personne est la même dans chacune de ses milliards de cellules somatiques et toutes les cellules possèdent les mêmes gènes. Pourtant, il existe différents types cellulaires qui ont des fonctions distinctes et qui s'agencent en tissus et en organe pour former l'organisme. L'expression différentielle de ces gènes est à l'origine des différences cellulaires et fonctionnelles. Les manifestations phénotypiques de traits complexes comme les MII impliquent l'interaction de plusieurs types de cellules et de tissus. Pour cette raison, les études génétiques seules ne sont pas

suffisantes, mais sont plutôt le premier maillon de la chaîne qui mène à la compréhension des mécanismes moléculaires sous-jacents à l'étiologie de ces maladies.

D'autant plus que les informations génétiques disponibles sont actuellement incomplètes. En effet, il existe une part d'héritabilité manquante qui se terre dans les gènes encore non identifiés, dans les variantes communes non couvertes ou mal couvertes par les technologies de génotypage actuelles et dans les variantes rares. Les nouvelles technologies de re-séquençage et l'intégration des plus récentes variantes cataloguées, notamment des variations en nombre de copies, promettent de résoudre une partie de l'énigme.

Les études de génomique fonctionnelle sont d'autres maillons de cette chaîne. Par exemple, les versions haut débit de mon étude d'expression, des criblages par délétion ou surexpression de gènes, font partie des techniques de bases utilisées pour comprendre la biologie de maladie comme le cancer, car elles permettent de lancer des hypothèses mesurables⁸⁴. De façon générale, les études fonctionnelles permettent aussi de confirmer les découvertes génétiques, d'approfondir la compréhension des interactions entre les facteurs de risques génétiques et de dévoiler des voies biologiques à l'origine des maladies. Par la suite, il devient possible de bloquer ou débloquent ces voies biologiques causatives via des cibles pharmacologiques. Les approches de biologie intégrative combinant plusieurs disciplines comme la génétique, la bio-informatique, la génomique fonctionnelle et la biologie moléculaire et cellulaire deviennent donc très importantes pour révéler les mécanismes moléculaires sous-jacents au développement des MII.

Dans le cadre de mon projet, j'ai voulu participer à cette chaîne de connaissance. J'ai voulu apporter mon maillon à travers la découverte génétique et l'étude fonctionnelle d'un facteur de risque et à ce titre je pense avoir réussi.

Conclusion

The primary goal of human genetics is to transform the treatment of common disease through an understanding of the underlying molecular pathways.

E.S. Lander⁸⁵

À mon avis, la recherche en génétique médicale s'apparente à une chasse au trésor qui aboutit à la découverte d'une autre série d'indices pour poursuivre la chasse. Qui plus est, dans le contexte des maladies complexes, il y a des dizaines voire des centaines de trésors à identifier. En vérité, l'identification de facteurs de risque génétiques n'est que la première phase qui permet de dévoiler des hypothèses mesurables. Pour que ce type de recherche se traduise en une amélioration des connaissances sur les mécanismes moléculaires sous-jacents le développement des maladies, les études de génomiques fonctionnelles sont des compagnons essentiels.

L'idée de départ de mon projet de doctorat était de bâtir une « histoire » la plus complète possible, en partant de l'identification de facteurs de risque génétique jusqu'à la caractérisation de leurs rôles dans la pathogenèse des MII. La recherche de facteurs de risque était ciblée au chromosome 19p, mais l'identification de *MAST3* fut tout de même une surprise. D'une part, parce que des dizaines de gènes connus pour être impliqués dans des fonctions immunitaires se trouvent dans le même locus et n'ont pas été identifiés et d'autre part, parce qu'on (dans le laboratoire en tout cas) n'avait jamais entendu parler de *MAST3*. Les études d'association actuelles basée sur le criblage de la variation sans cibler des gènes précis permettent les surprises. Ce qui est encore heureux parce qu'on est encore loin de comprendre l'expression, la fonction et les interactions de chacun des 20 500 gènes composant le génome humain.

Les MII sont parmi les maladies complexes les plus étudiées et les mieux connues. Tout de même, la susceptibilité génétique et les mécanismes moléculaires qui en sont à l'origine cachent encore des secrets que chaque nouvelle étude s'efforce de dévoiler, le but étant d'améliorer la santé humaine. Les sommités dans le domaine de la génétique s'entendent pour déclarer que la clef pour traiter les maladies se trouve dans la compréhension des voies biologiques qui les déclenchent. Je pense que chaque incursion dans l'univers fonctionnel derrière les facteurs de risque génétique est essentielle à la compréhension globale des interactions au cœur de ces voies biologiques. Déjà, les études de génétique se font plus rares et les chercheurs privilégient les projets intégratifs combinant génétique et autres disciplines de la biologie.

Avec mes recherches sur MAST3, j'ai participé à l'amélioration des connaissances sur les facteurs de risque aux maladies inflammatoires de l'intestin, j'ai identifié un régulateur de la voie NF- κ B et j'ai identifié un groupe de gènes immunitaires, régulés par MAST3, dont l'augmentation de l'expression est reliée à l'inflammation de la muqueuse intestinale typique des MII. J'ai l'espoir d'avoir participé à un projet collectif qui à moyen terme permettra d'améliorer le diagnostic et le traitement des 200 000 Canadiens qui souffrent de MII.

Bibliographie

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG *et al.* The sequence of the human genome. *Science* 2001; **291**(5507): 1304-1351.
3. HapMap. The International HapMap Project. *Nature* 2003; **426**(6968): 789-796.
4. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *American journal of human genetics* 2000; **66**(6): 1863-1870.
5. Labbe C, Goyette P, Lefebvre C, Stevens C, Green T, Tello-Ruiz MK *et al.* MAST3: a novel IBD risk factor that modulates TLR4 signaling. *Genes Immun* 2008; **9**(7): 602-612.
6. Goring HH, Terwilliger JD & Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *American journal of human genetics* 2001; **69**(6): 1357-1369.
7. Zollner S & Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *American journal of human genetics* 2007; **80**(4): 605-615.
8. Low JH, Williams FA, Yang X, Cullen S, Colley J, Ling KL *et al.* Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. *Inflamm Bowel Dis* 2004; **10**(3): 173-181.
9. van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD & Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum Mol Genet* 2004; **13**(7): 763-770.
10. McCauley JL, Li C, Jiang L, Olson LM, Crockett G, Gainer K *et al.* Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. *BMC Med Genet* 2005; **6**: 1.
11. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Fuller JH, Bennett PH *et al.* A locus influencing total serum cholesterol on chromosome 19p: results from an

- autosomal genomic scan of serum lipid concentrations in Pima Indians. *Arterioscler Thromb Vasc Biol* 2000; **20**(12): 2651-2656.
12. Kemlink D, Plazzi G, Vetrugno R, Provini F, Polo O, Stiasny-Kolster K *et al.* Suggestive evidence for linkage for restless legs syndrome on chromosome 19p13. *Neurogenetics* 2008; **9**(2): 75-82.
 13. Lee YA, Ruschendorf F, Windemuth C, Schmitt-Egenolf M, Stadelmann A, Nurnberg G *et al.* Genomewide scan in german families reveals evidence for a novel psoriasis-susceptibility locus on chromosome 19p13. *American journal of human genetics* 2000; **67**(4): 1020-1024.
 14. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J *et al.* The DNA sequence and biology of human chromosome 19. *Nature* 2004; **428**(6982): 529-535.
 15. Matsuzawa J, Sugimura K, Matsuda Y, Takazoe M, Ishizuka K, Mochizuki T *et al.* Association between K469E allele of intercellular adhesion molecule 1 gene and inflammatory bowel disease in a Japanese population. *Gut* 2003; **52**(1): 75-78.
 16. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM *et al.* Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* 2006; **131**(6): 1768-1774.
 17. Nunez C, Oliver J, Mendoza JL, Gomez-Garcia M, Pinero A, Taxonera C *et al.* MYO9B polymorphisms in patients with inflammatory bowel disease. *Gut* 2007; **56**(9): 1321-1322.
 18. Latiano A, Palmieri O, Valvano MR, D'Inca R, Caprilli R, Cucchiara S *et al.* The association of MYO9B gene in Italian patients with inflammatory bowel diseases. *Aliment Pharmacol Ther* 2008; **27**(3): 241-248.
 19. Cooney R, Cummings JR, Pathan S, Beckly J, Geremia A, Hancock L *et al.* Association between genetic variants in myosin IXB and Crohn's disease. *Inflamm Bowel Dis* 2009; **15**(7): 1014-1021.
 20. Wolters VM, Xu W, Zhao X, Walters TD, Griffiths AM, Silverberg MS *et al.* Replication of genetic variation in the MYO9B gene in Crohn's disease. *Hum Immunol* 2011; **72**(7): 592-597.
 21. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics* 2011; **42**(12): 1118-1125.

22. Altshuler D, Daly MJ & Lander ES. Genetic mapping in human disease. *Science* 2008; **322**(5903): 881-888.
23. Purcell S, Cherny SS & Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**(1): 149-150.
24. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics* 2011; **43**(3): 246-252.
25. Valiente M, Andres-Pons A, Gomar B, Torres J, Gil A, Tapparel C *et al.* Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. *The Journal of biological chemistry* 2005; **280**(32): 28936-28943.
26. Garland P, Quraishe S, French P & O'Connor V. Expression of the MAST family of serine/threonine kinases. *Brain Res* 2007.
27. Xiong H, Li H, Chen Y, Zhao J & Unkeless JC. Interaction of TRAF6 with MAST205 regulates NF-kappaB activation and MAST205 stability. *The Journal of biological chemistry* 2004; **279**(42): 43675-43683.
28. Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF *et al.* Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* 2005; **307**(5710): 734-738.
29. Bertrand MJ, Doiron K, Labbe K, Korneluk RG, Barker PA & Saleh M. Cellular inhibitors of apoptosis cIAP1 and cIAP2 are required for innate immunity signaling by the pattern recognition receptors NOD1 and NOD2. *Immunity* 2009; **30**(6): 789-801.
30. Peyrin-Biroulet L & Chamaillard M. NOD2 and defensins: translating innate to adaptive immunity in Crohn's disease. *J Endotoxin Res* 2007; **13**(3): 135-139.
31. Ray M, Yu S, Sharda DR, Wilson CB, Liu Q, Kaushal N *et al.* Inhibition of TLR4-induced IkappaB kinase activity by the RON receptor tyrosine kinase and its ligand, macrophage-stimulating protein. *J Immunol*; **185**(12): 7309-7316.
32. Zhang J, Stirling B, Temmerman ST, Ma CA, Fuss IJ, Derry JM *et al.* Impaired regulation of NF-kappaB and increased susceptibility to colitis-associated tumorigenesis in CYLD-deficient mice. *J Clin Invest* 2006; **116**(11): 3042-3049.

33. Rogler G, Brand K, Vogl D, Page S, Hofmeister R, Andus T *et al.* Nuclear factor kappaB is activated in macrophages and epithelial cells of inflamed intestinal mucosa. *Gastroenterology* 1998; **115**(2): 357-369.
34. Aupperle KR, Bennett BL, Boyle DL, Tak PP, Manning AM & Firestein GS. NF-kappa B regulation by I kappa B kinase in primary fibroblast-like synoviocytes. *J Immunol* 1999; **163**(1): 427-433.
35. Pallone F & Monteleone G. Mechanisms of tissue damage in inflammatory bowel disease. *Curr Opin Gastroenterol* 2001; **17**(4): 307-312.
36. Nenci A, Becker C, Wullaert A, Gareus R, van Loo G, Danese S *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* 2007; **446**(7135): 557-561.
37. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001; **98**(19): 10869-10874.
38. Baehner FL, Lee M, Demeure MJ, Bussey KJ, Kiefer JA & Barrett MT. Genomic signatures of cancer: basis for individualized risk assessment, selective staging and therapy. *J Surg Oncol* 2011; **103**(6): 563-573.
39. Bannantine JP & Talaat AM. Genomic and transcriptomic studies in *Mycobacterium avium* subspecies paratuberculosis. *Vet Immunol Immunopathol*; **138**(4): 303-311.
40. Tzur G, Israel A, Levy A, Benjamin H, Meiri E, Shufaro Y *et al.* Comprehensive gene and microRNA expression profiling reveals a role for microRNAs in human liver development. *PLoS One* 2009; **4**(10): e7511.
41. Hertzog P, Forster S & Samarajiwa S. Systems biology of interferon responses. *J Interferon Cytokine Res*; **31**(1): 5-11.
42. Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM *et al.* Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci U S A* 2003; **100**(21): 12247-12252.
43. Schreiber J, Jenner RG, Murray HL, Gerber GK, Gifford DK & Young RA. Coordinated binding of NF-kappaB family members in the response of human cells to lipopolysaccharide. *Proc Natl Acad Sci U S A* 2006; **103**(15): 5899-5904.
44. Billmann-Born S, Till A, Arlt A, Lipinski S, Sina C, Latiano A *et al.* Genome-wide expression profiling identifies an impairment of negative feedback signals

- in the Crohn's disease-associated NOD2 variant L1007fsinsC. *J Immunol* 2011; **186**(7): 4027-4038.
45. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics* 2008; **40**(8): 955-962.
 46. Tsukada Y, Nakamura T, Iimura M, Iizuka BE & Hayashi N. Cytokine profile in colonic mucosa of ulcerative colitis correlates with disease activity and response to granulocytapheresis. *Am J Gastroenterol* 2002; **97**(11): 2820-2828.
 47. Arijis I, De Hertogh G, Machiels K, Van Steen K, Lemaire K, Schraenen A *et al.* Mucosal gene expression of cell adhesion molecules, chemokines, and chemokine receptors in patients with inflammatory bowel disease before and after infliximab treatment. *Am J Gastroenterol* 2011; **106**(4): 748-761.
 48. Matsuda R, Koide T, Tokoro C, Yamamoto T, Godai T, Morohashi T *et al.* Quantitative cytokine mRNA expression profiles in the colonic mucosa of patients with steroid naive ulcerative colitis during active and quiescent disease. *Inflamm Bowel Dis* 2009; **15**(3): 328-334.
 49. Rampart M, Van Damme J, Zonnekeyn L & Herman AG. Granulocyte chemotactic protein/interleukin-8 induces plasma leakage and neutrophil accumulation in rabbit skin. *Am J Pathol* 1989; **135**(1): 21-25.
 50. Huber AR, Kunkel SL, Todd RF, 3rd & Weiss SJ. Regulation of transendothelial neutrophil migration by endogenous interleukin-8. *Science* 1991; **254**(5028): 99-102.
 51. Smith WB, Gamble JR, Clark-Lewis I & Vadas MA. Chemotactic desensitization of neutrophils demonstrates interleukin-8 (IL-8)-dependent and IL-8-independent mechanisms of transmigration through cytokine-activated endothelium. *Immunology* 1993; **78**(3): 491-497.
 52. Li A, Dubey S, Varney ML, Dave BJ & Singh RK. IL-8 directly enhanced endothelial cell survival, proliferation, and matrix metalloproteinases production and regulated angiogenesis. *J Immunol* 2003; **170**(6): 3369-3376.
 53. Chakrabarti S & Patel KD. Regulation of matrix metalloproteinase-9 release from IL-8-stimulated human neutrophils. *J Leukoc Biol* 2005; **78**(1): 279-288.
 54. Sengupta N & MacDonald TT. The role of matrix metalloproteinases in stromal/epithelial interactions in the gut. *Physiology (Bethesda)* 2007; **22**: 401-409.

55. Dieu-Nosjean MC, Vicari A, Lebecque S & Caux C. Regulation of dendritic cell trafficking: a process that involves the participation of selective chemokines. *J Leukoc Biol* 1999; **66**(2): 252-262.
56. Wang C, Kang SG, Lee J, Sun Z & Kim CH. The roles of CCR6 in migration of Th17 cells and regulation of effector T-cell balance in the gut. *Mucosal Immunol* 2009; **2**(2): 173-183.
57. 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): 1152-1167.
58. DeGrendele HC, Estess P, Picker LJ & Siegelman MH. CD44 and its ligand hyaluronate mediate rolling under physiologic flow: a novel lymphocyte-endothelial cell primary adhesion pathway. *J Exp Med* 1996; **183**(3): 1119-1130.
59. Yu Q & Stamenkovic I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis. *Genes Dev* 2000; **14**(2): 163-176.
60. Hart SP, Dougherty GJ, Haslett C & Dransfield I. CD44 regulates phagocytosis of apoptotic neutrophil granulocytes, but not apoptotic lymphocytes, by human macrophages. *J Immunol* 1997; **159**(2): 919-925.
61. Webb DS, Shimizu Y, Van Seventer GA, Shaw S & Gerrard TL. LFA-3, CD44, and CD45: physiologic triggers of human monocyte TNF and IL-1 release. *Science* 1990; **249**(4974): 1295-1297.
62. Huang W, Chiquet-Ehrismann R, Moyano JV, Garcia-Pardo A & Orend G. Interference of tenascin-C with syndecan-4 binding to fibronectin blocks cell adhesion and stimulates tumor cell proliferation. *Cancer research* 2001; **61**(23): 8586-8594.
63. Knaggs HE, Layton AM, Morris C, Wood EJ, Holland DB & Cunliffe WJ. Investigation of the expression of the extracellular matrix glycoproteins tenascin and fibronectin during acne vulgaris. *Br J Dermatol* 1994; **130**(5): 576-582.
64. Gerritsen MJ, Elbers ME, de Jong EM & van de Kerkhof PC. Recruitment of cycling epidermal cells and expression of filaggrin, involucrin and tenascin in the margin of the active psoriatic plaque, in the uninvolved skin of psoriatic patients and in the normal healthy skin. *J Dermatol Sci* 1997; **14**(3): 179-188.
65. Midwood K, Sacre S, Piccinini AM, Inglis J, Trebaul A, Chan E *et al*. Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for

maintaining inflammation in arthritic joint disease. *Nat Med* 2009; **15**(7): 774-780.

66. Fensterl V & Sen GC. The ISG56/IFIT1 gene family. *J Interferon Cytokine Res* 2011; **31**(1): 71-78.
67. Guo J, Hui DJ, Merrick WC & Sen GC. A new pathway of translational regulation mediated by eukaryotic initiation factor 3. *EMBO J* 2000; **19**(24): 6891-6899.
68. Lai KC, Chang KW, Liu CJ, Kao SY & Lee TC. IFN-induced protein with tetratricopeptide repeats 2 inhibits migration activity and increases survival of oral squamous cell carcinoma. *Mol Cancer Res* 2008; **6**(9): 1431-1439.
69. Xiao S, Li D, Zhu HQ, Song MG, Pan XR, Jia PM *et al.* RIG-G as a key mediator of the antiproliferative activity of interferon-related pathways through enhancing p21 and p27 proteins. *Proc Natl Acad Sci U S A* 2006; **103**(44): 16448-16453.
70. Park BS, Song DH, Kim HM, Choi BS, Lee H & Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 2009; **458**(7242): 1191-1195.
71. Boone DL, Turer EE, Lee EG, Ahmad RC, Wheeler MT, Tsui C *et al.* The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. *Nat Immunol* 2004; **5**(10): 1052-1060.
72. Ferreira DU & Komives EA. Molecular mechanisms of system control of NF-kappaB signaling by IkappaBalpha. *Biochemistry* 2010; **49**(8): 1560-1567.
73. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature genetics* 2011; **43**(11): 1066-1073.
74. The_1000genomes_Project_Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-1073.
75. Murani E, Ponsuksili S, Seyfert HM, Shi X & Wimmers K. Dual effect of a single nucleotide polymorphism in the first intron of the porcine secreted phosphoprotein 1 gene: allele-specific binding of C/EBP beta and activation of aberrant splicing. *BMC Mol Biol* 2009; **10**: 96.
76. Alberobello AT, Congedo V, Liu H, Cochran C, Skarulis MC, Forrest D *et al.* An intronic SNP in the thyroid hormone receptor beta gene is associated with pituitary cell-specific over-expression of a mutant thyroid hormone receptor

- beta2 (R338W) in the index case of pituitary-selective resistance to thyroid hormone. *J Transl Med* 2011; **9**: 144.
77. Henckaerts L, Van Steen K, Verstreken I, Cleynen I, Franke A, Schreiber S *et al*. Genetic risk profiling and prediction of disease course in Crohn's disease patients. *Clin Gastroenterol Hepatol* 2009; **7**(9): 972-980 e972.
78. Arijis I, Li K, Toedter G, Quintens R, Van Lommel L, Van Steen K *et al*. Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut* 2009; **58**(12): 1612-1619.
79. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al*. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**(5804): 1461-1463.
80. Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G *et al*. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nature genetics* 2008; **40**(11): 1319-1323.
81. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA *et al*. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature genetics* 2007; **39**(7): 830-832.
82. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K *et al*. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature genetics* 2007; **39**(2): 207-211.
83. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A *et al*. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics* 2007; **39**(5): 596-604.
84. Ashworth A & Bernards R. Using functional genetics to understand breast cancer biology. *Cold Spring Harb Perspect Biol* 2010; **2**(7): a003327.
85. Lander ES. Initial impact of the sequencing of the human genome. *Nature* 2011; **470**(7333): 187-197.

ANNEXES

Annexe I – loci MII

Tableau I1. Loci MII confirmés par les études d'association pangénomiques.

No.	SNP	Maladie	CHR	MC (6333 cas/15056 contrôles)					CU(6687 cases/19718 contrôles)				
				P	FCTRL	OR	G95	D95	P	FCTRL	OR	G95	D95
1	rs734999	CU	1	1,51E-01	0,52	1,03	0,99	1,08	1,21E-09	0,52	1,13	1,09	1,18
2	rs2797685	MC	1	2,69E-10	0,18	1,19	1,12	1,25	1,32E-01	0,18	1,04	0,99	1,10
3	rs35675666	CU	1	5,55E-02	0,84	1,07	1,00	1,15	1,09E-08	0,83	1,19	1,12	1,26
4	rs6426833	CU	1	3,63E-01	0,55	1,02	0,98	1,06	3,93E-35	0,54	1,30	1,25	1,35
5	rs7524102	CU	1	1,62E-02	0,82	1,07	1,01	1,13	1,04E-11	0,82	1,22	1,15	1,29
6	rs11209026	CU-MC	1	1,00E-64	0,93	2,67	2,37	3,01	5,12E-28	0,94	1,74	1,57	1,92
7	rs2476601	MC	1	4,47E-09	0,91	1,27	1,17	1,37	2,96E-01	0,90	0,96	0,90	1,03
8	rs3180018 (rs1142287)	MC	1	1,29E-09	0,24	1,16	1,11	1,22	1,91E-01	0,24	1,03	0,98	1,08
9	rs4656940	MC	1	6,17E-07	0,80	1,15	1,09	1,21	1,62E-02	0,80	1,06	1,01	1,12
10	rs1801274	CU	1	1,12E-02	0,52	1,05	1,01	1,10	2,16E-20	0,51	1,21	1,16	1,26
11	rs7517810	MC	1	1,51E-15	0,25	1,22	1,16	1,28	2,82E-01	0,24	0,97	0,93	1,02
12	rs1998598	MC	1	4,90E-09	0,31	1,14	1,09	1,20	1,80E-01	0,31	1,03	0,99	1,08
13	rs7554511	CU-MC	1	1,58E-07	0,73	1,14	1,08	1,19	2,04E-13	0,72	1,19	1,14	1,25
14	rs3024505	CU-MC	1	8,32E-09	0,16	1,18	1,12	1,25	5,76E-17	0,16	1,25	1,19	1,32
15	rs13428812	MC	2	1,41E-08	0,32	1,14	1,08	1,19	2,36E-03	0,32	1,07	1,02	1,12
16	rs780093	MC	2	1,10E-04	0,42	1,09	1,04	1,14	3,76E-04	0,40	1,08	1,03	1,12
17	rs10495903	MC	2	7,70E-08	0,13	1,18	1,11	1,26	7,01E-02	0,13	1,06	1,00	1,12
18	rs7608910	CU	2	3,11E-07	0,39	1,14	1,09	1,21	1,70E-14	0,39	1,19	1,14	1,24
18	rs10181042	MC	2	6,61E-09	0,42	1,14	1,09	1,19	7,55E-10	0,42	1,14	1,09	1,19
18	(0.60,0.83)												
19	rs2310173	CU	2	8,31E-05	0,46	1,09	1,04	1,14	8,44E-08	0,46	1,12	1,07	1,16
20	rs2058660	MC	2	1,58E-12	0,23	1,19	1,14	1,26	2,03E-03	0,23	1,08	1,03	1,13
21	rs6738825	MC	2	1,82E-07	0,48	1,12	1,07	1,17	5,86E-02	0,49	0,96	0,92	1,00
22	rs11676348	CU	2	2,57E-03	0,50	1,07	1,02	1,11	8,78E-09	0,49	1,13	1,08	1,17
23	rs7423615	MC	2	4,57E-09	0,19	1,17	1,11	1,23	9,16E-02	0,19	1,05	0,99	1,10
24	rs3792109	MC	2	6,76E-41	0,53	1,34	1,29	1,40	7,26E-02	0,53	1,04	1,00	1,08
25	rs4676406	CU	2	2,50E-03	0,50	1,07	1,02	1,11	8,32E-11	0,52	1,14	1,09	1,18
26	rs13073817	MC	3	8,20E-07	0,33	1,12	1,07	1,17	5,79E-02	0,34	1,04	1,00	1,09
27	rs9822268	CU	3	1,53E-16	0,30	1,21	1,16	1,27	1,60E-17	0,30	1,21	1,16	1,26
27	rs3197999	MC	3	6,17E-17	0,30	1,22	1,16	1,27	1,86E-17	0,30	1,21	1,16	1,26
27	(0.96,1.00)												
28	rs17388568	CU	4	1,69E-04	0,27	1,10	1,05	1,15	9,49E-07	0,27	1,12	1,07	1,17
29	rs11739663	CU	5	7,85E-01	0,76	0,99	0,94	1,04	2,80E-08	0,76	1,15	1,09	1,21
30	rs267939	CU	5	4,48E-04	0,37	1,08	1,03	1,13	9,67E-07	0,37	1,11	1,06	1,16
31	rs3194051	CU	5	3,06E-01	0,27	1,03	0,98	1,08	2,19E-06	0,27	1,12	1,07	1,17
32	rs6451493	CU	5	1,61E-27	0,60	1,35	1,28	1,43	1,78E-06	0,61	1,12	1,07	1,17

32	rs11742570	MC	5	7,08E-36	0,61	1,33	1,27	1,39	9,79E-07	0,61	1,11	1,07	1,16
32	(1.00,1.00)												
33	rs7702331	MC	5	2,00E-06	0,62	1,11	1,06	1,16	3,02E-01	0,61	1,02	0,98	1,07
34	rs2549794	MC	5	4,47E-11	0,40	1,16	1,11	1,21	6,19E-02	0,41	1,04	1,00	1,08
35	rs12521868	MC	5	1,41E-20	0,42	1,23	1,17	1,28	3,06E-03	0,42	1,06	1,02	1,11
36	rs254560	CU	5	1,15E-01	0,39	1,03	0,99	1,08	3,06E-07	0,40	1,11	1,07	1,16
37	rs11167764	MC	5	1,10E-09	0,79	1,18	1,12	1,25	1,40E-01	0,79	1,04	0,99	1,09
38	rs7714584	MC	5	7,76E-19	0,09	1,37	1,28	1,47	3,95E-04	0,09	1,14	1,06	1,22
39	rs6871626	CU	5	6,08E-12	0,34	1,15	1,10	1,20	1,02E-08	0,33	1,12	1,08	1,17
39	rs6556412	MC	5	5,37E-14	0,33	1,18	1,13	1,23	1,69E-05	0,34	1,09	1,05	1,14
39	(0.03,0.19)												
40	rs359457	MC	5	5,25E-08	0,57	1,13	1,08	1,18	1,30E-01	0,57	1,03	0,99	1,08
41	rs17309827	MC	6	6,16E-07	0,63	1,11	1,06	1,16	1,79E-01	0,62	1,03	0,99	1,07
42	rs6908425	MC	6	1,41E-08	0,78	1,17	1,11	1,23	7,75E-05	0,78	1,11	1,05	1,16
43	rs9268853	CU	6	8,53E-01	0,66	1,00	0,96	1,05	1,35E-55	0,66	1,40	1,34	1,47
43	rs1799964	MC	6	3,98E-11	0,21	1,19	1,13	1,25	8,82E-01	0,21	1,00	0,96	1,05
43	(0.05,0.35)												
44	rs943072	CU	6	2,94E-02	0,09	1,09	1,00	1,18	1,05E-06	0,10	1,17	1,09	1,25
45	rs1847472	MC	6	3,63E-06	0,66	1,11	1,06	1,16	6,04E-03	0,66	1,06	1,02	1,11
46	rs6911490	CU	6	4,28E-07	0,22	1,12	1,07	1,18	3,51E-07	0,21	1,13	1,07	1,18
46	rs6568421	MC	6	4,37E-08	0,30	1,13	1,07	1,18	4,36E-04	0,30	1,08	1,03	1,13
46	(0.33,0.74)												
47	rs6920220	CU	6	7,12E-02	0,20	1,05	1,00	1,11	6,38E-10	0,21	1,16	1,11	1,22
48	rs212388	MC	6	1,41E-07	0,41	1,12	1,07	1,17	3,78E-01	0,41	1,02	0,98	1,06
49	rs415890	MC	6	2,51E-12	0,52	1,17	1,12	1,22	2,96E-02	0,52	1,06	1,01	1,11
50	rs798502	CU	7	5,75E-01	0,71	1,01	0,97	1,06	1,21E-08	0,71	1,14	1,09	1,20
51	rs1456896	MC	7	1,20E-08	0,69	1,15	1,09	1,20	4,48E-02	0,69	1,04	1,00	1,09
52	rs4510766	CU	7	1,22E-01	0,56	1,04	0,99	1,10	2,00E-16	0,56	1,20	1,15	1,26
53	rs4728142	CU	7	7,98E-01	0,45	0,99	0,95	1,04	1,68E-06	0,45	1,10	1,06	1,15
54	rs4871611	MC	8	1,51E-12	0,61	1,17	1,12	1,23	1,12E-02	0,62	1,06	1,01	1,10
55	rs6651252	MC	8	2,29E-06	0,13	0,85	0,80	0,91	8,22E-02	0,13	1,05	0,99	1,12
56	rs10758669	CU-MC	9	1,00E-13	0,35	1,18	1,13	1,23	8,52E-13	0,35	1,16	1,11	1,21
57	rs4246905	CU	9	1,33E-15	0,71	1,21	1,15	1,28	4,77E-08	0,71	1,13	1,08	1,18
57	rs3810936	MC	9	1,00E-15	0,68	1,21	1,15	1,27	1,32E-07	0,68	1,12	1,07	1,17
57	(0.89,1.00)												
58	rs10781499	CU	9	3,49E-18	0,40	1,20	1,15	1,26	6,95E-13	0,40	1,16	1,11	1,21
58	rs4077515	MC	9	4,37E-19	0,41	1,21	1,16	1,26	9,56E-13	0,41	1,16	1,11	1,21
58	(0.98,1.00)												
59	rs12722489	MC	10	8,51E-06	0,85	1,15	1,08	1,22	8,92E-04	0,84	1,10	1,04	1,16
60	rs12261843	CU	10	1,87E-09	0,29	1,15	1,10	1,20	2,35E-08	0,29	1,13	1,08	1,18
60	rs12242110	MC	10	1,10E-09	0,32	1,15	1,10	1,20	3,86E-08	0,31	1,13	1,08	1,18

60	(0.76,0.96)												
61	rs1819658	MC	10	1,41E-07	0,79	1,15	1,09	1,22	3,35E-02	0,80	1,06	1,00	1,11
62	rs10761659	MC	10	4,37E-22	0,54	1,23	1,18	1,28	7,39E-06	0,54	1,10	1,05	1,14
63	rs1250550	MC	10	2,00E-10	0,68	1,16	1,10	1,21	2,25E-03	0,67	1,07	1,02	1,11
64	rs6584283	CU	10	7,18E-20	0,47	1,21	1,16	1,27	8,46E-21	0,47	1,21	1,16	1,26
64	rs4409764	MC	10	2,29E-20	0,49	1,22	1,17	1,27	4,01E-19	0,49	1,20	1,15	1,25
64	(0.92,0.98)												
65	rs907611	CU	11	2,77E-02	0,32	1,05	1,00	1,10	2,49E-08	0,32	1,13	1,08	1,18
66	rs102275	MC	11	7,24E-08	0,33	1,13	1,08	1,18	2,37E-01	0,33	1,03	0,98	1,07
67	rs694739	MC	11	3,38E-07	0,62	1,12	1,07	1,17	3,48E-03	0,62	1,06	1,02	1,11
68	rs2155219	CU	11	1,58E-12	0,51	1,16	1,11	1,21	6,33E-08	0,50	1,12	1,07	1,16
68	rs7927997	MC	11	5,62E-13	0,39	1,17	1,12	1,22	8,94E-05	0,39	1,08	1,04	1,13
68	(0.69,1.00)												
69	rs678170	CU	11	1,23E-01	0,66	1,04	0,99	1,08	6,88E-11	0,66	1,16	1,11	1,21
70	rs11564258	MC	12	6,17E-21	0,03	1,73	1,54	1,94	2,90E-04	0,02	1,28	1,12	1,45
71	rs7134599	CU	12	1,18E-03	0,39	1,07	1,03	1,12	1,06E-16	0,39	1,19	1,14	1,24
72	rs17085007	CU	13	7,87E-01	0,18	1,01	0,95	1,07	3,30E-09	0,18	1,16	1,10	1,22
73	rs941823	CU	13	1,31E-02	0,76	1,07	1,01	1,12	3,93E-07	0,76	1,13	1,08	1,19
74	rs2062305	MC	13	2,00E-06	0,48	1,11	1,06	1,16	9,35E-01	0,47	1,00	0,96	1,04
75	rs3764147	MC	13	1,41E-10	0,24	1,17	1,11	1,23	1,30E-02	0,24	1,06	1,01	1,11
76	rs4902642	MC	14	2,00E-07	0,58	1,12	1,07	1,17	2,74E-01	0,58	1,02	0,98	1,07
77	rs8005161	MC	14	1,29E-08	0,09	1,23	1,15	1,32	6,85E-04	0,09	1,13	1,05	1,21
78	rs17293632	MC	15	1,41E-13	0,24	1,19	1,14	1,25	9,52E-06	0,24	1,11	1,06	1,16
79	rs151181	MC	16	1,10E-10	0,38	1,15	1,10	1,20	5,31E-03	0,39	1,06	1,02	1,10
80	rs2076756	MC	16	3,98E-69	0,26	1,53	1,46	1,60	5,83E-01	0,25	1,01	0,97	1,06
81	rs6499188	CU	16	4,81E-03	0,74	1,07	1,02	1,13	3,97E-08	0,75	1,14	1,09	1,20
82	rs16940202	CU	16	9,33E-02	0,19	1,04	0,99	1,11	1,27E-12	0,18	1,18	1,12	1,24
83	rs3091315	MC	17	1,70E-13	0,72	1,20	1,14	1,26	1,88E-03	0,72	1,08	1,03	1,13
84	rs2872507	CU-MC	17	1,51E-09	0,46	1,14	1,09	1,19	5,44E-11	0,46	1,15	1,10	1,19
85	rs11871801	MC	17	2,51E-08	0,76	1,15	1,09	1,21	5,51E-04	0,75	1,09	1,04	1,14
86	rs1893217	MC	18	1,29E-14	0,15	1,25	1,18	1,32	4,78E-05	0,16	1,12	1,06	1,18
87	rs740495	MC	19	8,13E-12	0,25	1,16	1,10	1,21	3,75E-04	0,26	1,08	1,03	1,13
88	rs12720356	MC	19	9,20E-10	0,09	1,22	1,14	1,31	3,90E-06	0,08	1,17	1,09	1,26
89	rs736289	MC	19	2,69E-07	0,62	1,11	1,06	1,16	1,89E-05	0,61	1,08	1,03	1,12
90	rs281379	MC	19	8,60E-10	0,50	1,13	1,09	1,18	9,04E-01	0,50	1,00	0,96	1,04
91	rs6017342	CU	20	4,91E-01	0,55	1,01	0,96	1,07	1,09E-20	0,54	1,20	1,15	1,26
92	rs2297441	CU	20	1,83E-11	0,77	1,19	1,13	1,25	5,78E-08	0,76	1,14	1,09	1,20
92	rs4809330	MC	20	2,51E-12	0,69	1,18	1,13	1,24	3,87E-07	0,69	1,12	1,07	1,17
92	(0.54,0.92)												
93	rs1297265	CU	21	1,41E-08	0,57	1,16	1,10	1,22	1,73E-07	0,57	1,11	1,06	1,16
93	rs1736020	MC	21	9,33E-12	0,58	1,16	1,11	1,21	6,97E-07	0,58	1,11	1,06	1,16

93	(0.84,0.96)												
94	rs2836878	CU	21	3,22E-06	0,73	1,12	1,06	1,17	1,86E-22	0,74	1,25	1,20	1,32
95	rs2838519	CU-MC	21	2,09E-14	0,39	1,18	1,13	1,23	2,26E-08	0,39	1,12	1,08	1,17
96	rs181359	MC	22	6,31E-13	0,19	1,21	1,15	1,28	2,73E-05	0,19	1,11	1,06	1,17
97	rs713875	MC	22	5,70E-09	0,45	1,13	1,09	1,18	1,72E-03	0,46	1,08	1,03	1,14
98	rs2413583	MC	22	1,70E-10	0,83	1,20	1,13	1,28	1,15E-04	0,83	1,12	1,06	1,18
99	rs5771069	CU	22	3,49E-01	0,51	1,02	0,98	1,06	1,87E-07	0,52	1,11	1,07	1,16

Tableau I2. Gènes-candidats correspondant aux loci MC.

dbSNP ID	Chr.	Interval (Mb)	Gènes identifiés par la position génétique	Gènes identifié avec l'aide du 1000Genomes	Gènes identifiés avec GRAIL	Gènes identifié par eQTL (LOD score)
rs2476601	1p13	113.66 - 114.42		<i>PTPN22</i>	<i>PTPN22</i>	
rs11209026	1p31	67.13 - 67.54		<i>IL23R</i>	<i>IL23R</i>	
rs2797685	1p36	7.66 - 7.89	<i>VAMP3</i>	-	-	-
rs3180018	1q22	153.24 - 154.39	<i>SCAMP3, MUC1</i>	-	-	-
rs4656940	1q23	158.96 - 159.20	<i>ITLN1</i>		<i>CD244</i>	<i>CD244 (7.7)</i>
rs7517810	1q24	170.92 - 171.21			<i>TNFSF18, TNFSF4, FASLG</i>	
rs1998598	1q31	195.58 - 196.21	<i>DENND1B</i>	-	-	-
rs3024505	1q32	204.87 - 205.10			<i>IL10, IL19</i>	
rs7554511	1q32	199.11 - 199.32	<i>KIF21B</i>	<i>C1orf106</i>		
rs10181042	2p16	60.77 - 61.74	<i>REL</i>			<i>C2orf74 (9.6)</i>
rs10495903	2p21	43.30 - 43.80		<i>THADA</i>		
rs13428812	2p23	25.30 - 25.46	<i>DNMT3A</i>	-	-	-
rs780093	2p23	27.24-27.71		<i>GCKR</i>		
rs2058660	2q12	102.17 - 102.67	<i>IL18RAP</i>	-	-	-
rs6738825	2q33	197.85 - 198.67		<i>PLCL1</i>		
rs7423615	2q37	230.76 - 230.94				<i>SP140 (8.8)</i>
rs3792109	2q37	233.81 - 234.23		<i>ATG16L1</i>	<i>ATG16L1</i>	
rs3197999	3p21	48.16 - 51.73		<i>MST1, GPX1, BSN</i>		
rs13073817	3p24	18.58 - 18.86				
rs11742570	5p13	39.88 - 41.00	<i>PTGER4</i>	-	-	<i>PTGER4</i>
rs7702331	5q13	72.49 - 72.62	<i>TMEM174</i>	-	-	-
rs2549794	5q15	96.11 - 96.45		<i>ERAP2</i>		<i>LRAP (47.2)</i>
rs12521868	5q31	129.41 - 132.05		<i>SLC22A4</i>	<i>IRF1, CSF2, IL3</i>	<i>SLC22A5 (5.4)</i>
rs11167764	5q31	141.39 - 141.62	<i>NDFIP1</i>	-	-	-
rs7714584	5q33	150.01 - 150.38	<i>IRGM</i>	-	<i>IRGM</i>	-
rs6556412	5q33	158.43 - 158.88	<i>IL12B</i>	-	<i>IL12B</i>	-
rs359457	5q35	173.15 - 173.47				<i>CPEB4 (6.1)</i>
rs1799964	6p21	31.49 - 32.98	<i>HLA-DQA2</i>	<i>MCCD1, LTA</i>	<i>TNF, LST1, LTB, LTA, NCR3</i>	
rs6908425	6p22	20.60 - 21.25	<i>CDKAL1</i>	-	-	-
rs17309827	6p25	3.35 - 3.41				
rs1847472	6q15	90.86 - 91.14	<i>BACH2</i>	-	-	-
rs6568421	6q21	106.50 - 106.67	<i>PRDM1</i>	-	<i>PRDM1</i>	-
rs212388	6q25	159.26 - 159.46	<i>TAGAP</i>	-	-	-
rs415890	6q27	167.26 - 167.47	<i>CCR6</i>	-	<i>CCR6</i>	-
rs1456896	7p12	50.22 - 50.34	<i>IKZF1</i>	-	<i>IKZF1, ZBPB, FIGNL1</i>	-
rs6651252	8q24	129.56 - 129.67				

rs4871611	8q24	126.54 - 126.65				
rs10758669	9p24	4.93 - 5.29	JAK2			
rs3810936	9q32	116.47 - 116.74	TNFSF15		TNFSF15, TNFSF8	
rs4077515	9q34	138.27 - 138.54		CARD9, SNAPC4		CARD9 (12.4)
rs12242110	10p11	35.22 - 35.94			CREM	CREM (6.4)
rs12722489	10p15	6.07 - 6.21			IL2RA	
rs1819658	10q21	59.50 - 59.81	UBE2D1			
rs10761659	10q21	63.97 - 64.43	ZNF365			
rs1250550	10q22	80.67 - 80.77	ZMIZ1			
rs4409764	10q24	101.26 - 101.33	NKX2-3			
rs102275	11q12	61.28 - 61.44				FADS1 (5.0)
rs694739	11q13	63.58 - 64.05	PRDX5, ESRRRA			
rs7927997	11q13	75.70 - 76.04	C11orf30			
rs11564258	12q12	38.42 - 39.31	LRRK2	MUC19		
rs2062305	13q14	41.72 - 42.00				TNFSF11 (5.9)
rs3764147	13q14	43.13 - 43.54		C13orf31		
rs4902642	14q24	68.23 - 68.39	ZFP36L1			
rs8005161	14q35	87.28 - 87.71		GALC, GPR65	GPR65	
rs17293632	15q22	65.20 - 65.27	SMAD3			
rs151181	16p11	28.20 - 28.94		APOB48R, IL27, SULT1A2, SULT1A1, SH2B1		EIF3C (11.3)
rs2076756	16q12	49.02 - 49.41	NOD2	NOD2	NOD2	
rs3091315	17q12	29.51 - 29.70	CCL2, CCL7			
rs2872507	17q21	34.62 - 35.51		GSMDL, ZPBP2	IKZF3	ORMDL3 (20.3)
rs11871801	17q21	37.57 - 38.25	STAT3	MLX		
rs1893217	18p11	12.73 - 12.92	PTPN2			
rs12720356	19p13	10.26 - 10.50		TYK2	TYK2, ICAM1, ICAM3	
rs740495	19p13	1.04 - 1.13	GPX4, SBNO2			
rs736289	19q13	38.42 - 38.47				
rs281379	19q13	53.78 - 53.97		FUT2, RASIP1		
rs12242110	10p11	35.22 - 35.94			CREM	CREM (6.4)
rs12722489	10p15	6.07 - 6.21			IL2RA	
rs1819658	10q21	59.50 - 59.81	UBE2D1			
rs10761659	10q21	63.97 - 64.43	ZNF365			
rs1250550	10q22	80.67 - 80.77	ZMIZ1			
rs4409764	10q24	101.26 - 101.33	NKX2-3			
rs102275	11q12	61.28 - 61.44				FADS1 (5.0)
rs694739	11q13	63.58 - 64.05	PRDX5, ESRRRA			
rs7927997	11q13	75.70 - 76.04	C11orf30			
rs11564258	12q12	38.42 - 39.31	LRRK2	MUC19		

rs2062305	13q14	41.72 - 42.00				<i>TNFSF11 (5.9)</i>
rs3764147	13q14	43.13 - 43.54		<i>C13orf31</i>		
rs4902642	14q24	68.23 - 68.39	<i>ZFP36L1</i>			
rs8005161	14q35	87.28 - 87.71		<i>GALC, GPR65</i>	<i>GPR65</i>	
rs17293632	15q22	65.20 - 65.27	<i>SMAD3</i>			
rs151181	16p11	28.20 - 28.94		<i>APOB48R, IL27, SULT1A2, SULT1A1, SH2B1</i>		<i>EIF3C (11.3)</i>
rs2076756	16q12	49.02 - 49.41	<i>NOD2</i>	<i>NOD2</i>	<i>NOD2</i>	
rs3091315	17q12	29.51 - 29.70	<i>CCL2, CCL7</i>			
rs2872507	17q21	34.62 - 35.51		<i>GSM DL, ZPBP2</i>	<i>IKZF3</i>	<i>ORMDL3 (20.3)</i>
rs11871801	17q21	37.57 - 38.25	<i>STAT3</i>	<i>MLX</i>		
rs1893217	18p11	12.73 - 12.92	<i>PTPN2</i>			
rs12720356	19p13	10.26 - 10.50		<i>TYK2</i>	<i>TYK2, ICAM1, ICAM3</i>	
rs740495	19p13	1.04 - 1.13	<i>GPX4, SBNO2</i>			
rs736289	19q13	38.42 - 38.47				
rs281379	19q13	53.78 - 53.97		<i>FUT2, RASIP1</i>		
rs4809330	20q13	61.65 - 61.95		<i>RTEL1</i>		
rs1736020	21q21	15.62 - 15.77				
rs2838519	21q22	44.41 - 44.52	<i>ICOSLG</i>		<i>ICOSLG</i>	
rs181359	22q11	20.14 - 20.39		<i>YDJC</i>		
rs713875	22q12	28.23 - 29.00	<i>MTMR3</i>			
rs2413583	22q13	38.00 - 38.14	<i>MAP3K7IP1</i>			

Tableau I3. Gènes-candidats correspondant aux loci CU.

dbSNP ID	Chr.	Position (Mb)	Gènes	GRAIL	eQTL	nsSNP	Candidat par position
rs11209026	1p31	67.30 - 67.54	<i>IL23R, C1orf141</i>	<i>IL23R</i>		<i>IL23R</i>	<i>IL23R</i>
rs734999	1p36	2.39 - 2.80	<i>PLCH2, MMEL1, PANK4, C1orf93, HESS, TNFRSF14</i>	<i>TNFRSF14</i>	<i>TNFRSF14, MMEL1, PLCH2, C1orf93</i>	<i>MMEL1</i>	
rs35675666	1p36	7.83 - 8.13	<i>TNFRSF9, ERRF1, UTS2, PARK7, PER3</i>	<i>TNFRSF9†, ERFF1†</i>		<i>UTS2</i>	<i>PARK7</i>
rs7524102	1p36	22.54 - 22.61					
rs6426833	1p36	19.93 - 20.18	<i>TMCO4, RNF186, OTUD3, PLA2G2E, PLA2G2A</i>				
rs1801274	1q23	159.54 - 159.91	<i>MPZ, SDHC, LOC542502, C1orf192, FCGR2A, HSPA6, FCGR3A, FCGR2C, FCGR3B, FCGR2B</i>	<i>FCGR2A, FCGR2B</i>	<i>FCGR2B, FCGR2A, HSPA6</i>	<i>FCGR2A</i>	<i>FCGR2A</i>
rs7554511	1q32	199.06 - 199.33	<i>C1orf106, GPR25, KIF21B, CACNAIS, CAMSAP1L1</i>				<i>C1orf106</i>
rs3024505	1q32	204.85 - 205.11	<i>LGTN, IL19, DYRK3, IL20, IL10, MAPKAPK2</i>	<i>IL10, IL19</i>			
rs7608910	2p16	60.76 - 61.87	<i>PUS10, REL, CCDC139, XPO1, KIAA1841, PEX13, AHS2, USP34, PAPOLG</i>				<i>PUS10</i>
rs2310173	2q11	101.66 - 102.13	<i>MAP4K4, IL1R2, IL1R1</i>	<i>IL1R2</i>	<i>IL1R2</i>		
rs11676348	2q35	218.58 - 218.97	<i>TMBIM1, ARPC2, AAMP, GPBARI, PNKD, IL8RA, IL8RB, SLC11A1 (NRAMP1), C2orf62, RUFY4, CTDSPI</i>	<i>IL8RA, IL8RB</i>	<i>IL8RA, SLC11A1, IL8RB, AAMP, ARPC</i>		
rs4676406	2q37	241.20 - 241.32	<i>AQP12A, AQP12B, GPR35, KIF1A</i>	<i>GPR35</i>			
rs9822268	3p21	48.14 - 51.77	Plusieurs	<i>MST1</i>	<i>UBA7, MST1, APEH, AMIGO3, GMPFB</i>	<i>MST1, BSN</i>	<i>MST1</i>
rs17388568	4q27	123.20 - 123.78	<i>IL21, ADAD1, KIAA1109, IL2,</i>	<i>IL21, IL2</i>			<i>ADAD1</i>
rs3194051	5p13	35.83 - 36.07	<i>CAPSL, IL7R, UGT3A1, UGT3A2</i>	<i>IL7R</i>	<i>IL7R</i>	<i>IL7R</i>	<i>IL7R</i>

rs6451493*	5p13	40.32 - 40.85	<i>PTGER4, PRKAA1, TTC33</i>	<i>PTGER4†</i>			
rs267939	5p15	10.72 - 10.90	<i>DAP</i>		<i>DAP</i>		<i>DAP</i>
rs11739663	5p15	0.48 - 0.80	<i>Q9NX68_HUMAN, CEP72, PDCD6, Q71RB1_HUMAN, EXOC3, TPPP, SLC9A3, Q8N2X6_HUMAN, Q6P7N6_HUMAN,</i>		<i>EXOC3</i>		
rs254560	5q31	134.41 - 134.53					
rs6871626	5q33	158.46 - 158.86	<i>RNF145, IL12B, UBLCP1</i>	<i>IL12B</i>			
rs943072	6p21	43.88 - 43.92					
rs9268853	6p21	31.49 - 33.01	Plusieurs		<i>HLA-DRB5, HLA-DQA1, HLA-DRB1, HLA-DRA</i>	<i>BTNL2, HLA-DRB1, HLA-DQA1</i>	
rs6911490	6q21	106.51 - 106.67	<i>PRDMI</i>	<i>PRDMI</i>			
rs6920220	6q23	137.88 - 138.17					
rs798502	7p22	2.70 - 2.90	<i>GNA12, AMZ1</i>	<i>GNA12</i>	<i>GNA12</i>	<i>GNA12</i>	<i>GNA12</i>
rs4510766	7q22	107.20 - 107.39	<i>SLC26A3, DLD, LAMB1</i>				
rs4728142	7q32	128.33 - 128.56	<i>IRF5, TNPO3</i>	<i>IRF5</i>	<i>TNPO3</i>		
rs10758669	9p24	4.93 - 5.28	<i>INSL4, INSL6, JAK2, RLN2</i>	<i>JAK2</i>			
rs4246905	9q32	116.48 - 116.74	<i>TNFSF8, TNFSF15</i>	<i>TNFSF8, TNFSF15</i>	<i>TNFSF8</i>	<i>TNFSF15</i>	<i>TNFSF15</i>
rs10781499	9q34	138.27 - 138.55	<i>QSOX2, GPSM1, DNLZ, CARD9, SNAPC4, SDCCAG3, PMPCA, INPP5E, SEC16A, C9orf163, NOTCH1</i>		<i>CARD9, INPP5E, SDCCAG3, SEC16A, SNAPC4</i>	<i>CARD9, SNAPC4</i>	<i>CARD9</i>
rs12261843	10p11	35.22 - 35.94	<i>CREM, GJD4, CUL2, CCNY</i>		<i>CCNY</i>		<i>CCNY</i>
rs6584283	10q24	101.25 - 101.33	<i>NKX2-3</i>				
rs2155219	11q13	75.72 - 76.02	<i>C11orf30, PRKRIR (DAP4)</i>				
rs907611	11q15	1.82 - 1.93	<i>TNNI2, LSP1, MRPL23, TNNT3</i>		<i>LSP1</i>		
rs678170	11q23	113.76 - 114.08	<i>C11orf71, REXO2, FAM55D, FAM55A, RBM7</i>				
rs7134599	12q14	66.72 - 66.92	<i>IFNG, IL26, IL22</i>	<i>IFNG, IL26</i>			
rs17085007	13q12	26.39 - 26.46					
rs941823	13q13	39.90 - 39.95					
rs6499188	16q22	66.98 - 67.40	<i>SMPD3, ZFP90, CDH3, CDH1</i>		<i>ZFP90</i>		<i>ZFP90</i>
rs16940202	16q24	84.53 - 84.58					

rs2872507	17q12	34.62 - 35.51	<i>PERLD1, ZBPB2, NR1D1, GSDML, FBXL20, PSMD3, THRA, STARD3, ERBB2, NEUROD2, IKZF3, GSDMI, PPP1R1B, TCAP, MED1, GRB7, STAC2, CSF3, MED24, PNMT, ORMDL3, C17orf37, CRKRS</i>	<i>IKZF3</i>	<i>ORMDL3, IKZF3, PNMT</i>	<i>ZBPB2, GSDML</i>	
rs2297441	20q13	61.66 - 61.98	<i>C20orf195, PRIC285, GMEB2, STMN3, RTEL1, TNFRSF6B, ARFRP1, ZGPAT, LIME1, SLC2A4RG, ZBTB46, C20orf135, TPD52L2</i>		<i>SLC2A4RG, STMN3, ZBTB46, ZGPAT</i>	<i>RTEL1, TNFRSF6B</i>	<i>RTEL1</i>
rs6017342	20q13	42.49 - 42.70	<i>TTPAL, ADA, SERINC3, HNF4A, PKIG</i>		<i>SERINC3</i>		
rs1297265	21q11	15.62 - 15.77					
rs2838519	21q22	44.41 - 44.52	<i>DNMT3L, ICOSLG,</i>	<i>ICOSLG</i>			
rs2836878	21q22	39.34 - 39.41					
rs5771069	22q13	48.70 - 48.83	<i>ALG12, CRELD2, PIM3, IL17REL, TTLL8, MLC1</i>		<i>PIM3</i>	<i>IL17REL</i>	<i>IL17REL</i>

Annexe II –Additional figures for article 2

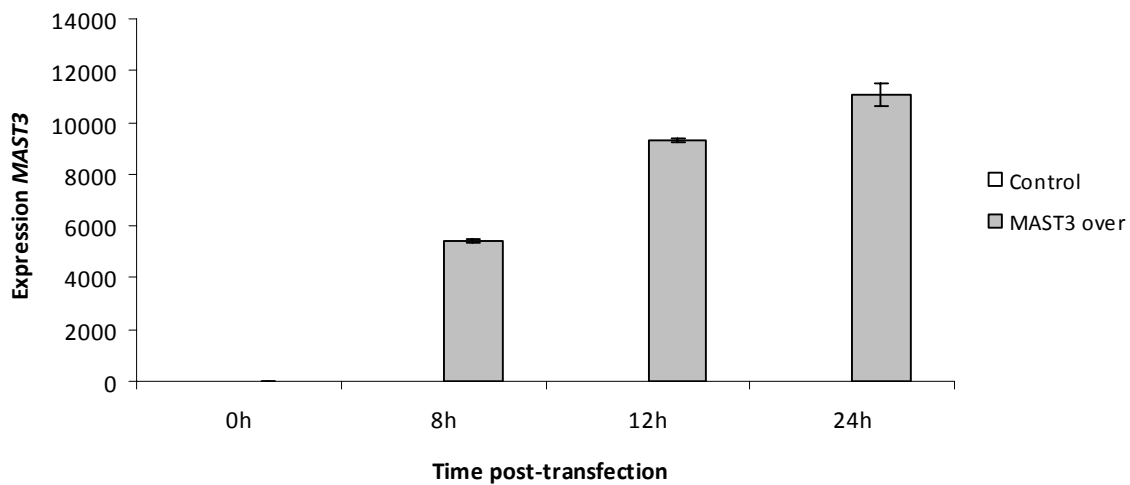


Figure II1. Expression of *MAST3* mRNA post-transfection in *MAST3* overexpressed cells.

The graph shows the gene expression tested by qPCR at different times post-transfection. Expression is normalized to *HPRT1* expression. Error bars represent standard deviation of 3 independent biological replicates.

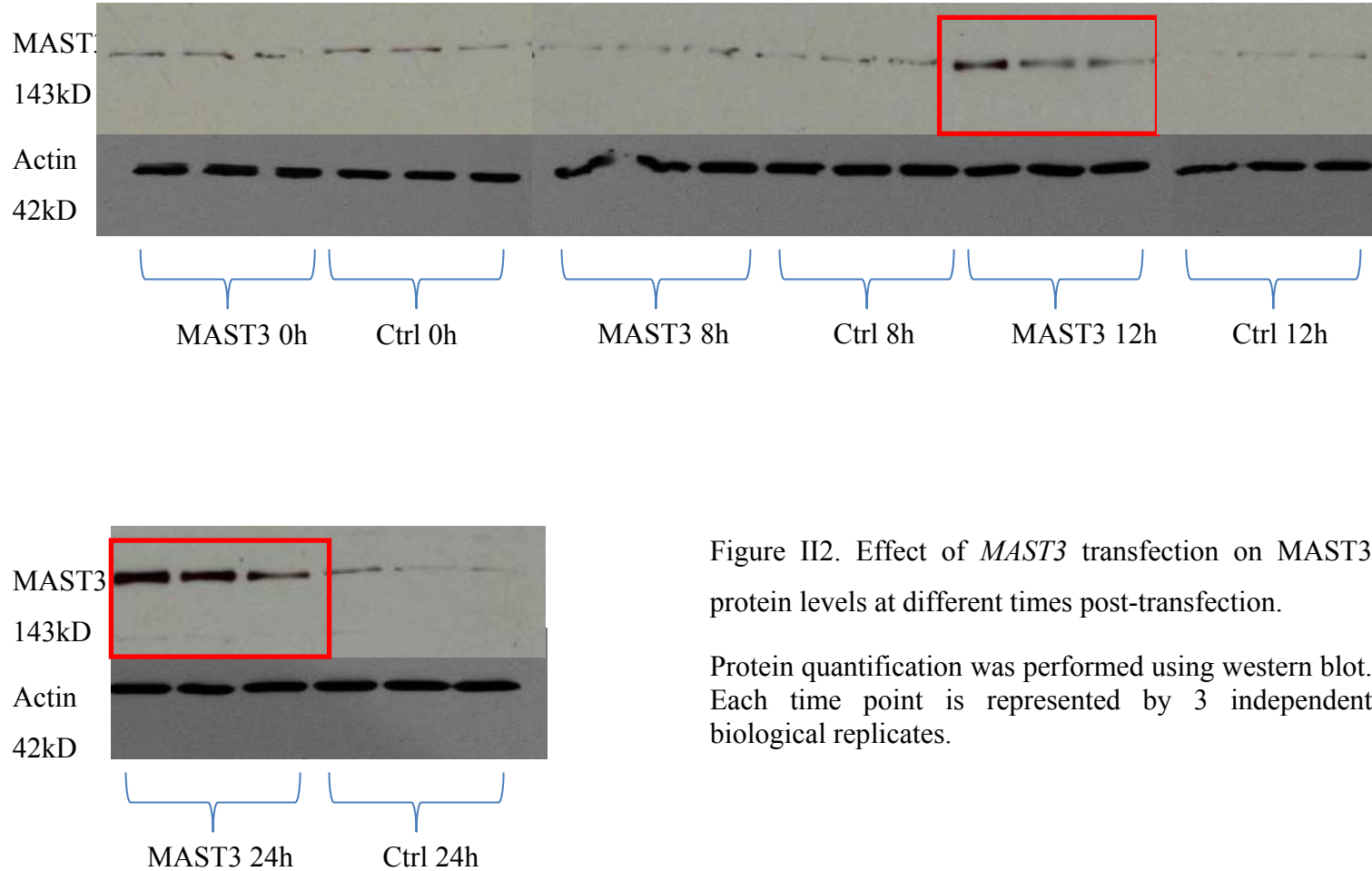


Figure II2. Effect of *MAST3* transfection on MAST3 protein levels at different times post-transfection.

Protein quantification was performed using western blot. Each time point is represented by 3 independent biological replicates.

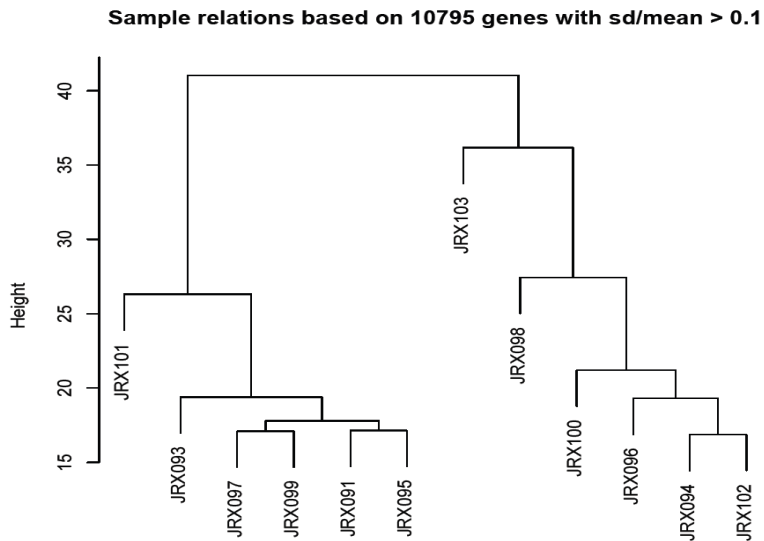


Figure II3. Unsupervised hierarchical clustering of normalized expression data.

The branches of the tree clearly separates the mock-transfected cell samples (to the left) from the samples of cells overexpressing *MAST3* (to the right) based on global gene expression.

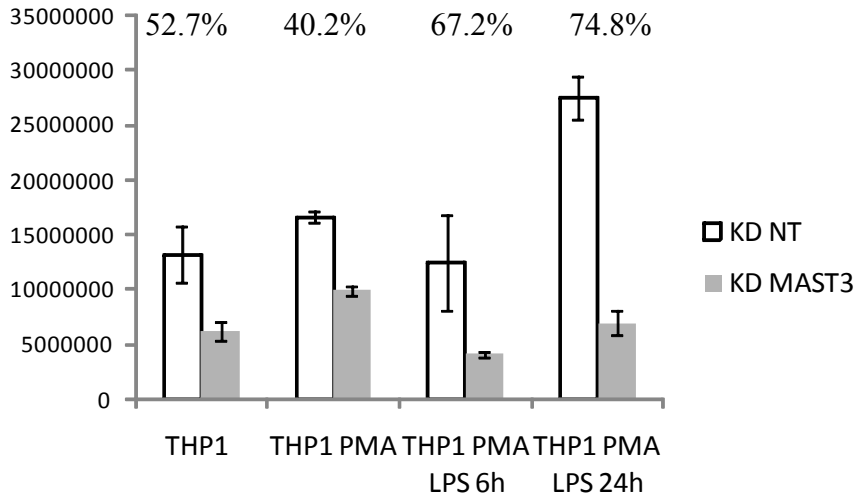


Figure II4. Effect of a KD of *MAST3* on *MAST3* mRNA levels.

Expression is normalized to *HPRT1* expression. Error bars represent the standard deviation between 3 independent biological replicates. The percentages represent the drop in expression in *MAST3* knocked down cells (KD *MAST3*) compared to non target knockdown control (KD NT)

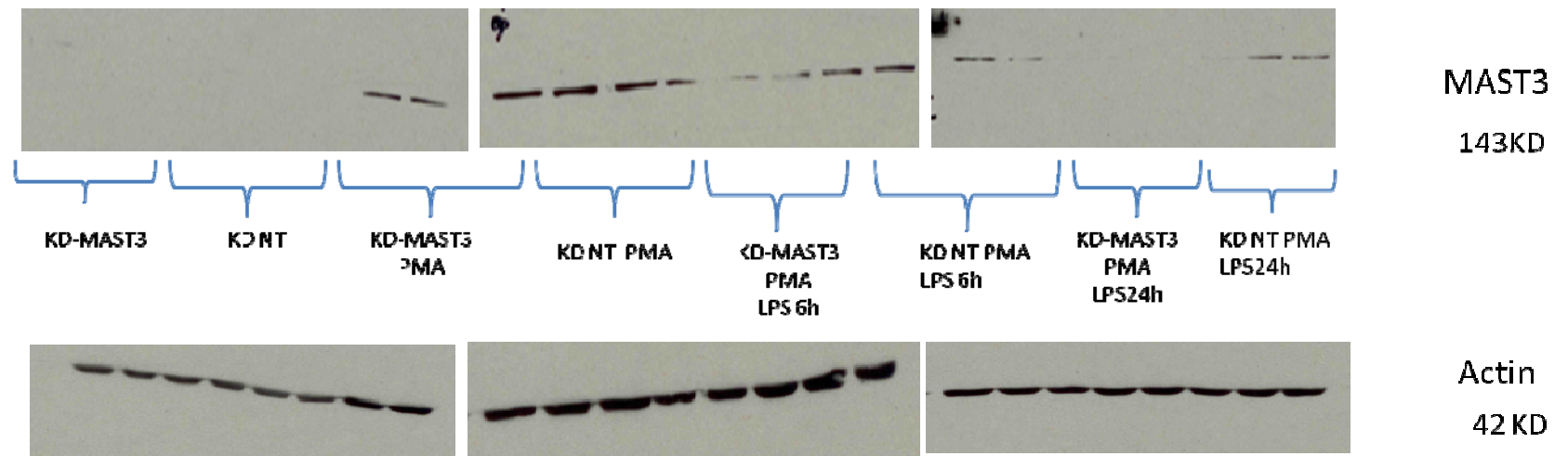
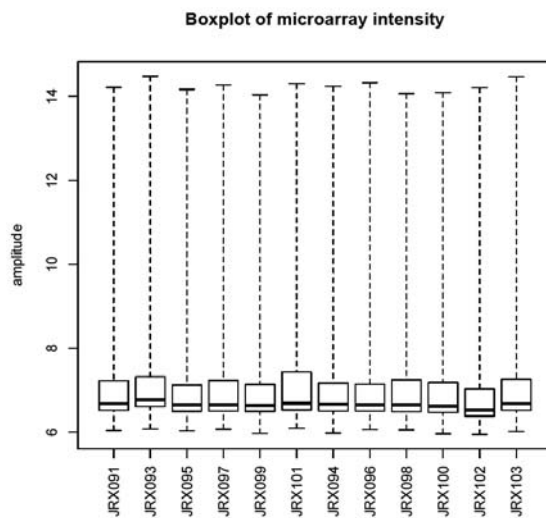


Figure II5. Effect of a stable *MAST3* knockdown on *MAST3* protein levels at different times post-transfection. Protein quantification was performed using western blot. Each time point is represented by 3 independent biological replicates. KD MAST3= knockdown *MAST3*, KD NT= non target control shRNA

A



B

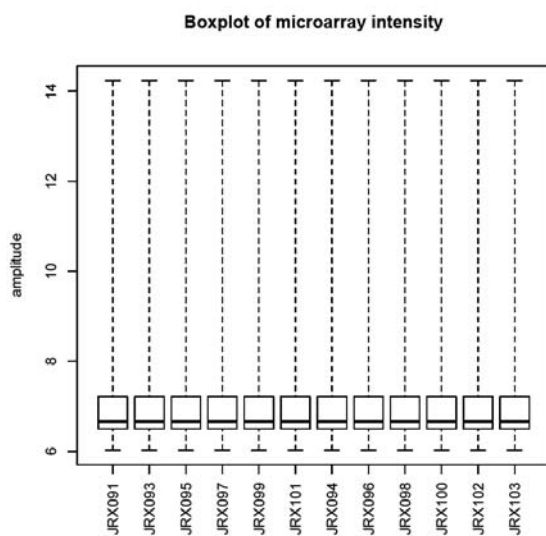
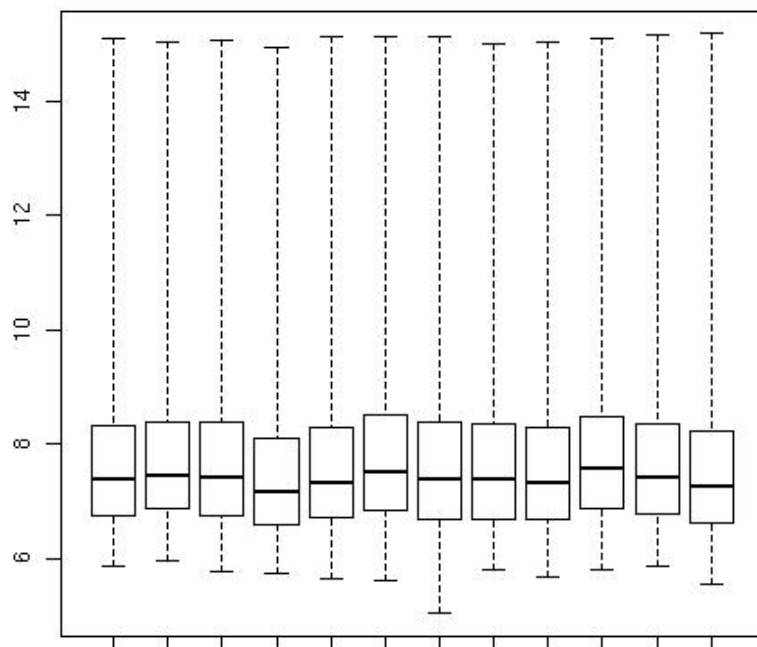


Figure II6. *MAST3* overexpression data normalization.

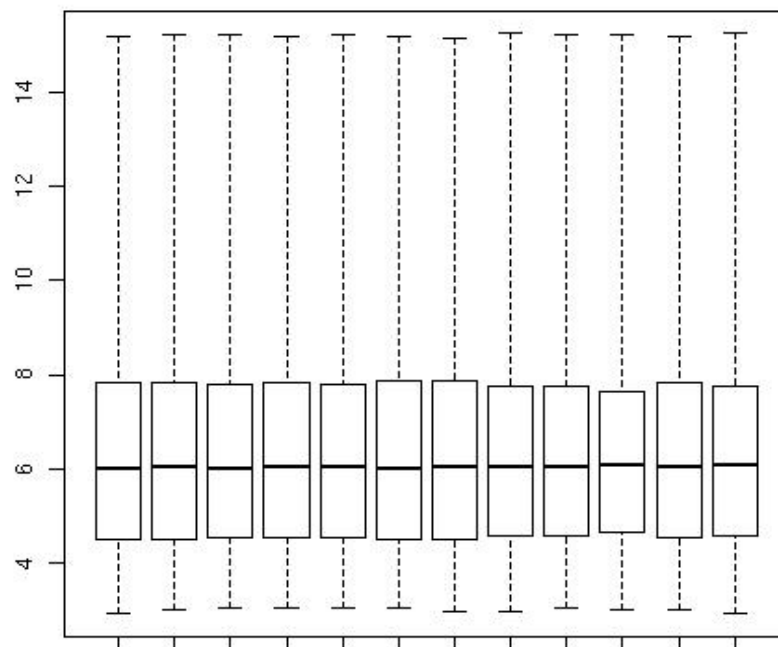
Quality control was performed on expression data. Raw expression data (A) was adjusted by \log_2 transformation and global mean normalization. The boxplot in B presents the adjusted data.

A (BC vs BPN batch 1)

Before normalization

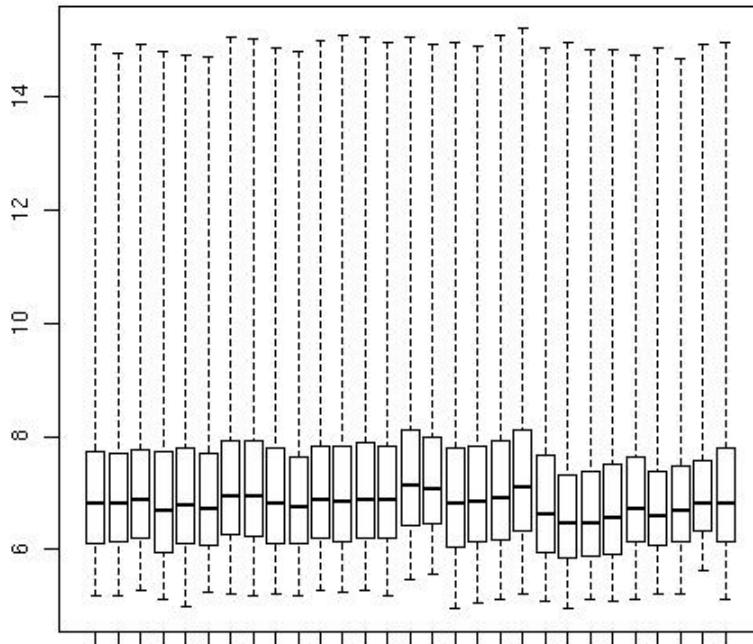


After normalization

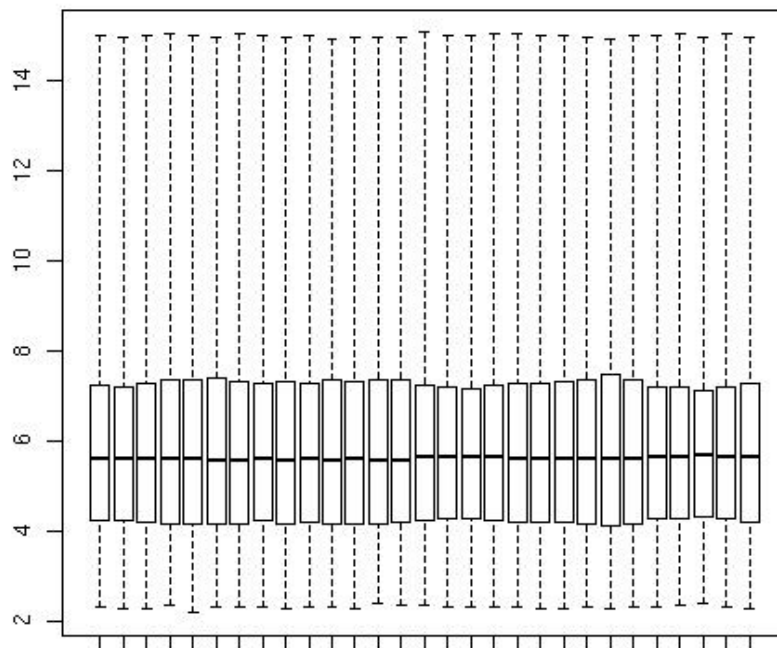


B (BC vs BPN batch 2)

Before normalization

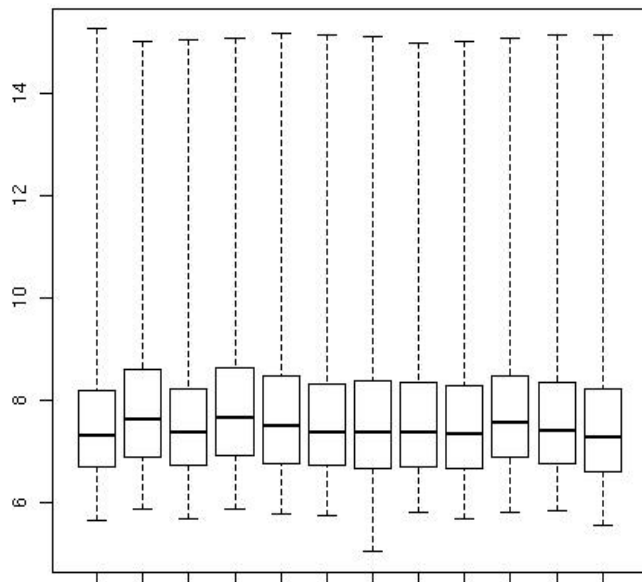


After normalization

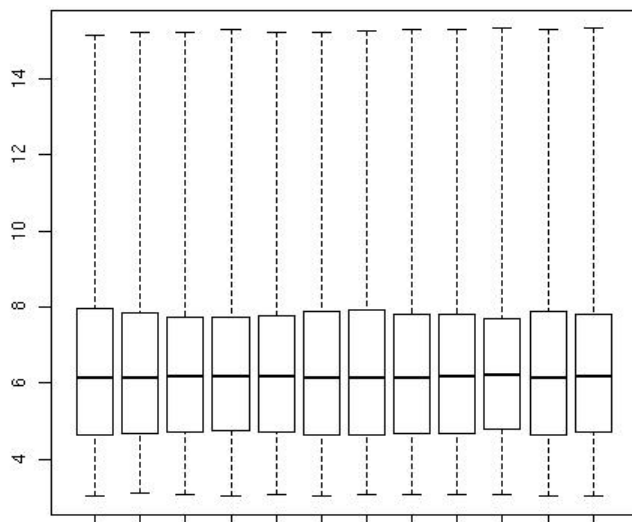


C (BPN vs BPI batch 1)

Before normalization

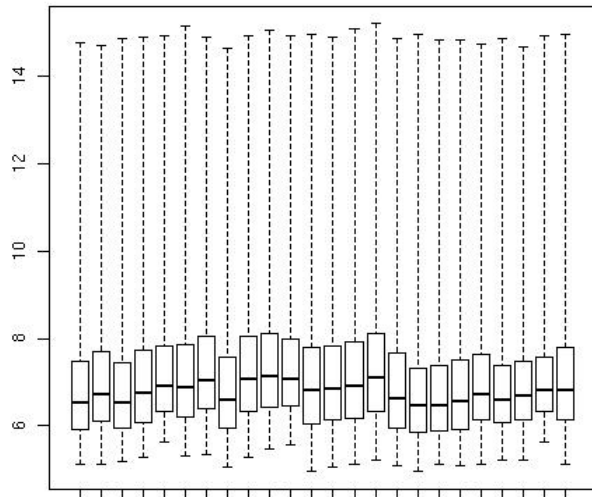


After normalization



D (BPN vs BPI batch 2)

Before normalization



After normalization

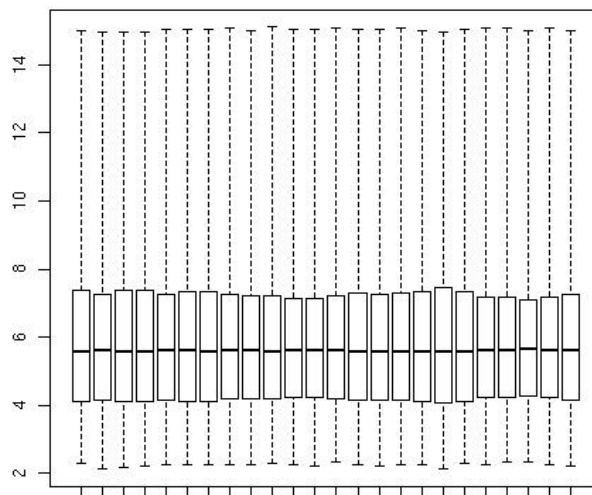


Figure II7. Biopsy expression data normalization

Raw expression was adjusted using quantile normalization with the medium algorithm. Before and After box plots represent global expression per sample.

Annexe III-CV**Catherine Labbé**

Candidate au doctorat

DOSSIER ACADÉMIQUE

Ph.D. en Sciences biomédicales Faculté de médecine, Université de Montréal, <i>Montréal, QC</i>	2005-2011
M.Sc. en Biologie moléculaire Faculté de médecine, Université de Montréal, <i>Montréal, QC</i>	2003-2005
B.Sc. en Biologie Faculté des Sciences, Université McGill, <i>Montréal, QC</i>	2000-2003

FORMATION EN RECHERCHE

Centre de recherche de l'Institut de Cardiologie de Montréal, <i>Montréal, QC</i> Étudiante graduée en génétique Thèse: MAST3 : facteur de risque génétique aux maladies inflammatoires de l'intestin et modulateur d'inflammation Directeur: John Rioux	2005-2011
Centre de recherche du CHU Ste- Justine, <i>Montréal, QC</i> Étudiante graduée en génétique des populations Thèse: Caractérisation de la variabilité dans les gènes du métabolisme du glutathion Directeur: Damien Labuda	2003-2005

PRIX ET DISTINCTIONS

Bourse de formation de doctorat (\$20 000/an pour 3 ans) Fonds de la recherche en santé du Québec (FRSQ)	2007-2009
1 ^{er} prix présentation orale Ph.D.en recherche fondamentale (\$1500) XIIe Journée de la recherche, Institut de cardiologie de Montréal, Montréal, QC	2008
1 ^{er} prix présentation par affiche Ph.D. en recherche fondamentale (\$1000) Réseau de Médecine Génétique Appliquée du Québec (RMGA) Meeting, Québec, QC	2008
Bourse de rédaction (\$400) Centre de recherche du CHU Sainte-Justine, Montréal, QC	2005

PUBLICATIONS RÉVISÉES PAR LES PAIRS

1. Labbé C, Boucher G, Foisy S, Alikashani A, Nkwimi H, David G, Beaudoin M, Goyette P, Charron G, Xavier RJ, Rioux JD (2011) Genome-wide expression profiling implicates a MAST3-regulated gene set in colonic mucosal inflammation of ulcerative colitis patients. *Inflammatory Bowel Diseases*. Published online October 12.
2. Labbé C, Goyette P, Lefebvre C, Stevens C, Green T, Tello-Ruiz MK, Cao Z, Landry A, Stempak J, Annese V, Latiano A, Brant S, Duerr R, Taylor K, Cho J, Steinhart H, Daly M, Silverberg M, Xavier R, Rioux JD (2008) MAST3: a Novel IBD Risk Factor that Modulates TLR4 Signalling. *Genes and Immunity*. Oct;9(7):602-12.
3. Labuda D, Labbé C, Langlois S, Lefebvre JF, Freytag V, Moreau C, Sawicki J, Beaulieu P, Pastinen T, Hudson TJ, Sinnett D (2007) Patterns of variation in DNA segments upstream of transcription start sites. *Hum Mutat*. May;28(5):441-50.

ARTICLES DE REVUE ET CHAPITRE DE LIVRES

1. Dedenus S, Labbé C, Phillips M., Tardif JC, Rioux JD, "Section 6: Chapter 38: Perspective on the future in Pharmacogenomics" in *Acute Coronary Syndromes 2nd Edition*, Chicago, Saunders, 2010
2. Labbé C, Rioux JD (2009) Démystifier le lupus à l'ère de la révolution pangénomique. *Le spécialiste*. 11 (1): 36-37.
3. Budarf M, Labbé C, David G, Rioux JD (2009) GWA studies: rewriting the story of IBD. *Trends in Genetics*. 25(3):137-46.
4. Goyette P, Labbé C, Trinh TT, Xavier RJ, Rioux JD (2007) Molecular pathogenesis of inflammatory bowel disease: genotypes, phenotypes and personalized medicine. *Ann Med*. 39(3):177-99.

CONGRÈS SCIENTIFIQUES

1. 12th International Congress of Human Genetics
Montréal, QC, Octobre 2011
Labbé C et al. Genome-wide expression profiling implicates a MAST3-regulated gene set in colonic mucosal inflammation of ulcerative colitis patients.
2. 78^e Congrès de l'ACFAS
Montréal, QC, Mai 2010
Labbé C et al. MAST3 et l'inflammation : de la génétique au fonctionnel.
3. Journées Génétiques du Réseau de Médecine Génétique Appliquée du Québec (RMGA)
Montréal, QC, Mai 2010
Labbé C et al. MAST3 et l'inflammation : de la génétique au fonctionnel.
4. American Society of Human Genetics 58th Annual Meeting
Honolulu, HI, Novembre 2009
Labbé C et al. Role of IBD Risk Factor MAST3 in TLR4 Signaling.

5. American Society of Human Genetics 58th Annual Meeting
Philadelphie, PA, Novembre 2008
Labbé C et al. MAST3: a Novel IBD Risk Factor that Modulates TLR4 Signalling.
6. Journées génétiques du Réseau de Médecine Génétique Appliquée du Québec (RMGA)
Québec, QC, Mai 2008
Labbé C et al. Le gène MAST3 associé aux maladies inflammatoires de l'intestin.
7. Premier colloque canadien annuel sur la génétique humaine
St-Sauveur, QC, Avril 2008
Labbé C et al. Coding Variants Implicate MAST3 in IBD Susceptibility.
8. American Society of Human Genetics 57th Annual Meeting
San Diego, CA, Octobre 2007
Labbé C et al. High Density Association Mapping of IBD6.
9. 12th Human Genome Organization Meeting
Montréal, QC, Mai 2007
Labbé C et al. Association Mapping of Inflammatory Bowel Disease Susceptibility Genes on Chromosome 19p.
10. Retraite annuelle Genome Québec
Montréal, QC, Octobre 2004
Labbé C et al. Polymorphisms in Genes of Glutathione Metabolism: Promoter Regions.
11. American Society of Human Genetics 54th Annual Meeting
Toronto, ON, Octobre 2004
Labbé C et al. Polymorphisms in Genes of Glutathione Metabolism: Promoter Regions.

