Article

A novel rare missense variation of the *NOD2* gene: evidences of implication in Crohn's disease

Sara FRADE-PROUD'HON-CLERC[1]*, Thomas SMOL[2,3], Frédéric FRENOIS[2], Olivier SAND[4], Emmanuel VAILLANT[4], Véronique DHENNIN[4],Amélie BONNEFOND[4,5], Philippe FROGUEL[4,5], Mathurin FUMERY[6], Nathalie GUILLON-DELLAC[7,8],Corinne GOWER-ROUSSEAU[7,8] and Francis VASSEUR[1]

[1] Univ. Lille, CHU Lille, EA2694 - Santé Publique : épidémiologie et qualité des soins, F-59000 Lille, France; sara.frade@chru-lille.fr, francis.vasseur@univ-lille2.fr
[2] Univ. Lille, EA 7364 – RADEME - Maladies RAres du DEveloppement embryonnaire et du MEtabolisme, F-59000 Lille, France; Frederic.FRENOIS@chru-lille.fr
[3] CHU Lille, Institut de Génétique Médicale, F-59000 Lille, France; Thomas.SMOL@chru-lille.fr
[4] CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France ; Amelie.bonnefond@cnrs.fr, philippe.froguel@cnrs.fr, veronique.dhennin@cnrs.fr, Emmanuel.Vaillant@cnrs.fr, Olivier.sand@cnrs.fr
[5] Department of Medicine, Section of Genomics of Common Disease, Imperial College London, London, United Kingdom
[6] Registre Epimad, Gastroenterology Unit, Amiens University Hospital, France; fumery.mathurin@chu-amiens.fr
[7] Registre Epimad, Service de Santé Publique, d'Epidémiologie, d'Economie de la Santé et de la Prévention, Maison Régionale de la Recherche Clinique, CHU Lille, France; corinne.gower@chru-lille.fr, nathalie.guillon@chru-lille.fr
[8] Inserm, UMR 995 – LIRIC, Université de Lille, France
* Correspondence: sara.frade@chru-lille.fr; Tel.: +33 3 20 44 41 45

Abstract: The *NOD2* gene, involved in innate immune responses to bacterial peptidoglycan, has been found to be strongly associated with Crohn's Disease, with an Odd Ratio ranging from 3 to 36. Families with 3 or more CD affected patients were related to high frequency of *NOD2* gene variations as R702W, G908R, 1007fs and were reported in EPIMAD Registry. However, some rare CD multiplex families were described without identification of common *NOD2* linked-to-disease variations. In order to identify new genetic variation(s) with a major effect on Crohn's disease (CD), whole exome sequencing was performed in available subjects comprising 4 patients on 2 generations affected with Crohn's disease without R702W and G908R variation, and 3 unaffected related subjects. A new rare and not yet reported missense variation of the *NOD2* gene, the N1010K, was detected and co-segregated across affected patients. *In silico* evaluation and modeling highlighted evidences for a deleterious effect of the N1010K variation regarding CD. Moreover cumulative characterization of N1010K and 1007fs as compound heterozygous state in two more severely CD family members strongly suggesting that the N1010K should be a new risk factor involved in Crohn's disease genetic susceptibility.

Keywords: Crohn's disease; *NOD2* gene; variation; WES.

1. Introduction

Crohn's disease is a chronic Inflammatory Bowel Disease (IBD) and results from the interaction of environmental factors, including the intestinal microbiota, with host immune mechanisms in genetically susceptible individuals. The *NOD2* gene, involved in innate immune responses to bacterial peptidoglycan, is strongly associated with Crohn's disease (CD) with Odd Ratio ranging from 3 to 36 and was initially identified through genetic linkage analyses [1]. Genome wide association studies (GWAS) have now reported more than 200 genetic susceptibility *loci*, only accounting for 20% of the contribution to the disease risk, suggesting that more *loci* remain to be discover. EPIMAD Registry covers a large area of Northern France (6 million inhabitants) and collects all incident CD cases. Data from multiplex families defined by 3 or more than 3 first degree relatives with CD have been recorded by EPIMAD. We previously reported 22 multiplex families and genotyping evidenced that most cases from these multiplex families were related to high frequency of *NOD2* R702W, G908R, L1007fs variations [2]. However rare CD multiplex families did not display high frequency of R702W, G908R, L1007fs variations. Thus in these families high prevalence of affected cases may rely on other major genetic susceptibility variant(s) that remained to be determined. In order to identify new genetic variations with major effect in CD, a Whole Exome Sequencing (WES) protocol has been initiated in a family with 4 CD cases among two generations. WES with intra-familial controls disclosed a new *NOD2* variation related with familial aggregation of the disease.

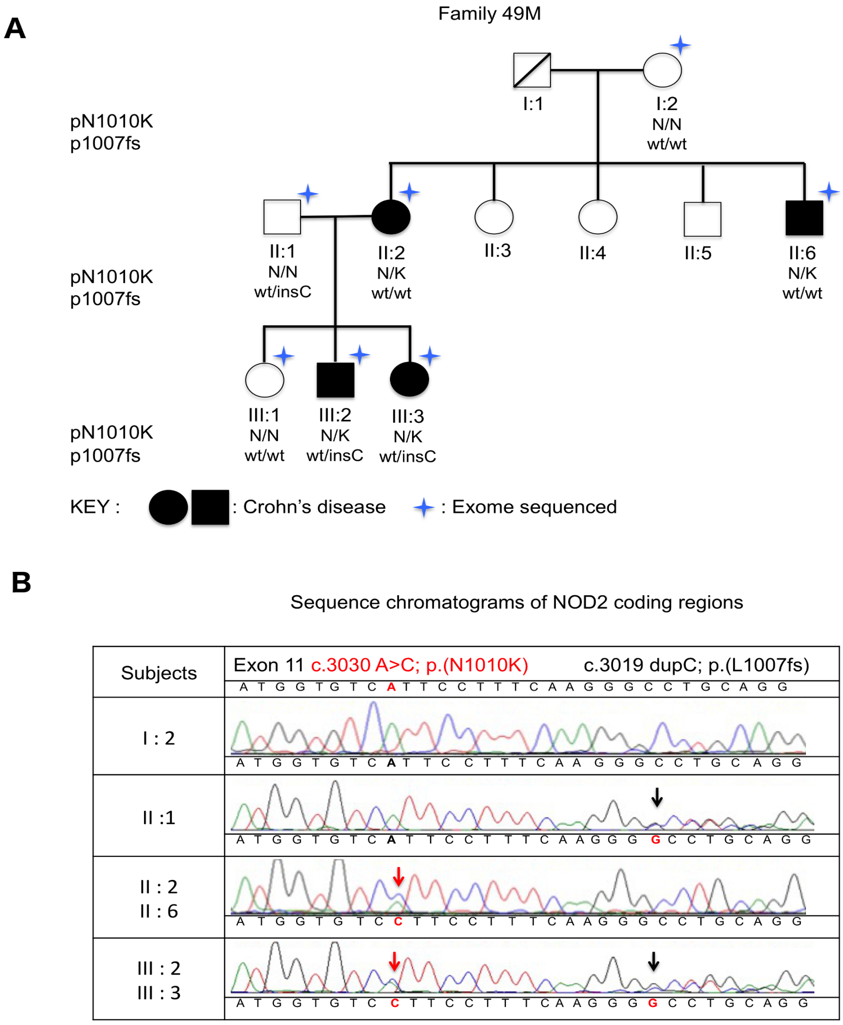

Figure 1. (**a**) Pedigree of family F49M with segregation of the c.3030A>C; p.(N1010K) and c.3019dupC; p.(L1007fs) variations.(**b**) Chromatograms for *NOD2* coding region in exon 11. The red arrows show the c.3030A>C nucleotide substitution consisting in the amino acid substitution N1010K. The black arrows show the c.3019dupC frameshift variation (rs2066847).

2. Results

To identify new coding variations related to CD, we performed WES on multiple members from a multiplex CD family. Seven individuals, including 4 patients affected with CD were successfully whole exome sequenced. Two variations within the *NOD2* gene, were identified. A novel heterozygous missense *NOD2* variation, the c.3030A>C; p.(N1010K), was identified in all affected members (figure 1A). This variation was localized in the last exon of the *NOD2* gene. The other significant variation was the well-known c.3019dupC; p.(L1007fs), rs2066847. The rs2066847 was detected in affected members III:2 and III:3 as compound heterozygous with c.3030A>C; p.(N1010K) and was present in unaffected father II:1 (figure 1A). Both variations were confirmed with Sanger sequencing (Figure 1B).

The new p.N1010K *NOD2* variation, as the p.L1007fs, was located in the Leucine-Rich Repeat (LRR) domain of NOD2 protein, which was already implicated in CD (Figure 2A)[3].

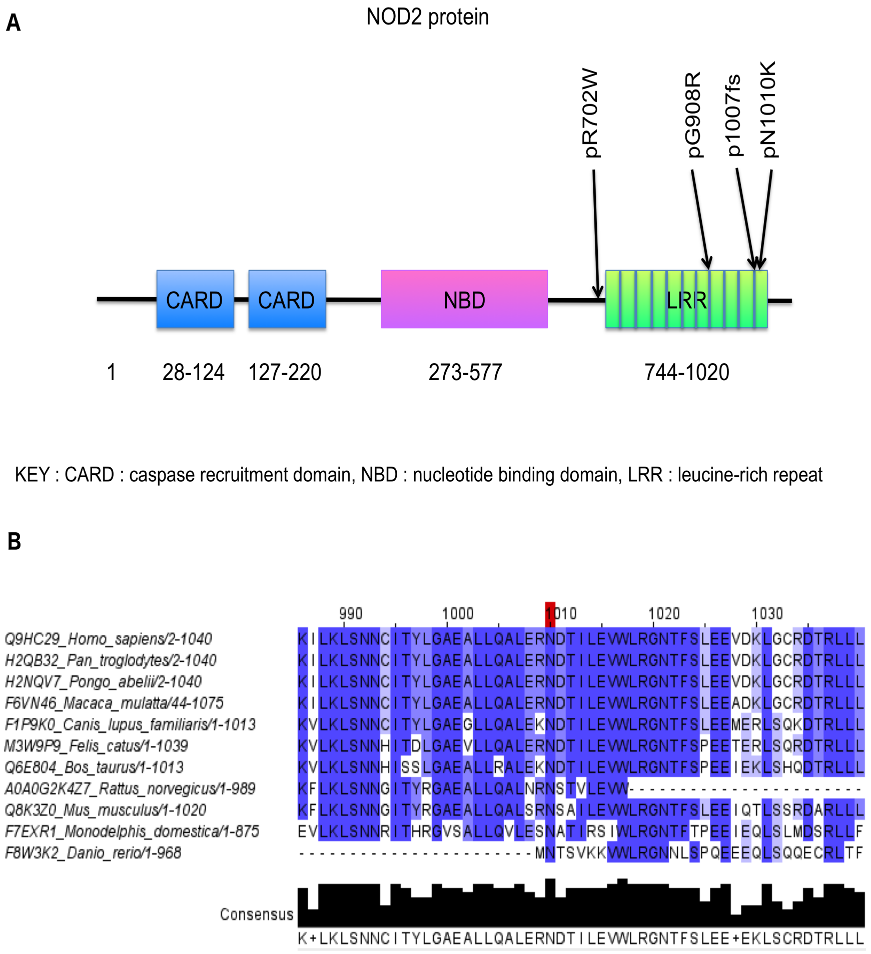
A
NOD2 protein
pR702W
pG908R
p1007fs
pN1010K
CARD
CARD
NBD
LRR
1
28-124
127-220
273-577
744-1020
KEY : CARD : caspase recruitment domain, NBD : nucleotide binding domain, LRR : leucine-rich repeat
B
990
1000
1010
1020
1030
Q9HC29_Homo_sapiens/2-1040 KILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
H2QB32_Pan_troglodytes/2-1040 KILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
H2NQV7_Pongo_abelii/2-1040 KILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
F6VN46_Macaca_mulatta/44-1075 KILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEADKLGCRDTRLLL
F1P9K0_Canis_lupus_familiaris/1-1013 KVLKLSNNCITYLGAEGLLQALEKNDTILEVWLRGNTFSLEEMERLSQKDTRLLL
M3W9P9_Felis_catus/1-1039 KVLKLSNNHITDLGAEVLLQALERNDTILEVWLRGNTFSPEETERLSQRDTRLLL
Q6E804_Bos_taurus/1-1013 KVLKLSNNHISSLGAEALLRALEKNDTILEVWLRGNTFSPEEIEKLSHQDTRLLL
A0A0G2K4Z7_Rattus_norvegicus/1-989 KFLKLSNNGITYRGAEALLQALNRNSTVLEVW-----------------------
Q8K3Z0_Mus_musculus/1-1020 KFLKLSNNGITYRGAEALLQALSRNSAILEVWLRGNTFSLEEIQTLSSRDARLLL
F7EXR1_Monodelphis_domestica/1-875 EVLKLSNNRITHRGVSALLQVLESNATIRSIWLRGNTFTPEEIEQLSLMDSRLLF
F8W3K2_Danio_rerio/1-968 -----------------------MNTSVKKVWLRGNNLSPQEEEQLSQQECRLTF
Consensus
K+LKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEE+EKLSCRDTRLLL

Figure 2. (**a**)Location of the p.R702W (rs2066844), p.G908R (rs2066845), p.L1007fs (rs2066847) and p.N1010K NOD2 protein-altering variations.(**b**)Multiple alignments for the amino acid sequence of the NOD2 proteins in 11 species in agreement with a conserved amino acid.

The heterozygous missense c.3030A>C; p.(N1010K) was absent from the public databases GnomAD, ExAC and Kaviar. Coverage metrics from WES samples in public databases were considered as correct: 99.79% of control samples presented a coverage > 20% for this region. This new *NOD2* variation appeared as a very rare genetic event. *In silico* predictions were strongly supportive of a deleterious effect for N1010K variation (Table 1).

Table 1. Comparison of N1010K variation and known variations associated with CD: R702W and G098R. Indicators of *in silico* prediction of the deleterious effects of *NOD2* variations. CADD Phred: global potential deleterious effect, SIFT: protein potential deleterious effect, PolyPhen2: protein domain potential deleterious effect, Physicochemical gap: physicochemical gap between the 2 AA (Grantham score), Modelisation gap: modelisation mid gap between α-carbons.

	R702W	**G908R**	**N1010K**
ExAC MAF in Non-Finnish CEU	0,04307	0,01187	0
GnomAD MAF	0,02355	0,007589	0
Kaviar MAF	0,2409	0,009595	0
CADD Phred	24,6	29,8	22,6
SIFT	0,01	0,01	0
PolyPhen2	0,72	0,986	0,996
Grantham Score	101	125	94
Modelisation gap	2,87 Å	3,34 Å	3,48 Å

In CADD-Phred and SIFT, the N1010K variation was predicted to be deleterious (respectively 22,6 and 0); in PolyPhen2, as possibly deleterious (0,996). The pathogenicity estimation by SIFT is based on conservation whereas PolyPhen2 consider also available biochemical information. Although asparagine and lysine are two hydrophilic amino acids, the impact of the N1010K substitution is considered to have a significant effect according to the Grantham score (94). The multiple alignment of NOD2 protein sequence showed a high conservation level of the N1010 amino acid among *vertebrata* members (Figure 2B).

These results were corroborated by the 3D protein modeling obtained by crystallographic simulation (Figure 3): the average displacement of each α-carbon of each amino acid was measured and made possible to quantify the overall deformation of the protein related with the mutation.

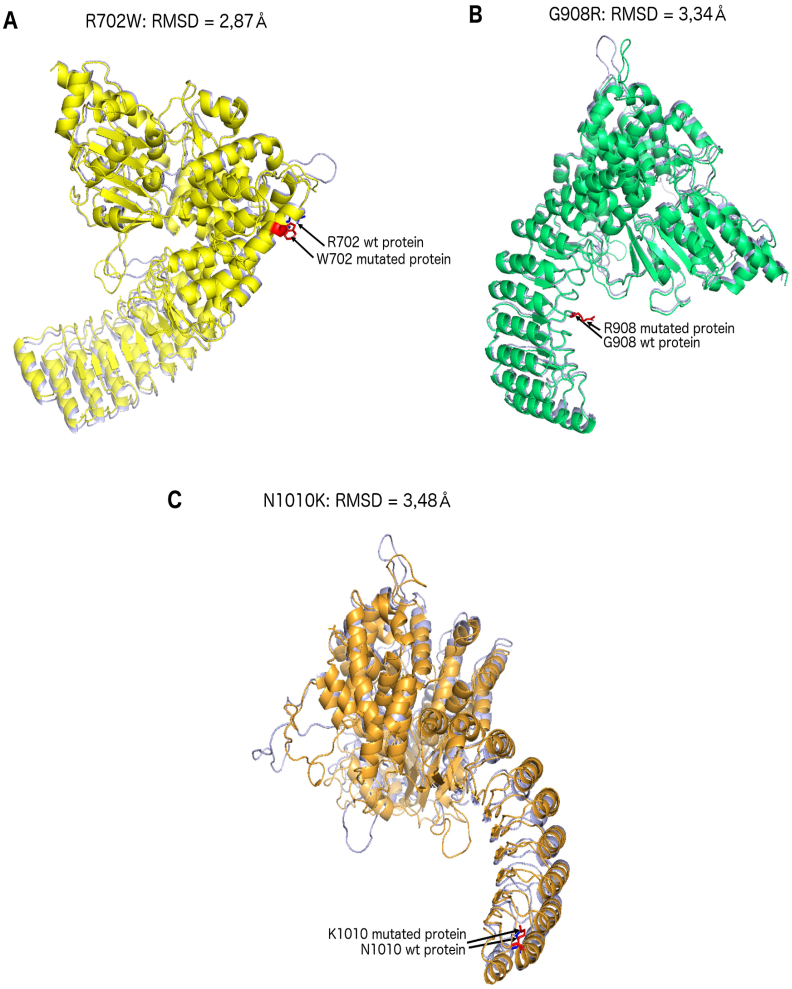

Figure 3. (**a**)Structural predictions of the human native NOD2 protein (amino-acids 219-1040) and the human R702W mutated NOD2 proteins (amino-acids 219-1040).(**b**)Structural predictions of the human native NOD2 protein (amino-acids 219-1040) and the human G908R mutated NOD2 proteins (amino-acids 219-1040). **c**)Structural predictions of the human native NOD2 protein (amino-acids 219-1040) and the human N1010K mutated NOD2 proteins (amino-acids 219-1040).

3. Discussion

Whole Exome Sequencing study with intra-familial controls allowed the detection of a new N1010K *NOD2* variation that may be related with familial aggregation of CD. *NOD2* gene encodes for a protein of the NOD-like receptors family (NLR) that contributes to the detection of intracellular bacteria and their destruction [4] and that is able to stimulate the inflammatory response through the activation of NF-κB. Three major variations altering the function of the C-terminal part of the protein have been reported as main genetic factors of CD susceptibility: p.R702W, p.G908R and p.L1007Pfs.

We could speculate that the N1010K *NOD2* variation impacts the NOD2 protein function. As well as R702W, G908R and L1007fs variations, the N1010K missense is located in the Leucine-Rich Repeat (LRR) domain of the NOD2 protein. This motif is evolutionarily conserved in many proteins and associated with innate immunity [5]. The LRR domain is known to be involved in the recognition of pathogen-associated molecular patterns including components of bacteria as the bacterial peptidoglycan muramyl dipeptide targeting by NOD2 protein [6]. Polymorphisms in the LRR domain are one of the most important genetic risk factors for the occurrence of CD [7][1].

The well-conserved Asparagine residue involved in the N1010K variation is a polar amino acid as Lysine. However, the proximity of a conserved Threonine amino acid T1012 within NOD2 protein sequence could form a [Asn-X-Ser/Thr] motif that could be considerd as a potential target for N-glycosylation making a substitution deleterious [8]. Moreover, the N1010K substitution is considered to have a significant effect according to the Grantham score based on the physico-chemical difference between Asparagine and Lysine. Regarding the highly conserved amino acid, the location within functional domain of NOD2 and the absence of N1010K variation in other databases, the N1010K variation is predicted to be deleterious with CADD Phred and SIFT *in silico* tools, and probably deleterious with PolyPhen2 *in silico* tool. The results were similar for the two known missense variations R702W and G908R (Table 2).

Protein modeling suggests that N1010K could be associated with an alteration of 3D structure of the human native NOD2 protein. There is no structural homology between the 3D predicted structure and the predicted hN1010K mutated NOD2 proteins as well as for hR702W and hG908R. Crystallographic analysis previously showed that LRR domain, between residues 745 and 1020, consisted of ten LRR units forming a horseshoe-like structure in a single curvature with alpha-helices at convex surface and beta-strand in concave faces. Thus, LRR interacts closely with HD1 and HD2 domains through 3D structure of NOD2 protein [9]. Maekawa and collaborators hypothesized that SNPs associated with CD located in LRR domain would disrupt the interaction between HD1 or HD2 and the LRR domain [9]. Therefore, N1010K could disrupt or attenuate the association between HD2 and LRR domain and could act as a loss-of-function variation. This assumption is reinforced by the significant conservation of residue N1010 and the absence of known variation in public databases.

Two out of the 4 patients were more severely affected than other family members with CD according to Montreal Classification with an earlier diagnosis [13]. Interestingly, both patients, III:2 and III:3, presented the recurrent L1007fs variation in addition to the N1010K variation in compound heterozygous state. The L1007fs variation results in a frameshift mutation that generates a truncated NOD2 protein which fails to co-localize to plasma membrane [3][10]. This mutation resulting in a truncated protein is a major genetic risk factor of CD [1][3]. Cumulative association between the two variations in LRR domain could explain the early occurence of CD in compound heterozygous patients III:2 and III:3. Cumulative effect of variants in combined heterozygous state for CD was suggested by Girardelli and collaborators, but they considered variations in two different genes: K953E in *NOD2* and S159G/G351R in *IL10RA* [11]. Homozygous variations of L1007fs were identified in patients with the largest response loss to muramyl dipeptide binding by NOD2. Same effect was not reported in case of association between L1007fs and R702W [12]. Considering proximity between L1007fs/N1010K in LRR domain and possible loss-of-function due to N1010K, similar impact to homozygous L1007fs could be hypothesized.

These results strongly suggest that the N1010K shoud be a new risk factor involved in Crohn's disease genetic susceptibility and together with the 1007fs may explain familial agregation of CD in the F49M family.

4. Materials and Methods

4.1. Subjects

One among the 22 CD multiplex families (family with 3 or more affected first degree relatives) from the population-based EPIMAD Registry was recruited. Although the 1007fs (rs2066847) mutation was present in the family but transmitted through a non affected husband (II:1), any of the R702W, G908R, L1007fs were present in affected members of generation II (figure 1A). The authors got an approval with waiver of informed consent for all subjects. This work was approved by the « Comité de Protection des Personnes Nord Ouest IV », which is the Institutional Review Board for

University Hospital of Lille. The « Comité de Protection des Personnes Nord Ouest IV » is registred by the Office for Human Research Protections Database (IORG0009553). We recruited 4 CD patients and 3 intra-familial control subjects covering 3 generations. The diagnostic criteria for CD were based on clinical, radiological, endoscopic, and histological findings, as described previously[13][14] , and phenotypes were defined according to the Montreal Classification of CD. Age at diagnosis, clinical presentation and phenotype of CD differed between patients of this family (Table 2). This project is registred in the Clinical Trial Database: NCT02851134.

Table 2. Montreal classification for CD. L1 Pure small bowel involvment; L2 pure colonic involvment; L3 Small and colonic involvments; B1 nonstricturing and nonpenetrating; B2 stricturing; B3 penetrating.

Patient identification	Age at diagnosis (y)	Location at diagnosis	Behavior at diagnosis
II:2 N1010K	30 / A2	L3	B2
II:6 N1010K	24 / A2	L1	B3
III:2 1007fs + N1010K	8 / A1	L3	B1
III:3 1007fs + N1010K	15 / A1	L3	B1

Genomic DNA was prepared from 10 ml of whole blood using the Autopure LS automate method following the manufacturers' protocols.

4.2. *Exome sequencing analysis and computer analyses*

Whole-exome sequencing [WES] was carried out in 7 persons from the F49M family : 4 CD affected and 3 intra-familial control subjects. For this purpose, we used a NimbleGenSeqCap E2 exome v3 capture in combination with Illumina next generation sequencing (on a HiSeq 4000 system using paired-end reads), following the manufacturers' protocols.

Sequence reads were mapped to the human reference genome (UCSC NCBI37/hg19) using BWA v0.7.13 software. The target was covered with a mean depth of 127,7 reads in th 7 samples. Variant detection was performed with GATK v3.3 software, and candidate variants were filtered out to fit a minimum depth of 8. Variants were then annotated with Ensembl database v75 using their Perl API, and only the non-synonymous coding, stop gain or loss, frameshift, splice site and miRNA variants were kept for further analysis. After selection of variants found only in all affected family members, the remaining ones were also annotated with dbSNP v135 and dbNSFP 3.0b2 (in silico functional predictions, various project allele frequencies, GO classification, expression and pathway data).

4.3. *Sanger sequencing confirmation of NOD2 variants*

The presence of *NOD2* variants was confirmed by Sanger sequencing, using standard protocol. The details regarding PCR primers and PCR conditions are available upon request.

4.4. *Structural Predictions : NOD2*

Structural predictions of the human native NOD2 protein (amino-acids 219-1040) and the human R702W, G908R and N1010K mutated NOD2 proteins (amino-acids 219-1040) was carried out with the M4T server (Multiple Mapping Method with Multiple Templates, ver.3.0) using the deduced amino-acid sequence of each proteins. The predicted 3D structure of proteins was compared with the 3D resolved structure of rabbit (Oryctolagus cuniculus, Oc) NOD2 (Oc NOD2 ; 86 identity sequence with the human native NOD2 protein) [9] using the molecular visualization system PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The topology of proteins is almost conserved in the 3D predicted structures as illustrated on Fig 3. To determine whether mutated NOD2 proteins was structurally homologous to the native protein, the protein SuperPose server v.1.0 [15] was used to measure the root-mean-square-deviation (*RMSD): the average distance between the alpha carbons atoms (the backbone atoms) of superimposed proteins. The RMSD value on the domain of amino acids 219-1040 for the hR702W, hG908R and hN1010K mutated NOD2 proteins was respectively equal to : 2,86 Å, 3,34 Å and 3.48 Å. If the RMSD is below 1,5 Å, two 3D structures or domains whose sequence alignment is over 30 per cent can be considered as almost homologous. The most significant result is correlated with the highest number of residues aligned. Regarding NOD2 proteins, there is no structural homology between the 3D predicted structure of the human native NOD2 protein and the human hR702W, hG908R and hN1010K mutated NOD2 proteins.

4.5. In silico Predictions and Annotations

Prediction of *in silico* pathogenicity for missenses was performed using the CADD software (version 1.3) with calculation of CADD phred score[18]. A variant was predicted to be pathogenic if the CADD phred score was above 20. Predictive tools commonly used for variant annotation including SIFT[17] and PolyPhen2[16] was also performed.

Acknowledgments: This work was supported by a grant from the Association de l'Etude des Anomalies Congénitales (AEAC). We would also like to thank all the patients and their family who have kindly contributed to this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CD: Crohn's disease
DOAJ: Directory of open access journals
GWAS: Genome wide association study
IBD: Inflammatory Bowel Disease
LD: linear dichroism
LRR: Leucine-Rich Repeat
MDPI: Multidisciplinary Digital Publishing Institute
NLR: NOD-like receptor
PCR : Polymerase Chain Reaction
RMSD: root-mean-square-deviation
TLA: Three letter acronym
WES: Whole Exome Sequencing

References

1. Hugot, J. P.; Chamaillard, M.; Zouali, H.; Lesage, S.; Cézard, J. P.; Belaiche, J.; Almer, S.; Tysk, C.; O'Morain, C. A.; Gassull, M.; Binder, V.; Finkel, Y.; Cortot, A.; Modigliani, R.; Laurent-Puig, P.; Gower-Rousseau, C.; Macry, J.; Colombel, J. F.; Sahbatou, M.; Thomas, G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **2001**, *411*, 599–603.
2. Vasseur, F.; Sendid, B.; Jouault, T.; Standaert-Vitse, A.; Dubuquoy, L.; Francois, N.; Gower-Rousseau, C.; Desreumaux, P.; Broly, F.; Vermeire, S.; Colombel, J.-F.; Poulain, D. Variants of NOD1 and NOD2 genes display opposite associations with familial risk of Crohn's disease and anti-saccharomyces cerevisiae antibody levels. *Inflammatory Bowel Diseases* **2012**, *18*, 430–438.
3. Ogura, Y.; Bonen, D. K.; Inohara, N.; Nicolae, D. L.; Chen, F. F.; Ramos, R.; Britton, H.; Moran, T.; Karaliuskas, R.; Duerr, R. H.; Achkar, J. P.; Brant, S. R.; Bayless, T. M.; Kirschner, B. S.; Hanauer, S. B.; Nuñez, G.; Cho, J. H. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.*Nature* **2001**, *411*, 603–606.
4. Philpott, DJ.; Sorbara, MT.; Robertson, SJ.; Croitoru, K.; Girardin, SE. NOD proteins: regulators of inflammation in health and disease.*Nat Rev Immunol* **2014**, *14*, 9–23.

5. Inohara; Chamaillard; McDonald, C.; Nuñez, G. NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. *Annu. Rev. Biochem.* **2005**, *74*, 355-383.
6. Inohara, N.; Ogura, Y.; Fontalba, A.; Gutierrez, O.; Pons, F.; Crespo, J.; Fukase, K.; Inamura, S.; Kusumoto, S.; Hashimoto, M.; Foster, S. J.; Moran, A. P.; Fernandez-Luna, J. L.; Nuñez, G. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J. Biol. Chem.* **2003**, *278*, 5509–5512.
7. Cho, J. H. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat. Rev. Immunol.* **2008**, *8*, 458–466.
8. Gavel, Y.; Heijne, von, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* **1990**, *3*, 433-442.
9. Maekawa, S.; Ohto, U.; Shibata, T.; Miyake, K.; Shimizu, T. Crystal structure of NOD2 and its implications in human disease.*Nat Commun* **2016**, *7*, 11813.
10. Barnich, N.; Hisamatsu, T.; Aguirre, J. E.; Xavier, R.; Reinecker, H.-C.; Podolsky, D. K. GRIM-19 interacts with nucleotide oligomerization domain 2 and serves as downstream effector of anti-bacterial function in intestinal epithelial cells. *J. Biol. Chem.* **2005**, *280*, 19021–19026.
11. Girardelli, M.; Vuch, J.; Tommasini, A.; Crovella, S.; Bianco, A. M. Novel missense mutation in the NOD2 gene in a patient with early onset ulcerative colitis: causal or chance association? *Int J Mol Sci* **2014**, *15*, 3834–3841.
12. Chen, Y.; Salem, M.; Boyd, M.; Bornholdt, J.; Li, Y.; Coskun, M.; Seidelin, J. B.; Sandelin, A.; Nielsen, O. H. Relation between NOD2 genotype and changes in innate signaling in Crohn's disease on mRNA and miRNA levels. *NPJ Genom Med* **2017**, *2*, 3.
13. Gower-Rousseau, C.; Salomez, J. L.; Dupas, J. L.; Marti, R.; Nuttens, M. C.; Votte, A.; Lemahieu, M.; Lemaire, B.; Colombel, J. F.; Cortot, A. Incidence of inflammatory bowel disease in northern France (1988-1990). *Gut* **1994**, *35*, 1433–1438.
14. Molinié, F.; Gower-Rousseau, C.; Yzet, T.; Merle, V.; Grandbastien, B.; Marti, R.; Lerebours, E.; Dupas, J. L.; Colombel, J. F.; Salomez, J. L.; Cortot, A. Opposite evolution in incidence of Crohn's disease and ulcerative colitis in Northern France (1988-1999). *Gut* **2004**, *53*, 843–848.
15. Maiti, R.; Van Domselaar, G. H.; Zhang, H.; Wishart, D. S. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **2004**, *117*, 761-771.
16. Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249.
17. Ng, P. C.; Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **2001**, *11*, 863–874.
18. Kircher, M.; Witten, D. M.; Jain, P.; O'Roak, B. J.; Cooper, G. M.; Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **2014**, *46*, 310–315.