npg

ARTICLE

Association analysis of copy numbers of FC-gamma receptor genes for rheumatoid arthritis and other immune-mediated phenotypes

Lude Franke[1,14], Hanane el Bannoudi[2,14], Diahann TSL Jansen[2], Klaas Kok[1], Gosia Trynka[3,4,5], Dorothee Diogo[3,4,5], Morris Swertz[1,6], Karin Fransen[1,7], Rachel Knevel[2], Javier Gutierrez-Achury[1], Lisbeth Ärlestig[8], Jeffrey D Greenberg[9], Joel Kremer[10], Dimitrios A Pappas[11], Alexandros Kanterakis[1,6], Rinse K Weersma[7], Annette HM van der Helm-van Mil[2], Viktor Guryev[12], Solbritt Rantapää-Dahlqvist[8], Peter K Gregersen[13], Robert M Plenge[3,4,5], Cisca Wijmenga[1], Tom W-J Huizinga[2], Andreea Ioan-Facsinay[2], Rene EM Toes[2,15] and Alexandra Zhernakova*[,1,2,15]

Segmental duplications (SDs) comprise about 5% of the human genome and are enriched for immune genes. SD loci often show copy numbers variations (CNV), which are difficult to tag with genotyping methods. CNV in the Fcγ receptor region (*FCGR*) has been suggested to be associated with rheumatic diseases. The objective of this study was to delineate association of *FCGR*-CNV with rheumatoid arthritis (RA), coeliac disease and Inflammatory bowel disease incidence. We developed a method to accurately quantify CNV in SD loci based on the intensity values from the Immunochip platform and applied it to the *FCGR* locus. We determined the method's validity using three independent assays: segregation analysis in families, arrayCGH, and whole genome sequencing. Our data showed the presence of two separate CNVs in the *FCGR* locus. The first region encodes *FCGR2A*, *FCGR3A* and part of *FCGR2C* gene, the second encodes another part of *FCGR2C*, *FCGR3B* and *FCGR2B*. Analysis of CNV status in 4578 individuals with RA and 5457 controls indicated association of duplications in the *FCGR3B* gene in antibody-negative RA ($P = 0.002$, OR = 1.43). Deletion in *FCGR3B* was associated with increased risk of antibody-positive RA, consistently with previous reports ($P = 0.023$, OR = 1.23). A clear genotype–phenotype relationship was observed: CNV polymorphisms of the *FCGR3A* gene correlated to CD16A expression (encoded by *FCGR3A*) on CD8 T-cells. In conclusion, our method allows determining the CNV status of the *FCGR* locus, we identified association of CNV in *FCGR3B* to RA and showed a functional relationship between CNV in the *FCGR3A* gene and CD16A expression.
European Journal of Human Genetics (2016) 24, 263–270; doi:10.1038/ejhg.2015.95; published online 13 May 2015

INTRODUCTION

Segmental duplications (SDs) are loci with two or more, highly similar, duplicated regions. These loci cover about 5% of the human genome and are enriched for immune-mediated genes.[1,2] One of the SDs encodes the low-affinity human FC-gamma receptors (FcγR). FcγR are glycoproteins that bind the Fc region of IgG and have a pivotal role in many immunological processes.[3–8] FcγR are expressed by various immune cell types and provide a critical link between the humoral and the cellular compartments of the immune system. The proteins encoded by FCGR genes (CD16, encoded by FCGR3, and CD32, encoded by FCGR2) are cell surface receptors that are involved in the process of phagocytosis and the clearing of immune complexes.

Genetic analysis of the FCGR locus in genome-wide association studies (GWAS) has been limited, because genes encoding the FcγR molecules are located in highly homologous SD blocks. The low-affinity FCGR gene family is located in blocks of two repeats of ~82 kb with >98% being identical. It includes three FCGR2 genes (FCGR2A (CD32A), FCGR2B (CD32B) and FCGR2C (CD32C)) and two FCGR3 genes (FCGR3A (CD16A) and FCGR3B (CD16B)). The genetic structure of the low-affinity FCGR gene family in relation to blocks of SD is presented in Figure 1a–c.

Because of the complex genetic structure of the *FCGR* locus, only the promoter and the first three exons of the *FCGR2A* gene are tagged by genetic variants present on current genome-wide genotyping

[1]Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands; [2]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands; [3]Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [4]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [6]Genomics Coordination Center, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; [7]Department of Gastroenterology and Hepatology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands; [8]Department of Public Health and Clinical Medicine/Rheumatology, Umeå University, Umeå, Sweden; [9]Department of Medicine, New York University School of Medicine, New York, New York, USA; [10]Department of Medicine, Albany Medical College, Albany, New York, USA; [11]Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York, USA; [12]Laboratory of Genome Structure and Ageing, European Research Institute for the Biology of Ageing, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands; [13]Feinstein Institute for Medical Research, Manhasset, New York, USA
*Correspondence: Dr A Zhernakova, Department of Genetics, University Medical Centre Groningen, P.O. Box 30001, Groningen 9700RB, The Netherlands. Tel: +31 50 3617100; Fax: +31 50 3617230; E-mail: a.zhernakova@umcg.nl
[14]These authors wish it to be known that, in their opinion, they should be regarded as joint first authors.
[15]These authors wish it to be known that, in their opinion, they should be regarded as joint last authors.
Received 31 July 2014; revised 27 March 2015; accepted 15 April 2015; published online 13 May 2015
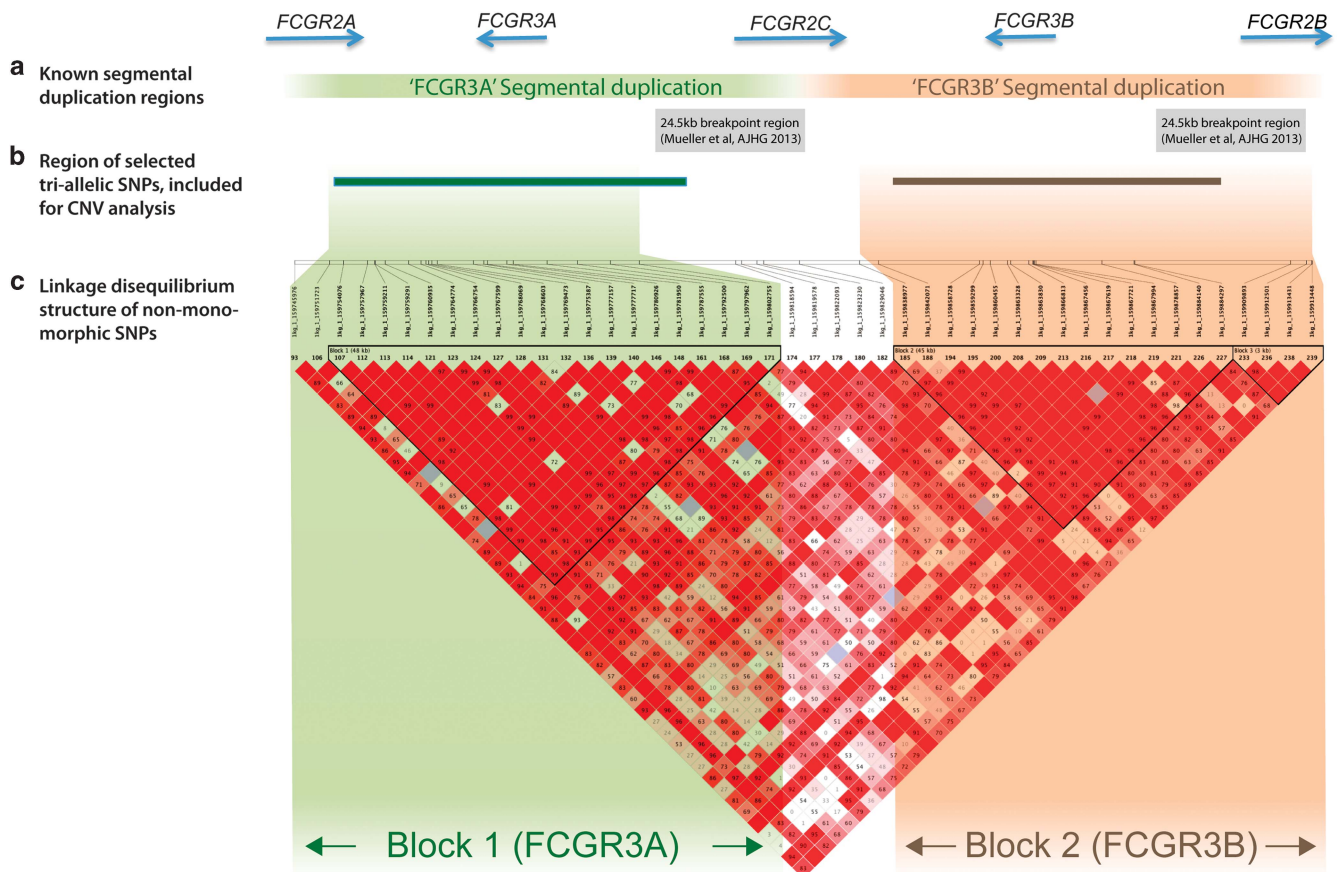
Figure 1 The genetic structure of the *FCGR* locus. The genetic structure of the low-affinity *FCGR* gene family in relation to blocks of SD and LD in the locus. The LD block is constructed based on analysis of 43 single-nucleotide polymorphisms that passed quality control and had MAF > 5% in Dutch RA and control cohorts.

platforms. SNPs in the *FCGR2A* gene (outside the SD locus) are strongly associated with inflammatory bowel disease (IBD) and in particular to its subgroup ulcerative colitis (UC).[9,10] Less strong association was reported with systemic lupus erythematosus (SLE) (rs1801274:A > G, $P = 6.78 \times 10^{-7}$)[11] and rheumatoid arthritis (RA) (rs12746613:C > T, $P = 2 \times 10^{-5}$).[12] The exact genetic risk of the whole locus contributing to disease is unknown and cannot be assessed by GWAS. However, several reports of deletions and duplications of FcγR genes have been published.[13–18] The relevance of copy number variants (CNV) in the FcγR locus in SLE has been strongly suggested by several studies,[4,15,18–21] whereas similar analyses in RA showed contradictory results.[18]

At least four experimental techniques have been proposed to quantify copy numbers within this SD. These include quantitative PCR,[19] comparative genomic hybridization (arrayCGH),[22] multiplex ligand-dependent probe amplification[23] and several variations of paralogous ratio tests (PRT), including the use of PRT in combination with quantitative sequence variant (QSV) assay.[17,24] However, none of these methods provides an accurate estimation of both the number and location of CNVs. Methods based on quantitative PCR give a continuous, rather than discrete, value of CNVs (reviewed in McKinney et al[18]). Other methods aim to design sets of primers and probes specific to the *FCGR3A* or *FCGR3B* genes, which is a complicated task given the high homology of these genes. The PRT method, including the genotyping of a reference diploid sequence on chromosome 18, is probably the most robust reported to date,[24] but it

does not allow identification of CNV boundaries. Recently, the combination of a PRT probe with QSV was applied to an RA case–control analysis, allowing an estimation of CNV in 1115 RA patients and 654 controls.[17] This study suggested that there is an association with lower CNV in the *FCGR3B* gene in RA patients.

The Immunochip is an Illumina array, which includes SNPs to fine map 186 distinct loci associated with at least 1 of the 12 immune-mediated diseases, including RA, CeD and IBD.[9,25–28] The low-affinity FcγR locus is also included in the Immunochip platform for fine-mapping. The Immunochip was not specifically designed to quantify CNVs, as its primary focus was on calling bi-allelic SNP genotypes. Previously, several methods of quantifying CNV changes based on intensity values of genotyping probes have been proposed.[29–37] As these methods rely upon raw intensity measures, it is crucial to properly account for any systematic differences that may exist (eg, batch effects), which is especially important when comparing many thousands of samples that have been hybridized in different laboratories. In this study, we developed a highly robust method to accurately quantify *FCGR*-CNV using the Immunochip, by extensively correcting for various confounders through principal component analysis (PCA). We confirmed our results using three independent methods (segregation analysis in families, arrayCGH and next-generation sequencing (NGS) analysis) and then applied our method to a cohort of 4578 RA patients as well as 5235 individuals with two other immune-mediated diseases (CeD and IBD) and 7941 controls. We identified association of CNVs in FCGR3B in RA. Functional

analysis showed that copy numbers in the *FCGR* locus have a clear effect on the expression of FcγRs in major blood cell types.

MATERIALS AND METHODS

Sample collection

After quality control, our sample collection included 4578 RA cases from three populations (the Netherlands, Sweden and the USA), 1477 IBD samples (900 CD and 577 UC, all from the Netherlands), 3758 CeD cases from five populations (the Netherlands, Poland, Spain, Italy and India) and 7941 matched population controls. In total, 17 754 samples were included in the case–control analysis. In addition, 285 offspring from the Genome of the Netherlands (GoNL, www.nlgenome.nl)[38] were included for segregation analysis, giving a total of 18 039 individuals (Table 1). Written informed consent was obtained from all subjects; the research was approved by the ethics committee or institutional review boards of all the participating institutions.

SNP genotyping

Samples were genotyped using the Immunochip according to Illumina's (San Diego, CA, USA) protocols at five laboratories (listed in Supplementary Methods). The final report file, including R intensity values, was extracted for all *FCGR* and reference SNPs, as listed in Supplementary Table S1.

Data quality control

The Illumina GenomeStudio GenTrain2.0 algorithm was used to cluster samples. Only samples with call rates >99% that also passed the QC for the primary Immunochip analysis in each disease (described in Jostins *et al*,[9] Eyre *et al*[25] and Trynka *et al*[26]) were included. We performed an extra check to exclude any duplicates or first-degree relatives in the combined analysis.

FCGR copy number count quantification

The algorithm for CNV estimation is described in detail in the Supplementary Methods section.

Identification of CNVs by arrayCGH

In 22 individuals, the complete *FCGR* locus was assessed by a dense set of arrayCGH probes, using a custom-designed array (Agilent, Santa Clara, CA, USA; ID 029465). In total, 2704 arrayCGH probes were included in the extended *FcγR* locus, of which 1171 were located in the SD (see Supplementary Methods for further details).

Sequencing analysis

We selected 20 individuals with various combinations of CNVs in blocks 1 and 2 from the GoNL study,[38] for whom, on average, 14x whole genome sequences were made and Immunochip SNP genotypings were available. We predicted their CNV status using a dynamic window approach (DWAC-Seq, http://tools.genomes.nl/dwac-seq.html)), see Supplementary Methods for details.

Statistical analysis

CNV quantification was performed using the customized Java software (available at https://github.com/molgenis/systemsgenetics/wiki/Copy-number-determination-using-ImmunoChip-intensity-data-for-the-FCGR-locus.

Association analysis was done by chi-square testing using SPSS (Armonk, NY, USA) and R. Correlation of genotypes and expression of the protein was performed in SPSS v19. Meta-analysis was performed using an inverse variance meta-analysis using R. Power calculations were performed using http://pngu.mgh.harvard.edu/~purcell/gpc/. The Breslow–Day test for genetic heterogeneity was performed using R. The haplotype analysis was done using Haploview,[39] default settings.

Functional studies

To stain CD16 molecules on different cell types, fresh leukocytes were isolated from the blood of 21 healthy individuals using HetaSep (StemCell Technologies, Vancouver, BC, Canada). Cells were stained with antibodies against CD3, CD4, CD8, CD14, CD16 and CD19.

To identify which isoform of CD16 (CD16a or CD16b) was expressed on CD8+ cells, leukocytes were left untreated or were treated with 5 U/ml phosphatidylinositol-specific phospholipase C (PI-PLC) for 1 h at 37 °C under constant mixing. After extensive washing, cells were stained with antibodies

Table 1 Sample collection

Population	Controls (% females)	Cases (% females)	CCPpos (% females)	CCPneg (% females)	CCPunkn	Total
RA cohort						
NL	1379 (44%)	648 (67%)	332 (67%)	300 (66%)	16	2027
SW	942 (68%)	675 (68%)	406 (67%)	190 (65%)	79	1617
US_NY	1275 (69%)	2238 (74%)	1567 (74%)	531 (76%)	140	3513
US_B	1861 (75%)	1017 (79%)	1006 (79%)	0	11	2878
Total	5457	4578	3311	1021	246	10035

Population	Controls (% females)	Cases (% females)	Total
Coeliac disease			
NL	1379 (44%)	1110 (68%)	2489
Italy	1255 (81%)	1374 (67%)	2629
Poland	532 (42%)	505 (63%)	1037
Spain	306 (53%)	545 (68%)	851
India	391 (16%)	224 (55%)	615
Total	3863	3758	7621

Population	Controls (% females)	IBD cases (% females)	CrD (% females)	UC (% females)	Total
IBD					
NL	1379 (44%)	1548 (65%)	900 (65%)	577 (48%)	2927

All study	Controls	Cases	Family members	All samples
Total	7941	9813	285	18039

against CD3, CD8, CD15 and CD16 – or an irrelevant isotype-matched antibody for 20 min at 4 °C. The expression of CD16 on CD8+ T cells, and neutrophils was assessed by FACS. Detailed information is described in Supplementary Methods.

Data sharing

The results of this study are submitted to DGVArchive (http://www.ebi.ac.uk/dgva/data-download), submission number estd222.

RESULTS

Haplotype structure of the *FCGR* locus

To gain insight into the haplotype structure of the *FCGR* locus, we first investigated the Immunochip Genome Studio cluster plots for 1159 SNPs located in the block of SD in Dutch RA cases and controls. Only 114 SNPs (9.8%) passed our quality criteria for SNP genotyping (minor allele frequency (MAF) > 0.1%, Hardy Weinberg equilibrium (HWE) *P*-value > 0.0001). Of these, 75 SNPs (6.5%) had a MAF > 1%, while only 43 SNPs (3.7%) passed the quality criteria with a MAF > 5%.

Haplotype analyses on SNPs with MAF > 5% showed strong linkage disequilibrium (LD) across the whole locus but identified two LD blocks. The split between the two blocks (*D*' = 0.68) corresponded approximately to the split between the two duplicated regions (Figure 1a–c). As the two CNV blocks included the complete sequence of the *FCGR3A* gene (left block) and the *FCGR3B* gene (right block), respectively, we will refer to these as the *FCGR3A* and *FCGR3B* CNV loci, although the borders of both CNVs are wider than the *FCGR3A* or 3B genes (see Figure 1a–c).

Visual inspection of all 1159 SNP clusters clearly indicated the presence of several SNPs with a CNV pattern (ie, more than three clusters were visible per SNP, see Supplementary Figure S1). In total, at least 13 SNPs with a CNV pattern were located in the *FCGR3A* block, and at least 11 SNPs showed a CNV pattern in the *FCGR3B* block (Supplementary Tables S2 and S3). Next, we developed an algorithm to estimate CNV in the *FCGR* locus based on analysing the fluorescent dye intensities of multiple SNPs genotyped on the Illumina Immunochip platform, as described in the Supplementary Methods.

We performed the analysis of CNVs in the *FCGR* locus in three ways: (1) All 1159 SNPs were included, indicating the average number of copies in both *FCGR3A* and *FCGR3B* blocks (Supplementary Table S1). (2) PCA of 13 CNV-like SNPs from the *FCGR3A* block (Supplementary Table S2). (3) PCA of 11 CNV-like SNPs from the *FCGR3B* block (Supplementary Table S3). In all the three analyses, the first principal component from the PCA of the *FCGR* SNP intensity data correlated the number of copies for each individual (Supplementary Figures S2a–c). We also investigated whether any of the first 10 PCs from the combined analysis of all 1159 *FCGR* SNPs reflected the CNV status of *FCGR3A* and/or *FCGR3B* blocks, and we found that the third PC does correspond to the CNV status of the *FCGR3A* block. Consistent with this observation, of the 100 top SNPs that explain most of the third PC, 97 are annotated in the *FCGR3A* block. Combining the first and third PCs from the 1159 SNP analysis showed the clearer cluster separations in all populations, concordant with the results obtained by analysing the *FCGR3A* and *FCGR3B* blocks (Figure 2). In consequence, we used this method to define the CNV status. In both duplicated blocks, three, four and more copies could not be undoubtfully separated from each other; they were therefore combined in one group of ≥ 3 copies.

After estimating the CNV status for each individual, we validated our CNV estimation algorithm by three independent methods: segregation analysis in families, arrayCGH, and whole genome sequencing.

Segregation in the families

We first studied the segregation of the *FCGR* CNVs in 257 trio families to verify the outcome of our new method. For the *FCGR3A* block, the inferred *FCGR3A* CNV genotypes of the 257 trios showed perfect Mendelian inheritance, indicating that our method correctly assigned these genotypes. The same was also true for the FCGR3B block, except for two unlikely events in the distribution of FCGR3B CNV, which could occur due to recombination or uneven distribution of CNV on parental chromosomes. The genotypes of all parent–child trios are presented in Supplementary Table S4.

ArrayCGH analysis

We next used arrayCGH to confirm the CNV quantification yielded by our method. We selected 22 individuals representing three different CNVs in the *FCGR3A* locus (1, 2 and ≥ 3 copies), and four CNV types in *FCGR3B* (0, 1, 2 and ≥ 3 copies) based on the PCA analysis. The selection and CNV genotypes of these 22 samples are indicated in Supplementary Table S5.

ArrayCGH analysis showed perfect correlation with the combined number of copies of *FCGR3A*+*FCGR3B* blocks. However, the arrayCGH method could not separate the CNVs seen in the two blocks, not even when the analysis was performed only on probes annotated as unique for either block. For example, individuals with a double deletion of the *FCGR3B* block showed a low signal across the whole *FCGR* locus (2/4 copies, Supplementary Figure S3a), while those with a single deletion of the *FCGR3B* or *FCGR3A* block showed a similar pattern of deletion on the whole *FCGR* block on arrayCGH analysis (3/4 copies) (Supplementary Figures S3b–c). Similarly, individuals with three copies of *FCGR3A* or *FCGR3B* showed similar arrayCGH patterns (Supplementary Figure S3d). We concluded that arrayCGH confirmed the results of our PCA analysis but that arrayCGH cannot be used to assess the CNV structure of the *FCGR* locus accurately.

Next-generation sequencing analysis

To further confirm our PCA-based method of estimating CNVs, we compared the results with those from NGS analysis. We selected 20 individuals with various CNV distributions in the *FCGR3A* and *FCGR3B* blocks and performed a CNV analysis of the whole genome sequence data using DWAC-Seq. Of the 40 CNVs included for confirmation analysis, we observed a perfect match between the results obtained by both methods (Supplementary Table S6).

Together, the results from segregation analysis, arrayCGH and sequencing suggest that our method accurately estimates the CNV structure in the *FCGR* locus.

Structural composition of the *FCGR* locus

We next analysed the structural composition of *FCGR* CNVs in the cohort of 17 754 individuals. The *FCGR3A* deletion is extremely rare: only 2 out of the 17 754 individuals carried a double deletion (0 copies) for *FCGR3A*, and 275 individuals (1.5%) harboured a deletion of one *FCGR3A* allele. CNV in *FCGR3B* is more common: we identified 46 out of the 17 754 (0.26%) individuals with double deletion of the *FCGR3B* locus, whereas 1304 subjects (7.4%) had a deletion of one copy of the *FCGR3B* block. There were no individuals with a double deletion of both genes, and only 9 out of the 17 754 individuals had a single-copy deletion in both the *FCGR3A* and *FCGR3B* loci. The frequency of CNV per population and in every disease group is presented in Supplementary Tables S7–S9.
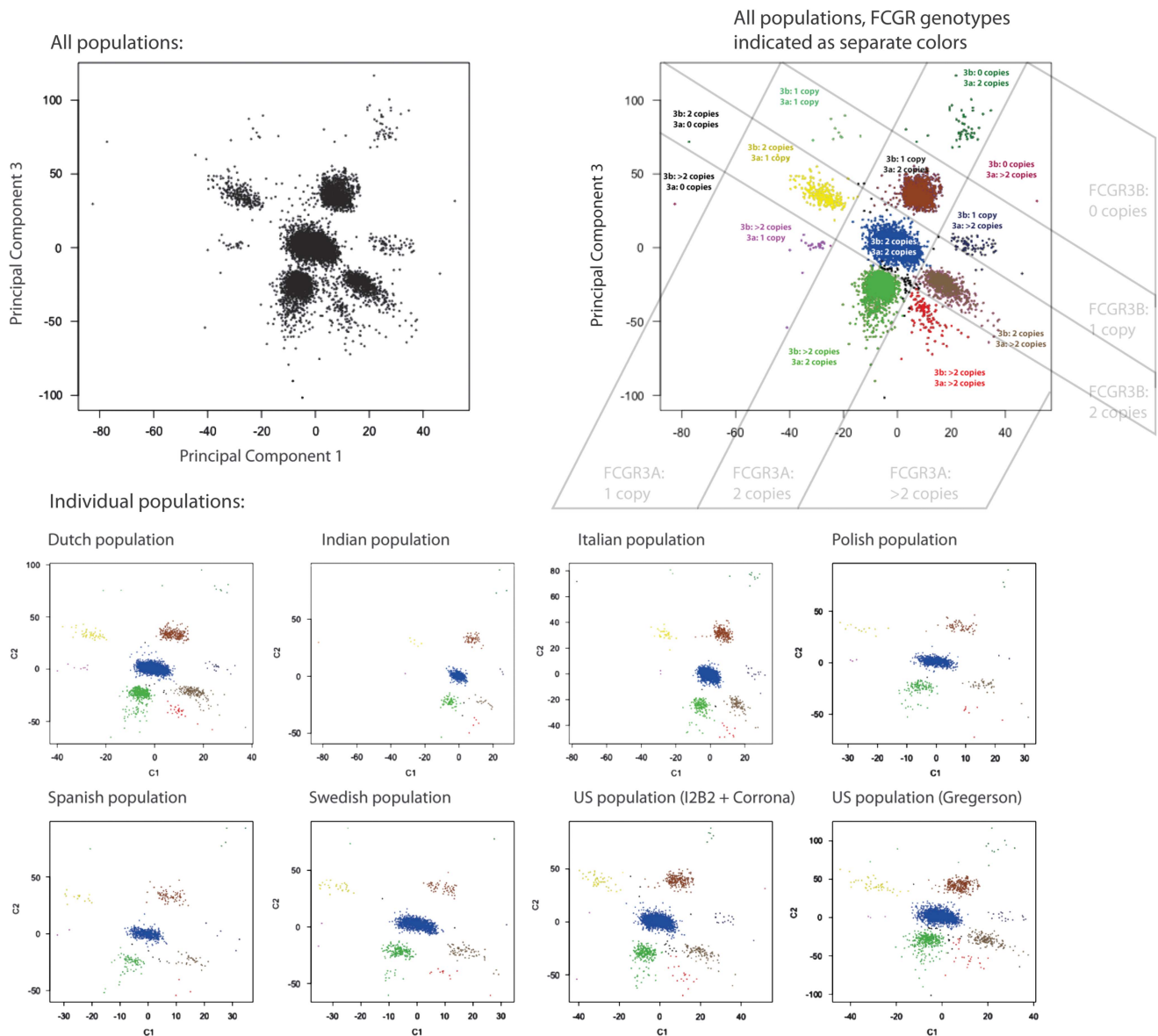
Figure 2 CNV frequency distributions of the *FCGR3A* and *FCGR3B* based on PC1 and PC3 of analysis of all 1159 single-nucleotide polymorphisms in the *FCGR* locus.

Association of CNV in *FCGR* locus with RA, IBD and CeD

We next investigated whether the CNVs in *FCGR* locus are associated with RA (including the CCP+ and CCP− groups), CeD and IBD. All analyses were performed separately per population and per disease (Supplementary Tables S7–S9) and then combined in a meta-analysis (Supplementary Table S10, Figure 3). We did not observe strong heterogeneity across populations (calculated by Breslow–Day test, Supplementary Table S11).

No significant association, in any group, was observed with deletions or duplications of *FCGR3A* block. Given the low frequency of FCGR3A deletions, it should be noted that our sample size only had sufficient power to determine strong effects of this rare variant (Supplementary Table S12).
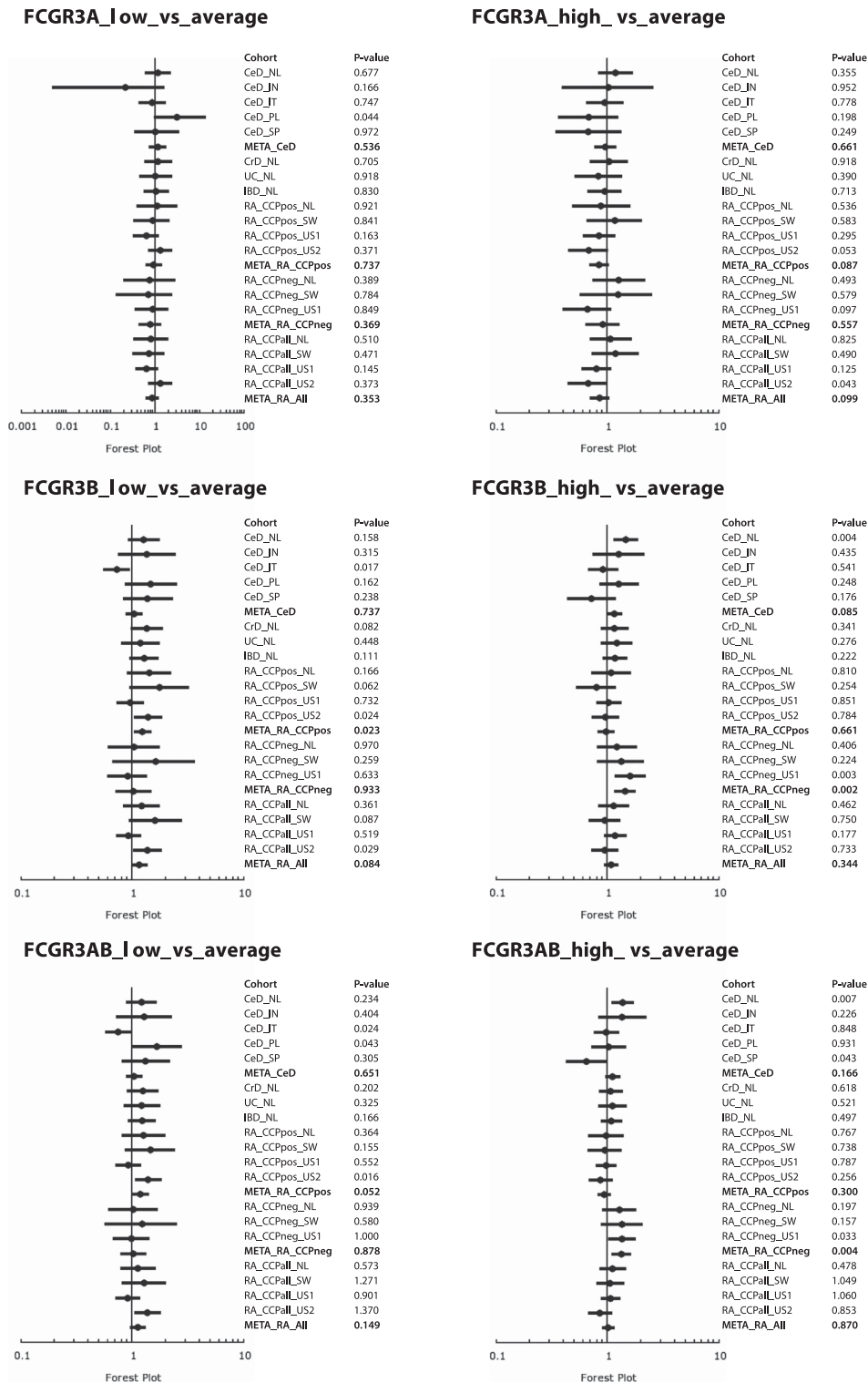
In the FCGR3B gene, we observed associations of CCP− RA with duplication of the *FCGR3B* locus, where an extra copy of the *FCGR3B* block showed a susceptible effect in CCP− RA ($P = 0.002$, OR 1.429,

95% CI (1.146–1.782)). A similar, less significant, trend was observed in CeD analysis ($P = 0.085$, OR = 1.149, 95% CI (0.981–1.345)).

Previous studies have indicated an association with deletion of the *FCGR3B* locus in CCP+ RA, although these results were inconsistent.[16–18] In our analysis, we identified nominally significant association with deletion of the *FCGR3B* locus, and the trend of association was consistent with the previous observations ($P = 0.023$, OR = 1.229, 95% CI (1.029–1.467); Figure 3 and Supplementary Table S10). In the combined analysis of *FCGR3A* and *FCGR3B* blocks, trend of association of high CNV with CCP− RA was observed ($P = 0.004$, OR = 1.328, 95% CI (1.093–1.614)).

Overall, only association of *FCGR3B* duplication with CCP− RA was significant after correction for multiple testing (Bonferroni; $P(\text{corr}) = 0.02$).

Although our method is capable of quantifying CNV in the *FCGR* locus properly, it is a fairly complicated procedure, requiring raw

FCGR3A_low_vs_average

FCGR3A_high_vs_average

FCGR3B_low_vs_average

FCGR3B_high_vs_average

FCGR3AB_low_vs_average

FCGR3AB_high_vs_average

Figure 3 Meta-analysis of *FCGR3A*, *FCGR3B* and *FCGR3A+*FCGR3B blocks in RA, CD and IBD.

intensity data of both the *FCGR* locus and other autosomal regions in order to accurately correct for batch effects. A much more straight-forward procedure would be possible if a normal bi-allelic SNP were in strong LD with the *FCGR3A* or *FCGR3B* loci. We therefore investigated whether any SNP in or around the locus could tag the

FCGR3A and/or *FCGR3B* CNVs. We found no SNP proxies to tag the CNV genotype (max $D' = 1$, $r^2 < 0.1$) (Supplementary Figure S4); we therefore concluded that CNV estimation algorithms need to be applied to genotype the *FCGR* locus properly. Given the previous reports of association of SNPs in the *FCGR* locus to RA, we have also

investigated the association of bi-allelic SNPs with RA, including the CCP+ and CCP− subgroup analysis. We looked for association in the locus chr1: 160,975,205-162143863 (from 500 kb left from start of FCGR2A till 500 kb right from end of FCGR3B gene). None of the SNPs were associated with RA with $P < 0.003$ (including analyses in the CCP+ and CCP− subgroups).

Expression of CD16 in relation to CNVs in FcγR locus. To identify the functional consequences of CNVs in *FCGR3A* and *FCGR3B*, we tested the expression of CD16 (FcγRIII) receptor on neutrophils, monocytes, B and T cells. Despite previous studies[17,40,41] reporting a correlation between neutrophil CD16b expression and *FCGR3B* copy number, we did not observe this correlation in our own samples (Supplementary Figure S5). However, we did observe a significant correlation of *FCGR3A* CNV and expression of CD16 on CD8+ T-cells (*P*-value for correlation $P = 0.0018$) (Figure 4). No correlation was observed for monocytes, B cells or CD4+ T cells. We confirmed that the Fc receptor expressed on this subtype of T cells is the CD16a form (encoded by *FCGR3A*) by using PI-PLC, an enzyme that cleaves the GPI-linked form of CD16 (ie, CD16b, encoded by *FCGR3B*). This enzyme was not able to cleave CD16 from CD8+ T cells, although it was able to cleave CD16 from neutrophils, indicating that neutrophils carry CD16b and CD8+ T cells carry CD16a (Supplementary Figure S6). Further characterization of CD8+CD16+T cells indicated that they carry the *αβ* T-cell receptor and have a terminally differentiated phenotype[42] based on the following characterization; CD45RA +CCR7−CD27–CD28− (data not shown).
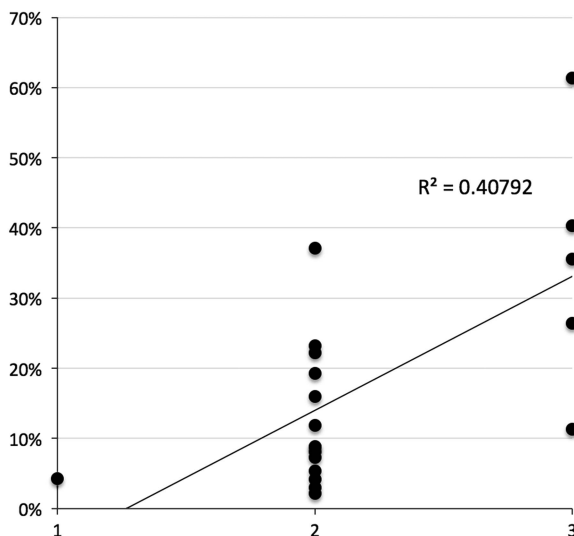
DISCUSSION

The analysis of genetic association in SD loci is challenging due to their complex genetic structure. In this study, we have developed a method for CNV analysis in SD loci and applied it to the *FCGR* locus. Our method is based on PCA of raw intensity values of SNPs in CNV regions. We confirmed our results with three independent methods: segregation analysis in families, arrayCGH, and sequencing analysis. We found that CNVs were segregated in families, as expected, and there was a good correlation of the results from our method with those from NSG analysis. We also concluded that arrayCGH method

Figure 4 Correlation of *FCGR3A* CNV and expression of CD16 on CD8 cells. X-axis: different numbers of copies of *FCGR3A*. y Axis: percentage of CD16+CD8+ cells.

is not accurate enough for fine-mapping CNVs in the *FCGR* locus, as it does not allow the specific identification of *FCGR3A* and *FCGR3B* CNVs. The CNV estimation from arrayCGH does, however, gives a good correlation with the total number of copies in both the *FCGR3A* and *FCGR3B* blocks. The boundaries of two CNV blocks identified in this study correspond approximately to the recent findings of breakpoints of *FCGR3B* deletion (24.5 kb region of identity in the *FCGR2C* and *FCGR2B* genes), identified by high-throughput sequencing analysis.[21]

Association of low copies of *FCGR3B* with RA was recently suggested in a study of a cohort of 1115 cases and 654 healthy individuals ($P = 0.028$; OR = 1.50). This association appeared stronger in CCP+ individuals ($P = 0.011$; OR = 1.61; 932 CCP+ cases included).[17] A similar trend, although not significant, was observed by another meta-analysis that included a comparable number of samples ($P = 0.15$; OR = 1.36).[18] We applied our method to a large cohort of 17 754 individuals with immune-mediated phenotypes and matched controls, including RA, CeD and IBD. As our study included 4578 RA patients (of which 3311 were CCP+ RA patients) and 5457 population-matched controls, we had sufficient power (Supplementary Table S12) to confirm the reported association mentioned above. We observed similar association with deletion in *FCGR3B* in the CCP+ subgroup of RA ($P = 0.023$, OR = 1.229), although this finding was not significant after correction for multiple testing. Similar results observed in three independent data sets suggest that association of deletion of *FCGR3B* in CCP+ RA patients is most likely a true-positive association and that the effect size of *FCGR3B* deletion on CCP+ RA is in the range of OR 1.2–1.5. We also identified an association in *FCGR3B* with the CCP− subgroup of RA, where high CNV in *FCGR3B* was associated with disease ($P = 0.002$, OR 1.429). Overall, this indicates that CNVs in the *FCGR*3B locus show a different effect on the autoantibody-positive and -negative subgroups of RA. CCP can interact with *FCGR3B* leading to activation of various immune cells and are the most specific biomarkers for RA. It is now clear that the genetic contribution to CCP-positive and -negative disease is different. Our observation that the contribution of CNVs in the *FCGR3B* appears distinct for CCP-positive and -negative RA is in line with these findings that also indicate that CCP-positive and -negative disease have a different aetiology.

We did observe a significant positive correlation of *FCGR3A* CNV and the expression of CD16 on CD8+ T-cells, which has not been reported before. Intriguingly, this correlation is not present for other immune cells that express CD16a, indicating a cell-specific regulation of the expression of this receptor. The role of CD8+ T-cell immunity in RA is presently unclear, and it is not known whether CD16 expression on CD8+ T cells is involved in the inflammatory process operative in RA. However, it would be interesting to know whether the CD16 expression of CD8 T cells differs between CD8+ T cells obtained from CCP-positive and -negative patients, as differential expression of CD16 in these two subgroups of RA would suggest a role for these cells in (CCP-positive) disease.

Overall, we have created a framework to identify deletions and duplications from intensity values of SNP genotypes on genome-wide association platforms. By confirming the results from our method with family segregation analysis, array CGH and sequencing analysis, we have shown that they are accurate and reliable and have identified a functional correlation between *FCGR3A* CNV with CD16 expression on CD8+ T cells. Our method can now be applied to other CNVs genotyped on the Immunochip platform or on any other dense SNP genotyping platform.

1 Bailey JA, Eichler EE: Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* 2006; **7**: 552–564.
2 Sharp AJ, Locke DP, McGrath SD *et al*: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; **77**: 78–88.
3 Dijstelbloem HM, Van de Winkel JG, Kallenberg CG: Inflammation in autoimmunity: receptors for IgG revisited. *Trends Immunol* 2001; **22**: 510–516.
4 Fanciulli M, Vyse TJ, Aitman TJ: Copy number variation of Fc gamma receptor genes and disease predisposition. *Cytogenet Genome Res* 2008; **123**: 161–168.
5 Von Bubnoff D, Novak N, Kraft S, Bieber T: The central role of FcepsilonRI in allergy. *Clin Exp Dermatol* 2003; **28**: 184–187.
6 Indik ZK, Park JG, Hunter S, Schreiber AD: The molecular dissection of Fc gamma receptor mediated phagocytosis. *Blood* 1995; **86**: 4389–4399.
7 Huizinga TW, Dolman KM, Van der Linden NJ *et al*: Phosphatidylinositol-linked FcRIII mediates exocytosis of neutrophil granule proteins, but does not mediate initiation of the respiratory burst. *J Immunol* 1990; **144**: 1432–1437.
8 Gessner JE, Heiken H, Tamm a, Schmidt RE: The IgG Fc receptor family. *Ann Hematol* 1998; **76**: 231–248.
9 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.
10 Anderson C, 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.
11 Harley JB, Alarcón-Riquelme ME, Criswell LA *et al*: Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008; **40**: 204–210.
12 Raychaudhuri S, Thomson BP, Remmers EF *et al*: Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. *Nat Genet* 2009; **41**: 1313–1318.
13 Nossent JC, Rischmueller M, Lester S: Low copy number of the Fc-γ receptor 3B gene FCGR3B is a risk factor for primary Sjogren's syndrome. *J Rheumatol* 2012; **39**: 2142–2147.
14 Niederer H a, Willcocks LC, Rayner TF *et al*: Copy number, linkage disequilibrium and disease association in the FCGR locus. *Hum Mol Genet* 2010; **19**: 3282–3294.
15 Fanciulli M, Norsworthy PJ, Petretto E *et al*: FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 2007; **39**: 721–723.
16 Thabet MM, Huizinga TWJ, Marques RB *et al*: Contribution of Fcgamma receptor IIIA gene 158 V/F polymorphism and copy number variation to the risk of ACPA-positive rheumatoid arthritis. *Ann Rheum Dis* 2009; **68**: 1775–1780.
17 Robinson JI, Carr IM, Cooper DL *et al*: Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis. *Hum Mutat* 2012; **33**: 741–749.
18 McKinney C, Merriman TR: Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes. *Hum Mol Genet* 2012; **21**: 2370–2376.
19 Aitman TJ, Dong R, Vyse TJ *et al*: Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 2006; **439**: 851–855.
20 Niederer HA, Clatworthy MR, Willcocks LC, Smith KGC: Fc γ RIIB, Fc γ RIIIB, and systemic lupus erythematosus. *Ann N Y Acad Sci* 2010; **1183**: 69–88.
21 Mueller M, Barros P, Witherden AS *et al*: Genomic pathology of SLE-associated copy-number variation at the FCGR2C/FCGR3B/FCGR2B locus. *Am J Hum Genet* 2013; **92**: 28–40.
22 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.
23 Breunis WB, Van Mirre E, Geissler J *et al*: Copy number variation at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B. *Hum Mutat* 2009; **30**: E640–E650.
24 Hollox EJ, Detering J-C, Dehnugara T: An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. *Hum Mutat* 2009; **30**: 477–484.
25 Eyre S, Bowes J, Diogo D *et al*: High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* 2012; **44**: 1336–1340.
26 Trynka G, Hunt KA, Bockett NA *et al*: Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; **43**: 1193–1201.
27 Cooper JD, Simmonds MJ, Walker NM *et al*: Seven newly identified loci for autoimmune thyroid disease. *Hum Mol Genet* 2012; **21**: 5202–5208.
28 Juran BD, Hirschfield GM, Invernizzi P *et al*: Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum Mol Genet* 2012; **21**: 5209–5221.
29 Wang K, Li M, Hadley D *et al*: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007; **17**: 1665–1674.
30 Colella S, Yau C, Taylor JM *et al*: QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007; **35**: 2013–2025.
31 Franke L, De Kovel CGF, Aulchenko YS *et al*: Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays. *Am J Hum Genet* 2008; **82**: 1316–1333.
32 Cardin N, Holmes C, Donnelly P, Marchini J: Bayesian hierarchical mixture modeling to assign copy number from a targeted CNV array. *Genet Epidemiol* 2011; **35**: 536–548.
33 Barnes C, Plagnol V, Fitzgerald T *et al*: A robust statistical method for case-control association testing with copy number variation. *Nat Genet* 2008; **40**: 1245–1252.
34 Pinto D, Darvishi K, Shi X *et al*: Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 2011; **29**: 512–520.
35 Korn JM, Kuruvilla FG, McCarroll SA *et al*: Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
36 Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA: Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat Genet* 2008; **40**: 1199–1203.
37 Kumasaka N, Fujisawa H, Hosono N *et al*: PlatinumCNV: a Bayesian Gaussian mixture model for genotyping copy number polymorphisms using SNP array signal intensity data. *Genet Epidemiol* 2011; **35**: 831–844.
38 Boomsma DI, Wijmenga C, Slagboom EP *et al*: The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014; **22**: 221–227.
39 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
40 Willcocks LC, Lyons PA, Clatworthy MR *et al*: Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med* 2008; **205**: 1573–1582.
41 Morris DL, Roberts AL, Witherden AS *et al*: Evidence for both copy number and allelic (NA1/NA2) risk at the FCGR3B locus in systemic lupus erythematosus. *Eur J Hum Genet* 2010; **18**: 1027–1031.
42 Björkström NK, Gonzalez VD, Malmberg K-J *et al*: Elevated numbers of Fc gamma RIIIA + (CD16+) effector CD8 T cells with NK cell-like function in chronic hepatitis C virus infection. *J Immunol* 2008; **181**: 4219–4228.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)