Computational Molecular Bioscience, 2024, 14, 146-191
https://www.scirp.org/journal/cmb
ISSN Online: 2165-3453
ISSN Print: 2165-3445

# Integration between Genomic and Computational Statistical Surveys for the Screening of SNP Genetic Variants in Inflammatory Bowel Disease (IBD) Pediatric Patients*

**Dago Dougba Noel[1,2,3,4#], Koffi N'Guessan Bénédicte Sonia[4,5], Dagnogo Olefongo[6], Daramcoum Wentoin Alimata Marie-Pierre[4], Mauro Giacomelli[1,2], Dagnogo Dramane[5], Eboulé Ago Eliane Rebecca[4], Yao Saraka Didier Martial[4], Diarrassouba Nafan[4], Giovanni Malerba[5], Raffaele Badolato[1,2]**

[1]Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy
[2]Angelo Nocivelli Institute of Molecular Medicine, Children's Hospital, ASST Spedali Civili, Brescia, Italy
[3]Unit of Biostatistics and Biomathematics & Unit of Bioinformatics, Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[4]Biological Sciences Training and Research Unit, Department of Genetic and Biochemistry, Peleforo Gon Coulibaly University, Korhogo, Côte d'Ivoire
[5]Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy
[6]Biosciences Training and Research Unit, Biology and Health Laboratory, Felix Houphouet-Boigny University, Abidjan, Côte d'Ivoire
Email: #dougbanoeldago@gmail.com, #dgnoel7@gmail.com

**How to cite this paper:** Noel, D.D., Sonia, K.N.B., Olefongo, D., Marie-Pierre, D.W.A., Giacomelli, M., Dramane, D., Rebecca, E.A.E., Martial, Y.S.D., Nafan, D., Malerba, G. and Badolato, R. (2024) Integration between Genomic and Computational Statistical Surveys for the Screening of SNP Genetic Variants in Inflammatory Bowel Disease (IBD) Pediatric Patients. *Computational Molecular Bioscience*, **14**, 146-191.
https://doi.org/10.4236/cmb.2024.143006

**Received:** August 7, 2024
**Accepted:** September 3, 2024
**Published:** September 29, 2024

## Abstract

Inflammatory bowel diseases (IBD) are complex multifactorial disorders that include Crohn's disease (CD) and ulcerative colitis (UC). Considering that IBD is a genetic and multifactorial disease, we screened for the distribution dynamism of IBD pathogenic genetic variants (single nucleotide polymorphisms; SNPs) and risk factors in four (4) IBD pediatric patients, by integrating both clinical exome sequencing and computational statistical approaches, aiming to categorize IBD patients in CD and UC phenotype. To this end, we first aligned genomic read sequences of these IBD patients to hg19 human genome by using bowtie 2 package. Next, we performed genetic variant calling analysis in terms of single nucleotide polymorphism (SNP) for genes covered by at least 20 read

*Genomic and Computational Statistical Surveys for the Screening of SNP Genetic Variants in Inflammatory Bowel Disease (IBD) Pediatric Patients.
#Corresponding author.

genomic sequences. Finally, we checked for biological and genomic functions of genes exhibiting statistically significant genetic variant (SNPs) by introducing *Fitcon* genomic parameter. Findings showed *Fitcon* parameter as normalizing IBD patient's population variability, as well as inducing a relative good clustering between IBD patients in terms of CD and UC phenotypes. Genomic analysis revealed a random distribution of risk factors and as well pathogenic SNPs genetic variants in the four IBD patient's genome, claiming to be involved in: i) Metabolic disorders, ii) Autoimmune deficiencies; iii) Crohn's disease pathways. Integration of genomic and computational statistical analysis supported a relative genetic variability regarding IBD patient population by processing IBD pathogenic SNP genetic variants as opposite to IBD risk factor variants. Interestingly, findings clearly allowed categorizing IBD patients in CD and UC phenotypes by applying *Fitcon* parameter in selecting IBD pathogenic genetic variants. Considering as a whole, the study suggested the efficiency of integrating clinical exome sequencing and computational statistical tools as a right approach in discriminating IBD phenotypes as well as improving inflammatory bowel disease (IBD) molecular diagnostic process.

## Keywords

Inflammatory Bowel Disease (IBD), Crohn Disease (CD), Ulcerative Colitis (UC), Clinical Exome Analysis, Computational Statistic, SNP Genetic Variants

## 1. Introduction

Inflammatory bowel disease (IBD) is a chronic, disabling, crypto-genetic disease that progresses in relapses interspersed with periods of remission. IBD constitutes one of the major problems in hepato-gastroenterology. Numerous studies reported gut microbes role in IBD, indicating gut microbiota as an essential component in the development of mucosal lesions [1]. However, it is still unclear whether a specific individual bacterial species might be causative of IBD or only contribute in exacerbation of IBD pathogenesis. In addition, according to [2], IBD can have a hereditary tendency and affects patients of all ages and genders, but it usually occurs before the age of 30, with a peak incidence in young people aged 14 to 24. Indeed, IBD refers to two rare bowel diseases including Crohn's disease (CD) and Ulcerative Colitis (UC) or Hemorrhagic Recto-Colitis, diagnosed based on precise clinical, endoscopic, radiological and histological criteria [3]. Thus, over the last few decades, the incidence of these two pathologies has changed profoundly according to a pattern specific to each of these diseases, making them definitively distinct from each other [4]. Indeed, the inflammation in Ulcerative Colitis is continuous and limited to the mucosal layer of the colon, whereas CD is characterized by segmental transmural lesions that can affect any part of the gastrointestinal tract [5]. IBD is becoming increasingly common, particularly in Crohn's disease, but there is no specific treatment for it, and its etiology is

unknown. Furthermore, UC and CD do not increase mortality, but due to their early onset and chronicity, they induce high morbidity that impairs patients' quality of life [4]. Similarly, the highest rates are traditionally reported in Northern and Western Europe and North America, whereas in Africa, South America and Asia (including China) the incidence of IBD has long been noted as low [4]. However, various studies have demonstrated the involvement, in varying proportions, associating genetic predisposition factors with environmental, immune factors, altered digestive bacterial flora and altered intestinal permeability, which contribute to the development of intestinal lesions [3] [6]. Knowing that genetic predisposition is likely to play a greater role in the onset of inflammatory bowel disease, numerous studies have been conducted in this direction [7]-[9]. Genome-wide association studies (GWAS) have identified new single nucleotide polymorphisms (SNPs), confirmed numerous IBD susceptibility loci and identified genes such as NOD2 [10] [11], autophagy related genes such as IRGM [12] and ATG16L1 [13] [14], and genes associated with autoimmune diseases such as IL23R [15] and PTPN2 [16]-[18]. Like many other researchers, we have also used next-generation sequencing (NGS) methods to improve our understanding of the genetic and molecular basis of inflammatory bowel diseases [6] [10] [11]. It is noteworthy to underline that, computational statistical analysis of NGS data and its applications in clinical oncology as well as in medicine gain popularity. Of note, rigorous statistical scheme and demarche is needed for a correct inference regarding structural and functional genomic data. Indeed, we previously developed a computational statistical script in assessing genetic variants in CD patients [6]. Since CD and UC represent the main components of IBD troubles exhibit a quite similar phenotype, the present study set out to examine and evaluate both IBD risk factor and pathogenic single nucleotide polymorphisms (SNPs) genetic variants in discerning IBD pathologies phenotypes based on bioinformatics (i.e. NGS) tool by integrating functional genomic data to the computational statistical analysis. To this end, we considered genomic DNA read sequences from a clinical exome sequencing experiment regarding four (4) pediatric patients with evident IBD phenotype including UC and CD of the "Spedali Civili" of Brescia in Italy, aiming to highlight genomic functions that fit well with each IBD phenotype by developing our own computational statistical script in R programming environment.

## 2. Material and Methods

### 2.1. Inflammatory Bowel Disease (IBD) Patient's Population

Genomic read sequences from clinical exome samples of four (4) pediatric patients with IBD phenotype were processed. IBD patients with acronyms IBD1, IBD3 exhibit Crohn's disease phenotype, while IBD2 and IBD4 patients display respectively ulcerative colitis and recto-colitis phenotypes (**Table 1**). Age of analyzed onset inflammatory bowel disease patients ranged from four (4) to 15 years. Indeed, 50% of IBD patients in this study were female and 50% male. In addition, IBD patients' blood samples, used in the present study were collected from February to August

2018 in reference center and shipped to our laboratory for clinical exome sequencing analysis [6]. DNA library preparation and clinical exome sequencing have been performed following the Illumina MiSeq sequencer manufacture.

**Table 1.** Anthropomorphic and clinical features of analyzed inflammatory bowel (IBD) disease pediatric patients.

| IBD Patients | IBD patient 1 | IBD patient 2 | IBD patient 3 | IBD patient 4 |
|---|---|---|---|---|
| **Age (years)** | 17 | 14 | 19 | 21 |
| **Age at onset (years)** | 11 | 8 | 4 | 15 |
| **Gender** | M | F | F | M |
| **IBD type** | Crohn's disease | Ulcerative colitis | Crohn's disease | Ulcerative recto colitis |
| **Other pathologies** | – | Recurrent infections | – | Autism |

## 2.2. Genomic Read Sequences Quality Control, Alignment and Genetic Variants Calling Procedures

We executed a quality control of genomic read sequences obtained from DNA clinical exome sequencing by running Fast-Q quality control package in R programming environment. Next, we performed genomic reads sequences alignment on hg19 human genome running Bowtie 2 package in Galaxy bioinformatics platform by setting standard parameters. We selected genomic read sequence with length ranking between 290 - 300 bp as well as exhibiting quality control score threshold ≥30 for the subjacent bioinformatics as well as genomic and computational statistical analysis (**Figure S1** and **Figure S2**). We retrieved and characterized genetic variants in term of single nucleotide polymorphism (SNP) for each IBD patient by processing BAM (compressed binary files storing sequence reads) files and their indexes from the process of genomic read sequences alignment to hg19 human genome by running in frame the following bioinformatics packages and script (see below) in the Galaxy platform as following:

- *Freebyes*
- *VCF allelic-primitive*
- *SNPEff-Eff*
- *VCF ToTab-delimited*
- *SNPSift-Extract-field*
- *Gemini load*
- *Gemini database-info*
- *Gemini query*

We carried out genetic variants calling analysis (VCF = Variants Calling Format) from the BAM files and their indices resulting from aligned genomic read sequences to hg19 reference genome. The bioinformatics packages used for the genetic variant calling analysis (VCF analysis) are as following *Freebyes* [19], *VCF allelic-primitive*, *SNPEff-Eff* [20], *VCF ToTab-delimited*, *SNPSift-Extract-field*, *Gemini load* [21], *Gemini database-info* and *Gemini query*. Of note, read sequences

quality control as well as alignment process, genetic variant calling analysis and their characterization were performed on Galaxy platform. Output files of this analysis are variant calling format files (VCF). Of note, Figure 1 describes and summarizes experimental protocol of bioinformatics and genomic analysis.

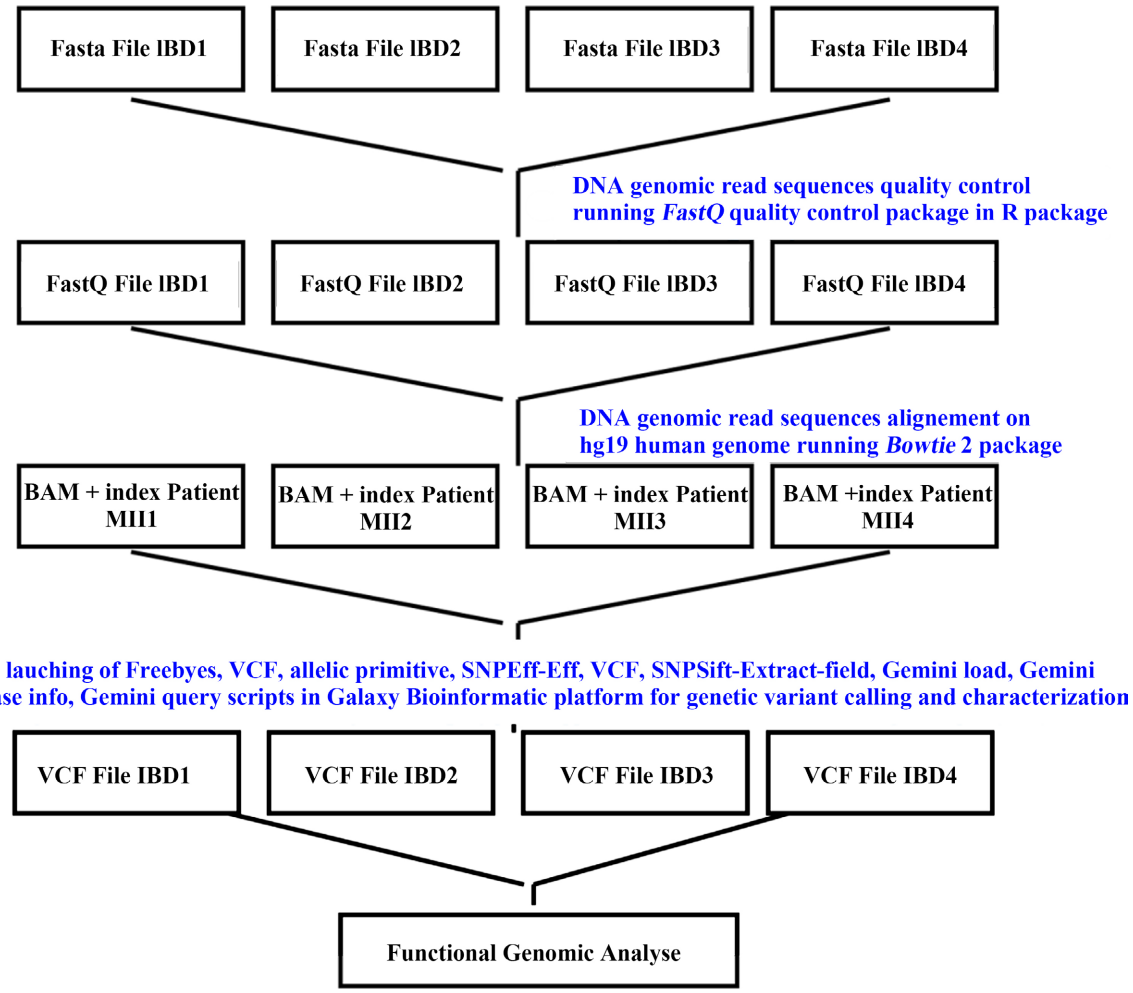


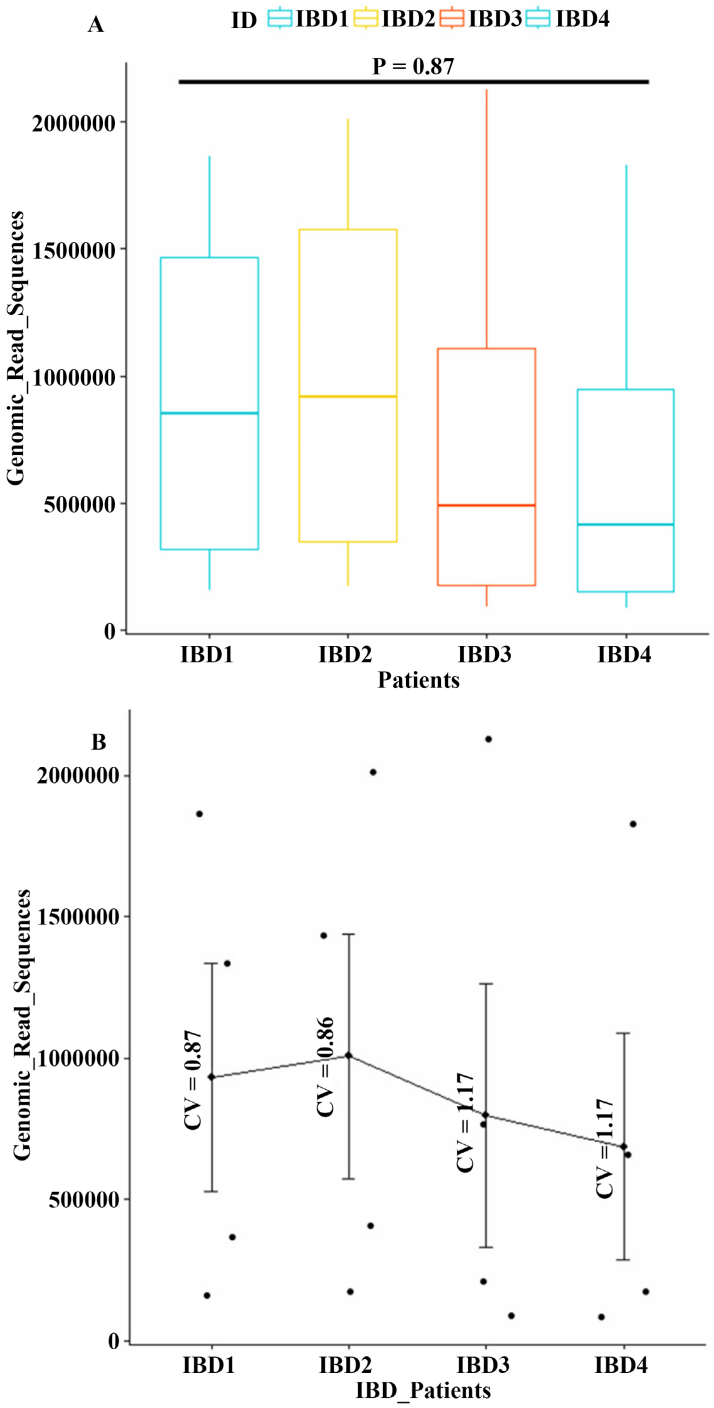

Figure 1. Bioinformatics and Genomic workflow experimental dispositive for assessing and processing inflammatory bowel disease (IBD) pediatric patients genomic read sequences for SNP genetic variants characterization and functional genomic analysis.

## 2.3. Assessment of IBD (CD and CU) Genetic Variants Genomic Functions

As a prelude to the statistical analyses, functional genomic data resulting regarding SNP retrieved from analyzed IBD patients were characterized and structured. Here, we were interested in genetic variants covered by at least 20 genomic read sequences. Indeed, SNP were classified in term of intronic and exonic mutations. Genomic functions regarding IBD patients SNP features are as following:

1) intronic mutation function: 3' and 5' un-transcribed regions (3' and 5' UTR), downstream gene variant, inter-genic region, intronic variant, upstream gene variant and splice region;

2) Exonic mutation function: induction of the **initiator** codon, nonsense mutation, loss of the stop codon, loss of the initiator codon, splice mutation and silent mutation.




(b)

**Figure 2.** Multivariate statistical analysis weighing inflammatory bowel disease patient's data variability in term of aligned genomic read sequences in prelude to genetic variants (SNP) calling survey.

### 2.4. Statistical and Functional Genomic Analysis of IBD Risk Factor and Pathogenic SNP Genetic Variants (SNPs)

We checked and quantified genomic functions regarding significant SNP retrieved from inflammatory bowel disease patients by processing data from previous obtained VCF file. Statistical analyses were performed in R programming environment (version 3.6.2). Statistical descriptive analyses included *Venn-Diagram*, *limma*, *and density plot* scripts with the purpose to characterize and assess genetic variants distribution in IBD patient population. We developed a script for endogenous variable assignment of variance for both selected IBD risk factor and pathogenic genetic variants. Analytical statistical analysis included the following statistical tests: i) Kruskal Wallis non-parametric test for variance analysis [22]; ii) Shapiro normality test; iii) Multiple pairwise comparison analysis (Turkey's statistical test); iv) Fisher exact test as well as v) Bartlett non-parametric test. Indeed, from the output of the Kruskal-Wallis test, we checked for a significant difference between groups. A significant Kruskal-Wallis test is generally followed up by Dunn's test to identify which groups are different. A test is considered significant at α threshold = 0.05 ($p \leq 0.05$). Genomic functions were assigned to retrieve pathogenic and risk factor SNP genetic variants by using ENSEMBL genomic database. The statistical parameter *Fitcon*, which indicates the probability of a given mutation having a significant impact on the phenotype, has been introduced to characterize IBD patients CD and UC phenotypes. The following *Fitcon* parameter values have been set for the: i) intronic genetic variants: risk factors and/or pathogens: $0.3 \leq Fitcon \leq 0.4$ and ii) exonic genetic variants: risk and/or pathogenic factors: $Fitcon \geq 0.6$.

## 3. Results

### 3.1. IBD Patients Genomic Read Sequences Quality Control and Alignment Statistical Comparative Analysis

We checked for IBD patients' genomic read quality control. Genomic read sequences quality Phred score is higher than 30 (**Figure S1**). Exome DNA genomic read sequences size raking between 100 - 300 bp (**Figure S2**). According to **Figure S2**, the majority of genomic sequences for the four analyzed IBD patients have a size of 300 bp. The MiSeq Illumina sequencer carried out sequencing experiment. The total number of obtained genomic read sequences of considered IBD patients is ranking between 1,830,550 - 2,128,442 sequences. Of note, the average rate of those sequences was estimated to 1,958,395 with a variation coefficient (CV) = 7.06%. Interestingly, more than 70% of aligned genomic read sequences result to be specific for their gene target, while 9% of them claim to recognize more than one gene model targets (**Table 2**). Taking together, 80% of IBD patients genomic read sequences have been aligned on the h19 human genome (**Table 2**). Of note, approximately 20% of the genomic read sequences have been not aligned on human h19 reference genome (**Table 2**). We performed a multivariate comparative statistical analysis regarding IBD patients basing on the proportion of aligned

and/or not aligned read sequences, aiming to assess read sequences distribution on h19 human genome (reference genome) in prelude to genetic variant calling as well as to subjacent functional genomic analysis. Multivariate comparative analysis shown no variance difference ($p = 0.87$) between the four analyzed IBD patients (Figure 2(A)). In the other words, inflammatory disease patients 1, 2, 3 and 4 exhibited high variance homogeneity considering aligned genomic read sequences of considered IBD pediatric patients on the h19 human genome (Figure 2(B)). Considering as a whole, IBD patients exhibited homogeneous features for the purpose of subjacent genetic variance calling and functional genomic analysis.

Table 2. Descriptive statistical analysis of IBD pediatric patients genomic read sequences aligned on hg19 human genome.

| Genomic read sequences | IBD Patient 1 (IBD1) | IBD Patient 2 (IBD2) | IBD Patient 3 (IBD3) | IBD Patient 4 (IBD4) |
|---|---|---|---|---|
| **Total number of analyzed genomic read sequences** | 1,862,628 | 2,011,960 | 2,128,442 | 1,830,550 |
| **Number and proportion (%) of read sequences aligned 1 time** | 1,335,762 (71.71%) | 1,431,974 (71.17%) | 766,939 (72.07%) | 656,309 (71.71%) |
| **Number and proportion (%) of read sequences aligned sequences > 1** | 159,094 (8.54%) | 173,284 (8.61%) | 89,295 (8.39%) | 85,580 (9.35%) |
| **Number and proportion (%) of read sequences aligned 0 times** | 367,772 (19.75%) | 406,702 (20.22%) | 207,987 (19.54%) | 17,338 (18.94%) |
| **Proportion (%) of aligned read sequences** | 80.25% | 79.78% | 80.46% | 81.06% |

## 3.2. Analysis of the Typology of Genetic Mutations Observed in the IBD Pediatric Patients' Population

Analysis revealed 6465 - 8492 genetic variants covered by at least 20 genomic read sequences for all analyzed IBD patients. IBD patient 4 recorded the lowest number of genetic mutations (6465) while IBD patient 3 recorded the highest number (8492) of genetic mutations. IBD patients 1 and 2 respectively reported 7782 and 7924 genetic variants (Table 3). Next, we discriminated the proportions of those genetic variants in term of homozygosity and heterozygosity's mutation. Findings revealed respectively 4037 and 2444 heterozygous and homozygous mutations in the IBD pediatric patient 1. The same analyzes showed respectively 4212 and 2639 heterozygous and homozygous mutations in the IBD pediatric 2. IBD pediatric patient 3 recorded respectively 4650 and 2722 heterozygous and homozygous mutations. Analysis of the typology of genetic mutation in the IBD pediatric patient 4, revealed respectively 3370 and 2215 heterozygous and homozygous mutation (Table 3). Of note, genetic mutations retrieved in the IBD population, mainly are single nucleotide polymorphisms (SNPs) inducing in some cases amino acid

changing. Thus, we recorded 3669, 3829, 4054 and 3347 amino acid changes respectively in IBD pediatric patients 1, 2, 3 and 4 (Table 3). We stimulated a genomic comparative analysis between IBD patients 1, 2, 3 and 4 basing on heterozygous and homozygous genetic variants inducing amino acid changes. Turkey's statistical test revealed a non-significant difference in term of variability ($p > 0.05$) between analyzed IBD pediatric patients 1, 2, 3 and 4 (Figure 3). This result would be in favor of the inflammatory bowel disease (IBD) phenotype shared by all of the patients.

Table 3. Analysis of inflammatory bowel disease pediatric population genetic variants inducing amino acid changes on polypeptide chain.

| | Genetic variants covered by reads sequences ≥20 | Heterozygous | Homozygous | Mutation Type | Amino Acid change |
|---|---|---|---|---|---|
| IBD Patient 1 | 7782 | 4037 | 2444 | SNP | 3669 |
| IBD Patient 2 | 7924 | 4212 | 2639 | SNP | 3829 |
| IBD Patient 3 | 8492 | 4650 | 2722 | SNP | 4054 |
| IBD Patient 4 | 6465 | 3370 | 2215 | SNP | 3347 |

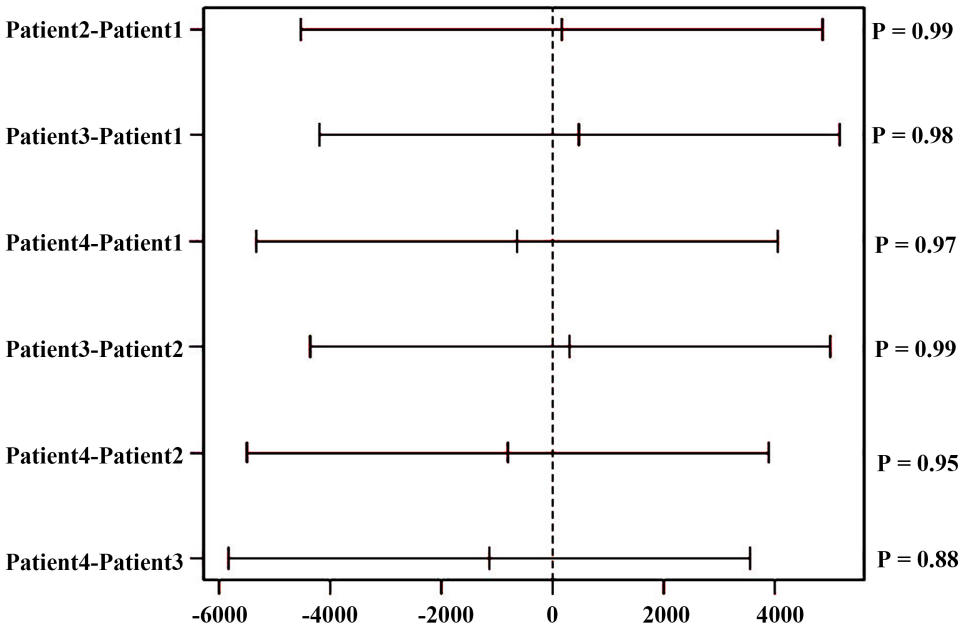



Figure 3. Genomic comparative analysis basing on homozygous and heterozygous mutation inducing amino acid change in inflammatory bowel disease pediatric patient 1, 2, 3 and 4.

### 3.3. Distribution of Genomic Functions Induced by Exonic Single Nucleotide Polymorphism (SNPs) in Influencing IBD Population Variability

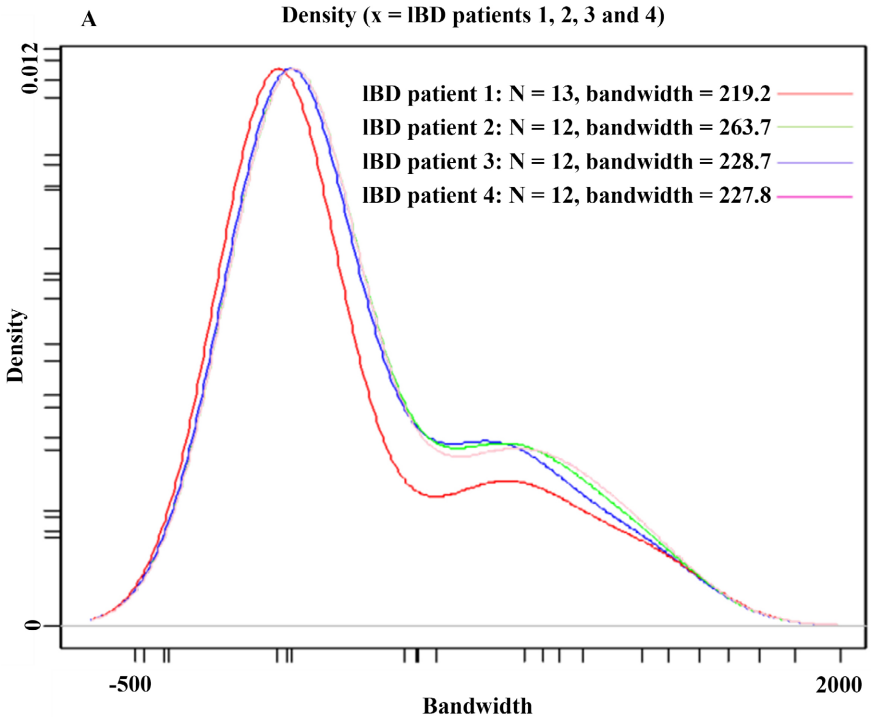
We checked for IBD population variability by assessing genomic functions induced

by exonic SNP genetic variants. Analysis revealed the following genomic functions: initiator codon induction, nonsense mutation, stop codon loss, initiator codon loss, splice mutation and silent mutation (Table 4). The recurrent genomic functions induces by exonic SNP in the analyzed IBD patients population are nonsense mutations and silent mutations, while the rare genomic functions are introduction of stop codon and gain of initiator codon respectively (Table 4). The rarest genomic functions in the present analysis are represented by i) stop codon loss in IBD patient 4 and ii) introduction of initiator codon in IBD patient 3. Interestingly, genomic analysis revealed a modest frequency of genetic mutations in several splice regions of the genomes. Indeed, Fisher's exact test showed a significant recurrence of splice mutations in heterozygote in contrast to splice mutations in homozygote ($p = 0.04$). The same test supported a non-significant difference for the genomic functions nonsense mutations and silent mutations in terms of hetero and homozygosity ($p = 0.84$). We checked for the normality distribution regarding heterozygous and homozygous mutations inducing amino acid (aa) changes in analyzed IBD population (Figure 4). Shapiro normality test suggested an asymmetric distribution ($p = 0.00$) of genomic functions associated to heterozygous and homozygous exonic SNP mutation in the IBD population (Figure 4(A)). Density plot graphic descriptive analysis, relative to those genomic functions distribution exhibited a similar profile between the four analyzed IBD patients. Kruskal Wallis non-parametric test evaluating variance difference on that IBD population in term of genomic functions distribution (33 genomic functions) revealed no significant variance difference ($p = 0.72$) between IBD patients 1, 2, 3 and 4 (Figure 4(B)). Interestingly, estimated eta-squared of the Kruskal Wallis test effect size suggested small effect (0.01 - <0.06) of above retrieved genomic functions distribution in affecting IBD population variability.

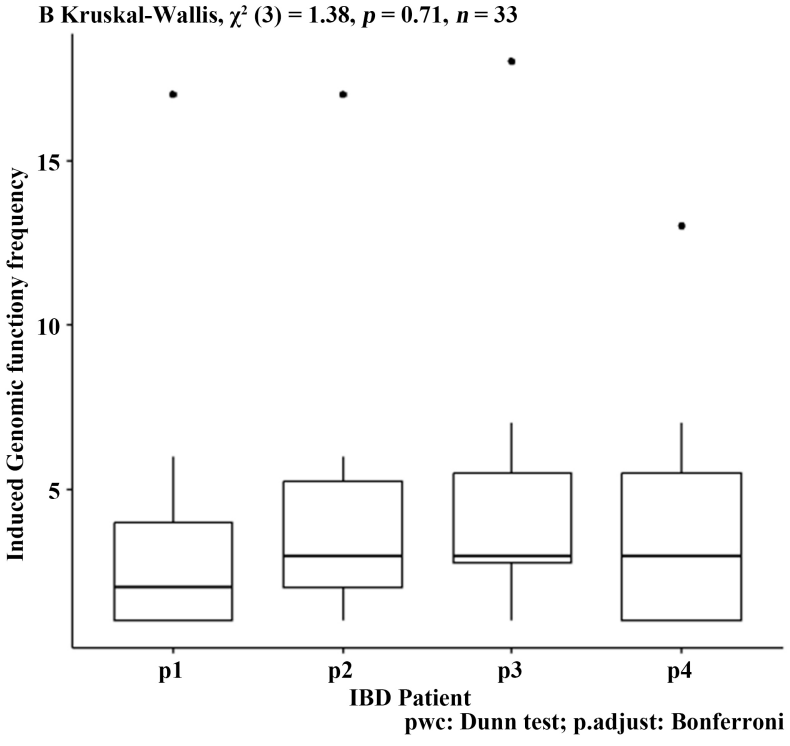
Table 4. Assessment heterozygous and homozygous SNPs exonic genetic mutations inducing and/or non-inducing amino acid (aa) changes for assessing inflammatory bowel disease (IBD) pediatric patients 1, 2, 3 and 4 variability.

| | Heterozygous mutations (SNPs) | | | | | | | | Homozygous mutations (SNPs) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **IBD patients** | IC | MM | SpM | IC L | ICG | SCI | SCL | SM | IC | MM | SpM | ICL | ICG | SCI | SCL | SM |
| **IBD Patient 1** | 0 | 958 | 17 | 2 | 1 | 6 | 1 | 1327 | 0 | 546 | 4 | 1 | 3 | 1 | 0 | 802 |
| **IBD Patient 2** | 0 | 988 | 17 | 1 | 2 | 6 | 2 | 1372 | 0 | 581 | 5 | 0 | 3 | 3 | 0 | 849 |
| **IBD Patient 3** | 1 | 1021 | 18 | 0 | 2 | 7 | 3 | 1487 | 0 | 618 | 5 | 0 | 3 | 3 | 0 | 886 |
| **IBD Patient 4** | 0 | 883 | 13 | 1 | 0 | 7 | 1 | 1185 | 0 | 496 | 5 | 0 | 4 | 2 | 1 | 749 |

IC = Initiator codon; MM = Missense mutation; SpM = Splice mutation; ICL = Initiator codon loss; ICG = Initiator codon gain; SCI = Stop codon introduction; SCL = Stop codon loss; SM = Silent mutation, IC = Initiator codon; MM = Missense mutation; SpM = Splice mutation; ICL = Initiator codon loss; ICG = Initiator codon gain; SCI = Stop codon introduction; SCL = Stop codon loss; SM = Silent mutation.




(a)




(b)

**Figure 4.** (A) Density plot accessing distribution of genomic functions induce by heterozygous and homozygous exonic SNP mutations in the IBD pediatric population (IBD pediatric patients 1, 2, 3 and 4). (B) Non-parametric test (Kruskal-Wallis test) assessing IBD patient's population variability by weighing homozygous and heterozygous exonic SNP mutation distribution in the IDB population genome.

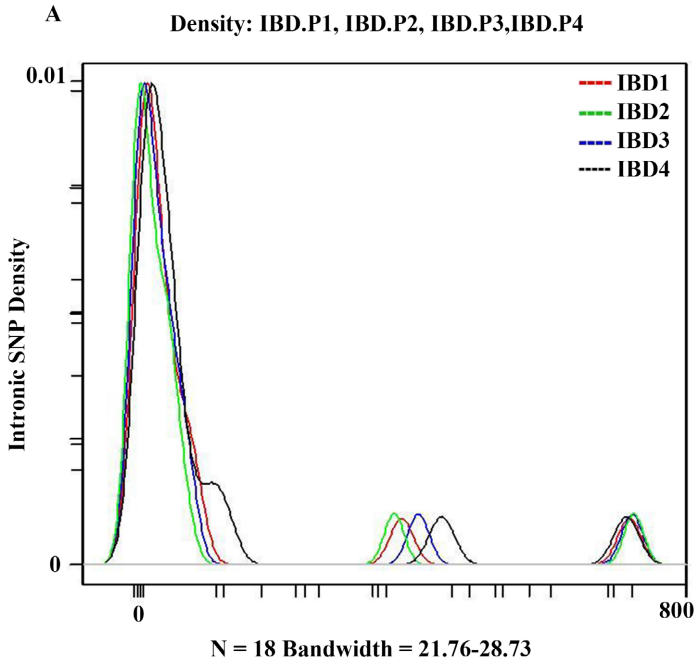
### 3.4. Analysis of the Distribution of Genomic Functions Induced by Intronic SNP Genetic Variants Influencing IBD Population's Variability

We analyzed the impact of intronic SNP genetic mutation on the genetic variability of IBD patients. Before doing that, we checked for exonic SNP distribution normality in the IBD population. Of note, genomic functions in terms of intronic SNPs genetic mutations were observed in the following genome regions: i) 3' and 5' non-transcribed regions (3' and 5' UTR), ii) downstream gene variant and iii) inter-genic regions, iv) intronic variant, v) upstream gene variant and vi) splice regions (Table 5). The majority of intronic SNPs retrieved in the IBD population are represented by i) intronic mutation variants and ii) mutation in splice and inter-genic regions (Table 5) that exhibited significant difference in term of heterozygous and homozygous mutations ($p < 0.05$). Thus, intronic gene variants are more frequent in heterozygosity than in homozygosity ($p = 0.02$). The same analysis suggests a greater number of mutations in heterozygous splice regions compared to those in homozygosity ($p = 0.0002$). Inter-gene mutations are more frequent in homozygosity ($p = 0.002$). Mutations in the 3' and 5' non-transcribed regions (the 3'UTR and 5'UTR genomic regions) have a more significant average frequency in heterozygosity than in homozygosity ($p < 0.05$) for each IDB patients (Table 5). Considering, as a whole IBD population seem to exhibit a variability in terms of intronic SNP variants detected in heterozygous and homozygous as opposite to those detected in exonic regions. We checked for the normality distribution of intronic SNP (genomic functions) in the IBD patient's population. Shapiro normality test revealed asymmetric distribution of intronic SNP (genomic functions) in the IBD population ($p < 0.05$) (Figure 5(A)). Of note, Kruskal-Wallis no-parametric test suggested no variance difference regarding IBD population by processing intronic SNP genomic functions distribution in IBD patients genome ($p > 0.05$) (Figure 5(B)).

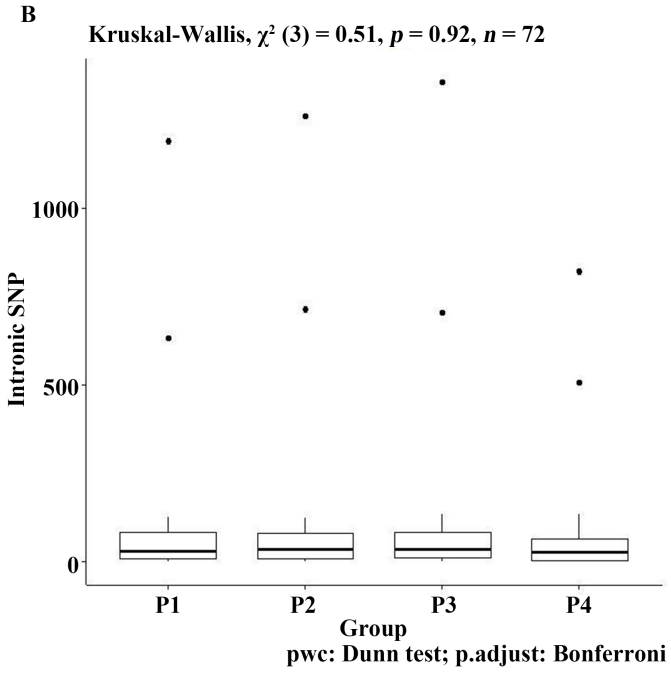
Table 5. Genomic functions retrieved from non-coding exonic and intronic SNPs in IBD pediatric patient's population (IBD1, IBD2, IBD3 and IBD4).

| | Heterozygous variants | | | | | | | | | | Homozygous variants | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3'UTR | 5'UTR | DGV | IGR | IV | SIC | SAS | SDS | SR | UGV | 3'UTR | 5'UTR | DGV | IGR | IV | SIC | SAS | SDS | SR | UGV |
| **P.1** | 57 | 74 | 3 | 88 | 1191 | 17 | 0 | 2 | 120 | 8 | 33 | 30 | 3 | 129 | 636 | 10 | 1 | 0 | 45 | 9 |
| **P.2** | 68 | 70 | 0 | 86 | 1262 | 26 | 2 | 1 | 120 | 8 | 34 | 36 | 3 | 124 | 716 | 13 | 2 | 0 | 66 | 10 |
| **P.3** | 80 | 82 | 3 | 86 | 1358 | 13 | 0 | 1 | 135 | 13 | 34 | 38 | 10 | 106 | 706 | 25 | 2 | 0 | 66 | 14 |
| **P.4** | 43 | 52 | 6 | 68 | 823 | 10 | 1 | 0 | 108 | 4 | 30 | 25 | 2 | 136 | 508 | 22 | 1 | 0 | 43 | 7 |

RNA message includes untranslated regions upstream (5'UTR) and downstream (3'UTR) of coding sequence; DGV = Downstream Gene Variant; IGR = Inter-genic Region; IV = Intronic Variant; SIC = Sequence variant that changes the non-coding exon sequence in a non-coding transcript; SAS = Splice Acceptor Site (3' end of intron); SDS = Splice Donor Site (5' end of intron); SR = Splice Region; UGV = Upstream Gene Variant.




(a)



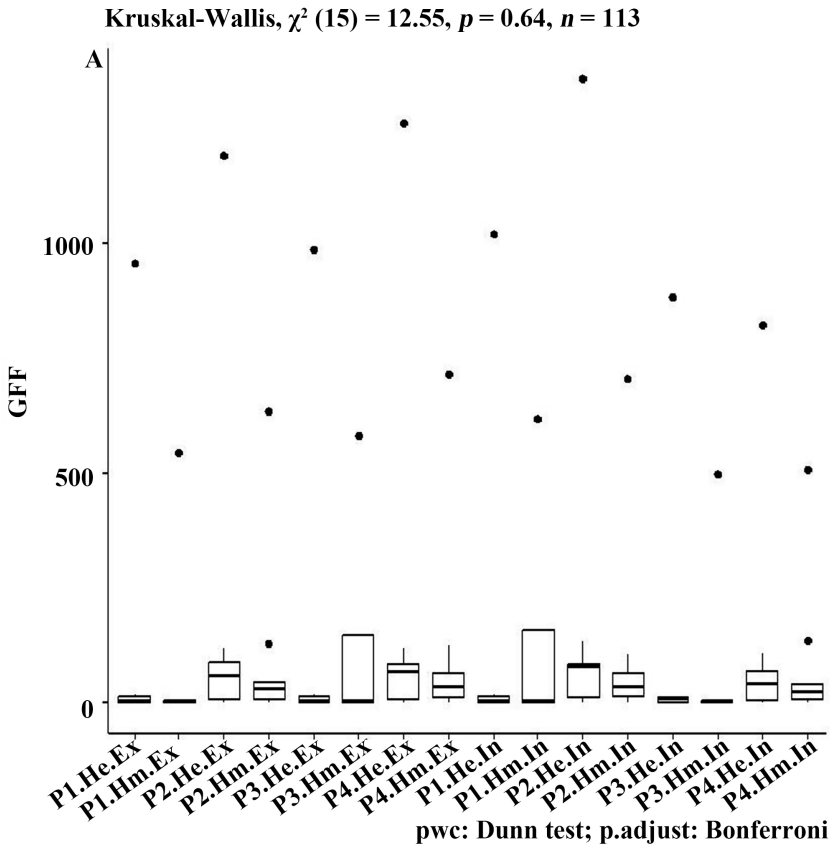

(b)

**Figure 5.** (A) Density plot referring intronic SNPs distribution in IBD patients; (B) Kruskal-Wallis non-parametric test assessing intronic SNPs genomic mutation distribution in influencing IBD.

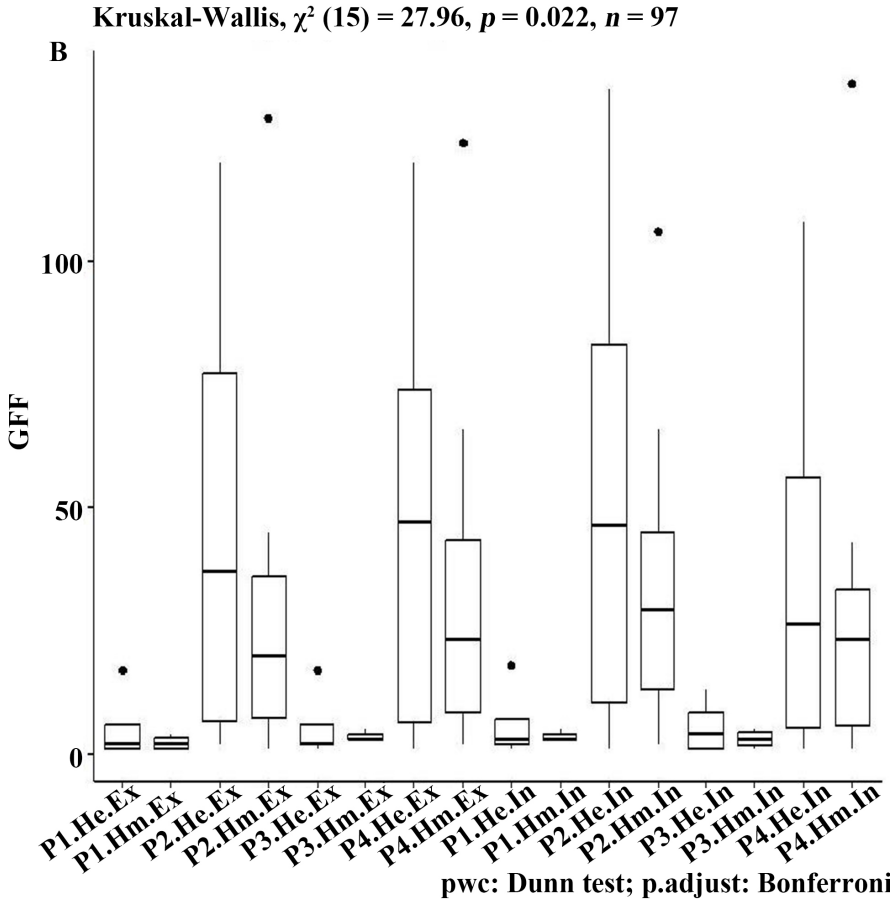
### 3.5. Clustering of Heterozygous and Homozygous SNP in Exonic and Intronic Genomic Regions for Evaluating IBD Pediatric Population Variability

We previously shown a significant variability in IBD pediatric population by

comparing heterozygous and homozygous SNPs variants in exonic genomic regions. Analysis revealed variability in terms of heterozygous non-coding region mutation (intronic SNPs) for splice, inter-genic and 3'UTR and 5'UTR genomic regions (Figure 5(A)). Herein we characterized IBD population by combining heterozygous and homozygous exonic and intronic SNP variants with the purpose to estimate that population variability. Shapiro normality test suggested random distribution regarding intronic and exonic homozygous and heterozygous SNPs in IBD patients' genome (Table 6). Of note, all analyzed IBD patients seem to exhibit the same statistic features regarding heterozygous and homozygous intronic and exonic genomic SNPs distribution (Table 6, Figure 6(A)). Interestingly, by removing intronic variants as well as exonic missense and silent SNP mutations, statistical analysis revealed significant variability in the inflammatory bowel disease pediatric patient's population (Figure 6(B) and Figure 7). Wilcoxon pairwise comparative analysis suggested significant difference between i) IBD patients 1 and 2 ($p = 0.04$), between ii) IBD patient 1 and 3 ($p = 0.02$), between iii) IBD patients 2 and 3 ($p = 0.03$), between iv) IDB patients 3 and 4 ($p = 0.02$) and between IBD 1 and 4 ($p = 0.04$) (Figure 7 and Table S1). Of note, intronic and non-coding exonic SNP exhibited a significant normality ($p < 0.05$) in the IBD patient 1, 2, 3 and 4 as opposite to the other analyzed SNP (Table S1).




(a)




(b)

**Figure 6.** Multivariate analysis weighing IBD population variability by Kruskal Wallis test by processing homozygous and heterozygous intronic and exonic SNP genetic variant (A). Multivariate statistical analysis assessing IBD population variability by removing missense, silent and intronic SNP mutation variants (B). GFF = acronym is for genomic function frequency induced by exonic and intronic SNPs.

**Table 6.** Shapiro test assessing exonic and intronic homozygote and heterozygote SNPs normality distribution in the IBD pediatric population.

| | IBD patients exonic SNPs | | | | IBD patients non-coding exonic and intronic SNPs | | | |
|---|---|---|---|---|---|---|---|---|
| | Heterozygous | | Homozygous | | Heterozygous | | Homozygous | |
| **IBD patient 1** | p | 3.068e−05 | p | 0.00015 | p | 4.703e−06 | p | 1.64e−05 |
| | w | 0.51 | w | 0.57 | w | 0.49 | w | 0.53 |
| **IBD patient 2** | p | 3.01e−05 | p | 0.0014 | p | 4.073e−06 | p | 1.379e−05 |
| | w | 0.51 | w | 0.63 | w | 0.48 | w | 0.53 |
| **IBD patient 3** | p | 3.071e−05 | p | 0.0013 | p | 4.244e−06 | p | 9.427e−06 |
| | w | 0.51 | w | 0.63 | w | 0.49 | w | 0.51 |
| **IBD patient 4** | p | 0.00019 | p | 0.00016 | p | 8.12e−06 | p | 4.492e−05 |
| | w | 0.56 | w | 0.56 | w | 0.51 | w | 0.57 |

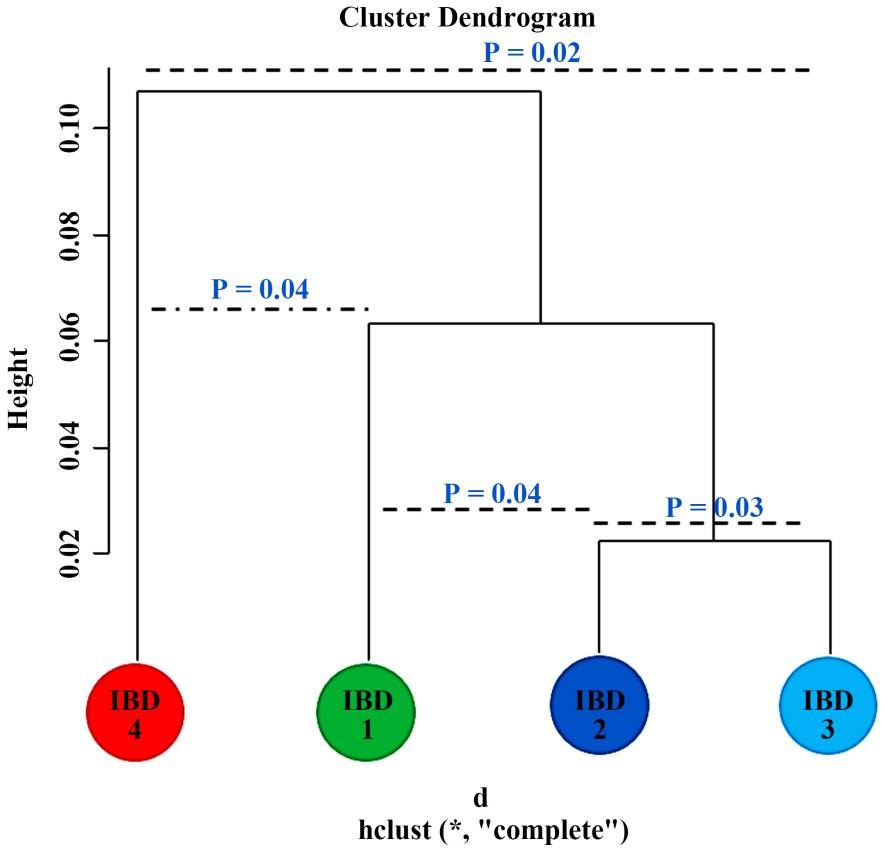



Figure 7. Hierarchical clustering analysis evaluating heterozygous and homozygous SNPs normality distribution in assessing IBD pediatric patient's variability.

### 3.6. Introduction of *Fitcon* Parameter for Measuring SNPs Inducing Significant Genomic Function Change in Influencing IBD Population Variability

We introduced *Fitcon* parameter with the purpose to reveal homozygote and heterozygote intronic as well as exonic SNPs that significantly influence IBD patient's genomic functions. Density plot analysis suggested asymmetric distribution of SNP influencing significantly IBD patient genomic functions by introducing *Fitcon* parameter (Figure 8(A)). The same descriptive statistical analysis exhibited similar distribution regarding heterozygote and homozygote SNPs genetic variants selected by *Fitcon* parameter in the IBD population (Figure 8(A)). Of note, hierarchical clustering analysis exhibit an apparent variability in the IBD pediatric population by considering homozygous and heterozygous intronic as well as exonic SNPs mutations that significantly influence IBD patient's genomic functions (Figure 8(B)). Interestingly, estimated eta-squared of the Kruskal Wallis test effect size suggested moderate effect of those SNPs genomic functions distribution in affecting IBD population variability. However, Kruskal Wallis test suggested *Fitcon* parameter as reducing IBD pediatric population variability by contrast to descriptive result of hierarchical clustering analysis (Figure S3) as well as Kruskal Wallis test effect size (Figure 8(B)).

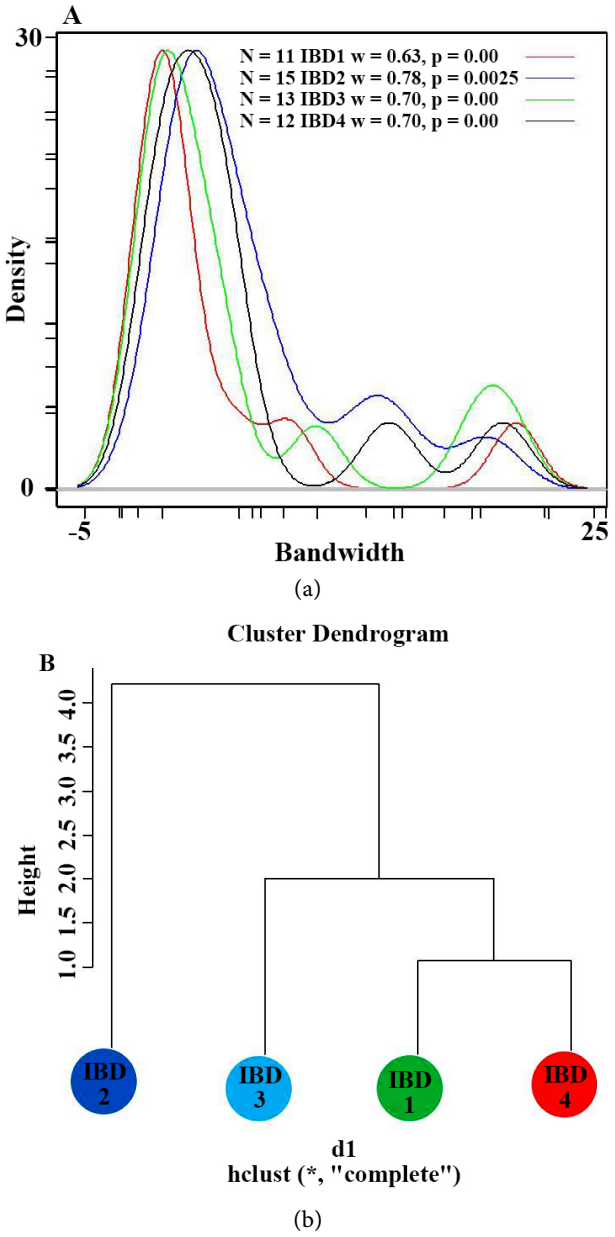



**Figure 8.** (A) Density plot descriptive statistical analysis measuring data normality by processing heterozygote and homozygote SNP selected by introducing *Fitcon* parameter. (B) Hierarchical clustering analysis evaluating IBD pediatric population variability.

### 3.7. Assessment of the Distribution of IBD Genetic Risk Factors and Pathogenic Variants in Weighing IBD Pediatric Population Variability

We analyzed the impact of pathogenic mutations and as well risk factor variants of inflammatory bowel diseases on the genetic diversity of a pediatric IBD patient population. Analysis revealed 29 and 64 IBD genetic pathogenic and risk factor variant respectively (Table 7 and Table S3 and Table S4). Of note, Shapiro normality test revealed an abnormal distribution of risk factor (Bandwidth = 0.15)

and as well pathogenic (Bandwidth = 0.17) SNPs genetic variants in the IBD population (Figure S4). Bartlett non-parametric test revealed a non-significant variance difference for analyzed IBD patient population by processing IBD genetic risk factor variants (SNPs) (p = 0.99) (Table S3). The same analysis revealed a relative significant variance difference in the IBD patients population by processing IBD pathogenic genetic variants by opposite genetic risk factor (p = 0.28) (Table S4). In addition, finding revealed that more than 95% of IBD pathogenic and risk factor genetic SNP mutations happened in the coding regions. Proportion survey regarding IBD risk factor and genetic pathogenic variants by introducing *Fitcon* parameter suggested more than 50% of exon mutations regarding IBD genetic pathogenic (55%) and risk factors (69%) displayed a significant probability value (*Fitcon* ≥ 0.6) in terms of significantly impacting IBD patient population phenotype (Table 7). Interestingly, finding revealed IBD pathogenic and risk factor SNP variants as exhibiting the same performances by introducing the *Fitcon* parameter in selecting both pathogenic and risk factor variants, confirming *Fitcon* parameter as normalizing factor in assessing IBD patient phenotype (Figure 9). Of note, genetic variability in IBD pediatric patients population through endogenous variable assignment of variance regarding normalized IBD pathogenic and risk factor variants suggested 2 IBD patients cluster groups as following, group 1 including IBD patients 1 and 3,while group 2 included IBD patients 2 and 4 (Figure 9).

Table 7. Proportion estimation of IBD risk factors and as well pathogenic SNP genetic variants affecting significantly IBD patient's phenotype by introducing *Fitcon* parameter

| | IBD genetic risk variants (SNP) | | | IBD pathogenic genetic variants (SNP) | | | |
|---|---|---|---|---|---|---|---|
| | Exon mutations | | Intron mutations | Exon mutations | | Intron mutations | |
| Total | Fitcon ≥ 0.60 | Total | 0.30 ≤ Fitcon ≤ 0.40 | Total | Fitcon ≥ 0.60 | Total | 0.30 ≤ Fitcon ≤ 0.40 |
| 64 | 35 | 2 | 2 | 29 | 20 | 2 | 2 |

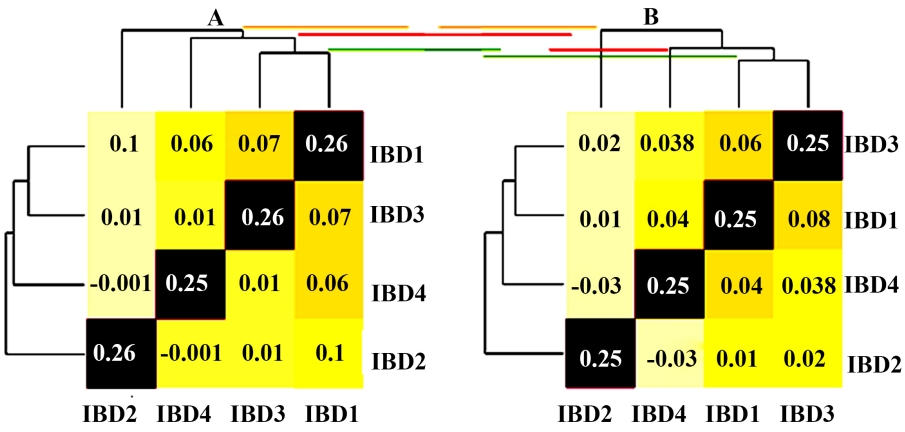



Figure 9. Evaluation of genetic variability in IBD pediatric patient's population through endogenous variable assignment of variance for selected IBD pathogenic SNPs genetic variants (A) and IBD risk factors (B) by introducing *Fitcon* parameter.

### 3.8. Distribution Analysis of Inflammatory Bowel Disease Risk Factors Genetic Variants (SNPs) in the Pediatric IBD Patient'S Population

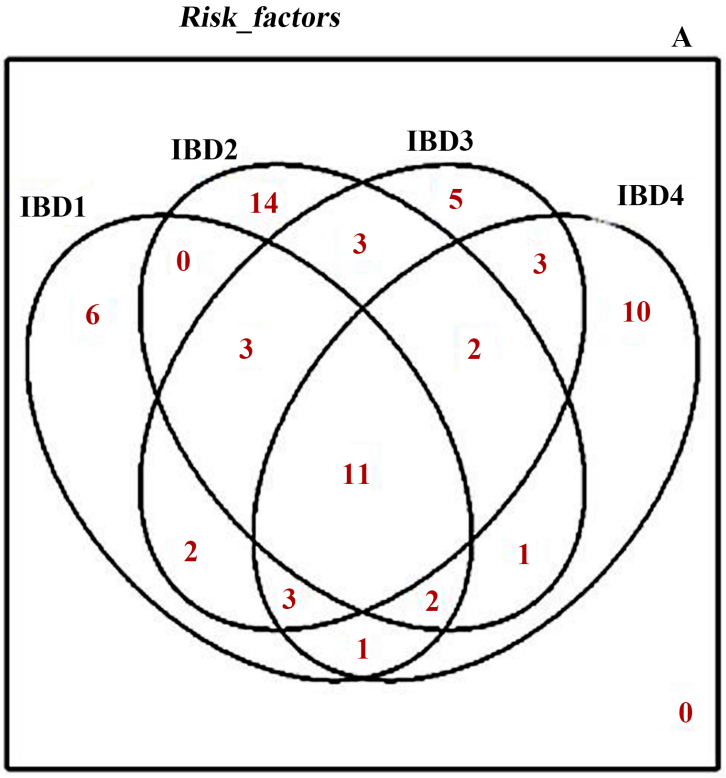
Findings revealed 66 genetic risk factor SNP variants including 14, 4 and 3 genetic risk factor variants associated with metabolic, inflammatory bowel and autoimmune diseases respectively, while 68% of revealed SNPs claim to be associated with other diseases (**Figure 10(A)** and **Table S3**). Of note, rs429358 (c.466 T > C/p. (Cys156Arg)) and rs1805097 (c.3170G > A/p. (Gly1057Asp)) genetic variants, respectively from APOE and IRS2 genes associated to metabolic disorder have been detected in IBD patient 1. IBD patient 2 exhibited three specific genetic variants associated to metabolic disorders i.e. rs4880 (c.47T > C/p. (Val16Ala)), rs13266634 (c.973C > T/p. (Arg325Trp)) and rs231775 (c.49A > G/p. (Thr17Ala)) respectively from SOD2, SLC30A8 and CTLA4 genes. rs2904552 (c.1292G > A/p.(Arg431His)) from gene PRODH involved in metabolic abnormalities, rs34911341 (c.152G > A/p.(Arg51Gln)) from GHRL gene, a susceptibility factor for obesity and rs1053874 (c.731 G > A/p.(Arg244Gln)) from DNASE1 gene, a susceptibility factor for systemic lupus erythematosus disease and body fat distribution have been specifically detected in IBD patient 4. The same analysis suggested rs12150220 homozygous variant (c.464 T > A/p. (Leu155His)) from NLRP1 gene involved in Vitiligo disease associated with multiple autoimmune diseases, rs373237 (c.935C > T/p. (Thr312Met)) from CX3CR1 gene and rs3732379 (c.841 G > A/p. (Val281Ile)) variant from CX3CR1 gene were revealed only in IBD patient 4 (**Table S3**). Of note, analysis showed inflammatory bowel disease genetic risk factors commonly shared in t the IBD pediatric population involved in metabolic, inflammatory bowel and autoimmune disorder. The homozygote genetic variants rs450046 (c.1562 G > A/p.(Arg521Gln)) and rs1799983 (c.894T > G/p.(Asp298Glu)) respectively from genes PRODH and NOS3 and the genetic variant rs237025 (c.163G > A/p.(Val55Met)) from SUMO4 gene as well as risk factors variant, rs180223 (c.2200 T > G/p(.Ser734Ala)) involved in metabolic syndrome, and homozygous genetic variant rs853326 (c.3082A > G/p.(Met1028Val)) from TG gene, associated with thyroid autoimmune diseases and rs2241880 genetic variant (c.898A > G/p.(Thr300Ala)) from ATG16L1 gene a risk factor for Crohn's disease and characteristic of inflammatory bowel diseases are shared by the four analyzed IBD patients population (**Table S3**). IBD patients 1, 2 and 3 exhibited homozygous genetic variants rs861539 (c.722C > T/p. (Thr241Met)) and rs1044498 (c.517A > C p. (Lys173Gln)), respectively from XRCC3 and ENPP1 gene, linked to metabolic disorders. IBD patients 2 and 3 share rs1799945 (c.187C > G/p. (His63Asp)) genetic variant of HFE gene, a susceptibility factor for metabolic abnormalities, while homozygous genetic variant rs1131454 (c.484G > A/p. (Gly162Ser)) of OAS1 gene, a susceptibility factor of metabolic abnormalities, is shared by IBD 1 and 3 patients (**Figure 10** and **Table S3**). Analysis revealed homozygous rs5219 (c.67A > G/p. (Lys23Glu)) genetic variant of KCNJ11 gene involved in abnormal metabolic disorders in IBD patients 1, 2 and 3 (**Figure 10**

and Table S3). IBD patients 2 and 4 share rs2066844 (c.2104C > T/p. (Arg702Trp)) genetic variant of NOD2 gene a Crohn disease risk factor and a well know IBD biomarker (Figure 10, Table S3 and Figure S5). IBD patients 2, 3 and 4 share homozygous rs7076156 (c.184A > G/p. (Thr62Ala)) genetic variant of ZNF365 gene, a risk factor of Crohn's disease. IBD patients 1, 3 and 4 share the homozygous rs2227564 (c.422T > C/p. (Leu141Pro)) and rs1169288 (c.79A > C/p. (Ile27Leu)) genetic variants respectively of PLAU and HNF1A genes that are risk factors of Crohn's disease and metabolic syndrome respectively.

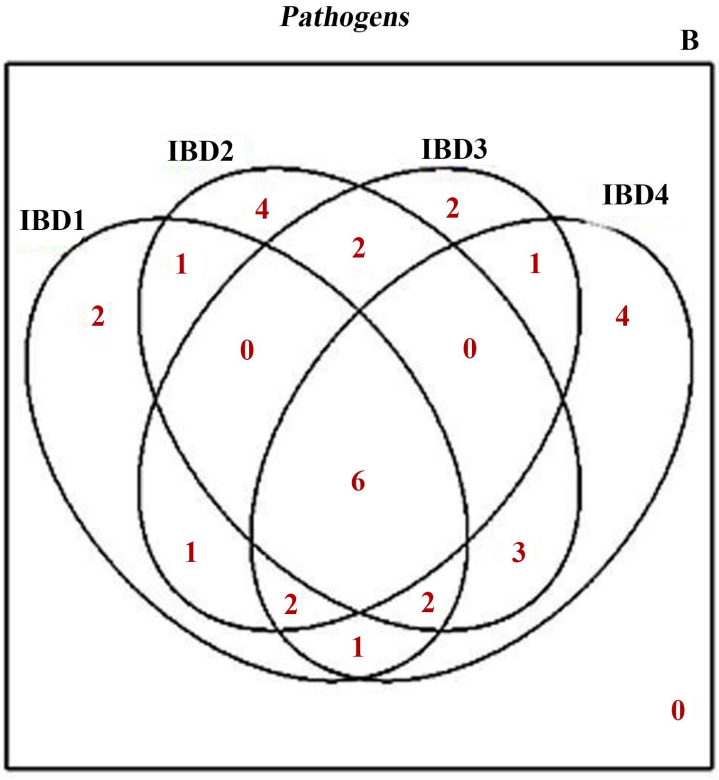
### 3.9. Distribution Analysis of Inflammatory Bowel Disease Pathogenic Genetic Variants (SNPs) in the Pediatric IBD Patient's Population

Finding revealed 31 IBD pathogenic genetic variants. Among them, nine (9) pathogenic genetic variants result to be involve in metabolic disorders, while one (1) pathogenic variant is associated to inflammatory bowel disease (Figure 10(B)). The same analysis revealed 3 pathogenic genetic variants associated with autoimmunity and 18 pathogenic variant linked to other diseases (Figure 10(B)). The analysis revealed heterozygous rs429358 (c.466T > C /p. (Cys156Arg)) pathogenic variant of APOE gene involved in metabolic abnormalities in IBD patient 1. Rs2904552 (c.1292G > A/p. (Arg431His)) pathogenic genetic variants of PRODH gene involved in metabolic abnormalities, rs3732378 (c.935 C > T/p. (Thr312Met)) and rs3732379 (c.841G > A/p. (Val281Ile)) pathogenic genetic variants of CX3CR1 gene, susceptibility factor for acquiring immunodeficiency syndrome, were revealed only in patient inflammatory bowel (IBD) patient 4 (Table S4). IBD patients 1, 2, 3 and 4 share six (6) pathogenic genetic variants (Figure 10(B)) and 4 of them i.e. rs450046 homozygous (c.1562G > A/p. (Arg521Gln)), rs820878 homozygous (c.185T > C/p. (Leu62Ser)), rs1169305 homozygous (c.1741 A > G/p. (Ser581Gly)) and rs1799983 homozygous (c.894T > G/p. (Asp298Glu)), are IBD pathogenic variants (SNPs) involved in metabolic pathologies (Table S4 and Figure 4(B)). Of note, IBD patients 3 and 4 share the pathogenic variant rs17580 (c.863A > T/p. (Glu288Val)) of SERPINA1 gene associated with the control of low-density lipoprotein cholesterol levels. IBD patients 2 and 3 shared rs1799945 (c.187C > G/p. (His63Asp)) pathogenic variant of HFE gene involved in metabolic disorders. The heterozygous rs10065172 pathogenic genetic variant (c.313C > T/p. (Leu105Leu)) of IRGM gene associated with inflammatory bowel disease was revealed in IBD patients 1, 2 and 4 patients (Figure S5). IBD patients 1, 3 and 4 share the rs10010131 and rs351855 genetic variant of WFS1 and FGFR4 genes respectively susceptible factor of metabolic syndrome and body fat distribution. The pathogenic rs1805010 genetic variant of IL4R gene involved in the acquired immunodeficiency process has been revealed in IBD patients 1, 2 and 4 (Table S4 and Figure 10(B)). Of note, statistical analysis evaluating variability in the IBD population by processing IBD pathogenic and as well risk factor genetic variants together highlights high similarity ($p > 0.05$) between Crohn's disease (CD) and

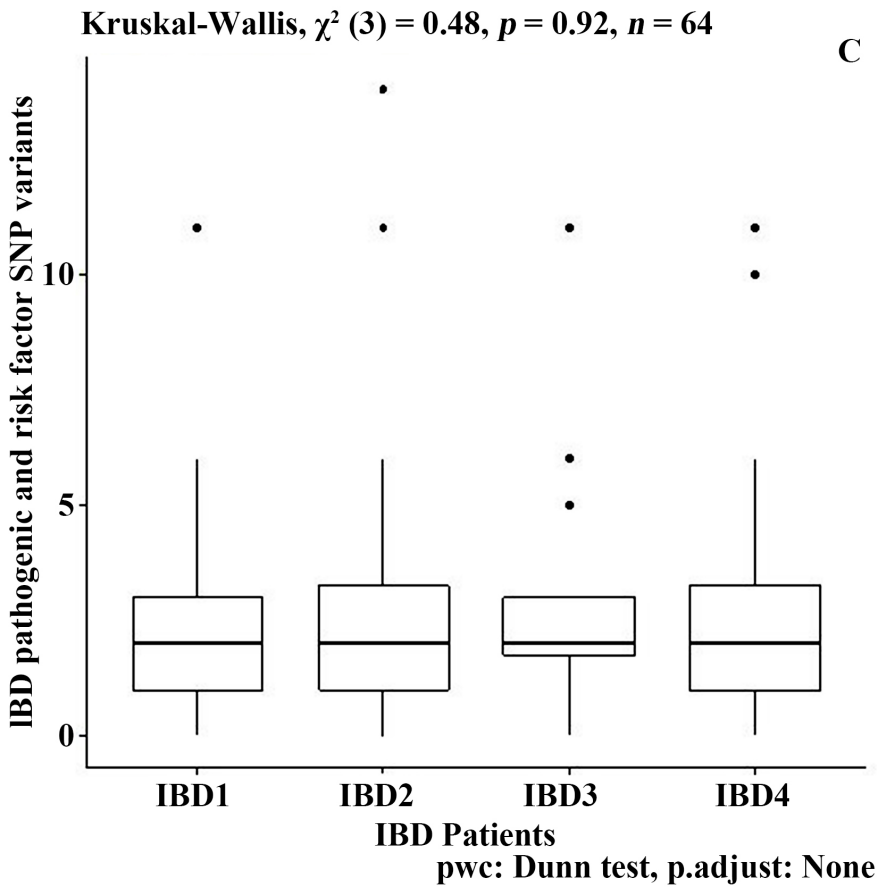
ulcerative colitis phenotype (Figure 10(C)).




(a)




(b)




(c)

**Figure 10.** Venn diagram assessing risk factor (A) and pathogenic (B) genetic variants (SNPs) distribution in inflammatory bowel disease patients 1, 2, 3 and 4. Kruskal-Wallis multivariate analysis assessing IBD patient's variability by analyzing pathogenic and risk factors genetic variants distribution in the inflammatory bowel disease pediatric patient population (C). IBD 1, 2, 3 and 4 referred to inflammatory bowel disease patient 1, 2, 3 and 4.

### 3.10. Multivariate Statistical Survey Evaluating IBD Patients' Variability by IBD Pathogenic and Risk Factors SNP Genetic Variants

Here we embarked in characterizing IBD pediatric patient's variability by processing separately IBD pathogenic and risk factor SNP genetic variants. Because of IBD pathogenic and risk factor SNP genetic variants displayed asymmetric distribution (Shapiro test, $p < 0.05$), we performed Kruskal-Wallis test assessing IBD patient population variability (Figure 11(A) and Figure 11(B), suggesting no significant variability between the four analyzed IBD pediatric patients (Figure 11(c) and Figure 11(D). However, a comparative analysis between IBD pathogenic SNP genetic variants ($p = 0.4$) and IBD SNP risk factors ($p = 0.76$) in evaluating IBD population variability, suggested a moderate aptitude of pathogenic genetic variants in categorizing IBD patients phenotype (Figure 11(C) and

Figure 11(D)). Interestingly Euclidean distance clustering analysis suggested a relative high aptitude of IBD SNP pathogenic genetic variants in categorizing IBD patient's in ulcerative colitis and Crohn's disease phenotype as opposite to IBD risk factor variants (Figure 12(A) and Figure 12(B)). In the other words, analyzed IBD patients exhibit the same phenotype by considering IBD risk factors as opposite to IBD pathogenic SNP genetic variants, that clustered together i) IBD patients 2 and 4 recognized as exhibiting ulcerative colitis phenotype and ii) IBD patients 1 and 3 patients with Crohn's disease phenotype (Figure 12(B)). This result confirm Fitcon parameter clustering analysis that suggested two IBD patients groups as following; i) IBD patients 1 and 3 and ii) IBD patients 2 and 4 (Figure 9).

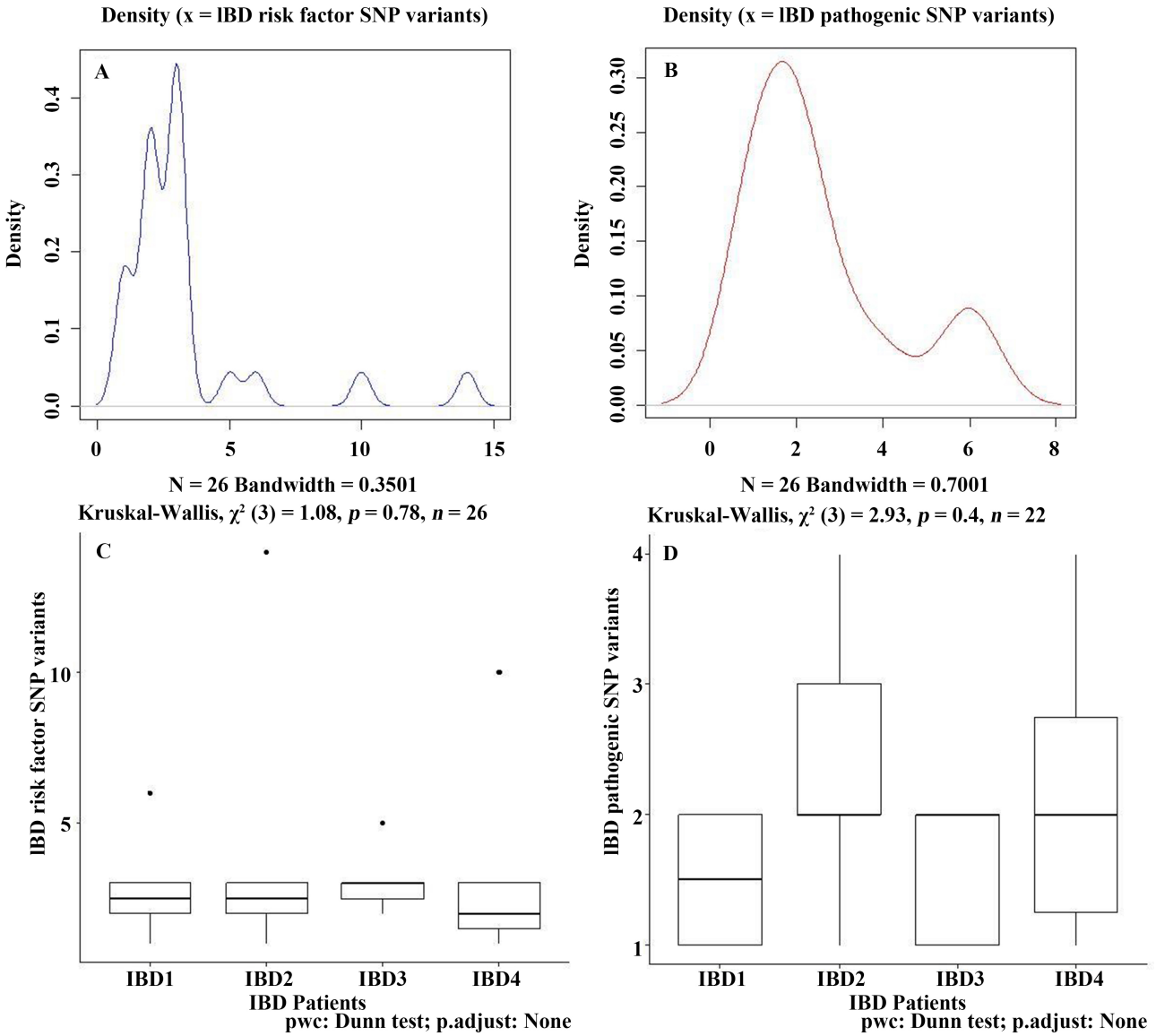



Figure 11. Inflammatory bowel disease pathogenic and risk factor SNP genetic variants distribution (A and B) in evaluating IBD patient population variability by non-parametric Kruskal-Wallis test (C and D).

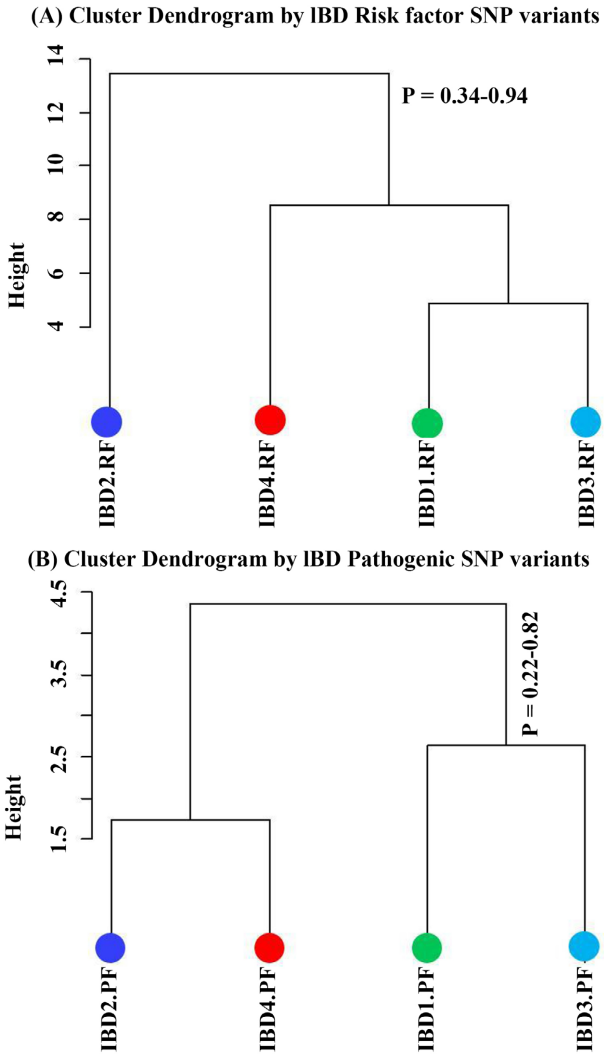



**Figure 12.** Inflammatory bowel disease patient dendrogram clustering analysis by Euclidean distance method for categorizing IBD patient basing on CD and UC, IBD phenotypes basing on IBD SNPs risk factors (A) and IBD pathogenic SNP genetic variants (B). RF and PF acronyms referred to IBD risk and pathogenic factors respectively.

## 4. Discussion

Inflammatory bowel diseases are multifactorial chronic diseases of the gastrointestinal tract including ulcerative colitis and Crohn's disease. Numerous studies have shown genetic susceptibility, gut microbiota as well as the immune system troubles as involving in the onset of intestinal diseases [3]. Many studies point to the presence of genetic, immunological, environmental, and microbiological factors and the interactions between them in the occurrence of IBD. Indeed, the first genetic factor that was linked with the occurrence of IBD i.e. CD was a mutation in the nucleotide oligomerization domain containing the protein 2 gene (NOD2). The NOD2 gene encodes a protein that functions as a receptor that recognizes components of the building wall of pathogenic bacteria. The main variants of

NOD2 mutations associated with CD are the following R702W and G908R [23]. Alterations of genes responsible for autophagy, e.g., ATG16L1 autophagy-related 16-like 1, LRRK2 repeat kinase rich in leucine, 2, and IRGM immune-related GTPase M, which can predispose to IBD, are also presented in the literature [24]-[26]. Of note, IL-10 receptor mutations (IL10RA and IL10RB) are associated with colitis [27] [28]. Study revealed some 240 gene loci associated with the predisposition and occurrence of inflammatory bowel disease confirming genetic susceptibility of IBD. It is noteworthy to underline that genomic analysis and as well, results interpretation due to the abundance and complexity of high-throughput sequencing NGS data, requires in many cases the use of computational statistical analysis. Indeed, several authors integrated genomic and computational statistical approaches in solving complex molecular genetics as well as molecular biology concerns [29]-[31]. Thus, to assess the role of genetic predisposition in the onset of inflammatory bowel disease phenotype, we screened through a computational statistic survey, for a pediatric IBD patient population by analyzing IBD risk factors and pathogenic genetic SNPs variants from clinical exome sequencing analysis.

Bioinformatics analysis suggested high quality of genomic sequences as well as high precision rate regarding those genomic sequences alignment to hg19 human genome guarantying right subjacent structural and functional genomic analysis [32]. Statistical analysis shows high homogeneity between analyzed IBD pediatric patients, by considering intronic, exonic homozygous and heterozygous pathogenic and risk factor genetic variants covered by at least 20 genomic sequence reads, since 20 per (20×) sequencing depth is enough to guarantee efficient genomic data analysis and interpretation [33]. Monitoring of biological samples variance homogeneity is an essential parameter in favor of data normalization, in prelude to statistical as well as to genomic comparative analysis **[30]** [31] [34]-[36]. We selected and quantified genomic functions related to the IBD SNPs pathogenic and risk factors genetic variants in IBD patients population, showing non-significant variance difference in the IBD patients phenotype by considering exonic SNPs genetic variants as opposite to IBD intronic variants. Indeed, computational statistical analysis suggested IBD population variability by clustering intronic variants in homozygote and heterozygote regarding several genomic functions (i.e. mutation in splice and inter-genic genomic region; mutation in 3' and 5' non-transcribed regions, 3'UTR and 5'UTR genomic regions) linked to IBD SNPs genetic variants. Koufariotis *et al*. (2018) [37] showed genetic variants in some functional classes, such as splice site regions, DNA methylated regions and long noncoding RNA as explaining more variance in complex animal genetic population. In the same tendency, Kryukov *et al*. (2007) study [38] suggested through their study most rare missense alleles in terms of intronic SNP mutation as deleterious in human in complex disease pathways. Of note, several studies have found that SNP in splice site regions are significantly associated with genetic variability [39]. Li *et al*. (2016) has delivered evidence that splicing quantitative trait loci (QTL) have major contributions to complex traits in humans; in fact, these

contributions are stated to be just as significant as variants that affect gene expression [40]. Several studies have attributed gene regulation and/or gene expression profile to polymorphisms within introns [41] [42]. Studies of genetic variation can successfully discriminate and identify functional elements in non-coding regions [43]. Considering as a whole, intronic mutations (SNP) are able to explain variability in characterizing and clustering genetic population phenotypes. However, findings revealed a relative aptitude of IBD pathogenic variants to induce variability in the analyzed IBD population by contrast risk factor variant parameter in clustering patients in both CD and UC phenotypes by introducing *Fitcon* parameter. Then, we checked for genomic functions significantly impacted by IBD risk factors as well as pathogenic variants by introducing *Fitcon* statistical probability parameter, measuring their impact on IBD patients' genome and as well phenotype. Interestingly, 57.81% of analyzed IBD risk factor SNP variants revealed by *Fitcon* parameter claimed to affect significantly IBD patient's phenotype. Of note, 3.12% of risk factor variants affecting significantly IBD patient's phenotype are intronic, while 54.69% of them result to be exonic variants. The same analysis suggested that 75.86% of IBD pathogenic variants selected by *Fitcon* statistical parameter give good contribution in clustering IBD patients in CD and UC phenotypes. Of note, 70% of IBD pathogenic genetic variants retrieved by *Fitcon* parameter are exonic, while 6.90% of them claimed to be intronic variants. Indeed, evaluating genetic variability in inflammatory bowel disease (IBD) pediatric patient's population through endogenous variable assignment of variance for selected IBD pathogenic and risk factor variants by introducing *Fitcon* parameter clustered IBD patients 1 and 3 exhibiting Chon's disease phenotype together as well as IBD patients 2 and 4 together, with respectively ulcerative colitis and ulcerative recto-colitis phenotype. Interestingly, clustering analysis based on endogenous variance assignment suggested relative strong correlation between IBD patient 1 and 3 with the Crohn diagnostic in comparison to IBD patients 2 and 4 clustering group with the ulcerative colitis and recto-colitis diagnostic. It is noteworthy to underline that variance difference between these two clustering groups in not statistically significantly different, for sure because the symptoms CD and UC are very similar. Then, non-significant genetic variability observed in the IBD population as previously mentioned could be explained by highly influential mutations of pathogenic variants such as the exonic variant synonyms rs10065172 = p.Leu105Leu of the Crohn's disease pathogenic IRGM gene [44] revealed in pediatric IBD 1, IBD 2 and IBD 4 patients. Of note, similar symptoms of both CD and UC troubles in the studied IBD population could be explained by genetic variant rs2227564 (c.422T > C/p. (Leu141Pro)) of PLAU gene, a risk factor for Crohn's disease expressed in IBD patients 1, 3 and 4 that exhibit CD and UC phenotypes respectively. The rs7076156 homozygous variant (c.184A > G/p. (Thr62Ala)) of the ZNF365 gene, a Crohn's disease susceptibility factor was reported in IBD pediatric 2 and 4 with CD phenotype as well as in IBD pediatric patient 3 with UC phenotype. In addition, the similarity between four analyzed IBD pediatric patients could be

supported by rs2241880 heterozygous SNP variant (p.Thr300Ala) of ATG16L1 gene that have retrieved in IBD patients 1, 2, 3 and 4. Indeed, IBD and CD phenotype susceptibility variant loci's have been retrieved in ATG16L1 gene involved in cellular autophagy process [6] [15]. However, performed clustering analysis focusing on significant pathogenic IBD variants suggested two phenotypes i.e. CD and UC in the analyzed IBD population. This result supports the subtle differences between IBD pathologies i.e. CD and UC, since Crohn's disease can cause inflammation anywhere in the gastrointestinal tract from the mouth to the anus while ulcerative colitis can cause inflammation and ulceration in the large intestine. Alongside intronic variants, our study has clearly indicated the impact and role of pathogenic genetic variants on the genetic variability of the IBD population since 6.90% of significantly detected IBD pathogenic variants in the four analyzed IBD patients are intronic. Findings we recorded the recurrence of the intronic rs2066844 variant (Arg702Trp) of NOD2 gene in IBD patients 2 and 4 patients exhibiting exclusively CD phenotype. Indeed, NOD2 gene is predominantly expressed by immune cells i.e. macrophages, lymphocytes and dendritic cells. Intestinal epithelial cells, known as Paneth cells, code for an intracellular receptor involved in the recognition of muramyl-dipeptide motifs found in the bacterial wall [45] [46]. Numerous studies have shown that the Arg702Trp variant of NOD2 gene, which affects the innate immune response, is one of the three best-known variants associated with inflammatory bowel infections specific to the Crohn's disease phenotype [7] [47]. So, as more non-coding sequence data becomes available, the genomic methods can be used to identify additional functional elements in the human genome and provide possible explanations for phenotypic associations. An interesting observation by Parkes (2007) **[48]** showed that genetic variants in the IRGM gene played a key role in the autophagy mechanism and were strongly correlated with the Crohn's disease phenotype [12]. Furthermore, IBD patient 2 reported IBD risk factor, rs231775 of CTLA4 gene implicated in metabolic disorder syndrome as well as susceptibility to systemic lupus erythematosus. The CTLA4 gene is expressed on the surface of T helper cells essential for the function of CD25+ CD4+ regulatory cells involved in the process of controlling intestinal inflammation [49]. It has also been shown that the CTLA4 gene variant rs231775 (g.49A > G) can control the phenotype of Crohn's disease [50]. Interestingly our study revealed several genetic variants associated with metabolic disorders in characterizing pediatric IBD patients [6]. Sztembis *et al.* (2018) [51] argued that patients with Crohn's disease had a different metabolic profile to those with ulcerative colitis. The same study showed that the occurrence of metabolic syndrome in patients with hemorrhagic colitis was higher in patients with the Crohn's disease phenotype [51] [52]. Of note, genetic variants associated with inflammatory bowel infections in the NOD2, ATG16L1 and IRGM genes affect cellular autophagy processes, so these genes indicate that alterations in the intracellular fate of bacteria are a central element in the pathogenicity of CD [53]. These observations suggest an interaction between the occurrence of inflammatory bowel disease in

general, and Crohn's disease in particular, and susceptibility to autoimmune and metabolic disorders in the pediatric IBD population under investigation [6]. Considering as a whole, our study clearly discriminating two distinct phenotypes i.e. CD and UC in the four (4) analyzed IBD pediatric patients confirming integration between genomic and computational statistical approaches as an acceptable practice aiming to improve molecular diagnostic of rare genetic disease and inflammatory bowel disease in particular. A limit of our study could be studied population sample size. However, an increasing of IBD patients sample can sturdily contribute in improving statistical significance in distinguishing both IBD phenotypes. Despite this, we proposed for the first time an integrative genomics and statistics approach for the phenotypic analysis of inflammatory bowel diseases through the clinical analysis of exome sequencing.

## 5. Conclusion

Inflammatory bowel diseases are multifactorial disorders influenced by genetic susceptibility, altered intestinal flora and immune dysfunction making challenging IBD phenotypes diagnosis. Our study provided integrative analysis including genomic, bioinformatics and computational statistical in improving IBD molecular diagnosis process allowing distinguishing clearly both IBD phenotypes i.e. Chron's disease and ulcerative colitis, by characterizing statistically IBD risk factors and as well pathogenic genetic variants, by performing clinical exome analysis regarding four (4) IBD pediatric patients.

## Authors' Contribution

RB set up the experimentation and the study. NDD proposed the protocol of the work as well as the organization of the article, figures and tables. NDD and KNBS wrote the article. MG, KNBS and DDN performed genomic data analysis and interpretation. NDD performed computational statistical analyses. NDD gave orientation for bioinformatics analysis. DD, KNBS and NDD performed the bioinformatics analyses. DO give a contribution to revising and adjusting bibliographic references. All authors revised the paper as well as approved final version of the article.

## Acknowledgements

Thank you to all Health Institute in Italy participating in this project. Thank you to the Institute of Molecular Medicine Angelo Nocivelli, University of Brescia and Children's Hospital, ASST Spedali Civili, Brescia, Italy.

## Conflicts of Interest

The authors declare no conflict of interest in this work.

## References

[1] Lepage, P., Häsler, R., Spehlmann, M.E., Rehman, A., Zvirbliene, A., Begun, A., *et al.* (2011) Twin Study Indicates Loss of Interaction between Microbiota and Mucosa of

Patients with Ulcerative Colitis. *Gastroenterology*, **141**, 227-236. https://doi.org/10.1053/j.gastro.2011.04.011

[2] Matricon, J. (2010) Immunopathogenèse des maladies inflammatoires chroniques de l'intestin. *Médecine/Sciences*, **26**, 405-410. https://doi.org/10.1051/medsci/2010264405

[3] Kökten, T., Hansmannel, F., Melhem, H. and Peyrin-Biroulet, L. (2016) Physiopathologie des maladies inflammatoires chroniques de l'intestin (MICI). *Hegel*, **2**, 119-129. https://doi.org/10.3917/heg.062.0119

[4] Corinne, G.R. (2012) Epidémiologie des maladies inflammatoires chroniques de l'Intestin en France: Apport du registre EPIMAD. Médecine humaine et pathologie. Doctorale Thèse, Université du Droit et de la Santé-Lille II.

[5] World Gastroenterology Organization Global Guidelines (2009) Maladies inflammatoires chroniques intestinales: Une approche globale juin 2009. https://www.worldgastroenterology.org/UserFiles/file/guidelines/inflammatory-bowel-disease-french-2009.pdf

[6] Noel, D.D., Marinella, P., Mauro, G., Tripodi, S.I., Pin, A., Serena, A., *et al.* (2021) Genetic Variants Assessing Crohn's Disease Pattern in Pediatric Inflammatory Bowel Disease Patients by a Clinical Exome Survey. *Bioinformatics and Biology Insights*, **15**, 1-9. https://doi.org/10.1177/11779322211055285

[7] Lesage, S., Zouali, H., Cézard, J., Colombel, J., Belaiche, J., Almer, S., *et al.* (2002) CARD15/NOD2 Mutational Analysis and Genotype-Phenotype Correlation in 612 Patients with Inflammatory Bowel Disease. *The American Journal of Human Genetics*, **70**, 845-857. https://doi.org/10.1086/339432

[8] Cho, J.H. (2008) Inflammatory Bowel Disease: Genetic and Epidemiologic Considerations. *World Journal of Gastroenterology*, **14**, 338-347. https://doi.org/10.3748/wjg.14.338

[9] Wagner, J., Sim, W.H., Ellis, J.A., Ong, E.K., Catto-Smith, A.G., Cameron, D.J.S., *et al.* (2010) Interaction of Crohn's Disease Susceptibility Genes in an Australian Paediatric Cohort. *PLOS ONE*, **5**, e15376. https://doi.org/10.1371/journal.pone.0015376

[10] Kullberg, B.J., Ferwerda, G., De Jong, D.J., Drenth, J.P.H., Joosten, L.A.B., Van der Meer, J.W.M., *et al.* (2008) Crohn's Disease Patients Homozygous for the 3020insC NOD2 Mutation Have a Defective NOD2/TLR4 Cross-Tolerance to Intestinal Stimuli. *Immunology*, **123**, 600-605. https://doi.org/10.1111/j.1365-2567.2007.02735.x

[11] Strober, W. and Watanabe, T. (2011) NOD2, an Intracellular Innate Immune Sensor Involved in Host Defense and Crohn's Disease. *Mucosal Immunology*, **4**, 484-495. https://doi.org/10.1038/mi.2011.29

[12] Singh, S.B., Davis, A.S., Taylor, G.A. and Deretic, V. (2006) Human IRGM Induces Autophagy to Eliminate Intracellular Mycobacteria. *Science*, **313**, 1438-1441. https://doi.org/10.1126/science.1129577

[13] Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., *et al.* (2006) A Genome-Wide Association Scan of Nonsynonymous SNPs Identifies a Susceptibility Variant for Crohn Disease in *ATG16L1*. *Nature Genetics*, **39**, 207-211. https://doi.org/10.1038/ng1954

[14] Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., *et al.* (2008) A Key Role for Autophagy and the Autophagy Gene *ATG16L1* in Mouse and Human Intestinal Paneth Cells. *Nature*, **456**, 259-263. https://doi.org/10.1038/nature07416

[15] Cummings, F.J.R., Ahmad, T., Geremia, A., Beckly, J., Cooney, R., Hancock, L., *et al.* (2007) Contribution of the Novel Inflammatory Bowel Disease Gene *IL23R* to Disease Susceptibility and Phenotype. *Inflammatory Bowel Diseases*, **13**, 1063-1068. https://doi.org/10.1002/ibd.20180

[16] Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., *et al*. (2006) A Genome-Wide Association Study Identifies *IL23R* as an Inflammatory Bowel Disease Gene. *Science*, **314**, 1461-1463. https://doi.org/10.1126/science.1135245

[17] Naser, S.A. (2012) Role of *ATG16L*, *NOD2* and *IL23R* in Crohn's Disease Pathogenesis. *World Journal of Gastroenterology*, **18**, 412-424. https://doi.org/10.3748/wjg.v18.i5.412

[18] Glas, J., Wagner, J., Seiderer, J., Olszak, T., Wetzke, M., Beigel, F., *et al*. (2012) *PTPN2* Gene Variants Are Associated with Susceptibility to Both Crohn's Disease and Ulcerative Colitis Supporting a Common Genetic Disease Background. *PLOS ONE*, **7**, e33682. https://doi.org/10.1371/journal.pone.0033682

[19] Garrison, E. et Marth, G. (2012) Détection de variantes basée sur l'haplotype à partir du séquençage à lecture courte. arXiv: 1207.3907.

[20] Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., *et al*. (2012) A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly*, **6**, 80-92. https://doi.org/10.4161/fly.19695

[21] Paila, U., Chapman, B.A., Kirchner, R. and Quinlan, A.R. (2013) GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. *PLOS Computational Biology*, **9**, e1003153. https://doi.org/10.1371/journal.pcbi.1003153

[22] Maciej, T. and Ewa, T. (2014) The Need to Report Effect Size Estimates Revisited. An Overview of Some Recommended Measures of Effect Size. *Trends in Sport Sciences*, **1**, 19-25.

[23] Eckmann, L. and Karin, M. (2005) NOD2 and Crohn's Disease: Loss or Gain of Function? *Immunity*, **22**, 661-667. https://doi.org/10.1016/j.immuni.2005.06.004

[24] Lauro, M.L., Burch, J.M. and Grimes, C.L. (2016) The Effect of NOD2 on the Microbiota in Crohn's Disease. *Current Opinion in Biotechnology*, **40**, 97-102. https://doi.org/10.1016/j.copbio.2016.02.028

[25] Matsuzawa-Ishimoto, Y., Shono, Y., Gomez, L.E., Hubbard-Lucey, V.M., Cammer, M., Neil, J., *et al*. (2017) Autophagy Protein *ATG16L1* Prevents Necroptosis in the Intestinal Epithelium. *Journal of Experimental Medicine*, **214**, 3687-3705. https://doi.org/10.1084/jem.20170558

[26] Foerster, E.G., Mukherjee, T., Cabral-Fernandes, L., Rocha, J.D.B., Girardin, S.E. and Philpott, D.J. (2021) How Autophagy Controls the Intestinal Epithelial Barrier. *Autophagy*, **18**, 86-103. https://doi.org/10.1080/15548627.2021.1909406

[27] Pigneur, B., Escher, J., Elawad, M., Lima, R., Buderus, S., Kierkus, J., *et al*. (2013) Phenotypic Characterization of Very Early-Onset IBD Due to Mutations in the IL10, IL10 Receptor Alpha or Beta Gene: A Survey of the GENIUS Working Group. *Inflammatory Bowel Diseases*, **19**, 2820-2828. https://doi.org/10.1097/01.mib.0000435439.22484.d3

[28] Ananthakrishnan, A.N. (2015) Epidemiology and Risk Factors for IBD. *Nature Reviews Gastroenterology & Hepatology*, **12**, 205-217. https://doi.org/10.1038/nrgastro.2015.34

[29] Asimit, J.L., Day-Williams, A.G., Morris, A.P. and Zeggini, E. (2012) ARIEL and AMELIA: Testing for an Accumulation of Rare Variants Using Next-Generation Sequencing Data. *Human Heredity*, **73**, 84-94. https://doi.org/10.1159/000336982

[30] Lange, K., Papp, J.C., Sinsheimer, J.S. and Sobel, E.M. (2014) Next-Generation Statistical Genetics: Modeling, Penalization, and Optimization in High-Dimensional Data. *Annual Review of Statistics and Its Application*, **1**, 279-300. https://doi.org/10.1146/annurev-statistics-022513-115638

[31] Noel, D.D., Nafan, D., Inza, J.F., Jean-Luc, A.M., Hermann-Desire, L., Didier, M.Y.S., *et al*. (2017) DEXseq and Cuffdiff Approaches Weighing Differential Spliced Genes Exons Modulation in Estrogen Receptor β (Erβ) Breast Cancer Cells. *African Journal of Biotechnology*, **16**, 1404-1427. https://doi.org/10.5897/ajb2016.15860

[32] Minoche, A.E., Dohm, J.C. and Himmelbauer, H. (2011) Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems. *Genome Biology*, **12**, R112. https://doi.org/10.1186/gb-2011-12-11-r112

[33] Farhan, S.M.K., Dilliott, A.A., Ghani, M., Sato, C., Liang, E., Zhang, M., *et al*. (2016) The ONDRISeq Panel: Custom-Designed Next-Generation Sequencing of Genes Related to Neurodegeneration. *npj Genomic Medicine*, **1**, Article No. 16032. https://doi.org/10.1038/npjgenmed.2016.32

[34] Dago, D.N., Scafoglio, C., Rinaldi, A., Memoli, D., Giurato, G., Nassa, G., *et al*. (2015) Estrogen Receptor Beta Impacts Hormone-Induced Alternative mRNA Splicing in Breast Cancer Cells. *BMC Genomics*, **16**, Article No. 367. https://doi.org/10.1186/s12864-015-1541-1

[35] Noel, D.D., Sonia, K.N.B., Martial, Y.S.D., *et al*. (2021) Assessment of Genetic Variability in an Inflammatory Bowel Disease Patients Population by a Clinical Exome Survey. *Proteomics Bioinformatics*, **3**, 1 p.

[36] Noel, D.D. (2021) Normality Assessment of Several Quantitative Data Transformation Procedures. *Biostatistics and Biometrics Open Access Journal*, **10**, Article 555786. https://doi.org/10.19080/bboaj.2021.10.555786

[37] Koufariotis, L.T., Chen, Y.P., Stothard, P. and Hayes, B.J. (2018) Variance Explained by Whole Genome Sequence Variants in Coding and Regulatory Genome Annotations for Six Dairy Traits. *BMC Genomics*, **19**, Article No. 237. https://doi.org/10.1186/s12864-018-4617-x

[38] Kryukov, G.V., Pennacchio, L.A. and Sunyaev, S.R. (2007) Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. *The American Journal of Human Genetics*, **80**, 727-739. https://doi.org/10.1086/513473

[39] Levenstien, M.A. and Klein, R.J. (2011) Predicting Functionally Important SNP Classes Based on Negative Selection. *BMC Bioinformatics*, **12**, Article No. 26. https://doi.org/10.1186/1471-2105-12-26

[40] Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., *et al*. (2016) RNA Splicing Is a Primary Link between Genetic Variation and Disease. *Science*, **352**, 600-604. https://doi.org/10.1126/science.aad9417

[41] Li, M. and Pritchard, P.H. (2000) Characterization of the Effects of Mutations in the Putative Branchpoint Sequence of Intron 4 on the Splicing within the Human Lecithin: Cholesterol Acyltransferase Gene. *Journal of Biological Chemistry*, **275**, 18079-18084. https://doi.org/10.1074/jbc.m910197199

[42] Artiga, M., Sáez, A., Romero, C., Sánchez-Beato, M., Mateo, M., Navas, C., *et al*. (2002) A Short Mutational Hot Spot in the First Intron of BCL-6 Is Associated with Increased BCL-6 Expression and with Longer Overall Survival in Large B-Cell Lymphomas. *The American Journal of Pathology*, **160**, 1371-1380. https://doi.org/10.1016/s0002-9440(10)62564-3

[43] Lomelin, D., Jorgenson, E. and Risch, N. (2009) Human Genetic Variation Recognizes Functional Elements in Noncoding Sequence. *Genome Research*, **20**, 311-319. https://doi.org/10.1101/gr.094151.109

[44] Glas, J., Seiderer, J., Bues, S., Stallhofer, J., Fries, C., Olszak, T., *et al*. (2013) IRGM Variants and Susceptibility to Inflammatory Bowel Disease in the German Population. *PLOS ONE*, **8**, e54338. https://doi.org/10.1371/journal.pone.0054338

[45] Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., *et al*. (2001) A Frameshift Mutation in NOD2 Associated with Susceptibility to Crohn's Disease. *Nature*, **411**, 603-606. https://doi.org/10.1038/35079114

[46] Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., *et al*. (2003) Host Recognition of Bacterial Muramyl Dipeptide Mediated through $NOD_2$. *Journal of Biological Chemistry*, **278**, 5509-5512. https://doi.org/10.1074/jbc.c200673200

[47] Hoefkens, E., Nys, K., John, J.M., Van Steen, K., Arijs, I., Van der Goten, J., *et al*. (2013) Genetic Association and Functional Role of Crohn Disease Risk Alleles Involved in Microbial Sensing, Autophagy, and Endoplasmic Reticulum (ER) Stress. *Autophagy*, **9**, 2046-2055. https://doi.org/10.4161/auto.26337

[48] Ajayi, T.A., Innes, C.L., Grimm, S.A., Rai, P., Finethy, R., Coers, J., *et al*. (2019) Crohn's Disease *IRGM* Risk Alleles Are Associated with Altered Gene Expression in Human Tissues. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, **316**, G95-G105. https://doi.org/10.1152/ajpgi.00196.2018

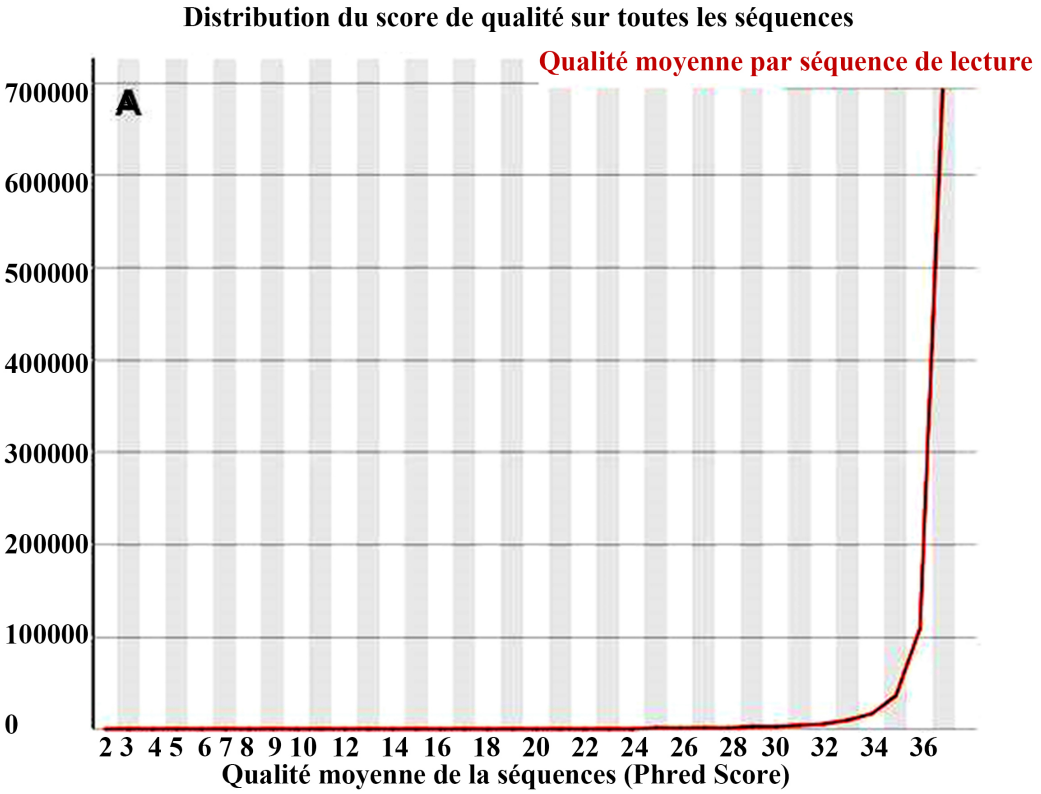
[49] Read, S., Malmström, V. and Powrie, F. (2000) Cytotoxic T Lymphocyte-Associated Antigen 4 Plays an Essential Role in the Function of $CD25^+CD4^+$ Regulatory Cells That Control Intestinal Inflammation. *The Journal of Experimental Medicine*, **192**, 295-302. https://doi.org/10.1084/jem.192.2.295

[50] Xia, B., Crusius, J.B.A., Zwiers, A., Bodegraven, A.A.V., Peña, A.S. and Wu, J. (2002) CTLA4 Gene Polymorphisms in Dutch and Chinese Patients with Inflammatory Bowel Disease. *Scandinavian Journal of Gastroenterology*, **37**, 1296-1300. https://doi.org/10.1080/003655202761020579

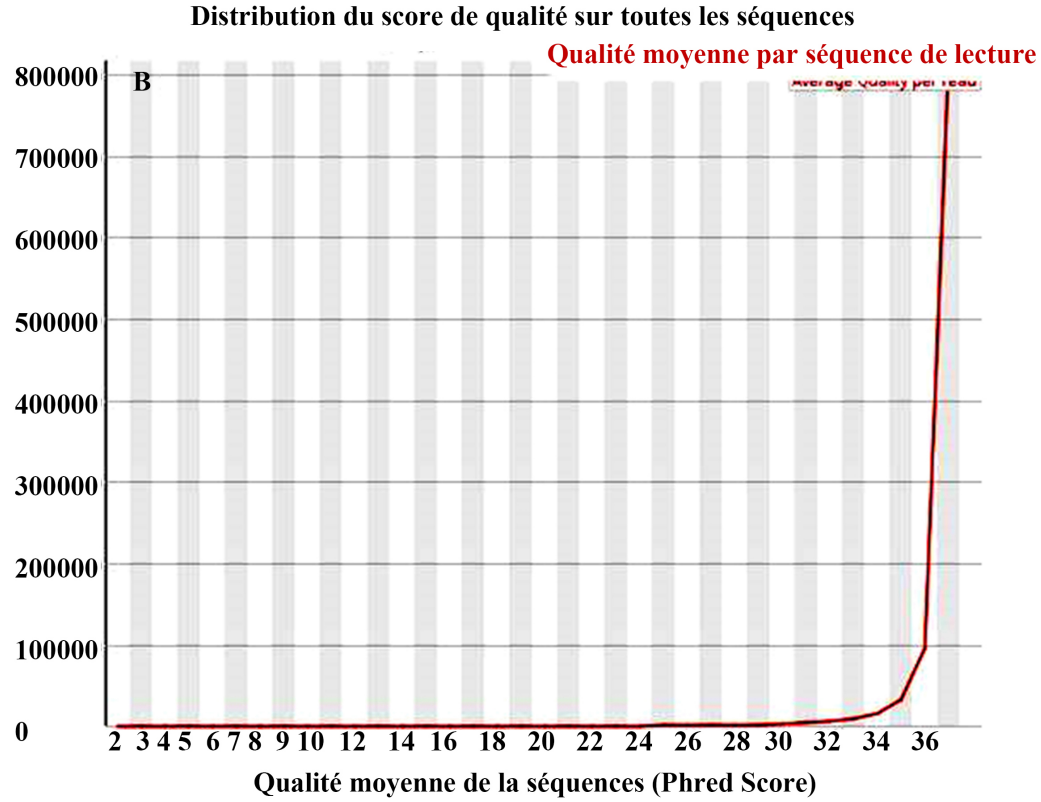
[51] Sztembis, J., Filip, R., Pękala, A., Kiela, P.R., Witas, B., Jarmakiewicz, S., *et al*. (2018) P834 Metabolic Syndrome Occurrence in Patients with Inflammatory Bowel Disease in Poland—Preliminary Results from the POLIBD Study. *Journal of Crohn's and Colitis*, **12**, S538-S538. https://doi.org/10.1093/ecco-jcc/jjx180.961

[52] Sasaki, M. and Klapproth, J.A. (2012) The Role of Bacteria in the Pathogenesis of Ulcerative Colitis. *Journal of Signal Transduction*, **2012**, 1-6. https://doi.org/10.1155/2012/704953

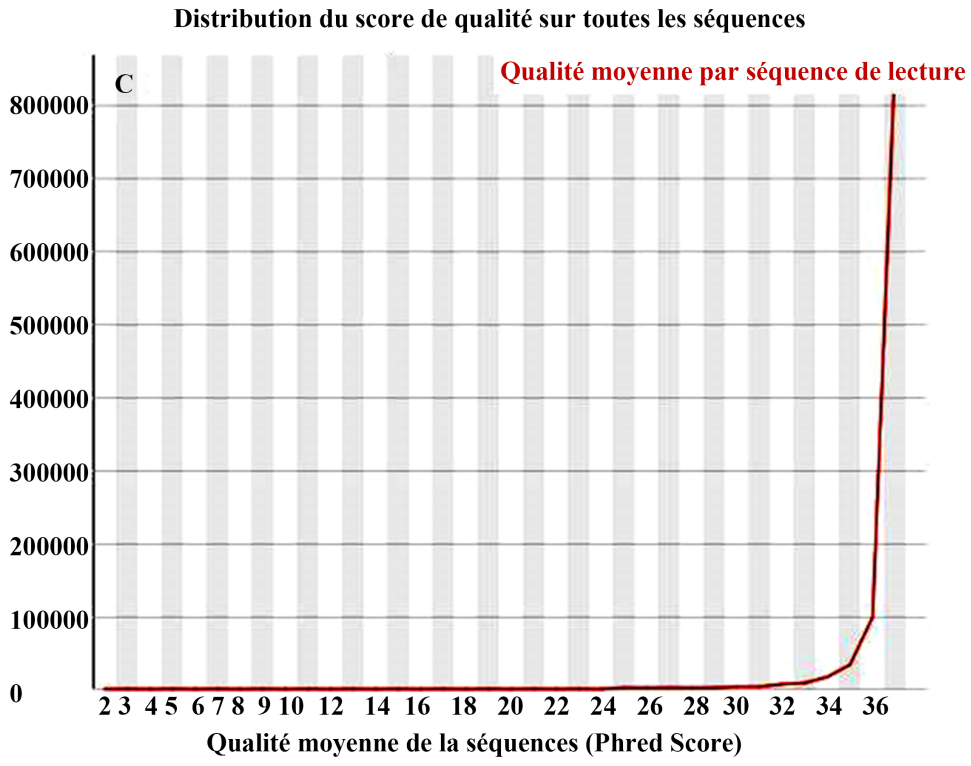
[53] Kuballa, P., Huett, A., Rioux, J.D., Daly, M.J. and Xavier, R.J. (2008) Impaired Autophagy of an Intracellular Pathogen Induced by a Crohn's Disease Associated *ATG16L1* Variant. *PLOS ONE*, **3**, e3391. https://doi.org/10.1371/journal.pone.0003391


Distribution du score de qualité sur toutes les séquences
Qualité moyenne par séquence de lecture
A
700000
600000
500000
400000
300000
200000
100000
0
2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 36
Qualité moyenne de la séquences (Phred Score)

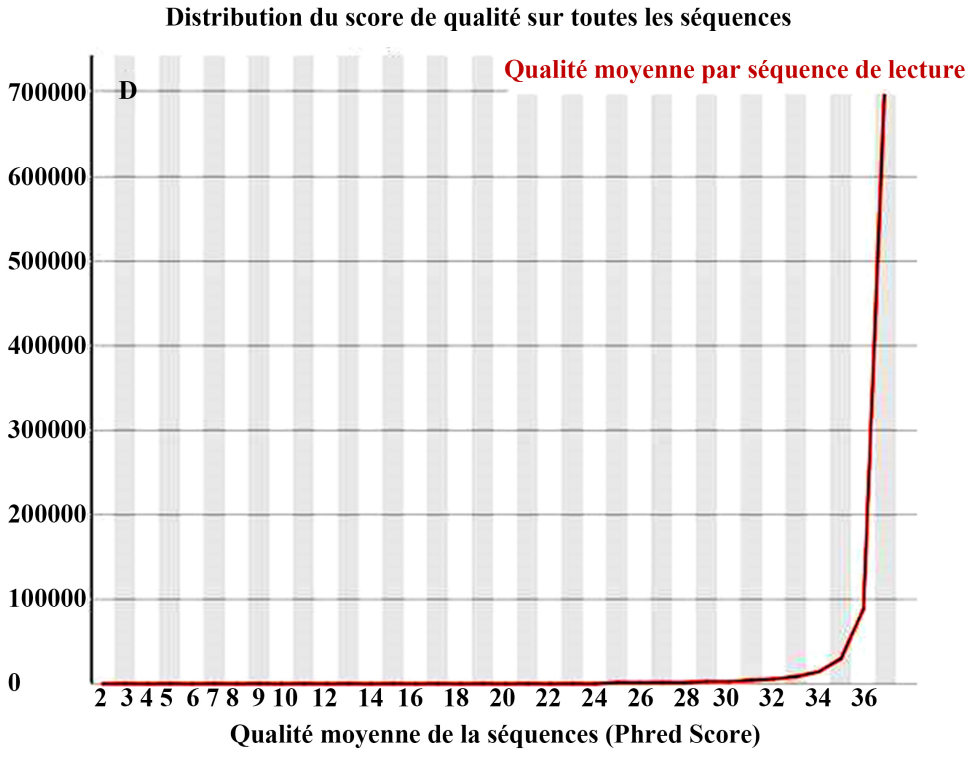
(a)


Distribution du score de qualité sur toutes les séquences
Qualité moyenne par séquence de lecture
B
800000
700000
600000
500000
400000
300000
200000
100000
0
2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 36
Qualité moyenne de la séquences (Phred Score)

(b)

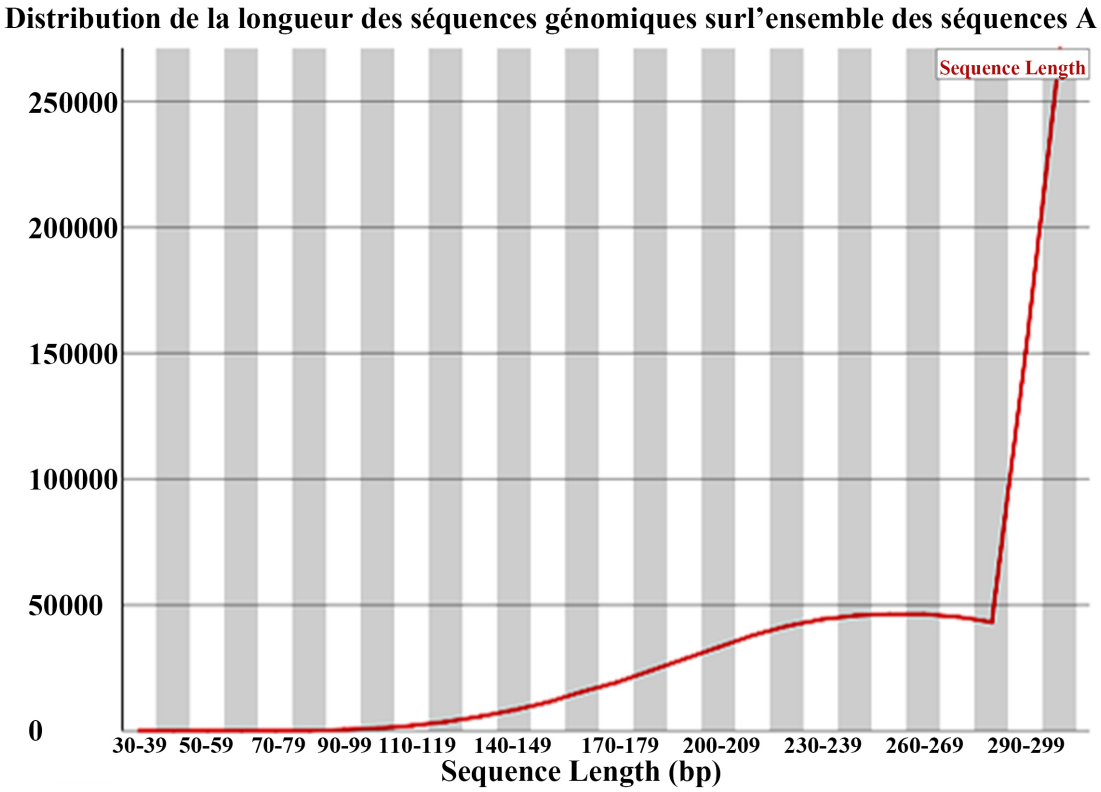



(c)

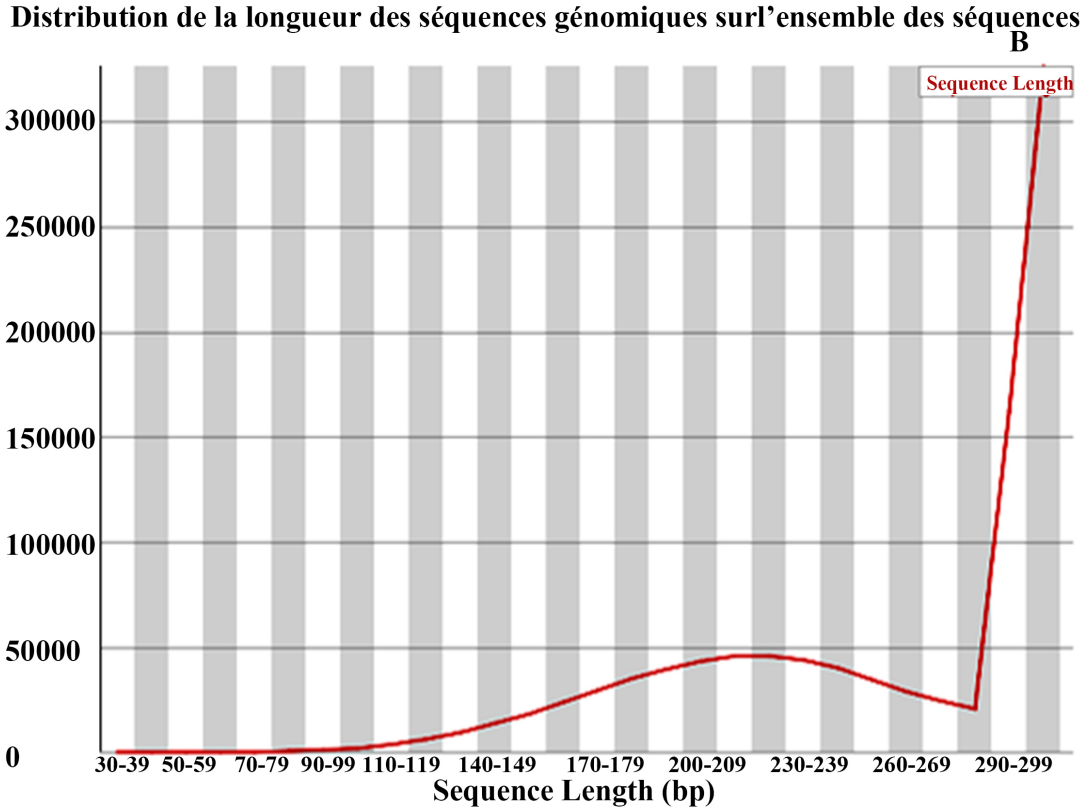



(d)

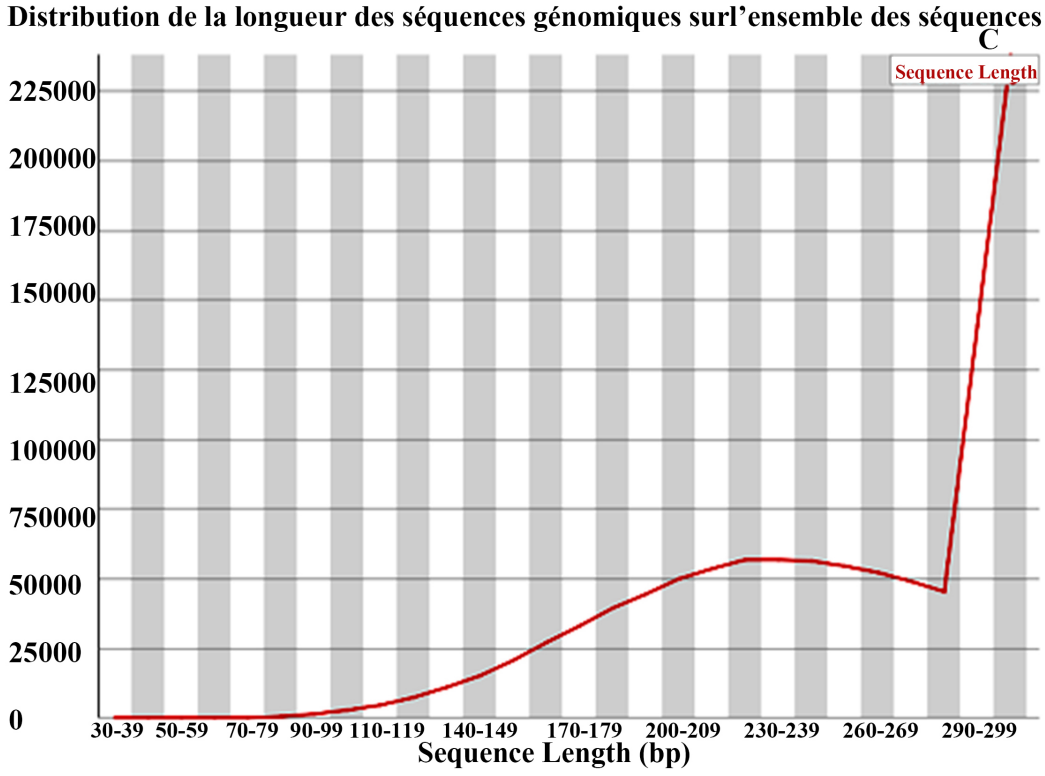
**Figure S1.** Quality control score and genomic read sequences distribution of IBD patients. Panels A, B, C and D refer to genomic read sequences quality control score as well as distribution respectively for IBB1, IBD2, IBD3 and IBD4 inflammatory bowel disease pediatric patients.


Distribution de la longueur des séquences génomiques surl'ensemble des séquences A
Sequence Length
250000
200000
150000
100000
50000
0
30-39 50-59 70-79 90-99 110-119 140-149 170-179 200-209 230-239 260-269 290-299
Sequence Length (bp)

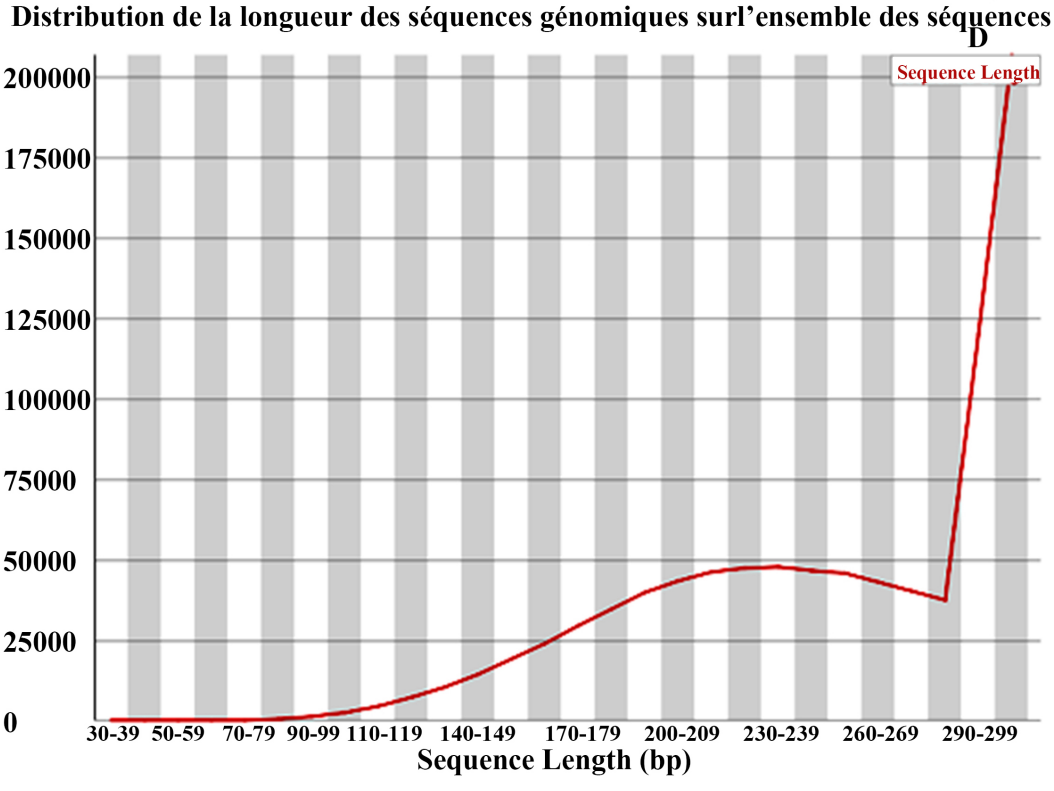
(a)


Distribution de la longueur des séquences génomiques surl'ensemble des séquences B
Sequence Length
300000
250000
200000
150000
100000
50000
0
30-39 50-59 70-79 90-99 110-119 140-149 170-179 200-209 230-239 260-269 290-299
Sequence Length (bp)

(b)




(c)




(d)

**Figure S2.** Assessment of the length of the four (4) IBD patients' clinical exome DNA genomic sequences by FastQ quality control package.

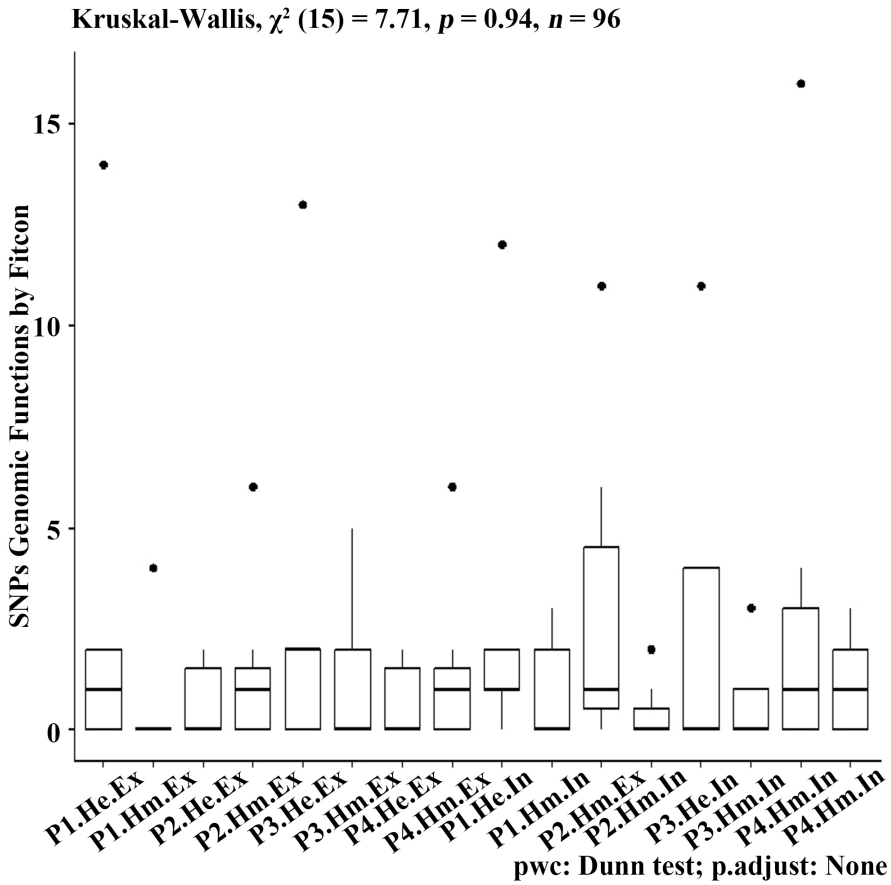



Figure S3. Kruskal-Wallis test evaluating the influence of SNPs genomic functions filtered by *Fitcon* statistical parameter on IBD pediatric population variability.

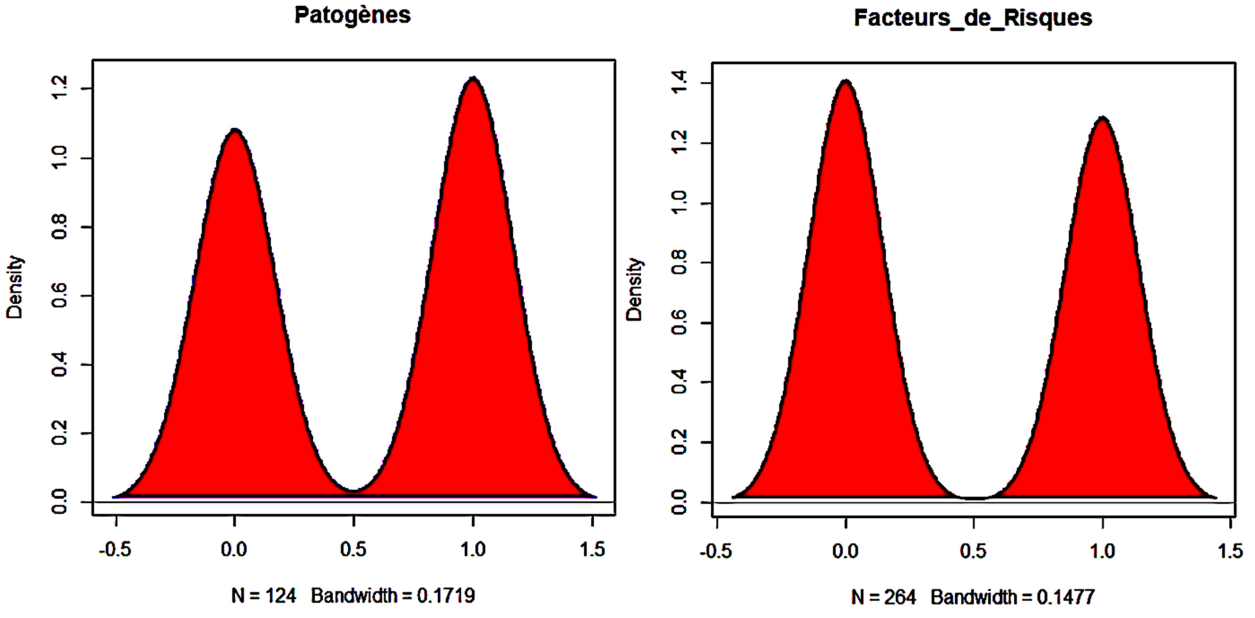



Figure S4. IBD pathogenic and risk factors SNP variants distribution in the four (4) IBD pediatric patients.

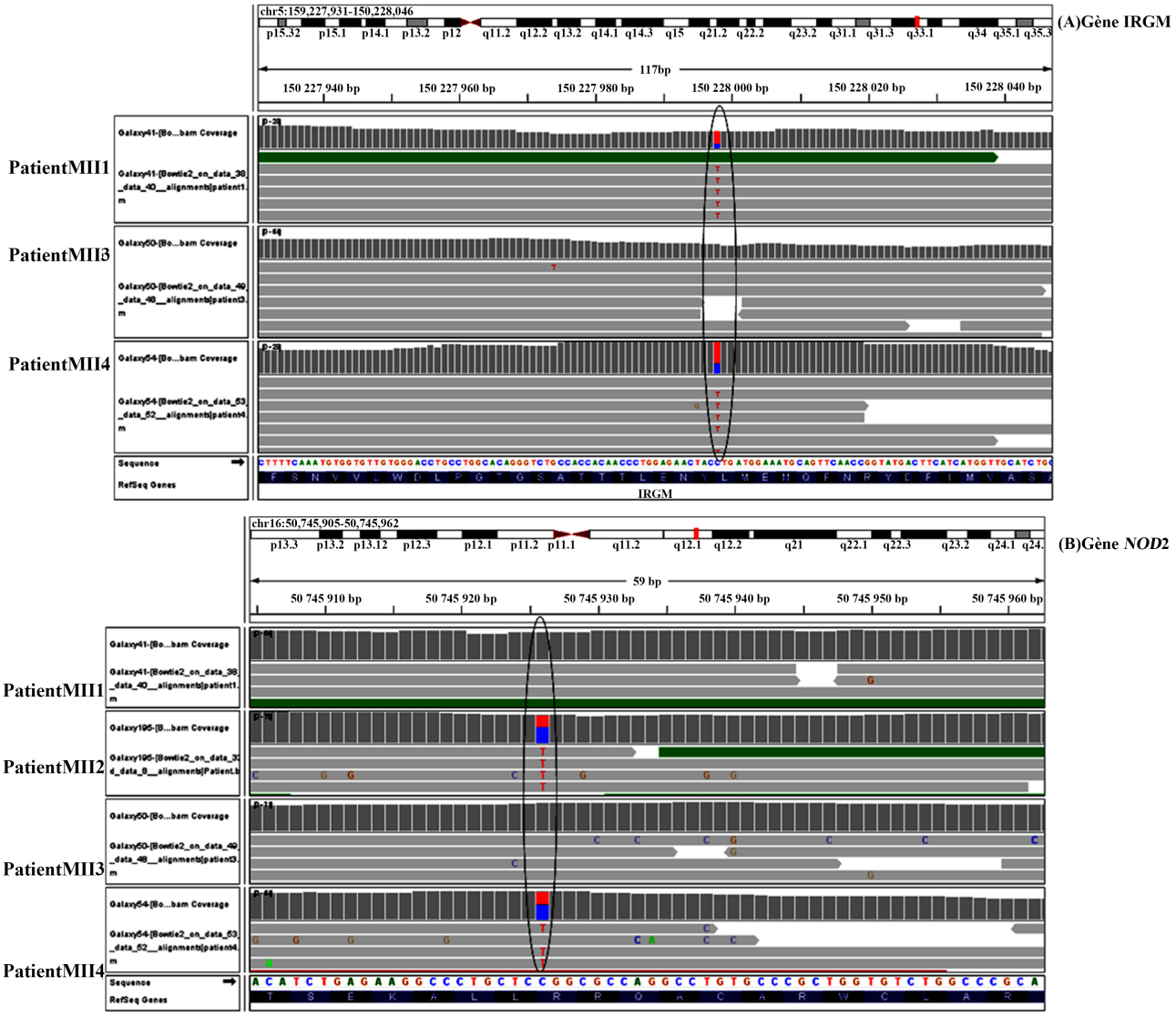



**Figure S5.** IGV illustrating NOD2 and IRGM genes IBD risk factor and pathogenic genetic variants (SNPs) respectively in the four (4) IBD patients population. Patient MII acronyms refers to IBD pediatric patients.

**Table S1.** Shapiro normality test for exonic, intronic and non-coding regions SNP (inducing genomic function changes) distribution in IBD patient's genome.

| | IBD patients exonic SNPs | | | | IBD patients non-coding exonic and intronic SNPs | | | |
|---|---|---|---|---|---|---|---|---|
| | Heterozygous | | Homozygous | | Heterozygous | | Homozygous | |
| **IBD patient 1** | p | 0.03 | p | 0.22 | p | 0.21 | p | 0.007 |
| | w | 0.75 | w | 0.85 | w | 0.88 | w | 0.74 |
| **IBD patient 2** | p | 0.03 | p | 2.2e-16 | p | 0.27 | p | 0.04 |
| | w | 0.80 | w | 0.75 | w | 0.89 | w | 0.81 |
| **IBD patient 3** | p | 0.09 | p | 2.2e-16 | p | 0.10 | p | 0.21 |
| | w | 0.80 | w | 0.63 | w | 0.85 | w | 0.88 |
| **IBD patient 4** | p | 0.27 | p | 0.71 | p | 0.16 | p | 0.003 |
| | w | 0.86 | w | 0.97 | w | 0.87 | w | 0.72 |

**Table S2.** Wilcoxon pairwise comparison test in assessing IBD population variability by analyzing homozygous (Hm), heterozygous (He) intronic (In) and exonic (Ex) IBD risk factors and pathogenic SNPs variants.

| Generated group assessing SNPs intronic and exonic homozygous and heterozygous genomic functions | N1 | N2 | Statistic | p | p adjusted | p adjusted Significance |
|---|---|---|---|---|---|---|
| **IBD P1.He.Ex vs. IBD P1.He.In** | 5 | 8 | 6 | 0.047 | 0.047 | * |
| **IBD P1.He. Ex vs. IBD P1.Hm.In** | 5 | 4 | 12 | 0.701 | 0.701 | ns |
| **IBD P1.He.Ex vs. IBD P1.Hm.In** | 5 | 8 | 8 | 0.09 | 0.09 | ns |
| **IBD P1.He.Ex vs. IBD P2.Hm.In** | 5 | 8 | 6.5 | 0.06 | 0.06 | ns |
| **IBD P1.He.Ex vs. IBD P3.Hm.In** | 5 | 8 | 4.5 | 0.05 | 0.03 | * |
| **IBD P1.He.Ex vs. IBD P1.Hm.In** | 8 | 4 | 28.5 | 0.04 | 0.04 | * |
| **IBD P1.He.In vs. IBD P3.Hm.In** | 8 | 8 | 33.5 | 0.92 | 0.92 | ns |
| **IBD P1.Hm.Ex vs. IBD P2.Hm.Ex** | 4 | 3 | 3 | 0.35 | 0.35 | ns |
| **IBD P1.Hm.Ex vs. IBD P2.Hm.In** | 4 | 8 | 3.5 | 0.041 | 0.041 | * |
| **IBD P1.Hm.Ex vs. IBD P3.He.Ex** | 4 | 5 | 6.5 | 0.45 | 0.45 | ns |
| **IBD P1Hm.Ex vs. IBD P3.Hm.Ex** | 4 | 3 | 3 | 0.35 | 0.35 | ns |
| **IBD P1.Hm.Ex vs.IBD P3.Hm.In** | 4 | 8 | 2 | 0.02 | 0.02 | * |
| **IBD P1.Hm.Ex vs. IBD P4.He.Ex** | 4 | 4 | 6 | 0.64 | 0.64 | ns |
| **IBD P1.Hm.Ex vs. IBD P4.He.In** | 4 | 8 | 3.5 | 0.04 | 0.04 | * |
| **IBD P2.He.Ex vs. IBD P3.He.Ex** | 5 | 5 | 10.5 | 0.75 | 0.75 | ns |
| **IBD P2.He.Ex vs. IBD P3.He.In** | 5 | 8 | 8.5 | 0.11 | 0.11 | ns |
| **IBD P2.He.Ex vs. IBD P3.Hm.Ex** | 5 | 3 | 6 | 0.76 | 0.76 | ns |
| **IBD P2.He.Ex vs. IBD P3.Hm.In** | 5 | 8 | 5 | 0.03 | 0.03 | * |
| **IBD P2.He.Ex vs. IBD P4.He.Ex** | 5 | 4 | 11 | 0.9 | 0.9 | ns |
| **IBD P3.He.Ex vs. IBD P3.Hm.Ex** | 5 | 3 | 7 | 1 | 1 | ns |
| **IBD P3.He.Ex vs. IBD P3.Hm.In** | 5 | 8 | 5.5 | 0.04 | 0.04 | * |
| **IBD P3.He.Ex vs. IBD P4.He.Ex** | 5 | 4 | 11.5 | 0.82 | 0.82 | ns |
| **IBD P3.Hm.Ex vs. IBD P4.Hm.In** | 3 | 8 | 6 | 0.26 | 0.26 | ns |
| **IBD P3.Hm.In vs. IBD P4.He.Ex** | 8 | 4 | 29 | 0.03 | 0.03 | * |
| **IBD P3.Hm.In vs. IBD P4.He.In** | 8 | 8 | 33.5 | 0.91 | 0.91 | ns |
| **IBD P3.Hm.In P4.Hm.Ex** | 8 | 4 | 29.5 | 0.02 | 0.02 | * |

**Table S3.** Summary of retrieved IBD risk factors SNPs variants in the four (4) IBD patients and genomic functions analysis by using ENSEMBL genomic database.

| rs_ID | Chr | Position | Genes | VT | Amino acid change | Phenotypes | MAF | MAF | Patients | Fitcons | I/E |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs34298354 | 1 | 247588053 | NLRP3 | HE | p.Ser436Ser | Chronic infantile neurological, Cutaneous and articular syndrome | 0,14 | 0,07 | 1,4 | 0.55 | E |
| rs5219 | 11 | 17409572 | KCNJ11 | HO | p.Lys23Glu | Body Mass Index, Type 2 diabetes, Blood pressure | 0,49 | 0,26 | 1,2,4 | 0.72 | E |
| rs1169288 | 12 | 121416650 | HNF1A | HE | p.Ile27Leu | Cholesterol, Diabetes MODY3 | 0,30 | 0,47 | 1,3,4 | 0.52 | E |
| rs1044498 | 6 | 132172368 | ENPP1 | HE | p.Lys173Gln | Type 2 diabetes, Obesity | 0,32 | 0,34 | 1,2,3 | 0.71 | E |
| rs237025 | 6 | 149721690 | SUMO4 | HE | p.Val55Met | Type 1 diabetes | 0,49 | 0,35 | 1,2,3,4 | 0.56 | E |
| rs180223 | 8 | 133900252 | TG | HE | p.Ser734Ala | Autoimmune thyroid disease | 0,49 | 0,32 | 1,2,3,4 | 0.55 | E |
| rs853326 | 8 | 133909974 | TG | HE | p.Met1028Val | Autoimmune thyroid disease | 0,49 | 0,33 | 1,2,3,4 | 0.49 | E |
| rs6280 | 3 | 113890815 | DRD3 | HE | p.Gly9Ser | Schizophrenia | 0,50 | 0,49 | 1,2,3,4 | 0.50 | E |
| rs1801968 | 9 | 132580901 | TOR1A | HE | p.Asp216His | Dystonia | 0,21 | 0,08 | 1 | 0.71 | E |
| rs1805097 | 13 | 110435231 | IRS2 | HE | p.Gly1057Asp | Type 2 diabetes | 0,48 | 0,28 | 1 | 0.65 | E |
| rs699 | 1 | 230845794 | AGT | HO | p.Met268Thr | Coronary artery disease, Renal dysplasia, High blood pressure | 0,5 | 0,29 | 1,3,4 | 0.49 | E |
| rs4833095 | 4 | 38799710 | TLR1 | HE | p.Asn248Ser | Asthma, Leprosy, Seasonal allergic rhinitis | 0,50 | 0,43 | 1,3 | 0.72 | E |
| rs1800858 | 10 | 43595968 | RET | HE | p.Ala45Ala | Neoplasm, Aganglionic mega-colon | 0,5 | 0,25 | 1,2,3,4 | 0.71 | E |
| rs2227564 | 10 | 75673101 | PLAU | HO | p.Leu141Pro | Crohn's disease, Inflammatory bowel disease | 0,5 | 0,22 | 1,3,4 | 0.67 | E |
| rs1051730 | 15 | 78894339 | CHRNA3 | HE | p.Tyr215Tyr | Lung adenocarcinoma; Lung cancer in smokers | 0,48 | 0,17 | 1,2,3,4 | 0.66 | E |
| rs1799983 | 7 | 150696111 | NOS3 | HO | p.Asp298Glu | Metabolic syndrome | 0,4 | 0,18 | 1,2,3,4 | 0.71 | E |
| rs16969968 | 15 | 78882925 | CHRNA5 | HE | p.Asp398Asn | Forced expiratory volume, lung cancer | 0,47 | 0,15 | 1,2,3 | 0.73 | E |
| rs2273535 | 20 | 54961541 | AURKA | HE | p.Phe31Ile | Colon cancer | 0,42 | 0,31 | 1,2,3,4 | 0.71 | E |
| rs1042713 | 5 | 148206440 | ADRB2 | HO | p.Gly16Arg | Asthma | 0,49 | 0,48 | 1,2,3,4 | 0.44 | E |

**Continued**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs2274700 | 1 | 196682947 | CFH | HE | p. Ala473Ala | Basal laminar drusen, Hemolytic-uremic syndrome, Age-related macular degeneration, Blood protein levels | 0,49 | 0,48 | 1 | 0.69 | E |
| rs2230199 | 19 | 6718387 | C3 | HE | p.Arg102Gly | Wet macular degeneration, Age-related macular degeneration | 0,28 | 0,09 | 1 | 0.61 | E |
| rs429358 | 19 | 45411941 | APOE | HE | p. Cys156Arg | Alzheimer's disease, Type 2 diabetes, Age-related macular degeneration, Kidney failure, Low-density lipoprotein cholesterol measurement | 0,38 | 0,15 | 1 | 0.63 | E |
| rs2066844 | 16 | 50745926 | NOD2 | HE | p.Arg702Trp | Asthma, Crohn's disease, Mouth ulcer, Large intestine inflammation, Colorectal adenoma. | 0,16 | 0,01 | 2,4 | 0.67 | E |
| rs231775 | 2 | 204732714 | CTLA4 | HE | p.Thr17Ala | Celiac disease, Alopecia, Systemic lupus erythematosus, Type I diabetes mellitus, Hashimoto's thyroiditis, Autoimmune thyroid disease | 0,48 | 0,43 | 2 | 0.49 | E |
| rs1061170 | 1 | 196659237 | CFH | HE | p.His402Tyr | Age-related macular degeneration, Atypical hemolytic-uremic syndrome, Blood protein levels | 0,49 | 0,27 | 1,2,4 | 0.71 | E |
| rs20541 | 5 | 131995964 | IL13 | HO | p.Gln144Arg | allergy, Asthma, Serum IgE measurement, Body height | 0,48 | 0,27 | 2 | 0.43 | E |
| rs116454156 | 10 | 95347041 | FFAR4 | HE | p.Arg270His | body mass index, Locus trait quantitative 10 (BMIQ10) | 0,13 | 0,01 | 2 | 0.49 | E |
| rs861539 | 14 | 104165753 | XRCC3 | HE | p.Thr241Met | Cutaneous melanoma | 0,45 | 0,22 | 1,2,3 | 0.67 | E |
| rs3743205 | 15 | 55790530 | DYX1C1 | HE | 0 | Dyslexia | 0,31 | 0,1 | 2 | 0 | I |

**Continued**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs1048661 | 15 | 74219546 | LOXL1 | HE | p.Arg141Leu | Exfoliation syndrome | 0,5 | 0,31 | 2 | 0.79 | E |
| rs6180 | 5 | 42719239 | GHR | HO | p.Ile551Leu | Familial hypercholesterolemia, Body height | 0,49 | 0,44 | 2,3,4 | 0.55 | E |
| rs1801131 | 1 | 11854476 | MTHFR | HE | p.Glu429Ala | Gastrointestinal tumor, Neoplasm of rectum and colon, Dysphagia, Rash, Vomiting, Large hands, Constipation | 0,47 | 0,25 | 2,3 | 0.73 | E |
| rs10993994 | 10 | 51549496 | TIMM23B | HE | 0 | Prostate cancer | 0 ,5 | 0,48 | 2 | 0.05 | I |
| rs1799945 | 6 | 26091179 | HFE | HE | p.His63Asp | Blood pressure, Pancreatic abnormalities, Microvascular complications of diabetes, Alzheimer's disease, Male infertility | 0,25 | 0,07 | 2,3 | 0.62 | E |
| rs4880 | 6 | 160113872 | SOD2 | HE | p.Val16Ala | Microvascular complications of diabetes | 0,50 | 0,41 | 2 | 0.44 | E |
| rs1800470 | 19 | 41858921 | TGFB1 | HE | p. Pro10Leu | Breast carcinoma, Pancreatic anomaly, Male infertility, | 0,5 | 0,45 | 3,4 | 0.56 | E |
| rs13266634 | 8 | 118184783 | SLC30A8 | HE | p.Arg325Trp | Type II and I diabetes | 0,49 | 0,26 | 2 | 0.45 | E |
| rs2073711 | 15 | 65494212 | CILP | HE | p.Ile395Thr | Lumbar disc disease | 0,42 | 0,50 | 2 | 0.52 | E |
| rs2236225 | 14 | 64908845 | MTHFD1 | HE | p.Arg653Gln | Neural tube defects | 0,5 | 0,34 | 2,3 | 0.71 | E |
| rs7076156 | 10 | 64415184 | ZNF365 | HO | p.Thr62Ala | Crohn's, Uric acid nephrolithiasis | 0,32 | 0,13 | 2,3,4 | 0.49 | E |
| rs2241880 | 2 | 234183368 | ATG16L1 | HE | p.Thr300Ala | Crohn's disease, Inflammatory bowel disease | 0 | 0,40 | 1,2,3,4 | 0.71 | E |
| rs5918 | 17 | 45360730 | ITGB3 | HE | p.Leu59Pro | Gastrointestinal hemorrhage, Myocardial infarction, Intracranial hemorrhage | 0,16 | 0,09 | 3 | 0.71 | E |
| rs2072493 | 1 | 223284599 | TLR5 | HE | p.Asn592Ser | Migraine, Anorexia, vomiting, Arthralgia, Diarrhea, Abdominal pain | 0,36 | 0,14 | 3 | 0.52 | E |
| rs1805008 | 16 | 89986144 | MC1R | HE | p.Arg160Trp | Cutaneous melanoma | 0,11 | 0,01 | 3 | 0.76 | E |

**Continued**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs17580 | 14 | 94847262 | SERPINA1 | HE | p.Glu288Val | Low-density lipoprotein cholesterol levels, BMI-adjusted sex hormone globulin levels | 0 | 0,02 | 3,4 | 0.55 | E |
| rs9344 | 11 | 69462910 | CCND1 | HE | p.Pro241Pro | Colorectal cancer, Multiple myeloma | 0,50 | 0,41 | 3 | 0.67 | E |
| rs12150220 | 17 | 5485367 | NLRP1 | HO | p.Leu155His | Vitiligo-associated multiple autoimmune disease | 0 ,47 | 0,19 | 4 | 0.73 | E |
| rs324981 | 7 | 34818113 | NPSR1 | HE | p. Asn107Ile | Asthma-related traits | 0,49 | 0,47 | 4 | 0.49 | E |
| rs1051931 | 6 | 46672943 | PLA2G7 | HO | p. Val379Ala | Asthma and Atopy | 0,39 | 0,19 | 1 | 0.62 | E |
| rs1800470 | 19 | 41858921 | TGFB1 | HE | p. Pro10Leu | NS | 0,5 | 0,45 | 4,3 | 0.56 | E |
| rs34911341 | 3 | 10331519 | GHRL | HE | p. Arg51Gln | Obesity; Metabolic syndrome | 0,04 | 0,01 | 4 | 0.53 | E |
| rs7080536 | 10 | 115348046 | HABP2 | HE | p. Gly534Glu | Factor VII Marburg I variable thrombophilia; Non-medullary thyroid cancer | 0,05 | 0,01 | 4 | 0.49 | E |
| rs41307846 | 1 | 1959699 | GABRD | HE | p. Arg220His | Generalized epilepsy | 0,05 | 0,01 | 4 | 0.65 | E |
| rs1053874 | 16 | 3707747 | DNASE1 | HE | p.Arg244Gln | Body mass index, systemic lupus erythematosus | 0,47 | 0,49 | 4 | 0.67 | E |
| rs486907 | 1 | 182554557 | RNASEL | HE | p.Arg462Gln | Prostate cancer | 0,40 | 0,23 | 4 | 0.65 | E |
| rs3732378 | 3 | 39307162 | CX3CR1 | HE | p.Thr312Met | Age-related macular degeneration, Coronary artery disease, HIV infection | 0,23 | 0,09 | 4 | 0.50 | E |
| rs3732379 | 3 | 39307256 | CX3CR1 | HE | p.Val281Ile | Coronary artery disease, HIV infection, Age-related macular degeneration | 0,32 | 0,14 | 4 | 0.50 | E |
| rs6050 | 4 | 155507590 | FGA | HE | p.Thr331Ala | Venous thromboembolism, fibrinogen measurement | 0,48 | 0,33 | 3 | 0.59 | E |
| rs1131454 | 12 | 113348870 | OAS1 | HO | p.Gly162Ser | Type 1 diabetes | 0,49 | 0,47 | 1,3 | 0.71 | E |
| rs10490924 | 10 | 124214448 | ARMS2 | HE | p.Ala69Ser | Wet macular degeneration, age-related macular degeneration | 0,43 | 0,29 | 2 | 0.50 | E |

**Continued**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| rs6003 | 1 | 197031021 | F13B | HO | p.Arg115His | Age-related macular degeneration with neovascularization; Age-related macular degeneration with Geographic Atrophy | 0,46 | 0,24 | 3 | 0.49 | E |
| rs35719940 | 5 | 1254594 | TERT | HE | p.Ala1062Thr | Dyskeratosis congenital, Leukemia, Idiopathic pulmonary fibrosis, Breast carcinoma | 0,03 | 0,01 | 2 | 0.70 | E |
| rs3135506 | 11 | 116662407 | APOA5 | HE | p.Ser19Trp | Hypertriglyceridemia | 0,21 | 0,06 | 3 | 0.62 | E |
| rs662 | 7 | 94937446 | PON1 | HE | p.Gln192Arg | Coronary artery disease | 0,50 | 0,46 | 3 | 0.49 | E |
| rs2904552 | 22 | 18905964 | PRODH | HE | p.Arg431His | Autosomal dominant inheritance, Hyperprolinemia type 1, Metabolism anomaly | 0,5 | 0,04 | 4 | 0.68 | E |
| rs450046 | 22 | 18901003 | PRODH | HO | p.Arg521Gln | Autosomal dominant inheritance, Hyperprolinemia type 1, Schizophrenia | 0,26 | 0,09 | 4,1,2,3 | 0.63 | E |

rs_ids = ID of single nucleotide polymorphisms, Chr = Chromosome, Variant Type = VT, Minor Allele Frequency = MAF, Highest Population Allele Frequency = MAF, I/E = Intonc or Exonic.

**Table S4.** Pathogenic genetic variants.

| rs_ID | Chr | Positions | Genes | VT | Amino acid change | Phenotypes | MAF | MAF | Patients | Fitcon | I/E |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs10010131 | 4 | 6292915 | WFS1 | HE | 0 | Type 2 diabetes | 0,43 | 0,27 | 1,3,4 | 0.09 | I |
| rs1137617 | 7 | 150648198 | KCNH2 | HE | p.Tyr652Tyr | Prolonged QT interval, Sudden cardiac death, Cardiovascular system abnormalities | 0,44 | 0,23 | 1,2,3,4 | 0.72 | E |
| rs1136743 | 11 | 18290859 | SAA1 | HO | p.Ala70Val | Amyloid serum variant | 0,5 | 0 | 1 | 0.55 | E |
| rs1805010 | 16 | 27356203 | IL4R | HE | p.Ile75Val | AIDS, Atopy | 0,5 | 0,46 | 1,2,4 | 0.71 | E |
| rs351855 | 5 | 176520243 | FGFR4 | HO | p.Gly388Arg | Cancer progression and tumor cell motility; Respiratory failure, Body mass index, Migraine | 0,49 | 0,30 | 1,3,4 | 0.70 | E |
| rs1052030 | 11 | 76853783 | MYO7A | HE | p.Leu16Ser | Hearing impairment, Usher syndrome | 0,5 | 0,49 | 4,1 | 0 | E |

**Continued**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs1169305 | 12 | 121437382 | HNF1A | HO | p.Ser581Gly | Maturity-onset diabetes in young people, MODY | 0,09 | 0,01 | 4,1,2,3 | 0.43 | E |
| rs1042522 | 17 | 7579472 | TP53 | HE | p.Pro72Arg | Pancreatic neoplasm, Colon cancer, | 0,5 | 0,46 | 4,1,3,2 | 0.72 | E |
| rs1799983 | 7 | 150696111 | NOS3 | HO | p.Asp298Glu | Metabolic syndrome | 0,4 | 0,18 | 1,2,3,4 | 0.71 | E |
| rs2228671 | 19 | 11210912 | LDLR | HO | p.Cys27Cys | LDL cholesterol, Total cholesterol measurement, Familial hypercholesterolemia | 0,14 | 0,06 | 1,3 | 0.65 | E |
| rs1801133 | 1 | 11856378 | MTHFR | HE | p.Ala222Val | High LDL cholesterol, Total cholesterol measurement, Familial hypercholesterolemia | 0,5 | 0,25 | 1,2 | 0.72 | E |
| rs429358 | 19 | 45411941 | APOE | HE | p.Cys156Arg | Alzheimer's disease, Type 2 diabetes; Age-related macular degeneration Renal failure, LDL cholesterol | 0,38 | 0,15 | 1 | 0.63 | E |
| rs10065172 | 5 | 150227998 | IRGM | HE | p.Leu105Leu | Inflammatory bowel disease | 0,5 | 0,30 | 4,2,1 | 0.55 | E |
| rs1061170 | 1 | 196659237 | CFH | HE | p.His402Tyr | Age-related macular degeneration. Atypical hemolytic-uremic syndrome, Blood protein levels | 0,49 | 0,27 | 1,2,4 | 0.71 | E |
| rs1064039 | 20 | 23618427 | CST3 | HE | p.Ala25Thr | Age-related macular degeneration | 0 ,38 | 0,21 | 3 | 0.73 | E |
| rs12406197 | 1 | 179545050 | NPHS2 | HE | 0 | Idiopathic nephrotic syndrome | 0 ,37 | 0 ,18 | 2 | 0.20 | I |
| rs820878 | 5 | 73981270 | HEXB | HO | p.Leu62Ser | Sandhoff's disease | 0,06 | 0,02 | 4,1,3,2 | 0.73 | E |
| rs2229992 | 5 | 112162854 | APC | HO | p.Tyr486Tyr | Hereditary cancer predisposition syndrome; Familial colorectal cancer | 0,49 | 0,49 | 2 | 0.71 | E |
| rs1799945 | 6 | 26091179 | HFE | HE | p.His63Asp | High blood pressure. Pancreatic abnormalities. Microvascular complications of diabetes Alzheimer's disease | 0,25 | 0,07 | 2,3 | 0.62 | E |
| rs17261572 | X | 119760629 | C1GALT1C1 | HO | p.Asp131Glu | Erythrocyte polyagglutination syndrome | 0,33 | 0,15 | 2 | 0 | E |

**Continued**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| rs41265017 | 1 | 156146640 | SEMA4A | HE | p.Arg713Gln | Penis hypoplasia, Diabetes mellitus type II, Retinal vascular anomaly, Skin pigmentation anomaly ,Macular degeneration, Blindness | 0,12 | 0,02 | 2,4 | 0.68 | E |
| rs11539444 | 12 | 32908518 | YARS2 | HE | p.Gly97Gly | Myopathy, Respiratory failure, Failure to thrive, Muscular weakness | 0,28 | 0,13 | 4,2 | 0.44 | E |
| rs1801131 | 1 | 11854476 | MTHFR | HE | p.Glu429Ala | Gastrointestinal tumor Neoplasm of rectum and colon, Dysphagia, Rash vomiting, Large hands, Constipation, Intestinal obstruction. | 0,47 | 0,25 | 2,3 | 0.73 | E |
| rs17580 | 14 | 94847262 | SERPINA1 | HE | p.Glu288Val | Low-density lipoprotein cholesterol levels, BMI-adjusted sex hormone globulin levels | 0 | 0,02 | 3,4 | 0.55 | E |
| rs2904552 | 22 | 18905964 | PRODH | HE | p.Arg431His | Autosomal dominant inheritance, Autistic behavior, Schizophrenia | 0,5 | 0,04 | 4 | 0.67 | E |
| rs199476099 | 16 | 2168022 | PKD1 | HE | p.Arg324Leu | Muscular weakness | 0,01 | 0,01 | 4 | 0.65 | E |
| rs3732378 | 3 | 39307162 | CX3CR1 | HE | p.Thr312Met | Age-related macular degeneration, Coronary heart disease, HIV infection | 0,23 | 0,09 | 4 | 0.50 | E |
| rs3732379 | 3 | 39307256 | CX3CR1 | HE | p.Val281Ile | Age-related macular degeneration, Coronary heart disease, HIV infection | 0,32 | 0,14 | 4 | 0.50 | E |
| rs450046 | 22 | 18901003 | PRODH | HO | p.Arg521Gln | Autosomal dominant transmission, schizophrenia | 0,26 | 0,09 | 4,1,2,3 | 0.63 | E |
| rs10509305 | 10 | 70645376 | STOX1 | HE | p.Glu608Asp | Preeclampsia | 0,31 | 0,14 | 2 | 0.62 | E |
| rs1805008 | 16 | 89986144 | MC1R | HE | p.Arg160Trp | Cutaneous melanoma | 0,11 | 0,01 | 3 | 0.76 | E |

rs_ids = ID of single nucleotide polymorphisms, Chr = Chromosome, Variant Type = VT, Minor Allele Frequency = MAF, Highest Population Allele Frequency = MAF, I/E = Intonc or Exonic.