HMG Advance Access published October 2, 2007

An insertion-deletion polymorphism in the Interferon Regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases

Vinciane Dideberg [1,6], Gudlaug Kristjansdottir [6], Lili Milani [6], Cécile Libioulle [3], Snaevar Sigurdsson [6], Edouard Louis [2], Ann-Christin Wiman [6], Séverine Vermeire [4], Paul Rutgeerts [4], Jacques Belaiche [2], Denis Franchimont [5], André Van Gossum [5], Vincent Bours [1], and Ann-Christine Syvänen [6*].

Departments of [1]Human Genetics, [2]Hepatology and Gastroenterology, CHU de Liège and GIGA-R,University of Liège, [3]Factorial and Molecular Genetics, GIGA-R, University of Liège, Liège; [4]Department of Gastroenterology, UZ Gasthuisberg, Leuven; [5]Department of Gastroenterology , Erasme University Hospital, Brussels, Belgium and [6]Molecular Medicine, Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

[*]To whom correspondence should be addressed at:

Department of Medical Sciences, Entrance 70, Third Floor, Research Department 2, Uppsala University, Academic Hospital, 75185-Uppsala, Sweden.

Tel: +46-18-6112959

Fax: +46-18-5536 01

E-mail: Ann-Christine.Syvanen@medsci.uu.se

Abstract

The interferon regulatory factor 5 (IRF5) gene encodes a transcription factor that plays an important role in the innate as well as in the cell-mediated immune responses. The IRF5 gene has been shown to be associated with systemic lupus erythematosus and rheumatoid arthritis. We studied whether the IRF5 gene is also associated with inflammatory bowel diseases (IBD), Crohn disease (CD) and ulcerative colitis (UC). Twelve polymorphisms in the IRF5 gene were genotyped in a cohort of 1007 IBD patients (748 CD and 241 UC) and 241 controls from Wallonia, Belgium. The same polymorphisms were genotyped in a confirmatory cohort of 311 controls and 687 IBD patients (488 CD and 192 UC) from Leuven, Belgium. A strong signal of association ($p = 1.9 \times 10^{-5}$, OR: 1.81 (1.37-2.39)) with IBD was observed for a 5bp indel (CGGGG) polymorphism in the promoter region of the IRF5 gene. The association was detectable ($p = 6.8 \times 10^{-4}$) also in CD patients, and was particularly strong among the UC patients ($p = 5.3 \times 10^{-8}$, OR 2.42 (1.76 -3.34)). The association of the CGGGG indel was confirmed in the second cohort ($p = 3.2 \times 10^{-5}$, OR 1.59 (1.28 – 1.98)). The insertion of one CGGGG unit is predicted to create an additional binding site for the transcription factor SP1. Using an electrophoretic mobility shift assay we show allele-specific differences in protein binding to this repetitive DNA-stretch, which suggest a potential function role for the CGGGG indel.

Introduction

The chronic inflammatory bowel diseases (IBD), Crohn disease (CD) and ulcerative colitis (UC), are common causes of gastrointestinal morbidity in Western countries. These diseases are caused by the interaction of genetic, immunologic and environmental factors (1). The role of genetic factors in IBD is supported by a strong familial clustering of the disease and by the significantly higher disease concordance between MZ twins than between DZ twins (2-6). A genetic predisposition to CD has been demonstrated and validated for few genes (CARD15, DLG5, IL23R, ATG16L1), and several genomic regions were high lighted by recent genome-wide association (GWA) scans with SNP markers (7-9). The associations of the CARD15 and the IL23R genes with CD were recently confirmed by a GWA study performed on a large case-control population by the Welcome Trust Case Control Consortium (WTCCC) (10). Moreover, this study identified four new strong association signals at the chromosomal locations 3p21, 5q33, 10q24 and 18p11, which were subsequently replicated in an independent cohort of CD patients (11).

In addition to being shared within the two IBD disease groups, CD and UC, some genes could also be involved in the pathogenesis of other autoimmune diseases. Autoimmune diseases affect up to 5% of the Western population, and they are characterized by a loss of self-tolerance leading to immune-mediated tissue destruction. Autoimmune diseases are multifactorial diseases that may have a shared genetic background, as supported by a higher frequency of immune–mediated disorders in families with immune diseases than in the general population and by the higher rate of co-occurrence of immune diseases in patients affected by immune disorders (12, 13). Some loci or genes are known to be shared between auto-immune diseases: the HLA region, the cytotoxic T-lymphocyte

antigen-4 (CTLA-4) gene in Graves' disease (GD), Hashimoto thyroiditis and type 1 diabetes (T1D) or the PTPN22 gene in rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), GD and T1D (14-16).The recent WTCCC GWA study reported several loci implicated in more than one disease. By grouping three auto-immune diseases (RA, CD and T1D), four regions on chromosomes 4, 10, 12 and 18 showed high association signals (10).Although the pathogenesis of IBD is uncertain, IBD can be considered as autoimmune diseases, or maybe more appropriately as “immune dysregulation diseases”(17). Mutations in genes implicated in the immune regulation and response could lead to an immune dysregulation that predisposes to immune diseases.

The interferon (IFN) regulatory factor (IRF) family comprises nine transcription factors involved in the defense against microbes, in cellular survival and hematopoietic development. Some members of the IRF family are activated in response to infections and are consequently involved in innate immunity (18-20). IRF3 and IRF7 are direct transducers of virus-mediated signaling, whereas IRF5 has a role in the response to kinases involved in viral infections or in Toll-like-receptor (TLR) signaling (21, 22). IRF5 acts as repressor or activator of type I interferon genes, but also as inducer of pro-inflammatory cytokines, such as interleukin-6 (IL-6), IL-12 and tumor-necrosis factor-alpha. IRF5 is crucial for the induction of these proinflammatory cytokines by the TLR – MYD88 pathway and plays a central role in the recently described synergism between the TLRs (18, 23-26). The implication of these cytokines in IBD is well documented, and has led to the development of specific therapeutic agents against these cytokines or their receptors (27).

The IRF5 gene, located on 7q32, displays a strong association with SLE (28). This association is well established and has been replicated in several independent patient cohorts (29). Moreover the IRF5 gene has recently been shown to be associated with rheumatoid arthritis (30). Given the role of IRF5 in innate and cell-mediated immunity and the described association with two important autoimmune diseases, SLE and RA, we postulated that the IRF5 gene could be part of the genetic background that leads to the development of multiple immune diseases, including IBD. To test this hypothesis, twelve polymorphisms in the IRF5 gene were genotyped in IBD patients from Belgium. Six of the polymorphisms displayed an association with IBD, with an exceptionally strong association signal for a 5 bp insertion-deletion polymorphism (CGGGG indel) in the promoter region of the gene. This association was observed with both CD and UC, and we were able to confirm these results in an independent cohort of IBD patients. We also demonstrated differential protein binding to the two alleles of the CGGGG indel using an electrophoretic mobility shift assay.

Results

Twelve polymorphisms in the IRF5 gene were tested for their association with IBD, CD or UC, in a cohort of Caucasian patients and controls from Wallonia in Belgium. These polymorphisms were selected for our study based on their previously described association with SLE or RA (28-30), or their potential functional roles (Table 1). The polymorphisms analyzed were 10 single nucleotide polymorphisms (SNPs) and two insertion/deletion polymorphisms (indels). Five SNPs (rs729302, rs4728142, rs3757385, rs2004640, rs3807306) and a 5 bp indel (CGGGG) are located in the promoter or first intron of the IRF5 gene. The CGGGG indel was included as a potentially functional variant because it is predicted to alter a binding site for the transcription factor SP1 ("TFSEARCH: Searching Transcription Factor Binding Sites", http://www.cbrc.jp/research/db/TFSEARCH.html/) (31). The other indel is a deletion of 30 bp in exon 6 that removes ten amino acids from the IRF5 protein (32). Figure 1 shows the location of the SNPs and indels in the IRF5 gene.

We found that six of the polymorphisms, including SNPs rs4728142, rs3807306, rs10954213, rs11770589 and both indels appear to be associated with IBD ($p < 0.05$), with the strongest association signal for the CGGGG indel ($p = 1.9 \times 10^{-5}$). The association analysis for each polymorphisms reveals strong association signals for the CGGGG indel in the UC subgroup ($p = 5.3 \times 10^{-8}$, OR 2.43 (1.76-3.34)). Association signals were also observed in the CD patients, particularly for the CGGGG indel ($p = 6.8 \times 10^{-4}$, OR 1.64 (1.23-2.17)) (Table 2). To confirm the association between the IRF5 polymorphisms and IBD, they were genotyped in an independent cohort of IBD patients

from Leuven, Belgium, and tested for their association with IBD, CD and UC (Table 3). The CGGGG indel that was the most strongly associated polymorphism in the cohort from Wallonia was also associated in the Leuven cohort ($p = 3.2 \times 10^{-5}$, OR 1,59 (1.28 - 1.98)), while no association signal was observed for the other polymorphisms. In the combined analysis of both cohorts (1661 IBD patients), the association signals from the CGGGG indel and the SNP rs4728142 are maintained, with a strong signal for the CGGGG indel ($p= 1.4 \times 10^{-8}$ OR 1.62 (1.37-1.92)). Both patient subgroups (CD and UC) were also considered in the combined analysis, with strong association signals for the CGGGG indel in both subgroups ($p = 3.3 \times 10^{-6}$ in CD and $p = 7.9 \times 10^{-10}$ in UC), and also a strong signal in UC for the SNP rs4728142 ($p = 4.2 \times 10^{-5}$). Table 4 shows the p-values and odds ratio (OR) for all the analyzed polymorphisms when the data from both cohorts are combined.

It is notable that two of the SNPs, rs2070197 located in the 3'UTR and rs12539741 located 5 kb downstream of the IRF5 gene do not show any signals for association with IBD in our cohorts, although these two SNPs which are in perfect LD with each other have been found to be strongly associated with SLE (see Table 1 and references therein). These two SNPs are in perfect LD with the HapMap SNP rs10488631 (32), which is in perfect LD with several additional HapMap SNPs distributed over a 100 kb region located in 3' of the IRF5 gene which contains the transportin 3 (TNPO3) gene. The pair-wise LD between these SNPs and the polymorphism that was most strongly associated with IBD, the CGGGG indel, is low with an r^2-value of 0.10 in our control groups from both cohorts (Figure 2). Figure 2 also shows that this CGGGG indel is correlated to the SNPs rs4728142 ($r^2 = 0.77$) and rs3807306 ($r^2 = 0.65$), that showed association signals

with IBD, and that the pairwise LD-values between these SNPs and the SNPs rs2070197 and rs12539741 is low (mean $r^2 = 0.10$, maximum 0.11).

We performed haplotype analysis of the six IRF5 polymorphisms located in the promoter and first intron of the IRF5 gene. Four major haplotypes (frequency > 10%) were identified. The haplotype analysis did not reveal any association signals that were stronger than the signal from the 5 bp CGGGG indel alone (data not shown). The frequency of the minor allele, corresponding to the insertion of CGGGG (the 4 x CGGGG allele) in the promoter region of the IRF5 gene, is higher in IBD patient than in the controls. A single copy of the minor (risk) allele was associated with IBD with an OR of 1.62 (1.37-1.91), while two copies of the risk allele were associated with an OR of 3.03 (2.04-4.49). Thus the CGGGG insertion confers a very strong risk for IBD, CD or UC in the Belgian population.

Given the *in silico* prediction that the insertion of CGGGG alters a binding site for the transcription factor SP1, we used EMSA to test for differential protein binding to the two alleles of the CGGGG indel using probes specific for 4 and 3 CGGGG repeat units. The EMSA revealed a clearly higher level of protein binding to the 4 x CGGGG probe than to the 3 x CGGGG probe (Figure 3)

Discussion

A genetic etiology in IBD is well established, but the complete set of the underlying genetic variants and their epistasis remain to be elucidated. For about 30 years, genetic factors predisposing to IBD have been searched for, beginning with the implication of different HLA phenotypes in the predisposition to CD (33). Recent GWA studies have identified several loci and genes that are putatively associated with CD (7-9, 34). Genome wide linkage studies and association studies of "CD susceptibility genes" on UC cohorts have demonstrated that CD and UC share some, but not all disease-predisposing genes (35, 36). For example, associations of the CARD15 and ATG16L genes have been reported only with CD, while an association signal for IL23R gene validated in CD appears also, although weaker, in UC (6, 37, 38). Here we report an association of polymorphisms in the IRF5 gene with both CD and UC. IRF5 is the first gene associated with UC to be validated in a confirmatory cohort, where a stronger association signal is observed in UC (OR = 1.93 (1.56 – 2.38)) than in CD (1.51 (1.27 – 1.80)). Several loci predisposing to IBD have already been explored in SLE, such as the IBD5 locus, the CARD15 gene and the discs large homolog 5 gene (DLG5), but no associations with SLE have been confirmed for these genes (39-41). Hence, our study provides the first confirmed evidence of a shared gene between SLE and IBD. Our findings of association with IBD diverge from those in SLE, where one group of SNPs in the promoter and first intron of the IRF5 gene and another group of SNPs tagged by the SNP rs2070197 or its proxies in the 3'-region downstream of the IRF5 gene, which contains the TNPO3 gene, seem to be independently associated with the disease (42). In our study we observe association signals with IBD from SNPs that belong to the first group of SNPs only,

while we do not observe any association between IBD and the latter group of SNPs. Thus our data indicate clearly the IRF5 gene as susceptibility factor for IBD, and reveal an unlikely role in IBD for SNPs located 3' of IRF5 and in the TNPO3 gene. A haplotype formed by four linked SNPs that belong to the first group of SNPs in the promoter region of IRF5 is associated with RA (30), but none of the SNPs located in the 3'-region of the IRF5 gene that are tagged by the SNP rs2070197 has to our knowledge been tested for their association with RA.

We observed the strongest signal of association with IBD for a 5bp CGGGG indel in the promoter region of IRF5. This CGGGG indel is part of a polymorphic repetitive DNA region that contains either 3 or 4 CGGGG units, where the insertion of an additional CGGGG unit (the 4 x CGGGG allele) is the risk allele for IBD. A possible function of the insertion of CGGGG as a cis-acting regulatory element is suggested by the prediction that it creates an additional binding site for the transcription factor SP1. In this polymorphic region 3 CGGGG repeats constitute 2 SP1 binding sites, and 4 CGGGG constitute 3 SP1 binding sites. Our data from EMSA performed on nuclear extract from blood cells in the present study support this prediction. We speculate that the creation of an additional binding site for the transcription factor SP1, could increase transcription of the IRF5 gene. However further functional studies are necessary to clarify if there is an increased expression of IRF5 as a consequence of a specific enhanced binding of SP1 to the 4 x CGGGG allele of IRF5 in tissues that are relevant for IBD. Indirect support for this suggestion comes from previously shown, constitutive binding of SP1 at two adjacent sites in IRF4, which has been observed in HTLV-I-infected cells (43). Constitutive binding of SP1 to IRF1 has also been observed, and is required for G1 activation of the

transcription in the prolactin signalling (44). A similar sequence (GGCGGGG) is present in the promoter of three IRF genes (IRF1, IRF4 and IRF5). The specific binding of SP1 to these sequences shown by supershift assays on IRF1 and IRF4 support the predicted binding of SP1 to the IRF5 gene (43, 44). Considering the role of IRF5 in the TLR pathway, increased expression of IRF5 could lead to an increased production of pro-inflammatory cytokines and the perpetuation of inflammation, but further studies on the role of IRF5 in the pathogenesis of IBD are required.

The association of IRF5 with SLE and RA has been reported (28, 30), but the causal variant(s) have not yet been identified in these diseases. Several pathogenic mechanisms have been proposed, such as differential immune response depending on the expressed IRF5 isoforms, enhanced cytokine production or alteration of apoptosis (30, 32). The polymorphisms in IRF5 are correlated and different polymorphisms or combinations of them could also result in different pathologies, demonstrating the complexity of the role of IRF5 in immune diseases. Here we show a strong association of IRF5 with two other immune diseases, CD and UC. A more complete characterization of the genetic variation of IRF5 genetic and functional studies are now necessary to fully understand the underlying mechanisms by which IRF5 is involved in IBD.

Materials and methods

Cohorts

Our cohort contained 1007 IBD patients from Wallonia, Belgium, of which 748 were diagnosed with Crohn disease (CD), 254 with ulcerative colitis (UC) and five with indeterminate colitis (IC). Healthy individuals (N = 241) attending the University Hospital of Liège and blood donors were used as controls. A second cohort of 687 IBD patients from Leuven, of which 488 were CD patients, 192 were UC patients and 7 were IC was also analyzed. A set of 311 volunteer blood donors served as unrelated control individuals for these patients. All patients and controls included in the study gave their informed consent. Ethical approvals for the study were obtained from the ethics committees of the University Hospitals of Liège and Leuven.

Genotyping

Ten SNPs and two insertion-deletion polymorphisms (indels) in the IRF5 gene were genotyped. Nine SNPs (rs729302 rs3757385, rs2004640, rs3807306, rs2070197, rs10954213, rs11770589, rs2280714, rs12539741) were genotyped by fluorescent minisequencing using the multiplex SNPstream system (Beckman Coulter) (45). The SNP rs4728142 was genotyped by a homogeneous template directed-dye terminator assay with fluorescence polarization detection (FP-TDI [Perkin Elmer]) (46). The two indels, a CGGGG indel in the promoter region of the IRF5 gene and a 30 bp indel in exon 6 of the IRF5 gene were genotyped after amplification with a fluorescent PCR primer, and the amplified fragments were analyzed using an ABI3100 sequencer (Applied Biosystems, Foster City, USA). The fragment analysis was performed using the

GeneMapper v.3.7. software (Applied Biosystems, Foster City, USA). The sequences of the PCR and minisequencing primers are provided in Supplementary Table 1. The genotype call rate was 96,7% and the reproducibility was 99,7%, according to replication of 15.6% of the genotypes.

Electrophoretic mobility shift assay (EMSA)

Complementary 5' biotinylated and unmodified 30-37 bp oligonucleotides were designed for both alleles of the CGGGG indel (3 x CGGGG probe: 5'-AGTGGATTCG**CGGGGCGGGGCGGG**GCACTGCCCGCGC-3' and 4 x CGGGG 5'-AGTGGATTCG**CGGGGCGGGGCGGGGCGGGG**CACTGCC-3'). The oligonucleotides were obtained from Integrated DNA Technologies (IDT Inc, Coralville, IA, USA). The complementary oligonucleotides were allowed to anneal in 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA to generate double stranded probes for the EMSA reaction. Twenty fmoles of the labeled double stranded probes were incubated with 3 µl of nuclear extract prepared from blood cells, using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Pierce Biotechnology, Rockford, IL, USA), in a freshly made binding buffer containing 12 mM HEPES pH 7.4, 5 mM $MgCl_2$, 60 mM KCl, 1% glycerol, 0.05% NP-40, 50 µg/µl BSA, 1 mM DTT, 0.5 mM EDTA with 50 ng/µl of poly(dI-dC)·poly(dI-dC) (Amersham Biosciences, Piscataway, NJ, USA) and Halt™ Protease Inhibitor Cocktail (Pierce Biotechnology, Rockford, IL, USA) in a final volume of 20 µl. The mixtures were incubated at room temperature for 20 minutes, and analyzed using electrophoresis on 6% polyacrylamide gels (Bio-Rad Laboratories, Hercules, CA, USA). The gels were run for 1.5 h at 100 V, followed by transfer to

Hybond-N+ nylon membranes (Buckinghamshire, England) in 0.5 x TBE for 1 h at 550 mA, using a Criterion Blotter (Bio-Rad Laboratories, Hercules, CA, USA). The Chemiluminescent Nucleic Acid Detection Module (Pierce Biotechnology, Rockford, IL, USA) was used to visualize the biotinylated oligonucleotide signals on the membranes using a ChemiDoc XRS system (Bio-Rad Laboratories, Hercules, CA, USA).

Statistical analysis

For each tested polymorphisms, the quality of the genotype data was confirmed for both cohorts by testing for Hardy-Weinberg equilibrium in the control samples, using the Chi-square test and Haploview 3.32 (47). Pairwise linkage disequilibrium values D'- and r^2- were determined using the Haploview v3.32 software. The haplotypes were constructed using the FastPhase software (48). Allele and genotypes frequencies were compared between patient and healthy controls by a Chi square test. For CD and UC, the allele frequencies were compared between the disease subgroups and all controls. Differences in haplotype frequencies were determined by Fisher's exact test. Odds ratios were calculated using the formula [*a(r)*/*b(r)*]/ [*a(nr)*/*b(nr)*], where *a* and *b* stand for the allele counts in patients and controls, respectively, with the risk-allele as *(r)* and the non-risk allele as *(nr)*. The combined analyses were performed by pooling the genotype data from the individual cases and controls from both cohorts.

Acknowledgments

We are grateful to all the clinicians that took part in patient recruitment. (University of Liège collaborators: Paul Mainguet, Faddy Mokaddem, Fernand Fontaine, Jacques Deflandre, and Hubert Demolin). Genotyping was performed using the equipment of the SNP technology platform at Uppsala University (www.genotyping.se). This work was supported by grants from the Fonds de la Fondation Léon Frédéricq, the University of Liège (CHU), the University Hospital of Liège, the Swedish Research Council for Medicine and the Swedish Knut and Alice Wallenberg Foundation. E. Louis is a Research Associate of the Fonds National de la Recherche Scientifique. S. Vermeire is a fellow of the Nationaal Fonds voor Wetenschappelijk Onderzoek.

References

1. Danese, S., Sans, M. and Fiocchi, C. (2004) Inflammatory bowel disease: the role of environmental factors. *Autoimmun. Rev.,* **3,** 394-400.
2. Kuster, W., Pascoe, L., Purrmann, J., Funk, S. and Majewski, F. (1989) The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. *Am. J. Med. Genet.,* **32,** 105-108.
3. Orholm, M., Binder, V., Sorensen, T.I., Rasmussen, L.P. and Kyvik, K.O. (2000) Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand. J. Gastroenterol.,* **35,** 1075-1081.
4. Russell RK, W.D., Satsangi J. (2004) Unravelling the complex genetics of inflammatory bowel disease. *Arch. Dis. Child.,* **89,** 598-603.
5. Thompson, N.P., Driscoll, R., Pounder, R.E. and Wakefield, A.J. (1996) Genetics versus environment in inflammatory bowel disease: results of a British twin study. *B.M.J.,* **312,** 95-96.
6. Lesage, S., Zouali, H., Cezard, J.P., Colombel, J.F., Belaiche, J., Almer, S., Tysk, C., O'Morain, C., Gassull, M., Binder, V. *et al.* (2002) CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am. J. Hum. Genet.,* **70,** 845-857.
7. Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J. *et al.* (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.,* **39,** 207-211.

8. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A. *et al.* (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene Novel susceptibility genes in inflammatory bowel disease. *Science,* **314,** 1461-1463. .

9. Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire, S., Dewit, O., de Vos, M., Dixon, A. *et al.* (2007) Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of PTGER4. *PLoS Genet.,* **3,** e58.

10. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* **447,** 661-678.

11. Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D. *et al.* (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.,* **39**,830-832.

12. Becker, K.G., R. M. Simon, et al. (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc. Natl. Acad. Sci. U S A,* **95,** 9979-9984.

13. Weng, X., Liu, L., Barcellos, L.F., Allison, J.E. and Herrinton, L.J. Clustering of Inflammatory Bowel Disease With Immune Mediated Diseases Among Members of a Northern California-Managed Care Organization. *Am. J. Gastroenterol.,* **102,** 1429-1435.

14. Shiina, T., Inoko, H. and Kulski, J.K. (2004) An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens,* **64,** 631-649.

15. Lee, Y.H., Rho, Y.H., Choi, S.J., Ji, J.D., Song, G.G., Nath, S.K. and Harley, J.B. (2007) The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis. *Rheumatology,* **46,** 49-56.

16. Kavvoura, F.K., Akamizu, T., Awata, T., Ban, Y., Chistiakov, D.A., Frydecka, I., Ghaderi, A., Gough, S.C., Hiromatsu, Y., Ploski, R. *et al.* (2007) CTLA-4 Gene Polymorphisms and Autoimmune Thyroid Disease: A Meta Analysis. *J. Clin. Endocrinol. Metab.***, 92**, 3162-3170.

17. Wen, Z. and Fiocchi, C. (2004) Inflammatory bowel disease: autoimmune or immune-mediated pathogenesis? *Clin. Dev. Immunol.,* **11,** 195-204.

18. Marie, I., Durbin, J.E. and Levy, D.E. (1998) Differential viral induction of distinct interferon-alpha genes by positive feedback through interferon regulatory factor-7. *Embo J.,* **17,** 6660-6669.

19. Lin, R., Heylbroeck, C., Pitha, P.M. and Hiscott, J. (1998) Virus-Dependent Phosphorylation of the IRF-3 Transcription Factor Regulates Nuclear Translocation, Transactivation Potential, and Proteasome-Mediated Degradation. *Mol. Cell. Biol.,* **18,** 2986-2996.

20. Zhang, L., Wu, L., Hong, K. and Pagano, J.S. (2001) Intracellular Signaling Molecules Activated by Epstein-Barr Virus for Induction of Interferon Regulatory Factor 7. *J. Virol.,* **75,** 12393-12401.

21. Cheng, T.-F., Brzostek, S., Ando, O., Van Scoy, S., Kumar, K.P. and Reich, N.C. (2006) Differential Activation of IFN Regulatory Factor (IRF)-3 and IRF-5 Transcription Factors during Viral Infection. *J. Immunol.,* **176,** 7462-7470.

22. Honda, K. and Taniguchi, T. (2006) IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.,* **6,** 644-658.

23. Barnes, B.J., Field, A.E. and Pitha-Rowe, P.M. (2003) Virus-induced Heterodimer Formation between IRF-5 and IRF-7 Modulates Assembly of the IFNA Enhanceosome in Vivo and Transcriptional Activity of IFNA Genes. *J. Biol. Chem.,* **278,** 16630-16641.

24. Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S.-i., Honda, K., Ohba, Y., Mak, T.W. *et al.* (2005) Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature,* **434,** 243-249.

25. O'Neill, L.A.J. and Bowie, A.G. (2007) The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.,* **7,** 353-364.

26. Ouyang, X., Negishi, H., Takeda, R., Fujita, Y., Taniguchi, T. and Honda, K. (2007) Cooperation between MyD88 and TRIF pathways in TLR synergy via IRF5\par activation. *Biochem. Biophys. Res. Commun.,* **354,** 1045-1051.

27. Nakamura, K., Honda, K., Mizutani, T., Akiho, H. and Harada, N. (2006) Novel strategies for the treatment of inflammatory bowel disease: selective inhibition of cytokines and adhesion molecules. *World J. Gastroenterol.,* **12,** 4628-4635.

28. Sigurdsson, S., Nordmark, G., Goring, H.H., Lindroos, K., Wiman, A.C., Sturfelt, G., Jonsen, A., Rantapaa-Dahlqvist, S., Moller, B., Kere, J. *et al.* (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.,* **76,** 528-537.

29. Graham, R.R., Kozyrev, S.V., Baechler, E.C., Reddy, M.V.P.L., Plenge, R.M., Bauer, J.W., Ortmann, W.A., Koeuth, T., Escribano, M.F.G., Collaborative Groupsthe Argentine and, S. *et al.* (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.,* **38,** 550-555.

30. Sigurdsson, S., Padyukov, L., Kurreeman, F., Liljedahl, U., Wiman, A.C., Alfredsson, L., Toes, R., Rönnelid, J., Klareskog, L. Huizinga, T. *et al*. (2007) Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthr. Rheum.,* **56,** 2202-2210.

31. Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A. *et al.* (1998) Databases on Transcriptional Regulation: TRANSFAC, TRRD, and COMPEL. . *Nucleic Acids Res.,* **26,** 364-370.

32. Graham, R.R., Kyogoku, C., Sigurdsson, S., Vlasova, I.A., Davies, L.R., Baechler, E.C., Plenge, R.M., Koeuth, T., Ortmann, W.A., Hom, G. *et al.* (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. U S A.,* **104,** 6758-6763. .

33. Bergman, L., Lindblom, J.B., Safwenberg, J. and Krause, U. (1976) HL-A frequencies in Crohn's disease and ulcerative colitis. *Tissue Antigens.,* **7,** 145-150.

34. Gaya, D.R., Russell, R.K., Nimmo, E.R. and Satsangi, J. New genes in inflammatory bowel disease: lessons for complex diseases? *The Lancet,* **367,** 1271-1284.

35. Cho, J.H., Nicolae, D.L., Gold, L.H., Fields, C.T., LaBuda, M.C., Rohal, P.M., Pickles, M.R., Qin, L., Fu, Y., Mann, J.S. *et al.* (1998) Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: Evidence for epistasis between 1p and IBD1. *PNAS,* **95,** 7502-7507.

36. Hampe, J., Shaw, S.H., Saiz, R., Leysens, N., Lantermann, A., Mascheretti, S., Lynch, N.J., MacPherson, A.J., Bridger, S., van Deventer, S. *et al.* (1999) Linkage of inflammatory bowel disease to human chromosome 6p. *Am. J. Hum. Genet.*, **65,** 1647-1655.

37. Cummings, J.R., Cooney R., Pathan, S., Anderson, C., Barrett,J., Beckly, J., Geremia, A., Hancock, L., Guo, C.,Ahmad, T. *et al.* (2007) Confirmation of the role of ATG16l1 as a Crohn's disease susceptibility gene. *Inflamm. Bowel Dis.*, **13,** 941-946.

38. Tremelling, M., Cummings, F., Fisher, S.A., Mansfield, J., Gwilliam, R., Keniry, A., Nimmo, E.R., Drummond, H., Onnie, C.M., Prescott, N.J. *et al.* (2007) IL23R Variation Determines Susceptibility But Not Disease Phenotype in Inflammatory Bowel Disease. *Gastroenterology,* **132,** 1657-1664.

39. De Jager, P.L., Graham, R., Farwell, L., Sawcer, S., Richardson, A., Behrens, T.W., Compston, A., Hafler, D.A., Kere, J., Vyse, T.J. *et al.* (2006) The role of

inflammatory bowel disease susceptibility loci in multiple sclerosis and systemic lupus erythematosus. *Genes Immun.,* **7,** 327-334.

40. Ferreiros-Vidal, I., Garcia-Meijide, J., Carreira, P., Barros, F., Carracedo, A., Gomez-Reino, J.J. and Gonzalez, A. (2003) The three most common CARD15 mutations associated with Crohn's disease and the chromosome 16 susceptibility locus for systemic lupus erythematosus. *Rheumatology,* **42,** 570-574.

41. Chong, W.P., Ip, W.K., Lau, C.S., Chan, T.M., Padyukov, L. and Lau, Y.L. (2004) Common NOD2 polymorphisms in Hong Kong Chinese patients with systemic lupus erythematosus. *Rheumatology,* **43,** 104-105.

42. Ferreiro-Neira I., Calaza M., Alonso-Perez E., Marchini M., Scorza R., Sebastiani G.D., Blanco F.J., Rego I., Pullmann R., Pullmann R. *et al.* (2007) Opposed independent effects and epistasis in the complex association of IRF5 to SLE. *Genes Immun.,* **8,** 429-438.

43. Sharma, S., Grandvaux, N., Mamane, Y., Genin, P., Azimi, N., Waldmann, T. and Hiscott, J. (2002) Regulation of IFN Regulatory Factor 4 Expression in Human T Cell Leukemia Virus-I-Transformed T Cells. *J. Immunol.,* **169,** 3120-3130.

44. McAlexander, M.B. and Yu-Lee, L.-y. (2001) Sp1 is required for prolactin activation of the interferon regulatory factor-1 gene. *Mol. Cell. Endocrinol.,* **184,** 135-141.

45. Bell, P.A., Chaturvedi, S., Gelfand, C.A., Huang, C.Y., Kochersperger, M., Kopla, R., Modica, F., Pohl, M., Varde, S., Zhao, R. *et al.* (2002) SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques.,* **Suppl,** 70-72, 74, 76-77.

46. Hsu, T.M., Chen, X., Duan, S., Miller, R.D. and Kwok, P.Y. (2001) Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques.*, **31,** 560, 562, 564-568.

47. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* **21,** 263-265.

48. Scheet, P. and Stephens, M. (2006) A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. *Am. J. Hum. Genet.,* **78,** 629-644.

49. Graham, D.S.C., Manku, H., Wagner, S., Reid, J., Timms, K., Gutin, A., Lanchbury, J.S. and Vyse, T.J. (2007) Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. *Hum. Mol. Genet.,* **16,** 579-591.

50. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature,* **437,** 1365-1369.

Legends to the figures

Figure 1

Illustration of the IRF5 gene. a. Exons are represented as boxes; coding exons in white. The genotyped polymorphisms and their location are represented. b. Sequences of the two insertions (underlined). The binding sites for the transcription factor SP1 are represented by boxes, the white box corresponding to the third SP1 binding site created by the insertion of the 5bp CGGGG.

Figure 2

Plot of pair-wise linkage disequilibrium values (r^2) between IRF5 polymorphisms. The LD plot has been generated with the "snp.plotter"-option of the R-software package.

Figure 3

Electrophoretic mobility shift assay for the CGGGG indel. Samples loaded in lanes 1-3 contain: 1) labeled probe only; 2) labeled probe and nuclear extract; 3) labeled probe, nuclear extract and the unlabeled probe, which is added as a competitor in 100-fold excess. The bands showing specific and differential protein binding are indicated with arrows.

Tables

Table 1. Potential effects and disease associations for the genotyped polymorphisms

SNP	Potential functional role	Association	References
rs729302	Promoter region	SLE	Sigurdsson *et al* (28), Graham R.R. *et al.* (29, 32)
rs4728142	Promoter region, variation in IRF5 expresion level	SLE	Graham RR *et al*. (32) (34), Graham D.S.G. *et al.* (49)
rs3757385	Promoter region	RA	Sigurdsson *et al.* (30)
CGGGG indel	Promoter region		Chromosome position 128365152 (NCBI build 36.2.)
rs2004640	Altered consensus splice donor site / intron, variation in IRF5 expresion level	SLE,RA	Sigurdsson *et al.* (28) (30) Graham R.R. *et al.* (29, 32),Graham D.S.G. et al. (49)
rs3807306	Intron	SLE,RA	Sigurdsson *et al* .(30), Graham D.S.G. *et al*. (49)
30bp indel	Alteration of PEST domain, variation in IRF5 expresion level		Graham R.R. *et al.*(32)
rs2070197	Variation in IRF5 expresion level	SLE	Graham R.R. *et al.* (32)
rs10954213	Altered length of he IRF5 3' UTR, variation in IRF5 expresion level	SLE	Graham R.R. *et al.* (32), Graham D.S.G. *et al.* (49)
rs11770589	3’ UTR		
rs2280714	Variation in IRF5 expresion level	SLE	Cheung V *et al*.(50), Graham R.R. *et al.* (29, 32)
rs12539741	Variation in IRF5 expresion level	SLE	Graham *et al.* (32).

Table 2: Single marker association results. Main cohort.

			Allele Frequencies				P values			Odds ratio (C.I. 0.95)		
SNP	**Alleles**	**Risk allele**	**Controls n=223**	**IBD Patients n=987**	**CD Patients n=731**	**UC Patients n=250**	**IBD Patients**	**CD Patients**	**UC Patients**	**IBD Patients**	**CD Patients**	**UC Patients**
rs729302	A/C	A	0.40	0.35	0.37	0.30	0.068	0.25	0.0027	0.80 (0.64-1.01)	0.87 (0.69-1.10)	0.64 (0.48-0.85)
rs4728142	A/G	A	0.36	0.45	0.42	0.51	0.0016	0.029	0.000010	1.46 (1.15-1.84)	1.32 (1.03-1.68)	1.90 (1.43-2.52)
rs3757385	G/T	G	0.42	0.37	0.38	0.35	0.072	0.15	0.028	0.81 (0.65-1.01)	0.84 (0.66-1.06)	0.73 (0.55-0.96)
CGGGG indel	¾ (CGGGG)	Ins	0.30	0.43	0.41	0.51	0.000019	0.00068	0.000000053	1.81 (1.37-2.39)	1.63 (1.23-2.17)	2.42 (1.76-3.34)
rs2004640	T/G	T	0.45	0.5	0.48	0.55	0.071	0.27	0.0036	1.23 (0.99-1.54)	1.15 (0.91-1.44)	1.51 (1.15-1.98)
rs3807306	T/G	T	0.40	0.48	0.46	0.53	0.0056	0.046	0.00023	1.37 (1.10-1.71)	1.27 (1.01-1.60)	1.69 (1.28-2.22)
30bp indel	½	Del	0.46	0.52	0.52	0.46	0.021	0.042	0.022	1.30 (1.04-1.62)	1.27 (1.01-1.60)	0.73 (0.55-0.95)
rs2070197	C/T	C	0.08	0.09	0.08	0.10	1	0.92	0.63	1.03 (0.69-1.53)	0.98 (0.65-1.48)	1.16 (0.71-1.86)
rs10954213	A/G	A	0.46	0.39	0.40	0.37	0.018	0.050	0.011	0.76 (0.61-0.95)	0.79 (0.63-0.99)	0.69 (0.52-0.91)
rs11770589	A/G	A	0.55	0.48	0.52	0.46	0.018	0.036	0.018	0.76 (0.61-0.95)	1.28 (1.02-1.6)	0.72 (0.55-0.94)
rs2280714	T/C	T	0.39	0.35	0.36	0.34	0.10	0.16	0.072	0.82 (0.66-1.03)	0.84 (0.67-1.06)	0.77 (0.58-1.02)
rs12539741	T/C	T	0.07	0.09	0.08	0.10	0.46	0.58	0.25	1.22 (0.79-1.90)	1.16 (0.74-1.83)	1.37 (0.82-2.29)

Table 3: Single marker association results. Confirmatory cohort.

			Allele Frequencies				P values			Odds ratio (C.I. 0.95)		
SNP	**Alleles**	**Risk allele**	**Controls n=311**	**IBD Patients n=674**	**CD Patients n=476**	**UC Patients n=179**	**IBD Patients**	**CD Patients**	**UC Patients**	**IBD Patients**	**CD Patients**	**UC Patients**
rs729302	A/C	A	0.36	0.36	0.37	0.34	0.91	0.54	0.56	1.01 (0.82-1.25)	1.07 (0.86-1.34)	0.91 (0.69-1.21)
rs4728142	A/G	A	0.39	0.43	0.42	0.43	0.17	0.29	0.26	1.15 (0.94-1.42)	1.13 (0.91-1.40)	1.17 (0.89-1.55)
rs3757385	G/T	G	0.38	0.37	0.37	0.40	0.79	0.74	0.57	0.97 (0.79-1.19)	0.96 (0.77-1.20)	1.09 (0.83-1.43)
CGGGG indel	¾ (CGGGG)	Ins	0.35	0.46	0.46	0.46	0.000032	0.00014	0.0023	1.59 (1.28-1.98)	1.57 (1.24-1.98)	1.59 (1.18-2.14)
rs2004640	T/G	T	0.49	0.50	0.50	0.50	0.72	0.83	0.89	1.04 (0.85-1.27)	1.03 (0.83-1.27)	1.02 (0.78-1.34)
rs3807306	T/G	T	0.47	0.49	0.48	0.5	0.51	0.71	0.41	1.07 (0.88-1.31)	1.04 (0.84-1.29)	1.13 (0.86-1.48)
30bp indel	½	Del	0.47	0.49	0.49	0.49	0.58	0.42	0.68	1.06 (0.87-1.29)	1.09 (0.88-1.3)	1.06 (0.82-1.39)
rs2070197	C/T	C	0.08	0.10	0.11	0.092	0.12	0.057	0.54	1.34 (0.94-1.91)	1.43 (0.99-2.09)	1.17 (0.73-1.89)
rs10954213	A/G	A	0.39	0.39	0.39	0.40	0.87	0.87	0.72	0.98 (0.79-1.20)	0.98 (0.79-1.20)	1.05 (0.80-1.39)
rs11770589	A/G	A	0.47	0.49	0.50	0.49	0.54	0.39	0.58	1.07 (0.87-1.3)	1.10 (0.89-1.36)	1.09 (0.83-1.42)
rs2280714	T/C	T	0.34	0.34	0.34	0.36	0.91	0.86	0.52	0.98 (0.79-1.21)	0.97 (0.78-1.22)	1.1 (0.83-1.45)
rs12539741	T/C	T	0.080	0.10	0.11	0.092	0.10	0.057	0.54	1.35 (0.94-1.92)	1.45 (1-2.10)	1.18 (0.73-1.89)

Table 4: Single marker association results case/controls from combined cohorts

			Allele Frequencies				P values			Odds ratio (C.I. 0.95)		
SNP	**Alleles**	**Risk allele**	**Controls n=534**	**IBD Patients n=1661**	**CD Patients n=1207**	**UC Patients n=429**	**IBD Patients**	**CD Patients**	**UC Patients**	**IBD Patients**	**CD Patients**	**UC Patients**
rs729302	A/C	A	0.38	0.36	0.37	0.32	0.26	0.81	0.011	0.91 (0.79-1.07)	0.98 (0.83-1.15)	0.77 (0.63-0.94)
rs4728142	A/G	A	0.38	0.42	0.42	0.48	0.0014	0.029	0.000042	1.28 (1.1-1.49)	1.20 (1.02-1.40)	1.50 (1.24-1.83)
rs3757385	G/T	G	0.40	0.37	0.37	0.37	0.19	0.30	0.28	0.90 (0.78-1.05)	0.92 (0.78-1.07)	0.90 (0.74-1.09)
CGGGG indel	¾ (CGGGG)	Ins	0.33	0.44	0.43	0.49	0.000000014	0.0000033	0.00000000079	1.62 (1.37-1.91)	1.51 (1.27-1.80)	1.93 (1.56-2.38)
rs2004640	T/G	T	0.47	0.50	0.49	0.53	0.16	0.46	0.027	1.11 (0.96-1.29)	1.06 (0.91-1.24)	1.24 (1.03-1.50)
rs3807306	T/G	T	0.44	0.48	0.47	0.52	0.027	0.17	0.0018	1.18 (1.02-1.37)	1.11 (0.95-1.30)	1.36 (1.12-1.64)
30bp indel	½	Del	0.50	0.48	0.49	0.47	0.30	0.51	0.27	0.92 (0.80-1.07)	0.95 (0.81-1.10)	0.90 (0.74-1.08)
rs2070197	C/T	C	0.081	0.094	0.094	0.094	0.27	0.27	0.39	1.17 (0.89-1.52)	1.17 (0.89-1.54)	1.17 (0.84-1.67)
rs10954213	A/G	A	0.42	0.39	0.39	0.38	0.11	0.21	0.15	0.88 (0.76-1.03)	0.90 (0.77-1.06)	0.87 (0.71-1.05)
rs11770589	A/G	A	0.50	0.49	0.49	0.48	0.30	0.48	0.29	0.92 (0.80-1.07)	0.95 (0.81-1.10)	0.90 (0.75-1.09)
rs2280714	T/C	T	0.36	0.35	0.35	0.35	0.31	0.41	0.48	0.92 (0.79-1.08)	0.93 (0.80-1.10)	0.93 (0.77-1.13)
rs12539741	T/C	T	0.077	0.095	0.096	0.095	0.11	0.11	0.19	1.25 (0.95-1.65)	1.26 (0.95-1.68)	1.26 (0.90-1.78)

Figures

Fig. 1

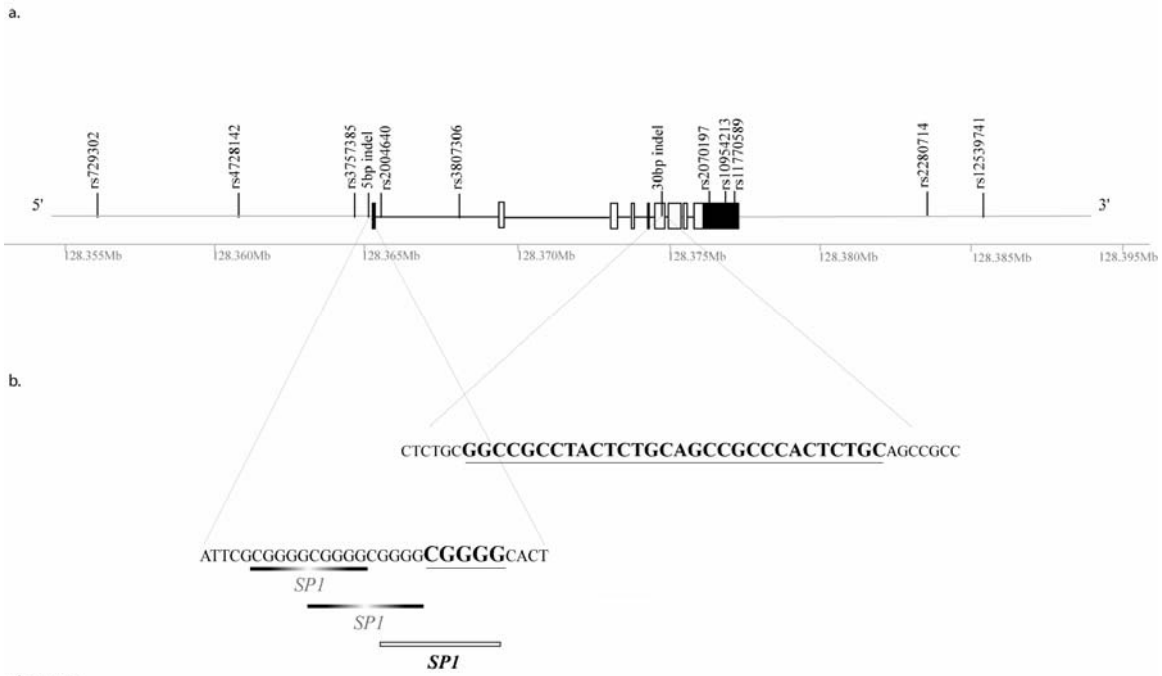


Figure 2

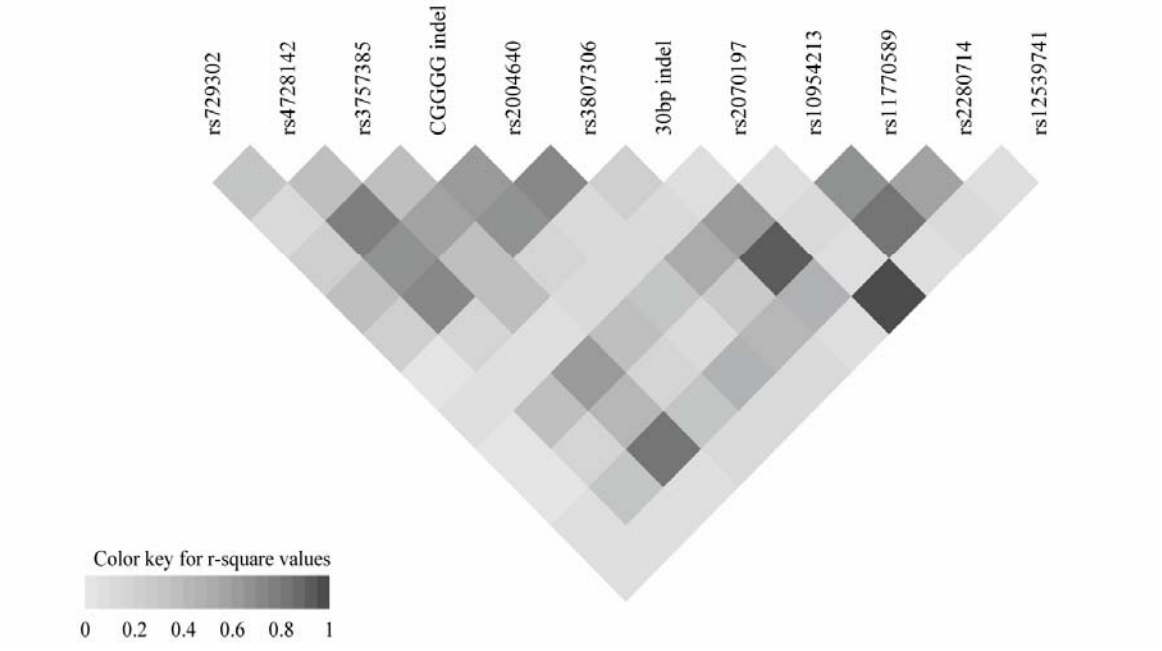
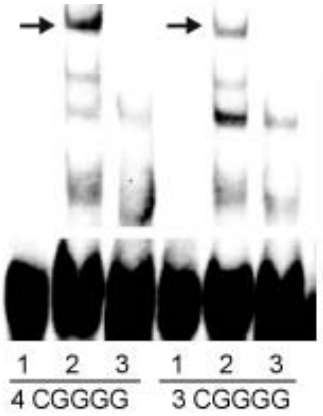


Fig.3



Abbreviations

CD: Crohn disease

CTLA-4: Cytotoxic T-lymphocyte antigen-4

GD: Graves' disease

GWA: Genome wide association

IBD: Inflammatory bowel diseases

IC: Indeterminate colitis

IFN: Interferon

IL: Interleukin

Indel: Insertion / deletion

IRF5: Interferon regulatory factor 5

RA: Rheumatoid arthritis

SLE: Systemic lupus erythematosus

SNP: Single nucleotide polymorphism

SP1: Specificity protein 1

T1D: Type 1 diabetes

TLR: Toll-like receptor

UC: ulcerative colitis

WTCCC: Welcome Trust Case Control Consortium