Genome-Wide Copy Number Variation Scan Identifies Complement Component C4 as Novel Susceptibility Gene for Crohn's Disease

Isabelle Cleynen, PhD, Peter Konings, MSc,[†,‡] Caroline Robberecht, PhD,[§] Debby Laukens, PhD,[‖] Leila Amininejad, PhD,[¶] Emilie Théâtre, PhD,[**,††] Kathleen Machiels, PhD,* Ingrid Arijs, PhD,* Paul Rutgeerts, PhD,[*,‡‡] Edouard Louis, PhD,[**,††] Denis Franchimont, PhD,[¶] Martine De Vos, PhD,[‖] Kristel Van Steen, PhD,[§§] Michel Georges, PhD,** Yves Moreau, PhD,[†,‡] Joris Vermeesch, PhD,[§] and Séverine Vermeire, PhD[*,‡‡]*

Background: The genetic component of Crohn's disease (CD) is well known, with 140 susceptibility loci identified so far. In addition to single nucleotide polymorphisms typically studied in genome-wide scans, copy number variation is responsible for a large proportion of human genetic variation.

Methods: We performed a genome-wide search for copy number variants associated with CD using array comparative genomic hybridization. One of the found regions was validated independently through real-time PCR. Serum levels of the found gene were measured in patients and control subjects.

Results: We found copy number differences for the *C4S* and *C4L* gene variants of complement component C4 in the central major histocompatibility complex region on chromosome 6p21. Specifically, we saw that CD patients tend to have lower *C4L* and higher *C4S* copies than control subjects ($P = 5.00 \times 10^{-03}$ and $P = 9.11 \times 10^{-04}$), which was independent of known associated classical HLA I and II alleles ($P = 7.68 \times 10^{-03}$ and $P = 6.29 \times 10^{-03}$). Although C4 serum levels were not different between patients and control subjects, the relationship between *C4* copy number and serum level was different for patients and control subjects with higher copy numbers leading to higher serum concentrations in control subjects, compared with CD patients ($P < 0.001$).

Conclusions: C4 is part of the classical activation pathway of the complement system, which is important for (auto)immunity. Low *C4L* or high C4S copy number, and corresponding effects on C4 serum level, could lead to an exaggerated response against infections, possibly leading to (auto)immune disease.

(Inflamm Bowel Dis 2016;22:505–515)

Key Words: Crohn's disease, copy number variation, complement component C4, HLA, inflammatory bowel disease

Crohn's disease (CD) is, besides ulcerative colitis, 1 of the 2 major subtypes of inflammatory bowel disease (IBD). CD is a chronic inflammatory disease of the gastrointestinal tract and affects 300 per 100,000 people of European ancestry with increasing prevalence in other populations.[1] Inflammation can be debilitating with symptoms, including severe abdominal pain, diarrhea,

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.ibdjournal.org).

Received for publication July 8, 2015; Accepted September 9, 2015.

From the *Department of Clinical and Experimental Medicine, TARGID, KU Leuven, Leuven, Belgium; [†]Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium; [‡]iMinds Department of Medical Information Technologies, KU Leuven, Leuven, Belgium; [§]Department of Human Genetics, KU Leuven, Leuven, Belgium; [‖]Department of Gastroenterology, Ghent University Hospital, Gent, Belgium; [¶]Erasmus Hospital, Free University of Brussels, Department of Gastroenterology, Brussels, Belgium; **Unit of Animal Genomics, Groupe Interdisciplinaire de Genoproteomique Appliquee (GIGA-R) and Faculty of Veterinary Medicine, University of Liege, Liege, Belgium; [††]Division of Gastroenterology, Centre Hospitalier Universitaire, Université de Liege, Liège, Belgium; [‡‡]Division of Gastroenterology, UZ Leuven, Leuven, Belgium; and [§§]Department of Electrical Engineering and Computer Science, Bioinformatics-Statistical Genetics, Montefiore Institute, Universty of Liège, Liège, Belgium.

Supported by the Research Foundation-Flanders (FWO) project grant G.0643.08N. I. Cleynen, I. Arijs and S. Vermeire are supported by a personal FWO grant. ULg (E. T., E.L. and M.G.) is funded by the Walloon Region (IPSEQ, Crohn & CIBLES projects), by the FEDER, by the Politique Scientifique Fédérale (IAP BeMGI), by the Fonds National de la Recherche Scientifique (FNRS) and by the Communauté Française de Belgique (ARC IBD@ULg). Y. Moreau is funded by the Research Council KU Leuven (GOA/10/09 MaNet, CoE PFV/10/016 SymBioSys), and by the Flemish Government: IWT (O&O ExaScience Life Pharma; ChemBioBridge; PhD grants); IOF (IOF_KP); Hercules Stichting [Hercules 3: PacBio RS; Hercules 1: The C1 single-cell auto prep system, BioMark HD System and IFC controllers (Fluidigm) for single-cell analyses]; iMinds Medical Information Technologies SBO 2014.

The authors have no conflicts of interest to disclose.

I. Cleynen, P. Konings, J. Vermeesch, and S. Vermeire have contributed equally.

Reprints: Isabelle Cleynen, PhD, O&N1, Herestraat 49, Box 602, B-3000 Leuven (e-mail: isabelle.cleynen@med.kuleuven.be).

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc.

DOI 10.1097/MIB.0000000000000623

Published online 23 November 2015.

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

and weight loss, and it has important social consequences. CD is a heterogeneous disorder with differences in disease location, behavior, and age of onset. The cause of IBD is multifactorial—environmental and genetic—and poorly understood.[2] The current hypothesis is that of an exaggerated mucosal immune response to the gut microbiota in a genetically susceptible host. Genome-wide association studies and subsequent meta-analyses identified a total of 140 CD susceptibility loci so far.[3–6] The most strongly associated genes are *NOD2* and *IL23R*, both involved in innate immunity and bacterial sensing; *ATG16L1* is an essential autophagy gene, and *PTGER4* is involved in epithelial restitution.[6] A strong association was also found to the major histocompatibility complex (MHC).[6,7] The MHC region is characterized by a very high gene density and genetic variation, and strong and long-range linkage disequilibrium, spanning over 3.6 Mb. This makes it difficult to identify which gene or genes are most relevant in the context of disease association. The region encodes a set of cell surface molecules also called the human leukocyte antigens (HLA) that present peptide antigens to the immune system.[8] There are 2 major classes of MHC molecules: class I and class II. Other genes that map within the MHC (MHC class III) include complement components (complement component C2, complement component C4, and factor B) and cytokines (tumor necrosis factor-α and lymphotoxin [LTA, LTB]).

In addition to single-nucleotide polymorphisms (SNPs) typically studied in genome-wide association studies, copy number variation (CNV) is responsible for a large proportion of human genetic variation.[9] Several CNVs contribute to susceptibility to (auto)immune and inflammatory disorders.[10–12] Associations of CNV to CD have been reported: a deletion polymorphism in the promoter region of the *IRGM* gene is associated with altered *IRGM* expression and with CD.[13] A genome-wide CNV association study replicated the *IRGM* deletion and found 2 additional CNVs: a not further fine-mapped CNV in the MHC region and a 900-kb inversion polymorphism on chromosome 17 (CNVR7113.6).[14]

To identify chromosomal areas of CNV associated with CD, we performed an exploratory genome-wide CNV analysis by array comparative genomic hybridization (aCGH) using an in-house developed 32K bacterial artificial chromosome (BAC) array. We identified several regions that vary in copy number in individuals with CD and further focused on 1 such region on chromosome 6p21 in the central MHC region through independent validation and functional follow-up.

MATERIALS AND METHODS

Development of a CNV-Specific BAC Array

Different platforms to detect CNVs have been developed and commercialized: clone (typically BACs but cosmids or complementary DNAs are also used) or oligonucleotide-based arrays that can only detect copy number variants and arrays interrogating both SNPs and CNVs. Each of these platforms has advantages and disadvantages for the study of CNVs. In the comprehensive study by Redon et al,[9] it was shown that both BAC-based arrays and the 500K Affymetrix SNP chip (Affymetrix, Santa Clara, CA) detect CNVs. However, only 30% of all CNVs overlapped, indicating that both platforms detect different types of CNVs. One reason is that complex genomic regions are selected against in oligonucleotide-based arrays (including the SNP arrays). Another reason may be the nature of the clones: stretches of 150 kb detect different genomic rearrangements as compared with small oligonucleotides. Because of our experience with BAC arrays, the observation that the larger CNVs are better detected using BAC arrays, and our knowledge about the sensitivity to detect mosaicism using BAC arrays (see below), we used BAC arrays. A CNV BAC array was developed in house containing 3906 clones. Clones were chosen to overlap with known copy number variable regions as reported in the study by Redon et al[9] and based on the database of genomic variants (http://projects.tcag.ca/variation/). Three thousand five hundred four clones were annotated in the database of genomic variants, and 156 overlapped with the most frequent polymorphic regions as described by Redon et al.[9] Five hundred forty BAC clones located in non–copy number variable regions were used for internal normalization. The majority of the BAC clones (n = 3210) came from the Children's hospital Oakland Research Institute, where the "BACPAC Resource Center" is located. One hundred sixteen clones were developed in house, and the remainder originated from the Wellcome Trust Sanger Institute. All clones were spotted at least in duplicate on the array.

Study Samples and Sample Preparation for aCGH Experiment

Seven hundred seventy unrelated patients with CD who are followed at the University Hospital Leuven and 345 healthy control subjects were included for the aCGH experiment. Diagnosis of CD was based on accepted clinical, endoscopic, radiologic, and histologic criteria.[15,16] All individuals were of European origin. Disease location was recorded according to the Montreal classification.[17] Of the CD patients, 315 (40.9%) were male and 455 (59.1%) were female; of the healthy control subjects, 164 (47.5%) were male and 181 (52.5%) were female. Two hundred seventy CD patients (35.1%) had ileal (L1), 85 (11.0%) had colonic (L2), and 415 (53.9%) had ileocolonic (L3) disease location. The median age at diagnosis was 24 years (interquartile range [IQR], 18–30 years). DNA was extracted from venous blood using a salting-out technique,[18] and stored at −80°C. DNA was diluted and concentration measured using NanoDrop (Thermo Scientific, Wilmington, DE).

We showed that BAC-based arrays can readily detect 5% mosaicism when up to 30 successive clones are involved, and local differences including single clones can detect a 20% mosaicism.[19,20] Therefore, we decided to cost-effectively pool patient DNA and control DNA for the aCGH experiment. We pooled DNA from 5 CD patients, each patient in a pool being of identical sex and disease location. Per patient pool, a random

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

5-sample control pool was put together, closely matched to the corresponding patient pool based on sex and age. To ensure that the pools are as homogeneous as possible, we constructed pools using the R package optMatch.[21] Control subjects could be reused in different pools but never in the same pool. Each control was used a median of 2 times (range, 1–9). A sex mismatch was never allowed, whereas the age difference in each matched pair was minimized (see Figure, Supplemental Digital Content 1, http://links.lww.com/IBD/B152). Matches were optimized using the Mahalanobis distance between patients and control subjects. Next, each case-control pair was assigned to a sample pool, based on location, sex, and a minimized age difference within the sample set, using a PAM clustering approach. We checked the resulting pools using box plots of age within each pool, stratified for location and sex (see Figure, Supplemental Digital Content 1, http://links.lww.com/IBD/B152). In total, we assembled 17 colon-only pools (5 male and 12 female), 54 ileum-only pools (27 male and 27 female), and 83 ileocolonic pools (31 male and 52 female).

Array Comparative Genomic Hybridization Experiment

An aCGH was performed as described previously by Vermeesch et al.[22] In brief, 150 ng genomic DNA from the patient pool and 150 ng of the matched control pool were differentially labeled by random priming with Cy3-dCTP and Cy5-dCTP (Amersham, GE Healthcare Life Sciences, Little Chalfont, UK) and cohybridized to the array overnight at 37°C in a humid chamber. Hybridized arrays were washed and scanned using an Agilent Microarray Scanner (Agilent Technologies, Santa Clara, CA). Image data were extracted using Agilent Feature Extraction software, and the data were analyzed using Agilent CGH analytics software. Self-hybridization was performed to test for clone-specific variability, and direct and dye swap assays were performed for each pool combination.

Quality Control and Data Analysis

We rejected 1 array after failing laboratory quality control and excluded 84 spots with a clone length <600 bases. Next, we excluded spots having a lower intensity than twice the median autosomal background from the analysis. The median rejected proportion was 3.5% for the green channel and 2.5% for the red channel (see Figure, Supplemental Digital Content 2, http://links.lww.com/IBD/B153). We then subtracted background intensity from foreground intensity,[19] after which we used 2-dimensional loess smoother to correct for spatial anomalies. We assessed preprocessing results using MA plots.[23] We averaged the intensity ratio per clone if 2 spots were available; otherwise we considered the available one as the measurement. We deleted a further 11 clones because they passed the quality control on less than 3 arrays or because they could not be mapped.

We summarized the results by averaging the $\log_2$ ratio for each clone. If in the different pools we repeatedly detect either below or above average intensity ratios for a particular locus, this would indicate a lower or higher copy number of this locus (see Table, Supplemental Digital Content 3, http://links.lww.com/IBD/B154). We heuristically defined outlying clones as those with a ratio higher or lower than 3 standard deviations away from the mean ($-0.13 >$ ratio > 0.13) (see Figure, Supplemental Digital Content 4, http://links.lww.com/IBD/B155). Clones were annotated using Ensembl release 54.[24] Data analysis was done using R (http://www.R-project.org/).

Validation and Replication Using Quantitative Real-Time PCR

We performed 3 TaqMan real-time PCR copy number assays, with probes specifically targeted to *C4A* (Hs07226349_cn), *C4B* (Hs07226350_cn), or *C4L* (QBLnoCOX_CCSIZO) (Life Technologies Corporation, Foster City, CA).[25,26] Validation samples included the individual patients and control subjects as used for aCGH, and an independent replication data set consisting of 1117 CD patients and 687 healthy control subjects from 3 other Belgian IBD centers (Liège University Hospital, Ghent University Hospital, and Brussels Erasme Hospital; Table 1). Diagnosis of CD was based on accepted clinical, endoscopic, radiologicl, and histological criteria.[15,16] All individuals were of European origin. Four replicates of each sample were amplified with the ABI7500 FAST (Life Technologies) starting from 20 ng DNA in a 20 μL reaction, according to manufacturer's instructions. RNAse-P was used as reference gene for relative quantification. Each 96-well plate contained negative control subjects (no DNA) and 2 positive control subjects (calibrator samples). Two independent runs were performed on samples that showed aberrant fold differences. Data were analyzed with the Copy Caller software version 1.0 from Life Technologies. We calculated total *C4* copy number as the sum of the *C4A* and *C4B* copy number, and *C4S* copy number as total *C4* minus *C4L* copy number. As an additional validation, the *C4L* copy number had to be smaller than *C4A* and *C4B* combined. Differences in copy number distribution between CD patients and healthy control subjects were tested with the Wilcoxon's test using R. Graphs were plotted using R.

Principal Components and HLA Data

A subset of 1540 CD patients and 900 control subjects were genotyped using immunochip in the context of the IIBDGC immunochip project.[6] For these individuals, principal components (PCs) were available to account for population structure in the association analyses. Association tests were performed on each C4 isotype (C4L, C4S, C4A, C4B, and total C4) conditional on the first 5 PCs in R.

High-density SNP data in the HLA region and imputed HLA alleles and amino acids were available in a subset of 1404 CD patients and 881 healthy control subjects, through the IIBDGC MHC project.[27] Correlation of C4 copy number with HLA SNPs, alleles, and amino acids was tested, and conditional logistic regression analysis was performed in R.

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 1. Real-Time Validation Sample Details

	Leuven	Brussel	Gent	Liège	Total
Controls					
n	313	229	243	215	1000
Sex, male (%)	145 (46.6)	33 (17.9)	102 (41.9)	71 (41.3)	351 (38.6)
Median age (IQR)	45.5 (37–57)	NA	36 (33–44)	44 (34–62)	41 (34–55)
CD					
n	685	399	461	257	1802
Sex, male (%)	248 (36.4)	155 (38.8)	198 (43.3)	108 (45.0)	709 (39.9)
Median age (IQR)	49 (40–59)	47 (38–59)	46 (35–56)	44 (34–57)	47 (37–58)
Median age at diagnosis (IQR)	24 (18–30)	26.5 (20–37)	25 (20–35)	23 (19–32)	25 (19–33)
Ileal disease location, n (%)	241 (35.2)	126 (32.5)	137 (30.7)	87 (36.7)	591 (33.7)
Colonic disease location, n (%)	74 (10.8)	72 (18.6)	91 (20.4)	41 (17.3)	278 (15.8)
Ileocolonic disease location, n (%)	369 (53.9)	190 (48.9)	218 (48.9)	109 (45.9)	886 (50.5)

IQR, interquartile range; NA, not available.

Determination of C4 Levels in Serum

We measured C4 serum levels in a subset of 60 CD patients and 116 healthy control subjects with nephelometry on an Immage instrument (Beckman Coulter, Fullerton, CA), at the Laboratory Medicine Department of the UZ Leuven. Serum levels were compared between cases and control subjects with Mann–Whitney *U* in R. Correlation of C4 serum levels with copy number was tested with regression analysis in R. We included case-control status and an interaction term for copy number and status to test for differential effects in cases and control subjects.

Ethical Considerations

The ethical boards of each separate recruiting institution approved for the study (approval number B322201213950/S53684 for University Hospital Leuven). Informed consent was obtained from all participants. All DNA samples and data in this study were handled anonymously.

RESULTS

Array Comparative Genomic Hybridization Identifies Several CNV Regions of Potential Interest in CD

After quality control, we had 3853 clones available for further analysis. The summary statistic for each clone (mean $\log_2$ ratio) is given in Table, Supplemental Digital Content 3, http://links.lww.com/IBD/B154. A negative $\log_2$ ratio represents a lower copy number in CD patients compared with healthy control subjects, whereas a positive ratio represents a higher copy number in CD patients compared with control subjects. One clone (RP11-758F23) had a positive outlying ratio of >0.13 (see Figure, Supplemental Digital Content 4, http://links.lww.com/IBD/B155). This clone is located in a gene desert on chromosome 14. Fifteen clones had a negative outlying ratio of <-0.13 (see Figure, Supplemental Digital Content 4, http://links.lww.com/IBD/B155). Table 2 lists the 20 clones with lowest negative mean $\log_2$ ratios. Five of these overlap with known IBD susceptibility loci, as indicated in Table 2.[6] The clone with the lowest ratio (RP1-34F7) is, together with 2 other top 10 clones (RP11–598J9 and RP11–427G15), is located on chromosome 6p21 in the central MHC region (Fig. 1). Another 3 overlapping top 20 clones (RP11-12F9, RP13-628D17, CTD-2129N23) are located on chromosome 10q24, encompassing *HIF1AN* and *PAX2*. One clone (RP13-650G11) overlapped with the *MUC3A/B* gene on chromosome 7q22.

Independent Replication Confirms Complement C4 as a New Susceptibility Gene for CD

We further focused on the region on chromosome 6p21. The deviating clones overlapped with the multiallelic copy number variable "RCCX" locus, which is named after the genes *STK19* (serine/threonine kinase 19; formerly *RP*), *C4* (complement component 4), *CYP21* (steroid 21-hydroxylase), and *TNX* (tenascin-X). *C4* is known to be copy number variable and can code for either a C4A or a C4B isotype protein.[28] For both C4A and C4B, the size varies between long (C4L) and short (C4S) depending on the presence or absence of the endogenous retroviral sequence HERV-K (C4).[29] We used different quantitative PCR (qPCR) assays with probes specifically targeting *C4A*, *C4B*, or *C4L*, to first test if individual CD patients from the aCGH experiment differed in *C4* copy number compared with healthy control subjects. Copy number frequencies for total *C4*, *C4A*, *C4B*, *C4L*, and *C4S* are given in Table 3. They were similar to those previously described for a German control population.[25] Raw genotype counts for all samples are given in Table, Supplemental Digital Content 5,

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 2. Top 20 BAC clones

Rank	Clone_Name	Mean $\log_2$ Ratio	Location (NCBI36)[a]	Genes
1	RP1-34F7	−0.203	6:32,031,967–32,200,774[b]	*C2, STK19, C4B, TNXA, C4A, CYP21A2, TNXB, ATF6B, FKBPL ….*
2	RP5-1050D4	−0.177	17:4,774,358–4,866,284	*GP1BA, SLC25A11, RNF167, PFN1, SPAG7, INCA1, KIF1C, GPR172B ….*
3	RP11-598J9	−0.171	6:32,138,352–32,298,363[b]	*TNXB, ATF6B, FKBPL, PRRT1, PPT2, AGER, RNF5, PBX2, NOTCH4 ….*
4	RP13-650G11	−0.169	7:100,363,784–100,392,763[b]	*MUC3A/3B*
5	RP11-787O14	−0.164	3:50,173,489–50,368,468	*GNAT1, SLC38A3, GNAI2, IFRD2, TUSC2, TUSC4, TMEM115 ….*
6	RP1-137D17	−0.152	6:169,488,216–169,576,118	Intergenic (*THBS2, WDR27*)
7	RP5-1090M5	−0.146	1:37,168,106–37,299,346	*GRIK3*
8	RP11-427G15	−0.146	6:31,995,533–32,151,955[b]	*C2, STK19, C4B, TNXA, C4A, CYP21A2, TNXB, ATF6B, FKBPL ….*
9	RP13-589F15	−0.145	7:149,250,848–149,305,409	Intergenic
10	RP5-1005L2	−0.145	20:44,369,220–44,380,337[b]	Intergenic (*CDH22, SLC35C2*)
11	RP11-12F9	−0.139	10:102,370,664–102,520,939	*HIF1AN, PAX2*
12	RP11-462E2	−0.137	12:52,696,607–52,866,711	*HOXC4, HOXC5, HOXC6, SMUG1*
13	RP13-628D17	−0.137	10:102,390,007–102,554,000	*HIF1AN, PAX2*
14	RP4-561P1	−0.134	1:34,1520,76–34,253,968	*CSMD2, HMGB4*
15	RP5-884C9	−0.134	1:38,263,828–38,379,441	*POU3F1*
16	CTD-2129N23	−0.129	10:102,512,527–102,606,367	*HIF1AN, PAX2*
17	RP11-44F19	−0.129	7:131,759,989–131,930,759	*PLXNA4*
18	RP13-529B24	−0.129	8:21,525,858–21,626,914	*GFRA2*
19	RP11-424K17	−0.127	19:60,210,061–60,373,766[b]	*NLRP7, GP6, RDH13, EPS8L1, LENG3, TNNT1*
20	RP11-33B3	−0.127	2:74,453,782–74,604,122	*DCTN1, WDR54, RTKN, WBP1, GCS1, LBX2 ….*

[a]Given as chromosome:basepair location start-stop (in NCBI b36 coordinates).
[b]Overlap Jostins et al.[6]

http://links.lww.com/IBD/B156. The pattern of *C4L* copy number distribution in patients differed significantly from that in healthy control subjects, with a shift toward a lower copy number for CD patients compared with control subjects ($P = 2.66 \times 10^{-02}$). *C4S* copy number also differed significantly between patients and control subjects ($P = 3.37 \times 10^{-02}$). Total *C4*, *C4A*, and *C4B* copy number distribution did not differ between CD patients and healthy control subjects.

We replicated these findings in an additional independent 1117 CD patients and 687 healthy control subjects from 3 other Belgian IBD centers (Table 1 for details on the samples, Table 3 for copy number frequencies, and see Table, Supplemental Digital

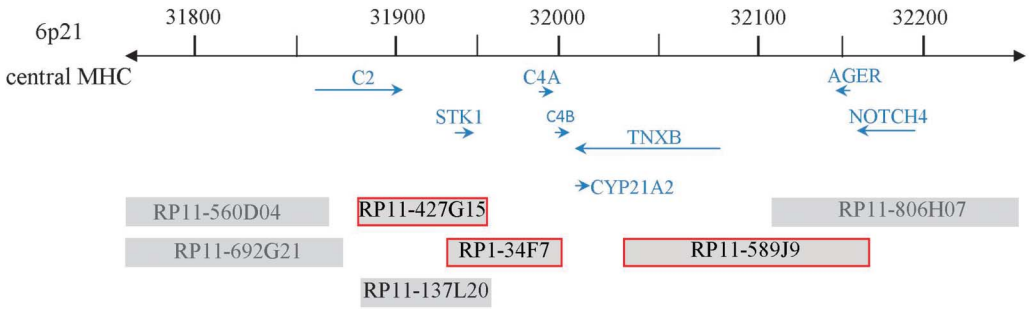


FIGURE 1. Overview of the 6p21 region. The relative position of the BAC clones (gray boxes) and genes (blue) (in GRCh37 coordinates) are shown. The genes *STK19* (serine/threonine kinase 19; formerly RP), *C4* (complement component 4), *CYP21* (steroid 21-hydroxylase), and *TNX* (tenascin-X) are part of the RCCX module. Red-framed BAC clones have a mean $\log_2$ ratio in the lowest top 20 (Table 2). RP11-137L20 has a mean $\log_2$ ratio of −0.08. BAC clones in gray text are not covered on the aCGH.

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 3. Copy Number Frequencies for *C4A*, *C4B*, *C4S*, *C4L*, and Total *C4* Per Center

	Leuven	Brussel	Gent	Liège	All	German Controls[a]
C4A						
CON						
0	0.01	0.00	0.02	0.01	0.01	0.01
1	0.12	0.15	0.13	0.14	0.13	0.18
2	0.56	0.53	0.58	0.65	0.58	0.56
3	0.28	0.27	0.26	0.17	0.25	0.22
4	0.02	0.04	0.00	0.03	0.02	0.02
5	0.01	0.00	0.00	0.00	0.00	0.00
CD						
0	0.02	0.01	0.02	0.01	0.01	
1	0.14	0.14	0.13	0.18	0.14	
2	0.57	0.60	0.56	0.61	0.58	
3	0.21	0.23	0.26	0.19	0.22	
4	0.06	0.02	0.02	0.02	0.04	
5	0.00	0.00	0.00	0.00	0.00	
C4B						
CON						
0	0.04	0.02	0.02	0.04	0.03	0.02
1	0.24	0.29	0.24	0.22	0.25	0.26
2	0.67	0.62	0.70	0.66	0.67	0.65
3	0.04	0.07	0.04	0.08	0.06	0.05
4	0.01	0.00	0.00	0.00	0.00	0.00
CD						
0	0.03	0.02	0.03	0.01	0.02	
1	0.26	0.24	0.27	0.24	0.25	
2	0.64	0.62	0.62	0.65	0.63	
3	0.07	0.10	0.07	0.10	0.08	
4	0.00	0.01	0.01	0.01	0.01	
C4S						
CON						
0	0.36	0.39	0.34	0.35	0.36	0.34
1	0.43	0.36	0.47	0.42	0.42	0.47
2	0.18	0.21	0.17	0.20	0.19	0.17
3	0.03	0.03	0.01	0.03	0.02	0.01
4	0.00	0.00	0.00	0.00	0.00	0.00
CD						
0	0.31	0.29	0.32	0.26	0.30	
1	0.43	0.46	0.44	0.50	0.45	
2	0.21	0.19	0.22	0.20	0.21	
3	0.05	0.05	0.02	0.03	0.04	
4	0.00	0.01	0.00	0.01	0.00	
C4L						
CON						
0	0.01	0.00	0.01	0.01	0.01	0.01
1	0.06	0.07	0.07	0.04	0.06	0.08
2	0.20	0.22	0.17	0.25	0.21	0.23
3	0.37	0.33	0.41	0.38	0.37	0.36
4	0.32	0.34	0.32	0.30	0.32	0.30
5	0.03	0.03	0.01	0.01	0.02	0.01
6	0.01	0.01	0.00	0.00	0.01	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
CD						
0	0.01	0.01	0.02	0.01	0.01	
1	0.07	0.06	0.06	0.06	0.06	
2	0.25	0.24	0.22	0.23	0.24	
3	0.37	0.44	0.39	0.44	0.40	
4	0.25	0.23	0.30	0.25	0.26	
5	0.04	0.03	0.02	0.01	0.03	
6	0.00	0.00	0.00	0.00	0.00	
7	0.00	0.00	0.00	0.00	0.00	
Total *C4*						
CON						
2	0.03	0.01	0.03	0.02	0.02	NA
3	0.19	0.25	0.20	0.21	0.21	NA
4	0.62	0.55	0.65	0.64	0.61	NA
5	0.15	0.16	0.11	0.11	0.13	NA
6	0.01	0.04	0.00	0.01	0.02	NA
7	0.01	0.00	0.00	0.00	0.00	NA
CD						
2	0.03	0.02	0.03	0.01	0.03	
3	0.22	0.20	0.19	0.24	0.21	
4	0.57	0.60	0.61	0.59	0.59	
5	0.16	0.15	0.15	0.15	0.15	
6	0.02	0.02	0.02	0.01	0.02	
7	0.00	0.00	0.00	0.00	0.00	

[a]Adapted from Flachsbart F, Caliebe A, Heinsen FA, et al. Investigation of complement component C4 copy number variation in human longevity. PLoS One. 2014;9:e86188.25 This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
NA, not available.

Content 5, http://links.lww.com/IBD/B156 for genotype counts for all samples), where we also saw a significant shift toward a higher *C4S* copy number in CD patients versus healthy control subjects ($P = 1.15 \times 10^{-02}$). *C4L* copy numbers were also shifted toward a lower number in CD patients compared with control subjects, but this was not significant ($P = 6.91 \times 10^{-02}$). In the combined data set of 1802 CD patients and 1000 healthy control subjects, both *C4L* and *C4S* copy number were significantly different between CD patients and healthy control subjects ($P = 5.01 \times 10^{-03}$ and $P = 9.11 \times 10^{-04}$; Fig. 2A, B). We did not see significant differences for *C4A*, *C4B*, or total *C4* in the replication or combined cohort (Fig. 2C–E). For none of the *C4* gene variants, there was a significant association with disease

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

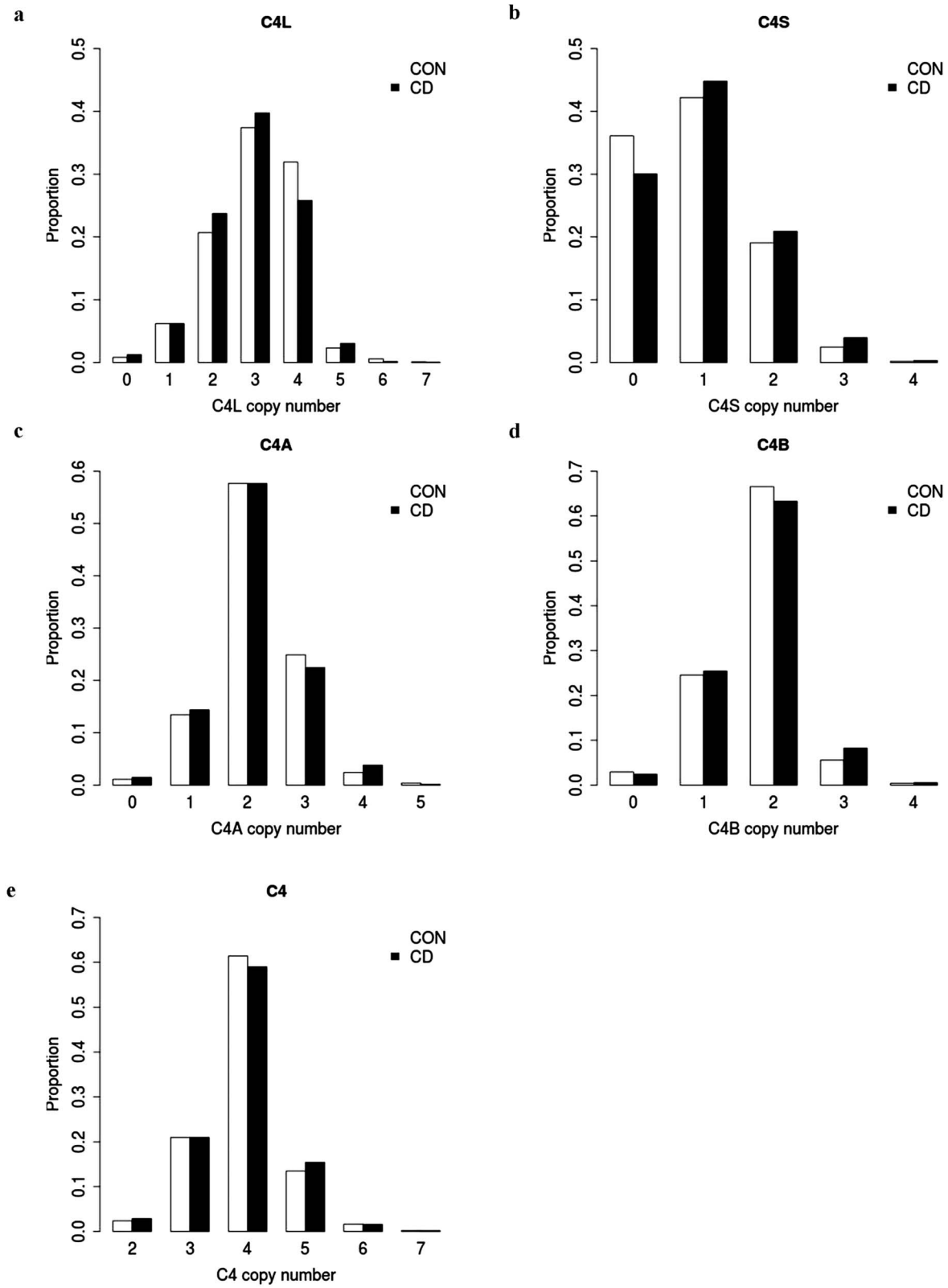


FIGURE 2. Distribution of C4 gene copy numbers in CD patients and healthy control subjects. Distribution pattern of *C4L* (a), *C4S* (b), *C4A* (c), *C4B* (d), and total *C4* (e) gene copy numbers.

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

subphenotype (disease location and age at diagnosis). Because the MHC region is highly variable geographically, we repeated the analysis controlling for PCs in 1540 cases and 900 control subjects in whom PCs were available. Both *C4S* and *C4L* remained significantly associated with CD after accounting for population structure ($P = 7.40 \times 10^{-03}$; 95% confidence interval [CI] = 0.89; IQR, 0.82–0.97 and $P = 7.94 \times 10^{-04}$; 95% CI = 1.20; IQR, 1.08–1.33, respectively).

Correlation of C4 Copy Number with HLA SNPs, Alleles, and Amino Acids

Given the location of *C4* within the MHC region, we checked the level of correlation with surrounding SNPs and class I and II HLA alleles and amino acids. The overall level of correlation between *C4* copy numbers and HLA SNPs/alleles was very low (median $r^2_{linear_regression} = 3.89 \times 10^{-03}$; IQR, 5.64×10^{-03} to 1.68×10^{-03}). The IIBDGC MHC project implicated 8 HLA class II alleles and 7 HLA class I alleles in CD, with a primary role for HLA-DRB1*01:03. We performed conditional logistic regression analysis for *C4* copy number and these 15 independently associated HLA class I and II alleles. Table 4 lists these 15 alleles, together with the level of correlation with the overall *C4* copy number. Conditional on all 15 independently associated HLA alleles, *C4L* and *C4S* copy number remained significantly associated with CD ($P = 7.68 \times 10^{-03}$; OR = 0.86; 95% CI, 0.77–0.96 and $P = 6.29 \times 10^{-03}$; OR = 1.21; 95% CI, 1.05–1.38, respectively). Also, when accounting for population structure, this association remains significant ($P = 5.95 \times 10^{-03}$; OR = 0.85; 95% CI, 0.76–0.95 and $P = 5.20 \times 10^{-03}$; OR = 1.21; 95% CI, 1.06–1.39, respectively).

TABLE 4. List of Known CD-Associated HLA Alleles

	HLA Allele	Correlation with *C4* Copy Number[a]
HLA class II	HLA-DRB1*01:03	1.97×10^{-04}
	HLA-DRB1*01:01	4.94×10^{-04}
	HLA-DRB1*03:01	2.38×10^{-01}
	HLA-DRB1*07:01	3.86×10^{-03}
	HLA-DRB1*08:01	2.52×10^{-05}
	HLA-DRB1*16:01	9.14×10^{-05}
	HLA-DRB1*13:01	1.16×10^{-03}
	HLA-DPA1*01:03	1.74×10^{-03}
HLA class I	HLA-C*06:02	7.57×10^{-03}
	HLA-B*14:02	5.65×10^{-02}
	HLA-B*35:03	1.77×10^{-04}
	HLA-B*52:01	1.59×10^{-04}
	HLA-B*35:02	2.78×10^{-03}
	HLA-A*03:01	8.35×10^{-04}
	HLA-C*14:02	4.55×10^{-04}

[a] r^2 value from linear regression.

Differential Impact of C4 Copy Number in CD Patients Compared with Healthy Control Subjects

To see if *C4* copy numbers correlated with C4 serum levels in CD patients and healthy control subjects, we measured C4 serum levels in both the groups. C4 serum concentrations ranged from 0.15 to 0.61 g/L in CD patients and from 0.12 to 0.75 g/L in healthy control subjects and were similar between both the groups ($median_{healthy_controls} = 0.32$; IQR; 0.26–0.42; $median_{CD} = 0.33$; IQR, 0.28–0.40; $P = 4.51 \times 10^{-01}$; Fig. 3A). However, C4 serum levels were clearly related to the *C4* copy number with a higher copy number associated with higher serum levels (Fig. 3B–D). But this relationship differed for patients and control subjects: a regression model including *C4* copy number, an indicator for patient status, and their interaction term showed that higher copy numbers lead to higher serum concentrations in control subjects compared with CD patients (adjusted $R^2 = 0.26$; $P < 0.001$).

DISCUSSION

Copy number variants have been predicted to have an important role in genetic susceptibility to common disease, and several CNV associations have been described.[10] This genome-wide CNV analysis found several CNV regions of potential interest in CD pathogenesis. We focused on the CNV association on chromosome 6p21 in the MHC region and found that low copy numbers of the *C4L* and high copy numbers of the *C4S* gene variants of complement component C4 are risk factors for CD. Complement component C4 copy numbers are also associated with other immune diseases: low *C4* copy numbers are risk factors for systemic lupus erythematosus[28,30] and an increased frequency of *C4B* deficiency is seen in rheumatoid arthritis.[31] Complement component C4 is essential for the activation of the classical and lectin complement pathways and as such, an important effector protein of the immune system. The complement pathway is involved in the clearance of immune complexes and in cytolysis or neutralization of invading microbes.[32,33] The 2 C4 isotypes, C4A and C4B, share >99% amino acid sequence identity and differ in only 5 bp in a 17-bp span. C4A proteins take part in removing immune complexes, whereas C4B proteins are active in the defense against intruders.[33,34] The exact function of C4L compared with C4S is not clear, but it was shown that C4L confers protection against exogenous retroviral attacks.[35,36] We observed a differential effect of total *C4* copy number on C4 serum levels, where higher copy numbers led to higher serum concentrations in control subjects than in patients.

We found only relatively low levels of correlation between *C4* copy number with surrounding SNPs and class I and II HLA alleles and amino acids. A previous study by Fernando et al[37] looking at the correlation between complement *C4* copy number and surrounding SNPs, and integrating HLA-DRB1 allele data, in the HapMap CEU cohort and the 1958 British Birth cohort subjects also found only moderate levels of correlation and thus do not support the use of SNP genotypes as proxies for complement *C4* copy number. Importantly, *C4L* and *C4S* remained

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

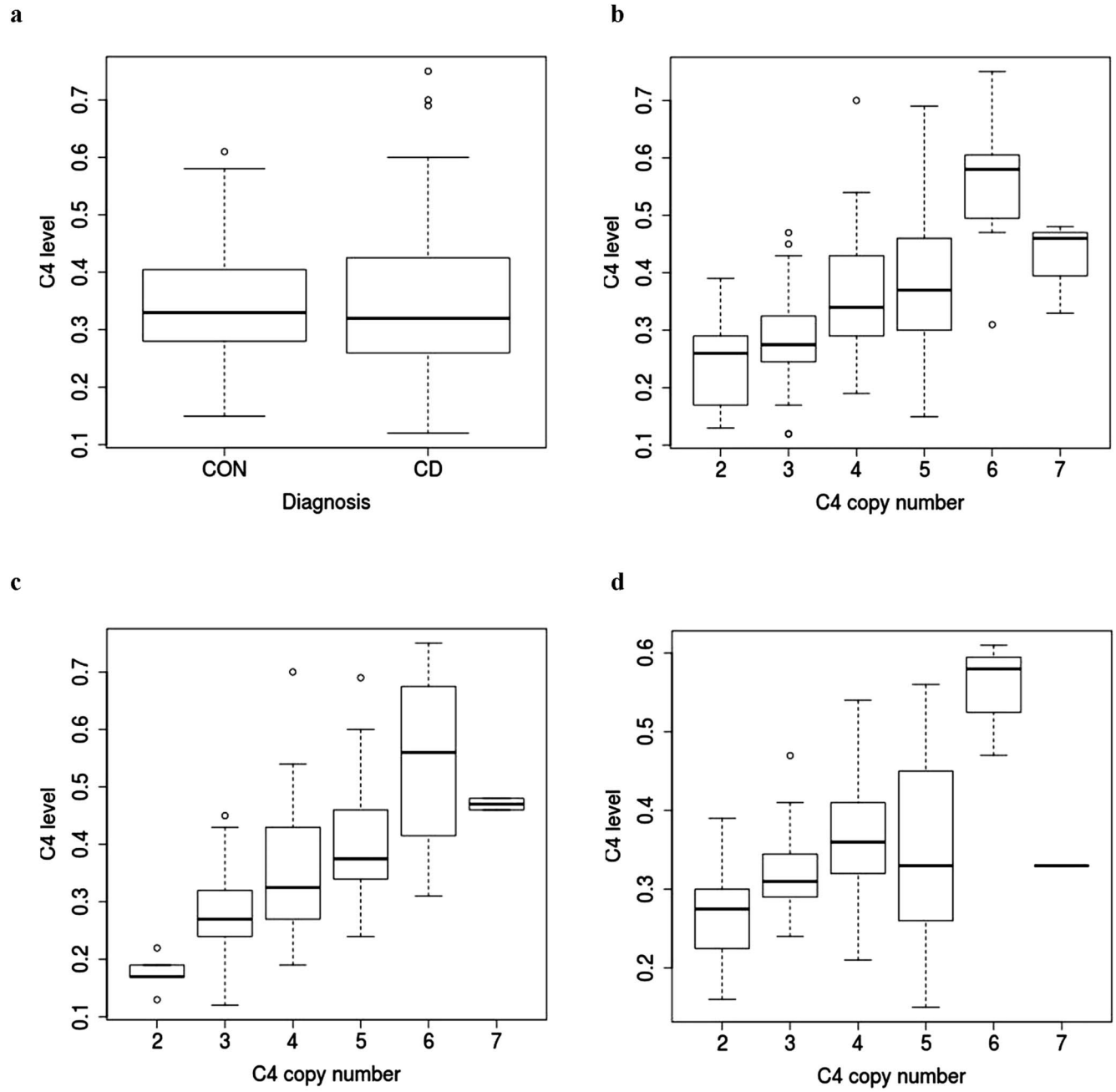


FIGURE 3. C4 serum levels in CD patients and healthy control subjects (a). C4 levels broken down by C4 copy number in all samples (b), healthy control subjects (c) and CD patients (d).

significantly associated with CD when conditioning on all known associated HLA alleles.

Of the other 20 BAC clones with lowest mean $\log_2$ ratios, 3 were located on chromosome 10q24, encompassing *HIF1AN* and *PAX2*, within 1 Mb of a known susceptibility locus.[6] Clone RP11-710A19, overlapping with the most 3′ of these clones (RP11-12F9), had a mean $\log_2$ ratio of −0.11. *PAX2* is expressed in a subset of lymphocytes[38] and represses human β-defensin1 expression.[39] Therefore, it will be interesting to further investigate this region, the more because a higher resolution CNV region as defined by Conrad et al[40] is situated in this region. One of the top 20 clones overlaps with the *MUC3A/B* gene on chromosome 7q22, a known IBD susceptibility locus.[6] The adjacent clone RP5-1059M17 had a mean $\log_2$ ratio of −0.11. *MUC3* is found to be less expressed in ileal mucosa of CD patients compared with healthy control subjects.[41] The *MUC3* gene contains long stretches of tandem repeats. Rare variable number of tandem repeats in *MUC3* are associated with ulcerative colitis.[42] The second

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

ranked clone (RP5-1050D4, mean $\log_2 = -0.18$) is located on chromosome 17p13, overlapping with among others *SLC25A11*, SPAG7, KIF1C, and *ENO3*. This region is prone to deletions and translocations in, for example, Asperger syndrome patients[43] and cancer patients.[44] *SLC25A11*, one of the genes in this region, interacts with *ATG16L1*—a known CD susceptibility gene—leading to IL1B secretion and antibacterial autophagy.[45] RP5-1005L2 and RP11-424K17 (rank 10 and 19) overlap with a known IBD susceptibility locus.[6] Different studies show gains and losses in these regions.[9,40,46,47] It remains to be determined if and how these predispose to CD.

Some studies showed an association of β-defensin 2 (*DEFB4*) copy number on chromosome 8 with CD: a lower copy number was seen in colonic CD patients compared with healthy control subjects,[48] but also an elevated *DEFB4* copy number was seen in CD, irrespective of intestinal location.[49] Other studies failed to show an association between *DEFB4* copy number and CD.[14,50] One clone from the BAC array overlaps with the *DEFB4* location on chromosome 8 (RP11-739E3) but with a mean $\log_2$ ratio of −0.004 does not show any evidence for differential copy numbers in cases versus controls in this study. Four BAC clones (RP11-263K13, RP11-678D15, RP13-511L2, and CTD2380M22) overlapped with the known *IRGM* 20-kb deletion polymorphism.[13] None of these showed a deviating mean $\log_2$ ratio. The CD-associated CNV region on chromosome 17 (CNVR7113.6) was not covered in the aCGH.[14] The fact that we could not replicate the known CD-associated CNV regions in the aCGH might be explained by the low resolution of the aCGH compared with the size of the known CNVs. The 4 clones overlapping with the *IRGM* deletion, for example, have a length range of 83.4 to 159.9 kb, whereas the *IRGM* deletion itself is 20 kb.

We acknowledge that our study has some limitations. We used BAC arrays for the CNV scan, which have a lower resolution than oligonucleotide arrays. We also lost some resolution by using a pooling strategy for the samples. Another limitation of this study is that technically it is very difficult to define the exact copy number of multiallelic CNVs as for the *C4* gene. Advantages and disadvantages of different methods are described in various studies.[26,51,52] We used real-time qPCR because it has a high sensitivity and throughput. However, qPCR results can be influenced by differential sample preparation, storage conditions, and DNA degradation. In addition, because qPCR rounds data to the nearest integer ("binning"), copy numbers can be misscored by several copies. We performed 4 replicates for each sample, which were repeated when not identical. As an additional validation, the *C4L* copy number had to be smaller than *C4A* and *C4B* combined. Another limitation of the qPCR method, instead of, for example, Southern blot techniques, is that we could not determine the precise localization and order of the long and short *C4* genes and the *C4A* and *C4B* genes in multimodular RCCX haplotypes. The RCCX segment can be present up to 4 times in the haploid genome,[53,54] and the precise localization and order of *C4A*/*C4B* and *C4S*/*C4L* genes within the multimodular RCCX segments might determine the association instead of the discrete *C4* copy number. RCCX modular variations/haplotypes have indeed been associated with disease.[28,30,54]

Our study identified complement component *C4* as a new susceptibility gene for CD, and we provide evidence for the complement system as an important contributor to CD pathogenesis. Also, we shed light on the MHC association observed in CD: in addition to the often studied class I and class II HLA SNPs, alleles, and amino acids, we show association of the class III HLA antigen C4 with CD.

ACKNOWLEDGMENTS

The authors thank all the individuals who contributed samples and the physicians and nursing staff who helped with recruitment. We also thank Nooshin Ardeshir Davani, Tamara Coopmans, Willem-Jan Wollants, Karolien Claes, and Sophie Organe for the technical support; and Vera Ballet for the database management and recruitment of patient samples. Phyllis Verstappen, Cindy Vandoren, and Ida Tassens are also thanked for their administrative support.

REFERENCES

1. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46–54.e42; quiz e30.
2. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature*. 2007;448:427–434.
3. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* 2011;43:246–252.
4. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008;40:955–962.
5. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* 2011;42:1118–1125.
6. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
7. Rioux JD, Goyette P, Vyse TJ, et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci U S A.* 2009;106:18680–18685.
8. Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet.* 2004;5:889–899.
9. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–454.
10. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 2007;3:1787–1799.
11. de Cid R, Riveira-Munoz E, Zeeuwen PL, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet.* 2009;41:211–215.
12. Hollox EJ, Huffmeier U, Zeeuwen PL, et al. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet.* 2008;40:23–25.
13. McCarroll SA, Huett A, Kuballa P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet.* 2008;40:1107–1112.
14. Craddock N, Hurles ME, Cardin N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*. 2010;464:713–720.
15. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl.* 1989;170:2–6; discussion 16–19.
16. Podolsky DK. Inflammatory bowel disease. *N Engl J Med.* 2002;347:417–429.
17. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease:

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.

report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol.* 2005;19(suppl A):5–36.
18. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16: 1215.
19. Le Caignec C, Spits C, Sermon K, et al. Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Res.* 2006;34:e68.
20. Menten B, Maas N, Thienpont B, et al. Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J Med Genet.* 2006;43:625–633.
21. Hansen BB, Klopfer SO. Optimal full matching and related designs via network flows. *J Comput Graph Stat.* 2006;15:609–627.
22. Vermeesch JR, Melotte C, Froyen G, et al. Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis. *J Histochem Cytochem.* 2005;53:413–422.
23. Konings P, Vanneste E, Jackmaert S, et al. Microarray analysis of copy number variation in single cells. *Nat Protoc.* 2012;7:281–310.
24. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Res.* 2013;41:D48–D55.
25. Flachsbart F, Caliebe A, Heinsen FA, et al. Investigation of complement component C4 copy number variation in human longevity. *PLoS One*. 2014;9:e86188.
26. Wu YL, Savelli SL, Yang Y, et al. Sensitive and specific real-time polymerase chain reaction assays to accurately determine copy number variations (CNVs) of human complement C4A, C4B, C4-long, C4-short, and RCCX modules: elucidation of C4 CNVs in 50 consanguineous subjects with defined HLA genotypes. *J Immunol.* 2007;179: 3012–3025.
27. Goyette P, Boucher G, Mallon D, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet.* 2015;47:172–179.
28. Wu YL, Yang Y, Chung EK, et al. Phenotypes, genotypes and disease susceptibility associated with gene copy number variations: complement C4 CNVs in European American healthy subjects and those with systemic lupus erythematosus. *Cytogenet Genome Res.* 2008;123:131–141.
29. Mack M, Bender K, Schneider PM. Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region. *Immunogenetics*. 2004;56:321–332.
30. Yang Y, Chung EK, Wu YL, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet.* 2007;80:1037–1054.
31. Rigby WF, Wu YL, Zan M, et al. Increased frequency of complement C4B deficiency in rheumatoid arthritis. *Arthritis Rheum.* 2012;64: 1338–1344.
32. Carroll MC. The role of complement and complement receptors in induction and regulation of immunity. *Annu Rev Immunol.* 1998;16:545–568.
33. Yu CY, Chung EK, Yang Y, et al. Dancing with complement C4 and the RP-C4-CYP21-TNX (RCCX) modules of the major histocompatibility complex. *Prog Nucleic Acid Res Mol Biol.* 2003;75:217–292.
34. Blanchong CA, Chung EK, Rupert KL, et al. Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. *Int Immunopharmacol.* 2001;1: 365–392.
35. Saxena K, Kitzmiller KJ, Wu YL, et al. Great genotypic and phenotypic diversities associated with copy-number variations of complement C4 and RP-C4-CYP21-TNX (RCCX) modules: a comparison of Asian-Indian and European American populations. *Mol Immunol.* 2009;46:1289–1303.
36. Yang Y, Chung EK, Zhou B, et al. Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities, and body mass index. *J Immunol.* 2003;171:2734–2745.
37. Fernando MM, Boteva L, Morris DL, et al. Assessment of complement C4 gene copy number using the paralog ratio test. *Hum Mutat.* 2010;31: 866–874.
38. Zhai QJ, Ozcan A, Hamilton C, et al. PAX-2 expression in non-neoplastic, primary neoplastic, and metastatic neoplastic tissue: a comprehensive immunohistochemical study. *Appl Immunohistochem Mol Morphol.* 2010;18:323–332.
39. Bose SK, Gibson W, Bullard RS, et al. PAX2 oncogene negatively regulates the expression of the host defense peptide human beta defensin-1 in prostate cancer. *Mol Immunol.* 2009;46:1140–1148.
40. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464:704–712.
41. Buisine MP, Desreumaux P, Debailleul V, et al. Abnormalities in mucin gene expression in Crohn's disease. *Inflamm Bowel Dis.* 1999;5:24–32.
42. Kyo K, Parkes M, Takei Y, et al. Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3. *Hum Mol Genet.* 1999;8:307–311.
43. Tentler D, Johannesson T, Johansson M, et al. A candidate region for Asperger syndrome defined by two 17p breakpoints. *Eur J Hum Genet.* 2003;11:189–195.
44. Tsuchiya E, Tanigami A, Ishikawa Y, et al. Three new regions on chromosome 17p13.3 distal to p53 with possible tumor suppressor gene involvement in lung cancer. *Jpn J Cancer Res.* 2000;91:589–596.
45. Lassen KG, Xavier RJ. An alteration in ATG16L1 stability in Crohn disease. *Autophagy*. 2014;10:1858–1860.
46. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453:56–64.
47. Perry GH, Yang F, Marques-Bonet T, et al. Copy number variation and evolution in humans and chimpanzees. *Genome Res.* 2008;18:1698–1710.
48. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet.* 2006;79:439–448.
49. Bentley RW, Pearson J, Gearry RB, et al. Association of higher DEFB4 genomic copy number with Crohn's disease. *Am J Gastroenterol.* 2010; 105:354–359.
50. Aldhous MC, Abu Bakar S, Prescott NJ, et al. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Hum Mol Genet.* 2010;19:4930–4938.
51. Cantsilieris S, White SJ. Correlating multiallelic copy number polymorphisms with disease susceptibility. *Hum Mutat.* 2013;34:1–13.
52. Cantsilieris S, Baird PN, White SJ. Molecular methods for genotyping complex copy number polymorphisms. *Genomics*. 2013;101:86–93.
53. Blanchong CA, Zhou B, Rupert KL, et al. Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in caucasians. The load of RCCX genetic diversity on major histocompatibility complex-associated disease. *J Exp Med.* 2000;191:2183–2196.
54. Yang Z, Mendoza AR, Welch TR, et al. Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). A mechanism for gene deletions and disease associations. *J Biol Chem.* 1999;274:12147–12156.

Copyright © 2015 Crohn's & Colitis Foundation of America, Inc. Unauthorized reproduction of this article is prohibited.