Association between genetic variants in the \textit{HNF4A} gene and childhood-onset Crohn’s disease

V Marcil\textsuperscript{1}, D Sinnett\textsuperscript{2}, E Seidman\textsuperscript{1,3}, F Boudreau\textsuperscript{4}, F-P Gendron\textsuperscript{3}, J-F Beaulieu\textsuperscript{3}, D Menard\textsuperscript{3}, M Lambert\textsuperscript{2}, A Bitton\textsuperscript{1}, R Sanchez\textsuperscript{2}, D Amre\textsuperscript{2} and E Levy\textsuperscript{3,4}

Hepatocyte nuclear factor 4 alpha (HNF4\textalpha), involved in glucose and lipid metabolism, has been linked to intestinal inflammation and abnormal mucosal permeability. Moreover, in a genome-wide association study, the HNF4\textalpha locus has been associated with ulcerative colitis. The objective of our study was to evaluate the association between HNF4\textalpha genetic variants and Crohn’s disease (CD) in two distinct Canadian pediatric cohorts. The sequencing of the HNF4\textalpha gene in 40 French Canadian patients led to the identification of 27 single nucleotide polymorphisms (SNPs) with a minor allele frequency >5%. To assess the impact of these SNPs on disease susceptibility, we first conducted a case–control discovery study on 358 subjects with CD and 542 controls. We then carried out a replication study in a separate cohort of 416 cases and 1208 controls. In the discovery cohort, the genotyping of the identified SNPs revealed that six were significantly associated with CD. Among them, rs1884613 was replicated in the second CD cohort (odds ratio (OR): 1.33; \(P < 0.012\)) and this association remained significant when both cohorts were combined and after correction for multiple testing (OR: 1.39; \(P < 0.004\)). An 8-marker P2 promoter haplotype containing rs1884613 was also found associated with CD (\(P < 2.09 \times 10^{-4}\) for combined cohorts). This is the first report showing that the HNF4\textalpha locus may be a common genetic determinant of childhood-onset CD. These findings highlight the importance of the intestinal epithelium and oxidative protection in the pathogenesis of CD.

\textit{Genes and Immunity} (2012) \textbf{13}, 556–565; doi:10.1038/gene.2012.37; published online 23 August 2012

\textbf{Keywords:} Crohn’s disease; HNF4\textalpha; genetic variants; oxidative stress

\section*{INTRODUCTION}

Inflammatory bowel disease (IBD) refers to two chronic inflammatory disorders affecting the intestinal mucosa: Crohn’s disease (CD, (MIM 266600)) and ulcerative colitis (UC, (MIM 191390)). CD is common in developed countries, with a prevalence estimated 1,2 at 100–300/100 000. The etiology of CD has not yet been elucidated, but is considered to involve a complex interaction between predisposing genes, environmental factors and impaired immune response to the commensal gut microbiome. The understanding of the genetic contribution to risk of CD has advanced enormously as a result of recent case–control and genome-wide association studies (GWAS).3–6 Indeed, GWAS,7 followed by deep sequencing of GWAS loci,8 have identified 85 distinct loci associated with the disease. However, the genes identified thus far only explain \(\approx 23\%\) of the genetic contribution to CD.\textsuperscript{7}

Hepatocyte nuclear factor 4 alpha (HNF4\textalpha, NR2A1) belongs to the nuclear hormone receptor superfamily.\textsuperscript{3} It is expressed in the liver, kidney, pancreatic islets and gut.\textsuperscript{9–11} HNF4\textalpha interacts with regulatory elements in promoters and enhancers of genes involved in cholesterol, fatty acid and glucose metabolism,\textsuperscript{12} and maintains glucose homeostasis by regulating gene expression in pancreatic \(\beta\) cells,\textsuperscript{13,14} while the P1 promoter is mainly active in liver cells.\textsuperscript{15,16} Both promoters appear to be effective in the intestine.\textsuperscript{20}

The key hepatic and pancreatic functions of HNF4\textalpha are well established. It activates gluconeogenesis in hepatocytes,\textsuperscript{21} maintains glucose homeostasis by regulating gene expression in pancreatic \(\beta\) cells,\textsuperscript{12,22} activates insulin genes through both direct and indirect mechanisms\textsuperscript{22,23} and regulates the expression of many genes, such as apolipoproteins.\textsuperscript{24} Rare loss-of-function mutations in the HNF4A gene cause a monogenic form of type 2 diabetes (T2D), type 1 maturity-onset diabetes of the young (MODY).\textsuperscript{25} Also, HNF4\textalpha has been reported to be associated with the risk of late-onset T2D in several populations.\textsuperscript{26–28} In the gut, HNF4\textalpha has a role in colonic development,\textsuperscript{29} lipid transport,\textsuperscript{30} as well as intestinal epithelial cell differentiation and phenotype expression.\textsuperscript{31,32} It has also been associated with susceptibility to abnormal intestinal permeability, inflammation and oxidative stress.\textsuperscript{33,34} Of particular relevance, a recent GWAS demonstrated associations between the 20q13.1.1 locus that harbors the HNF4A gene and risk of developing UC.\textsuperscript{35} Interestingly, no associations

\textsuperscript{1}Departments of Medicine and Pediatrics, Research Institute, McGill University, Montreal, Quebec, Canada; \textsuperscript{2}Department of Pediatrics, Research Centre, CHU-Sainte-Justine, Université de Montréal, Montreal, Quebec, Canada; \textsuperscript{3}Department of Anatomy and Cellular Biology, Faculty of Medicine and Health Sciences, Canadian Institutes for Health Research Team on the Digestive Epithelium, Université de Sherbrooke, Sherbrooke, Quebec, Canada and \textsuperscript{4}Department of Nutrition, Research Centre, CHU-Sainte-Justine, Université de Montréal, Montreal, Quebec, Canada. Correspondence: Dr E Levy, Department of Nutrition, Research Centre, GI-Nutrition Unit, CHU-Sainte-Justine, Université de Montréal, 3175 Côte Ste-Catherine, Montreal, Quebec, Canada H3T 1C5. E-mail: emile.levy@recherche-ste-justine.qc.ca

Received 7 May 2012; accepted 27 July 2012; published online 23 August 2012
with CD were found. In this study, we have hypothesized that HNF4A gene polymorphisms are associated with the risk of developing CD. We comprehensively examined the association between variants in and around the HNF4A gene and CD in two distinct cohorts of Canadian children.

RESULTS
SNP discovery by sequencing
To determine the single nucleotide polymorphism (SNP) content of HNF4A in our population, 30 selected fragments of the HNF4A gene were sequenced in 40 IBD French Canadian patients. As summarized in Table 1, sequencing of the gene led to the identification of 27 SNPs with a minor allele frequency >5%. Among the identified SNPs, one was non-synonymous (rs1800961, T130) and 26 were located either in intronic or in promoter regions. All SNPs had been previously reported in dbSNP (build 131). Most of the variants identified in this study were previously associated with the risk of developing T2D and dyslipidemia.

Table 1. Summary of the identified SNPs in the targeted HNF4A regions

| SNP          | Change | Position          | MAF |  
|--------------|--------|-------------------|-----|  
| rs4810424    | G>C    | 42,408,437 (9428 bp upstream exon 1D) | C = 0.203 |  
| rs1884613    | C>G    | 42,413,829 (4036 bp upstream exon 1D) | G = 0.186 |  
| rs1884614    | C>T    | 42,413,933 (3982 bp upstream exon 1D) | T = 0.186 |  
| rs6031543    | G>C    | 42,413,979 (3935 bp upstream exon 1D) | G = 0.157 |  
| rs2144908    | G>A    | 42,419,131 (intron 1A–1D) | A = 0.157 |  
| rs6031555    | C>T    | 42,423,085 (intron 1A–1D) | T = 0.190 |  
| rs6031551    | T>C    | 42,423,127 (intron 1A–1D) | C = 0.180 |  
| rs16988991   | G>A    | 42,423,191 (intron 1A–1D) | A = 0.203 |  
| rs6031552    | C>A    | 42,423,208 (intron 1A–1D) | A = 0.167 |  
| rs6103716    | A>C    | 42,433,044 (intron 1A–1D) | C = 0.242 |  
| rs6031558    | G>C    | 42,433,057 (intron 1A–1D) | C = 0.300 |  
| rs6130608    | T>C    | 42,457,422 (5933 bp upstream exon 1A) | T = 0.250 |  
| rs2425637    | G>T    | 42,457,463 (5892 bp upstream exon 1A) | G = 0.382 |  
| rs2425639    | A>G    | 42,460,924 (2431 bp upstream exon 1A) | A = 0.400 |  
| rs2425640    | A>G    | 42,461,451 (1904 bp upstream exon 1A) | A = 0.295 |  
| rs3212172    | A>G    | 42,461,804 (1551 bp upstream exon 1A) | G = 0.125 |  
| rs1800963    | A>C    | 42,462,699 (656 bp upstream exon 1A) | C = 0.379 |  
| rs2071997    | G>A    | 42,463,849 (661 bp upstream exon 1B) | A = 0.125 |  
| rs736824     | T>C    | 42,468,074 (intron 1A/1B-2) | C = 0.329 |  
| rs745975     | C>T    | 42,468,107 (intron 1A/1B-2) | T = 0.250 |  
| rs3212183    | C>T    | 42,468,552 (intron 2–3) | T = 0.487 |  
| rs3212184    | G>C    | 42,468,574 (intron 2–3) | C = 0.459 |  
| rs1885088    | G>A    | 42,472,454 (intron 3–4) | A = 0.222 |  
| rs1800961    | C>T    | 42,475,778 (exon 4) | T = 0.107 |  
| rs3212195    | G>A    | 42,476,509 (intron 4–5) | A = 0.190 |  
| rs3212198    | C>T    | 42,477,775 (intron 4–5) | T = 0.369 |  
| rs3818247    | G>T    | 42,490,894 (intron 9–10) | T = 0.283 |  

Abbreviation: MAF, minor allele frequency. SNPs revealed by sequencing with a frequency >5% in the inflammatory bowel disease French Canadian sample population (20 Crohn’s disease and 20 ulcerative colitis patients). The SNP positions are based on the March 2006 Homo sapiens high coverage assembly from the Genome Reference Consortium.

The relative positions of SNPs on the HNF4A locus are illustrated in Figure 1.

Genotyping for association with Crohn’s disease in discovery cohort
A total of 356 (271 French Canadian, 57 Jewish and 30 non-Caucasian) subjects with CD and 542 controls were included for genotyping. The descriptive and clinical characteristics of participants of the discovery cohort are shown in Table 2. There was a non-significant higher proportion of males among the cases (53.35%). The mean age at diagnosis (15.41 ± 7.63 years) was similar to age of controls (13.67 ± 2.72 years). Based on the Montreal Classification, most cases (n = 224, 62.57%) had ileocolonic location (L3 ± L4) and inflammatory disease (B1 ± P) (n = 287, 80.17%). The majority of the population was of Caucasian ancestry (n = 271, 75.70%).

Among the 27 SNPs identified, three could not be adequately genotyped owing to technical difficulties (rs2425640, rs16988991 and rs3212184). The remaining 24 SNPs were analyzed for association. Table 3 shows the distribution of the frequencies of the corresponding alleles in cases and controls. Six SNPs demonstrated significant associations with CD: rs4810424 (P < 0.007), rs1884613 (P < 0.004), rs1884614 (P < 0.005), rs2144908 (P < 0.003), rs3212172 (P < 0.044) and rs1800963 (P < 0.048). Analysis including only individuals of Caucasian ancestry revealed similar results. However, the associations for two SNPs (rs3212172 and rs1800963) were no longer significant probably owing to reduced power.

Genotyping for association with Crohn’s disease in replication cohort
For replication, we selected 10 SNPs significantly associated with CD in the single SNP and haplotype analyses of the discovery study. A total of 416 Caucasian subjects with CD and 1208 controls were included for genotyping. The descriptive and clinical characteristics of participants of the replication cohort are shown in Table 2. The proportion of males among the cases was higher (56.49%), but the difference was not significant. The mean age at diagnosis (15.41 ± 3.41 years) was similar to that of controls (12.71 ± 2.98 years). A high percentage of cases (n = 200, 48.08%) had ileocolonic location (L3 ± L4) and inflammatory disease (B1 ± P) (n = 365, 87.75%). All subjects in replication cohort were of Caucasian ancestry. Table 4 shows the distribution of the frequencies of the corresponding alleles in cases and controls. All SNPs were in Hardy–Weinberg equilibrium. Among the 10 SNPs genotyped for replication, rs1884613 remained significantly associated with CD (odds ratio (OR): 1.327; P < 0.012).

Single SNP analysis in combined cohorts
The descriptive and clinical characteristics of participants of the combined discovery and replication cohorts are shown in Table 2. Association analysis revealed a significant association for three of the six SNPs associated in the discovery cohort, namely rs1884613 (OR: 1.389, P < 0.0001), rs1884614 (OR: 1.295, P < 0.001) and rs2144908 (OR: 1.260, P < 0.012). After correction for multiple testing (40 tests), the association for rs1884613 remained significant (P < 0.004) and P < 0.04, respectively.

Haplotype analysis
Linkage disequilibrium (LD) analysis (Figure 2) showed that the SNPs were distributed within six major haplotype blocks: a first block including eight SNPs overlapping Promoter 2 and spanning on a 14-kb region (rs4810424, rs1884613, rs1884614, rs6031543, rs2144908, rs6031550, rs6031551 and rs6031552); a second block of two adjacent intronic SNPs (rs6103716 and rs6031558); a third
block of three SNPs (3 kb) in the intronic region between both promoters (rs6130608, rs2425637 and rs2425639); a fourth block of two SNPs (4 kb) also located in the intronic region between the two promoters (rs2071197 and rs736824); a fifth block of two intronic SNPs (rs745975 and rs3212183, respectively, in introns 1 and 2); and finally a sixth block of three SNPs (5 kb) located in introns 3 and 4 (rs3212195 and rs3212198). Table 5 shows the results of the haplotype analyses performed on the SNPs within each block of LD in the discovery cohort. One 8-marker haplotype was significantly associated with CD (haplotype CGTCACTC, $\chi^2 = 8.276$, $P < 0.004$). Subsequently, association analysis was replicated for the significant P2 promoter haplotype. In the replication cohort, the association with the CGTCACTC haplotype remained significant ($\chi^2 = 8.266$, $P < 0.004$) (Table 5). Combining both cohorts, the significant association was also replicated ($\chi^2 = 19.997$, $P < 7.755 \times 10^{-5}$), even after correcting for 27 other variables.
haplotype comparisons \((P < 2.09 \times 10^{-4})\). Moreover, a second haplotype was found significantly associated with CD (GCCCGTCA, \(\chi^2 = 4.038, P < 0.045\)) when both cohorts were combined.

Oxidant and antioxidant status

To assess the oxidative status of CD patients in comparison with controls and according to their rs188461 genotype, plasma malondialdehyde (MDA) was measured. Results show that MDA levels were significantly elevated in CD subjects compared with controls \((P < 0.0001)\) (Figure 3a), but no significant difference was noted when MDA levels were separated according to rs188461 genotype (Figure 3b).

Subjects’ antioxidant profile was assessed by measuring plasma retinol, \(\beta\)-carotene, \(\gamma\)-tocopherol and \(\alpha\)-tocopherol. Compared with controls, the plasma concentrations of \(\beta\)-carotene were reduced in CD \((P < 0.0001)\) (Figure 4a), whereas retinol (Figure 4b) and \(\gamma\)-tocopherol (Figure 4c) levels were elevated \((P < 0.001\) and \(P < 0.001\), respectively). No significant difference was observed in \(\alpha\)-tocopherol levels (Figure 4d). Figure 5 shows the differences in vitamin levels according to the rs188461 genotype in CD subjects. A tendency of lower levels of retinol, \(\gamma\)-tocopherol and \(\alpha\)-tocopherol was observed in the homozygote carriers of the rare allele (G), but the differences did not reach statistical significance. Importantly, a large inter-individual disparity was observed in these experiments.

**DISCUSSION**

This is the first study reporting an association between genetic variants in the HNF4A gene and risk for CD. In a discovery study, we found that six HNF4A SNPs were significantly associated with CD. In a replication study performed on distinct cohorts of CD subjects and controls, one SNP (rs1884613) remained significantly associated with CD. Combining both cohorts, the single SNP analysis demonstrated significant associations for three of the six SNPs (rs1884613, rs1884614 and rs2144908), due to the gain in power. The associations for rs1884613 and rs1884614 remained significant after correcting for multiple testing. Moreover, haplotype analyses underlined the association between CD and a 8-marker haplotype containing the SNPs found to be associated in the single SNP analysis.

### Table 3. Distribution of allele frequencies for controls and Crohn’s disease subjects in discovery cohort

| SNP       | Controls MAF | CD MAF | Odds ratio | 95% CI          | P-value |
|-----------|--------------|--------|------------|-----------------|---------|
| rs4810424 | 0.152        | 0.202  | 1.407      | 1.099–1.800     | 0.007*  |
| rs1884613 | 0.148        | 0.200  | 1.432      | 1.118–1.835     | 0.004*  |
| rs1884614 | 0.149        | 0.201  | 1.429      | 1.116–1.829     | 0.005*  |
| rs6031543 | 0.147        | 0.148  | 1.000      | 0.758–1.329     | 0.977   |
| rs2144908 | 0.148        | 0.204  | 1.467      | 1.143–1.882     | 0.003*  |
| rs6031550 | 0.229        | 0.204  | 0.865      | 0.690–1.084     | 0.209   |
| rs6031551 | 0.229        | 0.208  | 0.882      | 0.701–1.109     | 0.282   |
| rs6031552 | 0.225        | 0.193  | 0.823      | 0.650–1.042     | 0.106   |
| rs6130716 | 0.336        | 0.320  | 0.929      | 0.756–1.140     | 0.479   |
| rs6031558 | 0.324        | 0.303  | 0.910      | 0.741–1.119     | 0.371   |
| rs6130608 | 0.264        | 0.256  | 0.961      | 0.771–1.197     | 0.721   |
| rs2425637 | 0.464        | 0.469  | 1.019      | 0.843–1.232     | 0.845   |
| rs2425639 | 0.471        | 0.461  | 0.957      | 0.789–1.161     | 0.659   |
| rs3212172 | 0.141        | 0.178  | 1.300      | 1.007–1.680     | 0.044*  |
| rs1800963 | 0.392        | 0.441  | 1.218      | 1.002–1.480     | 0.048*  |
| rs2071197 | 0.093        | 0.097  | 1.042      | 0.757–1.434     | 0.802   |
| rs736824  | 0.389        | 0.395  | 1.027      | 0.841–1.255     | 0.791   |
| rs745975 | 0.231        | 0.239  | 1.045      | 0.829–1.317     | 0.710   |
| rs3212183 | 0.474        | 0.461  | 0.951      | 0.788–1.148     | 0.600   |
| rs1885088 | 0.229        | 0.234  | 1.025      | 0.817–1.286     | 0.828   |
| rs1800961 | 0.026        | 0.027  | 1.043      | 0.573–1.898     | 0.890   |
| rs3212195 | 0.207        | 0.237  | 1.184      | 0.924–1.517     | 0.181   |
| rs3212198 | 0.408        | 0.396  | 0.950      | 0.776–1.162     | 0.617   |
| rs3818247 | 0.349        | 0.354  | 1.019      | 0.834–1.245     | 0.851   |

Abbreviations: CD, Crohn’s disease; CI, confidence interval; MAF, minor allele frequency; SNP, single nucleotide polymorphism. Distribution of the genotyped SNPs minor allele frequencies. Allelic association for individual SNPs was carried out using logistic regression by fitting an additive model. *P < 0.05.

### Table 4. Distribution of allele frequencies for controls and Crohn’s disease subjects in replication and combined cohorts

| SNP       | Controls MAF | CD MAF | Odds ratio | 95% CI          | P-value |
|-----------|--------------|--------|------------|-----------------|---------|
| rs4810424 | 0.156        | 0.145  | 0.912      | 0.719–1.156     | 0.465   |
| rs1884613 | 0.139        | 0.175  | 1.327      | 1.060–1.635     | 0.012*  |
| rs1884614 | 0.152        | 0.175  | 1.196      | 0.957–1.472     | 0.121   |
| rs6031543 | 0.154        | 0.143  | 0.918      | 0.733–1.150     | 0.453   |
| rs2144908 | 0.147        | 0.156  | 1.071      | 0.854–1.344     | 0.551   |
| rs6031550 | 0.236        | 0.213  | 0.880      | 0.725–1.067     | 0.191*  |
| rs6031551 | 0.236        | 0.222  | 0.922      | 0.759–1.121     | 0.409   |
| rs6031552 | 0.237        | 0.214  | 0.881      | 0.728–1.067     | 0.188   |
| rs3212172 | 0.153        | 0.142  | 0.909      | 0.721–1.146     | 0.425   |
| rs1800963 | 0.410        | 0.399  | 0.956      | 0.809–1.129     | 0.596   |

Abbreviations: CD, Crohn’s disease; CI, confidence interval; MAF, minor allele frequency; SNP, single nucleotide polymorphism. Distribution of the genotyped SNPs minor allele frequencies. Allelic association for individual SNPs was carried out using logistic regression by fitting an additive model. *P < 0.05.
**The P-values adjusted for rs1884613 and rs1884614 retained significance after correcting for multiple comparisons (40 tests).
Iron-ascorbate-generating oxidative stress. In line with our cellular antioxidant defences and increased cellular vulnerability to gene and protein expression amplified lipid peroxidation, reduced intestinal epithelial cell model and found that reduced HNF4α rs6017342, which maps 5 kb distal to the 3'0 in a whole genome study. This association was seen at inflammatory cytokines. Darsigny et al. reported that loss of HNF4α affects colonic ion transport and causes chronic inflammation resembling IBD in a knockout mouse model. Finally, a crosstalk between HNF4α and NF-κB was reported, supporting its role in inflammation.

We believe our findings are of high interest in view of the association between the HNF4α region and the risk of UC revealed in a whole genome study. This association was seen at rs6017342, which maps 5 kb distal to the 3' untranslated region of the HNF4α gene, within a recombination hot spot. However, rs6017342 was not in high LD with the identified variant associated with CD in our study (rs1884613). In fact, none of the SNPs associated with CD in the discovery study were in strong LD with rs6017342, which can be explained by the fact that rs60317342 is located within a recombination hot spot. In addition, in the GWAS United Kingdom (UK) cohort, the rs60317342 locus did not show any association to CD, suggesting that different signals on the HNF4α gene are associated with different types of IBD. Hence, it is possible that the associations are independent and it is also probable that they may even be linked to different genes within the 12q12–13 region. Cryptic differences in the genetic structure of the French Canadian ‘founder’ population, compared with the UK population used in the GWAS, could also explain the different associations in the HNF4α locus. Moreover, it has been put forward that some genes/loci may be specific to early onset CD patients and that new variants in many genes could have been missed by GWAS in this specific population.

Under the control of its two promoters, the HNF4α gene encodes a total of nine isoforms with various 3' truncations. The liver-specific P1 promoter drives the expression of transcripts HNF4α1–6, which include exons 1A and 2–10 (HNF4α1–3) or exons 1A, 1B and 2–10 (HNF4α4–6). Transcripts HNF4α7–9 are expressed from the pancreatic P2 promoter located ~46 kb upstream of the HNF4α transcription start and exhibit splicing of the upstream exon 1D to exon 2, without the inclusion of sequences from either exon 1A or 1B. The observed genetic variations in our study suggest a contribution of the P2 promoter in HNF4α implication in regulating inflammatory processes.

In our study, the P2 promoter variant rs1884613 was the only one that was replicated in a second independent cohort of cases and controls. This P2 promoter genetic variant has been associated with type 2 diabetes mellitus (T2DM) in several studies, pointing out to HNF4α's potential role in inflammation. In fact, rs1884613 was found to be associated with T2DM in Ashkenazi, Mexican American, and Scandinavian populations. Moreover, a link between rs1884613 and insulin resistance was noted. However, the association with T2DM was not replicated in UK and a Finnish population, nor in a broader meta-analysis with additional populations.

The identification of HNF4α, which has been associated with MODY1 and T2DM, as a CD-susceptibility gene is in line with the recent concept of shared genetic determinants for clinically
distinct disorders.51 GWAS have identified several genes conferring susceptibility to multiple conditions, such as CD, ankylosing spondylitis, rheumatoid arthritis, systemic lupus erythematosus and type 1 diabetes.52 It has been suggested that there may be a general set of susceptibility genes for autoimmune, which are modulated by disease-specific genes, as well as the host’s human leukocyte antigen status. A specific combination of polymorphisms, combined with environmental factors, could determine the type of disease developed by a subject.53

To predict the effect of the P2 promoter SNP rs1884613, we investigated the impact on putative transcription factor-binding sites. Our in silico analyses show that variations in that SNP could theoretically modify the binding of the ras-responsive element binding protein 1 (RREB1), a transcription factor involved in DNA repair by modulating p53 transcription54 and associated with immune tolerance.55 Thus, studying the impact of rs1884613 and other P2 promoter SNPs on HNF4α gene expression and function might help understand the role of this gene in inflammation and IBD.

During liver development, HNF4α regulates the expression of cell adhesion proteins.56 It also provokes the expression of tight-junction adhesion molecules and the modulation of subcellular distribution of junction and cell polarity proteins, resulting in junction formation and epithelial polarization in embryonic carcinoma cells.57 Moreover, using an adult mouse model lacking HNF4α in the intestinal epithelium, HNF4α was shown to have a pivotal role in the homeostasis of the intestinal epithelium, in the epithelial cell architecture, and in intestinal barrier function.58 These results underline the potential role of HNF4α in epithelial integrity in IBD physiopathology.

In an attempt to explore the mechanisms behind the rs1884613-(G/G) haplotype, we measured oxidative stress biological markers in controls and CD subjects. CD patients displayed higher oxidative stress status, as documented by the elevated MDA levels and the reduced β-carotene. Yet, the average plasma γ-tocopherol was increased in subjects with CD; such elevation in CD was previously described in the literature.59 Although no significant difference was observed in MDA and vitamin levels in the case of rs1884613 genotype, an apparent trend was noted for the levels of retinol, γ-tocopherol and α-tocopherol when compared with CC and CG genotypes. Discriminatig patients according to C-reactive protein levels or disease activity could not contribute to explain the differences in antioxidant levels (data not shown). Given the limited number of patients with the rare genotype available in our study, larger cohorts are needed to focus on this aspect.

In conclusion, our results suggest that the HNF4A locus may be a common genetic determinant of CD, but its relative contribution may differ between populations. Further replication of these data in international IBD cohorts is necessary to estimate the effect of the HNF4α polymorphisms on risks for CD and UC. Functional studies are also necessary to investigate the impact of the aforementioned genetic variants on HNF4α protein functions.

### MATERIALS AND METHODS

**Subjects**

For the discovery cohort, patients were recruited from the IBD clinics of tertiary pediatric and adult hospitals in Montreal (CHU-Sainte-Justine, Montreal General, Royal Victoria and Montreal Children’s Hospitals) between 30 June 2008 and 30 January 2010. For the replication cohort, patients were those diagnosed and followed at the pediatric gastroenterology clinics of three hospitals across Canada: CHU-Sainte-Justine, Montreal; the British Columbia’s Children’s Hospital, Vancouver; and the Children’s Hospital of Eastern Ontario, Ottawa. These patients were recruited from 1 January 2003 to 30 June 2011. The diagnosis of CD was confirmed based on standard clinical, endoscopic, radiological and histopathological criteria.60,61 Clinical and demographic information acquired included age at diagnosis, gender and ethnicity. Disease location and clinical phenotype were classified according to World Gastroenterology Organization’s Montreal classification (L1, ileum; L2, colon; L3, illeocolon; L4, upper GI tract; B1, non-stricturing and non-penetrating; B2, structuring; B3, penetrating; p, perianal modifier).62 The designation of French Canadian, Jewish or other ethnicity was based on self-reported. Self-identified race/ethnicity has previously been shown to highly correlate with genetic cluster categories.63 For all patients, blood or saliva was collected for DNA analysis. Controls were chosen from the 1999 Quebec Child and Adolescent Health and Social Survey, a school-based survey of youth aged 9, 13 and 16 years providing DNA samples.64 The institutional Ethics Review Boards of all centers approved the study and informed consent was acquired from all participating subjects.

| Table 5. Distribution of haplotype frequencies for controls and Crohn’s disease subjects in discovery and combined cohorts |
|---------------------------------------------------------------|
| **Haplotype**         | **Frequency** | **Frequency** | **P-value** |
|-----------------------|---------------|---------------|-------------|
| **Discovery cohort**  |               |               |             |
| Block 1               |               |               |             |
| GCCCGCTC              | 0.616         | 0.585         | 0.193       |
| GTTCCTC               | 0.143         | 0.195         | 0.004*      |
| GCCGGTCA              | 0.145         | 0.134         | 0.532       |
| GCCGGTCA              | 0.075         | 0.062         | 0.265       |
| Block 2               |               |               |             |
| AG                    | 0.373         | 0.356         | 0.469       |
| CG                    | 0.314         | 0.343         | 0.193       |
| AC                    | 0.313         | 0.300         | 0.572       |
| Block 3               |               |               |             |
| TTG                   | 0.522         | 0.531         | 0.727       |
| CGA                   | 0.257         | 0.249         | 0.700       |
| TGA                   | 0.204         | 0.211         | 0.732       |
| Block 4               |               |               |             |
| GT                    | 0.606         | 0.603         | 0.897       |
| GC                    | 0.300         | 0.299         | 0.972       |
| AC                    | 0.091         | 0.097         | 0.681       |
| Block 5               |               |               |             |
| CC                    | 0.471         | 0.457         | 0.568       |
| CT                    | 0.304         | 0.306         | 0.909       |
| TT                    | 0.224         | 0.233         | 0.645       |
| Block 6               |               |               |             |
| GCC                    | 0.405         | 0.394         | 0.640       |
| GGT                   | 0.362         | 0.366         | 0.859       |
| AAT                   | 0.224         | 0.230         | 0.760       |
| **Replication cohort**|               |               |             |
| Block 1               |               |               |             |
| GCCCGCTC              | 0.593         | 0.608         | 0.453       |
| GTTCCTC               | 0.130         | 0.171         | 0.004*      |
| GCCGGTCA              | 0.140         | 0.143         | 0.869       |
| GCCGGTCA              | 0.089         | 0.071         | 0.128       |
| Combined cohorts      |               |               |             |
| Block 1               |               |               |             |
| GCCCGCTC              | 0.600         | 0.595         | 0.755       |
| GTTCCTC               | 0.136         | 0.185         | 7.755 × 10^-6** |
| GCCGGTCA              | 0.142         | 0.139         | 0.775       |
| GCCGGTCA              | 0.083         | 0.067         | 0.045*      |

Abbreviation: CD, Crohn’s disease. Haplotype analyses were performed on the SNPs within each block of high linkage disequilibrium. Haplotype analysis was carried out using HAPLOVIEW Software, version 3.11, with haplotype blocks created using the confidence interval feature. The association of specific haplotypes within blocks with the outcome was examined and P-values were estimated. *P < 0.05; **P < 2.09 × 10^-4 after correcting for 27 haplotype comparisons.
DNA extraction
Genomic DNA was prepared from white blood cells, total blood or saliva with the Puregene DNA Isolation kit (Gentra Systems, Qiagen Inc., Toronto, ON, Canada) using methods described by the manufacturer.

DNA variants detection by direct sequencing
To identify SNPs present in our population, we first sequenced the HNF4A gene in a total of 40 French Canadian patients diagnosed with childhood-onset IBD (20 CD and 20 UC patients). The sequencing targeted the coding regions, the P1 promoter region (2.5 kb upstream exon 1a) and other regions containing SNPs previously associated with the risk of developing diseases, such as T2D and dyslipidemia.26,36,37,50,64,65 In total, 30 fragments were sequenced. Genomic DNA (2 ng) was amplified in a total volume of 50 μl volume using 5 μl PCR buffer (10 mM Tris), 1.5 μl MgCl₂ (50 mM), 2 μl dNTPs (2.5 mM), 0.4 μM of each corresponding primer (25 μM) and 1.0 units of Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA). The PCR amplifications were performed using a GeneAmp PCR System 9700 (Applied Biosystems, Carlsbad, CA, USA) under the following profile: 35 cycles of amplification were used at 95 °C for 30 s, 58 °C for 45 s and 72 °C for 45 s. Amplicons were verified on standard ethidium bromide stained 1.5% agarose gel. The specific primers for each fragment and the amplicon size are available upon request. Amplified fragments were sent to the McGill University Genome Quebec Innovation Center in Montreal for sequencing using Applied Biosystem’s 3730xl DNA Analyzer technology. Complete sequences were aligned, assembled and compared using the MultiAlign software.66 Visual inspection of chromatograms was used for identification of each candidate SNP.

Genotyping
Discovery cohort. Based on sequencing results, identified SNPs were genotyped using the Luminex MAP/Autoplex Analyser CS1000 system (Perkin Elmer, Waltham, MA, USA). The 27 selected SNPs were amplified in a single multiplex assay and hybridized to Luminex MicroPlex-xTAG Microspheres67 for genotyping using allele-specific primer extension. Amplification and reaction conditions are available upon request. Allele calls were assessed and compiled using the Automatic Luminex Genotyping software.68 For quality control purposes, genotyping of a systematic random sample of 20% of the specimens was repeated.
Replication cohort. Replication genotyping was performed on the SNPs significantly associated with CD in the discovery study (in the single SNP and haplotype analyses). In total, 10 SNPs were genotyped using Sequenom-based primer-extension methods. These methods are designed for high-throughput SNP genotyping. The platform has a high assay conversion rate (85%), high genotyping success rate (95%) and minimal error rates (0.5–1%). Genotyping was carried out at the McGill University and Genome Quebec Innovation Center in Montreal.

Biological studies

Blood samples. In order to examine the levels of plasma MDA and antioxidant vitamins, blood samples were collected in tubes containing EDTA 1 g–1. Plasma was separated immediately by centrifugation (700 g for 20 min at 4 °C). CD patients were characterized according to their rs1884613 genotype.

Malondialdehyde. The amount of free MDA in plasma was determined by HPLC in 48 CD patients and 213 healthy controls using an improved method previously described by our unit.69

Antioxidant vitamins. The antioxidant profile was determined by measuring antioxidant vitamin levels (β-carotene, retinol, γ-tocopherol and α-tocopherol) in 45 CD patients and 112 healthy controls using an improved method previously described by our unit.70

In silico analysis

To explore the potential interaction between transcription factors and the HNF4α P2 promoter polymorphism rs1884613, we performed in silico analyses using the Genomatix MatInspector program (Genomatix Software GmbH, Munich, Germany) with a standard (0.75) core similarity. Transcription factor recognition site sequences were identified in the HNF4A gene region containing the SNP.

Statistical analysis

Potential genotyping errors were assessed using χ²-tests, which evaluate the deviation of each SNP from Hardy–Weinberg equilibrium. Allelic association for individual SNPs was carried out using logistic regression by fitting an additive model. Genotype and allele frequencies were compared between cases and controls using χ²-tests and Fisher’s exact tests where appropriate. OR and 95% confidence intervals were estimated. In addition to single SNP analysis, haplotype analysis was carried out. LD blocks were defined using the ‘single gamete rule’ implemented in the HAPLOVIEW Software, version 3.11.71 The association of specific haplotypes within blocks with the outcome was examined and P-values were estimated. For the biological studies, statistical differences were assessed by Anova and Student’s two-tailed t-test. P-values < 0.05 after correction for multiple hypotheses were considered significant in the genetic analysis based on the combined cohorts. Adjusting for multiple comparisons was made using Bonferroni methods separately for the single SNP and haplotype analysis for the combined analysis. For the single SNP analysis, we tested 24 SNPs in the discovery cohort, 10 in the replication cohort and 6 in the combined cohort, we therefore accounted for 40 comparisons. As for the haplotype analysis, we tested 19 haplotypes in the discovery cohort, 4 in the replication cohort and 4 in the combined cohort, thus we accounted for 27 comparisons.

Power estimations

Based on findings of the discovery cohort, the power required to replicate associations in an independent cohort was made after considering the observed allele frequencies and ORs, assuming an alpha level of significance of 0.05, an available case sample of ~450 cases and a control population of ~1300 subjects. Based on this pre-defined sample size, it was estimated that the replication cohort would have >80% power to replicate associations noted in the discovery cohort. Power analysis was carried out using QUANTO Software, version 1.2.4 (http://hydra.usc.edu/gxe).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by the Canadian Institutes of Health Research Team Grant (CTP-82942; ES, F8, FPG, JFB, DM and EL), the J. A. DeSe `ve Research Chair in Nutrition (EL), the Canada Research Chair in Immune Mediated Gastrointestinal Disorders (ES), the Canadian Institutes of Health Research Fellowship Award and The Richard and Edith Strauss Postdoctoral Fellowships Award in Medicine, McGill University (VM). A preliminary version of this study, presented by VM, was awarded with a Presidential Poster Award at the American College of Gastroenterology 2010 Annual Scientific Meeting. We thank Mrs Schohraya Spahis for her technical assistance.
56 Battle MA, Konopka G, Parviz F, Gaggl AL, Yang C, Sladek FM et al. Studies in 3523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 alpha (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. Diabetes 2007; 56: 3112–3117.

57 Satohisa S, Chiba H, Osanai M, Ohno S, Kojima T, Saito T et al. Hepatocyte nuclear factor 4 alpha P2 promoter variants associate with insulin resistance. Acta Biochim Pol 2011; 58: 179–186.

58 Cattin AL, Le BJ, Barreau F, Saint-Just S, Houllier A, Gonzalez FJ et al. Association testing of variants in the hepatocyte nuclear factor 4alpha gene with risk of type 2 diabetes in 7883 people. Diabetes 2005; 54: 886–892.

59 Genser D, Kang MH, Vogelsang H, Elmadfa I. Status of lipidsoluble antioxidants and TRAP in patients with Crohn’s disease and healthy controls. Eur J Clin Nutr 1999; 53: 675–679.

60 Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989; 170: 2–6.

61 Sands BE. From symptom to diagnosis: clinical distinctions among various forms of intestinal inflammation. Gastroenterology 2004; 126: 1518–1532.

62 Tang H, Quertermous T, Rodriguez B, Kardia SL, Zhu X, Brown A et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. Am J Hum Genet 2005; 76: 268–275.

63 Paradis G, Lambert M, O’Loughlin J, Lavallee C, Aubin J, Berthiaume P et al. The Quebec Child and Adolescent Health and Social Survey: design and methods of a cardiovascular risk factor survey for youth. Can J Cardiol 2003; 19: 523–531.

64 Andrulionyte L, Laukkonen Q, Chaisson JL, Laakso M. Single nucleotide polymorphisms of the HNF4alpha gene are associated with the conversion to type 2 diabetes mellitus: the STOP-NIDDM trial. J Mol Med 2006; 84: 701–708.

65 Muller YL, Infante AM, Hanson RL, Love-Gregory L, Knowler W, Bogardus C et al. Variants in hepatocyte nuclear factor 4alpha are modestly associated with type 2 diabetes in Pima Indians. Diabetes 2005; 54: 3035–3039.

66 Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 1988; 16: 10881–10890.

67 Koo SH, Ong TC, Chong KT, Lee CG, Chew FT, Lee EJ. Multiplexed genotyping of ABC transporter polymorphisms with the Bioplex suspension array. Biol Proced Online 2007; 9: 27–42.

68 Bourgey M, Latiere M, Richer C, Sinnett DALG. Automated genotype calling of luminex assays. PLoS One 2011; 6: e19368.

69 Courtois F, Suc I, Garofalo C, Ledoux M, Seidman E, Levy E. Iron-ascorbate alters the efficiency of Caco-2 cells to assemble and secrete lipoproteins. Am J Physiol Gastrointest Liver Physiol 2000; 279: G12–G19.

70 Levy E, Rizwan Y, Thibault L, Lepage G, Brunet S, Bouthillier L et al. Altered lipid profile, lipoprotein composition, and oxidant and antioxidant status in pediatric Crohn disease. Am J Clin Nutr 2000; 71: 807–815.

71 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263–265.