

# GÉNOMIQUE DES MALADIES INFLAMMATOIRES INTESTINALES

## Base pour une nouvelle classification moléculaire de ces maladies et source de nouvelles stratégies thérapeutiques

E. LOUIS (1), C. LIBIOLLE (2), C. REENAERS (3), J. BELAICHE (4), M. GEORGES (5)

**RÉSUMÉ :** Les maladies inflammatoires intestinales, la maladie de Crohn et la rectocolite ulcéro-hémorragique, sont des maladies complexes, multifactorielles et polygéniques. Des progrès considérables dans la connaissance du génome humain et les techniques de génotypage ont permis l'identification de plusieurs dizaines de gènes et loci associés à ces maladies. Ces découvertes lancent les bases d'une nouvelle classification moléculaire de ces maladies et laissent entrevoir de nouvelles possibilités de traitement plus spécifiques. Le présent article résume, dans le cadre de la réunion Synthèse CHU 2009, les contributions de l'équipe de recherche du CHU et de l'Université de Liège dans ce domaine.

**MOTS-CLÉS :** *Maladie de Crohn - Rectocolite ulcéro-hémorragique - Génétique*

**GENOMICS OF INFLAMMATORY BOWEL DISEASES : BASIS FOR A NEW MOLECULAR CLASSIFICATION AND NEW THERAPEUTIC STRATEGIES OF THESE DISEASES**

**SUMMARY :** Inflammatory Bowel Disease, Crohn's disease and ulcerative colitis, are complex, multifactorial, polygenic diseases. Huge progresses in the knowledge of human genome and genotyping techniques have allowed the identifications of dozens of genes and loci associated with these diseases. These discoveries set the path for a new molecular classification of these diseases and let us see new therapeutic possibilities. The present paper highlights, in the setting of the Synthèse CHU 2009 meeting, the contributions of the team of the CHU and university of Liège in this field.

**KEYWORDS :** *Crohn's disease - Ulcerative colitis - Genetics*

### INTRODUCTION

Les maladies inflammatoires intestinales (MICI) se caractérisent par une atteinte chronique inflammatoire, du tube digestif. Elles font partie de ce qu'il est maintenant courant d'appeler les «Immune Mediated Inflammatory Disorders (IMIDs)» au même titre, par exemple, que la polyarthrite rhumatoïde, les spondylarthropathies séronégatives, ou le psoriasis. On reconnaît classiquement deux grands types de MICI : la maladie de Crohn (MC) et la rectocolite ulcéro-hémorragique (RCUH). La première se caractérise par une atteinte multifocale et transmurale de la paroi digestive pouvant toucher n'importe quel segment de la bouche à l'anus. La deuxième se caractérise par une atteinte muqueuse d'un seul tenant, démarrante au rectum et remontant sur une longueur variable, mais n'atteignant que le côlon. Cette distinction est une réalité clinique, récemment confirmée par les avancées génétiques. Néanmoins il existe une grande variété de présentation, tant de la MC que de la RCUH avec des formes de passage entre ces deux affections. L'origine partiellement génétique de ces affections est suspectée depuis longtemps en raison de leur fréquence variable en fonction des ethnies, des formes familiales et

surtout de la concordance nettement plus élevée pour ces maladies parmi les jumeaux monozygotes que dizygotes. Après plusieurs études de gènes-candidats infructueuses, une première série de criblages génomiques par étude de liaison basée sur des marqueurs microsatellites a permis d'identifier dans les années 1990 une série de loci associés à la transmission des MICI (1). Néanmoins, ces loci contenaient en général un grand nombre de gènes (100-1.000) et ce n'est qu'en 2001 que le premier gène clairement associé à la MC a été découvert (2). Ce gène, NOD2/CARD15, est encore aujourd'hui le gène le plus fortement associé au développement de la MC. Dans le courant des années 2000, une meilleure connaissance de l'architecture du génome humain par son séquençage et par la caractérisation des blocs haplotypiques (projet HaMap) (3), l'avènement des techniques de génotypage à large échelle et la constitution de larges cohortes de patients par la création de collaborations internationales, ont permis la réalisation de criblages complets du génome par étude d'association (4). Ces criblages ont abouti à la mise en évidence de plusieurs dizaines de gènes associés aux MICI. Parallèlement, les études par gènes-candidats sont devenues plus efficaces grâce à une meilleure connaissance de la biologie de l'inflammation ainsi qu'à une meilleure capacité d'analyse de l'ensemble des variants principaux caractérisant les différents gènes de l'espèce humaine. Le service de Gastroentérologie du CHU de Liège, avec les laboratoires de génétique de GIGA-Research, et dans le cadre

(1) Maître de Recherche FNRS, (3) Chargé de Recherche FNRS, (4) Professeur, Service de Gastroentérologie, CHU de Liège.

(2) Chercheur, (5) Professeur, Unité de Génétique animale, GIGA-Research, Université de Liège.

de collaborations internationales, a été au centre de ces avancées. Le présent article résume les principales contributions de notre groupe de recherche à ces avancées.

#### **DÉCOUVERTE DE VARIANTS D'IRF5 ASSOCIÉS À LA RCUH (5)**

L'«Interferon Regulatory Factor 5» (IRF5) est un facteur de transcription jouant un rôle à la fois dans l'immunité innée et la réponse immunitaire à médiation cellulaire. Une association de variants au niveau de ce gène a déjà été mise en évidence dans le lupus érythémateux disséminé et la polyarthrite rhumatoïde. Dans le cadre d'une collaboration entre le service de Génétique humaine du CHU de Liège, le département de Médecine moléculaire d'Uppsala en Suède, le service de Gastroentérologie du CHU de Liège et celui du Gasthuisberg à Leuven, douze polymorphismes couvrant l'ensemble du gène IRF5 ont été génotypés sur deux cohortes successives de MICI et de contrôles (au total 1.694 MICI et 552 contrôles). Une association hautement significative a pu être mise en évidence et confirmée entre un variant de type indel dans le promoteur du gène IRF5 et les MICI. Cette association, retrouvée aussi bien avec la MC que la RCUH, était la plus forte avec cette deuxième maladie. Le variant associé aux MICI était différent de ceux précédemment associés au lupus ou à la polyarthrite. Il est susceptible de modifier la séquence d'un site de liaison pour un facteur de transcription SP1 et pourrait donc avoir une signification fonctionnelle et représenter une mutation causale. Cette découverte a été une des toutes premières mises en évidence d'un gène clairement associé à la RCUH.

#### **DÉCOUVERTE DE VARIANTS DE NLRP3 ASSOCIÉS À LA MALADIE DE CROHN (6)**

Le gène NLRP3 (=NALP3), codant pour la cryopyrine, fait partie de la même famille de protéines que NOD2/CARD15, partageant avec ceux-ci les domaines centraux NACHT/NAD et répétitions LRR carboxyterminales, mais s'en différenciant par un domaine amino-terminal «pyrine» (PYD) en lieu et place du domaine CARD. La cryopyrine fait partie de l'inflammasome, plateforme moléculaire régulant l'activation de la caspase-1 et le processus de production de l'IL-1 $\beta$ , deux médiateurs clés de l'inflammation. Des mutations de type «gain-de-fonction» sont responsables de trois formes de fièvre périodique familiale caractérisées par une surproduction d'IL-1 $\beta$ .

Afin d'étudier la contribution éventuelle de NLRP3 à la prédisposition à la MC, dans le cadre d'une collaboration entre l'Université McGill à Montréal, le service de Gastroentérologie de l'hôpital Erasme à Bruxelles, du Gasthuisberg à Leuven et celui du CHU de Liège, Villani et al. (6) ont d'abord effectué une analyse de type TDT dans 296 trios à l'aide de 47 SNP chevauchant le gène NLRP3. Une distorsion de transmission suggestive fut observée pour trois SNP localisés à 4.7 Kb en aval du gène NLRP3, séparés de celui-ci par le petit gène OR2B11 codant pour un récepteur olfactif. Ces trois SNP furent ensuite testés dans quatre autres cohortes totalisant 414 trios, 239 cas et 107 contrôles. L'association initialement détectée fut confirmée de façon remarquablement consistante. Un effort de séquençage identifia 79 nouveaux polymorphismes dont 24 furent génotypés dans l'ensemble des cohortes, confirmant l'association caractérisée pour le meilleur SNP par une valeur p de 10<sup>-8</sup> (répondant aux critères sévères des GWAS) et un OR de 1,78.

La position des SNP associés, en aval du principal gène candidat, suggérait fortement l'intervention d'un polymorphisme régulateur. Afin de tester cette hypothèse, les auteurs examinèrent l'effet des SNP associés sur les taux d'expression de NLRP3 dans des leucocytes et monocytes sanguins. Les allèles augmentant le risque de MC s'avéraient significativement associés à une diminution du taux d'expression de NLRP3 ainsi qu'à une diminution de la sécrétion d'IL-1 $\beta$  par les monocytes, particulièrement après stimulation au LPS.

Ces résultats plaident fortement en faveur de la contribution de variants fréquents du gène NLRP3 à la prédisposition à la MC, et confirme le rôle clé de mécanismes d'immunité innée dans sa pathogénie.

#### **DÉCOUVERTE DE VARIANTS NON CODANTS, POTENTIELLEMENT RÉGULATEURS DE L'EXPRESSION DE PTGER4, ASSOCIÉS À LA MALADIE DE CROHN (7)**

Un des premiers criblages du génome entier par étude d'association dans la maladie de Crohn a été réalisé par le groupe de recherche liégeois, associant le service de Gastroentérologie du CHU de Liège et le laboratoire de Génétique factorielle du GIGA-Research. Pour ce travail, la cohorte de patients a été étendue grâce à une collaboration avec les services de Gastroentérologie universitaires de Leuven, Gent et Bruxelles et les génotypages ont été réalisés en collaboration avec le Génopôle d'Evry.

Ce travail, outre la confirmation des associations avec le gène NOD2/CARD15, et celui du récepteur de l'IL23, a mis en évidence pour la première fois une association entre la MC et une région identifiée en 5p13. Cette découverte est particulièrement intrigante, car la région est dépourvue de gènes connus. Qui plus est, elle se situe dans un désert génique couvrant plus d'un million de paires de bases. Comment une région sans gènes pourrait-elle influencer la prédisposition à la MC ? Une hypothèse possible est que cette région contienne des éléments régulateurs du taux d'expression génique agissant à longue distance (tels que décrits par exemple pour les gènes de globines). En perturbant le fonctionnement de tels éléments régulateurs, les variants de prédispositions influenceraient le taux d'expression de leurs gènes cibles affectant ainsi la prédisposition à la MC. Afin de tester cette hypothèse, l'équipe de recherche a mis à profit des données de transcriptomiques générées sur des lignées lymphoblastoïdes de 400, individus par ailleurs génotypés pour 300.000 SNP. Elle démontre que, dans ces cellules, les polymorphismes associés à la MC ont un effet très significatif sur les taux d'expression de PTGER4 situés à plus de 250 Kb. Ce type d'effet est qualifié d'eQTL pour «Expression QTL». Le gène PTGER4 code pour le récepteur de type EP4 de la prostaglandine PGE2. Le rôle des prostanoïdes dans les phénomènes inflammatoires est bien documenté et, de plus, l'inactivation de PTGER4 chez la souris augmente très significativement leur sensibilité à la colite induite par le dextran sulphate. PTGER4 s'avère donc être un excellent gène candidat pour le nouveau locus de prédisposition en 5p13.

#### IDENTIFICATION PAR MÉTA-ANALYSE DE PLUSIEURS DIZAINES DE LOCI ET GÈNES ASSOCIÉS À LA MALADIE DE CROHN (8)

Une observation importante résultant des premiers criblages du génome entier par étude d'association est la faiblesse de quasi tous les effets observés, la majorité des OR associés aux gènes ou aux marqueurs identifiés étant inférieurs à 1,5. Une analyse statistique *a posteriori* indique que ces criblages individuels, y compris le plus important réalisé par le Wellcome Trust Case Control Consortium, n'avaient qu'une faible puissance (~0,30) de détection d'effets aussi modestes. Force est donc de conclure que la majorité des effets existants n'ont tout simplement pas été détectés. La façon la plus efficace d'augmenter la puissance de détection est tout simplement la force de l'union : analyser conjointement plusieurs GWAS. En 2008,

trois des équipes (l'équipe liégeoise associée au Génopôle d'Evry et avec le renfort de cohortes de patients de Leuven, Gent et Bruxelles, un consortium Nord-Américain financé par le NIH – et le Wellcome Trust Consortium-consortium Britannique) ayant réalisé des criblages individuels, s'unissent dans une telle méta-analyse d'un total de 3.230 patients et de 4.829 contrôles caucasiens. Afin de fusionner des données génotypiques obtenues avec des panels de SNP distincts (panels Illumina 300K pour les criblages américains et belgo-français *versus* panel Affymetrix 500K pour le criblage anglais), les équipes recourent à de l'imputation génotypique : elles vont effectuer du génotypage *in silico* pour des SNP typés réellement par leur partenaire mais pas par eux. Ce subterfuge est rendu possible grâce au projet HapMap, dans le cadre duquel une centaine d'individus caucasiens ont été génotypés pour plus de 3 millions de SNP comprenant tous les marqueurs utilisés pour les criblages du génome entier. Le déséquilibre de liaison entre marqueurs (c'est-à-dire les associations non aléatoires entre SNP) évalué en détail dans la population HapMap, permet de prédire avec précision le génotype pour des SNP non typés, grâce aux génotypes connus de SNP voisins. Des analyses d'association sont effectuées en prenant bien soin de ne pas créer de nouveaux problèmes de stratification. L'analyse des résultats révèle un excès important de régions hautement significatives. Les 75 régions les plus prometteuses ( $p < 5 \times 10^{-5}$ ) sont ensuite évaluées dans une cohorte de confirmation indépendante comprenant 2.325 cas, 1.809 contrôles et 1.339 trios. Cette analyse confirme 11 des 19 régions rapportées dans les criblages du génome entier individuels (dont toutes celles qui avaient été détectées dans plus d'une étude), et en identifie 20 nouvelles.

Pour chacun des loci de prédisposition identifiés, un intervalle de confiance est défini sur base de l'étendue du déséquilibre de liaison dans la région correspondante. La taille moyenne des intervalles de confiance est de 198 Kb, allant de 19 Kb pour le plus petit intervalle à 720 Kb pour le plus grand. Ces intervalles contiennent 3,7 gènes en moyenne, allant de 0 à 35. Pour quasi tous ces loci, gènes et mutations causales restent à identifier formellement. Néanmoins, plusieurs candidats positionnels sont particulièrement intrigants et semblent renforcer le rôle important de l'IL-23 et des cellules TH17. L'imputation de génotypes HapMap permet d'exclure une contribution prépondérante de SNP altérant la séquence protéique («SNP non-synonymes»), alors que des analyses du transcriptome en cel-

lules lymphoblastoïdes (telle que celle utilisée pour l'identification de PTGER4 (cfr. Plus haut), révèlent un enrichissement significatif en effets eQTL dans les régions associées à la MC, suggérant dès lors un rôle important pour des SNP régulateurs agissant sur les taux d'expression géniques. A l'issue de cette méta-analyse, ayant généré plus de 30 loci de prédisposition avec forts supports statistiques, la MC s'avère être la pathologie complexe la mieux comprise en termes de nombre de loci de prédisposition.

Dans la foulée, 4 criblages du génome entier par étude d'association ont été effectués dans la RCUH. Ceux-ci (9) ont permis l'identification d'association avec toute une série de gènes et loci déjà associés à la MC, mais également d'associations originales notamment avec le gène de l'interleukine-10 (IL-10). Au total, une vingtaine de gènes ou loci sont maintenant également associés à la RCUH.

## LA FLEUR GÉNÉTIQUE DES IMIDS

Une observation inattendue provenant de la méta-analyse mentionnée plus haut est la proportion non négligeable de loci affectant non seulement la prédisposition à la MC ou à la RCUH, mais également à d'autres maladies, certaines associées aux MICI dans le cadre des IMIDs et d'autres sans liens préalablement établis avec les MICI (10) : diabète de type I (PTPN2 et PTPN22), diabète de type II (CDKAL1), psoriasis (IL-23R, IL-12B), arthrite rhumatoïde (PTPN22), spondylite ankylosante (IL-23R), maladie de Parkinson (LRRK2), asthme (ORMDL3). Ces associations communes nous permettent d'avancer le concept de fleur génétique des IMIDs et autres maladies à composante immunitaire (Fig. 1) : chaque entité pathologique est représentée par un pétale. Certains gènes ou loci sont communs à deux ou plus de ces maladies (intersection des pétales). Le cœur de la fleur est constitué par des zones génétiques telles que le Complexe Majeur d'Histocompatibilité, pour lesquelles une association a été découverte avec quasi toutes ces maladies. Ces déterminismes génétiques communs permettent d'expliquer les agrégations familiales de certaines de ces maladies et soulignent certains mécanismes physiopathologiques communs. Compte tenu, par ailleurs, de la grande hétérogénéité génétique de chacune des maladies constituant les IMIDs, on pourrait même aller plus loin dans ce concept et avancer que certaines formes de ces maladies sont parfois génétiquement et moléculairement plus proches de certaines autres IMIDs que d'autres formes extrêmes de ces mêmes maladies. Ce

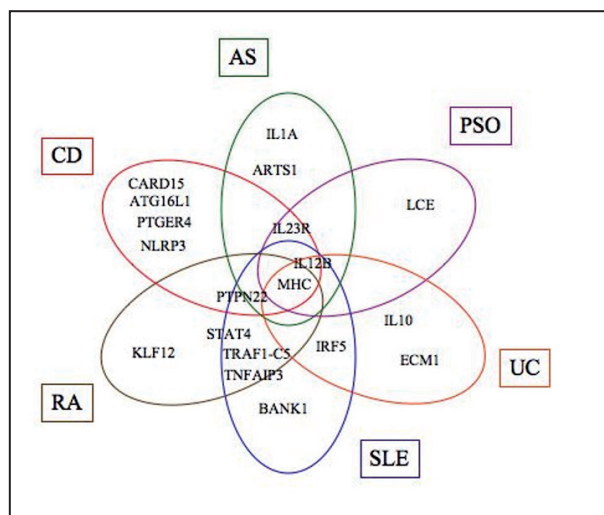


Figure 1. La fleur génétique des «Immune Mediated Inflammatory Disorders» illustre, de façon non exhaustive, le déterminisme génétique partiellement commun de ces affections. Chaque entité pathologique est représentée par un pétale. Certains gènes ou loci sont communs à deux ou plus de ces maladies (intersection des pétales). Le cœur de la fleur est constitué par des zones génétiques telles que le Complexe Majeur d'Histocompatibilité, pour lesquelles une association a été découverte avec quasi toutes ces maladies. Les gènes communs sont généralement impliqués dans l'induction et l'amplification non spécifique de la réaction immuno-inflammatoire. Les gènes plus spécifiques sont souvent impliqués dans l'homéostasie d'organe ou la réponse immunitaire spécifique d'organe.

CD = maladie de Crohn; UC = rectocolite ulcéro-hémorragique, RA = poly-Arthrite rhumatoïde; AS = spondylarthrite ankylosante, PSO = psoriasis; SLE = lupus érythémateux systémique.

concept pourrait nous amener à redéfinir des entités pathologiques moléculaires plutôt que cliniques. L'importance de la définition de ces entités pathologiques moléculaires pourrait être la définition d'entités plus homogènes et susceptibles de répondre de façon plus uniforme à certains agents pharmacologiques.

## NOUVELLES VOIES THÉRAPEUTIQUES

Toutes ces découvertes récentes faites dans le domaine des maladies complexes, et en particuliers des MICI, n'ont pas, à l'heure actuelle, d'impact diagnostique. Les risques relatifs associés à ces différents marqueurs génétiques apparaissent trop faibles pour avoir un quelconque intérêt dans le diagnostic. Par contre, ces découvertes contribuent de façon importante à une meilleure compréhension des mécanismes physiopathologiques sous-jacents à ces diverses entités. Dans la foulée, elles jettent les bases pour la recherche de nouveaux traitements plus spécifiques et physiopathologiques de ces maladies. Le meilleur exemple est constitué par les inhibiteurs de l'IL-23. Deux anticorps dirigés contre la sous-unité p40 commune à l'IL-12 et l'IL-23 font actuellement l'objet d'études de phases 2 et 3 dans les MICI (11). Ces anticorps font également l'objet d'études prometteuses dans d'autres IMIDs, dont le psoriasis. Une autre piste

encourageante est certainement celle de l'IL-10, notamment dans la RCUH, entité dans laquelle une association a été clairement établie avec des variants de ce gène. Néanmoins, des études réalisées avec de l'IL-10 humaine recombinante réalisées notamment dans la MC, il y a une dizaine d'années, avaient donné des résultats décevants. Les données génétiques récentes indiquent que cette voie de traitement ne doit peut-être pas être abandonnée, mais que le mode de délivrance de l'IL-10 doit probablement être repensé. Ce qui importe en effet, c'est la concentration d'IL-10 au voisinage de l'épithélium intestinal et des cellules immunitaires de la muqueuse. Dans cet esprit, la production d'IL-10 par des probiotiques (12), ou la capacité à stimuler la production d'IL-10 par des probiotiques, pourrait être une voie plus adéquate pour les MICI que l'injection systémique d'IL-10 recombinante. Dans cet esprit, notre groupe travaille, en collaboration avec la *spin-off* de l'Université de Liège THT, à l'identification de souches probiotiques stimulant particulièrement la production d'IL-10 par les cellules immunitaires.

## CONCLUSION

Des études récentes, auxquelles a participé un groupe de recherche liégeois associant principalement le service de Gastroentérologie du CHU de Liège et le laboratoire de Génomique animale du GIGA-Research, ont permis de mettre en évidence plusieurs dizaines de gènes ou loci associés aux MICI. Ces découvertes représentent une avancée considérable dans la connaissance de l'architecture de la prédisposition génétique aux maladies complexes, et plus particulièrement dans la compréhension de la physiopathologie des MICI. Elles jettent les bases pour une nouvelle classification moléculaire de ces maladies et pour de nouvelles voies thérapeutiques plus spécifiques.

## BIBLIOGRAPHIE

1. Mathew CG, Lewis CM.— Genetics of inflammatory bowel disease : progress and prospects. *Hum Mol Genet*, 2004, **13**, 161-168.
2. Hugot JP, Chamaillard M, Zouali H, et al.— Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 2001, **411**, 599-603.
3. International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, **437**, 1299-1320.
4. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007, **447**, 661-678.
5. Dideberg V, Kristjansdottir G, Milani L, et al.— An insertion-deletion polymorphism in the interferon regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. *Hum Mol Genet*, 2007, **16**, 3008-3016.
6. Villani AC, Lemire M, Fortin G, et al.— Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. *Nat Genet*, 2009, **41**, 71-76.
7. Libioulle C, Louis E, Hansoul S, 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**, 538-543.
8. Barrett JC, Hansoul S, Nicolae DL, et al.— Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*, 2008, **40**, 955-962.
9. Louis E, Libioulle C, Reenaers C, et al.— Genetics of Ulcerative colitis : the come-back of interleukin 10. *Gut*, 2009, in press.
10. Lettre G, Rioux J.— Autoimmune diseases : insights from genome-wide association studies. *Hum Mol Genet*, 2008, **17**, 116-121.
11. Sandborn WJ, Feagan BG, Fedorak RN, et al.— A randomized trial of Ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. *Gastroenterology*, 2008, **135**, 1130-1141.
12. Braat H, Rottiers P, Hommes DW, et al.— A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. *Clin Gastroenterol Hepatol*, 2006, **4**, 754-759.

Les demandes de tirés à part sont à adresser au Pr. E. Louis, Service de Gastroentérologie, CHU de Liège, 4000 Liège, Belgique.  
Email : edouard.louis@ulg.ac.be