



HAL
open science

Assessment of complement C4 gene copy number using the paralog ratio test

Michelle M A Fernando, Lora Boteva, David L Morris, Bi Zhou, Yee Ling Wu, Marja-Liisa Lokki, C. Yung Yu, John D Rioux, Ed Hollox, Timothy Vyse

► To cite this version:

Michelle M A Fernando, Lora Boteva, David L Morris, Bi Zhou, Yee Ling Wu, et al.. Assessment of complement C4 gene copy number using the paralog ratio test. *Human Mutation*, 2010, 31 (7), pp.866. 10.1002/humu.21259 . hal-00552384

HAL Id: hal-00552384

<https://hal.science/hal-00552384v1>

Submitted on 6 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Assessment of complement *C4* gene copy number using the paralog ratio test

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0048.R1
Wiley - Manuscript type:	Methods
Date Submitted by the Author:	25-Mar-2010
Complete List of Authors:	<p>Fernando, Michelle; Section of Rheumatology, Imperial College London, Faculty of Medicine</p> <p>Boteva, Lora; Section of Rheumatology, Imperial College London, Faculty of Medicine</p> <p>Morris, David; Section of Rheumatology, Imperial College London, Faculty of Medicine</p> <p>Zhou, Bi; Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital</p> <p>Wu, Yee Ling; Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital</p> <p>Lokki, Marja-Liisa; Transplantation Laboratory, Haartman Institute, University of Helsinki</p> <p>Yu, C. Yung; Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital</p> <p>Rioux, John; Montréal Heart Institute; Université de Montréal</p> <p>Hollox, Ed; University of Leicester, Genetics</p> <p>Vyse, Timothy; Section of Rheumatology, Imperial College London, Faculty of Medicine</p>
Key Words:	lupus, complement C4, copy number variation, paralog ratio test



1
2
3 **Assessment of complement *C4* gene copy number using the paralog ratio test**
4
5
6
7

8 Michelle M A Fernando¹, Lora Boteva¹, David L Morris¹, Bi Zhou², Yee Ling Wu², Marja-Liisa
9 Lokki³, Chack Yung Yu², John D Rioux^{4,5}, Edward J Hollox⁶, Timothy J Vyse^{1*}
10
11
12
13
14

15 ¹Section of Rheumatology, Faculty of Medicine, Imperial College London, London, United
16 Kingdom
17
18

19 ²Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's
20 Hospital, Columbus, Ohio, USA
21
22
23

24 ³Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland
25
26

27 ⁴Montréal Heart Institute, Montréal, Quebec, Canada
28

29 ⁵Université de Montréal, Montréal, Quebec, Canada
30

31 ⁶Department of Genetics, University of Leicester, Leicester, United Kingdom
32
33
34
35

36 *Corresponding author
37

38 Corresponding author details:
39

40 Professor Timothy J Vyse
41
42

43 Address: Section of Rheumatology, Faculty of Medicine, Imperial College London, London,
44 W12 0NN, United Kingdom
45
46

47 Email: t.vyse@imperial.ac.uk
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

The complement *C4* locus is in the class III region of the MHC and exhibits copy number variation. Complement *C4* null alleles have shown association with a number of diseases including systemic lupus erythematosus (SLE). However, most studies to date have used protein immunophenotyping and not direct interrogation of the genome to determine *C4* null allele status. Moreover, a lack of accurate *C4* gene copy number (GCN) estimation and tight linkage disequilibrium across the disease-associated MHC haplotypes has confounded attempts to establish whether or not these associations are causal. We have therefore developed a high through-put paralog ratio test (PRT) in association with two restriction enzyme digest variant ratio tests (REDVRs) to determine total *C4* GCN, *C4A* GCN and *C4B* GCN. In the densely genotyped CEU cohort we show that this method is accurate and reproducible when compared to gold standard Southern blot copy number estimation with a discrepancy rate of 9%. We find a broad range of *C4* GCNs in the CEU and the 1958 British Birth Cohort populations under study. In addition, SNP-*C4* CNV analyses show only moderate levels of correlation and therefore do not support the use of SNP genotypes as proxies for complement *C4* GCN.

Keywords

Complement C4

Copy number variation

Lupus

Paralog ratio test

Introduction

The human complement *C4* locus is in the class III region of the major histocompatibility complex (MHC) on the short arm of chromosome 6 and exhibits genetic complexity.

Complement *C4* genes show segmental duplication as part of mono-, bi-, tri-, or quadrimodular RCCX cassettes (**Figure 1**). Hence, in theory, two to eight copies of *C4* genes may be present in a diploid human genome; with each chromosome 6 comprising 1 to 4 copies of a single *C4* gene.

The *C4* gene exists as either of two forms: *C4A* (acidic) (MIM: 120810) or *C4B* (basic) (MIM: 120820), each of which is polymorphic in itself. At the nucleotide level *C4A* and *C4B* share 99% sequence homology over 41 exons. Each isotype is defined by five nucleotide changes in exon 26, which contribute to four isotype-specific amino acid residues from 1120 to 1125: **PCPVLD** for *C4A* and **LSPVIH** for *C4B* (Yu, 1991; Yu, et al., 1986). The *C4A* and *C4B* proteins differ in chemical reactivity. *C4A* preferentially binds to amino groups, forming amide bonds with proteins such as immune complexes. *C4B* demonstrates greater haemolytic activity in certain immunoassays compared to *C4A* and has a higher affinity for hydroxyl groups (Isenman and Young, 1984; Law, et al., 1984) Thus, *C4A* has a longer half-life against hydrolysis compared to *C4B* (Dodds, et al., 1996).

C4 genes may also vary in size, occurring as long (*C4L*) or short (*C4S*) forms. The long (21 kb) or short (14.6 kb) forms of the *C4* gene are determined by the presence or absence of a 6.4 kb insertion of human endogenous retrovirus, *HERV-K(C4)*, into intron 9 (Dangel, et al., 1994; Yu, et al., 1986). In a given population of European ancestry, zero to seven copies of the *C4* gene may be present in a diploid genome. *C4A* genes vary in copy number from zero to five, and *C4B*

1
2
3 genes from zero to four The most common copy number counts for *C4* in a European-derived
4
5
6 diploid genome are four copies of total *C4*, two copies of *C4A* and two copies of *C4B*
7
8 (Blanchong, et al., 2000; Yang, et al., 2007; Yang, et al., 2003).
9

10
11
12 The RCCX module (described by (Shen, et al., 1994) and (Yang, et al., 1999)) comprises four
13
14 genes encoded in tandem: the serine/threonine kinase gene, *RP1* (or *STK19*, MIM: 604977),
15
16 complement *C4* (*C4A* or *C4B*), cytochrome P450 steroid 21-hydroxylase, *CYP21A2* (MIM:
17
18 201910), and the extracellular matrix protein, *TNXB* (MIM: 600985). The breakpoints for each
19
20 duplicated module are identical (Shen, et al., 1994; Yang, et al., 1999). The *C4* gene in each
21
22 RCCX module is usually functional and codes for a C4A or C4B protein. In contrast, the *RP* and
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
TNX genes of duplicated RCCX modules are typically non-functional pseudogenes as a
consequence of partial sequences, known as *RP2* (or *STK19P*) and *TNXA* respectively.
Additional *CYP21* genes may be functional (*CYP21A2*) or non-functional (*CYP21A1P*) (Saxena,
et al., 2009; Yang, et al., 2007).

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
The presence of *C4* null alleles, *C4A*Q0* and *C4B*Q0*, was inferred from the absence of C4A or
C4B proteins from serum or plasma respectively. Partial deficiency of C4A or C4B, was used to
describe the phenomenon by which one isotype was expressed at a level about half of the other.
To date, it is known that the absence and the unequal serum/plasma protein levels of C4A and
C4B can be caused by the physical absence or non-sense mutations of the corresponding gene,
the unequal number of *C4A* and *C4B* genes in a diploid genome, and the differential protein
expression levels by the long and short *C4* genes. Notably, the *HLA-B*08-DRB1*0301*
haplotype, also known as the ancestral haplotype, AH8.1, is known to contain a single short *C4*
gene encoding for a C4B protein but no *C4A* gene (Awdeh, et al., 1983; Carroll, et al., 1985;

1
2
3 Chung, et al., 2002; Dawkins, et al., 1999). Non-sense mutations in *C4A* genes leading to the
4 absence of *C4A* protein include a 2 bp CT insertion into codon 1232 of exon 29, a G to A
5 substitution at the donor site of the intron 28 splice junction, a 1 bp deletion in exon 20, a 2 bp
6 deletion in exon 13 and a 1 bp deletion in exon 13 on a variety of haplotypic backgrounds
7 (Barba, et al., 1993; Wu, et al., 2009; Wu, et al., 2008). When *C4* null alleles are assessed by
8 typing *C4* gene copy numbers, the homozygous *C4B* null state is observed in 2-10%, the
9 homozygous *C4A* null state is seen in approximately 1% and heterozygous *C4A* or *C4B* null
10 alleles occur in approximately 45-56% of Europeans ((Seppanen, et al., 2006a; Yang, et al.,
11 2007); MF, LB, TJV unpublished). Complete homozygous deficiency of *C4* is extremely rare
12 and 28 cases have been reported to date (Pickering, et al., 2000; Wu, et al., 2009).
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 Clinically, the presence of *C4A* and *C4B* null alleles which result in “partial *C4* deficiency” have
30 shown association with the autoimmune disease, systemic lupus erythematosus (SLE)
31 (Christiansen, et al., 1983; Fielder, et al., 1983; Naves, et al., 1998; Pickering and Walport,
32 2000). In addition, an increased frequency of *C4* null alleles as determined by
33 immunophenotyping has been observed in a variety of other diseases including systemic
34 sclerosis, Henoch-Schönlein purpura, glomerulonephritis, myasthenia gravis, and chronic active
35 hepatitis (Briggs, et al., 1993; Christiansen, et al., 1991; Franciotta, et al., 2001; Vergani, et al.,
36 1985). Furthermore, *C4A* or *C4B* null alleles have shown association with reduced life
37 expectancy, myocardial infarction, Henoch-Schönlein purpura, autism and increased
38 susceptibility to microbial infection (Arason, et al., 2007; Kramer, et al., 1989; Odell, et al.,
39 2005; Seppanen, et al., 2006b; Stefansson Thors, et al., 2005). Accurate ascertainment of *C4A*
40 and *C4B* copy numbers were lacking in most studies where null allele states have generally been
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 inferred from the absence of the encoded protein. In addition, *C4* null alleles are in strong LD
4 with specific extended haplotypes, for example, *A*01-B*08-(C4A*Q0)-C4B1-DRB1*0301* and
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
*A*30-B*18-C4A3-(C4B*Q0)-DRB1*0301* (the “Basque” haplotype), so to date it has not been
possible to establish the identity of causal variation located within these extended haplotypes. In
the particular case of SLE, it has not yet been possible to distinguish association observed with
*HLA-DRB1*0301* and *C4A* null alleles due to tight LD on the disease-associated haplotype.

Current genetic association studies including genome-wide association scans utilize SNP-based
genotyping technologies given the abundance of polymorphic markers available. However, it is
known that SNP typing often fails to identify regions of CNV due to assay failure or exclusion
for deviation from both Hardy-Weinberg equilibrium and Mendelian inheritance. Moreover,
CNVs are not always in LD with SNPs and therefore not tagged by them (Hollox, et al., 2009;
Locke, et al., 2006). Hence, alternative methods must be used to interrogate the genome in order
to assess the impact of CNVs.

Southern blot techniques have been used to successfully determine copy number at the
complement *C4* locus (Yang, et al., 2007). However, these experiments require large quantities
of DNA (typically 10-15 micrograms per sample) and are time-consuming (three to ten days per
batch). Moderate to high-throughput strategies to determine complement *C4* gene copy number,
in conjunction with data from high-density genome-wide association scans that are now feasible,
would therefore prove invaluable in unraveling the complex genetic associations observed at the
MHC.

1
2
3 Southern blot techniques may currently be seen as a gold standard for the determination of copy
4 number variation in the genome given that genomic DNA is the substrate for experimentation,
5 while all other strategies use PCR amplification from genomic DNA. Unequal amplification of
6 test and reference sequences may therefore produce errors in copy number determination.
7
8 However, Southern blot techniques are not amenable to moderate let alone high-throughput copy
9 number estimation given the reasons outlined above. Hence there is a need for the development
10 of methods of assessing copy number that maintain the accuracy of Southern blots without the
11 attendant pitfalls. We therefore sought to develop such an assay using principles based on the
12 paralog ratio test (PRT) and restriction enzyme digest variant ratio determination (REDVR)
13 (Aldred, et al., 2005; Armour, et al., 2007). The utility of the PRT has been established given the
14 success of this method in determining copy number and disease association at the *FCGR3* locus
15 and the beta-defensin locus (Hollox, et al., 2009; Hollox, et al., 2008).
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Materials and Methods

Study cohorts

CEU

Genomic DNA from 89 CEU HapMap samples was obtained from Coriell Cell Repositories and used for the PRT, Southern blot and qPCR experiments for complement *C4* copy number determination.

1958 British Birth Cohort

In order to validate our results in the CEU cohort, we used the PRT and REDVRs to estimate *C4* GCN from the genomic DNA of 163 unrelated subjects from the 1958 British Birth Cohort. All 163 subjects had been previously genotyped to high-density at the MHC using a custom Illumina panel (Rioux, et al., 2009). 2-digit *HLA-DRB1* (MIM: 142857) genotypes were available for 132/163 (81%) of the subjects. The SNP and *HLA-DRB1* genotype data were used for SNP-CNV correlation analyses.

Complement C4 PRT

The principles of the paralog ratio test have been described elsewhere (Armour, et al., 2007). Briefly, identical primer pairs co-amplify a copy-variable “test” region and a non-copy-variable “reference” region using PCR. The resulting amplicons will differ in size and can be distinguished by capillary electrophoresis. Experiments are performed in duplicate by using two different fluorescent dyes to label the same primer of the pair. Raw copy number estimates can be determined by calculating the ratio of the area under the peak between test and reference amplicons. Data from a number of different plates can be pooled as the raw ratios from each

1
2
3 plate are normalized using control samples of known complement *C4* copy number, thus
4
5 correcting for inter-experiment variation. The normalized raw PRT and REDVR data are then
6
7 put forward for cluster analysis.
8
9

10
11
12 Specifically for the complement *C4* PRT, primer pairs were obtained by mining the complement
13
14 *C4* locus for repeat sequences using the Self-Chain track in the University of California, Santa
15
16 Cruz (UCSC) genome browser. Primer pairs were chosen such that they resulted in the co-
17
18 amplification of the test sequence and only one other sequence in the genome which was located
19
20 within a non-copy-variable region. We did not find identical primer pairs capable of co-
21
22 amplifying the copy variable complement *C4* locus and a reference locus. Instead we used
23
24 primer pairs that differed by one base-pair in the reverse primer for the test (chromosome 6:
25
26 CAGGGAAGGCTTCCTG) and reference (chromosome 19: CAGGGAGGGCTTCCTG) loci.
27
28 The labeled forward primer remained invariant (FAM- or HEX-CCTCTGGGCCTTTGTA). The
29
30 PCR products for the test and reference loci differed by only one base-pair (125 bp chromosome
31
32 6 test sequence and 124 bp chromosome 19 reference sequence) and hence required restriction
33
34 enzyme digestion prior to separation by capillary electrophoresis in order to distinguish test and
35
36 reference amplicons. The reference sequence but not the test sequence contained an AluI
37
38 restriction site yielding a 78 bp amplicon for the reference locus and an uncut 125 bp amplicon
39
40 for the test locus (**Figure 2 and Supp Figure S1**).
41
42
43
44
45
46
47
48
49

50
51 Experiments were performed in a 96-well format using 1 μ l of sample genomic DNA per well at
52
53 a concentration of approximately 10 ng/ μ l. Each 96-well plate contained 8 control samples for
54
55 standardization and normalization including samples from three cell lines sequenced at the
56
57
58
59
60

1
2
3 MHC: COX (total C4 CN of 2, C4A CN of 0, C4B CN of 2), QBL (total C4 CN of 2, C4A CN
4 of 2, C4B CN of 0) and PGF (total C4 CN of 4, C4A CN of 2, C4B CN of 2) (Traherne, et al.,
5 2006). The PCR mix totaled 10 μ l and comprised template DNA with 0.6 μ M of FAM- or HEX-
6 labeled forward primer, 0.6 μ M of chromosome 6 reverse primer, 0.1 μ M of chromosome 19
7 reverse primer, in a buffer with final concentrations of 50 mM Tris-HCl, 12.5 mM ammonium
8 sulphate, 1.4 mM magnesium chloride, 125 μ g/ml bovine serum albumin (BSA), 7.5 mM 2-
9 mercaptoethanol, 200 mM each dNTP and 0.5 units *Taq* DNA polymerase. The PCR
10 amplification conditions were: initial denaturation step at 95⁰C for 2 minutes, then 30 cycles of
11 denaturation at 95⁰C for 30 seconds, annealing at 60⁰C for 30 seconds and extension at 70⁰C
12 for 30 seconds, then a single chase phase of 56⁰C for 1 minute and finally 70⁰C for 20 minutes
13 to reduce levels of single-stranded DNA. Next, 2 μ l of the PCR product was digested with 5
14 units of AluI (New England Biolabs/NEB), 1 μ l 10x NEB Buffer 2 and sterile water to a total
15 volume of 10 μ l at 37⁰C for 4 hours, then 65⁰C for 1 hour (AluI inactivation) and then incubated
16 for 16 hours at 10⁰C. 2 μ l of each of the separately digested FAM and HEX reactions was then
17 added to 10 μ l of deionised formamide and analyzed by electrophoresis on a capillary sequencer
18 (ABI 3730xl DNA Analyzer, Applied Biosystems). The ratio of the peak areas for the 125 bp
19 chromosome 6 test product and that of the 78 bp chromosome 19 reference product (125/78) was
20 calculated for each FAM- and HEX-labeled reaction using GeneMapper software (Applied
21 Biosystems). The result was accepted if the coefficient of variation (standard deviation divided
22 by the mean) between FAM and HEX reactions was <0.15. All samples passed these criteria.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

*REDVRs**REDVR A: C4A and C4B copy number determination*

REDVR A was able to distinguish *C4A* and *C4B* by virtue of a *NlaIV* restriction site within the *C4A* gene created by a C (*C4A*)/T (*C4B*) nucleotide substitution that code for one of the *C4A/C4B* isotypic residues (Yu and Campbell, 1987). Thus, *C4B* does not harbour the restriction site and produces a 158 bp fragment, while *C4A* is cut by *NlaIV* and yields a 91 bp fragment. Primers spanning this site were designed in Primer3 (<http://frodo.wi.mit.edu/>, (Rozen and Skaletsky, 2000)). The PCR cycling conditions are as described for the PRT except that 0.5 μM of each forward (CTGAGAAACTGCAGGAGACATC) and labeled reverse primer (FAM-GAAGGGGCAAAGAGAGTCCT) was used, the annealing temperature was 62⁰C and 28 cycles were used for amplification. Next, 2 μl of the PCR product was digested with 5 units *NlaIV*, 1 μl of 10x NEB Buffer 4, and 0.1 μl of 100x BSA to a total of 10 μl at 37⁰C for four hours and then 10⁰C for 16 hours. 2 μl of the digestion product was mixed with 10 μl deionised formamide and analyzed by electrophoresis on a capillary sequencer.

REDVR B: Copy number determination of paralogous sequence variant, rs17855807

REDVR B was able to distinguish a paralogous sequence variant (PSV) in the form of a non-synonymous SNP (*rs17855807* A/G, N1176S) within the complement *C4* gene. The variant creates an *AluI* restriction site when the G allele (serine) is present. Primers spanning this site were designed in Primer3. The PCR cycling conditions are as described for REDVR A with 0.5 μM of each labeled forward (HEX-CCCGGCTCTCTCCCTTTTTC) and reverse primer (TTGGTCAGTGTCAGGGCATA); 26 cycles were used for amplification. Next, 1 μl of the PCR product was digested with 5 units *AluI*, 1 μl of 10x NEB Buffer 2, to a total of 10 μl at

1
2
3 37⁰C for four hours, then 65⁰C and then 10⁰C for 16 hours. 2 µl of the digestion product was
4
5 mixed with 10 µl deionised formamide and analyzed by electrophoresis on a capillary sequencer.
6

7
8 The cut: uncut ratios for both REDVRs were calculated as described for the PRT.
9

10 11 12 *Copy number estimation*

13
14
15 We used a clustering algorithm for determining integer copy number estimates from our
16
17 empirical data (**Figure 3**). We model the copy number (PRT) and REDVR A data as a mixture of
18
19 multivariate normally distributed variables, assuming that the PRT and REDVR A variables are
20
21 independent. We model the REDVR B data as a mixture of mixtures, as each possible copy
22
23 number can result in various REDVR B distributions (3 copies, for example, would have an
24
25 expected REDVR B ratio of 0, 0.5 or 2 with probabilities of 0.25, 0.375 and 0.375 respectively
26
27 assuming that the frequencies of the two REDVR B alleles are equal (0.5)). With respect to the
28
29 total copy number (PRT), REDVR A estimates (non-null) and REDVR B, each cluster is
30
31 assumed to be multivariate normal with unknown means and each observation belongs to one of
32
33 eight clusters. For observations with REDVR A-null data we only use PRT and REDVR B, as
34
35 REDVR A is always 0, so the null data can only belong to one of three clusters (CN/PRT=2, 3,
36
37 4). For REDVR B data each cluster is assumed to be a mixture over a point mass at zero and
38
39 normally distributed variables with means and mixture probabilities as described in **Supp Table**
40
41 **S1**. We define prior distributions on the means of each normal distribution and specify the
42
43 coefficient of variation. The priors require a mean and standard deviation for our belief about the
44
45 cluster mean which we declare in **Supp Table S1**. These values are based on known copy
46
47 number estimates obtained from the Southern blot CEU data paired with PRT and REDVR raw
48
49 ratios. Each observation is assumed to belong to one of the clusters defined by a cluster mean (or
50
51
52
53
54
55
56
57
58
59
60

1
2
3 means for the REDVR) and a cluster variance which we define with respect to a coefficient of
4 variation (the variance is proportional to the mean). The coefficient of variation was estimated by
5 samples of paired readings for each variable. Cluster membership is unknown and each cluster
6 has prior probabilities as defined in **Supp Table S1**. This method gives us the posterior
7 probability of every cluster and some measure of our uncertainty. Inference can also be
8 performed without reference to posterior probabilities by using Bayes Factors (BFs). In our data
9 this would have not made any difference as all BFs were in favor of the cluster chosen for each
10 observation and all BFs were >10 , 83 samples had $BF > 100$ and 62 had $BF > 10^{10}$. This shows
11 the robustness of the approach to the choice of prior probabilities on each cluster.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 *Southern blot for complement C4 gene copy number determination*

28 The Southern blots were performed as previously described (Chung, et al., 2005). Total C4 gene
29 copy numbers and RCCX modules were determined by TaqI RFLP, and C4A to C4B gene copy
30 ratios were determined by PshAI-PvuII RFLP for the 89 CEU samples.
31
32
33
34
35
36
37
38

39 *Quantitative-PCR for complement C4 gene copy number determination*

40 C4A and C4B copy numbers were performed by isotype-specific genomic real-time PCR. We
41 used unlabeled primers with SYBR Green QPCR (Stratagene, Cedar Creek, Texas, USA) or
42 Absolute QPCR SYBR GREEN MIX (Abgene, Epsom, UK) according to the manufacturers'
43 instructions with minor modifications (Vauhkonen H et al. submitted). C4A, C4B and beta-actin
44 primers were based on published primer sequences (Barba GM et al. 1994, Montgomery RA and
45 Dietz HC, 1997).
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

HLA-DRB1 typing in CEU

HLA-DRB1 genotype data was unavailable for 24 of the 89 CEU samples under study. These samples were therefore genotyped for *HLA-DRB1* to 4-digit resolution at the Anthony Nolan Trust, London, UK using a bead-based sequence-specific oligonucleotide probe (SSOP) protocol (Luminex). These results (**Supp Table S2**) were amalgamated with previous 4-digit *HLA-DRB1* typing (<http://www.inflamngen.org/>) and put forward for further analysis.

SNP-CNV correlation

In order to determine the relationship between the complement *C4* locus and surrounding SNPs we calculated the correlation coefficient, r^2 , between SNP genotypes and integer copy number estimates for *C4A* and *C4B* using standard linear regression. The SNP genotypes were coded 0, 1 or 2 (0 = homozygous for minor allele, 1 = heterozygous, 2 = homozygous for major allele). We also integrated *HLA-DRB1* allele data for both cohorts in the analysis. We converted the *HLA-DRB1* data into SNP genotypes for this analysis. We used 2-digit *HLA-DRB1* data for 1958 British Birth Cohort subjects as 4-digit data was not available for the majority of the samples. We used 4-digit *HLA-DRB1* data for the CEU cohort as these data were available for all subjects. We used independent data from founders only in these analyses. A p-value was calculated for the regression from which r^2 arises in order to determine the statistical significance of the correlation coefficient between a SNP and *C4A* or *C4B* copy number. We verified our results using more standard estimates of linkage disequilibrium, namely D-prime and r^2 , using 2-copy individuals only. To calculate the correlation between 0 copy individuals (that is individuals lacking either *C4A* or *C4B* genes) for *C4A* or *C4B* and surrounding SNPs, we analysed unrelated individuals with either 0, 1 or 2 copies of the respective genes. If we regard 0 copy as the 'minor allele', we

1
2
3 code the genotypes as [1,1], [1,2], or [2,2] respectively. This assumes that all two-copy
4 individuals have one copy on each chromosome. D-prime and r^2 values for these data were
5
6 calculated using Haploview. The virtually identical correlation coefficients between SNPs and
7
8 *C4* GCN using standard linear regression and pair-wise LD measures in Haploview indicate that
9
10 this assumption is valid. In the CEU cohort, we obtained genotype data for SNPs 1 Mb telomeric
11
12 and 1 Mb centromeric of the complement *C4* locus from HapMap for this analysis
13
14 (<http://www.hapmap.org/>). We used SNP data from the IMAGEN consortium study for the 1958
15
16 British Birth Cohort subjects across the same region as well as the entire SNP data from 26.1 Mb
17
18 to 33.5 Mb (Rioux, et al., 2009). We plotted the correlation coefficient, r^2 , against the
19
20 chromosomal position of each SNP for *C4A* and *C4B* copy number estimates (**Figure 4 and**
21
22 **Supp Tables S3-S6**). The size of the square representing each SNP is inversely proportional to
23
24 the p-value.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results

Assessment of complement C4 copy number in CEU samples demonstrates a broad range of copy number variation

We determined total complement *C4* gene copy number (GCN), *C4A* GCN and *C4B* GCN in 89 HapMap CEU samples using our novel PRT assay in association with two REDVRs (see Methods), Southern blot analysis (Chung, et al., 2005) and a quantitative PCR (qPCR) technique (Vauhkonen H et al. submitted), in order to assess *C4* gene copy number variability in this extensively genotyped cohort. Each method was performed in a blinded manner such that details of family structure and the results of the other strategies were not known to the experimenter at the time of typing. Results were then collated and analyzed (**Supp Table S7 (a) and (b)**). In this cohort, using Southern Blot data as the gold standard values, total complement *C4* gene copy number ranges from 2 to 5, with *C4A* varying between 0 and 4, and *C4B* between 0 and 3. We did not detect any Mendelian inconsistencies within the known trios.

We observe 100% concordance between PRT and Southern blot assays for samples that are homozygous for zero copies of either *C4A* or *C4B*. As expected, the “error” rate for copy number estimation using the PRT increases with copy number (if we use Southern blot copy number estimation as the gold standard) (**Supp Table S8**). **It should be noted that as Southern blot copy number determination is used as the gold standard, we therefore assume no errors with this technique and attribute all discrepancies to our method.** The overall **discrepancy** rate for the PRT compared to the Southern blot is 9% (8/89) and is comparable with other PRT assays (Armour, et al., 2007; Hollox, et al., 2009). Moreover, this assay is extremely accurate at low copy number

1
2
3 estimation which is important given that disease associations are seen with complement C4
4 deficiency states. The CEU cohort was also typed using a qPCR method for assessing total C4,
5
6 C4A and C4B copy number (Vauhkonen H et al. submitted). The discrepancy rate for this assay
7
8 was 13%.
9
10
11

12
13
14
15 *Correlation between surrounding SNPs and complement C4 copy number in CEU and a subset*
16
17 *of the 1958 British Birth Cohort*
18

19
20 We wanted to investigate the correlation between surrounding SNPs and complement C4 copy
21 number in the CEU population under study. Specifically, we wanted to determine whether we
22 could use SNP genotypes as surrogate markers for the absence (zero copies) of complement C4A
23 and C4B genes that we have assayed, thereby obviating the need to specifically evaluate C4 gene
24 copy number in future studies of this region. The CEU cohort has been extensively genotyped as
25 part of the ongoing HapMap project (The International HapMap Consortium, 2003) and hence is
26 an ideal population with which to assess such correlations. Standard linear regression in CEU
27 founders (that is, unrelated parents and not children) revealed evidence of correlation between
28 surrounding SNPs and integer copy number values for C4A and C4B (Figure 4, panels (a) and
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(b) and Supp Tables S3 and S4). Levels of correlation were greater for C4A copy number estimates compared to C4B. The six SNPs demonstrating the highest levels of correlation for C4A two-copy individuals, $r^2 = 0.62$, are all in strong LD and are located in the class III gene, *TNXB*. Of interest, the *HLA-DRB1*0301* allele showed only modest correlation in this C4A dataset, with an r^2 of 0.32 (r^2 of 0.34 between *DRB1*0301* and total C4 integer copy number). No other *HLA-DRB1* alleles showed significant correlation with total C4, C4A or C4B copy number values. For C4B two-copy individuals, there are two SNPs, *rs9332730* and *rs7774197*,

1
2
3 which show the greatest correlation with an r^2 of 0.45. These SNPs are located in the introns of
4 the genes *C2* and *TNXB* respectively and are correlated with each other (r^2 0.79). There were 165
5 SNPs showing correlation coefficients greater than 0.3 with *C4A*, while this figure reduced to 6
6 SNPs for *C4B*. These levels of SNP-CNV correlation do not support the use of SNP genotypes as
7 proxies or tagSNPs for complement *C4* gene copy number. In particular, there are no surrogate
8 SNPs for the absence of *C4A* or *C4B* genes in this cohort. Hence SNP studies will not tag *C4*
9 CNV and such variation will be missed by GWAS.
10
11
12
13
14
15
16
17
18
19
20
21

22 In order to corroborate our results, we genotyped 163 individuals of northern European ancestry
23 from the 1958 British Birth Cohort, for *C4* GCN using the PRT/REDVR method. High-density
24 SNP data at the MHC was available for all 163 subjects from a previous study (Rioux, et al.,
25 2009). Two-digit *HLA-DRB1* genotypes were available in 81% (132/163) of the cohort. In this
26 dataset, we show that total *C4* GCN varies between 2 and 5, while *C4A* GCN ranges from 0-4
27 and *C4B* GCN ranges from 0-4. SNP-CNV correlation analyses corroborate the CEU results
28 showing higher levels of SNP correlation with *C4A* compared with *C4B* (Figure 4, panels (c)
29 and (d) and Supp Tables S5 and S6). Furthermore, taking into account the SNPs that are
30 common to both datasets, the SNP showing the greatest correlation with *C4A*, (*rs1269852*, $r^2 =$
31 0.74, $p < 1.0 \times 10^{-16}$) is equivalent to the top marker in CEU. The greater correlation values for
32 *C4A* GCN likely represents the larger size of the 1958 cohort, while the lower values for *C4B*
33 GCN correlations reflect greater haplotypic variability in this larger cohort.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Discussion

Complement C4 is an important component of the classical and mannose-binding lectin complement cascades. Activation of these pathways stimulates innate and adaptive immune responses to microbes. Through opsonisation, C4 also participates in the clearance of circulating immune complexes and apoptotic debris. Further, C4 has been shown to reduce the activation threshold for B cells, enhancing T cell dependent antibody production (Fischer, et al., 1996). Inherited and acquired deficiencies of complement C4 are associated with the development of SLE.

Current estimates suggest that approximately half of all CNVs may be tagged by SNPs and that common CNVs (>5%) can be effectively tagged by SNPs of similar frequency (McCarroll, 2008; Redon, et al., 2006). This means that alternative methods must be used to interrogate the genome to assess the remaining CNVs. Future genetic studies will therefore need to incorporate strategies that allow assessment SNP variation as well as CNV. Initial data from the International Haplotype Project has shown that two randomly chosen genomes differed by 0.1% with regard to SNP diversity (<http://www.hapmap.org/>). The impact of CNV on genomic variation has revised this figure upwards, such that the majority of variation is now thought to derive from CNV (Levy, et al., 2007; Wheeler, et al., 2008). Indeed a recent CGH study calculated a cumulative CNV locus length of 24 Mb (0.78% of the genome) when comparing two genomes (Conrad, et al., 2009).

1
2
3 The greater level of SNP correlation with HLA haplotypes lacking a *C4A* gene compared to
4 those without a *C4B* gene reflects the greater number of haplotypes on which a *C4B* gene is
5 absent compared to *C4A*. The mechanisms responsible for this difference are unknown and may
6 reflect differential selective pressures between *C4A* and *C4B* or the SNP haplotypes they were
7 formed on, a higher mutation rate for haplotypes without *C4B* genes or population processes
8 such as genetic drift altering the frequencies of haplotypes in whom *C4* genes are absent.
9 Previous studies, principally performed on homozygous cell lines, have shown that absent *C4B*
10 genes occur on many different haplotypic backgrounds, while absent *C4A* genes seem to occur
11 on a more limited set of haplotypes. These data suggest that both *C4A* and *C4B* gene deletions
12 are recurrent having occurred on different haplotypic backgrounds.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 We have shown that our C4 PRT is accurate, reliable and reproducible particularly at low *C4*
30 copy number estimation. The assay is high through-put and requires 40-60 ng of genomic DNA
31 when performed in duplicate. We have also developed a clustering algorithm that enables
32 quantification of the accuracy of a call for each processed sample. In addition, individuals in
33 whom *C4A* and *C4B* genes are absent can be rapidly and accurately assessed by performing
34 REDVR A alone. We plan to use these assays to evaluate complement *C4* GCN in the context of
35 high-density SNP data at the MHC in SLE cohorts of differing ancestry. Consequently we hope
36 to determine whether complement *C4* gene disease associations are independent or in fact
37 secondary to LD with causal variants.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Author contributions

TJV, MF and JDR conceived the study design. EJH, MF and LB developed the *C4* PRT and REDVR assays. LB performed the PRT and REDVR assays. MF, LB and DLM analyzed the data. BZ and YLW performed the Southern blot experiments. YLW and CYY analyzed the Southern blot data. CYY provided CEU DNA. M-L L analyzed the qPCR data. MF wrote the paper with contributions from LB, DLM, M-L L, CYY, EJH and TJV.

Funding

MF and LB are supported by a Clinician Scientist Fellowship from Arthritis Research UK (Grant 18239). DLM is supported by an Arthritis Research UK Project Grant (Grant 17761). The IMAGEN consortium is supported by Grant AI067152 from the National Institutes of Allergy and Infectious Diseases. We acknowledge the use of DNA from the 1958 British Birth Cohort (D. Strachan, S. Ring, W. McArdle and M. Pembrey) funded by the Medical Research Council Grant G0000934 and Wellcome Trust Grant 068545/Z/02. CYY is funded by NIH grant AR054459.

Conflict of Interest

The authors declare no conflict of interest.

References

- Aldred PM, Hollox EJ, Armour JA. 2005. Copy number polymorphism and expression level variation of the human alpha-defensin genes DEFA1 and DEFA3. *Hum Mol Genet* 14(14):2045-52.
- Arason GJ, Kramer J, Blasko B, Kolka R, Thorbjornsdottir P, Einarsdottir K, Sigfusdottir A, Sigurdarson ST, Sigurdsson G, Ronai Z and others. 2007. Smoking and a complement gene polymorphism interact in promoting cardiovascular disease morbidity and mortality. *Clin Exp Immunol* 149(1):132-8.
- Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ. 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res* 35(3):e19.
- Awdeh ZL, Raum D, Yunis EJ, Alper CA. 1983. Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc Natl Acad Sci U S A* 80(1):259-63.
- Barba G, Rittner C, Schneider PM. 1993. Genetic basis of human complement C4A deficiency. Detection of a point mutation leading to nonexpression. *J Clin Invest* 91(4):1681-6.
- Blanchong CA, Zhou B, Rupert KL, Chung EK, Jones KN, Sotos JF, Zipf WB, Rennebohm RM, Yung Yu C. 2000. 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* 191(12):2183-96.
- Briggs D, Stephens C, Vaughan R, Welsh K, Black C. 1993. A molecular and serologic analysis of the major histocompatibility complex and complement component C4 in systemic sclerosis. *Arthritis Rheum* 36(7):943-54.
- Carroll MC, Palsdottir A, Belt KT, Porter RR. 1985. Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. *Embo J* 4(10):2547-52.
- Christiansen FT, Dawkins RL, Uko G, McCluskey J, Kay PH, Zilko PJ. 1983. Complement allotyping in SLE: association with C4A null. *Aust N Z J Med* 13(5):483-8.
- Christiansen FT, Zhang WJ, Griffiths M, Mallal SA, Dawkins RL. 1991. Major histocompatibility complex (MHC) complement deficiency, ancestral haplotypes and systemic lupus erythematosus (SLE): C4 deficiency explains some but not all of the influence of the MHC. *J Rheumatol* 18(9):1350-8.
- Chung EK, Wu YL, Yang Y, Zhou B, Yu CY. 2005. Human complement components C4A and C4B: complex genotypes and phenotypes. In: Coligan JE, Bierer BE, Margulis DH, Shevach EM, Strober W, editors. *Current Protocols in Immunology*. Edison, New Jersey: John Wiley & Sons, Inc. p 13.8.1-13.8.36.
- Chung EK, Yang Y, Rennebohm RM, Lokki ML, Higgins GC, Jones KN, Zhou B, Blanchong CA, Yu CY. 2002. Genetic sophistication of human complement components C4A and C4B and RP-C4-CYP21-TNX (RCCX) modules in the major histocompatibility complex. *Am J Hum Genet* 71(4):823-37.
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P and others. 2009. Origins and functional impact of copy number variation in the human genome. *Nature*.
- Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu LC, Yu CY. 1994. The dichotomous size variation of human complement C4 genes is mediated by a novel

- 1
2
3 family of endogenous retroviruses, which also establishes species-specific genomic
4 patterns among Old World primates. *Immunogenetics* 40(6):425-36.
- 5 Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J. 1999.
6 Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses
7 and disease. *Immunol Rev* 167:275-304.
- 8
9 Dodds AW, Ren XD, Willis AC, Law SK. 1996. The reaction mechanism of the internal
10 thioester in the human complement component C4. *Nature* 379(6561):177-9.
- 11 Fielder AH, Walport MJ, Batchelor JR, Rynes RI, Black CM, Dodi IA, Hughes GR. 1983.
12 Family study of the major histocompatibility complex in patients with systemic lupus
13 erythematosus: importance of null alleles of C4A and C4B in determining disease
14 susceptibility. *Br Med J (Clin Res Ed)* 286(6363):425-428.
- 15
16 Fischer MB, Ma M, Goerg S, Zhou X, Xia J, Finco O, Han S, Kelsoe G, Howard RG, Rothstein
17 TL and others. 1996. Regulation of the B cell response to T-dependent antigens by
18 classical pathway complement. *J Immunol* 157(2):549-56.
- 19
20 Franciotta D, Cuccia M, Dondi E, Piccolo G, Cosi V. 2001. Polymorphic markers in MHC class
21 II/III region: a study on Italian patients with myasthenia gravis. *J Neurol Sci* 190(1-2):11-
22 6.
- 23
24 Hollox EJ, Detering JC, Dehnugara T. 2009. An integrated approach for measuring copy number
25 variation at the FCGR3 (CD16) locus. *Hum Mutat* 30(3):477-84.
- 26
27 Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC,
28 Traupe H, de Jongh G, den Heijer M and others. 2008. Psoriasis is associated with
29 increased beta-defensin genomic copy number. *Nat Genet* 40(1):23-5.
- 30
31 Isenman DE, Young JR. 1984. The molecular basis for the difference in immune hemolysis
32 activity of the Chido and Rodgers isotypes of human complement component C4. *J
33 Immunol* 132(6):3019-27.
- 34
35 Kramer J, Rajczyk K, Fust G. 1989. Low incidence of null alleles of the fourth component of
36 complement (C4) in elderly people. *Immunol Lett* 20(1):83-5.
- 37
38 Law SK, Dodds AW, Porter RR. 1984. A comparison of the properties of two classes, C4A and
39 C4B, of the human complement component C4. *Embo J* 3(8):1819-23.
- 40
41 Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF,
42 Denisov G and others. 2007. The diploid genome sequence of an individual human. *PLoS
43 Biol* 5(10):e254.
- 44
45 Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S,
46 Albertson DG, Pinkel D, Altshuler DM and others. 2006. Linkage disequilibrium and
47 heritability of copy-number polymorphisms within duplicated regions of the human
48 genome. *Am J Hum Genet* 79(2):275-90.
- 49
50 McCarroll SA. 2008. Extending genome-wide association studies to copy-number variation.
51 *Hum Mol Genet* 17(R2):R135-42.
- 52
53 Naves M, Hajeer AH, Teh LS, Davies EJ, Ordi-Ros J, Perez-Pemen P, Vilardel-Tarres M,
54 Thomson W, Worthington J, Ollier WE. 1998. Complement C4B null allele status
55 confers risk for systemic lupus erythematosus in a Spanish population. *Eur J
56 Immunogenet* 25(4):317-20.
- 57
58 Odell D, Maciulis A, Cutler A, Warren L, McMahon WM, Coon H, Stubbs G, Henley K, Torres
59 A. 2005. Confirmation of the association of the C4B null allele in autism. *Hum Immunol*
60 66(2):140-5.

- 1
2
3 Pickering MC, Botto M, Taylor PR, Lachmann PJ, Walport MJ. 2000. Systemic lupus
4 erythematosus, complement deficiency, and apoptosis. *Adv Immunol* 76:227-324.
- 5 Pickering MC, Walport MJ. 2000. Links between complement abnormalities and systemic lupus
6 erythematosus. *Rheumatology (Oxford)* 39(2):133-41.
- 7
8 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH,
9 Carson AR, Chen W and others. 2006. Global variation in copy number in the human
10 genome. *Nature* 444(7118):444-54.
- 11 Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T, De Jager PL, Foisy S,
12 Wang J, de Bakker PI and others. 2009. Mapping of multiple susceptibility variants
13 within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci U S A*.
- 14 Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist
15 programmers. *Methods Mol Biol* 132:365-86.
- 16
17 Saxena K, Kitzmiller KJ, Wu YL, Zhou B, Esack N, Hiremath L, Chung EK, Yang Y, Yu CY.
18 2009. Great genotypic and phenotypic diversities associated with copy-number variations
19 of complement C4 and RP-C4-CYP21-TNX (RCCX) modules: a comparison of Asian-
20 Indian and European American populations. *Mol Immunol* 46(7):1289-303.
- 21
22 Seppanen M, Meri S, Notkola IL, Seppala IJ, Hiltunen-Back E, Sarvas H, Lappalainen M,
23 Valimaa H, Palikhe A, Valtonen VV and others. 2006a. Subtly impaired humoral
24 immunity predisposes to frequently recurring genital herpes simplex virus type 2
25 infection and herpetic neuralgia. *J Infect Dis* 194(5):571-8.
- 26
27 Seppanen M, Suvilehto J, Lokki ML, Notkola IL, Jarvinen A, Jarva H, Seppala I, Tahkokallio O,
28 Malmberg H, Meri S and others. 2006b. Immunoglobulins and complement factor C4 in
29 adult rhinosinusitis. *Clin Exp Immunol* 145(2):219-27.
- 30
31 Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY.
32 1994. Structure and genetics of the partially duplicated gene RP located immediately
33 upstream of the complement C4A and the C4B genes in the HLA class III region.
34 Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene
35 duplication. *J Biol Chem* 269(11):8466-76.
- 36
37 Stefansson Thors V, Kolka R, Sigurdardottir SL, Edvardsson VO, Arason G, Haraldsson A.
38 2005. Increased frequency of C4B*Q0 alleles in patients with Henoch-Schonlein purpura.
39 *Scand J Immunol* 61(3):274-8.
- 40
41 The International HapMap Consortium. 2003. The International HapMap Project. *Nature*. p 789-
42 96.
- 43
44 Traherne JA, Horton R, Roberts AN, Miretti MM, Hurles ME, Stewart CA, Ashurst JL, Atrazhev
45 AM, Coghill P, Palmer S and others. 2006. Genetic analysis of completely sequenced
46 disease-associated MHC haplotypes identifies shuffling of segments in recent human
47 history. *PLoS Genet* 2(1):e9.
- 48
49 Vergani D, Wells L, Larcher VF, Nasaruddin BA, Davies ET, Mieli-Vergani G, Mowat AP.
50 1985. Genetically determined low C4: a predisposing factor to autoimmune chronic
51 active hepatitis. *Lancet* 2(8450):294-8.
- 52
53 Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ,
54 Makhijani V, Roth GT and others. 2008. The complete genome of an individual by
55 massively parallel DNA sequencing. *Nature* 452(7189):872-6.
- 56
57 Wu YL, Hauptmann G, Viguier M, Yu CY. 2009. Molecular basis of complete complement C4
58 deficiency in two North-African families with systemic lupus erythematosus. *Genes*
59 *Immun*.
- 60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Wu YL, Yang Y, Chung EK, Zhou B, Kitzmiller KJ, Savelli SL, Nagaraja HN, Birmingham DJ, Tsao BP, Rovin BH and others. 2008. 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* 123(1-4):131-41.
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K and others. 2007. 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* 80(6):1037-54.
- Yang Y, Chung EK, Zhou B, Blanchong CA, Yu CY, Fust G, Kovacs M, Vatay A, Szalai C, Karadi I and others. 2003. 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* 171(5):2734-45.
- Yang Z, Mendoza AR, Welch TR, Zipf WB, Yu CY. 1999. 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* 274(17):12147-56.
- Yu CY. 1991. The complete exon-intron structure of a human complement component C4A gene. DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene. *J Immunol* 146(3):1057-66.
- Yu CY, Belt KT, Giles CM, Campbell RD, Porter RR. 1986. Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. *EMBO J* 5(11):2873-81.
- Yu CY, Campbell RD. 1987. Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles. *Immunogenetics* 25(6):383-90.

Figure Legends

Figure 1

The structural organisation of the RCCX module

The RCCX cassette may occur as mono-, bi-, tri- or quadrimodular sequences (examples shown) and comprises four genes encoded in tandem: the serine/threonine kinase gene, *RPI* (or *STK19*), complement *C4* (*C4A* or *C4B*), cytochrome P450 steroid 21-hydroxylase, *CYP21A2*, and the extracellular matrix protein, *TNXB*. The *C4* gene in each RCCX module is usually functional. In contrast, the *RP* and *TNX* genes of duplicated RCCX modules are typically non-functional pseudogenes, known as *RP2* (or *STK19P*) and *TNXA* respectively. Additional *CYP21* genes may be functional (*CYP21B/CYP21A2*) or non-functional (*CYP21A/CYP21A1P*). (*adapted from* (Yang, *et al.*, 1999) and (Chung, *et al.*, 2002)).

Figure 2

Modular RCCX cassette on 6p23.1 illustrating location of PRT and REDVR amplicons

Figure 3

Cluster plots of raw PRT values against raw REDVR A values in CEU and the 1958 British Birth Cohort

These cluster plots illustrate raw PRT values (x-axis) against raw REDVR A values (y-axis) for the CEU cohort (top panel) and the 1958 British Birth Cohort (bottom panel). The coloured circles represent unique clusters as shown in the key.

Figure 4

SNP-C4 CNV correlation plots in CEU and the 1958 British Birth Cohort

(a) Correlation of *C4A* two-copy founders with surrounding SNPs in CEU

The correlation coefficient, r^2 , of SNPs 1 Mb centromeric and 1 Mb telomeric of the complement *C4* locus is plotted against genomic position. Each SNP is represented as a square. The size of the square is inversely proportional to the p-value for the correlation coefficient of each SNP. The squares are colour-coded to represent SNPs with similar correlation coefficients.

(b) Correlation of *C4B* two-copy founders with surrounding SNPs in CEU

The correlation coefficient, r^2 , of SNPs 1 Mb centromeric and 1 Mb telomeric of the complement *C4* locus is plotted against genomic position. Each SNP is represented as a square. The size of the square is inversely proportional to the p-value for the correlation coefficient of each SNP. The squares are colour-coded to represent SNPs with similar correlation coefficients.

(c) Correlation of *C4A* two-copy founders with surrounding SNPs in the 1958 British Birth Cohort

The correlation coefficient, r^2 , of SNPs 1 Mb centromeric and 1 Mb telomeric of the complement *C4* locus is plotted against genomic position. Each SNP is represented as a square. The size of the square is inversely proportional to the p-value for the correlation coefficient of each SNP. The squares are colour-coded to represent SNPs with similar correlation coefficients.

(d) Correlation of *C4B* two-copy founders with surrounding SNPs in the 1958 British Birth Cohort

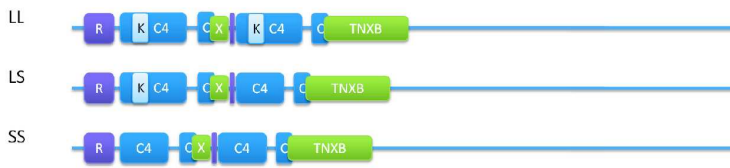
The correlation coefficient, r^2 , of SNPs 1 Mb centromeric and 1 Mb telomeric of the complement *C4* locus is plotted against genomic position. Each SNP is represented as a square. The size of the square is inversely proportional to the p-value for the correlation coefficient of each SNP. The squares are colour-coded to represent SNPs with similar correlation coefficients.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Monomodular RCCX



Bimodular RCCX



Trimodular RCCX



Quadrimodular RCCX

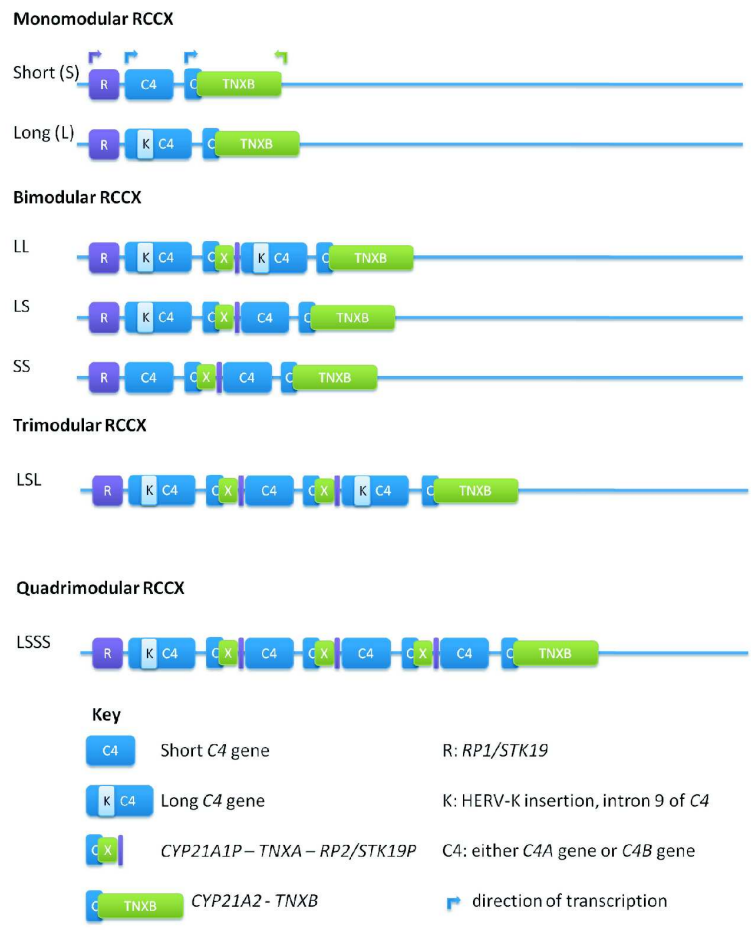


Key

- C4 Short C4 gene R: *RP1/STK19*
- K C4 Long C4 gene K: HERV-K insertion, intron 9 of C4
- C X *CYP21A1P - TNXA - RP2/STK19P* C4: either C4A gene or C4B gene
- C TNXB *CYP21A2 - TNXB* direction of transcription

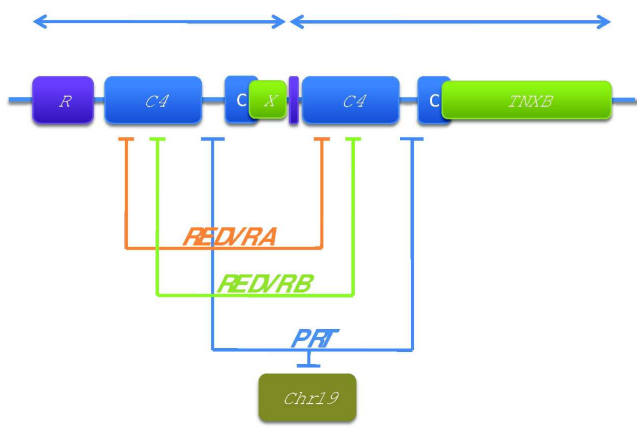
The structural organisation of the RCCX module
215x222mm (600 x 600 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



The structural organisation of the RCCX module
215x222mm (600 x 600 DPI)

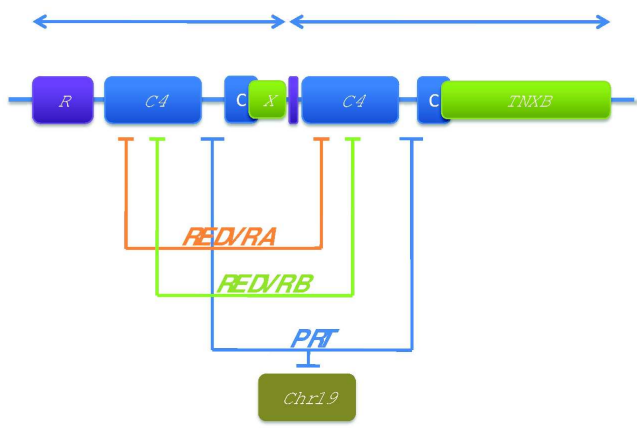
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Modular RCCX cassette on 6p23.1 illustrating location of PRT and REDVR amplicons
254x190mm (600 x 600 DPI)

Review

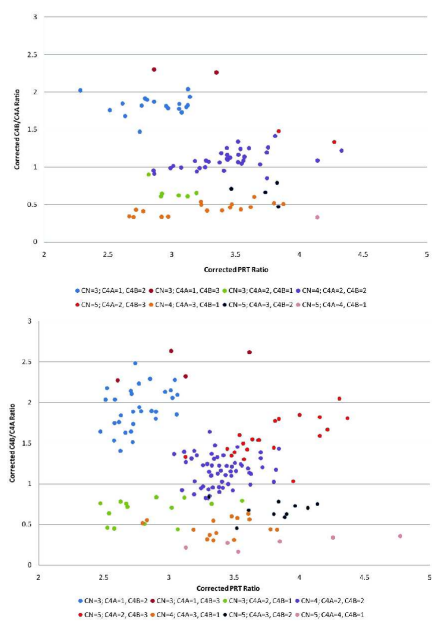
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Modular RCCX cassette on 6p23.1 illustrating location of PRT and REDVR amplicons
254x190mm (600 x 600 DPI)

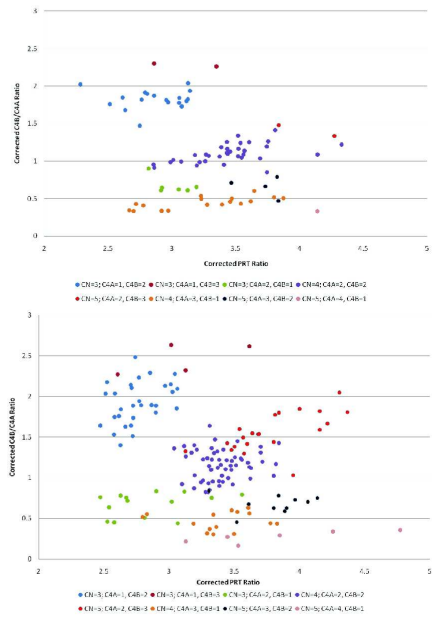
Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



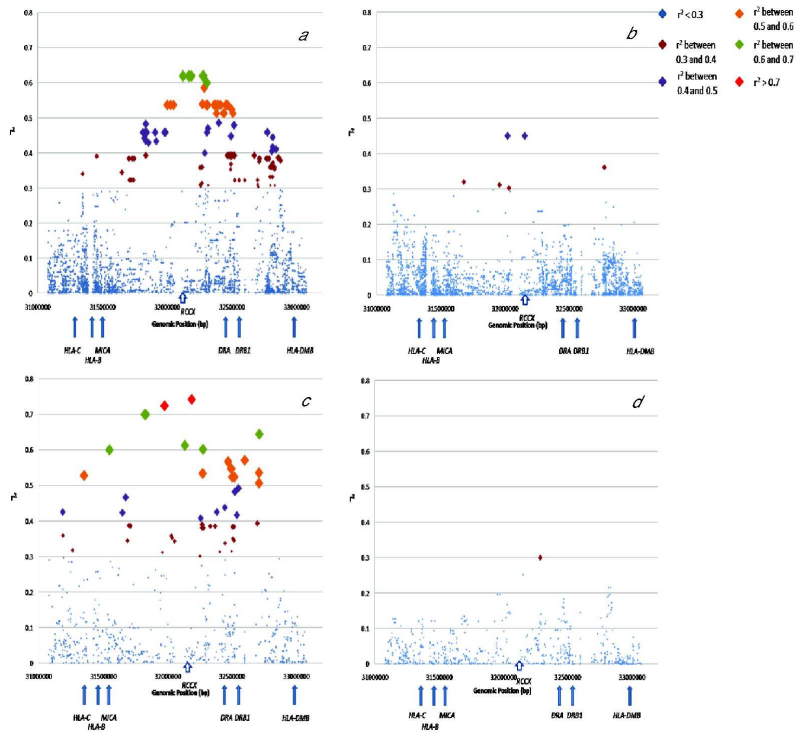
Cluster plots of raw PRT values against raw REDVR A values in CEU and the 1958 British Birth Cohort
254x190mm (600 x 600 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Cluster plots of raw PRT values against raw REDVR A values in CEU and the 1958 British Birth Cohort
254x190mm (600 x 600 DPI)

Review

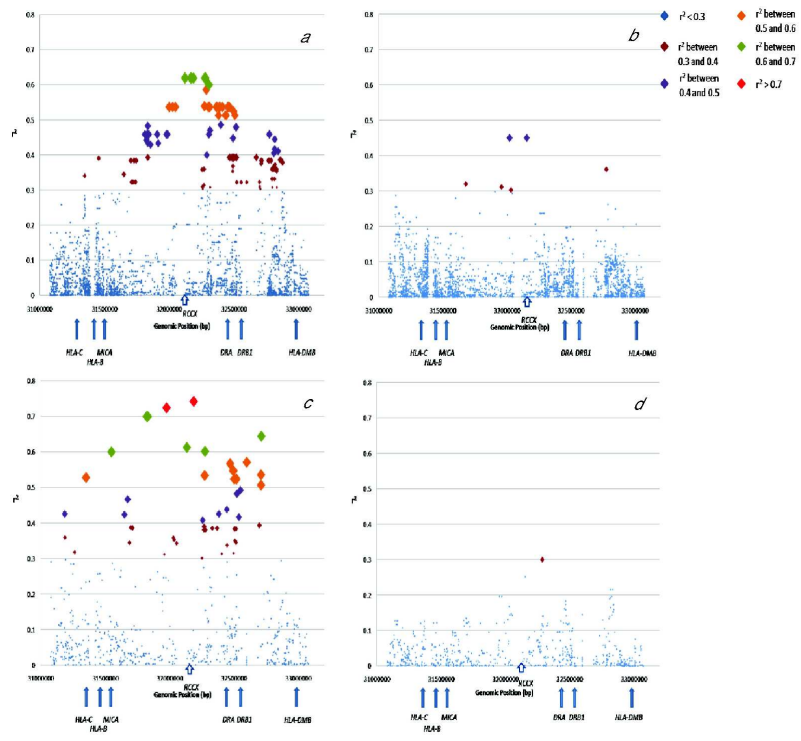


SNP-C4 CNV correlation plots in CEU and the 1958 British Birth Cohort
254x190mm (600 x 600 DPI)

Review

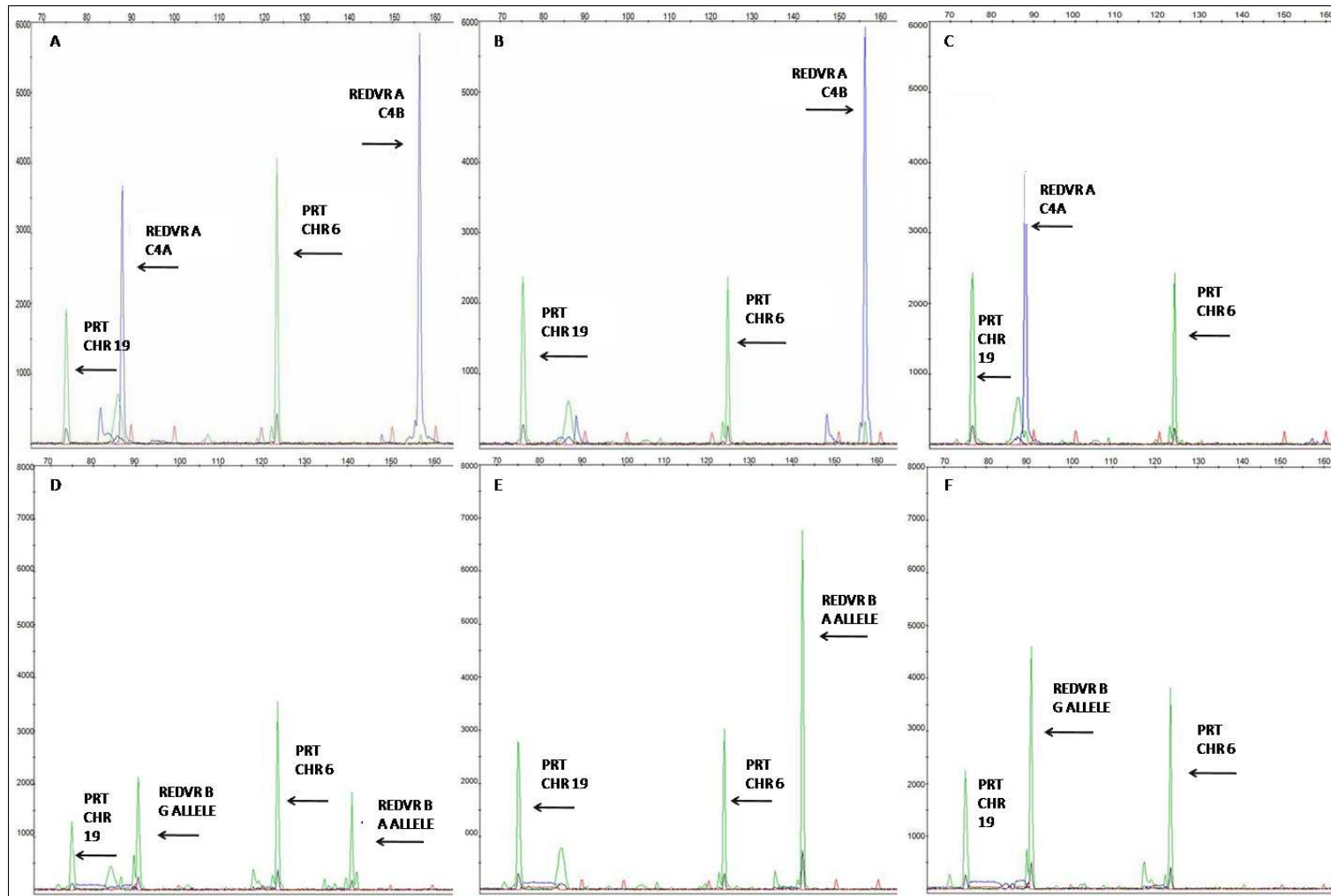
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



SNP-C4 CNV correlation plots in CEU and the 1958 British Birth Cohort
254x190mm (600 x 600 DPI)

Review



Supplementary Figure S1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47**Supplementary Figure S1 Legend**

Panels A - C illustrate representative Genescan outputs for the PRT and REDVR A. In panel A, the PRT test (125 bp chromosome 6) and reference (78 bp chromosome 19) amplicons show a peak area ratio of 2:1 corresponding to invariant reference copy number of two and total complement *C4* gene copy number of 4. Areas for the test (125 bp chromosome 6) and reference (78 bp chromosome 19) amplicons are 29875 and 18178 respectively. REDVR A amplicons show a 1:1 ratio for *C4A* (91 bp) and *C4B* (158 bp). Areas for the non-specific peak at ~82 bp – 2304, *C4A* 91 bp peak – 16179, *C4B* 158 bp peak – 21843. Panel B depicts a PRT with a 1:1 ratio for test and reference amplicons. REDVR A illustrates an individual who is homozygous for the *C4A* gene deletion. The area for the *C4B* 158 bp peak in this individual is 9891. Panel C shows an individual who is homozygous for the *C4B* gene deletion and a PRT with a 1:1 ratio for test and reference amplicons. The area for the *C4A* 91 bp peak in this individual is 20148. Panels D – F illustrate representative Genescan outputs for the PRT and REDVR B. Panel D shows a PRT with a peak area ratio of 2:1 and REDVR B with a 1:1 ratio for the A allele (144 bp) and the G allele (94 bp). The areas for the A allele 144 bp peak – 10441, and the G allele 94 bp peak – 13602. Panel E shows a PRT with a 1:1 ratio. REDVR B illustrates an individual in whom the A allele is absent (zero copies or null for A allele). The area for the G allele 94 bp peak in this individual is 11860. Panel F illustrates an individual in whom the G allele is absent (zero copies or null for G allele). The area for the A allele 144 bp peak in this individual is 19011.

Supplementary Table S1

(a) Prior means (standard deviation) of cluster means

Total C4	C4A	C4B	Cluster	PRT Mean	REDVR A Mean
2				2.5(0.033)	
2				2.5(0.033)	
3	1	2	21	2.9(0.033)	1.85(0.0167)
3	2	1	12	2.9(0.033)	0.65(0.0167)
4	1	3	31	3.4(0.033)	2.5(0.0167)
4	2	2	22	3.4(0.033)	1.2(0.0167)
4	3	1	13	3.4(0.033)	0.5(0.0167)
5	4	1	41	4.0(0.033)	0.25(0.0167)
5	2	3	32	4.0(0.033)	1.5(0.0167)
5	3	2	23	4.0(0.033)	0.66(0.0167)

(b) Cluster probabilities

Cluster	2:1	3:1	1:2	2:2	1:3	2:3	1:4	3:2
Prior Probability	0.2	0.02	0.15	0.45	0.15	0.01	0.01	0.01

Prior means (standard deviation) of cluster means

The coefficient of variation (standard deviation/mean) for the PRT, REDVR A and REDVR B variables were 0.035, 0.046 and 0.054 respectively. The standard deviation (s.d.) of 0.033 implies that we are confident that the cluster mean is not more than 0.1 (3 s.d.'s) above or below our prior mean. While the s.d. of 0.0167 on REDVR A mean implies that we are confident that the REDVR A cluster means are within 0.05 of our prior mean. For REDVR B, expectations and mixing probabilities are; 2 copies with expectations of (0, 1), $p=(0.5, 0.5)$, 3 copies with expectations of (0, 0.5, 2), $p=(0.25, 0.375, 0.375)$, 4 copies with expectations of (0, 0.33, 1, 3) $p=(0.125, 0.25, 0.375, 0.25)$, 5 copies with expectations of (0, 0.25, 0.67, 1.5, 4) $p=(0.0625, 0.15625, 0.3125, 0.3125, 0.15625)$, The s.d. on each non-zero mean is 0.0167 implying a similar

1
2
3
4
5 level of precision in these REDVR B means as for the REDVR A means. For the REDVR A-null observations each cluster (PRT=2, 3, 4) has the
6 same prior probability.
7

8
9 Cluster Coding:

10 21: Total Copy Number = 3, C4A Copy No=1, C4B Copy No=2
11 31: Total Copy Number= 4, C4A Copy No=1, C4B Copy No=3
12 12: Total Copy Number=3, C4A Copy No=2, C4B Copy No=1
13 13: Total Copy Number=4, C4A Copy No=3, C4B Copy No=1
14 22: Total Copy Number=4, C4A Copy No=2, C4B Copy No=2
15 32: Total Copy Number=5, C4A Copy No=2, C4B Copy No=3
16 23: Total Copy Number=5, C4A Copy No=3, C4B Copy No=2
17 14: Total Copy Number=5, C4A Copy No=4, C4B Copy No=1
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Supplementary Table S2

HLA-DRB1 genotype data on 24 CEU samples without previous data

Sample ID	Sample name	HLA DRB1_1	HLA DRB1_2	Comments
1341-13	NA06993	DRB1*0404/23	*0701/16	DRB1*0423 and 0716 not seen
1345-13	NA07345	DRB1*1501	*0701	
1350-11	NA11830	DRB1*1303	*0701/16	DRB1*0716 not seen
1358-12	NA12717	DRB1*0101	*1401/54	Unable to split DRB1*1401/54
1447-1	NA12752	DRB1*1501	*0701	
1447-9	NA12760	DRB1*0701/16	*0701/16	
1447-10	NA12761	DRB1*1501	*0301	
1447-11	NA12762	DRB1*0401	*0801	
1447-12	NA12763	DRB1*0103	*1501	
1454-1	NA12801	DRB1*1301	*0701/16	DRB1*0716 not seen
1454-2	NA12802	DRB1*0103/17	*0407/14	DRB1*0117 and 0414 not seen
1454-12	NA12812	DRB1*0401	*1301	
1454-13	NA12813	DRB1*0701/16	*0701/16	DRB1*0716 not seen
1454-14	NA12814	DRB1*0103	*1501	
1454-15	NA12815	DRB1*0407	*1301	
1459-1	NA12864	DRB1*1501	*0301	
1459-2	NA12865	DRB1*1501	*1601	
1459-9	NA12872	DRB1*0301	*0401	
1459-10	NA12873	DRB1*1501	*1501	
1459-11	NA12874	DRB1*1601	*0801	
1459-12	NA12875	DRB1*1501	*0801	
1463-2	NA12878	DRB1*0101	*0301	

1463-15	NA12891	DRB1*1501	*0301
1463-16	NA12892	DRB1*0101	*0101

Supplementary Table S3

Correlation of complement *C4A* two-copy individuals with surrounding SNPs in CEU

SNP	Position	r ²	p value
rs433061	32122806	0.6191	2.41E-10
rs1150753	32167845	0.6191	2.41E-10
rs1150752	32172704	0.6191	2.41E-10
rs1269852	32188169	0.6191	2.41E-10
rs3134942	32276749	0.6191	2.41E-10
rs3131296	32280971	0.6191	2.41E-10
rs3130297	32306959	0.6000	1.09E-09
rs3132956	32287416	0.5859	5.58E-09
rs2071278	32273422	0.5388	1.43E-08
rs497309	32000463	0.5365	1.59E-08
rs1270942	32026839	0.5365	1.59E-08
rs389884	32048876	0.5365	1.59E-08
rs3096690	32302608	0.5365	1.59E-08
rs6901158	32313920	0.5365	1.59E-08
rs9268144	32365177	0.5365	1.59E-08
rs7775397	32369230	0.5365	1.59E-08
rs9268208	32388569	0.5365	1.59E-08
rs9268219	32392086	0.5365	1.59E-08
rs1265754	32411670	0.5365	1.59E-08

rs2894254	32453667	0.5365	1.59E-08
rs3117103	32457535	0.5365	1.59E-08
rs3129950	32466179	0.5365	1.59E-08
rs3135363	32497626	0.5241	2.80E-08
rs9268177	32382860	0.5132	6.60E-08
rs2395149	32433540	0.5132	6.60E-08
rs3129927	32441805	0.5132	6.60E-08
rs3129856	32507315	0.5132	6.60E-08
rs9268235	32398186	0.4862	2.89E-07
rs3117575	31834232	0.4829	4.63E-07
rs3135394	32516475	0.4794	7.43E-07
rs3130304	32315159	0.4702	2.80E-07
rs1794282	32774504	0.4589	4.41E-07
rs3131383	31812273	0.4587	4.44E-07
rs3131379	31829012	0.4587	4.44E-07
rs3117574	31833209	0.4587	4.44E-07
rs3131378	31833264	0.4587	4.44E-07
rs3115671	31842324	0.4587	4.44E-07
rs7746496	31907055	0.4587	4.44E-07
rs558702	31978305	0.4587	4.44E-07
rs1314000	31986412	0.4587	4.44E-07
rs377763	32307122	0.4587	4.44E-07
rs3135378	32493077	0.4479	9.28E-07
rs9276375	32815563	0.4452	1.89E-06
rs3130484	31823861	0.4423	8.48E-07
rs3115672	31835876	0.4337	1.58E-06
rs3130679	31915519	0.4337	1.58E-06

1				
2				
3				
4				
5	rs599707	31916415	0.4337	1.58E-06
6	rs915652	31857121	0.4296	2.47E-06
7	rs9276357	32814537	0.4168	2.94E-06
8				
9	rs9276456	32823921	0.4114	4.72E-06
10	rs9276601	32842282	0.4109	2.76E-06
11	rs2859090	32808811	0.4056	4.40E-06
12				
13	rs3131290	32291153	0.4000	7.01E-06
14	rs3117577	31835453	0.3933	5.24E-06
15	rs3117097	32466667	0.3933	5.24E-06
16	rs3129953	32469799	0.3933	5.24E-06
17				
18	rs1980493	32471193	0.3933	5.24E-06
19	rs3129959	32483774	0.3933	5.24E-06
20	rs3129963	32488186	0.3933	5.24E-06
21	rs9268530	32491201	0.3933	5.24E-06
22	rs9268534	32491285	0.3933	5.24E-06
23				
24	rs6908056	32491338	0.3933	5.24E-06
25	rs6930933	32491388	0.3933	5.24E-06
26				
27	rs3135382	32491419	0.3933	5.24E-06
28	rs2001100	32491466	0.3933	5.24E-06
29	rs2001099	32491611	0.3933	5.24E-06
30	rs2001097	32491836	0.3933	5.24E-06
31				
32	rs2395159	32492055	0.3933	5.24E-06
33	rs2227138	32492478	0.3933	5.24E-06
34	rs3135380	32492655	0.3933	5.24E-06
35				
36	rs3135376	32493448	0.3933	5.24E-06
37	rs3135375	32493551	0.3933	5.24E-06
38	rs3135374	32493583	0.3933	5.24E-06
39				
40				
41				
42				
43				
44				
45				
46				
47				

rs3135373	32493631	0.3933	5.24E-06
rs3135372	32493760	0.3933	5.24E-06
rs2187820	32493851	0.3933	5.24E-06
rs2157336	32494039	0.3933	5.24E-06
rs3135368	32495256	0.3933	5.24E-06
rs2395161	32495730	0.3933	5.24E-06
rs2395162	32495758	0.3933	5.24E-06
rs2395164	32495838	0.3933	5.24E-06
rs2395167	32496286	0.3933	5.24E-06
rs2213580	32496552	0.3933	5.24E-06
rs3135366	32496687	0.3933	5.24E-06
rs6457577	32499494	0.3933	5.24E-06
rs3135353	32500855	0.3933	5.24E-06
rs2395171	32502515	0.3933	5.24E-06
rs3129861	32509510	0.3933	5.24E-06
rs2239806	32519285	0.3933	5.24E-06
rs2239805	32519354	0.3933	5.24E-06
rs1051336	32520570	0.3933	5.24E-06
rs1041885	32520787	0.3933	5.24E-06
rs4599680	32674272	0.3933	5.24E-06
rs4143333	31456498	0.3906	1.62E-05
rs2001098	32491815	0.3884	8.05E-06
rs9276665	32855480	0.3861	6.77E-06
rs4947349	32855768	0.3861	6.77E-06
rs9276689	32859940	0.3861	6.77E-06
rs7762279	32863268	0.3861	6.77E-06
rs3132450	31704117	0.3841	7.25E-06

1				
2				
3				
4				
5	rs3117582	31728499	0.3841	7.25E-06
6	rs9267531	31744721	0.3841	7.25E-06
7	rs2040410	32710676	0.3841	7.25E-06
8	rs2187668	32713862	0.3841	7.25E-06
9	rs2856674	32767623	0.3841	7.25E-06
10	rs9275294	32771333	0.3841	7.25E-06
11	rs9275576	32787304	0.3841	7.25E-06
12	rs4947350	32875598	0.3786	8.80E-06
13	rs2040406	32710985	0.3754	1.61E-05
14	rs9276401	32817182	0.3720	2.94E-05
15	rs3135381	32492321	0.3678	1.63E-05
16	rs9276190	32807328	0.3661	1.36E-05
17	rs204990	32269408	0.3593	1.71E-05
18	rs204989	32269830	0.3593	1.71E-05
19	rs9275652	32793800	0.3593	1.71E-05
20	rs9275659	32794081	0.3593	1.71E-05
21	rs9275686	32795548	0.3593	1.71E-05
22	rs9275738	32796817	0.3593	1.71E-05
23	rs763026	32799723	0.3593	1.71E-05
24	rs9275936	32800911	0.3593	1.71E-05
25	rs9275970	32801593	0.3593	1.71E-05
26	rs9276132	32805733	0.3593	1.71E-05
27	rs9276149	32806134	0.3593	1.71E-05
28	rs9276171	32806896	0.3593	1.71E-05
29	rs9276172	32806925	0.3593	1.71E-05
30	rs2858884	32808061	0.3593	1.71E-05
31	rs6913505	32810359	0.3593	1.71E-05
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				

1				
2				
3				
4				
5	rs2859078	32810427	0.3593	1.71E-05
6	rs2859076	32810564	0.3593	1.71E-05
7	rs9276313	32812684	0.3593	1.71E-05
8	rs5021448	32815684	0.3593	1.71E-05
9	rs9276425	32819028	0.3593	1.71E-05
10	rs9276435	32821845	0.3593	1.71E-05
11	rs9276556	32831211	0.3593	1.71E-05
12	rs3096689	32253977	0.3579	2.27E-05
13	rs9276550	32830221	0.3535	2.62E-05
14	rs6930571	32491186	0.3529	5.32E-05
15	rs1800629	31651010	0.3441	2.86E-05
16	rs9264636	31346276	0.3402	7.82E-05
17	rs9275690	32795698	0.3316	6.61E-05
18	rs9276351	32814394	0.3316	6.61E-05
19	rs3130626	31706468	0.3228	5.75E-05
20	rs2736157	31708799	0.3228	5.75E-05
21	rs3115663	31709822	0.3228	5.75E-05
22	rs9267522	31711749	0.3228	5.75E-05
23	rs3130628	31717251	0.3228	5.75E-05
24	rs3130047	31718458	0.3228	5.75E-05
25	rs3117583	31727555	0.3228	5.75E-05
26	rs3130618	31740113	0.3228	5.75E-05
27	rs3117579	31741475	0.3228	5.75E-05
28	rs3129763	32698903	0.3228	5.75E-05
29	rs7452076	32523058	0.3223	5.84E-05
30	rs7749057	32556882	0.3223	5.84E-05
31	DRB1_[301]	32600006	0.3223	5.84E-05
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				

rs204991	32269344	0.3139	7.65E-05
rs9271612	32699632	0.3108	1.03E-04
rs3130349	32255674	0.3095	8.82E-05
rs8365	32256381	0.3095	8.82E-05
rs3131300	32259912	0.3095	8.82E-05
rs424232	32316302	0.3077	1.13E-04
rs9275741	32796850	0.3073	1.39E-04
rs9276555	32830732	0.3071	1.70E-04
rs9271613	32699644	0.3066	1.42E-04
rs1800625	32260420	0.3034	1.30E-04
rs9271850	32703038	0.3027	1.60E-04

Supplementary Table S4

Correlation of complement *C4B* two-copy individuals with surrounding SNPs in CEU

SNP	Position	r ²	p value
rs9332730	32019988	0.4504	1.11E-08
rs7774197	32154253	0.4504	1.11E-08
rs17427599	32775342	0.3614	1.53E-06
rs2736186	31680312	0.3199	4.54E-06
rs555007	31958311	0.3115	6.44E-06
rs9501161	32032306	0.3026	9.30E-06

Supplementary Table S5

Correlation of complement *C4A* two-copy individuals with surrounding SNPs in the 1958 British Birth Cohort

SNP	Position	r ²	p value
rs1269852	32188168	0.7418	<1.00E-16
rs558702	31978304	0.7237	<1.00E-16
rs3131378	31833263	0.6996	<1.00E-16
rs3131379	31829011	0.6996	<1.00E-16
rs3130484	31823860	0.6996	<1.00E-16
rs2187668	32713861	0.6443	<1.00E-16
rs1150758	32136126	0.6126	<1.00E-16
rs3134942	32276748	0.6016	<1.00E-16
rs3131643	31550760	0.6001	<1.00E-16
DRB1_[3]	32600004	0.5713	<1.00E-16
rs1980493	32471192	0.5672	<1.00E-16
rs3135366	32496686	0.5471	<1.00E-16
rs2187820	32493850	0.5471	<1.00E-16
rs2040410	32710675	0.5356	<1.00E-16
rs2071278	32273421	0.5338	<1.00E-16
rs2524069	31352767	0.5278	<1.00E-16
rs2239805	32519353	0.5237	<1.00E-16
rs2395171	32502514	0.5237	<1.00E-16
rs2040406	32710984	0.5063	<1.00E-16
rs6901541	32550238	0.4924	<1.00E-16
rs7452076	32523058	0.4823	<1.00E-16
rs2857595	31676447	0.4664	<1.00E-16
rs3129939	32444743	0.4382	1.11E-16

1				
2				
3				
4				
5	rs7341328	32383171	0.4254	2.22E-16
6	rs2233974	31187994	0.4254	2.22E-16
7	rs1800629	31651009	0.4235	3.33E-16
8	rs7766843	32538706	0.4164	5.55E-16
9	rs4434496	32538485	0.4164	5.55E-16
10	rs1800625	32260419	0.4078	1.44E-15
11	rs3130349	32255673	0.4078	1.44E-15
12	rs3129763	32698902	0.3935	8.33E-15
13	rs204991	32269343	0.3902	8.99E-15
14	rs9267522	31711748	0.3871	1.22E-14
15	rs3130622	31700502	0.3871	1.22E-14
16	rs10885	31712569	0.3853	1.88E-14
17	rs3749966	32369484	0.3851	1.51E-14
18	rs3115569	32332117	0.3851	1.51E-14
19	rs3129876	32515989	0.3834	1.78E-14
20	rs3135340	32506849	0.3834	1.78E-14
21	rs3129845	32504254	0.3834	1.78E-14
22	rs3132935	32279052	0.3803	2.43E-14
23	rs3134605	32267933	0.3803	2.43E-14
24	rs2233956	31189183	0.359	1.97E-13
25	rs630379	32030232	0.358	2.15E-13
26	rs437179	32036992	0.3527	3.58E-13
27	rs440454	32035320	0.3527	3.58E-13
28	rs3129859	32508916	0.35	4.65E-13
29	rs3129878	32516712	0.345	7.47E-13
30	rs3132451	31690003	0.3444	7.89E-13
31	rs389883	32055438	0.3429	9.08E-13
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				

rs2050189	32447624	0.3373	1.53E-12
rs1265181	31263763	0.3177	9.39E-12
rs3135363	32497625	0.3145	1.25E-11
rs3129949	32406791	0.3127	1.48E-11
rs652888	31959212	0.3111	1.70E-11
rs3134945	32254469	0.3007	4.31E-11
rs3131297	32248982	0.3007	4.31E-11

Supplementary Table S6

Correlation of complement *C4B* two-copy individuals with surrounding SNPs in the 1958 British Birth Cohort

SNP	Position	r^2	p value
rs2071285	32288408	0.3001	1.06E-11

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Supplementary Table S7

(a) Comparison of Southern blot data and PRT C4 CNV data in 89 CEU samples with *HLA-DRB1* genotypes

Sample ID	Father	Mother	DRB1 1	DRB1 2	SB Total C4	SB C4A	SB C4B	Total C4-PRT	C4A-PRT	C4B-PRT	Probability-PRT
1408-13	0	0	407	701	5	3	2	3	2	1	1
1454-12	0	0	401	1301	5	4	1	4	3	1	1
1347-15	0	0	401	1501	4	2	2	5	2	3	1
1463-16	0	0	101	101	4	3	1	5	4	1	1
1459-2	1459-11	1459-12	1501	1601	4	3	1	5	3	2	0.99
1375-12	0	0	701	1404	3	2	1	4	3	1	1
1420-10	0	0	101	1501	3	1	2	4	1	3	1
1447-9	0	0	701	701	4	2	2	3	2	1	1
1362-15	0	0	402	404	5	3	2	5	3	2	1
1362-16	0	0	1301	1501	4	2	2	4	2	2	1
1340-1	1340-9	1340-10	404	1101	4	2	2	4	2	2	1
1346-1	1346-11	1346-12	101	1301	4	4	0	4	4	0	1
1416-11	0	0	701	1501	4	2	2	4	2	2	1
1334-12	0	0	101	1501	4	3	1	4	3	1	1
1350-11	0	0	1303	701	5	2	3	5	2	3	1
1362-2	1362-15	1362-16	404	1301	4	2	2	4	2	2	1
1454-2	1454-14	1454-15	103	407	4	3	1	4	3	1	1
1454-14	0	0	103	1501	4	2	2	4	2	2	1
1362-1	1362-13	1362-14	101	101	4	2	2	4	2	2	1
1459-10	0	0	1501	1501	4	2	2	4	2	2	1
1454-1	1454-12	1454-13	1301	701	5	3	2	5	3	2	1
1347-14	0	0	1501	1501	4	2	2	4	2	2	0.7
1349-13	0	0	401	1501	4	3	1	4	3	1	1

1												
2												
3												
4												
5	1341-11	0	0	401	701	4	3	1	4	3	1	1
6	1345-13	0	0	1501	701	4	2	2	4	2	2	1
7	1420-11	0	0	1501	1501	4	2	2	4	2	2	1
8	1349-14	0	0	404	701	4	2	2	4	2	2	1
9	1334-11	0	0	1101	1501	4	2	2	4	2	2	1
10	1349-2	1349-13	1349-14	401	404	4	3	1	4	3	1	1
11	1447-12	0	0	1501	103	4	2	2	4	2	2	1
12	1344-1	1344-12	1344-13	404	701	4	2	2	4	2	2	1
13	1447-2	1447-11	1447-12			4	2	2	4	2	2	1
14	1420-12	0	0	1501	404	4	2	2	4	2	2	1
15	1334-2	1334-12	1334-13	101	101	4	3	1	4	3	1	1
16	1350-1	1350-10	1350-11	401	1303	5	3	2	5	3	2	1
17	1346-11	0	0	101	404	4	3	1	4	3	1	1
18	1344-12	0	0	404	1501	4	2	2	4	2	2	1
19	1416-1	1416-11	1416-12	701	1302	4	2	2	4	2	2	1
20	1347-2	1347-14	1347-15	401	1501	4	2	2	4	2	2	1
21	1344-13	0	0	701	301	4	2	2	4	2	2	1
22	1334-10	0	0	101	407	4	4	0	4	4	0	1
23	1375-11	0	0	404	404	4	2	2	4	2	2	1
24	1334-1	1334-10	1334-11	407	1101	4	3	1	4	3	1	1
25	1345-2	1345-12	1345-13	1501	404	4	2	2	4	2	2	1
26	1416-12	0	0	401	1302	4	2	2	4	2	2	1
27	1341-13	0	0	404	701	4	1	3	4	1	3	1
28	1340-9	0	0	101	404	4	2	2	4	2	2	1
29	1350-10	0	0	401	401	4	3	1	4	3	1	1
30	1420-2	1420-11	1420-12	1501	1501	4	2	2	4	2	2	1
31	1454-13	0	0	701	701	4	2	2	4	2	2	1
32												
33												
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												
47												

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1454-15	0	0	407	1301	4	3	1	4	3	1	1
1447-1	1447-9	1447-10	701	1501	4	2	2	4	2	2	1
1341-14	0	0	1501	1501	4	2	2	4	2	2	1
1341-12	0	0	1401	1501	3	2	1	3	2	1	1
1459-12	0	0	1501	801	4	2	2	4	2	2	1
1408-1	1408-10	1408-11	301	1201	3	1	2	3	1	2	1
1459-1	1459-9	1459-10	1501	301	3	1	2	3	1	2	1
1375-2	1375-11	1375-12	404	1404	3	2	1	3	2	1	1
1447-10	0	0	1501	301	3	1	2	3	1	2	1
1459-9	0	0	301	401	3	1	2	3	1	2	1
1350-13	0	0	1301	1501	3	1	2	3	1	2	1
1362-14	0	0	101	1501	4	2	2	4	2	2	1
1463-15	0	0	1501	301	3	1	2	3	1	2	1
1341-2	1341-13	1341-14	404	1501	3	1	2	3	1	2	1
1408-12	0	0	301	401	3	2	1	3	2	1	1
1341-1	1341-11	1341-12	401	1501	3	3	0	3	3	0	0.98
1362-13	0	0	101	301	4	2	2	4	2	2	1
1447-11	0	0	801	401	4	2	2	4	2	2	1
1358-12	0	0	101	1401/54	4	3	1	4	3	1	1
1345-12	0	0	404	301	3	1	2	3	1	2	1
1444-14	0	0	401	301	3	1	2	3	1	2	1
1408-2	1408-12	1408-13	301	407	3	2	1	3	2	1	1
1408-10	0	0	301	404	3	1	2	3	1	2	1
1444-13	0	0	1501	404	4	2	2	4	2	2	1
1334-13	0	0	101	401	4	2	2	4	2	2	1
1420-9	0	0	1501	401	3	1	2	3	1	2	1
1340-10	0	0	301	1101	3	1	2	3	1	2	1

1358-1	1358-11	1358-12	101	405	4	3	1	4	3	1	1
1444-2	1444-13	1444-14	301	404	3	1	2	3	1	2	1
1340-12	0	0	301	701	3	1	2	3	1	2	1
1459-11	0	0	1601	801	4	3	1	4	3	1	1
1346-12	0	0	1301	1501	4	3	1	4	3	1	1
1358-11	0	0	101	405	4	3	1	4	3	1	1
1350-2	1350-12	1350-13	301	1301	2	0	2	2	0	2	0.94
1340-11	0	0	404	701	3	1	2	3	1	2	1
1350-12	0	0	301	1501	3	1	2	3	1	2	1
1463-2	1463-15	1463-16	101	301	3	1	2	3	1	2	1
1340-2	1340-11	1340-12	301	404	2	0	2	2	0	2	1
1420-1	1420-9	1420-10	1501	1501	3	1	2	3	1	2	1

(b) Comparison of Southern blot data with raw and clustered C4 CNV PRT data in 89 CEU samples with *HLA-DRB1* genotypes and alternate CEU sample IDs

Sample ID	Father	Mother	alt ID	DRB1 1	DRB1 2	Total C4-PRT (raw ratio)	REDVR A ratio (C4B/C4A)	REDVR B ratio	SBTotal C4	SB C4A	SB C4B	Total C4-PRT	C4A-PRT	C4B-PRT	PRT-Probability
1408-13	0	0	NA12156	407	701	2.926	0.643	0.51	5	3	2	3	2	1	1
1454-12	0	0	NA12812	401	1301	2.922	0.329	3.087	5	4	1	4	3	1	1
1347-15	0	0	NA11882	401	1501	4.277	1.336	0.776	4	2	2	5	2	3	1
1463-16	0	0	NA12892	101	101	4.14	0.326	0.23	4	3	1	5	4	1	1
1459-2	1459-11	1459-12	NA12865	1501	1601	3.837	0.471	0.672	4	3	1	5	3	2	0.99221
1375-12	0	0	NA12234	701	1404	3.23	0.494	1.256	3	2	1	4	3	1	1
1420-10	0	0	NA12004	101	1501	2.863	2.295	0.73	3	1	2	4	1	3	1
1447-9	0	0	NA12760	701	701	2.818	0.902	2.233	4	2	2	3	2	1	0.99687
1362-15	0	0	NA11994	402	404	3.825	0.787	1.276	5	3	2	5	3	2	1
1362-16	0	0	NA11995	1301	1501	3.46	1.124	0.836	4	2	2	4	2	2	1
1340-1	1340-9	1340-10	NA07029	404	1101	4.331	1.222	0.98	4	2	2	4	2	2	0.99843
1346-1	1346-11	1346-12	NA10857	101	1301	4.184	C4B null	1.154	4	4	0	4	4	0	1
1416-11	0	0	NA12248	701	1501	4.145	1.085	1.079	4	2	2	4	2	2	1

1																
2																
3																
4																
5	1334-12	0	0	NA12146	101	1501	3.874	0.504	0.282	4	3	1	4	3	1	0.99608
6	1350-11	0	0	NA11830	1303	701	3.841	1.48	1.273	5	2	3	5	2	3	1
7	1362-2	1362-15	1362-16	NA10861	404	1301	3.81	1.416	0.899	4	2	2	4	2	2	0.99978
8	1454-2	1454-14	1454-15	NA12802	103	407	3.801	0.52	0.304	4	3	1	4	3	1	0.99992
9	1454-14	0	0	NA12814	103	1501	3.754	1.258	0.901	4	2	2	4	2	2	1
10	1362-1	1362-13	1362-14	NA10860	101	101	3.747	1.195	0.916	4	2	2	4	2	2	1
11	1459-10	0	0	NA12873	1501	1501	3.746	0.855	0.935	4	2	2	4	2	2	1
12	1454-1	1454-12	1454-13	NA12801	1301	701	3.735	0.664	1.205	5	3	2	5	3	2	1
13	1347-14	0	0	NA11881	1501	1501	3.695	1.037	0.639	4	2	2	4	2	2	0.70218
14	1349-13	0	0	NA11839	401	1501	3.647	0.604	2.654	4	3	1	4	3	1	1
15	1341-11	0	0	NA07034	401	701	3.622	0.467	1	4	3	1	4	3	1	1
16	1345-13	0	0	NA07345	1501	701	3.605	1.249	2.437	4	2	2	4	2	2	1
17	1420-11	0	0	NA12005	1501	1501	3.571	1.133	0.833	4	2	2	4	2	2	1
18	1349-14	0	0	NA11840	404	701	3.562	1.084	2.787	4	2	2	4	2	2	1
19	1334-11	0	0	NA12145	1101	1501	3.547	1.046	0.769	4	2	2	4	2	2	1
20	1349-2	1349-13	1349-14	NA10854	401	404	3.54	0.43	2.452	4	3	1	4	3	1	1
21	1447-12	0	0	NA12763	1501	103	3.539	1.239	1.01	4	2	2	4	2	2	1
22	1344-1	1344-12	1344-13	NA10851	404	701	3.525	1.166	2.5	4	2	2	4	2	2	1
23	1447-2	1447-11	1447-12	NA12753			3.522	1.341	1.05	4	2	2	4	2	2	1
24	1420-12	0	0	NA12006	1501	404	3.521	1.066	0.816	4	2	2	4	2	2	1
25	1334-2	1334-12	1334-13	NA10847	101	101	3.47	0.498	0.333	4	3	1	4	3	1	1
26	1350-1	1350-10	1350-11	NA10856	401	1303	3.469	0.706	3.049	5	3	2	5	3	2	0.99657
27	1346-11	0	0	NA12043	101	404	3.461	0.46	0.321	4	3	1	4	3	1	1
28	1344-12	0	0	NA12056	404	1501	3.44	1.098	0.934	4	2	2	4	2	2	1
29	1416-1	1416-11	1416-12	NA10835	701	1302	3.437	1.158	0.799	4	2	2	4	2	2	1
30	1347-2	1347-14	1347-15	NA10859	401	1501	3.434	1.256	0.887	4	2	2	4	2	2	1
31	1344-13	0	0	NA12057	701	301	3.43	1.113	2.281	4	2	2	4	2	2	1
32	1334-10	0	0	NA12144	101	407	3.425	C4B null	0	4	4	0	4	4	0	0.99998
33	1375-11	0	0	NA12264	404	404	3.41	0.954	0.967	4	2	2	4	2	2	1
34	1334-1	1334-10	1334-11	NA10846	407	1101	3.396	0.423	0.383	4	3	1	4	3	1	1
35	1345-2	1345-12	1345-13	NA07348	1501	404	3.39	1.186	0.911	4	2	2	4	2	2	1
36	1416-12	0	0	NA12249	401	1302	3.374	1.062	1.187	4	2	2	4	2	2	1
37	1341-13	0	0	NA06993	404	701	3.352	2.257	0.429	4	1	3	4	1	3	0.99983
38	1340-9	0	0	NA06994	101	404	3.29	1.071	0.927	4	2	2	4	2	2	1
39	1350-10	0	0	NA11829	401	401	3.28	0.415	2.326	4	3	1	4	3	1	1
40	1420-2	1420-11	1420-12	NA10839	1501	1501	3.271	1.085	1.004	4	2	2	4	2	2	1
41																
42																
43																
44																
45																
46																
47																

1																
2																
3																
4																
5	1454-13	0	0	NA12813	701	701	3.265	0.995	2.628	4	2	2	4	2	2	1
6	1454-15	0	0	NA12815	407	1301	3.228	0.533	0.335	4	3	1	4	3	1	1
7	1447-1	1447-9	1447-10	NA12752	701	1501	3.221	0.988	0.391	4	2	2	4	2	2	1
8	1341-14	0	0	NA06985	1501	1501	3.197	0.943	2.463	4	2	2	4	2	2	1
9	1341-12	0	0	NA07055	1401	1501	3.192	0.653	1.992	3	2	1	3	2	1	1
10	1459-12	0	0	NA12875	1501	801	3.181	1.083	0.906	4	2	2	4	2	2	1
11	1408-1	1408-10	1408-11	NA10830	301	1201	3.144	1.937	0.43	3	1	2	3	1	2	1
12	1459-1	1459-9	1459-10	NA12864	1501	301	3.13	2.032	0.429	3	1	2	3	1	2	1
13	1375-2	1375-11	1375-12	NA10863	404	1404	3.127	0.608	1.736	3	2	1	3	2	1	1
14	1447-10	0	0	NA12761	1501	301	3.125	1.82	0.42	3	1	2	3	1	2	1
15	1459-9	0	0	NA12872	301	401	3.113	1.794	0.402	3	1	2	3	1	2	1
16	1350-13	0	0	NA11832	1301	1501	3.08	1.724	0.409	3	1	2	3	1	2	1
17	1362-14	0	0	NA11993	101	1501	3.072	0.99	0.78	4	2	2	4	2	2	1
18	1463-15	0	0	NA12891	1501	301	3.06	1.778	0.427	3	1	2	3	1	2	1
19	1341-2	1341-13	1341-14	NA06991	404	1501	3.058	1.843	0.395	3	1	2	3	1	2	1
20	1408-12	0	0	NA12155	301	401	3.057	0.625	2.146	3	2	1	3	2	1	1
21	1341-1	1341-11	1341-12	NA07048	401	1501	3.04	C4B null	0	3	3	0	3	3	0	0.98479
22	1362-13	0	0	NA11992	101	301	3.011	1.017	0.785	4	2	2	4	2	2	1
23	1447-11	0	0	NA12762	801	401	2.993	0.989	0.759	4	2	2	4	2	2	1
24	1358-12	0	0	NA12717	101	1401/5 4	2.975	0.337	0	4	3	1	4	3	1	1
25	1345-12	0	0	NA07357	404	301	2.971	1.779	0.448	3	1	2	3	1	2	1
26	1444-14	0	0	NA12751	401	301	2.957	1.81	0.449	3	1	2	3	1	2	1
27	1408-2	1408-12	1408-13	NA10831	301	407	2.915	0.608	0	3	2	1	3	2	1	1
28	1408-10	0	0	NA12154	301	404	2.864	1.871	0.374	3	1	2	3	1	2	1
29	1444-13	0	0	NA12750	1501	404	2.862	0.915	0.726	4	2	2	4	2	2	1
30	1334-13	0	0	NA12239	101	401	2.857	0.951	0.765	4	2	2	4	2	2	1
31	1420-9	0	0	NA12003	1501	401	2.806	1.895	1.689	3	1	2	3	1	2	1
32	1340-10	0	0	NA07000	301	1101	2.793	1.909	0.413	3	1	2	3	1	2	1
33	1358-1	1358-11	1358-12	NA12707	101	405	2.779	0.405	0	4	3	1	4	3	1	0.99949
34	1444-2	1444-13	1444-14	NA12740	301	404	2.764	1.815	0	3	1	2	3	1	2	1
35	1340-12	0	0	NA07056	301	701	2.748	1.467	1.577	3	1	2	3	1	2	1
36	1459-11	0	0	NA12874	1601	801	2.718	0.424	0.225	4	3	1	4	3	1	1
37	1346-12	0	0	NA12044	1301	1501	2.703	0.332	2.455	4	3	1	4	3	1	1
38	1358-11	0	0	NA12716	101	405	2.667	0.342	1.11	4	3	1	4	3	1	1
39	1350-2	1350-12	1350-13	NA10855	301	1301	2.665	C4A null	0	2	0	2	2	0	2	0.93531
40	1340-11	0	0	NA07022	404	701	2.635	1.674	0.366	3	1	2	3	1	2	1
41																
42																
43																
44																
45																
46																
47																

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1350-12	0	0	NA11831	301	1501	2.614	1.845	0.334	3	1	2	3	1	2	1
1463-2	1463-15	1463-16	NA12878	101	301	2.515	1.755	0.361	3	1	2	3	1	2	1
1340-2	1340-11	1340-12	NA07019	301	404	2.352	C4A null	0	2	0	2	2	0	2	1
1420-1	1420-9	1420-10	NA10838	1501	1501	2.283	2.017	0.438	3	1	2	3	1	2	1

Supplementary Table S8
Discordant samples between Southern blot and PRT C4 CNV data in 89 CEU samples

Copy No	2		3			4			5		4+
No of Samples	2		25			56			6		62
Discrepancies (%)	0 (0)		2 (8)			4 (7.2)			2 (33.3)		6 (9.6)
C4A:C4B	0:2	3:0	1:2	2:1	4:0	3:1	2:2	1:3	1:4	3:2	2:3
No of Samples	2	1	19	5	2	17	36	1	1	4	1
Discrepancies (%)	0 (0)	0(0)	1(5.26)	1 (20)	0(0)	2 (11.7)	2 (5.55)	0(0)	1(100)	1(25)	0 (0)