World Journal of Gastroenterology

Online Submissions: http://www.wjgnet.com/esps/
wjg@wjgnet.com
doi:10.3748/wjg.v19.i40.6721

World J Gastroenterol 2013 October 28; 19(40): 6721-6729
ISSN 1007-9327 (print) ISSN 2219-2840 (online)
© 2013 Baishideng. All rights reserved.

TOPIC HIGHLIGHT

WJG 20th Anniversary Special Issues (3): Inflammatory bowel disease

Impact of exome sequencing in inflammatory bowel disease

Christopher J Cardinale, Judith R Kelsen, Robert N Baldassano, Hakon Hakonarson

Christopher J Cardinale, Hakon Hakonarson, Center for Applied Genomics, Children's Hospital of Philadelphia, Abramson Research Center Suite 1216, Philadelphia, PA 19104, United States
Judith R Kelsen, Robert N Baldassano, Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, United States
Author contributions: Cardinale CJ, Kelsen JR, Baldassano RN, Hakonarson H wrote and edited the manuscript.
Supported by A Senior Research Award from the Crohn's to Cardinale CJ; Colitis Foundation of America to Hakonarson H; and a special purpose fund from the Edmunds Family Foundation for Ulcerative Colitis Studies to Baldassano RN
Correspondence to: Hakon Hakonarson, MD, PhD, Director, Center for Applied Genomics, Children's Hospital of Philadelphia, Abramson Research Center Suite 1216, 3615 Civic Center Blvd, Philadelphia, PA 19104,
United States. hakonarson@chop.edu
Telephone: +1-267-4266047 Fax: +1-267-4260363
Received: August 11, 2013 Revised: September 11, 2013
Accepted: September 16, 2013
Published online: October 28, 2013

Abstract

Approaches to understanding the genetic contribution to inflammatory bowel disease (IBD) have continuously evolved from family- and population-based epidemiology, to linkage analysis, and most recently, to genome-wide association studies (GWAS). The next stage in this evolution seems to be the sequencing of the exome, that is, the regions of the human genome which encode proteins. The GWAS approach has been very fruitful in identifying at least 163 loci as being associated with IBD, and now, exome sequencing promises to take our genetic understanding to the next level. In this review we will discuss the possible contributions that can be made by an exome sequencing approach both at the individual patient level to aid with disease diagnosis and future therapies, as well as in advancing knowledge of the pathogenesis of IBD.

© 2013 Baishideng. All rights reserved.

Key words: Sequencing; Exome; Genetics; Inflammatory bowel disease

Core tip: The genetic understanding of inflammatory bowel disease (IBD) has progressed over the last twenty years as new technologies and analytic techniques have become available. The nascent revolution in next-generation sequencing will enable us to sequence the exome - all the protein coding genes in the genome - in thousands of individuals. This review discusses the implications of this new approach for diagnosis in very early onset IBD and as a tool to gain understanding of the hereditary basis of the common polygenic form of the disease at the population level.

Cardinale CJ, Kelsen JR, Baldassano RN, Hakonarson H. Impact of exome sequencing in inflammatory bowel disease. *World J Gastroenterol* 2013; 19(40): 6721-6729 Available from: URL: http://www.wjgnet.com/1007-9327/full/v19/i40/6721.htm DOI: http://dx.doi.org/10.3748/wjg.v19.i40.6721

INTRODUCTION

The inflammatory bowel diseases (IBDs) consist of two main types of pathology: Crohn's disease and ulcerative colitis. Over the preceding decades, genetic epidemiology of twins and families indicated that these diseases have a strong genetic component, but that they do not segregate according to a Mendelian pattern of inheritance such as autosomal dominant, autosomal recessive, or X-linked[1]. Twin studies of Crohn's have shown a concordance of 20%-50% for monozygotic twins and 0%-7% for dizygotic twins[2]. For ulcerative colitis the concordance is 14%-19% for monozygotic and 0%-7% for dizygotic[2]. The fact that the monozygotic concor-

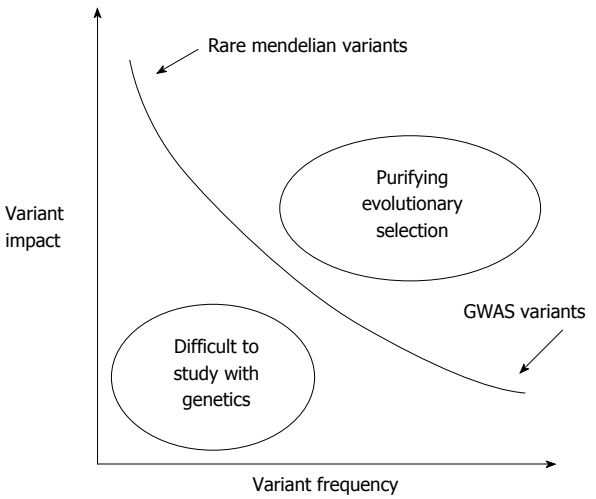

Figure 1 Relationship between variant frequency and functional impact. Rare and highly damaging variants such as those associated with familial forms of very early onset inflammatory bowel disease (IBD) tend to occur rarely in the population. They are unable to achieve high allele frequencies due to negative evolutionary selection. Variants captured by genome-wide association studies (GWAS), and which account for much of the population attributable risk of IBD, tend to be quite common but have small functional consequences (typically OR < 1.2). Rare variants with small impacts are difficult to assess statistically with the tools of genetics. Exome sequencing is intended to fill in the middle part of the curve: less common variants that have moderate impacts.

dance is well below 100% shows that there are strong environmental contributions and that there is incomplete penetrance of the genetic susceptibility loci. At the same time, the risk is considerably elevated compared to the general population. Supported by the results of recent genome-wide association studies, the most commonly accepted model of IBD susceptibility is a multifactorial model in which polygenic inheritance at hundreds of genetic loci, each with small effects, contribute along with non-genetic factors, such as diet and microbiome composition.

One of the first successful approaches to identifying specific risk genes was family-based linkage analysis. This approach seeks to identify chromosomal regions containing causative genes on the basis of recombinations within a family between a microsatellite marker and the trait of interest. Six loci were identified using linkage analysis, including the *IBD3* locus containing the human leukocyte antigen complex on chromosome 6, and the *IBD1* locus, the single largest genetic risk factor for Crohn's, which contains the nucleotide-binding oligomerization domain protein 2 (*NOD2*) gene on chromosome 16[3-5].

The next technology to make a major impact in IBD genetics has been genome-wide association studies (GWAS). These studies involve genotyping hundreds of thousands of single nucleotide polymorphisms (SNPs) throughout the entire genome in order to find direct association between a specific polymorphism and the case/control status. The first successful study found an association between the interleukin *(IL)23R* locus and Crohn's disease in addition to replicating the *NOD2* association[6]. Expanding the number of cases and controls in the cohort as genotyping prices dropped resulted in identification of *ATG16L1*[7], *IRGM*, *MST1*, *NKX2-3*, and *PTPN2*[8,9]. The first IBD GWAS studies in a pediatric cohort were reported by our group, highlighting associations with *TNFRSF6B* and *IL27*[10,11]. As studies have grown more powered with increased cohort sizes, genotype imputation techniques[12], and international collaboration through the IBD Genetics Consortium, the tally of associated loci for Crohn's and ulcerative colitis has risen to 163 in the latest meta-analysis[13], demonstrating unequivocally the polygenic nature of IBD inheritance. Notably, the distribution of SNPs genotyped in a GWAS study covers intergenic as well as exonic and intronic regions, so that polymorphisms which predominately affect the regulation of gene expression through transcriptional control can be assessed. Analysis of data from the ENCODE consortium has advanced the notion that much of the heritability of complex disorders originates in these non-coding regulatory regions of the genome[14]. There is no assumption in GWAS that the susceptibility or protective variants are confined to amino acid substitutions in proteins, the type of variation that would be found in exome sequencing. However, a major disadvantage of GWAS studies is that they are much more attuned to detecting common variation, that is, greater than 5% minor allele frequency for a SNP. It is worthy of note that IBD has generated a greater number of associations than any form of pathology studied genetically to date, leading some to suggest that evolutionary selective pressures for variants in the genes underlying the immune response drove autoimmune-risk alleles to relatively high frequencies, a phenomenon known as balancing selection[15]. The greater sensitivity of GWAS towards common variants is one reason among many that GWAS studies have only been able to account for a fraction of the heritability of polygenic diseases such as IBD[16]. It is becoming increasingly clear that there is more to the story than the common disease-common variant hypothesis[17], and that rare variants, detectable only through sequencing, must also play a role[18,19]. Moreover, these coding variants are more likely to have high ORs, greater penetrance, and to be amenable to follow-up by functional experimentation[18]. Figure 1 illustrates the relationship between variant frequency and the phenotypic impact of the variant. Highly disruptive mutations will not rise to high frequency due to purifying selection. Exome sequencing is an ideal technology to fill in the intermediate frequency range of variants which may have stronger impacts than the weak associations detected by common GWAS variants.

PROCESS OF EXOME SEQUENCING

With current technology, sequencing the whole 3 billion-base pair genome at the required depth of coverage to make rare variant calls is an expensive process, making it impractical to use on the scale required to implicate

less-common variants in IBD. Statistically validating less common variants with phenotypic impact would require GWAS-sized datasets comprising thousands of cases and controls[20]. A more practical alternative that has arisen since the emergence of next generation sequencing technology is to sequence the exome, the 1% of the genome that encodes protein. It has been estimated that 85% of monogenic, Mendelian disorders are the result of alterations in protein amino acid sequence[21], supporting the idea that exon-focused sequencing will yield the most functionally interesting variants.

The most common way to fractionate the genome for exome sequencing is in-solution hybridization. This can be accomplished by shearing the DNA into small 200-300 bp fragments by ultrasonic or enzymatic methods followed by ligation of common adapter sequences to the 3' and 5' ends of the fragments so that the sequencing primer can anneal. This whole-genome library is captured by hybridization in solution with 50-120 nucleotide-long "baits" that are complementary to the exon sequence being targeted. The library-bound baits are bound to magnetic beads and the non-coding DNA is washed out. The captured fragments are eluted and amplified by PCR. Next generation sequencing instruments that utilize the exome library include the HiSeq and MiSeq (Illumina, Inc.) and Ion Torrent Proton (Life Technologies). The instruments sequence by synthesis with a DNA polymerase, analyzing the incorporation of the next nucleotide by fluorescence imaging with modified nucleotides (Illumina) or by electrical measurement of the protons produced by the incorporation of nucleotides (Ion Torrent). This generates a short "read" typically 100-200 bp in length, significantly shorter than the 700-bp reads produced by traditional Sanger sequencing. The reads are furnished as a list of sequences accompanied by quality metrics, known as a FASTQ file. One of these instruments can generate 20-60 gigabases of sequence per day.

The FASTQ file is analyzed by a read-mapping program, such as the popular Burrows-Wheeler aligner[22], which matches these short reads with a reference genome. The alignment is stored in a common file format called BAM which is interpretable by a variety of analysis tools for visualization and variant identification. When enough independent reads have been aligned at the same nucleotide location in the genome, usually at least 20 reads, a variant calling application, for instance, Genome Analysis Toolkit[23,24], the variant caller for the 1000 Genomes Project, can be used to decide if the site matches the reference sequence or contains an alternate nucleotide. The variant calls can be collected in a variety of formats, typically the Variant Call File (VCF). A range of statistical analyses can be performed on the VCF files for each exome, including annotating them for function (missense, indel, synonymous) and likely impact of the variant (damaging, tolerated) using tools such as ANNOVAR[25], Sorting Intolerant From Tolerant[26,27], and PolyPhen[28]. These tools use evolutionary conservation of the gene across diverse species as well as the chemistry of the amino acid substitution to generate a predictive score of each variant's potential impact. The software tools also integrate information about the frequency of the variants in the general population using databases such as the The National Heart, Lung, and Blood Institute Exome Sequencing Project[29] and dbSNP[30], since it is most likely that a damaging and impactful mutation would be quite rare due to purifying evolutionary selection.

In large case/control studies, the coding variants will be rarer than the polymorphisms identified through GWAS, so that any individual rare variant is unlikely to achieve a threshold of statistical significance[20]. Therefore, a variety of groups have developed methods to aggregate all of the rare variants in a gene and test them collectively in order to identify a rare variant burden in cases compared with controls or to detect an unusual distribution of variant frequencies between cases and controls for a given gene[31]. A number of these tests have the feature of being able to detect association in the presence of a mixture of risk, protective, and neutral variation[32].

ROLE FOR EXOME SEQUENCING IN IBD

Individual- and family-based sequencing for clinical use

In attempting to identify a role for exome sequencing in inflammatory bowel disease we can appreciate two scenarios where it might be used. The first scenario is that of an individual patient or family with an atypical clinical presentation whose diagnosis or therapeutic decision may be influenced by genetic information. This can be seen in the very young children who present with clinical symptoms of IBD, known as very early onset IBD (VEO-IBD). These children frequently present with a more severe disease and often with a phenotype that is distinct from older children and adults, including extensive colonic disease unresponsive to standard therapy. These findings suggest distinct etiopathogenic pathways. In one well-known case, a 15-mo-old child presented with failure to thrive and perianal fistulae that was refractory to medical care. His disease progressed to pancolonic involvement, however the terminal ileum and upper tract were spared. This early age of onset and severity suggested a severe perturbation of the immune system[33]. He underwent numerous surgical procedures and treatment with immunosuppressive drugs, as well as targeted genetic and immunologic testing that did not yield a recognizable diagnosis or remission of symptoms. Sequencing of the child's exome revealed that this patient had an exceedingly rare mutation on the X chromosome in the *XIAP* gene, a potent regulator of the inflammatory response. He was treated by bone marrow transplant resulting in resolution of his disease. In our own IBD center at the Children's Hospital of Philadelphia, we encountered a 5-mo-old patient with colonic inflammatory bowel disease. She presented with severe disease that was unresponsive to medical therapy. Her

course was complicated by frequent episodes of dehydration and she became transfusion dependent despite various treatments. Exome sequencing in this patient revealed a mutation in the *MEFV* gene, resulting in a diagnosis of familial Mediterranean fever. The patient was referred to a pediatric rheumatology specialist and is being successfully treated for FMF with colchicine.

These successes highlight the critical role of exome sequencing in carefully selected patients by providing diagnoses that can guide treatment. Factors that suggest a patient may have a rare genetic perturbation that might be elucidated by exome sequencing would include early onset of disease, unusual severity, familial pattern of transmission, and a refractory response to standard therapies. In these cases, collecting DNA samples from parents so that exome sequencing in a trio setting can be performed is of high value. This will allow the identification of *de novo* variants as well as aiding in the elimination of the numerous false positive variant calls that exome sequencing generates by checking for non-Mendelian transmission of mutations. If a Mendelian inheritance model can be specified, as in the case of a consanguineous family which is likely to be autosomal recessive, such information can be of great help in narrowing down the causal variant. Homozygosity mapping in two consanguineous families was successfully used to identify mutations in the IL-10 receptor genes that resulted in severe VEO-IBD unresponsive to therapy. With this discovery, the disease resolved with bone marrow transplant[34]. This critical finding has been replicated in larger cohorts of patients with VEO-IBD and has shed light on an important pathway in the development of VEO-IBD[35]. A further appeal of applying exome sequencing in a family setting is in identifying novel monogenic causes of IBD that might yield an unexpected insight into the biology of disease, thereby directing interest towards novel targets for therapeutic development. An example would be the development of monoclonal antibodies that dramatically lower low-density lipoprotein (LDL) cholesterol by inhibiting proprotein convertase subtilisin kexin 9, a protein that was found to be deficient in a small number of individuals which genetically very low LDL[36].

Somatic mutations

An area where exome sequencing has been impactful is in the sequencing of cancer tissue exomes in comparison with the patient's inherited exome. Some studies have been successful in identifying somatic "driver mutations" which are essential for the growth of the tumor, which can spur the development of chemotherapeutic interventions that will target the cancer specifically[37,38]. Great interest has sprung up around the promise of personalized, or precision, medicine for cancer driven by the somatic genomics of tumors[39]. Whether sequencing of intestinal biopsies in IBD present an avenue to identity somatic mutations that may be critically important for microbiome interaction is yet to be determined, but studies are underway that are addressing this possibility.

Exome sequencing as a research tool to complement GWAS

The second scenario in which exome sequencing can be impactful is as a research tool to augment GWAS in uncovering novel susceptibility loci and specific coding variants in the typical polygenic form of Crohn's and ulcerative colitis. Whether exome sequencing will succeed in this role to the same degree as GWAS is still controversial. It is clear that identifying genes carrying a burden of exonic rare variants in a disease with the highly polygenic architecture of IBD will require GWAS-sized cohorts, that is, ones consisting of tens of thousands of cases and controls[40]. The high cost and labor intensity of such an effort currently makes these studies prohibitively expensive to all but the most resource-rich groups. Nevertheless, some groups have succeeded in finding rare variant associations through sequencing at the phenotypic extremes of several complex traits in carefully selected candidate genes such as *ANGPTL4* and *ANGPTL5*[41] or *LPL*[42] in triglycerides, *SLC12A1* in blood pressure[43], and *IFIH1* in type 1 diabetes[44]. Targeted next-generation sequencing in IBD has even produced some rare variant associations by following up GWAS hits, such as coding mutations that reduce signaling through the IL-23 receptor[45]. Targeted next-generation sequencing by Rivas *et al*[46] identified additional *NOD2* and *IL23R* coding variants, as well as novel coding variants in *CARD9*, *IL18RAP*, *CUL2*, *C1orf106*, *PTPN22* and *MUC19*. Our group has also recently identified rare nonsynonymous variants in the *TNFRSF6B* gene in IBD patients with pediatric onset disease[47], suggesting that this could be true for other GWAS loci as well. Table 1 summarizes genes that have been shown to have nonsynonymous variants with disease relevance in IBD.

Despite the success of these candidate gene efforts, doubts remain about how practical rare variant studies will be when applied to the entire exome. Most of the rare variant associations identified so far in candidate gene sequencing would not meet the stringent Bonferroni correction for multiple testing on an exome scale, estimated to be a $P < 2.5 \times 10^{-6}$[51]. Investigators must also consider that supporting novel rare variant associations requires replication in additional cohorts since rare variants are often population-specific and frequencies can vary in very inhomogeneous ways in spatially structured populations[48]. This type of confounding, known as population stratification, can lead to spurious associations. Therefore, replication would likely require additional sequencing of large cohorts since genotyping of specific variants would likely not be useful in a different geography or ethnicity[49], although the replication sequencing might be limited only to genes of interest in the discovery cohort.

Concerns about the likelihood of uncovering a substantial amount of heritability in common autoimmune diseases was raised by a recent report by Hunt *et al*[50]. This effort selected 25 risk genes that were identified in GWAS of at least two different common autoimmune

Table 1 Genes implicated as having coding variants in inflammatory bowel disease

Gene symbol	Description	Ref.
ADAM17	Disintegrin and metalloproteinase domain-containing protein 17	[57]
ATG16L1	Autophagy related 16-like 1	[7]
C1orf106	Chromosome 1 open reading frame 106	[46]
CARD9	Caspase recruitment domain family, member 9	[46]
CUL2	Cullin 2	[46]
DLG5	Discs-large 5	[58]
HEATR3	HEAT repeat containing 3	[59]
IL10RA and *B*	Interleukin-10 receptor	[34]
IL18RAP	Interleukin 18 receptor accessory protein	[46]
IL23R	Interleukin 23 receptor	[45]
LRBA	LPS-responsive vesicle trafficking, beach and anchor containing	[60]
MDR1	Multi-drug resistance 1	[61]
MUC19	Mucin 19, oligomeric	
NCF2	Neutrophil cytosolic factor 2	[62]
NDP52	Antigen nuclear dot 52 kDa Protein	[55]
NOD2	Nucleotide-binding oligomerization domain containing 2	[3,4]
PRDM1	PR domain containing 1, with Zinc finger domain	[55]
PTN22	Protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	[46]
SLC22A4	Solute carrier family 22 (organic cation/ergothioneine transporter), member 4	[63]
TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	[47]

diseases. The exons of these 25 genes were sequenced with excellent coverage in a cohort of 24892 subjects with six autoimmune disease phenotypes and 17019 controls. They found that the great majority of variants uncovered occurred in a single subject. Five aggregating gene-based tests (rather than individual variant-based tests) were used to identify rare-variant enrichment for any of the genes but none were found to be statistically significant. The authors concluded that there was little support for large-scale whole-exome sequencing projects in common autoimmune diseases. While this report may portend that the impact from rare coding variants is negligible, there are some limitations to the study that leave open the possibility for meaningful discovery. The study considered only 25 genes, while an exome-based approach would survey all 20000 human protein-coding genes. It is likely that the risk conferred by these 25 *GWAS* genes is carried by non-coding variation, while the risk at some subset of loci in the genome could be carried by rare coding variation that are not captured by GWAS SNPs. This is particularly true for variants in the intermediate 0.5%-5% frequency range which could be impactful in aggregate while escaping detection in GWAS studies due to the weak linkage disequilibrium for variants in this frequency range with common variants[51]. Indeed, Hunt *et al*[50] did identify three risk mutations at approximately the 5% minor allele frequency. Furthermore, the abundant singleton mutations could still identify IBD risk genes through the use of statistical tests which weight mutations more or less heavily depending on their frequency, as the very rare variants are the most likely to be impactful functionally. Several methods such as adaptive sum tests[52], Sequence Kernel Association Test[53], and Variable Threshold[54] tests have been developed specifically for sustaining a high statistical power with rare variants. Finally, the six autoimmune disease phenotypes were quite heterogeneous in their pathologic nature, ranging from IBD to autoimmune thyroid disease to multiple sclerosis. These diseases have distinct mechanisms with different rare variants underlying them, possibly in none of the 25 genes sequenced. Therefore, it is arguably diluting the power to detect rare variant association by combining diverse diseases.

A recent study by Ellinghaus *et al*[55] utilized exome sequencing to identify a role for missense variants in *PRDM1* and *NDP52* in Crohn's disease. Variants in these two genes were discovered in a cohort of 42 whole-exome sequenced individuals, with discovered variants being prioritized by functional impact scores and presence within GWAS-delineated loci. Over 20000 combined Crohn's and ulcerative colitis cases and controls were genotyped to establish that two variants, p.Ser354Asn in *PRDM1* and p.Val248Ala in *NDP52* were associated with IBD. Functional studies showed that the *PRDM1* mutant increased T cell proliferation and cytokine secretion while the *NDP52* mutant impaired the ability of the protein to downregulate nuclear factor kappa B signaling in toll-like receptor signaling pathways. This paper provides an example of how exome sequencing, even in a modest cohort, can refine GWAS signals and uncover less common risk variants, especially when coupled with functional validation.

Sequencing and risk prediction

Our group recently developed a machine-learning approach to predicting risk for IBD using data from the International IBD Genetics Consortium's ImmunoChip project[56]. The ImmunoChip assays 200000 SNPs with very dense coverage in genomic regions that have been associated with autoimmune disease through genome-wide association studies. Due to the ImmunoChip's wide spectrum of variants and the large number of cases and controls genotyped in the project, it was possible to use a penalized logistic regression model to predict risk for IBD with area under the curve of 0.86 for Crohn's disease and 0.83 for ulcerative colitis. With the coming availability of large-scale whole exome data we expect that risk prediction can be improved further and may achieve clinically-useful levels with the comprehensive catalog of variation that would be produced through eventual whole-genome sequencing.

CONCLUSION

We can predict with some confidence that exome sequencing will have a place in IBD in a patient- or family-based settings where features of the clinical presentation

suggest a likely monogenic, Mendelian basis for the disease. Personalized medicine based on the patient's genome in these carefully selected cases is no longer a far-off dream but a nascent reality. More uncertain are the prospects of large-scale exome sequencing projects for discovery of population-scale heritability for such a common and highly polygenic disease. Theoretical arguments can be made to support either position, but the debate can only be resolved by experimental testing of the common disease-rare variant hypothesis. Exome sequencing of rare variants may not collectively yield much explanation of the population attributable risk of disease, but it has great potential to highlight the key players in the pathogenesis of disease along with variants amenable to functional study and thereby influence the development of potent new therapeutics.

REFERENCES

1 **Taylor KD**, Rotter JI, Yang HY. Genetics of inflammatory bowel disease. 2nd ed. Targan SR, Shanahan F, Karp LC, editors. Inflammatory bowel disease: from bench to bedside. New York: Springer, 2005: 21-65

2 **Halme L**, Paavola-Sakki P, Turunen U, Lappalainen M, Farkkila M, Kontula K. Family and twin studies in inflammatory bowel disease. *World J Gastroenterol* 2006; **12**: 3668-3672 [PMID: 16773682]

3 **Hugot JP**, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599-603 [PMID: 11385576 DOI: 10.1038/35079107]

4 **Ogura Y**, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nuñez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**: 603-606 [PMID: 11385577 DOI: 10.1038/35079114]

5 **Hugot JP**, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996; **379**: 821-823 [PMID: 8587604 DOI: 10.1038/379821a0]

6 **Duerr RH**, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006; **314**: 1461-1463 [PMID: 17068223 DOI: 10.1126/science.1135245]

7 **Hampe J**, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Günther S, Prescott NJ, Onnie CM, Häsler R, Sipos B, Fölsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007; **39**: 207-211 [PMID: 17200669 DOI: 10.1038/ng1954]

8 **Burton PR**, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Barrett JC, Burton PR, Davison D, Donnelly P, Easton D, Evans D, Leung HT, Marchini JL, Morris AP, Spencer CC, Tobin MD, Cardon LR, Clayton DG, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Todd JA, Ouwehand WH, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskvina V, Nikolov I, O'Donovan MC, Owen MJ, Craddock N, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop DT, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Burton PR, Dixon RJ, Mangino M, Suzanne S, Tobin MD, Thompson JR, Samani NJ, Bredin F, Tremelling M, Parkes M, Drummond H, Lees CW, Nimmo ER, Satsangi J, Fisher SA, Forbes A, Lewis CM, Onnie CM, Prescott NJ, Sanderson J, Mathew CG, Barbour J, Mohiuddin MK, Todhunter CE, Mansfield JC, Ahmad T, Cummings FR, Jewell DP, Webster J, Brown MJ, Clayton DG, Lathrop GM, Connell J, Dominczak A, Samani NJ, Marcano CA, Burke B, Dobson R, Gungadoo J, Lee KL, Munroe PB, Newhouse SJ, Onipinla A, Wallace C, Xue M, Caulfield M, Farrall M, Barton A, Bruce IN, Donovan H, Eyre S, Gilbert PD, Hider SL, Hinks AM, John SL, Potter C, Silman AJ, Symmmons DP, Thomson W, Worthington J, Clayton DG, Dunger DB, Nutland S, Stevens HE, Walker NM, Widmer B, Todd JA, Frayling TA, Freathy RM, Lango H, Perry JR, Shields BM, Weedon MN, Hattersley AT, Hitman GA, Walker M, Elliott KS, Groves CJ, Lindgren CM, Rayner NW, Timpson NJ, Zeggini E, McCarthy MI, Newport M, Sirugo G, Lyons E, Vannberg F, Hill AV, Bradbury LA, Farrar C, Pointon JJ, Wordsworth P, Brown MA, Franklyn JA, Heward JM, Simmonds MJ, Gough SC, Seal S, Stratton MR, Rahman N, Ban M, Goris A, Sawcer SJ, Compston A, Conway D, Jallow M, Newport M, Sirugo G, Rockett KA, Kwiatowski DP, Bumpstead SJ, Chaney A, Downes K, Ghori MJ, Gwilliam R, Hunt SE, Inouye M, Keniry A, King E, McGinnis R, Potter S, Ravindrarajah R, Whittaker P, Widden C, Withers D, Deloukas P, Leung HT, Nutland S, Stevens HE, Walker NM, Todd JA, Easton D, Clayton DG, Burton PR, Tobin MD, Barrett JC, Evans D, Morris AP, Cardon LR, Cardin NJ, Davison D, Ferreira T, Pereira-Gale J, Hallgrimsdottir IB, Howie BN, Marchini JL, Spencer CC, Su Z, Teo YY, Vukcevic D, Donnelly P, Bentley D, Brown MA, Gordon LR, Caulfield M, Clayton DG, Compston A, Craddock N, Deloukas P, Donnelly P, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Kwiatkowski DP, Mathew C, McCarthy MI, Ouwehand WH, Parkes M, Pembrey M, Rahman N, Samani NJ, Stratton MR, Todd JA, Worthington J. Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661-678 [PMID: 17554300 DOI: 10.1038/nature05911]

9 **Parkes M**, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Cardon L, Mathew CG. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007; **39**: 830-832 [PMID: 17554261 DOI: 10.1038/ng2061]

10 **Imielinski M**, Baldassano RN, Griffiths A, Russell RK, Annese V, Dubinsky M, Kugathasan S, Bradfield JP, Walters TD, Sleiman P, Kim CE, Muise A, Wang K, Glessner JT, Saeed S, Zhang H, Frackelton EC, Hou C, Flory JH, Otieno G, Chiavacci RM, Grundmeier R, Castro M, Latiano A, Dallapiccola B, Stempak J, Abrams DJ, Taylor K, McGovern D, Silber G, Wrobel I, Quiros A, Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmuda MM, Bitton A, Dassopoulos T, Datta LW, Green T,

Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwillam R, Tremelling M, Delukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ, Heyman MB, Ferry GD, Kirschner B, Lee J, Essers J, Grand R, Stephens M, Levine A, Piccoli D, Van Limbergen J, Cucchiara S, Monos DS, Guthery SL, Denson L, Wilson DC, Grant SF, Daly M, Silverberg MS, Satsangi J, Hakonarson H. Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat Genet* 2009; **41**: 1335-1340 [PMID: 19915574 DOI: 10.1038/ng.489]

11 **Kugathasan S**, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, Cucchiara S, Kim CE, Frackelton EC, Annaiah K, Glessner JT, Santa E, Willson T, Eckert AW, Bonkowski E, Shaner JL, Smith RM, Otieno FG, Peterson N, Abrams DJ, Chiavacci RM, Grundmeier R, Mamula P, Tomer G, Piccoli DA, Monos DS, Annese V, Denson LA, Grant SF, Hakonarson H. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* 2008; **40**: 1211-1215 [PMID: 18758464 DOI: 10.1038/ng.203]

12 **Marchini J**, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906-913 [PMID: 17572673 DOI: 10.1038/ng2088]

13 **Jostins L**, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, Essers J, Mitrovic M, Ning K, Cleynen I, Theatre E, Spain SL, Raychaudhuri S, Goyette P, Wei Z, Abraham C, Achkar JP, Ahmad T, Amininejad L, Ananthakrishnan AN, Andersen V, Andrews JM, Baidoo L, Balschun T, Bampton PA, Bitton A, Boucher G, Brand S, Büning C, Cohain A, Cichon S, D'Amato M, De Jong D, Devaney KL, Dubinsky M, Edwards C, Ellinghaus D, Ferguson LR, Franchimont D, Fransen K, Gearry R, Georges M, Gieger C, Glas J, Haritunians T, Hart A, Hawkey C, Hedl M, Hu X, Karlsen TH, Kupcinskas L, Kugathasan S, Latiano A, Laukens D, Lawrance IC, Lees CW, Louis E, Mahy G, Mansfield J, Morgan AR, Mowat C, Newman W, Palmieri O, Ponsioen CY, Potocnik U, Prescott NJ, Regueiro M, Rotter JI, Russell RK, Sanderson JD, Sans M, Satsangi J, Schreiber S, Simms LA, Sventoraityte J, Targan SR, Taylor KD, Tremelling M, Verspaget HW, De Vos M, Wijmenga C, Wilson DC, Winkelmann J, Xavier RJ, Zeissig S, Zhang B, Zhang CK, Zhao H, Silverberg MS, Annese V, Hakonarson H, Brant SR, Radford-Smith G, Mathew CG, Rioux JD, Schadt EE, Daly MJ, Franke A, Parkes M, Vermeire S, Barrett JC, Cho JH. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012; **491**: 119-124 [PMID: 23128233 DOI: 10.1038/nature11582]

14 **Maurano MT**, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190-1195 [PMID: 22955828 DOI: 10.1126/science.1222794]

15 **Wang K**, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, Annese V, Dubinsky M, Rotter JI, Russell RK, Bradfield JP, Sleiman PM, Glessner JT, Walters T, Hou C, Kim C, Frackelton EC, Garris M, Doran J, Romano C, Catassi C, Van Limbergen J, Guthery SL, Denson L, Piccoli D, Silverberg MS, Stanley CA, Monos D, Wilson DC, Griffiths A, Grant SF, Satsangi J, Polychronakos C, Hakonarson H. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum Mol Genet* 2010; **19**: 2059-2067 [PMID: 20176734 DOI: 10.1093/hmg/ddq078]

16 **Manolio TA**, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747-753 [PMID: 19812666 DOI: 10.1038/nature08494]

17 **Reich DE**, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502-510 [PMID: 11525833]

18 **Bodmer W**, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**: 695-701 [PMID: 18509313 DOI: 10.1038/ng.f.136]

19 **Pritchard JK**. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**: 124-137 [PMID: 11404818 DOI: 10.1086/321272]

20 **Tennessen JA**, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012; **337**: 64-69 [PMID: 22604720 DOI: 10.1126/science.1219240]

21 **Choi M**, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009; **106**: 19096-19101 [PMID: 19861545 DOI: 10.1073/pnas.0910672106]

22 **Li H**, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-1760 [PMID: 19451168 DOI: 10.1093/bioinformatics/btp324]

23 **DePristo MA**, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-498 [PMID: 21478889 DOI: 10.1038/ng.806]

24 **McKenna A**, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-1303 [PMID: 20644199 DOI: 10.1101/gr.107524.110]

25 **Wang K**, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164 [PMID: 20601685 DOI: 10.1093/nar/gkq603]

26 **Ng PC**, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001; **11**: 863-874 [PMID: 11337480 DOI: 10.1101/gr.176601]

27 **Ng PC**, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812-3814 [PMID: 12824425]

28 **Adzhubei IA**, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248-249 [PMID: 20354512 DOI: 10.1038/nmeth0410-248]

29 **Fu W**, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, Akey JM. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013; **493**: 216-220 [PMID: 23201682 DOI: 10.1038/na-

ture11690]
30 **Sherry ST**, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999; **9**: 677-679 [PMID: 10447503]
31 **Li B**, Liu DJ, Leal SM. Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Genet* 2013; **Chapter 1**: Unit 1.26 [PMID: 23853079 DOI: 10.1002/0471142905.hg0126s78]
32 **Neale BM**, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. Testing for an unusual distribution of rare variants. *PLoS Genet* 2011; **7**: e1001322 [PMID: 21408211 DOI: 10.1371/journal.pgen.1001322]
33 **Worthey EA**, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 2011; **13**: 255-262 [PMID: 21173700 DOI: 10.1097/GIM.0b013e3182088158]
34 **Glocker EO**, Kotlarz D, Boztug K, Gertz EM, Schäffer AA, Noyan F, Perro M, Diestelhorst J, Allroth A, Murugan D, Hätscher N, Pfeifer D, Sykora KW, Sauer M, Kreipe H, Lacher M, Nustede R, Woellner C, Baumann U, Salzer U, Koletzko S, Shah N, Segal AW, Sauerbrey A, Buderus S, Snapper SB, Grimbacher B, Klein C. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med* 2009; **361**: 2033-2045 [PMID: 19890111 DOI: 10.1056/NEJMoa0907206]
35 **Moran CJ**, Walters TD, Guo CH, Kugathasan S, Klein C, Turner D, Wolters VM, Bandsma RH, Mouzaki M, Zachos M, Langer JC, Cutz E, Benseler SM, Roifman CM, Silverberg MS, Griffiths AM, Snapper SB, Muise AM. IL-10R polymorphisms are associated with very-early-onset ulcerative colitis. *Inflamm Bowel Dis* 2013; **19**: 115-123 [PMID: 22550014 DOI: 10.1002/ibd.22974]
36 **Cohen J**, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 2005; **37**: 161-165 [PMID: 15654334 DOI: 10.1038/ng1509]
37 **Agrawal N**, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Treviño L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 2011; **333**: 1154-1157 [PMID: 21798897 DOI: 10.1126/science.1206923]
38 **Varela I**, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, Tuveson DA, Perez-Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell PJ, Teh BT, Stratton MR, Futreal PA. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 2011; **469**: 539-542 [PMID: 21248752 DOI: 10.1038/nature09639]
39 **Chin L**, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med* 2011; **17**: 297-303 [PMID: 21383744 DOI: 10.1038/nm.2323]
40 **Kiezun A**, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Lehner T, Shugart YY, Price AL, de Bakker PI, Purcell SM, Sunyaev SR. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012; **44**: 623-630 [PMID: 22641211 DOI: 10.1038/ng.2303]
41 **Romeo S**, Yin W, Kozlitina J, Pennacchio LA, Boerwinkle E, Hobbs HH, Cohen JC. Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest* 2009; **119**: 70-79 [PMID: 19075393 DOI: 10.1172/JCI37118]
42 **Johansen CT**, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 2010; **42**: 684-687 [PMID: 20657596 DOI: 10.1038/ng.628]
43 **Ji W**, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008; **40**: 592-599 [PMID: 18391953 DOI: 10.1038/ng.118]
44 **Nejentsev S**, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009; **324**: 387-389 [PMID: 19264985 DOI: 10.1126/science.1167728]
45 **Momozawa Y**, Mni M, Nakamura K, Coppieters W, Almer S, Amininejad L, Cleynen I, Colombel JF, de Rijk P, Dewit O, Finkel Y, Gassull MA, Goossens D, Laukens D, Lémann M, Libioulle C, O'Morain C, Reenaers C, Rutgeerts P, Tysk C, Zelenika D, Lathrop M, Del-Favero J, Hugot JP, de Vos M, Franchimont D, Vermeire S, Louis E, Georges M. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat Genet* 2011; **43**: 43-47 [PMID: 21151126 DOI: 10.1038/ng.733]
46 **Rivas MA**, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N, Fennell T, Kirby A, Latiano A, Goyette P, Green T, Halfvarson J, Haritunians T, Korn JM, Kuruvilla F, Lagacé C, Neale B, Lo KS, Schumm P, Törkvist L, Dubinsky MC, Brant SR, Silverberg MS, Duerr RH, Altshuler D, Gabriel S, Lettre G, Franke A, D'Amato M, McGovern DP, Cho JH, Rioux JD, Xavier RJ, Daly MJ. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011; **43**: 1066-1073 [PMID: 21983784 DOI: 10.1038/ng.952]
47 **Cardinale CJ**, Wei Z, Panossian S, Wang F, Kim CE, Mentch FD, Chiavacci RM, Kachelries KE, Pandey R, Grant SFA, Baldassano RN, Hakonarson H. Targeted resequencing identifies defective variants of Decoy Receptor 3 in pediatric-onset inflammatory bowel disease. *Genes Immun* 2013; In press [DOI: 10.1038/gene.2013.43]
48 **Mathieson I**, McVean G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* 2012; **44**: 243-246 [PMID: 22306651 DOI: 10.1038/ng.1074]
49 **Liu DJ**, Leal SM. Replication strategies for rare variant complex trait association studies via next-generation sequencing. *Am J Hum Genet* 2010; **87**: 790-801 [PMID: 21129725 DOI: 10.1016/j.ajhg.2010.10.025]
50 **Hunt KA**, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SC, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, Trembath RC, Walker NM, Wozniak E, Todd JA, Simpson MA, Plagnol V, van Heel DA. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* 2013;

498: 232-235 [PMID: 23698362 DOI: 10.1038/nature12170]
51 **Terwilliger JD**, Hiekkalinna T. An utter refutation of the "fundamental theorem of the HapMap". *Eur J Hum Genet* 2006; **14**: 426-437 [PMID: 16479260 DOI: 10.1038/sj.ejhg.5201583]
52 **Pan W**, Shen X. Adaptive tests for association analysis of rare variants. *Genet Epidemiol* 2011; **35**: 381-388 [PMID: 21520272 DOI: 10.1002/gepi.20586]
53 **Wu MC**, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82-93 [PMID: 21737059 DOI: 10.1016/j.ajhg.2011.05.029]
54 **Price AL**, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 2010; **86**: 832-838 [PMID: 20471002 DOI: 10.1016/j.ajhg.2010.04.005]
55 **Ellinghaus D**, Zhang H, Zeissig S, Lipinski S, Till A, Jiang T, Stade B, Bromberg Y, Ellinghaus E, Keller A, Rivas MA, Skieceviciene J, Doncheva NT, Liu X, Liu Q, Jiang F, Forster M, Mayr G, Albrecht M, Häsler R, Boehm BO, Goodall J, Berzuini CR, Lee J, Andersen V, Vogel U, Kupcinskas L, Kayser M, Krawczak M, Nikolaus S, Weersma RK, Ponsioen CY, Sans M, Wijmenga C, Strachan DP, McArdle WL, Vermeire S, Rutgeerts P, Sanderson JD, Mathew CG, Vatn MH, Wang J, Nöthen MM, Duerr RH, Büning C, Brand S, Glas J, Winkelmann J, Illig T, Latiano A, Annese V, Halfvarson J, D' Amato M, Daly MJ, Nothnagel M, Karlsen TH, Subramani S, Rosenstiel P, Schreiber S, Parkes M, Franke A. Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* 2013; **145**: 339-347 [PMID: 23624108 DOI: 10.1053/j.gastro.2013.04.040]
56 **Wei Z**, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, Kim C, Mentch F, Van Steen K, Visscher PM, Baldassano RN, Hakonarson H; the International IBD Genetics Consortium. Large Sample Size, Wide Variant Spectrum, and Advanced Machine-Learning Technique Boost Risk Prediction for Inflammatory Bowel Disease. *Am J Hum Genet* 2013; Epub ahead of print [PMID: 23731541 DOI: 10.1016/j.ajhg.2013.05.002]
57 **Blaydon DC**, Biancheri P, Di WL, Plagnol V, Cabral RM, Brooke MA, van Heel DA, Ruschendorf F, Toynbee M, Walne A, O'Toole EA, Martin JE, Lindley K, Vulliamy T, Abrams DJ, MacDonald TT, Harper JI, Kelsell DP. Inflammatory skin and bowel disease linked to ADAM17 deletion. *N Engl J Med* 2011; **365**: 1502-1508 [PMID: 22010916 DOI: 10.1056/NEJMoa1100721]
58 **Russell RK**, Drummond HE, Nimmo ER, Anderson N, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset WM, Mahdi G, Satsangi J. The contribution of the DLG5 113A variant in early-onset inflammatory bowel disease. *J Pediatr* 2007; **150**: 268-273 [PMID: 17307543 DOI: 10.1016/j.jpeds.2006.12.010]
59 **Zhang W**, Hui KY, Gusev A, Warner N, Ng SM, Ferguson J, Choi M, Burberry A, Abraham C, Mayer L, Desnick RJ, Cardinale CJ, Hakonarson H, Waterman M, Chowers Y, Karban A, Brant SR, Silverberg MS, Gregersen PK, Katz S, Lifton RP, Zhao H, Nuñez G, Pe'er I, Peter I, Cho JH. Extended haplotype association study in Crohn's disease identifies a novel, Ashkenazi Jewish-specific missense mutation in the NF-κB pathway gene, HEATR3. *Genes Immun* 2013; **14**: 310-316 [PMID: 23615072 DOI: 10.1038/gene.2013.19]
60 **Alangari A**, Alsultan A, Adly N, Massaad MJ, Kiani IS, Aljebreen A, Raddaoui E, Almomen AK, Al-Muhsen S, Geha RS, Alkuraya FS. LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency. *J Allergy Clin Immunol* 2012; **130**: 481-488.e2 [PMID: 22721650 DOI: 10.1016/j.jaci.2012.05.043]
61 **Brant SR**, Panhuysen CI, Nicolae D, Reddy DM, Bonen DK, Karaliukas R, Zhang L, Swanson E, Datta LW, Moran T, Ravenhill G, Duerr RH, Achkar JP, Karban AS, Cho JH. MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *Am J Hum Genet* 2003; **73**: 1282-1292 [PMID: 14610718 DOI: 10.1086/379927]
62 **Muise AM**, Xu W, Guo CH, Walters TD, Wolters VM, Fattouh R, Lam GY, Hu P, Murchie R, Sherlock M, Gana JC, Russell RK, Glogauer M, Duerr RH, Cho JH, Lees CW, Satsangi J, Wilson DC, Paterson AD, Griffiths AM, Silverberg MS, Brumell JH. NADPH oxidase complex and IBD candidate gene studies: identification of a rare variant in NCF2 that results in reduced binding to RAC2. *Gut* 2012; **61**: 1028-1035 [PMID: 21900546 DOI: 10.1136/gutjnl-2011-300078]
63 **Peltekova VD**, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X, Newman B, Van Oene M, Cescon D, Greenberg G, Griffiths AM, St George-Hyslop PH, Siminovitch KA. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004; **36**: 471-475 [PMID: 15107849 DOI: 10.1038/ng1339]

P- Reviewers Decorti G, Fitzpatrick LR, Gazouli M, Yamamoto S
S- Editor Gou SX **L- Editor** A **E- Editor** Liu XM

百世登

Baishideng®

Published by **Baishideng Publishing Group Co., Limited**
Flat C, 23/F., Lucky Plaza,
315-321 Lockhart Road, Wan Chai, Hong Kong, China
Fax: +852-65557188
Telephone: +852-31779906
E-mail: bpgoffice@wjgnet.com
http://www.wjgnet.com

ISSN 1007-9327

© 2013 Baishideng. All rights reserved.