Reviewer's report

Title: Age-of-onset-dependent Influence of NOD2/CARD15 gene variants on Disease behaviour and treatment in Crohn's disease

Version: 1 Date: 8 November 2012

Reviewer: Francis Vasseur

Reviewer's report:


#1 Gene names should be typed in italic.

#2 In the results part of the abstract (page 3) it is mentioned that "Chronic active and moderate to severe course of disease is associated with younger age of onset (p=0.457)". Is it the true pvalue or is there a typo error?

#3 In the results part of the abstract (page 3) it is mentioned that "Interestingly, osteoporosis found more frequently in patients carrying NOD2 SNPs than in the wild type group". I should have written "...was found more frequently..."

#4 In the abstract the authors first use the term "SNP" to define the genetic variants and in the abstract conclusion use the term "polymorphism", then "mutation" later in the manuscript. It should be more convenient to use the same term throughout the whole manuscript.

#5 Page 4 it is mentioned "Polymorphisms in the leucine-rich repeats (LRR) region of the NOD2/CARD15 gene were identified as an independent risk factor". If "polymorphisms" is at the plural, the sentence should be "...as independent risk factors..."

#6 Page 4, regarding the severe demineralization and its evaluation using zscores, the authors should precise what parameter is (are) zscored.

#7 On page 5 the authors wrote "There is increasing evidence that inflammation may significantly contribute to a high-turnover mineral bone loss". This assumption should be supported by relevant references.

#8 On page 5 the authors draw a link between disease behavior and NOD2 genotypes, "as NF-B activation induces TNF, IL-6 and IL-1". The relationship of cause and effect is far from being evident and would remain to be supported by further explanations and literature.

#9 On page 5 at the end of the introduction section as the preterit has been used, the term "hypothesized" should be used.

#10 As there are 85 pediatric and 116 adult-onset CD patients, the term "large
dataset” is perhaps excessive.

#11 Page 5 in the material and methods section which criteria have been used, the Montreal or the Vienna classification? Moreover it is not clear if phenotypic parameters were those at diagnosis or those at follow-up or both? And if follow-up was performed what was follow-up duration. Did the authors collected data regarding anoperineal lesions that may be regarded as severe phenotype?

#11 the authors defined "underweight" using BMI. It would be more logical to define underweight using weight via the weight zscore determined using the Cole's method. BMI defines "corpulence" and not weight.

#12 What is the relevance of disease activity indexes that reflect acute phenomena rather than the chronic activity of the disease. In the background of such a study integration of the time activity (i.e. area under the curve) would better reflect what the authors try to link with NOD2 genotypes.

#12 Page 6 the authors wrote " a hospitalization of more than two weeks per year was also given to describe the percentage of severe ill patients". As they used the term "also", what were the criteria also used to define severe disease? Were accounted only hospitalizations related with CD?

#13 If the notion of zscore is well known, for a better readability of the paper, the authors should offer at least limited explanations regarding the "T-score".

#14 Page 7, the term "snp" should be in upper case.

#15 In the statistical analysis section, what is the chi quadrate test? As may be found in the german version of Wikipedia, it seems that it is indeed the chi2 test. If this is true, the authors should use the chi2 test.

#16 In the statistical analysis section, why the authors use a parametric test (t-test) to compare two groups and a non parametric one (Wilcoxon test) for three groups?

#17 In the statistical analysis section, fore descriptive statistics the authors used mean and SD. If parameters under study were not fitting a normal distribution they better use median and interquartile range for descriptive statistics.

#18 In the statistical analysis section, what do the authors mean with " the descriptive impact of our data-set". What was the definition of "chronic active", of "high active course of the disease".

#19 On page 8, results section, "retrospective" should be used instead of "retro-perspective".

#20 On page 8, results section, why "'s" is appended to SNP? The authors wrote that " NOD2 SNP's were found". As SNP defines a polymorphism and does not imply the wild type rather than the variant allele, the authors should write "SNP variant alleles were found" or "SNP minor alleles were found".

#21 In the supplemental table they report alleles counts in pediatric and adult onset patients. There is a great confusion in presenting this table as it is difficult to understand in the No of variant alleles column, pediatric onset sub column, heterozygous sub column, what is the meaning of 31 heterozygous alleles (or patients?) in a column that reports the number of alleles. Briefly an allele is
neither heterozygous. The same confusion is noticed for the adult onset part of the table and the homozygous and compound (compound heterozygote should be better) parts of the table. Moreover including % of patients in this table reporting the number of alleles add more confusion to this confused table.

#22 In the results section why the authors precised that only 14 patients were older than 40 at time of diagnosis as this sentence does not seem to be related neither with what is written before nor what is written after.

#23 The common allele suggesting the wild type one what do the authors mean with the "common NOD2 allele variants"?

#24 Did the authors compared allelic frequencies between the two groups of patients?

#25 The authors wrote that "7.1% of the pediatric-onset patients are homozygous in contrast to only 2.5% and 9.4% compound heterozygous versus 5.9% in the adult-onset group", did they objectified the reality of these differences with a statistical test?

#26 the authors wrote " There was no statistical difference regarding age at diagnosis, chronological age, average disease duration or gender" and refered to table 1. Even when going to table 1 the reader at last understand that there is no difference according to genotype. This should be mentioned in the text.

#27 In table 1 how the authors explain there are11 NOD2 variant allele carriers and 14 WT NOD2 genotype patients aged between 17 and 40 years in the pediatric onset group? In this table it is not correct to define the NOD2 genotype by the term "SNP".

#28 In table 1 are localizations and behaviors at diagnosis or at maximal follow up? Likewise for the EIMs mentioned on page 9.

#29 In the "Disease Localization and Behavior of NOD2 SNP Carriers in the Pediatric- and Adult-onset CD Cohort" section page 8, the sentence "NOD2 SNPs were associated...." is incorrect and should be replaced by "patients with at least one NOD2 variant allele...". In the same sentence what is the significance level of the 68% vs 48%?

#30 Likewise the term "NOD2 positive patients" is unclear, positive for what? Using patients with at least one NOD2 variant allele should be more correct. In the same way two and three lines lines below are used the incorrect sentence "NOD2 SNP pediatric-onset group". Similar abnormal denominations occur throughout the manuscript and should be corrected accordingly.

#31 On page 8 the authors wrote " Extraintestinal manifestation, need for surgery, complications and osteoporosis are more frequent in pediatric-onset than in adult-onset CD patients with NOD2 variants but did not reach significance". What must be understood, a suggestive trend of association (p=0.08?) suggesting a putative but undetected (lack of power?) difference, or a true unsignificant difference (p=0.4). This is quite difference in terms of discussion and hypotheses that may be drawn. The affirmation by the authors that these are "more frequent" is an over interpretation of the data.
#32 in figure 2 why the authors use "growth delay" on the figure "growth failure" in the legend? This is confusing. Where are the bars for growth failure at diagnosis in figure 2?

#33 On page 10, what is a "wt carrier"? Referring to what is assumed to be a variant allele carrier: a patient having at least one variant allele, a "wt carrier" should be a patient with at least one wt allele. Thus heterozygous genotypes should be included in this term.

#34 The authors should add on fig 3 that PCDAI scores refers to pediatric onset patients. Likewise for CDAI scores for adult ones.

#35 Referring to fig 1A the authors claim that BMD zscore is lower in the pediatric-onset group of patients with NOD2 variant allele. There is no mention of this BMD parameter in fig 1A and no statistical result to support this affirmation. Only osteoporosis is presented in fig 1A and the difference is unsignificant. Thus this affirmation is an over interpretation of the data. Although the text refers to fig 1A about osteopenia, where is osteopenia in fig1A.

#36 On page 10, where do the new 78 pediatric onset CD patients come from? And what are the significance levels according to their BMD zscores, osteoporosis.... parameters?

#37 In table 2 it is necessary to precise what is the "Z<-1". We hypothesize it is BMD zscore (?) as compared with the neighbour column. However the authors should demonstrate a minimum of rigor in their presentation of the data.

#38 On page 10 the authors wrote: "...underweight at diagnosis and during follow up...", what do they mean with "during follow up"? Either they report data and patients phenotypes recorded at maximal follow up or if it is during follow up they should precise after which time they recorded the data.

#39 In figure 5 the authors have pooled the two groups of patients (pediatric and adult onset patients) in order to compare the age at onset according to the genotype at the NOD2 locus. As presented and handled throughout the manuscript these two groups of patients are quite separated and were investigated separately. These two groups are significantly different regarding a lot of phenotypes and regarding allele frequencies of NOD2 gene variants. Thus they should not be pooled in an analysis. It is very likely that according to the recruitment of these two groups, the age distribution of the pool displays a bimodal distribution that reflects the two groups and thus would demonstrate that (1) pooling is not possible and (2) the non normal distribution of the parameter makes the parametric t-test misfits to compare means. The results presented in figure 5 panel A may only reflect an higher prevalence of NOD2 gene variants in pediatric than in adult onset CD patients. Thus all comparisons reported on figure 5 that rely on age at onset (for the pooled patients) may not be considered.

#40 In the discussion section the authors report « high prevalence of NOD2/CARD15 SNPs with 44% in the pediatric-- and 42% in adult--onset CD ». Were these frequencies significantly different as later the authors wrote that it « …confirms our hypothesis that pediatric--onset Crohn disease has a greater degree of genetic influence ». Moreover do these percentages reflect patients or
alleles? It is likely that they describe percentages of patients but the text is ambiguous.

#41 The authors report in the discussion section that "...homozygosity and compound heterozygosity for NOD2 SNPs was found more frequent in the pediatric--onset CD group as reported by others". Did the authors tried to ascertain this affirmation using a comparison frequency statistical test?

#42 How the authors have taken into account the possible phenotypic bias that relies upon the well known fact that endoscopic explorations are often more complete and more frequent in pediatric rather than in adult onset CD patients. A similar bias may have occured in the frequency of CDAI measure frequency.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests' below