

1 **Mutation spectrum of *NOD2* reveals recessive inheritance as a main driver of**
2 **Early Onset Crohn's Disease**

3

4 Julie E. Horowitz¹, Neil Warner^{2,3}, Jeffrey Staples¹, Eileen Crowley^{2,3}, Ryan Murchie^{2,3},
5 Cristopher Van Hout¹, Alejandra Klauer King¹, Karoline Fiedler^{2,3}, Jeffrey G. Reid¹, John D.
6 Overton¹, Alan R. Shuldiner¹, Aris Baras¹, Geisinger-Regeneron DiscovEHR Collaboration,
7 Frederick E. Dewey¹, Anne Griffiths², Omri Gottesman¹, Aleixo M. Muise^{2,3,4,*}, Claudia
8 Gonzaga-Jauregui^{1,*}

9

10 ¹ Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA

11 ² SickKids Inflammatory Bowel Disease Center, Hospital for Sick Children, Toronto, ON,
12 Canada

13 ³ Cell Biology Program, Research Institute, Hospital for Sick Children, Toronto, ON, Canada.

14 ⁴ Departments of Pediatrics and Biochemistry, University of Toronto, Toronto, ON, Canada.

15 *Co-corresponding authors

16

17 Correspondence should be addressed to:

18 **Aleixo Muise MD, Ph.D.**

19 555 University Ave.

20 The Hospital for Sick Children

21 Toronto, ON, Canada, M5G 1X8

22 Email: aleixo.muise@utoronto.ca

23 Phone: 416-813-7735

24 Fax: 416-813-6531

25

26

27 **Claudia Gonzaga-Jauregui, Ph.D.**

28 Regeneron Genetics Center, Regeneron Pharmaceuticals Inc.

29 777 Saw Mill River Rd.

30 Tarrytown, NY, USA

31 Email: clau.gonzagajauregui@regeneron.com

32 Phone: 914-847-3451

33

34 **Abstract:**

35 Inflammatory bowel disease (IBD), clinically defined as Crohn's disease (CD), ulcerative colitis
36 (UC), or IBD-unclassified, results in chronic inflammation of the gastrointestinal tract in
37 genetically susceptible hosts. Pediatric onset IBD represents $\geq 25\%$ of all IBD diagnoses and
38 often presents with intestinal stricturing, perianal disease, and failed response to conventional
39 treatments. *NOD2* was the first and is the most replicated locus associated with adult IBD, to
40 date. To determine the role of *NOD2* and other genes in pediatric IBD, we performed whole-
41 exome sequencing on a cohort of 1,183 patients with pediatric onset IBD (ages 0-18.5 years).
42 We identified 92 probands who were homozygous or compound heterozygous for rare and low
43 frequency *NOD2* variants accounting for approximately 8% of our cohort, suggesting a
44 Mendelian recessive inheritance pattern of disease. Additionally, we investigated the
45 contribution of recessive inheritance of *NOD2* alleles in adult IBD patients from the Regeneron
46 Genetics Center (RGC)-Geisinger Health System DiscovEHR study, which links whole exome
47 sequences to longitudinal electronic health records (EHRs) from 51,289 participants. We found
48 that ~7% of cases in this adult IBD cohort, including ~10% of CD cases, can be attributed to
49 recessive inheritance of *NOD2* variants, confirming the observations from our pediatric IBD
50 cohort. Exploration of EHR data showed that 14% of these adult IBD patients obtained their
51 initial IBD diagnosis before 18 years of age, consistent with early onset disease. Collectively,
52 our findings show that recessive inheritance of rare and low frequency deleterious *NOD2*
53 variants account for 7-10% of CD cases and implicate *NOD2* as a Mendelian disease gene for
54 early onset Crohn's Disease.

55

56 **Author Summary:**

57
58 Pediatric onset inflammatory bowel disease (IBD) represents $\geq 25\%$ of IBD diagnoses; yet the
59 genetic architecture of early onset IBD remains largely uncharacterized. To investigate this, we
60 performed whole-exome sequencing and rare variant analysis on a cohort of 1,183 pediatric
61 onset IBD patients. We found that 8% of patients in our cohort were homozygous or compound
62 heterozygous for rare or low frequency deleterious variants in the nucleotide binding and
63 oligomerization domain containing 2 (*NOD2*) gene. Further investigation of whole-exome
64 sequencing of a large clinical cohort of adult IBD patients uncovered recessive inheritance of
65 rare and low frequency *NOD2* variants in 7% of cases and that the relative risk for *NOD2* variant
66 homozygosity has likely been underestimated. While it has been reported that having >1 *NOD2*
67 risk alleles is associated with increased susceptibility to Crohn's Disease (CD), our data formally
68 demonstrate what has long been suspected: recessive inheritance of *NOD2* alleles is a
69 mechanistic driver of early onset IBD, specifically CD, likely due to loss of *NOD2* protein
70 function. Our data suggest that a subset of IBD-CD patients with early disease onset is
71 characterized by recessive inheritance of *NOD2* alleles, which has important implications for the
72 screening, diagnosis, and treatment of IBD.

73

74

75

76

77 **Introduction:**

78
79 Inflammatory bowel disease (IBD) is a chronic inflammatory condition of the
80 gastrointestinal (GI) tract that arises as part of an inappropriate response to commensal or
81 pathogenic microbiota in a genetically susceptible individual (1-4). IBD encompasses Crohn's
82 Disease (CD); ulcerative colitis (UC); and IBD unclassified (IBDU). The etiology of IBD is
83 complex and has been attributed to defects in a number of cellular pathways including pathogen
84 sensing, autophagy, maintenance of immune homeostasis, and intestinal barrier function,
85 among other processes (3-18).

86 Great effort has been invested into defining the genetic factors that confer IBD
87 susceptibility. To date, >200 unique loci have been associated with IBD through genome-wide
88 association studies (GWAS), primarily in adult populations (19, 20). Nearly all the identified
89 susceptibility loci exhibit low effect sizes (ORs ~1.0-1.5) individually (19), and collectively
90 account for less than 20% of the heritable risk for IBD (19, 21). These observations support a
91 complex disease model in which common variants of modest effect sizes interact with
92 environmental factors including diet, smoking, and the intestinal microbiome (22, 23) to give rise
93 to IBD susceptibility (4).

94 The earliest and most replicated genetic associations with IBD (24-26) correspond to a
95 locus on chromosome 16 that encompasses the nucleotide-binding and oligomerization domain-
96 containing 2 (*NOD2*) gene, with an average allelic odds ratio across multiple studies of 3.1 (19,
97 20). *NOD2* encodes an intracellular microbial sensor that recognizes muramyl dipeptide (MDP)
98 motifs found on bacterial peptidoglycans (27, 28). Upon activation, *NOD2* protein signals
99 through the NF- κ B family of proteins (29) to modulate transcription of genes encoding pro-
100 inflammatory cytokines IL-8, TNF- α , and IL-1 β (30-32), among others. Variation in *NOD2*
101 accounts for approximately 20% of the genetic risk among CD cases, with three variants -
102 p.R702W (ExAC MAF= 0.0227 across all populations), p.G908R (ExAC MAF= 0.0099), and

103 p.L1007fs (ExAC MAF= 0.0131) - accounting for over 80% of the disease-causing mutations in
104 *NOD2* associated with adult CD (33), albeit not with UC; and particularly ileal versus colonic CD
105 (34). These three “common” risk variants, typically observed in a heterozygous state, are
106 predicted loss-of-function alleles that impair NF- κ B activation in response to MDP ligands, *in*
107 *vitro* (28, 35-37).

108 With the assumption that genetic risk has a disproportionate effect over environmental
109 risk in early onset disease, recent studies have focused on pediatric IBD cases (diagnosed
110 <18y) (38). Pediatric IBD patients comprise 20-25% of all IBD cases and are typically more
111 clinically severe than adult-onset patients, often exhibiting disease of the upper GI tract, small
112 bowel inflammation, and perianal disease as well as failure to thrive and poor clinical response
113 (4, 39). Results from GWAS conducted in this group of severely affected patients indicate that
114 associated loci in early onset IBD significantly overlap with adult IBD loci, including both the
115 *NOD2* locus and an additional 28 CD-specific loci previously implicated in adult-onset IBD (40-
116 42). As the mechanism for these “common” IBD susceptibility loci in the pathogenesis of early
117 onset IBD remains unclear (43), we performed whole-exome sequencing and rare variant
118 analysis on a cohort of 1,183 pediatric onset IBD patients to elucidate the role of rare protein
119 coding variation in IBD-associated genes, specifically *NOD2*, in this disease.

120

121 **Results and Discussion:**

122 We performed trio-based analysis of 492 complete trios using a proband-based
123 analytical pipeline to identify all recessive (compound heterozygous and homozygous) and *de*
124 *novo* variants of interest in the affected probands. In our initial analyses, we identified 10
125 families with recessive (compound heterozygous or homozygous), rare variants ($2\% \leq \text{MAF}$) in
126 *NOD2*, all with a diagnosis of CD. We observed that some of the rare variants in these probands
127 were inherited in *trans* from previously-reported CD risk alleles, mainly the p.G908R missense

128 variant. We identified two individuals who are compound heterozygous for the p.G908R risk
129 allele in *trans* with a less common *NOD2* CD risk variant (p.N852S) in one case and a novel
130 truncating indel (p.S506Vfs*73) in the second case (Table S1, Fam008 and Fam009). The
131 observation of a CD-associated *NOD2* risk allele in *trans* from other rare or novel alleles led us
132 to survey the rest of the probands, including singletons and those part of incomplete trios, for
133 recessive inheritance, either in a homozygous or compound heterozygous manner, of *NOD2*
134 variants, but expanding our allelic range to low-frequency variants ($2\% \leq \text{MAF} \leq 5\%$). Through this
135 approach we identified 108 probands with putative recessive *NOD2* variants. Visual inspection
136 of sequence reads and orthogonal confirmation through Sanger sequencing excluded 13
137 probands with variants inherited in *cis* from an unaffected parent or heterozygous variants that
138 were initially called as homozygous due to low coverage of the region and skewed allelic
139 balance. Of note, we identified 5 probands carrying p.L1007fs and p.M863V risk variants, 4 of
140 which were confirmed to occur in *cis* and were inherited from an unaffected parent. The
141 remaining case with p.L1007fs and p.M863V was a singleton and thus phase could not be
142 determined. These two variants segregate in *cis* within the same haplotype, as confirmed by
143 segregation within the trios and as previously observed (44). Therefore, we excluded these 4
144 probands from our final count of recessively-inherited *NOD2* variants. Similarly, we identified 3
145 probands from 3 complete trios segregating the p.S431L and p.V793M reported risk variants in
146 *cis* inherited from an unaffected carrier parent; these probands were also excluded. Three
147 additional probands were excluded on the basis of a re-evaluation of the phenotype that
148 excluded a clinical diagnosis of IBD.

149 Thus, we identified 92 probands with confirmed recessive *NOD2* variants within our
150 pediatric onset IBD cohort. These include: 25 probands carrying homozygous variants, 41
151 probands with confirmed compound heterozygous variants, and an additional 26 singleton
152 probands with putative compound heterozygous variants where phasing could not be performed
153 (Table S1, Figure S1). The majority of the compound heterozygous individuals (65/67) carry a

154 known *NOD2* CD-risk allele in addition to either another known *NOD2* CD-risk allele or a novel
155 *NOD2* variant, including some truncating loss-of-function variants supporting loss or impaired
156 function of *NOD2* in the pathophysiology of CD (6). In total, 92 of 1,183 (7.8%) of the probands
157 in our pediatric onset IBD cohort conformed to a recessive, Mendelian inheritance mode for
158 *NOD2* rare and low frequency ($MAF \leq 5\%$) deleterious variants (Table 1, Figure 1, Table S1, and
159 Table S4).

160 The 92 pediatric patients homozygous for *NOD2* mutations were predominantly male
161 (71%) with a median age at diagnosis of 12.5 years (Table S1). At diagnosis, 83% displayed
162 diagnostic features of Crohn's disease. 23% of the cohort displayed a constellation of extra-
163 intestinal manifestations, mainly large joint arthritis, chronic recurrent multifocal osteomyelitis,
164 recurrent fevers, erythema nodosum, and pyoderma gangrenosum. Only 6% of the cohort
165 showed significant perianal disease (namely fistulae and abscesses, skin tags and fissures
166 were not considered as perianal disease) (Table S1). Per the Montreal classification of IBD (45),
167 44% of the overall cohort of patients presented with ileal disease at diagnosis (L1). 25%
168 presented with ileocolonic disease (L3) and 10% displayed features of colonic inflammation only
169 (L2). Isolated upper disease was only present in 2% of the cohort (L4). We observed a
170 progression of ileal disease in 21% (18.5% stricturing; 2.5% penetrating) with 21% requiring a
171 resection. On review of the Crohn's disease patients only, 50% displayed L1 disease (terminal
172 ileal +/- limited cecal disease), 32.9% L3 (ileocolonic) disease; 86.8% B1 (non-stricturing, non-
173 penetrating) disease, 10.5% B2 (stricturing) disease (Table S1).

174 Given the substantial contribution of recessive *NOD2* variants to CD in our pediatric
175 onset IBD cohort and the known contribution of *NOD2* to adult CD, we next investigated the
176 contribution of *NOD2* recessivity in a large clinical population. For this, we examined a cohort of
177 adult IBD patients from the RGC - Geisinger Health System (GHS) DiscovEHR study (46) . A
178 key feature of the DiscovEHR study is the ability to link genomic sequence data to de-identified
179 electronic health records (EHRs). Within this cohort, we identified 984 patients (of 51,289 total

180 sequenced DiscovEHR patient-participants) with a diagnosis of IBD, defined as having a
181 problem list entry or an encounter diagnosis entered for two separate clinical encounters on
182 separate calendar days for the ICD-9 codes 555* (Regional enteritis) or 556* (Ulcerative
183 enterocolitis). For our analysis, we surveyed all instances of homozygous *NOD2* rare and low
184 frequency variants ($MAF \leq 5\%$); the same parameters applied to our pediatric IBD probands.
185 Among patients with an IBD diagnosis, we identified 18 individuals who are either homozygous
186 for the p.R702W risk allele (N=10) or homozygous for the p.L1007fs allele (N=8) (Table 2,
187 Figure S2). We did not identify any p.G908R homozygous individuals with an IBD diagnosis in
188 this cohort. Next, we looked for instances of putative compound heterozygosity among these
189 adult IBD DiscovEHR patients. To investigate this, we searched for occurrences of two or more
190 of the three most prevalent *NOD2* risk alleles (p.R702W, p.G908R, or p.L1007fs) in these
191 individuals. We identified putative compound heterozygosity for the three main CD risk alleles,
192 p.R702W/p.G908R (N=6), p.G908R/p.L1007fs (N=5), and p.R702W/p.L1007fs (N=11) (Table
193 2). We also observed instances of putative compound heterozygosity for each of the three main
194 CD risk alleles along with either a rarer CD risk allele or a novel allele or two rare alleles in *trans*
195 (N=24), parallel to the findings in our pediatric IBD cohort. Using familial relationships and
196 pedigree reconstruction (47), we were able to confirm appropriate segregation for 32 of the 64
197 DiscovEHR recessive *NOD2* variant carriers with IBD, including *trans* inheritance in 13 putative
198 compound heterozygotes (Figure S2). The other 32 were singleton cases where phase could
199 not be confirmed. Overall, we identified 64 homozygous or putative compound heterozygous
200 *NOD2* variant carriers in the DiscovEHR IBD cohort, accounting for 6.5% of patients with an IBD
201 diagnosis in this clinical population (Figure 1, Table S4).

202 We were also able to evaluate longitudinal de-identified medical records for all patients
203 within the DiscovEHR IBD cohort. According to their EHR data, 21 patients received diagnoses
204 of both UC and CD. To clarify these diagnoses, we performed an evaluation of EHR information
205 (which includes demographics, encounter and problem list diagnosis codes, procedure codes,

206 and medications) for all 64 homozygous or compound heterozygous *NOD2* patients with an IBD
207 diagnosis. Through this review, 6 homozygotes exhibited a conflicting diagnosis of CD, of which
208 5 were resolved as CD and 1 could not be defined; 16 compound heterozygotes exhibited a
209 conflicting diagnosis of CD of which 6 were resolved as CD and 10 were resolved as UC (Table
210 S3). In total, we found that 17/18 (94.4%) of homozygous *NOD2* individuals and 33/46 (71.7%)
211 compound heterozygous had a diagnosis of CD and that 9.9% of all CD cases in this cohort
212 could be attributed to homozygous or compound heterozygous variants in *NOD2*. We next
213 investigated age of disease onset using the first recorded date of an IBD diagnosis in the EHR.
214 We identified 6 carriers of recessive *NOD2* variants (9.4% of our recessive *NOD2* patients with
215 IBD) who were diagnosed with IBD prior to 18 years of age. We also identified an additional 11
216 carriers of recessive *NOD2* variants diagnosed with IBD prior to age 30 years, which is at or
217 below the average age of IBD diagnosis (48) and is consistent with earlier disease onset (Table
218 S3). Of note, our RGC-GHS DiscovEHR data extends to a median of 14 years (and maximum of
219 20 years) of electronically recorded medical information, concurrent with the adoption of the
220 EHR by the Geisinger Health System. Since 72.4% of our cohort is currently over the age of 50
221 years, we cannot determine whether the age of onset for IBD occurred prior to the first
222 electronically recorded date of an IBD diagnosis for many recessive *NOD2* patients; thus it is
223 possible that other individuals with homozygous or compound heterozygous variants in *NOD2*
224 might have had pediatric-onset disease that was not captured in the EHR.

225 Finally, given the recessive inheritance of *NOD2* variants observed in both our pediatric
226 onset and adult IBD cohorts, we estimated the disease risk for the three main known CD risk
227 alleles (p.R702W, p.G908R, and p.L1007fs) in our adult IBD case cohort and their effect sizes
228 using additive, genotypic, and recessive genetic models. Under an additive model, we observed
229 similar effect sizes for each of the 3 variants [OR=1.43 (1.20-1.71 95%CI, P-value 4.63×10^{-5}) for
230 p.R702W; OR=1.56 (1.18-2.06 95%CI, P-value 1.54×10^{-3}) for p.G908R; and OR=1.84 (1.50-
231 2.26 95%CI, P-value 1.69×10^{-9}) for p.L1007fs], consistent with previously reported low to

232 moderate effect sizes for each allele by GWAS (19) and the most recent data available in the
233 IBD Exomes Portal (49) (Table 3, Figure 2). However, for the two risk alleles with homozygous
234 cases, in the genotypic model – which estimates distinct effect sizes for heterozygous and
235 homozygous carriers – we observe substantially larger effects in homozygotes versus
236 heterozygotes for the p.R702W variant (Het OR= 1.30 [1.06-1.58 95%CI], P-value 8.77×10^{-3} ,
237 versus Hom OR= 4.02 [2.17-7.45 95% CI], P-value 6.86×10^{-6}) and the p.L1007fs variant (Het
238 OR= 1.63 [1.30-2.04 95%CI], P-value 1.67×10^{-5} , versus Hom OR= 10.15 [4.75-21.69 95% CI],
239 P-value 1.38×10^{-12}). We also calculated the effect sizes using a recessive model for these two
240 variants and the 22 compound heterozygotes carrying any combination of the 3 CD risk alleles.
241 We found that recessive effect sizes for the p.R702W and p.L1007fs variants were similar to
242 those observed under the homozygous genotypic model (OR=3.91 [2.11-7.24 95% CI], P-value
243 2.86×10^{-6} , and OR=9.81 [4.59-20.94 95% CI], P-value 3.80×10^{-13} , respectively) (Table 3, Figure
244 2). Further, under the recessive model we observed that the effect size for the compound
245 heterozygotes was also significant (OR=4.35 [2.80-6.75 95% CI], P-value= 8.14×10^{-13}),
246 consistent with our previous observations (Table 3, Figure 2). Finally, we calculated the
247 combined contribution of the 3 CD risk alleles under the different genetic models: additive
248 (OR=1.64 [1.45-1.86 95%CI], P-value 4.58×10^{-15}), genotypic (Het OR= 1.49 [1.28-1.73 95%CI],
249 P-value 2.75×10^{-7} , versus Hom OR= 5.24 [3.77-7.27 95% CI], P-value 4.31×10^{-22}), and
250 recessive (OR=4.81 [3.47-6.67 95% CI], P-value= 1.63×10^{-25}) (Table 3, Figure 2). Collectively,
251 these analyses show substantially larger effects for *NOD2* homozygotes and compound
252 heterozygotes than heterozygotes and indicate that the genetic contribution of *NOD2* alleles, in
253 a subset of IBD patients, is consistent with a recessive disease model.

254 These observations are in line with previous analyses and meta-analyses of CD cohorts
255 where individuals carrying any one of the main three CD associated risk alleles (p.R702W,
256 p.G908R, or p.L1007fs) have 2-4 fold increased risk for developing CD (36), whereas carriers of
257 two or more of the same *NOD2* variants have a 15-40 fold increased risk for developing CD (33,

258 50, 51), exhibiting disease of the terminal ileum (34), and earlier diagnosis (by an average of 3
259 years) (33). Our observations support these studies but highlight a subset of IBD cases
260 molecularly defined by recessive inheritance of *NOD2* alleles that exhibit markedly increased
261 risk for CD with significantly earlier age of onset (mean age of onset among recessive *NOD2*
262 carriers in the DiscovEHR IBD cohort: 43.4y; mean age of onset in the DiscovEHR IBD cohort:
263 51.5y; P-value: 4.0×10^{-4} by unpaired t test).

264 Further, while we observe a low effect size for single allele carriers, based on our allelic
265 effect size calculations for each of the 3 main CD risk alleles in our DiscovEHR cohort (Table 3,
266 Figure 2), we hypothesize that homozygous and compound heterozygous *NOD2* individuals
267 included in large IBD GWAS cohorts have likely contributed to a large proportion of the relative
268 risk calculations for IBD, specifically for CD, under additive models, and that homozygous effect
269 sizes have been largely underappreciated or underreported. It is possible that stratification or
270 conditional statistical analysis of these large and heterogeneous cohorts based on *NOD2*
271 genotypes may increase power to detect other loci that contribute to IBD.

272 While our observations strongly support recessive inheritance of *NOD2* variants as a
273 driver of early onset Crohn's disease, we observed incomplete penetrance, as evidenced by
274 homozygous or compound heterozygous *NOD2* variant carriers that do not have a clinical
275 presentation of IBD (52-54). Penetrance and expressivity are two major genetic concepts that
276 play into the onset of the phenotype and the clinical presentation of monogenic diseases (55). In
277 the case of IBD, penetrance is known to be incomplete and clinical presentation is extremely
278 variable. Further, the contribution of additional environmental triggers that may enhance disease
279 onset and/or severity in an already genetically-compromised individual should not be
280 underestimated, especially considering that the loss of epithelial barrier function occurring
281 during IBD allows for host exposure to up to 10^{14} gut microbiota (56, 57). Even in cases of
282 monogenic IBD, such as IL-10 receptor deficiency (58-60), intestinal flora are required for
283 disease presentation in murine disease models (61-63). Furthermore, variation in genes

284 involved in NOD2-dependent signaling pathways, including *XIAP* (64-66) and *TRIM22* (67),
285 result in Mendelian forms of IBD. For *XIAP*, and most likely *TRIM22*, viral triggers are required
286 for disease onset and progression, and *XIAP* mutations have variable penetrance, with only a
287 small percentage of *XIAP*-deficiency patients developing CD (age of onset between 3 months
288 and 40 years (52)). As *NOD2*-deficient hosts are more susceptible to the pathogenic effects of a
289 changing intestinal microenvironment (68), the contribution of either discrete or continuous
290 gene-environment exposures may further explain heterogeneity in onset and presentation of
291 disease for genetically-sensitized recessive *NOD2* carriers.

292 Given the wide variability in clinical presentation of IBD (54), we cannot exclude the
293 possibility that recessive *NOD2* carriers exhibit subclinical phenotypes not formally diagnosed
294 as IBD or that they may eventually develop IBD. It is additionally possible that recessive *NOD2*
295 carriers in the DiscovEHR cohort have a diagnosis of IBD that has not been captured in the
296 EHR. Future investigation into the medical histories of recessive *NOD2* carriers may shed light
297 on this variable expressivity or incomplete capture of medical information. We also cannot
298 exclude the possibility that recessive *NOD2* carriers possess additional genes or alleles that
299 either contribute to disease onset and severity or, alternatively, provide protection or reduced
300 expressivity of the phenotype. Identification of these genetic modifiers warrants future
301 investigation both to unveil additional IBD-risk associated loci for early onset UC and CD cases
302 and to identify protective genes and alleles that can be used to derive therapeutic avenues for
303 IBD treatment and management.

304 In summary, in a cohort of 1,183 pediatric and early onset IBD patients, we report
305 recessive inheritance of rare and low frequency variants in *NOD2* accounting for about 8% of
306 probands. We assessed the contribution of *NOD2* recessive inheritance in a broader,
307 heterogeneous cohort of adult IBD patients, similar to those recruited for GWAS, and found that
308 recessive inheritance of variants in *NOD2* account for 6.5% of these IBD patients, including
309 9.9% of CD cases. Thus, recessive inheritance of rare and low frequency *NOD2* variants

310 explain a substantial proportion of CD cases in a pediatric cohort and a large clinical population,
311 with significantly earlier age of disease onset. Consistently, both pediatric and adult CD exhibit a
312 broad spectrum of clinical presentation, suggesting a shared etiology across age groups, at
313 least in the subgroup defined by recessive *NOD2*-driven CD. Our findings indicate that
314 deleterious *NOD2* variants should be considered as strong predictors of IBD-CD onset and
315 implicate *NOD2* as a Mendelian disease gene for early onset IBD, specifically for a molecularly
316 defined subset of Crohn's disease patients.

317

318 **Materials and Methods:**

319 We performed whole exome sequencing on a cohort of 1,183 probands with pediatric
320 onset IBD (ages 0-18.5 years), including their affected and unaffected parents and siblings,
321 where available (total samples = 2,704). All individuals were consented for genetic studies
322 under an IRB-approved protocol by the Toronto Hospital for Sick Children. Sample preparation,
323 whole exome sequencing, and sequence data production were performed at the Regeneron
324 Genetics Center (RGC) as previously described (46). In brief, 1ug of high-quality genomic DNA
325 for exome capture using the NimbleGen VCRome 2.1 design. Captured libraries were
326 sequenced on the Illumina HiSeq 2500 platform with v4 chemistry using paired-end 75 bp
327 reads. Exome sequencing was performed such that >85% of the bases were covered at 20x or
328 greater. Raw sequence reads were mapped and aligned to the GRCh37/hg19 human genome
329 reference assembly, and called variants were annotated and analyzed using an RGC developed
330 pipeline. Briefly, variants were filtered based on their observed minor allele frequencies at a
331 <2% cutoff using the internal RGC database and other population control databases (ExAC (69)
332 and NHLBI's ESP (70)) to filter out common polymorphisms and high frequency, likely benign
333 variants in consideration of disease prevalence.

334

335 **Conflict of Interest Disclosure**

336 J.E.H. is a postdoctoral fellow at the Regeneron Genetics Center, Regeneron
337 Pharmaceuticals, Inc. J.S., C.V.H., A.K.K., J.G.R., J.D.O., A.R.S., A.B., F.E.D, O.G., and C.G.J
338 are full-time employees of the Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc.
339 and receive stock options as part of compensation. N.W., R.M., K.F., A.G., and A.M.M have no
340 conflicts to disclose.

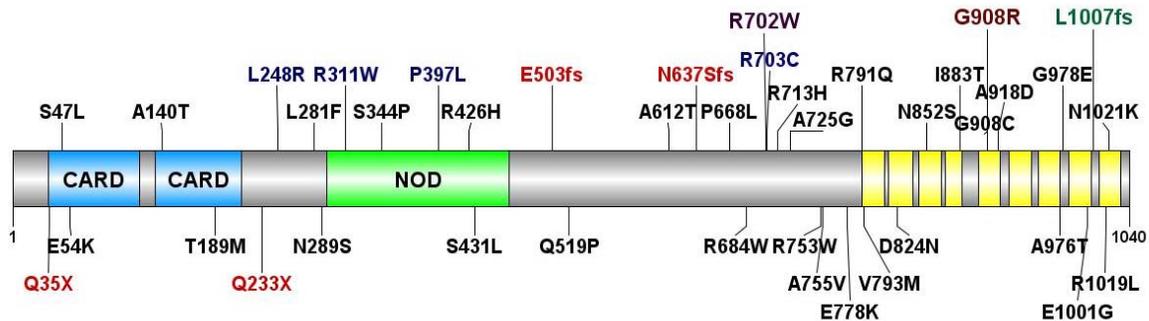
341

342 **Acknowledgements**

343 We are thankful to the patients and families who participated in this study. A.M.M. is funded by a
344 CIHR – Operating Grant (MOP119457) and the Leona M. and Harry B. Helmsley Charitable
345 Trust to study VEOIBD.

346

347 **Figures and Tables**



348

349 **Figure 1. Mutation spectrum of *NOD2* in inflammatory bowel disease (IBD) patients.**

350 *NOD2* variation identified in patients with pediatric early onset IBD (upper) and adult IBD cohort

351 from the RGC-GHS DiscovEHR collaboration (lower). Variants in blue were observed in both

352 cohorts; variants in red are predicted loss-of-function that result in nonsense mediated decay.

353 The three “common” low-frequency Crohn’s Disease risk variants are highlighted: R702W

354 (purple), G908R (brown), and L1007fs (green). Also depicted are the *NOD2* protein structural

355 domains: two caspase activation and recruitment domains (CARD), a nucleotide binding and

356 oligomerization (NOD) domain, and leucine rich repeat domains in yellow.

357

358 **Table 1. Mutation spectrum of recessive *NOD2* variants in an EO-IBD cohort.**

NOD2 Variant (p.)	# EO-IBD Probands	Mean Age (Range)	% CD Dx	Tissue Involvement		
				Colon	Ileum	Perianal
<u>Compound Heterozygous</u>						
Rare/Rare	4T; 1S	12.5 (9.0-14.6)	88.9%	80.0%	60.0%	40.0%
Common/Rare	1Q; 9T; 6D; 14S	12.6 (5.5-16.5)	93.1%	57.1%	82.1%	25.0%
Common/Common	17T; 5D; 10S	11.8 (2.1-18.5)	90.6%	56.3%	90.6%	25.0%
<u>Homozygous</u>						
Rare	1D; 2S	9.6 (4.2-15.1)	66.7%	33.3%	33.3%	33.3%
p.R702W	1Q; 2T; 3D; 1S	11.7 (5.8-13.8)	100%	85.7%	57.1%	42.9%
p.G908R	2T; 1D; 2S	10.6 (7.7-13.7)	80.0%	40.0%	40.0%	0.0%
p.L1007fs	4T; 2D; 4S	12.1 (5.9-13.4)	100%	60.0%	90.0%	20.0%

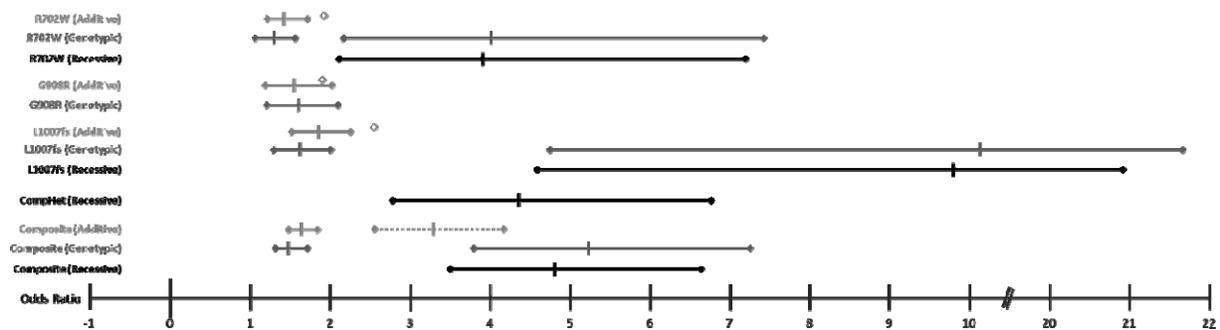
359
 360 Common NOD2 variants refer to the three main low-frequency Crohn's Disease risk variants
 361 p.R702W, p.G908R, and p.L1007fs; Rare NOD2 variants refer to other low-frequency variants
 362 (MAF_≤5%). Abbreviations: Q, quartet; T, trio; D, duo; S, singleton; Dx, diagnosis
 363

364
 365 **Table 2. Mutation spectrum of recessive *NOD2* variants in the RGC-GHS DiscovEHR adult**
 366 **IBD cohort.**

NOD2 Variants	# IBD patients	Mean Age (Range)	# CD Dx	% CD Dx	# Male	% Male	# Perianal	% Perianal
<u>Homozygous</u>								
p.R702W	10	47.3 (16.0-76.3)	10	100%	5	50.0%	1	10.0%
p.G908R	0	-	-	-	-	-	-	-
p.L1007fs	8	40.25 (11.0-75.0)	8	100%	4	50.0%	3	37.5%
<u>Compound Heterozygous</u>								
<u>Common/Common</u>								
p.R702W/p.G908R	6	48.5 (11.4-54.2)	3	50.0%	5	83.3%	1	16.7%
p.R702W/p.L1007fs	11	38.5 (21.5-69.2)	8	72.7%	3	27.3%	2	18.2%
p.G908R/p.L1007fs	5	35.2 (20.0-52.4)	4	80.0%	2	40.0%	0	0.0%
<u>Common/Rare</u>								
	16	40.3 (10.8-66.1)	8	50.0%	8	50.0%	6	37.5%
<u>Rare/Rare</u>								
	8	60.9 (30.8-78.7)	4	50.0%	3	37.5%	1	12.5%

367 Common NOD2 variants refer to the three main low-frequency Crohn's Disease risk variants
 368 p.R702W, p.G908R, and p.L1007fs; Rare NOD2 variants refer to other low-frequency variants
 369 (MAF_≤5%). Abbreviations: Dx, diagnosis.
 370

372 **Figure 2. Graphical representation of Odds Ratio (OR) point estimates and 95%**
373 **confidence intervals (CI) for the three main CD risk alleles (p.R702W, p.G908R, p.L1007fs)**
374 **under additive, genotypic, and recessive genetic models** (corresponding to values in Table
375 3). The dotted line in the Composite panel depicts the calculated CI with corresponding
376 calculated OR for 2 alleles under an additive genetic model; of note the point estimate (2xOR) is
377 outside of the 95% CI for the Composite genotypic homozygous and recessive models.
378 Diamonds correspond to estimated OR values for these same variants in the IBD Exomes
379 Browser(49); no confidence intervals are provided.



382 **Table 3. OR calculations for 3 NOD2 CD risk alleles (p.R702W, p.G908R, p.L1007fs),**
 383 **composite, and compound heterozygous combinations in the DiscovEHR cohort.** There
 384 were no homozygotes for the p.G908R variant affected with IBD in our cohort; therefore, no
 385 genotypic homozygous and recessive ORs could be calculated. The ‘Composite *NOD2*’
 386 calculations account for all alleles and genotypes for the 3 CD risk variants in the different
 387 genetic models.

<i>NOD2</i> Variant	DiscovEHR MAF	DiscovEHR Controls (N=50,305)	DiscovEHR IBD Cases (N=984)	Additive Model OR [95% CI] (P-value)	Genotypic Model (Heterozygous) OR [95% CI] (P-value)	Genotypic Model (Homozygous) OR [95% CI] (P-value)	Recessive Model OR [95% CI] (P-value)	ExAC MAF	IBD Exomes OR (P-value)
p.R702W	0.050	Het=4,843; Hom=156	Het=116; Hom=10	1.43 [1.20-1.71] (4.63x10 ⁻³)	1.30 [1.06-1.58] (0.008765)	4.02 [2.17-7.45] (6.86x10 ⁻⁶)	3.91 [2.11-7.24] (2.86x10 ⁻⁶)	0.035	1.92 (<1x10 ⁻¹⁶)
p.G908R	0.017	Het=1,735; Hom=16	Het=52; Hom=0	1.56 [1.18-2.06] (0.001544)	1.61 [1.21-2.13] (0.00087)	NA	NA	0.012	1.91 (<1x10 ⁻¹⁶)
p.L1007fs	0.029	Het=2,894; Hom=50	Het=86; Hom=8	1.84 [1.50-2.26] (1.69x10 ⁻³)	1.63 [1.30-2.04] (1.67x10 ⁻³)	10.15 [4.75-21.69] (1.38x10 ⁻¹²)	9.80 [4.59-20.94] (3.80x10 ⁻¹³)	0.018	2.57 (<1x10 ⁻¹⁶)
Compound Heterozygous	.	N=285	N=22	.	.	.	4.35 [2.80-6.75] (8.14x10 ⁻¹³)	.	.
Composite <i>NOD2</i>	.	Het=8955; Rec=450	Het=232; Rec=41	1.64 [1.45-1.86] (4.58x10 ⁻¹⁵) 3.29 [2.56-4.23] (2 alleles predicted)	1.49 [1.28-1.73] (2.75x10 ⁻⁷)	5.24 [3.77-7.27] (4.31x10 ⁻²²)	4.81 [3.47-6.67] (1.63x10 ⁻²⁵)	.	.

388

389

390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439

References

1. Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med*. 2009;361(21):2066-78.
2. Van Limbergen J, Wilson DC, Satsangi J. The genetics of Crohn's disease. *Annu Rev Genomics Hum Genet*. 2009;10:89-116.
3. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nature reviews Immunology*. 2008;8(6):458-66.
4. Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. *Annu Rev Immunol*. 2010;28:573-621.
5. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703-7.
6. Cho JH, Abraham C. Inflammatory bowel disease genetics: Nod2. *Annu Rev Med*. 2007;58:401-16.
7. Abraham C, Cho JH. Functional consequences of NOD2 (CARD15) mutations. *Inflamm Bowel Dis*. 2006;12(7):641-50.
8. Chu H, Khosravi A, Kusumawardhani IP, Kwon AH, Vasconcelos AC, Cunha LD, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science*. 2016;352(6289):1116-20.
9. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*. 2007;39(2):207-11.
10. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*. 2007;39(7):830-2.
11. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007;39(5):596-604.
12. Abraham C, Cho JH. IL-23 and autoimmunity: new insights into the pathogenesis of inflammatory bowel disease. *Annu Rev Med*. 2009;60:97-110.
13. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461-3.
14. Cattin AL, Le Beyec J, Barreau F, Saint-Just S, Houllier A, Gonzalez FJ, et al. Hepatocyte nuclear factor 4alpha, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol Cell Biol*. 2009;29(23):6294-308.
15. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011;474(7351):307-17.
16. Muise AM, Walters TD, Glowacka WK, Griffiths AM, Ngan BY, Lan H, et al. Polymorphisms in E-cadherin (CDH1) result in a mis-localised cytoplasmic protein that is associated with Crohn's disease. *Gut*. 2009;58(8):1121-7.
17. Scharl M, Paul G, Weber A, Jung BC, Docherty MJ, Hausmann M, et al. Protection of epithelial barrier function by the Crohn's disease associated gene protein tyrosine phosphatase n2. *Gastroenterology*. 2009;137(6):2030-40 e5.
18. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H, et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell*. 2008;134(5):743-56.
19. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119-24.

- 440 20. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association
441 analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared
442 genetic risk across populations. *Nat Genet.* 2015;47(9):979-86.
- 443 21. McGovern DP, Kugathasan S, Cho JH. Genetics of Inflammatory Bowel Diseases.
444 *Gastroenterology.* 2015;149(5):1163-76 e2.
- 445 22. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and
446 health. *Nat Rev Gastroenterol Hepatol.* 2012;9(10):577-89.
- 447 23. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current
448 status and the future ahead. *Gastroenterology.* 2014;146(6):1489-99.
- 449 24. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al. Association
450 of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature.*
451 2001;411(6837):599-603.
- 452 25. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, et al.
453 Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature.*
454 1996;379(6568):821-3.
- 455 26. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift
456 mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature.*
457 2001;411(6837):603-6.
- 458 27. Girardin SE, Boneca IG, Viala J, Chamaillard M, Labigne A, Thomas G, et al. Nod2 is a
459 general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem.*
460 2003;278(11):8869-72.
- 461 28. Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J, et al. Host recognition
462 of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol*
463 *Chem.* 2003;278(8):5509-12.
- 464 29. Van Limbergen J, Radford-Smith G, Satsangi J. Advances in IBD genetics. *Nat Rev*
465 *Gastroenterol Hepatol.* 2014;11(6):372-85.
- 466 30. Inohara, Chamaillard, McDonald C, Nunez G. NOD-LRR proteins: role in host-microbial
467 interactions and inflammatory disease. *Annu Rev Biochem.* 2005;74:355-83.
- 468 31. Strober W, Murray PJ, Kitani A, Watanabe T. Signalling pathways and molecular
469 interactions of NOD1 and NOD2. *Nat Rev Immunol.* 2006;6(1):9-20.
- 470 32. Watanabe T, Kitani A, Strober W. NOD2 regulation of Toll-like receptor responses and
471 the pathogenesis of Crohn's disease. *Gut.* 2005;54(11):1515-8.
- 472 33. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, et al. CARD15/NOD2
473 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel
474 disease. *Am J Hum Genet.* 2002;70(4):845-57.
- 475 34. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited
476 determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study.
477 *Lancet.* 2016;387(10014):156-67.
- 478 35. Bonen DK, Ogura Y, Nicolae DL, Inohara N, Saab L, Tanabe T, et al. Crohn's disease-
479 associated NOD2 variants share a signaling defect in response to lipopolysaccharide and
480 peptidoglycan. *Gastroenterology.* 2003;124(1):140-6.
- 481 36. Chamaillard M, Philpott D, Girardin SE, Zouali H, Lesage S, Chareyre F, et al. Gene-
482 environment interaction modulated by allelic heterogeneity in inflammatory diseases. *Proc Natl*
483 *Acad Sci U S A.* 2003;100(6):3455-60.
- 484 37. van Heel DA, Ghosh S, Butler M, Hunt KA, Lundberg AM, Ahmad T, et al. Muramyl
485 dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. *Lancet.*
486 2005;365(9473):1794-6.
- 487 38. Moran CJ, Klein C, Muise AM, Snapper SB. Very early-onset inflammatory bowel
488 disease: gaining insight through focused discovery. *Inflamm Bowel Dis.* 2015;21(5):1166-75.
- 489 39. Rosen MJ, Dhawan A, Saeed SA. Inflammatory Bowel Disease in Children and
490 Adolescents. *JAMA Pediatr.* 2015;169(11):1053-60.

- 491 40. Cutler DJ, Zwick ME, Okou DT, Prahalad S, Walters T, Guthery SL, et al. Dissecting
492 Allele Architecture of Early Onset IBD Using High-Density Genotyping. *PLoS One*.
493 2015;10(6):e0128074.
- 494 41. Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, Dubinsky M, et al.
495 Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat*
496 *Genet*. 2009;41(12):1335-40.
- 497 42. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, et
498 al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease.
499 *Nature genetics*. 2008;40(10):1211-5.
- 500 43. Paul T, Birnbaum A, Pal DK, Pittman N, Ceballos C, LeLeiko NS, et al. Distinct
501 phenotype of early childhood inflammatory bowel disease. *J Clin Gastroenterol*. 2006;40(7):583-
502 6.
- 503 44. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep
504 resequencing of GWAS loci identifies independent rare variants associated with inflammatory
505 bowel disease. *Nat Genet*. 2011;43(11):1066-73.
- 506 45. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward
507 an integrated clinical, molecular and serological classification of inflammatory bowel disease:
508 report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J*
509 *Gastroenterol*. 2005;19 Suppl A:5A-36A.
- 510 46. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al.
511 Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from
512 the DiscovEHR study. *Science*. 2016;354(6319).
- 513 47. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE. PRIMUS: rapid
514 reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum*
515 *Genet*. 2014;95(5):553-64.
- 516 48. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of
517 inflammatory bowel diseases. *Gastroenterology*. 2011;140(6):1785-94.
- 518 49. Portal IE. Cambridge, MA [Available from: <http://ibd.broadinstitute.org/>].
- 519 50. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP. Differential effects of
520 NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis.
521 *Am J Gastroenterol*. 2004;99(12):2393-404.
- 522 51. Hugot JP, Zaccaria I, Cavanaugh J, Yang H, Vermeire S, Lappalainen M, et al.
523 Prevalence of CARD15/NOD2 mutations in Caucasian healthy people. *Am J Gastroenterol*.
524 2007;102(6):1259-67.
- 525 52. Uhlig HH, Schwerd T, Koletzko S, Shah N, Kammermeier J, Elkadri A, et al. The
526 diagnostic approach to monogenic very early onset inflammatory bowel disease.
527 *Gastroenterology*. 2014;147(5):990-1007 e3.
- 528 53. Kammermeier J, Dziubak R, Pescarin M, Drury S, Godwin H, Reeve K, et al. Phenotypic
529 and genotypic characterisation of inflammatory bowel disease presenting before the age of two
530 years. *J Crohns Colitis*. 2016.
- 531 54. Ray A, Dittel BN. Interrelatedness between dysbiosis in the gut microbiota due to
532 immunodeficiency and disease penetrance of colitis. *Immunology*. 2015;146(3):359-68.
- 533 55. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where
534 genotype is not predictive of phenotype: towards an understanding of the molecular basis of
535 reduced penetrance in human inherited disease. *Hum Genet*. 2013;132(10):1077-130.
- 536 56. Mankertz J, Schulzke JD. Altered permeability in inflammatory bowel disease:
537 pathophysiology and clinical implications. *Curr Opin Gastroenterol*. 2007;23(4):379-83.
- 538 57. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal
539 disease: understanding a hidden metabolic organ. *Therap Adv Gastroenterol*. 2013;6(4):295-
540 308.

- 541 58. Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schaffer AA, Noyan F, et al. Inflammatory
542 bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*.
543 2009;361(21):2033-45.
- 544 59. Kotlarz D, Beier R, Murugan D, Diestelhorst J, Jensen O, Boztug K, et al. Loss of
545 interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and
546 therapy. *Gastroenterology*. 2012;143(2):347-55.
- 547 60. Muise AM, Snapper SB, Kugathasan S. The age of gene discovery in very early onset
548 inflammatory bowel disease. *Gastroenterology*. 2012;143(2):285-8.
- 549 61. Kuhn R, Lohler J, Rennick D, Rajewsky K, Muller W. Interleukin-10-deficient mice
550 develop chronic enterocolitis. *Cell*. 1993;75(2):263-74.
- 551 62. Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, et al. Resident
552 enteric bacteria are necessary for development of spontaneous colitis and immune system
553 activation in interleukin-10-deficient mice. *Infect Immun*. 1998;66(11):5224-31.
- 554 63. Yang I, Eibach D, Kops F, Brenneke B, Woltemate S, Schulze J, et al. Intestinal
555 microbiota composition of interleukin-10 deficient C57BL/6J mice and susceptibility to
556 *Helicobacter hepaticus*-induced colitis. *PLoS One*. 2013;8(8):e70783.
- 557 64. Abbott DW, Yang Y, Hutti JE, Madhavarapu S, Kelliher MA, Cantley LC. Coordinated
558 regulation of Toll-like receptor and NOD2 signaling by K63-linked polyubiquitin chains. *Mol Cell*
559 *Biol*. 2007;27(17):6012-25.
- 560 65. Damgaard RB, Fiil BK, Speckmann C, Yabal M, zur Stadt U, Bekker-Jensen S, et al.
561 Disease-causing mutations in the XIAP BIR2 domain impair NOD2-dependent immune
562 signalling. *EMBO Mol Med*. 2013;5(8):1278-95.
- 563 66. Krieg A, Correa RG, Garrison JB, Le Negrato G, Welsh K, Huang Z, et al. XIAP
564 mediates NOD signaling via interaction with RIP2. *Proc Natl Acad Sci U S A*.
565 2009;106(34):14524-9.
- 566 67. Li Q, Lee CH, Peters LA, Mastropaolo LA, Thoeni C, Elkadri A, et al. Variants in TRIM22
567 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel
568 Disease. *Gastroenterology*. 2016;150(5):1196-207.
- 569 68. Ramanan D, Bowcutt R, Lee SC, Tang MS, Kurtz ZD, Ding Y, et al. Helminth infection
570 promotes colonization resistance via type 2 immunity. *Science*. 2016;352(6285):608-12.
- 571 69. Lek M. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* [Internet].
572 2015. Available from: <http://biorxiv.org/content/early/2015/10/30/030338>.
- 573 70. Exome Variant Server [Internet]. [cited July, 2016]. Available from:
574 <http://evs.gs.washington.edu/EVS/>.
575