Published in final edited form as:

Mol Immunol. 2009 April ; 46(7): 1289–1303. doi:10.1016/j.molimm.2008.11.018.

Great Genotypic and Phenotypic Diversities Associated with Copy-Number Variations of Complement *C4* and *RP-C4-CYP21-TNX* (RCCX) Modules: a Comparison of Asian Indian and European American Populations

Kapil Saxena¹, Kathryn J. Kitzmiller^{1,2}, Yee Ling Wu^{1,2}, Bi Zhou¹, Nazreen Esack¹, Leena Hiremath³, Erwin K. Chung^{1,4}, Yan Yang^{1,4}, and C. Yung Yu^{1,2,4,*}

¹ Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive, Columbus Ohio 43205

² Department of Pediatrics, The Ohio State University, Columbus, Ohio

³ Department of Internal Medicine, The Ohio State University, Columbus, Ohio

⁴ Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, Ohio

Abstract

Inter-individual gene copy-number variations (CNVs) probably afford human populations the flexibility to respond to a variety of environmental challenges, but also lead to differential disease predispositions. We investigated gene CNVs for complement component *C4* and steroid 21-hydroxylase from the *RP-C4-CYP21-TNX* (RCCX) modules located in the major histocompatibility complex among healthy Asian-Indian Americans (AIA) and compared them to European Americans. A combination of definitive techniques that yielded cross-confirmatory results was used. The medium gene copy-numbers for *C4* and its isotypes, acidic *C4A* and basic *C4B*, were 4, 2 and 2, respectively, but their frequencies were only 53–56%. The distribution patterns for total *C4* and *C4A* are skewed towards the high copy-number side. For example, the frequency of AIA-subjects with three copies of *C4A* (30.7%) was 3.92-fold of those with a single copy (7.83%). The monomodal-short haplotype with a single *C4B* gene and the absence of *C4A*, which is in linkage- disequilibrium with HLA *DRB1*0301* in Europeans and a strong risk factor for autoimmune diseases, has a frequency of 0.012 in AIA but 0.106 among healthy European Americans ($p=6.6\times10^{-8}$). The copy-number and the size of *C4* genes strongly determine the plasma *C4* protein concentrations. Parallel variations in copy-numbers of *CYP21A* (*CYP21A1P*) and *TNXA* with total *C4* were also observed. Notably, 13.1% of AIA-subjects had three copies of the functional *CYP21B*, which were likely generated by recombinations between monomodal and bimodal RCCX haplotypes. The high copy-numbers of *C4* and the high frequency of RCCX recombinants offer important insights to the prevalence of autoimmune and genetic diseases.

*Address correspondence to C. Yung Yu, D.Phil., email: E-mail: yuc@chi.osu.edu.

Gene symbols are italicized, protein symbols are in regular fonts.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

CNV; complement C4A and C4B; congenital adrenal hyperplasia; extracellular matrix protein tenascin-X; graft rejections and C4 polymorphisms; recombinations; steroid CYP21-hydroxylase; systemic lupus erythematosus

1. Introduction

Among healthy human subjects of each gender, the number of nuclear genes has been presumed to be constant. Such concept of constant gene copy-numbers is being gradually revised in the past twenty years, albeit unnoticeably. For examples, complement C4 and amylase manifest frequent and heritable gene copy-number variations among different healthy individuals (Bank *et al.*, 1992; Blanchong *et al.*, 2000; Carroll *et al.*, 1984; Chung *et al.*, 2002a; Dangel *et al.*, 1994; Perry *et al.*, 2007; Shen *et al.*, 1994; Wu *et al.*, 2007; Yang *et al.*, 2003; Yang *et al.*, 2007; Yang *et al.*, 1999; Yu, 1991; Yu *et al.*, 2003; Yu and Campbell, 1987). Recent advent of whole genome microarray studies and personal genomic DNA sequencing revealed numerous genomic loci with duplications of DNA segments > 1 kb in size; many of those segments contain protein-coding genes (Lupski, 2007; McCarroll, 2008; Redon *et al.*, 2006; Sebat *et al.*, 2004). However, the physiologic impact of common CNVs in health and in disease remain conjectural or largely unknown. Little knowledge is available about the details and consequences of each common CNV locus among human populations. To understand the effects of CNVs, it is essential to characterize the genetic compositions for each CNV locus and the patterns of variations among different subjects and populations, and to determine the associated qualitative and quantitative diversities of protein products created by CNV.

Complement C4 is an important component protein in the classical and mannose-binding lectin complement activation pathways, which are main effectors of the adaptive and innate immune responses against microbial infections, respectively (Walport, 2001; Yu *et al.*, 2003). Deposition of activated C4 on immune complexes opsonizes them for immunoclearance through binding to complement receptor CR1 on the red blood cells and disposal through phagocytosis by Kupffer cells in the liver (Cornacoff *et al.*, 1983; Yu *et al.*, 2007), and facilitates B-cell activation in germinal centers of lymph nodes (Carroll, 2004). Polymorphic plasma protein variants with acidic C4A that migrated faster, and basic variants that migrated slower, were observed decades ago using non-denaturing, high voltage agarose gel electrophoresis (Awdeh *et al.*, 1979; Mauff *et al.*, 1983; Sim and Cross, 1986).

The discovery of gene copy-number variations for human complement component C4 took a relatively crooked route. The modes of inheritance were debated among the single co-dominant locus, polygenic loci, and two-locus C4A-C4B model (Awdeh and Alper, 1980; O'Neill *et al.*, 1978; Roos *et al.*, 1982; Teisberg *et al.*, 1976). The two-locus model was favored as it was supported by most (but not all) experimental data. When genomic clones for human *C4A* and *C4B* genes in the MHC class III region were isolated, characterized and sequenced (Carroll *et al.*, 1984; Yu *et al.*, 1986; Yu, 1991), it was soon found that cytochrome P450 steroid 21-hydroxylase genes *CYP21A* (also known as *CYP21A1P*) and *CYP21B* (also known as *CYP21A2*) were present approximately 3.0 kb downstream of *C4A* and *C4B*, respectively (Carroll *et al.*, 1985a; Rodrigues *et al.*, 1987; White *et al.*, 1984; Yu, 1991). Further work revealed that the serine/threonine kinase gene *RPI* (also known as *STK19*) is located 613 bp upstream of the first *C4* gene locus (Shen *et al.*, 1994), and the gene for extracellular matrix protein tenascin *TNXB*, while organized in the opposite transcriptional orientation, overlaps to the 3' region of *CYP21B*. A gene fragment of 911-bp known as *RP2* that matches to the 3' region of *RPI*, and another gene fragment of 4.5 kb in size known as *TNXA* that corresponds to intron 32 to exon 45 of *TNXB*, are found in every duplicated *C4-CYP21* complex. Such

discrete four-gene duplication unit is termed the RCCX (*RP-C4-CYP21-TNX*) module (Shen *et al.*, 1994; Yang *et al.*, 1999; Yu *et al.*, 2003) (Figure 1).

Cumulative studies in the past fifteen years revealed that one to four RCCX modules are present in the central region of the MHC. Therefore, the copy-number of *C4* genes in a diploid genome varies from 2 to 8 (Chung *et al.*, 2002b; Chung *et al.*, 2002a; Wu *et al.*, 2007; Yang *et al.*, 2007). In each duplicated RCCX, the *C4* gene is usually functional and is either a long or a short gene. The long gene is created by the integration of an ancient endogenous retrovirus HERV-K(C4) into intron 9 of the *C4* gene (Chu *et al.*, 1995; Dangel *et al.*, 1994) (Figure 1, panel B). Each *C4* gene can code for a polymorphic C4A or a polymorphic C4B protein. The *CYP21* gene in the duplicated RCCX module is generally a mutant gene *CYP21A* (also known as *CYP21A1P*) that has acquired three deleterious mutations in exons 3, 7 and 8, plus multiple point mutations in the coding and non-coding regions (Higashi *et al.*, 1986). In comparison with the intact gene *TNXB*, nine single nucleotide polymorphisms and a 120-bp deletion are present in the 4.5 kb *TNXA* gene segment that is present in each duplicated RCCX module. The location of the 120-bp deletion in *TNXA* corresponds to the sequence between intron 36 and exon 36 in *TNXB* (Gitelman *et al.*, 1992; Yang *et al.*, 1999).

In a study of healthy European Americans, the frequency of monomodular RCCX haplotypes with single and intact genes for *RP1*, *C4*, *CYP21B* and *TNXB* (mono-L and mono-S) is 0.149; bimodular RCCX with organization of *RP1-C4-[CYP21A-TNXA-RP2-C4]-CYP21B-TNXB* has a frequency of 0.769, while trimodular haplotype with *RP1-C4-[CYP21A-TNXA-RP2-C4]₂-CYP21B-TNXB* has a frequency of 0.088. The copy-number of *C4* genes in a diploid among different human subjects varies between 2 to 6 in the European American population, with 60.7% of the healthy subjects having four copies of *C4* genes, 26.1% having three copies, and 9.8% having five copies. A similar pattern of *C4* gene copy-number variation was also observed in the Hungarian population (Yang *et al.*, 2003).

We postulate that copy-number variation of *C4* genes creates a diversity of intrinsic strengths in the immune effector system. Such inherent structural variations probably empower individuals in a population the capability to respond to a variety of environmental challenges effectively, although they could also predispose some subjects to autoimmune or genetic diseases. Therefore, it is of interest to note that among the patients with systemic lupus erythematosus (SLE) of European ancestry, the copy-number of total *C4* or *C4A* genes are significantly reduced. Remarkably, 32.9% of the SLE patients have a homozygous or heterozygous deficiency of *C4A* (compared to 18.9% in matched healthy controls; $p=0.00014$) (Yang *et al.*, 2007). The high prevalence *C4A* deficiency was caused by increased frequency of a monomodular RCCX haplotype with a single *C4B* gene (mono-S; SLE patient: 0.169, healthy controls: 0.106).

Concurrent with *C4* gene CNV is the CNV for *CYP21* gene that codes for the steroid 21-hydroxylase. The 21-hydroxylase is an essential enzyme for the biosynthesis of cortisols that is important for blood sugar homeostasis and responses to stress, and aldosterone that is essential for the regulation of electrolytes by the kidneys and the maintenance of blood pressure. A wide spectrum of defects ranging from partial impairment of enzymatic activities to complete deficiencies of the steroid 21-hydroxylase that contribute to the mild non-classical congenital adrenal hyperplasia (CAH) with hyperandrogenisation, the severe classical CAH with salt-losing disease that can be life-threatening, and/or the simple virilizing phenotypes (Goncalves *et al.*, 2007). Complete deficiency of 21-hydroxylase can be caused by the absence of the *CYP21B* gene in monomodular RCCX haplotypes containing a single *CYP21A* pseudogene, or through genetic recombination or gene conversion-like events that resulted in the presence of two *CYP21A* pseudogenes in bimodular RCCX haplotypes, or the acquisition of deleterious mutations to *CYP21B* gene (Blanchong *et al.*, 2000; Chung *et al.*, 2002a; Yang *et al.*, 1999).

The incidence of classical CAH varies widely with 1 in 13,000–15,000 live births in the US, 1 in 2,575 in India (Rama Devi and Naushad, 2004) to 1 in 280 in Yupik Eskimos in Alaska (Pang *et al.*, 1988; Wilson *et al.*, 2007). Acquisitions of missense mutations or SNPs from *CYP21A* to *CYP21B* reduce the 21-hydroxylase enzymatic activities and probably contribute to non-classical CAH phenotypes, which has a prevalence of 1% in the population of New York City, and remarkably, 1 in 27 among Ashkenazi Jews (New, 2006).

We undertake a series of studies to determine the genotypic and phenotypic diversities created by common gene CNVs in different human populations in order to understand the roles of CNVs in complex and genetic diseases. Here we report the great inter-individual polymorphisms of the RCCX modules and *C4* gene copy-numbers in one of the largest human ethnic groups, the Asian-Indians.

2. Materials and Methods

Study Populations

The healthy Asian-Indian American (AIA) group comprised 168 subjects who originated from India, Pakistan, Bangladesh, and Sri Lanka. These subjects were mainly recruited during the Annual Indian Festivals in central Ohio, plus friends and colleagues of the Nationwide Children's Hospital (NCH) in Columbus, Ohio. Informed consent from each participant was obtained according to approved protocol by the Institutional Human Subject Review Board, NCH. From each individual, 10 ml of peripheral blood was collected in EDTA (purple-top) tubes. The mean age ($\pm$ SD) of the AIA group was 35.9 ± 14.1 years. The comparison group consisted of 440 healthy European Americans from central Ohio with an average age 36.8 ± 11.5 years. The studied subjects and their first-degree relatives did not have a history of autoimmune disease (Yang *et al.*, 2007).

Genotyping of RCCX modules

EDTA-blood samples were processed according to standard protocols to harvest plasma and genomic DNA (Chung *et al.*, 2005). The copy-number variations and genotypes of the RCCX modules were determined by two independent approaches. First, *TaqI* genomic Southern blot analyses were applied to determine the constituents and variations of *RP-C4-CYP21-TNX* (RCCX) modules. Three different hybridization probes were used. The first probe detected the presence and relative number of an *RP1* gene linked to a long *C4* gene (7.0 kb), an *RP1* gene linked to a short *C4* gene (6.4 kb), an *RP2* gene fragment linked to a long *C4* gene (6.0 kb), or an *RP2* gene fragment linked to a short *C4* gene (5.4 kb). The second probe elucidated the presence and relative number of cytochrome-P450 21-hydroxylase *CYP21B* (3.7 kb; also known as *CYP21A2*), and its non-functional mutant gene *CYP21A* that is characterized by three deleterious indels and point mutations (3.2 kb; also known as *CYP21A1P*). The third probe elucidated the presence and relative number of extracellular matrix protein tenascin *TNXB* (2.5 kb), and the truncated gene fragment *TNXA* that corresponds to the 3' region of *TNX* (2.4 kb). *TNXA* is characterized by a 120-bp deletion that corresponds to the region between exon 36 and intron 36 of *TNXB*.

Second, large molecular weight genomic DNA from leukocytes in agarose plugs were prepared, digested with *PmeI* restriction enzyme, resolved by pulsed field gel electrophoresis under conditions that maximize resolution of DNA fragments between 20 and 350 kb in size, and processed after Southern blot analysis. The breakpoints RCCX modular duplications are present at the 3' region of the *RP1* gene, and at intron 32 of the *TNXB* gene. Two *PmeI* restriction sites are located outside the duplicated regions, by which one is present in complement factor *B* gene, and the other at the 5' region of the *TNXB* gene. Therefore, the size of the *PmeI* fragment (s) represent the number of RCCX modules in haplotypes. A monomodular RCCX haplotype

with a short *C4* gene has a *PmeI* fragment of 107 kb in size, while that with a long *C4* gene has a fragment size of 113 kb. An addition of one RCCX module increases the fragment size by 26.3 kb if it contains a short *C4* gene, and by 32.7 kb if it contains a long *C4* gene. Thus, bimodular haplotype with two long *C4* genes (LL) has a *PmeI* fragment size of 148 kb, while that with one long and one short *C4* gene (LS), 139 kb. A trimodular haplotype with three long *C4* genes (LLL) has a *PmeI* fragment of 178 kb in size, and that with one long and two short *C4* genes, 172 kb. Thus, *PmeI*-PFGE yields information on the total number of RCCX modules or copy-numbers of *C4* genes, and elucidates the haplotypes of RCCX modules and *C4* gene copy-numbers in a subject (Chung *et al.*, 2002b; Chung *et al.*, 2002a).

Third, human subjects with high copy-numbers or with ambiguous copy-numbers were further interrogated by TaqMan-based real-time PCR assays that elucidated the copy-number of RCCX modules, the copy-number of *C4A* genes, and the copy-number of *C4B* genes (Wu *et al.*, 2007).

Genotyping of C4A and C4B

The C4 protein isotypes are defined by amino acid sequence PCPVLD 1101–1106 for C4A, and LSPVIH 1101–1106 for C4B (Yu *et al.*, 1986). These changes are the results of five single nucleotide polymorphisms within a sequence of 20 nucleotides in exon 26. These polymorphic nucleotides for C4A can be recognized by restriction enzyme *PshAI*. Using *PshAI*-*PvuII* digested genomic DNA for Southern blot analyses using a C4d-specific probe, the *C4A* is represented by a 1.7 kb restriction fragment, and *C4B* by a 2.2 kb restriction fragment (Chung *et al.*, 2002b).

Allotyping of C4A and C4B protein variants

EDTA-plasma samples were digested with neuraminidase and carboxyl peptidase B to remove heterogeneities in glycosylations (Awdeh and Alper, 1980), and incomplete processing of the carboxyl terminals for the alpha and beta chains in C4 proteins (Sim and Cross, 1986), respectively. The polymorphic variants of C4A and C4B proteins are resolved by high voltage agarose gel electrophoresis that is non-denaturing and separate protein allotypes through their gross difference in electric charges. The C4 proteins were immunofixed with goat anti-C4 sera and the agarose gel was washed to remove diffusible proteins, and stained with SimplyBlue Safestain (Invitrogen), as described previously (Awdeh and Alper, 1980; Chung *et al.*, 2005; Sim and Cross, 1986). The relative band intensities for C4A and C4B allotypes were scanned by densitometry and quantified by ImageQuant software. C4A and C4B protein concentrations were calculated from total C4 concentrations determined by radial immunodiffusion.

Radial immunodiffusion of plasma C4

Total C4 plasma protein concentrations were determined by single radial immunodiffusion assays using kits from the Binding Site (U.K).

Determination of TNXA-XB recombinants by TaqI and PshAI RFLPs

Except for cases with DNA recombinations, all human subjects have one copy of *TNXB* on each copy chromosome 6, while that of *TNXA* varies from 0 to 1, 2 or 3, dependent on the number of RCCX duplication modules present in a haplotype. A recombination at the 3' regions can transfer the intact 120-bp sequence from *TNXB* to *TNXA*. Using a specific 500-bp probe spanning exons 35–37, the presence of such 5'-*TNXA*-*XB*-3' recombinant increases the relative band intensity of the 2.5 kb *TaqI* fragment. The presence of such recombinant can also be detected by a specific, 3.45 kb *PshAI* restriction fragment in genomic Southern blot analysis.

Statistical Analyses

Copy-number variants were compared using Chi-Square analysis, and multiple continuous variables using One-Way-Analysis-of-Variance (ANOVA). Numeric data from individual groups were also compared using Student's t-test. Statistical programs and graphing software include Excel, JMP 7.0, and Graph Pad Prism 5.0.

3. Results

Determination of RCCX modular variations, total C4, C4A and C4B gene copy-numbers

We determined the copy-number variations RCCX constituents *RP1* and *R2*, long *C4* and short *C4*, *CYP21B* and *CYP21A*, and *TNXB* and *TNXA* in the AIA-samples by *TaqI* RFLP. The relative copy-numbers of *C4A* and *C4B* genes were determined by *PshAI-PvuII* RFLP and their respective numbers calculated from total *C4*. High copy-numbers and ambiguous results were further clarified by three different real-time PCR amplicons, which independently determined the copy-numbers of RCCX modules, *C4A* and *C4B* genes. The protein polymorphisms of *C4A* and *C4B* allotypes were determined by immunofixation and their relative band intensities quantified by ImageQuant software. The plasma *C4* protein concentrations from each blood sample were determined by radial immunodiffusion. The protein concentrations of *C4A* and *C4B* were calculated from total *C4*.

Figure 2 panel A illustrates the complex patterns of *TaqI* RFLP present in eleven Asian Indian subjects. AIA-48 was characterized by the presence of has three RCCX modules with LL and L haplotypes (LL/L), as the 7.0 kb *RP1-C4L TaqI* fragment was twice as intense as the 6.0 kb *RP2-C4L TaqI* fragment. Such interpretation is corroborated by the presence of two *CYP21B* and one *CYP21A*, as the 3.7 kb *TaqI* fragment for *CYP21B* was twice as intense as the 3.2 kb *CYP21A* fragment; and the presence of two *TNXB* and one *TNXA*, as the 2.5 kb *TNXB* fragment was twice as intense as the 2.4 kb *TNXA* fragment. *PshAI-PvuII* RFLP (pane B) showed that the 2.2 kb fragment specific for *C4B* in AIA-48 was twice as intense as the 1.7 kb fragment specific for *C4A*. *C4* protein allotyping experiments (panel C) revealed that AIA-48 has C4A3 and C4B1, the band intensities for C4B1 was approximately twice as intense as that of C4A3. Radial immunodiffusion assay revealed that the plasma *C4* protein concentration of AIA-48 was 20.2 mg/dL. ImageQuant of the C4A3 and C4B1 protein bands in allotyping gel showed a ratio of 0.391 to 0.609 and thus the *C4A* and *C4B* protein concentrations were calculated to be 7.9 and 12.3 mg/dL, respectively.

Interpretations for the *C4* genotypes and phenotypes for the eleven subjects shown in Figure 2 are detailed in Supplementary-Table 1. Remarkably, AIA-61 appeared to possess seven copies of *C4* genes, by which 6 copies were *C4A* and one copy was *C4B*. The *C4B* protein of AIA-61 in the allotyping gel was almost invisible, suggesting a probable mutation of the *C4B* gene. AIA-58, AIA-62 and AIA-64 each had five copies of *C4* genes but their *C4A* to *C4B* ratio were 3:2, 5:0 and 2:3, respectively. The absence of *C4B* gene and *C4B* protein in AIA-62 was conspicuous in the *PshAI-PvuII* RFLP (panel B), *C4* allotyping (panel C), and real-time PCR (Supplementary-Table 1).

Six subjects had four copies of *C4* genes, AIA-53, -56, -57, -65, -66, and -67, who had different combinations of RCCX modules and *C4A* and *C4B* gene copy-numbers per diploid genome (gene dosage) producing differential quantities of *C4A* and *C4B* proteins. AIA-53, -56 and -57 had RCCX modules LS/LS, LL/LS and LL/LL, respectively. Each of these three subjects had two *C4A* and two *C4B*, coding for C4A3 and C4B1. In AIA-57, the *C4A* and *C4B* genes all belonged to the long form and the protein band intensities for C4A3 and C4B1 in allotyping gel were almost identical. In AIA-53, there were two long genes and two short genes. The *C4B* protein band intensity was 32.5% higher than that of *C4A*, implying that human subjects with

equal number of long genes and short genes, and equal number of *C4A* and *C4B* genes, produced larger quantities of *C4B* proteins than *C4A* proteins. In AIA-56, there were three long genes and one short gene, the *C4B* protein was 22.2% higher than that of *C4A*. This infers a dosage effect of short *C4* genes on the relative higher expression of *C4B* than *C4A*.

In AIA-65, -66 and -67, the RCCX haplotypes were LL/LS, LL/LL and LL/LS respectively, and their corresponding *C4A* to *C4B* gene ratios were 3:1, 3:1 and 2:2. Polymorphic protein variant C4A2 was present in AIA-65 and AIA-66, and C4A6 was present in AIA-67 (and AIA-58), in addition to the common C4A3. This phenomenon revealed that among AIA-subjects with same *C4* gene copy-number, the *C4* genes could either long or short, each *C4* gene could code for a *C4A* or a *C4B* protein, and the corresponding *C4A* or *C4B* protein could be polymorphic.

Demonstration of RCCX length Variants by *PmeI*-PFGE

To confirm the high copy gene numbers of *C4* and elucidate their haplotypes, we performed *PmeI*-PFGE on the AIA-samples. Figure 3 illustrates results of such experiments in eight AIA-subjects, together with three controls which were previously demonstrated to be homozygous trimodular LLL/LLL (c008, lane 5), monomodular L/L (EW, lane 10) and monomodular S/S (c071, lane 11) (Chung *et al.*, 2002a).

Homozygous trimodular LSL (or LLS, the order of the short gene in the second and third module not defined) was found in AIA-45 (lane 6), bimodular LL in AIA-05 (lane 8) and bimodular LS in AIA-39 (lane 9). The others contained heterozygous RCCX length variants as each was marked by the presence of two *PmeI* fragments corresponding to distinct RCCX haplotypes.

Results by *TaqI* RFLP revealed that AIA-61 had five long *C4* genes and two short *C4* genes. *PmeI*-PFGE showed that these seven genes were organized with quadrimodular L(LLS) in one haplotype and trimodular LSL in the other haplotype (lane 2). *TaqI* RFLPs showed that AIA-27 and AIA-19 each had 6 copies of *C4* genes in a diploid genome, with four long genes and two short genes for AIA-27, and two long genes and four short genes for AIA-19. *PmeI* PFGE revealed that the *C4* genes in AIA-27 were organized in quadrimodular L(LLS) and bimodular LS (lane 3), while those of AIA-19 were organized in quadrimodular LSSS and bimodular LS (lane 4).

TaqI RFLPs showed that AIA-62 and AIA-38 each had five *C4* genes. Four long genes and one short gene were present in the former; two long and three short *C4* genes in the latter. *PmeI*-PFGE showed that AIA-62 had trimodular LSL and bimodular LL (lane 1), and AIA-38 had trimodular LSS and bimodular LS (lane 7).

Copy-number variations of human complement *C4* in AIA

CNV of *C4* genes—The copy-number of *C4* genes in a diploid genome ranged from 2 to 7 among the AIA-subjects. The distribution and frequencies of *C4* GCN and RCCX haplotypes are presented in Table 1. The median *C4* GCN was 4, whose frequency was 0.530. The overall distribution (Figure 4) biased toward higher copy-numbers with double the frequency of 5 genes (0.286) versus 3 genes (0.137). Additionally, there were a greater amount of individuals with 6 and 7 genes (0.042) versus 2 genes (0.006). The propensity toward higher copy-number of total *C4* is reflected in the gene copy index (GCI) of 4.23 ± 0.77 , which is greater than the median 4 in healthy Asian Indians.

CNVs of *C4A* and *C4B*—*C4A* and *C4B* gene copy-numbers were determined in 166 AIA-subjects. A total of 391 *C4A* genes (frequency: 0.556) and 312 *C4B* genes (frequency: 0.444) were found.

The distribution of *C4A* genes resembled to that of total *C4*, with higher copy-numbers appearing more frequently than lower copy-numbers. *C4A* displayed a large distribution ranging from 0 to 6 copies per diploid genome, with slightly greater than half of the population centered at the median of 2 genes (0.548). Less than one-tenth of the population (0.0783) contained one copy of *C4A*, with a single individual lacking *C4A* completely. The remaining individuals comprised 3 or more *C4A* genes. Of these, 3 copies appeared most frequently with a frequency of 0.307, and 4, 5 or 6 copies occupied a combined frequency of 0.06. The GCI of *C4A* is 2.36 ± 0.79 .

C4B GCN distribution contrasted that of *C4A* with a predilection toward lower copy-numbers. *C4B* copy-number ranged from 0–4 copies with the majority containing 2 copies (0.560). Approximately one-fourth of AIA carried a single copy of *C4B* (0.259). Three AIA-subjects (0.018) lacked *C4B* genes and proteins entirely. The lower frequency of individuals with either 3 or 4 copies of *C4B* (0.151 and 0.012, respectively) contributed to a GCI of *C4B* slightly less than the median (1.87 ± 0.72).

Copy-number variation of *C4L* and *C4S*

Long and short *C4* genes were elucidated in 165 Asian Indian subjects. A total of 699 *C4* genes were accounted for, of which 520 were long genes (frequency: 0.744) and 179 were short genes (frequency: 0.256). The GCN distribution of long *C4* genes in AIA ranged from 0 to 6 total copies and followed a pattern similar to that for both *C4* and *C4A*, skewed toward increased frequency of higher GCN. Nearly half the population (0.494) presented a median copy-number of 3 *C4L*. Higher GCN was evidenced by the 31.7% of AIA with *more* than 3 copies of long genes (4 copies, 0.274; 5 copies, 0.043). A smaller proportion carried 2 copies of *C4L* (0.177), and two individuals contained 0 or only 1 copy of long *C4* gene (0.012). The gene copy index of long *C4* is 3.152 ± 0.82 .

AIA-subjects frequently carried 0, 1 or 2 copies of short *C4* genes. The corresponding frequencies were 0.317, 0.354 and 0.268, respectively. The presence of both 3 and 4 copies of short *C4* extended the GCI over the median to 1.085 ± 0.94 (3 copies, 0.049; 4 copies, 0.012).

RCCX modular variations

All AIA-subjects contained at least two RCCX modules in a diploid genome, with the presence of at least a single module and up to four modules on each Chromosome 6. Monomodular L had a frequency of 0.071, while monomodular S had a frequency of 0.012. Bimodular LL was the most frequent haplotype with a frequency of 0.432, which was followed by bimodular LS with a frequency of 0.301. Trimodular RCCX were present in 17.3% of the haplotypes (frequencies: LLL, 0.039; LSL, 0.054; LSS, 0.080). Quadrimodular RCCX haplotypes were present in four subjects (two LLLS and two LSSS) with a combined frequency of 0.012.

Phenotypes of complement component proteins *C4A* and *C4B*

Polymorphic variants—*C4* protein allotypes were elucidated in 160 AIA-subjects based on results of immunofixation experiments of plasma *C4*, and verified by *C4A* and *C4B* genotyping through *TaqI* RFLP and *PshAI-PvuII* RFLP on each subject (Table 2). A total of 671 *C4* protein phenotypes were identified, by which 372 were *C4A* and 299 were *C4B*.

Among *C4A*, *C4A3* was the most common allotype, which had a frequency of 0.839. Minor common variants of *C4A* included *C4A2* and *C4A6*, *each* of which had a frequency close to

0.06. Rare minor variants included C4A4 (frequency: 0.013) and C4A1 (frequency: 0.022). Mutant *C4A* genes without protein products were observed in two AIA-subjects.

Among C4B, C4B1 was the most common allotype that had a frequency of 0.866. C4B2 was the main C4B minor variant with a frequency of 0.094. Other detectable minor variants included C4B92, C4B5, C4B3 and C4B96. Mutant *C4B* gene with undetectable C4B protein product was suspected in one AIA-subject.

The most common combination of C4A and C4B proteins present in a subject is C4A3 and C4B1, irrespective of their corresponding gene dosages. Such combination has a frequency of 56.0% in AIA. Non C4A3-C4B1 combinations have a combined frequency of 44%.

Protein concentrations of total C4, C4A, and C4B—C4 plasma protein concentrations in healthy Asian Indians were determined by radial immunodiffusion experiments. Total C4 protein concentrations ranged from 11.6 to 72.4 mg/dL. Mean C4 protein levels measured 36.0 ± 10.4 mg/dL, with a similar median value of 35.4 mg/dL. The larger constituent of C4 protein, C4A, ranged from 0 (no *C4A* genes) to 40.1 mg/dL with a mean of 19.1 ± 6.1 mg/dL. A mean of 16.8 ± 7.7 mg/dL for C4B was determined. Mirroring C4A protein range, levels of C4B ranged from 0 to 39.3 mg/dL.

The effect of *C4* gene copy-number on protein concentration was accessed by comparing the protein levels among the various copy-number groups or RCCX haplotype present (Table 2). Figure 5 demonstrates increased copy-number of total *C4*, *C4A* or *C4B* equated with increased corresponding protein levels. In Figure 5A, individuals with 2–3 copies of *C4* produced a mean C4 protein level of 26.2 ± 7.0 mg/dL, which was less than those with 4 copies of *C4* (34.0 ± 7.8 mg/dL). Likewise subjects with five *C4* genes had mean plasma C4 protein concentration increased to 42.6 ± 10.1 mg/dL. Additional copies of *C4* also demonstrated increased effect (6–7 copies: 49.0 ± 11.1). As shown by analyses of variance (ANOVA), the differences in mean plasma protein concentrations from among the gene copy-number groups are highly significant ($p=6.6 \times 10^{-14}$).

As expected, the copy-number of *C4A* genes positively influenced the amount of C4A protein being produced (Figure 4B). A mean protein level of 13.4 ± 4.2 mg/dL was represented in individuals with a single copy of *C4A*, which was significantly reduced compared to individuals with 2 copies of *C4A*, producing an average of 17.2 ± 4.1 mg/dL. Those with 3 and then 4–6 copies of *C4A* genes also demonstrated significant increases (3: 23.2 ± 5.5 mg/dL, 4–6: 27.3 ± 6.3 mg/dL). ANOVA of plasma C4A protein concentrations with *C4A* gene copy-numbers yielded a p-value of 2.5×10^{-16} .

Similarly, the copy-numbers of *C4B* genes strongly affect the concentrations of plasma C4B proteins. AIA-subjects with single copy of *C4B* genes had a mean C4B concentration of 9.7 ± 3.0 mg/dL. Those with two copies of *C4B*, 17.6 ± 4.7 mg/dL; and ≥ 3 copies of *C4B*, 27.3 ± 7.0 mg/dL. ANOVA of plasma C4B protein concentrations with *C4B* gene copy-numbers yielded a p-value of 1.4×10^{-30} .

Effects of long and short C4 genes on total C4, C4A and C4B protein levels—To investigate the effect of long and short genes on the C4 protein expression levels among the AIA-subjects, we compared the total C4, C4A and C4B protein concentrations among subjects with equal gene copy-numbers of *C4A* and *C4B*. Fifty-five AIA-subjects were identified to have four copies of *C4* genes and their *C4A* and *C4B* gene copy-numbers were both at 2. The long and short gene haplotypes were LL/LL, LL/LS and LS/LS. The group with 4 copies of *C4L* (LL/LL) presented with the lowest protein concentrations of total C4 (30.8 ± 7.2 mg/dL). When replacing one long gene with a short gene as in haplotypes LL/LS, the mean C4 protein

levels increased to 34.6 ± 6.8 mg/dL. When 2 long genes were exchanged with 2 short genes, as in LS/LS, the mean C4 protein level raised to 41.3 ± 7.4 mg/dL, which was significantly different from that with LL/LL (t-test, $p=0.0003$) (Table 2).

We further asked whether the positive effects of short *C4* genes on C4 protein expression levels affected C4A, C4B or both. Among the AIA-subjects with two copies of *C4A* and two copies of *C4B*, the presence of 0, 1 and 2 copies of short genes only had modest effects on the mean protein concentration of C4A, which increased from 15.9 ± 3.5 mg/dL to 17.0 ± 4.2 and 18.3 ± 3.2 mg/dL in each group, respectively. By contrast, the copy-numbers of short genes had drastic effects on the protein concentrations of C4B. In the 2 *C4A* + 2 *C4B* group, the mean concentrations of C4B protein with 0, 1, and 2 copies of short genes were 15.5 ± 3.7 , 17.3 ± 4.3 and 23.1 ± 5.5 mg/dL, respectively. Highly significant difference in C4B protein levels were observed between the LL/LL and the LS/LS group (t-test, $p=0.00006$) (Table 2).

Gene copy-number variations of CYP21 and TNX in the RCCX modules

The increase of *C4* gene copy-numbers concurs with neighboring genes *CYP21*, and gene fragments *TNXA* and *RP2*. In a bimodular RCCX, the duplicated *CYP21* is frequently the pseudogene *CYP21A* (also known as *CYP21A1P*). However, subsequent recombinations between *CYP21A* from a bimodular haplotype and *CYP21B* from a monomodular haplotype could convert the *CYP21A* in a bimodular haplotype to *CYP21B*, yielding the presence two functional *CYP21B* in one recombinant, and the presence of a *CYP21A* mutant gene and no *CYP21B* in the reciprocal recombinant. In trimodular and quadrimodular RCCX, the duplicated *CYP21* can be either *CYP21A* or *CYP21B* or both. Similarly, the *TNXA* gene segment that overlaps with *CYP21A* at the 3' regions usually has a 120-bp deletion. Recombinations between *TNXB* and *TNXA* at the 3' region can add the 120-bp sequence to the *TNXA* gene segment, creating an additional *TNXB*-like gene resembling *XB-S*. From the *TaqI* RFLP the presence of *TNXA+120* was marked by higher intensities of the 2.5 kb *TaqI* restriction fragment, as shown in AIA-56 and AIA-57 (Figure 2, panel A). We have further developed a definitive method to detect the presence of the *TNXA+120* recombinants by *PshAI* RFLP, by which a *TNXA-TNXB* recombinant is represented by the presence of a novel 3.5 kb *PshAI* restriction fragment.

The copy-number distributions of *CYP21* and *TNX* are shown Figure 6 and Table 4. The distribution patterns of *CYP21A* and *TNXA* were highly analogous, and largely resembled that of total *C4*.

It was found that 22 out of 166 AIA-subjects had three copies of *CYP21B* (carrier frequency: 0.131). Remarkably, 15 of these 22 subjects were also characterized by the concurrent presence of the *TNXA+120* (carrier frequency for *21B-21B* plus *TNXA+120*: 0.09; Figure 7). An example of such haplotype is present in AIA-56 shown in Figure 2. Bimodular haplotypes with *CYP21B-CYP21B* without the involvement of *TNXA+120* were found only in four AIA-subjects. Bimodular haplotypes with *TNXA+120* haplotypes but no concurrent presence of *CYP21B-CYP21B* were observed in 8 subjects (carrier frequency: 0.048), by which one homozygous case was shown in AIA-57 (lane 4, panel A, Figure 2; lane 5, panel C, Figure 7).

Discussions

We performed a study of the complement *C4* and RCCX copy-number variations in one of the world's largest ethnic groups, the Asian Indians. Ethnically, socially and culturally, India is renowned for her multiplicity and complexity, and harmonious coexistence of various entities. The current Indian gene pool include the Dravidians in the southern-most states and immigrants to the northern India such as the Aryans from Central Asia between 2000 and 1400 B.C., the Greeks in 400 B.C., Arabs in 800 A.D., the Turks, the Afghans and the Moghuls in 1500 A.D (Papiha, 1996). Here we show a parallel complexity and multiplicity on genotypic and

phenotypic diversities of complement *C4* and constituents of RCCX modules among the Asian Indian Americans (AIA).

Genotypic variations of total *C4*, *C4A* and *C4B*

One to four copies of *C4* genes or RCCX modules per MHC haplotype are frequently present. Slightly over half (53.0%) of the AIA each has four copies of *C4* genes per diploid genome, and the rest have *C4* gene copy-numbers spreading between 2 and 7. The distribution curve of *C4* gene copy-number is tilted towards the *high* gene copy-number side as the frequency of subjects with five copies (28.6%) was twice more than those with three copies (13.7%). The gene-copy index or the mean copy-number of total *C4* among the Asian Indians is 4.23 ± 0.77 , and those for *C4A* and *C4B* are 2.36 ± 0.79 and 1.87 ± 0.72 , respectively. The medium copy-number for *C4A* and *C4B* are both at 2. Similar to total *C4*, the distribution of *C4A* is heavily skewed towards the *high* copy-number side as subjects with three *C4A* genes (30.7%) outnumber those with one *C4A* gene (7.8%). On the other hand, the distribution of *C4B* gene copy-number is slightly tilted towards the *low* copy-number side, as 25.9% of AIA had one *C4B*, compared to 15.1% having three *C4B*.

The copy-number and size variations of *C4* genes create a repertoire of physical or length variants among the AIA-subjects. Ten different physical variants (monomodular L and S; bimodular LL and LS; trimodular LLL, LSL and LSS; quadrimodular LLLL, LSSS and LLLS) have been detected among the AIA-subjects by *PmeI*-PFGE and by *TaqI* RFLP with Southern blot analyses. Such RCCX length variants could generate misalignments between homologous chromosomes during meiosis, leading to unequal crossovers.

We had previously determined the copy-number variations of complement *C4* and its associated RCCX modules in healthy subjects of European ancestry (Table 1) (Blanchong *et al.*, 2000; Yang *et al.*, 2003; Yang *et al.*, 2007). In stark contrast to Asian-Indians, the distribution of *C4* gene copy-number in European Americans is skewed towards the side with *lower* copy-numbers instead (panel A, Figure 8). The GCI of total *C4* is *below* the medium, 3.85 ± 0.69 . The frequency of subjects with three *C4* genes is 2.7 times greater than that with five *C4* genes (three genes: 26.1%; five genes: 9.8%), a phenomenon that is opposite to that in AIA. Unlike the AIA-subjects that have *C4A* gene copy-numbers skewed towards the high end with increased frequencies of 3 to 6 copies, the distribution of *C4A* genes among European Americans is more balanced between the low and high copy-numbers, and follows to a sigmoidal curve more closely (panel C, Figure 8).

Comparing the RCCX modular haplotypes between Americans of Asian-Indian and European ancestries, we observed *lower* frequencies of monomodular structures and *higher* frequencies of trimodular and quadrimodular structures in AIA. Of particular interest is the monomodular-short (mono-S) haplotype with a single *C4B* gene and the *absence* of *C4A*, which is central to the European ancestral-haplotype with HLA *A1 B8 DR3* (AH8.1) and is strongly associated with increased risk of autoimmune diseases including SLE and type I diabetes mellitus (Awdeh *et al.*, 1983; Carroll *et al.*, 1985b; Dawkins *et al.*, 1999; Horton *et al.*, 2008; Stewart *et al.*, 2004; Yang *et al.*, 2007). Such mono-S haplotype has a frequency of 10.6% in healthy subjects (panel B, Figure 8), and 16.9% in SLE patients of European ancestry (Yang *et al.*, 2007). Remarkably, only three subjects with mono-S haplotypes (two heterozygous and one homozygous; haplotype frequency: 1.2%) were detected among our AIA-cohort ($\chi^2=29.2$, $p=6.6 \times 10^{-8}$). The prevalence of SLE in India was reported to be relatively low, 3.2 per 100,000 (Malaviya *et al.*, 1993), compared to an overall rate between 14.6 and 50.8 per 100,000 among US subjects (Rus *et al.*, 2007). The relatively low frequency of mono-S, or the high gene copy-number of *C4A*, probably reduces the risk of SLE among Asian-Indian subjects.

Quantitative and qualitative diversities of C4 proteins and implications

The *C4* gene copy-number variation leads to a large range of C4 plasma protein concentrations from 11.6 to 72.4 mg/dL among the healthy Asian-Indian subjects, and the presence of polymorphic variants of C4A and C4B proteins. The *C4* gene copy-numbers and gene size are important determining factors plasma C4 protein concentrations. Linear correlations between total *C4*, *C4A* and *C4B* gene copy-numbers with their corresponding plasma protein concentration were observed. Also, the presence of short *C4* genes significantly increased the plasma C4 protein concentrations, particularly C4B. In a logistic regression model of determinants for plasma C4 proteins, *C4* gene copy-number has an F-ratio of 13.8 ($p=2.9 \times 10^{-6}$), and copy-number of short *C4* genes has an F-ratio of 6.3 ($p=0.0005$). We reason that the quantitative and qualitative diversities of C4 would provide the flexibility among different subjects to react to environmental and microbial challenges. The relatively higher copy-number of *C4* genes among the Asian-Indians would infer a powerful effector arm of the innate and adaptive immune system.

Phenotyping experiments of plasma C4 proteins showed that C4A3 allotypes and C4B1 allotypes were the most common variants in each class of C4 proteins. C4A3 has a frequency of 0.839 among C4A; and C4B1, 0.866 among C4B. AIA-subjects with C4A3 plus C4B1 and no other C4 protein variants carry a frequency of 56.0%. In other words, 44% of AIA-subjects had other C4 variants in addition to C4A3 and C4B1, or did not have both C4A3 and C4B1. This is relevant because blood transfusion patients can make alloantibodies against mismatched donor C4 protein variants, which contributed to the presence of Chido and Rodgers blood groups (Giles *et al.*, 1988; Longster and Giles, 1976; Middleton and Crooksten, 1972; Robson *et al.*, 1989; Yu *et al.*, 1986; Yu *et al.*, 1988). The recent application of immunosuppression drugs has significantly increased the success of organ transplantations. However, a proportion of transplants are rejected because of humoral alloreactivity against grafts such as kidneys. Many rejected kidney transplants are characterized by depositions of C4d on peritubular capillaries (Bohmig *et al.*, 2008; Feucht, 2003; Feucht and Mihatsch, 2005; Ranjan *et al.*, 2008). C4d is an activation and degradation product for complement C4. While plasma C4 are mainly synthesized by the livers, the kidneys, adrenal and the thyroid glands, heart, small intestine, ovary and thymus also synthesize considerable quantities of C4 (Berger *et al.*, 2005; Seelen *et al.*, 1993; Yu *et al.*, 2003). There is a relatively high likelihood for a mismatch of C4A and/or C4B protein allotypes produced by the donor organ and the corresponding transplant recipient. Whether the *polymorphic* C4 protein variant(s) from the donor organ would elicit and/or aggravate alloimmune responses leading to a graft rejection by the recipient deserves detailed investigations.

Recombinations between RCCX length variants as a mechanism for genetic diseases

Our studies of RCCX modules also shed light on the genetics of cytochrome P450 21-hydroxylase CYP21 and extracellular matrix protein tenascin TNX. *TaqI* genomic RFLP revealed concurrent copy-number variations of pseudogene *CYP21A* and gene fragment *TNXA* with total *C4*. *CYP21A* and *TNXA* are present in MHC haplotypes with two or more RCCX modules. The copy-numbers of these two genetic elements both varied between 0 and 5 (Figure 6), and their distributions are highly analogous to that of total *C4* (Figure 4). It is striking to note that 13.1% of the Asian- Indian study population (22 subjects) had 3 copies of the functional *CYP21B* gene. Besides three subjects who had trimodular RCCX structures with *21A-21B-21B* configurations, the remaining 19 subjects were from bimodular RCCX haplotypes with *21B-21B* configurations. Remarkably, 15 of these bimodular *21B-21B* haplotypes also had an additional marker for the 120-bp addition to the *TNXA* gene segment. Such *CYP21B-TNXA+120* structures were likely generated through recombinations or gene conversion-like events between *TNXA* from a bimodular RCCX haplotype and *TNXB* from a monomodular RCCX haplotype because of misalignments between these RCCX length

variants during meiosis (Rupert *et al.*, 1999; Yang *et al.*, 1999; Yu *et al.*, 2000). Acquisition by *TNXA* of the 120-bp sequence between exon 36 and intron 36 from *TNXB* without the apparent involvement of *CYP21B* was found in an additional 8 MHC haplotypes. At this stage, the physiologic impact of higher steroid 21-hydroxylase activity due to high copy-numbers of *CYP21B* has not been investigated, nor has the presence of a novel protein coded by *TNXA*+120, which resembles *XB-S* (Kato *et al.*, 2008; Tee *et al.*, 1995). The reciprocal products for the unequal recombinations would be haplotypes with *CYP21A* pseudogene only, and/or *TNXB* gene that is missing the 120-bp sequence between exon 36-intron 36. We and others showed the presence of such recombinants in patients with congenital adrenal hyperplasia (Blanchong *et al.*, 2000; Yang *et al.*, 1999) and patients with Ehlers Danlos Syndrome (Burch *et al.*, 1997). The steroid 21-hydroxylase (CAH) is essential in the biosynthesis of the stress hormone hydrocortisone, and the salt-retaining hormone aldosterone. Deficiencies of 21-hydroxylase leads to CAH with a range of disease severity from salt-losing phenotype that is life-threatening, simple virilizing, ambiguous genitalia and male secondary sex characters among females (White and Speiser, 2000). It is of interest to point out that India has relatively high rate of CAH (Incidence: 1 in 2575 new-born). The high frequencies of the reciprocal recombination products, *CYP21B-CYP21B* and/or *TNXA*+120, plus the presence of many length variants with different long and short *C4* genes coding for C4A or C4B among AIA, are indicative of high rates of unequal recombinations between RCCX haplotypes.

In summary, our results demonstrate great diversities associated with gene copy-number variations of complement *C4*, steroid 21-hydroxylase *CYP21* and tenascin *TNX*. It also offers an explanation for the low prevalence of SLE but the high incidence of CAH among Asian Indians.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to express our sincere gratitude to the blood donors. We are indebted to Dr. Yaoling Shu for assistance. This work was supported by grants 1R01 AR050078 and 1R01 AR054459 from the NIAMS, NIAID, NIDDK, and the Office of the Director, the National Institutes of Health, USA, and by Lupus Foundation of America.

ABBREVIATIONS IN THIS MANUSCRIPT

AIA	Asian-Indian Americans
CAH	congenital adrenal hyperplasia
CNV	copy-number variation
L	monomodular RCCX with a single long <i>C4</i> gene
LL and LS	bimodular RCCX haplotypes with two long, or with one long and one short, <i>C4</i> genes, respectively
mono-S or S	monomodular RCCX with a single short <i>C4</i> gene

SLE

systemic lupus erythematosus

TNXA+120*TNXA-XB* recombinant in which the *TNXA* incorporated the 120-bp sequences from *TNXB***References**

- Awdeh ZL, Alper CA. Inherited structural polymorphism of the fourth component of human complement. *Proc Natl Acad Sci USA* 1980;77:3576–3580. [PubMed: 6932037]
- Awdeh ZL, Raum D, Alper CA. Genetic polymorphism of human complement C4 and detection of heterozygotes. *Nature* 1979;282:205–208. [PubMed: 492334]
- Awdeh ZL, Raum D, Yunis EJ, Alper CA. Extended HLA/complement allele haplotypes: Evidence for T/t-like complex in man. *Proc Natl Acad Sci USA* 1983;80:259. [PubMed: 6401863]
- Bank RA, Hettema EH, Muijs MA, Pals G, Arwert F, Boomsma DI, Pronk JC. Variation in gene copy number and polymorphism of the human salivary amylase isoenzyme system in Caucasians. *Hum Genet* 1992;89:213–222. [PubMed: 1375184]
- Berger SP, Roos A, Daha MR. Complement and the kidney: what the nephrologist needs to know in 2006? *Nephrol Dial Transplant* 2005;20(12):2613–2619. [PubMed: 16204271]
- Blanchong CA, Zhou B, Rupert KL, Chung EK, Jones KN, Sotos JF, Rennebohm RM, Yu CY. 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 MHC-associated disease. *J Exp Med* 2000;191:2183–2196. [PubMed: 10859342]
- Bohmig GA, Bartel G, Wahrmann M. Antibodies, isotypes and complement in allograft rejection. *Curr Opin Organ Transplant* 2008;13:411–418. [PubMed: 18685338]
- Burch GH, Gong Y, Liu W, Dettman R, Curry CJ, Smith L, Miller WL, Bristow J. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nature Genetics* 1997;17:104–108. [PubMed: 9288108]
- Carroll MC. A protective role for innate immunity in systemic lupus erythematosus. *Nat Rev Immunol* 2004;4:825–831. [PubMed: 15459673]
- Carroll MC, Campbell RD, Bentley DR, Porter RR. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature* 1984;307:237–241. [PubMed: 6559257]
- Carroll MC, Campbell RD, Porter RR. Mapping of steroid 21-hydroxylase genes adjacent to complement C4 genes in HLA, the major histocompatibility complex in man. *Proc Natl Acad Sci U S A* 1985a; 82:521–525. [PubMed: 3871526]
- Carroll MC, Palsdotir A, Belt KT, Porter RR. Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. *EMBO J* 1985b;4:2547–2552. [PubMed: 2996881]
- Chu X, Rittner C, Schneider PM. Length polymorphism of the human complement component C4 gene is due to an ancient retroviral integration. *Exp Clin Immunogenet* 1995;12:74–81. [PubMed: 7576718]
- Chung EK.; Wu, YL.; Yang, Y.; Zhou, B.; Yu, CY. Human complement components C4A and C4B genetic diversities: complex genotypes and phenotypes. In: Coligan, JE.; Bierer, BE.; Margulis, DH.; Shevach, EM.; Strober, W., editors. *Current Protocols in Immunology*. John Wiley & Sons, Inc.; Edison, NJ: 2005. 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. Genetic sophistication of human complement *C4A* and *C4B* and *RP-C4-CYP21-TNX* (RCCX) modules in the major histocompatibility complex (MHC). *Am J Hum Genet* 2002a;71:823–837. [PubMed: 12226794]
- Chung EK, Yang Y, Rupert KL, Jones KN, Rennebohm RM, Blanchong CA, Yu CY. Determining the one, two, three or four long and short loci of human complement *C4* in a major histocompatibility complex haplotype encoding for C4A or C4B proteins. *Am J Hum Genet* 2002b;71:810–822. [PubMed: 12224044]

- Cornacoff JB, Hebert LA, Smead WL, VanAman ME, Birmingham DJ, Waxman FJ. Primate erythrocyte-immune complex-clearing mechanism. *J Clin Invest* 1983;71:236–247. [PubMed: 6822663]
- Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu LC, Yu CY. The dichotomous size variation of human complement C4 gene is mediated by a novel family of endogenous retroviruses which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 1994;40:425–436. [PubMed: 7545960]
- Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev* 1999;167:275–304. [PubMed: 10319268]
- Feucht HE. Complement C4d in graft capillaries -- the missing link in the recognition of humoral alloreactivity. *Am J Transplant* 2003;3:646–652. [PubMed: 12780555]
- Feucht HE, Mihatsch MJ. Diagnostic value of C4d in renal biopsies. *Curr Opin Nephrol Hypertens* 2005;14:592–598. [PubMed: 16205481]
- Giles CM, Uring-Lambert B, Goetz J, Hauptmann G, Fielder AHL, Ollier W, Rittner C, Robson T. Antigenic determinants expressed by human C4 allotypes: a study of 325 families provides evidence for the structural antigenic model. *Immunogenetics* 1988;27:442–448. [PubMed: 2453461]
- Gitelman SE, Bristow J, Miller WL. Mechanism and consequences of the duplication of the human C4/P450c21/gene X locus. *Mol Cell Biol* 1992;12:2124–2134. [PubMed: 1373808]
- Goncalves J, Friaes A, Moura L. Congenital adrenal hyperplasia: focus on the molecular basis of 21-hydroxylase deficiency. *Expert Rev Mol Med* 2007;9:1–23. [PubMed: 17466088]
- Higashi Y, Yoshioka H, Yamane M, Gotoh O, Fujii-Kuriyama Y. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: A pseudogene and a genuine gene. *Proc Natl Acad Sci USA* 1986;83:2841–2845. [PubMed: 3486422]
- Horton R, Gibson R, Coghill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, de Jong PJ, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* 2008;60:1–18. [PubMed: 18193213]
- Kato A, Endo T, Abiko S, Ariga H, Matsumoto K. Induction of truncated form of tenascin-X (XB-S) through dissociation of HDAC1 from SP-1/HDAC1 complex in response to hypoxic conditions. *Exp Cell Res* 2008;314:2661–2673. [PubMed: 18588874]
- Longster G, Giles CM. A new specificity, anti-Rg^a, reacting with a red cell or serum antigen. *Vox Sang* 1976;30:175–180. [PubMed: 1251581]
- Lupski JR. Structural variation in the human genome. *N Engl J Med* 2007;356:1169–1171. [PubMed: 17360997]
- Malaviya AN, Singh RR, Singh YN, Kapoor SK, Kumar A. Prevalence of systemic lupus erythematosus in India. *Lupus* 1993;2:115–118. [PubMed: 8330032]
- Mauff G, Alper CA, Awdeh Z, Batchelor JR, Bertrams T, Bruun-Petersen G, Dawkins RL, Demant P, Edwards J, Grosee-Wilde H, Hauptmann G, Klouda P, Iam L, Mollenhauer E, Nerl C, Olaisen B, O'Neill G, Rittner C, Ross MH, Skanes V, Teisberg N, Wells L. Statement on the nomenclature of human C4 allotypes. *Immunobiology* 1983;164:184–191. [PubMed: 6852863]
- McCarroll SA. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 2008;17:R135–R142. [PubMed: 18852202]
- Middleton J, Crooksten M. Chido substance in plasma. *Vox Sang* 1972;23:256–261. [PubMed: 5081019]
- New MI. Extensive clinical experience: nonclassical 21-hydroxylase deficiency. *J Clin Endocrinol Metab* 2006;91:4205–4214. [PubMed: 16912124]
- O'Neill GJ, Yang SY, DuPont B. Two HLA-linked loci controlling the fourth component of human complement. *Proc Natl Acad Sci U S A* 1978;75:5165–5169. [PubMed: 368809]
- Pang SY, Wallace MA, Hofman L, Thuline HC, Dorche C, Lyon IC, Dobbins RH, Kling S, Fujieda K, Suwa S. Worldwide experience in newborn screening for classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Pediatrics* 1988;81:866–874. [PubMed: 3259306]
- Papiha SS. Genetic variation in India. *Human Biology* 1996;68:607–628.

- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 2007;39:1256–1260. [PubMed: 17828263]
- Rama Devi AR, Naushad SM. Newborn screening in India. *Indian J Pediatr* 2004;71:157–160. [PubMed: 15053381]
- Ranjan P, Nada R, Jha V, Sakhuja V, Joshi K. The role of C4d immunostaining in the evaluation of the causes of renal allograft dysfunction. *Nephrol Dial Transplant* 2008;23:1735–1741. [PubMed: 18065805]
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature* 2006;444(7118):444–454. [PubMed: 17122850]
- Robson T, Heard RNS, Giles CM. An epitope on the C4 β light (L) chains detected by human anti-Rg; its relationship with β -chain polymorphism and MHC associations. *Immunogenetics* 1989;30:344–349. [PubMed: 2478460]
- Rodrigues NR, Dunham I, Yu CY, Carroll MC, Porter RR, Campbell RD. Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. *EMBO J* 1987;6:1653–1661. [PubMed: 3038528]
- Roos MH, Mollenhauer E, Demant P, Rittner C. A molecular basis for the two locus model of human complement component C4. *Nature* 1982;298:854–855. [PubMed: 6180321]
- Rupert KL, Rennebohm RM, Yu CY. An unequal crossover between the RCCX modules of the human MHC leading to the presence of a *CYP21B* gene and a tenascin *TNXB/TNXA-RP2* recombinant between *C4A* and *C4B* genes in a patient with juvenile rheumatoid arthritis. *Exp Clin Immunogenet* 1999;16:81–97. [PubMed: 10343159]
- Rus, v; Maury, EE.; Hochberg, MC. The Epidemiology of Systemic Lupus Erythematosus. In: Wallace, DJ.; Hahn, BH., editors. *Dubois' Lupus Erythematosus*. Lippincott Williams & Wilkins; Philadelphia: 2007. p. 34-44.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science* 2004;305:525–528. [PubMed: 15273396]
- Seelen MAJ, Brooimans RA, van der Dorp FJ, van Es LA, Daha MR. Interferon- γ mediates stimulation of complement C4 biosynthesis in human proximal tubular epithelial cells. *Kidney Int* 1993;44:50–57. [PubMed: 8355466]
- Shen LM, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel A, Carroll MC, Zipf W, Yu CY. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and C4B genes in the HLA class III region: Molecular cloning, exon-intron structure, composite retroposon and breakpoint of gene duplication. *J Biol Chem* 1994;269:8466–8476. [PubMed: 8132574]
- Sim E, Cross S. Phenotyping of human complement component C4, a class III HLA antigen. *Biochem J* 1986;239:763–767. [PubMed: 3103606]
- Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, Forbes S, Halls K, Howson JM, Humphray SJ, Hunt S, Mungall AJ, Osoegawa K, Palmer S, Roberts AN, Rogers J, Sims S, Wang Y, Wilming LG, Elliott JF, de Jong PJ, Sawcer S, Todd JA, Trowsdale J, Beck S. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 2004;14:1176–1187. [PubMed: 15140828]
- Tee MK, Thomson AA, Bristow J, Miller WL. Sequences promoting the transcription of the human XA gene overlapping P450c21A correctly predict the presence of a novel, adrenal-specific, truncated form of tenascin-X. *Genomics* 1995;28:171–178. [PubMed: 8530023]
- Teisberg P, Akesson I, Olaisen B, Gedde-Dahl T Jr, Thorsby E. Genetic polymorphism of C4 in man and localisation of a structural C4 locus to the HLA gene complex of chromosome 6. *Nature* 1976;264:253–254. [PubMed: 1088823]

- Walport MJ. Complement-part I. *New Engl J Med* 2001;344:1058–1066. [PubMed: 11287977]
- White PC, New MI, DuPont B. HLA-linked congenital adrenal hyperplasia results from a defective gene encoding a cytochrome P-450 specific for steroid 21-hydroxylation. *Proc Natl Acad Sci USA* 1984;81:7505–7509. [PubMed: 6334310]
- White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev* 2000;21:245–291. [PubMed: 10857554]
- Wilson RC, Nimkarn S, Dumic M, Obeid J, Azar MR, Najmabadi H, Saffari F, New MI. Ethnic-specific distribution of mutations in 716 patients with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Mol Genet Metab* 2007;90:414–421. [PubMed: 17275379]
- Wu YL, Savelli SL, Yang Y, Zhou B, Rovin BH, Birmingham DJ, Nagaraja HN, Hebert LA, Yu CY. Sensitive and specific real-time PCR Assays to accurately determine copy-number variations (CNVs) of human complement *C4A*, *C4B*, *C4-Long*, *C4-Short* and *RCCX* modules: Elucidation of *C4* CNVs in 50 consanguineous subjects with defined HLA genotypes. *J Immunol* 2007;179:3012–3025. [PubMed: 17709516]
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride K, Higgins GC, Rennebohm RM, Rice RR, Hackshaw KV, Roubey RA, Grossman JM, Tsao BP, Birmingham DJ, Rovin BH, Hebert LA, Yu CY. Gene copy number variation and associated polymorphisms of complement component C4 in human systemic erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against European American SLE disease susceptibility. *Am J Hum Genet* 2007;80:1037–1054. [PubMed: 17503323]
- Yang Y, Chung EK, Zhou B, Blanchong CA, Yu CY, Füst G, Kovács M, Vatay A, Szalai C, Karádi I, Varga L. Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with *C4* gene size and polygenic variations, hemolytic activities and body mass index. *J Immunol* 2003;171:2734–2745. [PubMed: 12928427]
- Yang Z, Mendoza AR, Welch TR, Zipf WB, Yu CY. Modular variations of HLA class III genes for serine/threonine kinase RP, complement C4, steroid 21-hydroxylase CYP21 and tenascin TNX (*RCCX*): a mechanism for gene deletions and disease associations. *J Biol Chem* 1999;274:12147–12156. [PubMed: 10207042]
- Yu CY. The complete exon-intron structure of a human complement component C4A gene: DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene. *J Immunol* 1991;146:1057–1066. [PubMed: 1988494]
- Yu CY, Belt KT, Giles CM, Campbell RD, Porter RR. Structural basis of the polymorphism of human complement component C4A and C4B: gene size, reactivity and antigenicity. *EMBO J* 1986;5:2873–2881. [PubMed: 2431902]
- Yu CY, Campbell RD. 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* 1987;25:383–390. [PubMed: 2439447]
- Yu CY, Campbell RD, Porter RR. A structural model for the location of the Rodgers and the Chido antigenic determinants and their correlation with the human complement C4A/C4B isotypes. *Immunogenetics* 1988;27:399–405. [PubMed: 2453459]
- Yu CY, Chung EK, Yang Y, Blanchong CA, Jacobsen N, Saxena K, Yang Z, Miller W, Varga L, Fust G. Dancing with complement C4 and the RP-C4-CYP21-TNX (*RCCX*) modules of the major histocompatibility complex. *Progr Nucl Acid Res Mol Biol* 2003;75:217–292.
- Yu, CY.; Hauptmann, G.; Yang, Y.; Wu, YL.; Birmingham, DJ.; Rovin, BH.; Hebert, LA. Complement deficiencies in human systemic lupus erythematosus (SLE) and SLE nephritis: epidemiology and pathogenesis. In: Tsokos, GC., editor. *Systemic Lupus Erythematosus: A Companion to Rheumatology*. Elsevier; Philadelphia: 2007. p. 203–213.
- Yu CY, Yang Z, Blanchong CA, Miller W. The human and mouse MHC class III region: a parade of the centromeric segment with 21 genes. *Immunol Today* 2000;21:320–328. [PubMed: 10871871]

A. Copy number variation of *C4* genes and RCCX modules B. Exon-intron structure and size variation of *C4* genes C. *TaqI* RFLP and *PmeI* PFGE to determine variations of RCCX modules and *C4* genes **Figure 1.**

An introduction to the copy-number and size variations of RCCX modules and complement *C4*. *Panel A*, A molecular haplotype map showing the copy-number variations of *RP-C4-CYP21-TNX* (RCCX) in the class III region of the major histocompatibility complex (MHC) from complement *C2* to *CREB-BP* (Yu *et al.*, 2003). Horizontal arrows show the gene orientations. *Panel B*, exon-intron structure of long and short *C4* genes (Dangel *et al.*, 1994). A human endogenous retrovirus *HERV-K(C4)* is present in the ninth intron of the long *C4* gene. *Panel C*. *TaqI* RFLP and *PmeI*-PFGE to show the monomodular, bimodular and trimodular RCCX haplotypes. S, RCCX haplotype with short *C4* gene; L, RCCX haplotype with long *C4* gene. S/S, homozygous RCCX with short *C4* genes; L/L, homozygous RCCX

with long *C4* genes; LL/LL, homozygous bimodular haplotypes each with two long *C4* genes. LLL/LLL, homozygous trimodular haplotypes each with three long *C4* genes (Chung *et al.*, 2002b).

Figure 2.

Determination of RCCX modular variations, *C4A* and *C4B* genotypes and phenotypes in Asian Indian Americans. *Panel A*. Genomic *TaqI* RFLP to demonstrate the variation patterns of *RP*, *C4*, *CYP21* and *TNX*: *RP1-C4L* (7.0 kb), *RP2-C4L* (6.0 kb), *RP2-C4S* (5.4 kb), *CYP21B* (3.7 kb), *CYP21A* (3.2 kb), *TNXB* (2.5 kb), *TNXA* (2.4 kb), and *TNXA+120* (2.5 kb). The RCCX modules are shown on top of each lane. *Panel B*, Genomic *PshAI-PvuII* RFLP to determine the relative gene dosages of *C4A* and *C4B*. The ratio of *C4A* to *C4B* genes are marked below each lane. *Panel C*. Immunofixation experiment of EDTA-plasma with goat anti-human C4 antibodies to show the C4A and C4B protein polymorphisms based on gross differences in

electric charges of each protein. Interpretations of *TaqI* RFLPs, *PshAI-PvuII* RFLPs and immunofixation experiments are shown in Supplementary Table 1.

Figure 3.

Pulsed-field gel electrophoresis (PFGE) of *PmeI* digested genomic DNA and Southern blot analysis to show length variants of RCCX modules. Intact genomic DNA in agarose blocks from peripheral blood lymphocytes were digested with *PmeI* restriction enzyme, resolved by PFGE, Southern blotted and hybridized to a *C4* genomic probe. c008, EW and c071 are controls with known RCCX haplotypes. An addition of an RCCX module increased the *PmeI* fragment size by 32.7 kb if the module contains a long *C4* gene, or by 26.3 kb if the module contains a short *C4* gene. Assignments of RCCX haplotypes in each AIA samples were achieved after considering complementary data from *TaqI*-RFLP that revealed the numbers of long and short *C4* genes. M, (mid-range) molecular weight marker.

A. Gene copy-number variations of complement *C4* genes in diploid genomes

B. Mono-, bi-, tri- and quadri-modular variations of RCCX haplotypes

Figure 4.

Variations of complement *C4* gene copy-numbers and RCCX modules in Asian-Indian Americans. L and S, monomodular haplotypes with long and short *C4* genes, respectively; LL and LS, bimodular haplotypes with two long *C4* genes, and with one long and one short *C4* genes, respectively; LLL, LSL (or LLS), LSS, trimodular haplotypes with three long *C4* genes, two long and one short *C4* genes, and one long and two short *C4* genes, respectively; quad, quadrimodular haplotypes.

Figure 5.

Correlations of plasma C4 protein concentrations with *C4* gene copy-numbers and long and short *C4* genes. *Panels A, B and C*, scatter plots plasma C4 protein with total *C4* gene copy-numbers (A), C4A protein with *C4A* gene copy-numbers (B), and C4B proteins with *C4B* gene copy-numbers, respectively. Mean protein levels are shown by *horizontal lines*. *Panels D, E, F*, comparisons of total C4, C4A and C4B protein concentrations among AIA-subjects defined four *C4* genes and defined RCCX haplotypes. AIA-subjects with two *C4A* and two *C4B* genes were chosen for analyses. These subjects were stratified to three groups according to their RCCX haplotypes: LL/LL, LL/LS and LS/LS. The p-value in each panel was derived from analysis of variance (ANOVA); ns, not significant.

Figure 6.

Copy-number variations of *CYP21A* (*CYP21A1P*), *CYP21B* (*CYP21A2*), *TNXA* with 2.4 kb *TaqI* fragment (T2.4), *TNXB* plus recombinant *TNXA-XB* with a 2.5 kb *TaqI* fragment (*TNXA* +120 or T2.5) among healthy AIA-subjects.

A. A comparison of DNA and amino acid sequences at Exon 36 of *TNXB* and *TNXA*

B. Gene organizations and *Psh*AI sites at the breakpoint regions of the RCCX

C. *Psh*AI RFLP of *TNXX* 3' regions

Figure 7.

High frequency of *TNXA-XB* recombinants among AIA subjects. *Panel A*, A comparison of DNA sequences between *TNXB* and *TNXA* at exon 36 region. An 120-bp deletion is present in *TNXA* that spans the exon 36-intron 36 junction. *Panel B*, A comparison of *Psh*AI restriction maps at the intergenic regions for *CYP21A-TNXA* (left), and *CYP21B-TNXB* (right panel). The location of the 120-bp deletion in *TNXA* is shown by an *inverted* triangle. Using a DNA probe spanning the exon 36-intron 36 junction for *Psh*AI genomic Southern blot analysis, *TNXB* is represented by a 5.4 kb (P5.4) and a 6.1 kb (P6.1) *Psh*AI restriction fragments. An 8.8 kb (P8.8) *Psh*AI restriction fragment that corresponds to *TNXA-RP2* sequences will be present in regular bi-, tri- or quadrimodular RCCX haplotypes. For *TNXA-XB* recombinants, a 5.4 kb fragment that contains sequences for *CYP21B* and *TNXA+120*, plus a novel 3.5 kb (P3.5) fragment that corresponds to *TNXA+120* and *RP2* sequences are predicted. *Panel C*, *Psh*AI-RFLP to detect *TNXA-XB* recombinants in 16 healthy AIA-subjects. The number above each *lane* represents AIA-subject code. Six subjects (AIA-51, 52, 53, 58, 78, 80) had regular RCCX modular structures. The remaining ten subjects had *TNXA-XB* recombinations. Note that AIA-subjects 57, 99, 105 and 118 did not have the regular *CYP21A-TNXA* restriction fragments that can be caused by the presence of monomodular RCCX, and/or *TNXA+120*.

Figure 8.

A comparison of *C4* gene copy-numbers and RCCX modular variations between healthy Asian-Indian (AIA) and European American (EUA) subjects. Frequencies of AIA-subjects are shown in burgundy and EUA-subjects in gray. Note that AIA-subjects had high frequencies of 5 copies of total *C4* (panel A), 3 copies of *C4A* (panel C), 3 copies of *C4B* (panel D), when compared with the European counterparts. By contrast, AIA had low frequencies of subjects with 3 copies of total *C4* (panel A), 1 copy of *C4A* (panel C), 0 and 1 copy of long *C4* genes, and monomolecular RCCX haplotype with short *C4* gene (panel B), when compared with their European American counterparts.

Table 1
Variations of C4 gene copy-numbers in diploid genomes and RCCX haplotypes in healthy Americans of Asian Indian and European ancestries.

Asian-Indian American					European American	
Copy-number	N	Frequency	N	Frequency	Statistics	
Total C4 genes						
2	1	0.006	6	0.014	$\chi^2 = 40.9$; p= 2.8×10^{-8} t-test, p= 6.8×10^{-8}	
3	23	0.137	115	0.261		
4	89	0.530	267	0.607		
5	48	0.286	43	0.098		
6,7	7	0.042	9	0.020		
Total	168	4.23 ± 0.77	440	3.85 ± 0.69		
gene copy index						
C4A genes						
0	1	0.006	3	0.007	$\chi^2 = 14.4$; p=0.006 t-test, p=0.0003	
1	13	0.078	79	0.182		
2	87	0.548	243	0.559		
3	51	0.307	96	0.221		
4,5,6	10	0.060	14	0.033		
Total	166	2.36 ± 0.79	435	2.094 ± 0.758		
gene copy index						
C4B genes						
0	3	0.018	10	0.023	$\chi^2 = 14.8$; p=0.005 t-test, p=0.048	
1	43	0.259	116	0.267		
2	93	0.560	281	0.646		
3	25	0.151	27	0.062		
4	2	0.012	1	0.002		
Total	166	1.87 ± 0.72	435	1.75 ± 0.61		
gene copy index						
Long C4 genes (C4L)						
0,1	2	0.012	30	0.069		
2	32	0.192	108	0.246		

Asian-Indian American					European American		
Copy-number	N	Frequency	N	Frequency	Statistics		
3	81	0.485	166	0.378	$\chi^2 = 12.2$; p=0.016 t-test, p=0.029		
4	45	0.269	117	0.267			
5,6	7	0.042	18	0.041			
Total	167		439				
gene copy index		3.13± 0.83		2.97± 1.01			
Short C4 genes (C4S)							
0	51	0.305	151	0.343	$\chi^2 = 12.1$; p=0.017 t-test, p=0.012		
1	60	0.359	196	0.445			
2	46	0.275	82	0.186			
3	8	0.048	10	0.023			
4	2	0.012	1	0.002			
Total	167		440				
gene copy index		1.10 ±0.94		0.90 ± 0.80			
RCCX haplotypes							
Haplotypes	N	Frequency	N	Frequency	Statistics		
L	24	0.071	38	0.043	$\chi^2 = 79.2$; p=2.1 × 10 ⁻¹²		
S	4	0.012	93	0.106			
LL	145	0.432	443	0.503			
LS	101	0.301	235	0.267			
LLL	13	0.039	32	0.036			
LSL	18	0.054	10	0.011			
LSS	27	0.080	27	0.031			
Quadrimodular	4	0.012	1	0.001			
Total	336		879				

Quadrimodular haplotypes: LLLL, LLLS and LSSS.

Table 2
Quantitative variations of plasma C4, C4A and C4B with C4, C4A and C4B gene copy-numbers (GCN) and gene sizes.

Total C4			C4A			C4B		
GCN	n	[C4] mg/dL	GCN	n	[C4A] mg/dL	GCN	n	[C4B] mg/dL
2-3	23	26.2 ± 7.0	0	1	0	0	3	0
4	88	34.0 ± 7.8	1	11	13.4 ± 4.2	1	39	9.7 ± 3.0
5	48	42.6 ± 10.1	2	84	17.2 ± 4.1	2	75	17.6 ± 4.7
≥6	7	49.0 ± 11.1	3	41	23.2 ± 5.5	≥3	29	27.3 ± 7.0
Total	166		≥4	9	27.3 ± 6.3		146	
ANOVA		p=6.6×10 ⁻¹⁴			p=2.5×10 ⁻¹⁶			p=1.4×10 ⁻³⁰
Genotypes	n	[C4] mg/dL		n	[C4A] mg/dL		n	[C4B] mg/dL
1. LL / LL	22	30.8 ± 7.2		20	15.9 ± 3.5		20	15.5 ± 3.7
2. LL / LS	23	34.6 ± 6.8		22	17.0 ± 4.2		22	17.3 ± 4.3
3. LS / LS	10	41.3 ± 7.4		9	18.3 ± 3.2		9	23.1 ± 5.5
Total	55			51			51	
t-test, p	1 vs 2	0.08			0.35			0.17
	1 vs 3	0.0003			0.12			0.00006
	2 vs 3	0.014			0.4			0.0015

n, number of subjects.

Table 3
Frequencies of C4A and C4B protein allotypes among Asian-Indian Americans

C4A allotypes*	n	frequency among C4A	C4B allotypes**	n	frequency among C4B
A-mutation†	2	0.005	B-mutation†	1	0.003
A1	8	0.022	B1	259	0.866
A2	23	0.062	B2	28	0.094
A3	312	0.839	B3	2	0.007
A4	5	0.013	B5, B7	3	0.010
A5 or A6	22	0.059	B91, B92	5	0.017
Total	372		B96	1	0.003
				299	

* One subject with homozygous C4A deficiency because of absence of C4A genes.
** Three subjects with homozygous C4B deficiency because of absence of C4B genes.
† With suspected mutation of the corresponding C4A or C4B gene.
Number (frequency) of AIA subjects with both C4A3 and C4B1 allotypes and no other allotypes: 94 (56.0%).
Number (frequency) of AIA subjects with C4A and C4B allotypes other than C4A3 and C4B1: 74 (44%).

Table 4
Copy-number variations and recombinants of *CYP21* and *TNX* in RCCX.

Gene	Copy no.	N	Frequency
<i>CYP21A</i> (T3.2)			
	0	5	0.030
	1	30	0.181
	2	83	0.500
	3	42	0.253
	4	5	0.030
	5	1	0.006
		166	
<i>CYP21B</i> (T3.7)			
	2	145	0.863
	3	22	0.131
	4	1	0.006
		166	
<i>TNXA</i> (T2.4) *			
	0	8	0.048
	1	29	0.174
	2	76	0.455
	3	47	0.281
	4	6	0.036
	5	1	0.006
		167	
<i>TNXB</i> and <i>TNXA</i> (T2.5) **			
<i>TNXB</i>	2	146	0.874
<i>TNXA</i> + <i>I20</i> , hetero	3	19	0.114
<i>TNXA</i> + <i>I20</i> , homo	4	2	0.012
		167	
Recombinants		N	Haplotype frequency
<i>21B-21B</i> only (T3.7)		4	0.012
<i>21B-21B</i> with <i>TNXA</i> + <i>I20</i> (T3.7 and T2.5)		15	0.045
<i>TNXA</i> + <i>I20</i> only (T2.5)		8	0.024

* T2.4, *TNXA* gene fragment characterized by a 120-bp deletion and a 2.4 kb *TaqI* fragment

** T2.5, includes all *TNXB*, and *TNXA-XB* recombinants that incorporated the 120-bp sequence from *TNXB*.

Hetero, heterozygous; homo, homozygous.