Sensitive and Specific Real-Time Polymerase Chain Reaction Assays to Accurately Determine Copy Number Variations (CNVs) of Human Complement *C4A*, *C4B*, *C4-Long*, *C4-Short*, and RCCX Modules: Elucidation of *C4* CNVs in 50 Consanguineous Subjects with Defined HLA Genotypes[1]

Yee Ling Wu,*† Stephanie L. Savelli,*‡ Yan Yang,* Bi Zhou,* Brad H. Rovin,§ Daniel J. Birmingham,§ Haikady N. Nagaraja,¶ Lee A. Hebert,§ and C. Yung Yu[2]*†‡

Recent comparative genome hybridization studies revealed that hundreds to thousands of human genomic loci can have interindividual copy number variations (CNVs). One of such CNV loci in the HLA codes for the immune effector protein complement component C4. Sensitive, specific, and accurate assays to interrogate the *C4* CNV and its associated polymorphisms by using submicrogram quantities of genomic DNA are needed for high throughput epidemiologic studies of *C4* CNVs in autoimmune, infectious, and neurological diseases. Quantitative real-time PCR (qPCR) assays were developed using TaqMan chemistry and based on sequences specific for *C4A* and *C4B* genes, structural characteristics corresponding to the long and short forms of *C4* genes, and the breakpoint region of *RP-C4-CYP21-TNX* (RCCX) modular duplication. Assignments for gene copy numbers were achieved by relative standard curve methods using cloned *C4* genomic DNA covering 6 logs of DNA concentrations for calibrations. The accuracies of test results were cross-confirmed internally in each sample, as the sum of *C4A* plus *C4B* equals to the sum of *C4L* plus *C4S* or the total copy number of RCCX modules. These qPCR assays were applied to determine *C4* CNVs from samples of 50 consangu*i*neous subjects who were mostly homozygous in HLA genotypes. The results revealed eight HLA haplotypes with single *C4* genes in monomodular RCCX that are associated with multiple autoimmune and infectious diseases and 32 bimodular, 4 trimodular, and one quadrimodular RCCX. These *C4* qPCR assays are proven to be robust, sensitive, and reliable, as they have contributed to the elucidation of *C4* CNVs in >1000 human samples with autoimmune and neurological diseases. ***The Journal of Immunology,*** **2007, 179: 3012–3025.**

Recent advances in comparative genomic hybridization studies confirm earlier observations on the prevalence of copy number variations (CNVs)[3] as a source of inherent genetic diversity and reveal that >3000 loci in human genomes can have structural variations >1 kb in size (1–4). Most of these CNV loci are engaged in gene-environment interaction, especially in immunologic and sensory functions. Through our extensive molecular genetic studies of human complement C4 and *RP-C4-CYP21-TNX* (RCCX) modules and from recently published comparative genomic hybridization results, it appears that the occurrence of CNVs are often discretely segmental and can involve multiple neighboring genes (5–8). Although the duplicated sequences among different segments are highly homologous with 95–99.9% sequence identities, there are often secondary polymorphic sequences that can lead to variations in protein structures or functions or patterns of gene expression. Therefore, characterization of CNVs and their associated polymorphisms are important, as they may provide one of the key missing links for understanding the genetic basis of quantitative traits and the different susceptibilities to autoimmune and infectious diseases.

Technical challenge exists and hinders rapid progress in studies of inherent CNVs in human diseases. Unlike somatic gene amplifications in oncogenesis or gene expression studies that involve a change in the copy numbers of target sequences by more than one order of magnitude, the inherent gene copy difference among different individuals is usually within the range of 1–10 copies in a diploid genome. It is essential to be able to accurately distinguish the subtle and yet discrete differences such as one, two, three, or four copies of a target gene among different subjects. There is also a need to distinguish and determine the copy numbers of polymorphic variants of target genes that share a high degree of sequence identities. To date, the most definitive method to elucidate CNVs in the range of 5–500 kb are probably: 1) long range mapping using genomic DNA digested by rare restriction enzyme cutters

*Center for Molecular and Human Genetics, Columbus Children's Research Institute, Columbus, OH 43205; and †Integrated Biomedical Science Graduate Program, ‡Department of Pediatrics, §Department of Internal Medicine, and ¶Department of Statistics, Ohio State University, Columbus, OH 43210

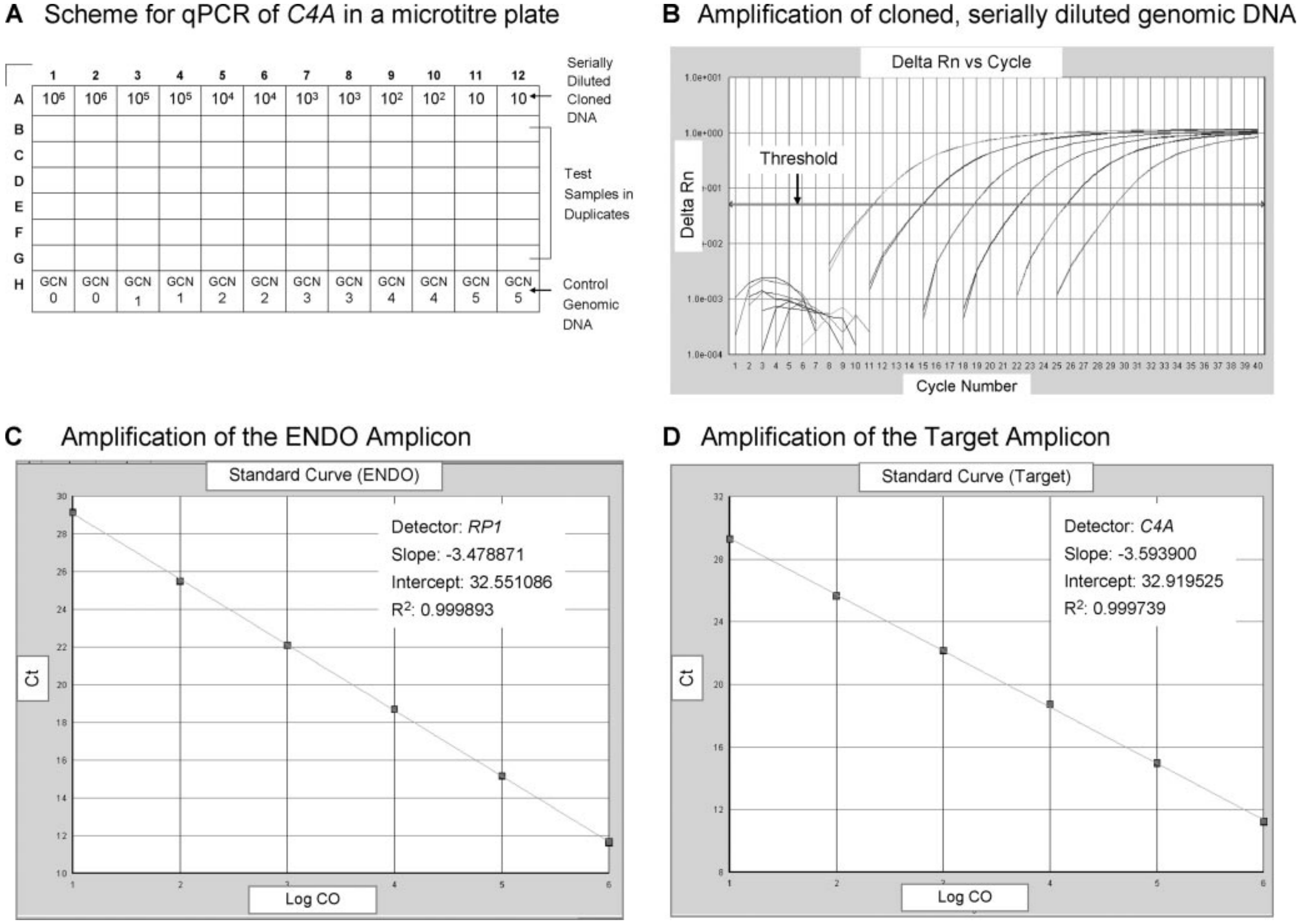
Received for publication April 12, 2007. Accepted for publication June 22, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant R01 AR050078 and National Institute of Diabetes and Digestive and Kidney Diseases Grant P01 DK55546 from the National Institutes of Health.

[2] Address correspondence and reprint requests to Dr. C. Yung Yu, Room W402, Columbus Children's Research Institute, 700 Children's Drive, Columbus Ohio 43205. E-mail address: cyu@chi.osu.edu

[3] Abbreviations used in this paper: CNV, copy number variation; AH, ancestral haplotype; *C4L*, long *C4* gene with endogenous retrovirus HERV-K(C4); *C4S*, short *C4* gene without HERV-K(C4); C_T, threshold cycle; ENDO, endogenous control; GCN, gene copy number; LLD, long-range linkage disequilibrium; LTR, long terminal repeat; qPCR, quantitative real-time PCR; RCCX, *RP-C4-CYP21-TNX* module; PFGE, pulsed field gel electrophoresis; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism; TNX, tenascin-X



	1	2	3	4	5	6	7	8	9	10	11	12
A	10^6	10^6	10^5	10^5	10^4	10^4	10^3	10^3	10^2	10^2	10	10
B												
C												
D												
E												
F												
G												
H	GCN 0	GCN 0	GCN 1	GCN 1	GCN 2	GCN 2	GCN 3	GCN 3	GCN 4	GCN 4	GCN 5	GCN 5

FIGURE 1. Real-time PCR using the relative standard curve method to quantify and characterize GCN variations. *A*, Sample setup plate for a qPCR assay. In a typical experiment, *row A* is reserved for a serially diluted cosmid sample containing equal copy numbers of the ENDO and target genes to construct two standard curves. *Row H* is reserved for genomic DNA samples containing different known copy numbers of the target gene to construct a calibration curve. *Rows B–G*, Reserved for test samples with an unknown *C4* copy number. All samples are assayed in duplicates. *B*, Amplification of serially diluted genomic DNA. In the absolute quantification (standard curve) setup window of SDS software (Applied Biosystems) the most concentrated dilution is assumed to contain 1,000,000 copies (input in "quantity") of both the ENDO and target genes, and each subsequent dilution is 10-fold diluted from the previous one. The diluted cosmid gives C_T ranges from approximately 10 to 30. *C* and *D*, Amplifications of the ENDO and target Amplicons. After the PCR, the SDS software automatically constructs a standard curve for the ENDO (*C*) and the target (*D*) by plotting the "absolute" copies of ENDO and target genes in each dilution against their corresponding observed C_T. Based on these two standard curves, the "absolute" copies of the ENDO and target genes are computed for samples in *row B–H.* Because in each diploid sample the GCN of our ENDO gene (*RP1*) is two, the GCN of the target gene is two times the molar ratio of the target gene to the ENDO gene, which is the quotient of the "absolute" copies of target gene to that of the ENDO gene. At higher GCNs of each target gene there is an intrinsic underestimation of the actual GCN. To correct such an underestimation, for each plate, we construct a calibration curve by plotting the observed *C4* copy number vs their actual *C4* copy number, producing a linear line with an equation in the format of $Y = mX + b$ where Y is the actual copy number of the target gene and X is the observed or calculated GCN. The GCNs for all unknown samples were calibrated based on this equation and rounded up to the closest integer.

that are not affected by DNA methylation and resolved by pulsed field gel electrophoresis (PFGE); and 2) carefully designed genomic RFLP analyses with appropriate probes for hybridization. The long range mapping technique yields information on the number of segmental duplications on each haplotype. Deliberately designed genomic RFLP analyses may give detailed information on structural variation within duplicated modules or complexes. Major limitations for these Southern blot-based strategies, however, include the time-consuming procedures that require 3–14 days for one complete cycle of experiment and the necessity of having a relatively large quantity of high m.w. genomic DNA in the range of 5–10 micrograms for each reaction. Large-scale epidemiologic studies of CNVs in human diseases usually require sensitive and high throughput methods. Thus, the high sensitivity and the fast performance of real-time PCR becomes an attractive alternative to help determine the CNVs of well-characterized genomic loci.

In a real-time PCR, a fluorescence dye is incorporated into the amplified DNA and emits light proportional to the number of amplified copies as the PCR proceeds. The kinetics of the entire PCR is recorded as the light emitted during each PCR cycle is being detected by the real-time machine. Quantification of the initial amount of DNA present in a reaction is based on the number of cycles it takes to reach a threshold (C_T), at which the fluorescence is increased significantly and passes an arbitrarily defined value (9). During a real-time PCR, the greater the initial amount of DNA template present, the sooner the reaction reaches the fluorescence threshold and the smaller the C_T observed (Fig. 1*B*). Several dye chemistries are available to monitor the kinetics of the amplification process. The TaqMan dye chemistry is a preferred choice

Table I. *Designs of amplicons used to determine the copy numbers of human C4A, C4B, long and short C4 genes, and RCCX modules*

Amplicon	Forward Primer	Reverse Primer	Probe	Amplicon Position[a] (Length in Nucleotides)	Annealing Temperature (°C)
C4A	C4F2	C4A32	C4AB	7489 to 7629 (141)	60
C4B	C4BF	C4BR2	C4AB3	7549 to 7640 (92)	59
C4L	C4Fin95	C4L-3LTR-R	C4in95	2502; to 20 bp into 3′ LTR of HERV-K(C4) (102)	60
C4S	C4Fin95	C4Sin9R-2	C4in95	2502 to 2604 (103)	59
TNXA-RP2	XA-RP2F2	XA-RP2R3	XA-RP2	3372 to 3440 (69)[b]	60
RP1 (ENDO)	RP1E4F	RP1E4R	RP1	−2598 to −2528 (71)	59–60

[a] Numbering is based on Fig. 9 of Ref. 7 except for *TNXA-RP2*.
[b] Numbering is based on data for GenBank accession no. L26263.

because of its specificity and precision in quantitative PCR (10). Therefore, we investigated the feasibility of applying real-time PCR using the TaqMan chemistry to determine the CNV of the human complement *C4*.

The complement component C4 plays a critical role in the classical and mannose-binding lectin activation pathways of the complement system, which is an effector arm for both adaptive and innate immunity (7, 11). Human C4 is polymorphic. At the gene level, *C4* exhibits copy number and size variations (5, 12–14). *C4* genes are located in the class III region of the MHC (also known as the human leukocyte Ag HLA in humans) on chromosome 6p21.3. The *C4* gene is a constituent of the RCCX module that manifests segmental duplication involving four consecutive genes coding for the serine/threonine kinase RP (or STK19), the complement component C4, steroid 21 hydroxylase CYP21, and extracellular matrix protein tenascin-X (TNX) (6, 13, 15–17). One to four copies of RCCX modules or *C4* genes have been reported in a MHC haplotype (5, 14). Any of those *C4* genes can be a long gene (*C4L*) or a short gene (*C4S*), and each *C4* gene can either encode for an acidic C4A protein or a basic C4B protein (8, 18–20). The overall gene copy numbers (GCNs) of total *C4, C4A*, or *C4B* may confer different strengths of innate or adaptive immunity or different susceptibilities to autoimmune or infectious diseases. For examples, deficiency of C4A has been implicated as a risk factor for systemic lupus erythematosus (SLE) in European Americans (20–22), and deficiency of C4B has been linked to a higher prevalence of recurrent infections (23).

The human MHC is noted for the high degree of polymorphisms of the class I and class II genes, the long range disequilibrium of classes I, III, and II alleles in certain ancestral haplotypes, and the association with numerous autoimmune and complex diseases (24–26). There have been intensive investigations in the past three decades to determine polymorphic markers, including microsatellites and single nucleotide polymorphisms (SNPs), as surrogates of MHC haplotypes that are highly significant for transplantation matching and for genetic studies of MHC-associated diseases (27–33). Unfortunately, there are still no accurate markers in the MHC that can accurately reflect the CNVs of *C4A* and *C4B* or RCCX modules, and the *C4* CNVs in most human MHC ancestral haplotypes remain relatively uncharacterized.

From our cumulative studies of *C4* genotypes and phenotypes from >2000 healthy subjects and SLE patients of different ethnic groups, it becomes clear that the copy number of total *C4* genes mainly varies between two and seven, that of *C4A* genes between zero and five, that of *C4B* genes between zero and four, that of long *C4* genes between zero and six, and that of short *C4* genes between zero and four (5, 8, 14, 19, 20). Our objectives are to design and validate methods that can definitively elucidate each of these genetic features.

The most informative methods for elucidating *C4* GCN variation are Southern blot analyses of the following: 1) *Pme*I-digested genomic DNA in agarose plugs resolved by PFGE that deciphers the copy number and size of RCCX modules in haplotypes; 2) *Taq*I genomic RFLP that reveals the configurations and numbers of the long and short *C4* genes plus the copy numbers of the neighboring genes *CYP21A* and *CYP21B* and *TNXA* and *TNXB* (5, 8, 18, 34); and 3) *Psh*AI-*Pvu*II genomic RFLP that discloses the molar ratio of *C4A* and *C4B* genes (35–37). To reduce the amounts of genomic DNA needed for *C4* genotyping, alternative methods including module-specific PCR and labeled primer, single-cycle, DNA polymerization PCR were developed to determine the total number of *C4* genes and the relative dosage of *C4A* and *C4B*, respectively (35, 38). However, these methods are relatively laborious and time consuming and therefore not favorable for high throughput epidemiologic studies.

In this study, we present five different quantitative real-time PCR (qPCR) assays that allow an in-depth interrogation of the CNVs of *C4A* and *C4B*, *C4L* and *C4S*, and RCCX modules. These methods have been vigorously tested in selected genomic DNA samples with *C4* GCN varying from 2 to 7 and from common human cell lines. In addition, the CNVs of *C4A*, *C4B*, *C4L*, *C4S*, and RCCX modules in 50 consanguineous subjects that contain many MHC ancestral haplotypes have been elucidated by these qPCR strategies.

Table II. *Primer and probe sequences used to determine the copy numbers of human C4A, C4B, long and short C4 genes, and RCCX modules*

Primer Name	Primer Sequences
C4F2	5′-CCTTTGTGTTGAAGGTCCTGAGTT-3′
C4A32	5′-TCCTGTCTAACACTGGACAGGGGT-3′
C4BF	5′-TGCAGGAGACATCTAACTGGCTTCT-3′
C4BR2	5′-CATGCTCCTATGTATCACTGGAGAGA-3′
C4Fin95	5′-TTGCTCGTTCTGCTCATTCCTT-3′
C4L-3′ LTR-R	5′-GTTGAGGCTGGTCCCCAACA-3′
C4Sin9R-2	5′-GGCGCAGGCTGCTGTATT-3′
XA-RP2F2	5′-TCCTGCAGTCATCTTTGTCTTCAG-3′
XA-RP2R3	5′-GAGCTGCAGATGGGATACCTTTAA-3′
RP1E4F	5′-GACCAAATGACACAGACCTTTGG-3′
RP1E4R	5′-GACTTTGGTTGGTTCCACAAGTC-3′

Probe Name	Probe Sequence
C4AB	VIC-CCA GGA GCA GGT AGG AGG CTC GC
C4AB3	VIC-AGC AGG CTG ACG GC
C4in95	VIC-CTC CTC CAG TGG ACA TG
XA-RP2	VIC-CCA AAT GCA CAA GTA CT
RP1	FAM-AGG GAC TCA GAA ATC ACG T

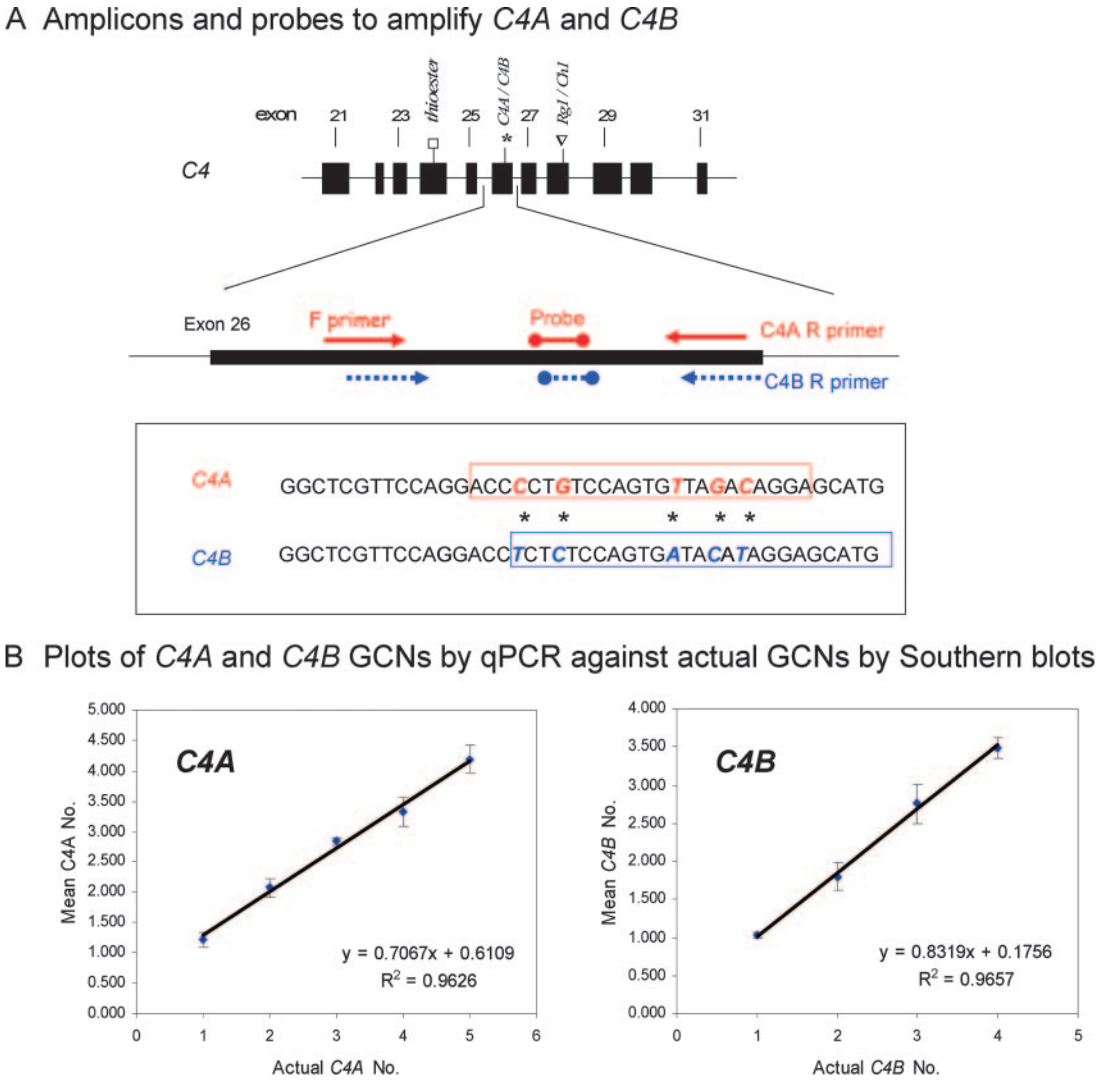


FIGURE 2. Determination of *C4A* and *C4B* GCNs. *A*, Schematic representations of amplicon designs for *C4A* and *C4B* genes. Double-headed bars represent the TaqMan probes; arrows depict the PCR primers. The five-nucleotide sequence polymorphisms specific for the *C4A* gene and the *C4B* gene are located in exon 26 (marked by asterisks and italics). The reversed primer (R primer) for each amplicon (*C4A* and *C4B*) was designed to completely incorporate all five *C4A* or *C4B* isotypic nucleotides to ensure specific amplification. Detailed sequences of the two *C4* isotypes are shown in the box. Two slightly different forward primers (F primer) and TaqMan probes were designed to optimize the amplification efficiency. *B*, Real-time PCR assay quantifying the copy number of *C4A* and *C4B* by the relative standard curve method. In each panel representative results from one 96-well microtiter plate with 28 selective samples from different *C4* GCN groups were presented. The mean and SD of *C4* GCNs from multiple samples determined by qPCR were plotted against their actual GCNs elucidated by Southern blot analyses. A linear line showing a high correlation coefficient ($R^2 > 0.96$) was produced for both the *C4A* and *C4B* assays. The linear equations generated were used to correct the underestimation of the observed GCNs determined by real-time PCR assays to give the best estimation of the samples with higher GCNs (see Table III).

Materials and Methods

Genomic DNA samples

Genomic DNA was isolated from the peripheral EDTA-blood of consented donors (20, 36) or from cultured cells using the Puregene DNA isolation kit (Gentra Systems) following the manufacturer's instructions. *C4* GCNs of each sample of DNA were previously determined by *Taq*I RFLP-Southern blot analysis and *Pvu*II-*Psh*AI RFLP-Southern blot analysis. The RCCX modules for each sample were confirmed by PFGE (14, 35, 37). Genomic DNA samples from consanguineous subjects were purchased from the International Histocompatibility Working Group (IHWG). Details of the HLA class I and II genotypes for the samples can be found at http://www.ihwg.org/cellbank/dna/refpan_hla_consang_table.html.

Cosmid DNAs

Three genomic DNA cosmids containing human complement *C4* and its neighboring genes were used for the qPCR assays with the standard curve method. Cos 3A3 contains a genomic DNA fragment spanning from *SKI2W* to the 3′ end of a long *C4A* gene (13, 39, 40). Cos KEM-1 was isolated from a subject with a monomodular short *C4B* (41). Cos 8 was isolated from a genomic library made from MOLT4 and contains genomic DNA fragment with a long *C4* gene, a short *C4* gene, and *TNXA-RP2* sequence at intergenic region between the two *C4* genes (12, 42, 43). In addition to containing the target gene, these cosmids also contain sequences for the endogenous control gene *RP1*.

Real-time PCR using TaqMan dye chemistry

All of our real-time PCR assays used TaqMan minor groove binder (MGB) probes (Applied Biosystems). The target probes were VIC labeled and the endogenous control probe was FAM labeled. Each reaction consisted of each of the forward and reverse primers (0.5–1 μM) for both the target and control amplicons, 100 nM the target probe and endogenous control probe, 15 ng of test genomic DNA (diluted to ~5 ng/μl and used at 3 μl per reaction), 2× TaqMan universal PCR master mix (P/N 4324018; Applied Biosystems). Final volume was adjusted to 10 μl with molecular grade water. Each sample was analyzed in triplicate and reactions were conducted in a MicroAmp fast 96-well optical reaction plate (P/N 4346906; Applied Biosystems) sealed with an optical adhesive cover (P/N 4311971;

Table III. *Mean C4 gene dosage values as determined by real-time PCR in 28 selected samples with known GCNs*[a]

Assay	Copy Number	No. of Samples (*n*)	Mean	SD	Observed/Actual Ratio	Adjusted Mean	Adjusted Observed/Actual Ratio
C4A	0	1	NA	NA	NA	NA	NA
	1	4	1.22	0.12	1.22	0.86	0.86
	2	8	2.08	0.16	1.04	2.07	1.04
	3	6	2.84	0.07	0.95	3.15	1.05
	4	5	3.33	0.23	0.83	3.84	0.96
	5	3	4.19	0.23	0.84	5.06	1.01
C4B	0	5	NA	NA	NA	NA	NA
	1	8	1.03	0.05	1.03	1.03	1.03
	2	9	1.80	0.18	0.90	1.95	0.97
	3	3	2.76	0.25	0.92	3.10	1.03
	4	2	3.48	0.14	0.87	3.98	0.99
C4L	0	1	NA	NA	NA	NA	NA
	1	2	1.05	0.10	1.05	0.94	0.94
	2	7	1.93	0.16	0.97	2.06	1.03
	3	6	2.62	0.24	0.87	2.93	0.98
	4	6	3.42	0.32	0.86	3.95	0.99
	5	4	4.40	0.30	0.88	5.19	1.04
	6	2	4.89	0.26	0.82	5.82	0.97
C4S	0	8	NA	NA	NA	NA	NA
	1	9	1.11	0.09	1.11	1.01	1.01
	2	4	1.87	0.15	0.94	1.91	0.95
	3	4	2.88	0.12	0.96	3.10	1.03
	4	2	3.60	0.04	0.90	3.93	0.98
TNXA-RP2[b]	0 (2)	4	NA	NA	NA	NA	NA
	1 (3)	3	0.99 (2.99)	0.05	0.99	0.83 (2.83)	0.83
	2 (4)	6	1.97 (3.97)	0.10	0.98	2.08 (4.08)	1.04
	3 (5)	9	2.72 (4.72)	0.23	0.91	3.05 (5.05)	1.02
	4 (6)	5	3.44 (5.44)	0.39	0.86	3.96 (5.96)	0.99
	5 (7)	1	4.12 (6.12)		0.823	4.83 (6.83)	0.97

[a] The term "Observed/Actual Ratio" refers to the ratio of the observed GCN to the actual GCN as calculated by real-time PCR. NA, Not available or not applicable (no C_T data was available when the target gene was absent). Adjusted, A value after inverse prediction based on the calibration curve.

[b] Number in parentheses represents total *C4*.

Applied Biosystems). Real-time PCR was performed using the ABI 7500 fast real-time PCR system using PCR cycles of 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 59–60°C for 1 min. Specific features for each amplicon are shown in Table 1. Sequences for PCR primers and probes are shown in Table II.

The relative standard curve method of real-time PCR

To calculate the copy number of the target genes, the relative standard curve method was used. To set the 96-well plate for quantification using this method, we selected the "absolute quantification (standard curve) assay option in SDS software (Applied Biosystems). To construct the standard curve for each plate, a serially diluted "standard" sample is required. In this case, we serially diluted selected cosmid samples with defined *C4* and *RP1* gene by 6 logs of DNA concentrations that cover a C_T range of ~10–30. Note that in this method the "absolute copy number" for the cosmid sample in each dilution is not necessarily required because we are only interested in deciphering the molar ratio of the target gene to the endogenous control (ENDO) gene, and all of the cosmid samples used contain equal copy numbers of the target and ENDO genes. For example, a sample for the lowest dilution can be assumed to have 10^6 copies of the target gene and the ENDO gene, the next 10-fold diluted sample has 10^5 copies, the next one has 10^4 copies, and so on. Each test genomic DNA sample was diluted to ~5 ng/μl, and 3 μl was used for each reaction. This quantity will yield an ENDO C_T of 24–26. After the PCR was completed, we allowed the machine to automatically set the fluorescence threshold for both the target and ENDO gene to calculate the C_T for each sample. A standard curve is generated with the log ("absolute" copy number) of DNA for the ENDO or the target genes vs their corresponding C_T for each dilution (Fig. 1, *C* and *D*). Based on these standard curves, the initial "absolute" copy numbers of each of ENDO and target genes in the test samples were calculated by the SDS software. Because the target and ENDO genes were amplified from the same diluted DNA sample, the ratio of the "absolute" copy number of the target gene to the "absolute" copy number of the endogenous control gene represents the molar ratio of the target gene to the endogenous control. Because the copy number of our endogenous control *RP1* in a diploid genome is always 2, this ratio multiplied by 2 is the GCN of our target genes per diploid genome. Our individual qPCR assays are capable of accurately distinguishing GCN from zero to three copies. However, at higher gene dosages (four copies or more) there is an intrinsic underestimation of the actual GCN. To correct such an underestimation, we include samples with the known GCN of the target gene in each plate (Fig. 1*D*). A calibration curve (similar to Figs. 2*B*, 3*B*, and 4*B*) is constructed by plotting the calculated GCN vs the actual GCN of these samples. Using the inverse predicted equation generated from this calibration curve, the GCNs of test samples are calculated and rounded to the closest integers.

SDS software and statistic programs

The C_T was recorded by the SDS version 1.3.1 software accompanying the ABI 7500 fast real-time PCR system (Applied Biosystems). For assays based on relative standard curved methods, we imported data with C_T and calculated an "absolute" copy number generated by the SDS software and performed additional calculations in Excel (Microsoft). The calibration equations (i.e., Figs. 2*B*, 3*B*, and 4*B*) used to adjust the intrinsic underestimation of the observed GCNs to the actual GCNs were generated as the best fitted linear line in Excel (Microsoft).

Results

Determination of human C4A and C4B GCNs by the relative standard curve method

Five nucleotide changes within a span of 20 bp in exon 26 contribute to the isotype-specific sequences for *C4A* and *C4B* (Fig. 2*A*). Based on these sequences, amplicons specific for *C4A* and

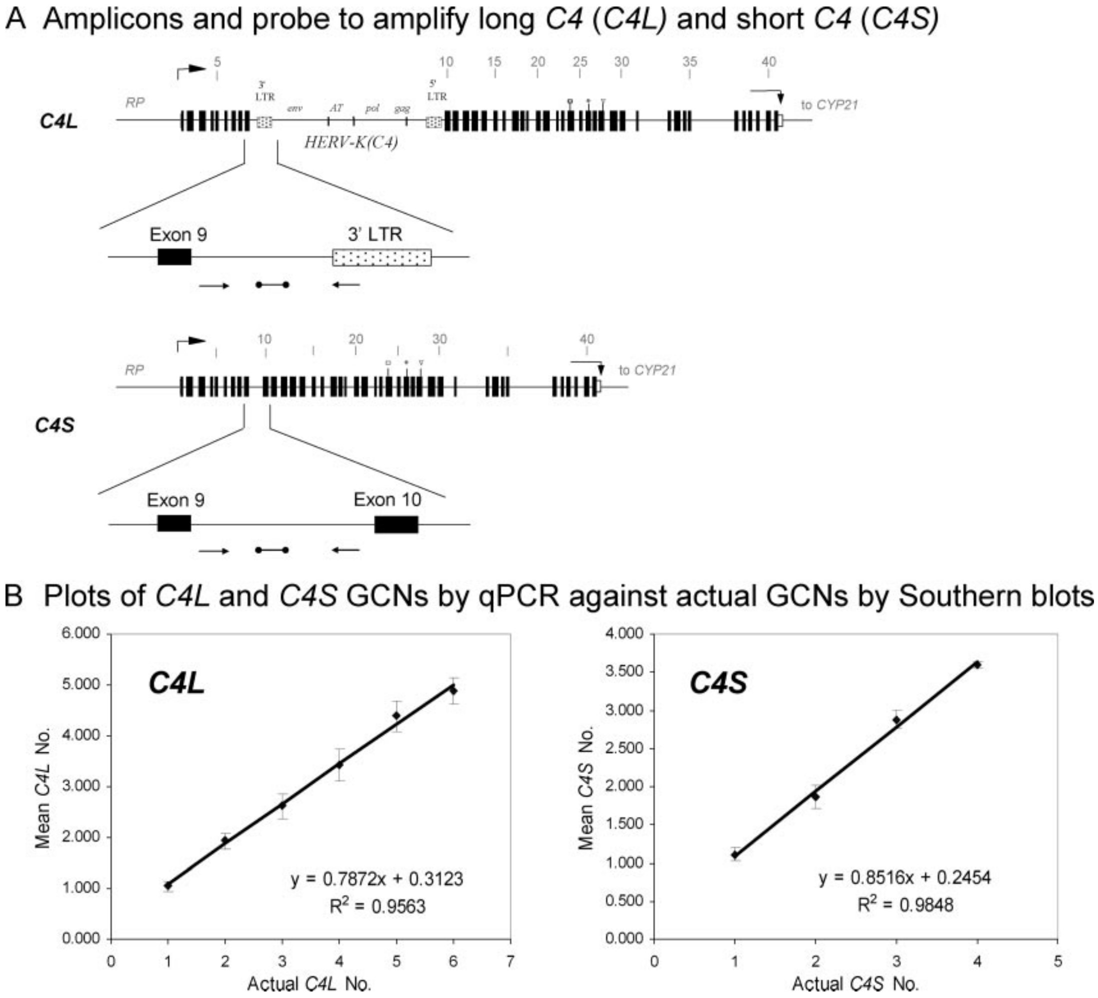


FIGURE 3. Determination of GCNs for the long and short *C4* genes, *C4L* and *C4S*. *A*, Schematic diagram showing amplicons for *C4L* and *C4S*. Double-headed bars represent the TaqMan probes and the arrows depict PCR primers. The difference between a long *C4* gene and a short *C4* gene is the insertion of the endogenous retrovirus *HERV-K(C4)* into the intron 9 of the long *C4*. A common forward primer and a common TaqMan probe were used for both *C4L* and *C4S* amplifications, while a reverse primer annealed to the 3′ LTR of the *HERV-K(C4)* was used for specific amplification of the long gene and a reverse primer spanning the putative insertion site of the endogenous retrovirus was used to amplify the short gene. *B*, Real-time PCR assay quantifying the copy numbers of *C4L* and *C4S* by the relative standard curve method. In each panel, representative results from the same 96-well microtiter plate with 28 selective samples from different *C4* GCN groups were presented. The mean and the SD of *C4* GCNs from multiple samples determined by real-time PCR were plotted against their actual GCNs. The linear equations generated were used to correct the underestimation of the observed GCNs determined by real-time PCR assay to give the best estimation of the samples with higher GCNs (see Table III).

C4B were designed. The GCNs (GCNs) of *C4A* and *C4B* are determined by independent assays. In each assay, a reverse primer (sequence showed in box) discriminately anchors to the isotypic sequence of *C4A* or *C4B* (Fig. 2*A*). In a typical reaction to interrogate the copy number of a target gene in a genomic DNA sample, there are two amplicons: one for an endogenous control gene and one for a target gene. The copy number of endogenous control gene is invariable, while the copy number of targets varies but can be deduced when its molar ratio to the endogenous control (ENDO) gene is known. In our assays, exon 4 of the *RP1* gene, a region that has no known sequence polymorphism or duplicated regions in the human genome, was used to design the ENDO amplicon.

The specificity of each primer set for *C4A* or *C4B* was tested and confirmed in reactions using genomic DNA samples that have both *C4A* and *C4B*, *C4A* only, and *C4B* only. To illustrate the accuracy of the quantification of *C4A* and *C4B* GCNs, we selected 28 human genomic DNA samples that had good representation of every GCN group for *C4A* or *C4B*. The GCNs of those samples were previously defined by *Pme*I-PFGE and genomic *Taq*I RFLP for the total copy number of *C4* genes and by *Psh*AI/*Pvu*II RFLP for the relative copy numbers of *C4A* and *C4B* (20).

In Fig. 2*B*, the mean *C4A* and *C4B* GCNs from DNA samples determined by real-time PCR (i.e., observed copy number) were plotted against the actual copy number determined by Southern blot analyses. Values with SD are shown in Table III. For the *C4A* GCN, individuals with 0–5 copies per diploid genome have been identified, and the most common *C4A* GCN group is 2. As in most real-time PCR quantification assays, when the GCN of the target amplicon is 0, depending on where the fluorescence threshold is set an undetectable or very high threshold cycle number for the target (i.e., more than five cycles later than the C_T of ENDO) will be observed. These samples can be easily identified as a homozygous deficiency for that particular target gene. Using the relative standard curve method, the observed copy numbers for *C4A* GCN groups 1, 2, 3, 4, and 5 were 1.22 ± 0.12, 2.08 ± 0.16, 2.84 ± 0.07, 3.33 ± 0.23, and 4.19 ± 0.23, respectively. Therefore, the *C4A* assay can accurately distinguish among zero, one, two, and three copies of *C4A* genes. At the higher *C4A* GCN groups 4 and

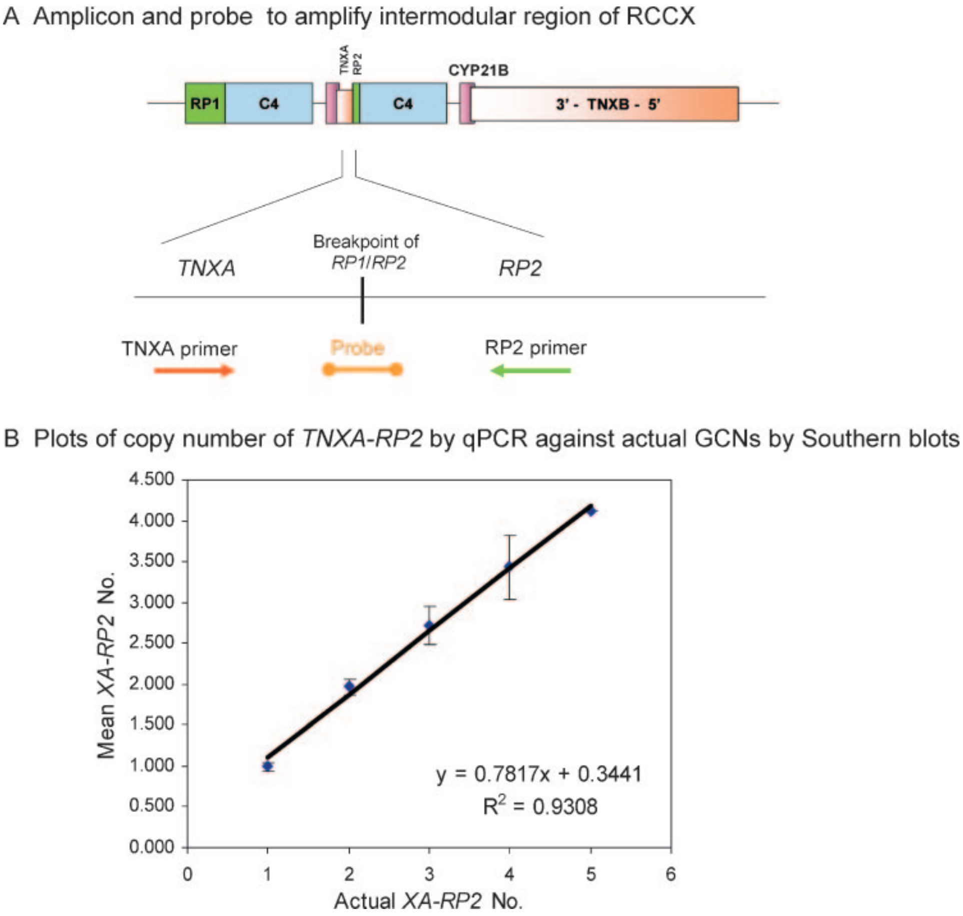


FIGURE 4. Determination of copy numbers of RCCX modules by elucidating the number of RCCX intermodular regions (*TNXA-RP2*). *A*, Schematic design of the amplification of the *TNXA-RP2* junction. Double-headed bars represent the TaqMan probes and arrows depict the PCR primers. *A*, *TNXA-RP2* (XA-RP2) specific amplicon employs a forward primer that amplifies *TNXA*, a reverse primer that amplifies the *RP2* gene, and a TaqMan probe that spans the breakpoint sequence for *TNXA* and *RP2*. In each MHC haplotype there is ($n - 1$) copy of *TNXA* and *RP2* where n is the number of RCCX modules or total *C4* genes. *B*, *XA-RP2* copy numbers quantified by real-time PCR assay using the relative standard curve method. The results presented are from 28 selective samples with various copy numbers of *XA-RP2*. The mean and the SD of GCNs from multiple samples determined by real-time PCR were plotted against their actual GCNs as determined by Southern blot analyses. The linear equation generated was used to correct underestimation at higher gene dosages (see Table III).

5, there is a tendency of underestimation as well as decreased resolution between the neighboring groups.

For *C4B* GCN, 0–4 copies per diploid genome have been observed and the most common GCN group is 2. In the current assay, the mean observed *C4B* copy numbers for GCN groups 1, 2, 3, and 4 were 1.03 ± 0.05, 1.80 ± 0.18, 2.76 ± 0.25, and 3.48 ± 0.14, respectively. Similar to the case for *C4A*, high accuracies are obtained for GCN groups 0, 1, 2, and 3, but an underestimation is present at a GCN equal to four.

To correct the intrinsic underestimation of the actual GCNs, especially for the higher end, a calibration curve is constructed by finding the best fitted line between the observed mean GCN against their actual GCN. A linear equation in the form of $Y = mX + b$ with Y being the observed GCN and X being the actual GCN is produced. Using this equation, an inverse prediction equation is obtained with $X' = (Y - b)/m$. X' calculated by this method is adjusted for the underestimation seen in GCN above 2 and therefore more closely represents the actual GCNs. The adjusted mean based on this method that is listed in Table III showed improved estimation to the actual GCN with an overall observed to actual GCN quotient of >0.95 for GCNs greater than one.

Determining the copy numbers of C4L and C4S

A human *C4* gene can either be a long gene of 21 kb or a short gene of 14.6 kb. The difference is due to the insertion of an endogenous retroviral element, *HERV-K*(*C4*), into intron 9 of a long *C4* gene. To selectively amplify a long gene or a short gene, we use the distinct sequences created by the insertion of *HERV-K*(*C4*). A common forward primer and a common TaqMan probe are used for both the *C4L* and the *C4S* assays (Fig. 3*A*). To selectively amplify the *C4L* gene, a reverse primer that anchors at the 3′ long terminal repeat (LTR) of *HERV-K*(*C4*) is used. Because there is no insertion of *HERV-K(C4)* in a short *C4* gene, *C4S* is not amplified by this primer set. Likewise, for the specific amplification of C4*S* a reverse primer is designed to span the putative integration site that would have been the location of *HERV-K*(*C4*) in a long gene and, therefore, is incapable of amplifying the long *C4*. Both primer sets for *C4L* and *C4S* were tested and proven to be specific for their target genes by using genomic DNA samples with a long *C4* only and samples with a short *C4* only.

The GCN assays for long and short *C4* were performed in a similar manner as those for *C4A* and *C4B* by using the *RP1* exon 4 amplicon as an endogenous control. Copy numbers were calculated based on the relative standard curve method. The GCN of *C4L* observed in an individual ranges from 0 to 6, whereas the GCN of *C4S* ranges from 0 to 4. In the GCN assays of *C4L*, the observed mean copy numbers for GCN groups 1, 2, 3, 4, 5, and 6 are 1.05 ± 0.10, 1.93 ± 0.16, 2.62 ± 0.24, 3.42 ± 0.32, 4.40 ± 0.30, and 4.89 ± 0.26, respectively. Accurate results are obtained for *C4L* GCNs 0, 1, 2 and 3, but those with 4, 5, and 6 are undervalued. After adjustments using the calibration curve (Fig. 3*B*), the GCN of *C4L* can be deduced with higher precision by using the ratio of adjusted observed value to actual value (adjusted observed/actual ratio) of 1.00 ± 0.06 for all GCN groups (Table III).

For the GCN assays of *C4S*, the observed mean copy numbers for GCN groups 1, 2, 3, and 4 are 1.11 ± 0.09, 1.87 ± 0.15, 2.88 ± 0.12, and 3.60 ± 0.04, respectively. Along with the relatively smaller SD (<8%) and its smaller potential range (0–4 copies), this assay worked with high resolution for all GCN groups. As in the *C4A*, *C4B*, and *C4L* assays, the effect of an intrinsic underestimation of the high GCN groups existed but is adjustable by using a calibration equation (Fig. 3*B* and Table III).

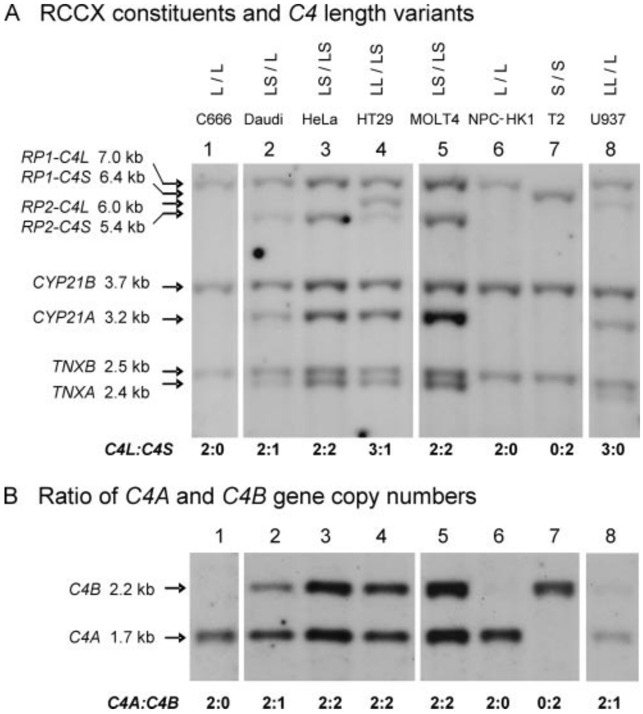


FIGURE 5. RFLP Southern analyses for copy numbers of RCCX modules and *C4A* and *C4B* genes in eight human cell lines. *A*, *Taq*I genomic RFLP reveals the configurations and numbers of long and short *C4* genes plus the copy numbers of the neighboring genes *CYP21A* and *CYP21B* and *TNXA* and *TNXB* in each cell line DNA. *B*, *Psh*AI-*Pvu*II genomic RFLP determines the ratio of *C4A* and *C4B* GCNs.

Table IV. *Comparison of the results of C4 gene copy numbers as determined by RFLP Southern blot analyses and real-time PCR in common cell lines*

	Copy Numbers of *C4* and RCCX Determined by Genomic RFLP				
Sample	*C4A*	*C4B*	*C4L*	*C4S*	Total *C4* (*XA-RP2*)
C666	2	0	2	0	2 (0)
Daudi	2	1	2	1	3 (1)
HeLa	2	2	2	2	4 (2)
HT29	2	2	3	1	4 (2)
MOLT4	2	2	2	2	4 (2)
NPC-HK1	2	0	2	0	2 (0)
T2	0	2	0	2	2 (0)
U937	2	1	3	0	3 (1)

	Copy Numbers of *C4* and RCCX Determined by Real-Time PCR				
Sample	*C4A*	*C4B*	*C4L*	*C4S*	Total *C4* (*XA-RP2*)
C666	1.97	0	1.99	0	2 (0)
Daudi	1.66	1.25	2.24	1.29	3.29 (1.29)
HeLa	1.86	1.96	2.10	1.99	4.01 (2.01)
HT29	1.92	2.14	3.50	0.92	4.11 (2.11)
MOLT4	1.81	1.87	1.92	1.78	3.84 (1.84)
NPC-HK1	1.74	0	1.87	0	2 (0)
T2	0	1.89	0	1.72	2 (0)
U937	1.83	1.30	2.95	0	3.19 (1.19)

Quantifying the number of TNXA-RP2 junctions to corroborate the total number of C4 genes

The GCN of total *C4* can be deduced separately by summing the copy numbers of *C4A* and *C4B* or the copy numbers of *C4L* and *C4S* from the same subject. Under conditions of experimental errors or a high copy number in *C4A*, *C4B*, *C4L*, or *C4S*, ambiguous results can be obtained. Therefore, we designed an independent assay to interrogate the GCN of total *C4* based on the unique structural feature of the RCCX modular variation. This assay specifically quantifies the junction of the *TNXA* and *RP2* genes in any duplicated RCCX module by using a *TNXA*-specific forward primer and a *RP2*-specific reverse primer and coupling with a TaqMan probe that spans the junction of *TNXA* and *RP2* (Fig. 4*A*). In a monomodular RCCX, *TNXA* and *RP2* are absent and hence there is no amplification of the amplicon for the *TNXA-RP2* junction. In a bimodular, a trimodular, or a quadrimodular RCCX, one, two, or three copies of *TNXA-RP2* junctions are present, respectively. A subject homozygous for monomodular RCCX with a total of two *C4* genes has no *TNXA-RP2* junction; a subject heterozygous for a monomodular and a bimodular RCCX with a total of three *C4* genes has one copy of the *TNXA-RP2* junction; a subject homozygous for bimodular RCCX with a total of four *C4* genes has two copies of the *TNXA-RP2* junction, and so on. Therefore, by using the characteristic modular duplication of RCCX we can quantify the total *C4* gene dosages independently by interrogating the copy number of *TNXA-RP2* junction(s). The relatively lower copy number of *TNXA-RP2* junctions improves the accuracy and resolution of subjects with high *C4* gene dosages as it brings down the copy number being interrogated by two. To date, individuals with 2–7 copies of *C4* genes have been identified. The *XA-RP2* junction copy number therefore ranges from 0 to 5. In the *TNXA-RP2* GCN assay the observed mean copy numbers for one, two, three, four, and five copies are 0.99 ± 0.05, 1.97 ± 0.10, 2.72 ± 0.23, 3.44 ± 0.39, and 4.12 respectively. The respective observed vs actual GCN quotients were 0.99, 0.98, 0.91, 0.86, and 0.82. After adjusting for underestimations at higher GCN groups, the adjusted observed/actual GCN quotients improve to >0.98 (for GCN > 1). Using this assay in combination with the complementary *C4A* and *C4B* and *C4L* and *C4S* assays, we can consistently resolve the CNVs of *C4* and RCCX modules with a high degree of accuracy.

CNVs of C4 and RCCX modules in common cell lines by real-time PCR and genomic Southern blot analyses

To facilitate the molecular genetic studies of the role of the *C4* gene CNV in human diseases, we characterize the CNVs of *C4A*, *C4B*, *C4L*, *C4S*, and RCCX in eight common cell lines that can be used to construct standard curves or as control samples to construct an internal calibration curve. The cell lines are the B lymphoid cell line Daudi, the T lymphoid cell line MOLT4, the monocyte cell line U937, the cervical carcinoma cell line HeLa, the intestine cell line HT29, the teratocarcinoma cell line T2, and the nasopharyngeal carcinoma cell lines C666 and NPC-HK1 (44). These cell lines were characterized by quantitative *C4* and *RCCX* real-time PCR assays, *Taq*I genomic RFLP, and *Psh*AI/*Pvu*II genomic RFLP (Fig. 5). The detailed *C4* gene copy numbers of these cell lines are listed in Table IV along with the observed *C4* GCNs obtained in the real-time PCR assays. The assignments for the copy numbers of *C4* genes and RCCX modules for these cell lines were based on the results from the real-time assays in which the results determined by the genomic Southern blots were not revealed until after the assignments were made. The results from the real-time PCR assays and the genomic Southern blot analyses were 100% congruent. Briefly, the *C4* GCNs 2, 3, and 4 are each represented by three cell lines in each group. These cell lines contain zero, one, or two

Table V. *CNVs of complement* C4A, C4B, C4L, C4S, *and RCCX modules in IHWG consanguineous panel*[a]

RCCX haplotypes	IHW No.	Name	*Probable AH*	*A**	*Cw**	*B**	*C4A*	*C4B*	*C4L*	*C4S*	Total *C4*	RCCX-C4[b]	*DRB1**	*DRB3**	*DRB4**	*DRB5**	*DQA1**	*DQB1**	*DPA1**	*DPB1**
Monomodular																				
01	9020	QBL	18.2	2601	0501	1801	2	0	2	0	2	L-A	0301	0202			0501	0201	0103	0202
02	9039	JVM		0201	0501	1801	2	0	2	0	2	L-A	1102	0202			0501	0301	01	020102
03	9006	WT100BIS	35.3	1101	0401	3501	2	0	2	0	2	L-A	0101				0101	0501	020102	010101
04	9047	PLH	47.1	0301	0602	4701	2	0	2	0	2	L-A	0701		0101		0201	0201	0103	1501
05	9068	BM9		0201	0401	3501	0	2	2	0	2	L-B	0801				0401	0402	01	020102
06	9050	MOU	44.3	2902	1601	440301	0	2	2	0	2	L-B	0701		0101		0201	0202	0103	020102
07	9022	COX	8.1	0101	0701	0801	0	2	0	2	2	S-B	0301	0101			0501	0201	01	0301
08	9023	VAVY	8.1	0101	0701	0801	0	2	0	2	2	S-B	0301	0101			050101	0201	0201	0101
Bimodular																				
09	9029	WT51		2301	0802	1401	4	0	4	0	4	LL-AA	0401		0101		03	0302	0103	020102
10	9030	JHAF	51.1	310102	1502	510101	4	0	4	0	4	LL-AA	0407		0103		0301	0301	0103	0301
11	9065	HHKB		0301	0702	0702	2	2	4	0	4	LL-AB	1301	0101			0103	0603	01	0401
12	9031	BOLETH		0201	0304	1501	2	2	4	0	4	LL-AB	0401		0103		030101	0302	0103	0401
13	9032	BSM		0201	0304	1501	2	2	4	0	4	LL-AB	040101		01030101		03	0302	01	020102
14	9104	DHIF		3101	1203		2	2	4	0	4	LL-AB	11							04
15	9105	FPAF		0101	0401	3502	2	2	4	0	4	LL-AB	110401	0202			0103	0603	020101	020102
16	9042	TISI		2402	0401	3508	2	2	4	0	4	LL-AB	1103	0202			0501	0301	0103	0402
17	9035	JBUSH		3201	1203	3801	2	2	4	0	4	LL-AB	1101	020201			050101	0301		0401
18	9098	MT14B		3101	0304	4001	2	2	4	0	4	LL-AB	0404		0101		03	0302		0402
19	9084	CALOGERO		0201	020202	4002	2	2	4	0	4	LL-AB	1601				0102			0401
20	9063	WT47	44.4	3201	0501	4402	2	2	4	0	4	LL-AB	1302	0301			0102	0604	01	1601
21	9004	JThom		0201	0102	270502	2	2	4	0	4	LL-AB	0101				0101	0501	01	0401
22	9015	WT24		0201	020202	270502	2	2	4	0	4	LL-AB	16						01	0301
23	9033	BM14		0301	0702	0702	2	2	2	2	4	LS-AB	0401		0101		03	0302	01	0401
24	9034	SAVC		0301	0702	0702	**2**	2	2	2	4	LS-AB	0401		0101		03	0302	020101	1001
25	9048	LBUF	13.1	3001	0602	1302	2	2	2	2	4	LS-AB	070101		0101		0201	0201	020101	1701
26	9096	LBF	13.1	3001	0602	1302	2	2	2	2	4	LS-AB	0701		0101		0201	0201	02	1701
27	9049	IBW9		3301	0802	1402	2	2	2	2	4	LS-AB	0701				0201		02	0101
28	9072	SPACHECO		3101	0102	1501	2	2	2	2	4	LS-AB	080201				0401	0402	01	0402
29	9101	SPL		3101	0102	1501	2	2	2	2	4	LS-AB	080201				0401	0402	01	0402
30	9038	BM16		0201	0701	1801	2	2	2	2	4	LS-AB	1201	0202			050103	0301	0103	020102
31	9075	DKB		2402	0304	4001	2	2	2	2	4	LS-AB	090102		0103		0302	030302	0103	0401
32	9043	BM21		0101	1701	4101	2	2	2	2	4	LS-AB	1101	0202			0501	0301	0201	1001
33	9036	SPO		0201	0501	4402	2	2	2	2	4	LS-AB	1101	0202			0102	0502	01	020102
34	9040	BM15		0101	0701	4901	2	2	2	2	4	LS-AB	1102	0202			0501	0301	01	0301
35	9092	BM92		2501	0102	5101	2	2	2	2	4	LS-AB	0404		0101		03	0302	0103	0402
36	9010	AMAI		6802	0401	5301	2	2	2	2	4	LS-AB	1503			0101	010201	0602	0301	0402
37	9052	DBB	57.1	0201	0602	5701	2	2	2	2	4	LS-AB	0701		0103102N		0201	030302	0103	0401
38	9102	ARBO		0301	0602	5802	2	2	2	2	4	LS-AB	09							01
39	9051	PITOUT	44.2	2902	1601	440301	2	2	2	2	4	LS-AB	0701		0101		0201	0201	01	0401
40	9069	MADURA		0201	0304	4001	2	2	0	4	4	SS-AB	0801				0401	0402	01	0401
Trimodular																				
41	9061	31227 ABO		0201	0701	1801	4	2	6	0	6	LLL-AAB	1401	0202			0104	0503	01	0401
42	9064	AMALA		0217	0303	1501	2	4	6	0	6	LLL-ABB	1402	0101			0503	0301	0103	0402
43	9099	LZL		0217	0303	1501	2	4	6	0	6	LLL-ABB	1402	0101			0503	0301	01	0402
44	9016	RML		0204	1502	510101	2	4	6	0	6	LLL-ABB	1602			0202	05013	0301	0103	0402

(*Table continues*)

Table V. (*Continued*)

RCCX haplotypes	IHW No.	Name	*Probable AH*	*A**	*Cw**	*B**	*C4A*	*C4B*	*C4L*	*C4S*	Total *C4*	RCCX-C4[b]	*DRB1**	*DRB3**	*DRB4**	*DRB5**	*DQA1**	*DQB1**	*DPA1**	*DPB1**
Quadrimodular																				
45	9060	CB6B		0101	0303	1501	4	4	6	2	8	LLLS-AABB	1301	0202			0103	0603	020201	1901
Heterozygous																				
46	9062	WDV		0201	1203	3801	2	2	4	0	4	LL-AB LL-AB	1301	0101			0103	0603	01	020102; 0401
47	9044	BRIP		2402	0701, 1504	1517 5101	2	2	4	0	4	LL-AB LL-AB	1101	0202			0501	0301		0201; 0402
48	9017	WT8	7.1?	0301	0702	0702	1	2	3	0	3	LL-AB L-B	1501							
49	9005	HOM2		0301	0102	270502	3	2	4	1	5	LLS-AAB LS-AB	0101				0101	0501	01	0401
50	9106	MANIKA		0301	0602	5001	2	3	2	3	5	LSS-ABB LS-AB	0701		0101		0201	0201; 030302	01; 0201	0401; 1301

[a] Modified from http://www.ihwg.org/cellbank/dna/refpan_hla_consang_table.html.

[b] The order of long *C4* (L), short *C4* (S), *C4A* (A), and *C4B* (B) is not determined and they are listed alphabetically.

copies of *C4A*, *C4B*, and *C4S*, and zero, two, or three copies of *C4L*. It is noteworthy that T2 had two *C4B* but no *C4A* genes, while C666 and NPC-HK1 each had two *C4A* but no *C4B* genes. Daudi and U937 both had two *C4A* genes and one *C4B* gene. HeLa, HT29, and MOLT4 each had two *C4A* and two *C4B* genes.

CNVs of C4A and C4B in IHWG consanguineous cell lines with homozygous and heterozygous MHC haplotypes

We applied the *C4* qPCR methods to determine the *C4* and RCCX CNVs in 50 genomic DNA samples derived from cell lines of consanguineous subjects whose HLA class I and class II alleles have been defined by the IHWG. Because of the consanguinity, the HLA haplotypes for these selected cell lines are overwhelmingly homozygous and, therefore, the GCNs for total *C4*, *C4A*, *C4B*, *C4L*, and *C4S* are expected to be in even numbers or 0 if the corresponding locus is absent. Results of the qPCR assays revealed that the copy number of total *C4* genes in a diploid genome among these samples varied from two to eight (Table V).

Forty-five DNA samples were found consistently homozygous across the entire MHC, including the *C4* and RCCX modules. Eight samples contain homozygous monomodular RCCX haplotypes with single *C4* genes. Among them, four were monomodular-long with *C4A*, *L-C4A* (IHW-9020, HLA *B*1801-DRB1*0301*; IHW-9039, HLA *B*1801-DRB1*1102*; IHW-9006, HLA *B*3501-DRB1*0101*; and IHW-9047, HLA *B*4701-DRB1*0701*), two were monomodular-long with *C4B*, *L-C4B* (IHW-9068, HLA *B*3501-DRB1*0801*; and IHW-9050, HLA *B*440301-DRB1*0701)*, and two were monomodular-short with *C4B*, *S-C4B* (IHW-9022, HLA *B*0801-DRB1* 0301-DPB1*0301*; and IHW-9023, HLA *B*0801-DRB1*0301-DPB1*0101*). The results are in keeping with the phenomenon that the *C4* gene in monomodular-long RCCX can either be *C4A* (*L-C4A*) or *C4B* (*L-C4B*), while the *C4* gene in monomodular-short is usually *C4B (S-C4B*).

For bimodular or multimodular RCCX haplotypes, the orders of the long and short *C4* genes and the *C4A* and *C4B* genes with respect to HLA class I and class II genes cannot be determined by qPCR and they are therefore listed alphabetically. Thirty-two samples contained homozygous bimodular RCCX haplotypes. Among them, 14 haplotypes contained two long genes (LL), 17 contained one long gene and one short gene (LS), and only one contained two short genes (SS). Two LL haplotypes had isoexpression of *C4A* from both RCCX modules (LL-AA; IHW-9029, HLA *B*1401-DRB1*0401*; and IHW9030, HLA *B*510101-DRB1*0407*). Each of the other 30 bimodular RCCX haplotypes consisted of one *C4A* and one *C4B* (LL-AB; LS-AB), including the haplotype with two short genes (SS-AB; IHW9069, HLA *B*4001*-DRB1*0801). This phenomenon underscores the relatively high prevalence of bimodular haplotypes with one *C4A* and one *C4B* in the MHC, which contributed to the earlier two-loci model for *C4* genetics (45–47).

Four samples contained homozygous trimodular LLL haplotypes. One haplotype consisted of two *C4A* and one *C4B* (*LLL-AAB*; IHW-9061, HLA *B*1801-DRB1*1401*); the other three each had one *C4A* and two *C4B* (*LLL-ABB*; IHW-9016, HLA *B*510101-DRB1*1602*; IHW-9064 and IHW-9099, both are HLA *B*1501-DRB1*1402*). In other words, each of these four samples had six copies of *C4* genes in a diploid genome. There were four *C4A* plus two *C4B* genes in one sample and two *C4A* plus four *C4B* genes in each of the other three samples.

Remarkably, sample IHW-9060 with HLA *B*1501-DRB1*1301* was homozygous for a quadrimodular RCCX haplotype (i.e., a total of eight *C4* genes in a diploid genome). This haplotype contained three long and one short *C4* genes coding for two C4A proteins and two C4B proteins (*LLLS-AABB*). This is first example

Table VI. *Conservations of MHC haplotypes in class I, class III, and class II regions*

Region	Characteristic Features
1. Entire MHC	
AMALA (IHW-9064) and LZL (IHW-9099)	
HLA	*A*0217 Cw*0303 B*1501 DRB1*1402 DRB3*0101 DQA1*0503 DQB1*0301 DPA1*01 DPB1*0402*
RCCX-C4	LLL-ABB
LBUF (IHW-9048) and LBF (IHW-9096)	
HLA	*A*3001 Cw*0602 B*1302 DRB1*0701 DRB4*0101 DQA1*0201 DQB1*0201 DPA1*02 DPB1*1701*
RCCX-C4	LS-AB
SPACHECO (IHW-9072) and SPL (IHW-9101)	
HLA	*A*3101 Cw*0102 B*1501 DRB1*080201 DQA1*0401 DQB1*0402 DPA1*01 DPB1*0402*
RCCX-C4	LS-AB
2. Three MHC regions	
COX (IHW-9022) and VAVY (IHW-9023), HLA-A to HLA DQB1	
HLA	*A*0101 Cw*0701 B*0801 DRB1*0301 DRB3*0101 DQA1*0501 DQB1*0201*
RCCX-C4	S-B
BOLETH (IHW-9031) and BSM (IHW-9032), HLA-A to HLA-DPA1	
HLA	*A*0201 Cw*0304 B*1501 DRB1*0401 DRB4*0103 DQA1*03 DQB1*0302 DPA1*01*
RXXC-C4	LL-AB
BM14 (IHW-9033) and SAVC (IHW-9034), HLA-A to DQB1	
HLA	*A*0301 Cw*0702 B*0702 DRB1*0401 DRB4*0101 DQA1*03 DQB1*0302*
RCCX-C4	LS-AB
SPACHECO (IHW-9072) and SPL (IHW-9101), HLA-B to DPB1	
HLA	*B*1501 DRB1*080201 DQA1*0401 DQB1*0402 DPA1*01 DPB1*0402*
RCCX-C4	LS-AB
3. Two MHC regions	
QBL (IHW-9020) and JVM (IHW-9039), HLA-Cw to RCCX	
HLA	*Cw*0501 B*1801*
RCCX-C4	L-A
JBUSH (IHW-9035) and WDV (IHW-9062), HLA-Cw to RCCX	
HLA	*Cw*1203 B*3801*
RCCX-C4	LL-AB
4. Class I and class II alleles but NOT in class III (RCCX-C4)	
MOU (IHW-9050) and PITOUT (IHW-9051)	
HLA	*A*2902 Cw*1601 B*440301 DRB1*0701 DRB4*0101 DQA1*0201*
RCCX, C4	L-B, monomodular RCCX with C4B only for MOU LS-AB, bimodular RCCX with both C4A and C4B for PITOUT

demonstrating the presence of eight *C4* genes (four *C4A* plus four *C4B*) in a human sample.

The remaining five samples in the IHW consanguineous panel did not appear to be homozygous in the HLA haplotypes, as different alleles were present in one or more of the class I, II, or III genes.

Conservation of MHC haplotypes

Long-range linkage disequilibrium (LLD) among alleles of class I, III, and II genes with conserved sequences or identical genetic markers spanning hundreds to thousands of kilobases is a remarkable feature of many haplotypes of the MHC (32, 33, 48). Such LLD is conspicuous when we analyze the CNVs of RCCX-C4 and class I and class II alleles in the HLA consanguineous panel. There are three pairs of samples that contained virtually identical genetic markers throughout the entire MHC: IHW-9048 (LBUF) and IHW-9096 (LBF), IHW-9072 (SPACHECO) and IHW-9101 (SPL) and IHW-9064 (AMALA) and IHW-9099 (LZL). There are four pairs of samples that contained continuously identical gene alleles spanning across the three MHC regions and two pairs of samples that contained identical alleles spanning two MHC regions (from HLA-Cw to RCCX-C4) (Table VI). In contrast, we did observe an exception on two samples, MOU (IHW9050) and PITOUT (IHW9051) (Table VI). Despite the high similarities of the flanking class I and class II genes, MOU had monomodular-long with a *C4B* gene (L-B) and PITOUT had bimodular long-short with a *C4A* and a *C4B* gene (LS-AB).

Discussion

Patients with autoimmune or infectious disease often present with hypocomplementemia, especially low C4 and/or low C3. Such a phenomenon can be caused by a massive activation and consumption of complement proteins during a disease state and/or an inherent deficiency or low GCNs of total *C4*, *C4A*, or *C4B*. Therefore, elucidation of the patient's *C4* genotypes and the serial variations of plasma or serum C4 protein concentrations and the cell-bound products of activated C4 can facilitate more accurate disease diagnosis and determination of the disease state, including flares, remissions, and future prognosis (49–51). In this article, we focus on developing a series of sensitive and specific assays using real-time PCR to rapidly determine *C4* genotypic diversity.

In designing the qPCR strategies, we consider specificity, accuracy, cost, and convenience. For amplicons to determine the GCNs of *C4A* and *C4B*, the specificities for the A and B isotypes are based on two reverse primers that incorporate specific sequences of five SNPs within 20 nucleotides close to the 3′ end of exon 26. For determining the copy numbers of the long *C4* and short *C4* genes we use a common probe and a common forward primer that hybridize to the 5′ region of intron 9. The specificities for the long and short genes were built into the reverse primers. The reverse primer for the long *C4* gene amplicon is located in the 3′ LTR of the endogenous retrovirus HERV-K(C4), while that of short *C4* gene amplicon is located downstream of the putative HERV-K(C4) integration site. We designed an additional method that yields the copy number of RCCX modules and therefore the number of total *C4* genes by interrogating the number of junctions for *TNXA-RP2*. In these *C4* GCN assays we use the same endogenous control ENDO amplicon that is designed at an invariable region of *RP1* exon 4, which is 7.9 kb upstream of the breakpoint of RCCX modular duplication at *RP1* exon 7 and is independent of the CNVs of the *C4* and RCCX modules (15). The ENDO amplicon has a constant copy number of two in a diploid genome among all human subjects. It was tested not to interfere with the amplification in any of the five target amplicons described here. Therefore, PCRs for the ENDO and target gene amplicons were performed in the same reaction mixture to ensure accurate quantification of the molar ratio of the target gene to the ENDO gene.

An important prerequisite pertaining to a qPCR of human CNVs is a means to ascertain the accuracy of experimental results, even under the constraint of a limited supply of genomic DNA samples. We addressed this issue by creating five different real-time PCR amplicons, each of which yields relevant and complementary data. The copy number of total *C4* genes equals the copy number of *C4A* plus *C4B* or the copy number of *C4L* plus *C4S*. Agreements of data for total *C4* genes from three independent sources on the same sample increase the confidence of experimental accuracy.

In using real-time PCR for quantification assays a frequently used approach is the $\Delta\Delta C_T$ method, in which the difference between the endogenous control and the target amplicon of a single calibrator is used as a reference and the fold changes are based on the ratio of each sample's ΔC_T to the ΔC_T of the calibrator. However, this method requires almost identical amplification efficiency approaching 100% for both the endogenous control and the target amplicon. Otherwise, the results are sensitive to a slight variation in DNA concentration (35, 52). In *C4* gene dosage quantification assays the specificities of each assay are restricted to sequences that define the A and B isotypes or the long and short genes. We found it exceedingly difficult to design target amplicons that have a high amplification efficiency identical to that of an ENDO amplicon, and the $\Delta\Delta C_T$ strategy tended to yield ambiguous or inaccurate results that are not desirable. We solved such drawback through the application of relative standard curve methods for quantification. For each amplicon to determine the CNV of *C4A*, *C4B*, *C4L*, *C4S*, or RCCX modules, we assign the copy number of target genes after two levels of calibrations. In the first level we use cloned genomic DNA covering 6 logs of DNA concentrations for the ENDO and target amplicons. Such calibration allows the calculation of copy numbers of the target amplicon relative to the ENDO amplicon at a specific DNA concentration and therefore minimizes the discrepancies in amplification efficiencies for the ENDO and target amplicons caused by variations among test DNA concentrations when a single calibrator at a single concentration is used. In our experience, unequivocal results are obtained for low copy numbers of the target genes but a slight underestimation is an intrinsic tendency for subjects with high copy numbers of a target gene. In the second level of calibration we correct the intrinsic underestimation of high GCN groups by creating a calibration curve for observed and actual GCNs among all GCN groups. We have applied these *C4* qPCR assays to determine the *C4* CNVs in >1000 human samples with autoimmune or neurological diseases (S. L. Savelli, R. A. S. Roubey, Y. W. Wu, G. Buxton, and C. Y. Yu, manuscript in preparation; K. Mayilyan, D. R. Weinberger, Y. L. Wu, B. Kolachana, and C. Y. Yu, manuscript in preparation). This technique has been proven to be robust, sensitive, and reliable. We observed that in ~5% of samples the data from the five independent assays may not be in total agreement. Under such scenarios we usually repeat those assays with the inconsistent data and, if possible, seek data for C4A and C4B phenotypes of the same subjects and from family members to support the final assignment.

The quality of DNA is an important factor in performing qPCR assays. In our experience, partially degraded DNA often yields conflicting results between complementary assays. Because DNA is rather unstable in a diluted state we recommend diluting genomic DNA using Tris-EDTA buffer rather than water and performing the required experiments within 2 wk after sample dilution. We also observed that whole genome-amplified DNA yielded a wild variation of target GCNs and is therefore not suitable for qPCR of CNV.

To facilitate application of the qPCR techniques for *C4* genotyping, we have elucidated the genetic diversities of *C4* in eight common human cell lines including Daudi, HeLa, HT29, MOLT4, and U937. Genomic DNA samples of these cell lines can be used as controls for calibration purposes. The monocyte cell line U937 is commonly used for studies of complement *C4* gene expression (53, 54). This cell line contains two *C4A* and one *C4B* genes and, therefore, the expression of total *C4*, *C4A*, and *C4B* transcripts would have to be interpreted together with its genotypic background.

The IHWG consanguineous panel is overwhelmingly homozygous for HLA class I and class II alleles and is therefore ideal for the discovery and haplotyping of SNPs and the characterization of CNVs of *C4* and RCCX in defined HLA haplotypes. The former is exemplified by the MHC Haplotype Project in which eight representative haplotypes are being sequenced and characterized (http://www.sanger.ac.uk/HGP/Chr6/MHC/) (32, 33). The latter is being demonstrated in this study for the presence of monomodular, bimodular, trimodular, and quadrimodular RCCX haplotypes with different copy numbers of long and short *C4* genes, each either coded for C4A or C4B. The results illustrate CNV as a mechanism for generating the genetic diversity of an important immune effector protein. As we and others had shown previously, the phenotypic outcome for the sophisticated genotypic diversity of complement *C4* is a wide range of plasma or serum C4 proteins among different subjects and two isotypes (C4A and C4B) with multiple protein variants (allotypes) that can have different physiologic functions (8, 14, 19, 55–60).

The mechanism leading to the LLD of numerous polymorphic markers as large “frozen blocks” of genomic sequences in the MHC is not known (24, 33). The length variations caused by interindividual CNVs at the class II DRB region and the class III RCCX-C4 region could create mismatches during meiosis and play a role in suppressing productive recombinations among certain haplotypes (5, 8, 61). The LLD of MHC alleles on chromosome 6 that persists in human populations is also known as an ancestral haplotype (AH). Some of the MHC ancestral haplotypes are associated with autoimmune or genetic diseases (24, 26). For example, AH47.1 (IHW-9047) with HLA *B*4701*, RCCX: *L-C4A*, *DRB1*0701* is associated with congenital adrenal hyperplasia (5, 62), AH8.1 (IHW-9022, 9023) with HLA *B*0801*, RCCX: *S-C4B*, *DRB1*0301* (DR3), and *DQB1*0201* is associated with SLE and type 1 diabetes mellitus, AH7.1 with HLA *B*0702*, RCCX: *LL-C4B-C4A*, *DRB1*1501* (DR2) is associated with SLE and multiple sclerosis (24, 63), and AH57.1 with HLA *B*5701, RCCX: LS-AB, DRB1*0701* is associated with psoriasis (24, 30, 64). In-depth characterization of genetic variations of all MHC genes including the polymorphisms of complement factor B and C2 (65, 66) and the constituents of RCCX modules using the consanguineous panel would prove highly informative and help the understanding of the genetic basis of MHC-associated diseases.

HLA-DR3 has been consistently implicated as a risk factor in SLE (24, 28, 30), but HLA haplotypes with DR3 can have different RCCX or *C4A* and *C4B* gene contents. In the IHW consanguineous panel were three samples with DR3 haplotypes: two with HLA-*B*0801*, RCCX monomodular-short with a single *C4B* gene (*S-C4B*) and the absence of *C4A* (COX and VAVY) and the third with HLA-*B*1801,* monomodular-long RCCX with a single *C4A* gene (*L-C4A*) and the absence of *C4B* (QBL). In European Americans we found that the absence or low GCN of *C4A*, but not of *C4B*, is a risk factor for SLE disease susceptibility. By contrast, a high GCN of *C4A* is a protective factor against the onset of the systemic autoimmune disease (20). To examine genetic risk factors in the HLA-associated diseases, it is prudent to elucidate the status of complement *C4A* and *C4B* CNVs in addition to SNPs of MHC genes and the conventional class I and class II alleles.

Acknowledgments

We thank the late Dr. Dolly P. Huang (Chinese University of Hong Kong) for nasopharyngeal carcinoma cell lines C666 and NPC-HK1 and Dr. Douglas Rain from Applied Biosystems for helpful discussion.

Disclosures

The authors have no financial conflict of interest.

References

1. Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, et al. 2004. Large-scale copy number polymorphism in the human genome. *Science* 305: 525–528.
2. Iafrate, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer, and C. Lee. 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* 36: 949–951.
3. Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, et al. 2006. Global variation in copy number in the human genome. *Nature* 444: 444–454.
4. Wong, K. K., R. J. deLeeuw, N. S. Dosanjh, L. R. Kimm, Z. Cheng, D. E. Horsman, C. MacAulay, R. T. Ng, C. J. Brown, E. E. Eichler, and W. L. Lam. 2007. A comprehensive analysis of common copy number variations in the human genome. *Am. J. Hum. Genet.* 80: 91–104.
5. Yang, Z., A. R. Mendoza, T. R. Welch, W. B. Zipf, and C. Y. Yu. 1999. 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.* 274: 12147–12156.
6. Shen, L. M., L. C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. Dangel, M. C. Carroll, W. Zipf, and C. Y. Yu. 1994. 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.* 269: 8466–8476.
7. Yu, C. Y., E. K. Chung, Y. Yang, C. A. Blanchong, N. Jacobsen, K. Saxena, Z. Yang, W. Miller, L. Varga, and G. Fust. 2003. Dancing with complement C4 and the RP-C4-CYP21-TNX (RCCX) modules of the major histocompatibility complex. *Prog. Nucleic Acid Res. Mol. Biol.* 75: 217–292.
8. Blanchong, C. A., B. Zhou, K. L. Rupert, E. K. Chung, K. N. Jones, J. F. Sotos, R. M. Rennebohm, and C. Y. Yu. 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 MHC-associated disease. *J. Exp. Med.* 191: 2183–2196.
9. Livak, K. J., S. J. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.* 4: 357–362.
10. Heid, C. A., J. Stevens, K. J. Livak, and P. M. Williams. 1996. Real time quantitative PCR. *Genome Res.* 6: 986–994.
11. Walport, M. J. 2001. Complement- part II. *New Engl. J. Med.* 344: 1140–1144.
12. Dangel, A. W., A. R. Mendoza, B. J. Baker, C. M. Daniel, M. C. Carroll, L.-C. Wu, and C. Y. Yu. 1994. 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* 40: 425–436.
13. Yu, C. Y. 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: 1057–1066.
14. Chung, E. K., Y. Yang, R. M. Rennebohm, M. L. Lokki, G. C. Higgins, K. N. Jones, B. Zhou, C. A. Blanchong, and C. Y. Yu. 2002. 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.* 71: 823–837.
15. Yu, C. Y., Z. Yang, C. A. Blanchong, and W. Miller. 2000. The human and mouse MHC class III region: a parade of the centromeric segment with 21 genes *Immunol. Today* 21: 320–328.
16. Gitelman, S. E., J. Bristow, and W. L. Miller. 1992. Mechanism and consequences of the duplication of the human C4/P450c21/gene X locus. *Mol. Cell. Biol.* 12: 2124–2134.
17. Higashi, Y., H. Yoshioka, M. Yamane, O. Gotoh, and Y. Fujii-Kuriyama. 1986. 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* 83: 2841–2845.
18. Yu, C. Y., and R. D. Campbell. 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: 383–390.
19. Yang, Y., E. K. Chung, B. Zhou, C. A. Blanchong, C. Y. Yu, G. Füst, M. Kovács, A. Vatay, C. Szalai, I. Karádi, and L. Varga. 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: 2734–2745.
20. Yang, Y., E. K. Chung, Y. L. Wu, S. L. Savelli, H. N. Nagaraja, B. Zhou, M. Hebert, K. N. Jones, Y. Shu, K. Kitzmiller, et al. 2007. 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.* 80: 1037–1054.
21. Yang, Y., E. K. Chung, B. Zhou, K. Lhotta, L. A. Hebert, D. J. Birmingham, B. H. Rovin, and C. Y. Yu. 2004. The intricate role of complement C4 in human systemic lupus erythematosus. *Curr. Dir. Autoimmun.* 7: 98–132.
22. Fielder, A. H. L., M. J. Walport, J. R. Batchelor, R. I. Rynes, C. M. Black, I. A. Dodi, and G. R. V. Hughes. 1983. Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. *Br. Med. J.* 286: 425–428.
23. Bishof, N. A., T. R. Welch, and L. S. Beischel. 1990. C4B Deficiency: a risk factor for bacteremia with encapsulated organisms. *J. Infect. Dis.* 162: 248–250.
24. Dawkins, R., C. Leelayuwat, S. Gaudieri, G. Tay, J. Hui, S. Cattley, P. Martinez, and J. Kulski. 1999. Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol. Rev.* 167: 275–304.
25. Martinez, O. P., N. Longman-Jacobsen, R. Davies, E. K. Chung, Y. Yang, S. Gaudieri, R. L. Dawkins, and C. Y. Yu. 2001. Genetics of human complement component C4 and evolution the central MHC. *Front. Biosci.* 6: D904–D913.
26. Awdeh, Z. L., D. Raum, E. J. Yunis, and C. A. Alper. 1983. Extended HLA/complement allele haplotypes: Evidence for T/t-like complex in man. *Proc. Natl. Acad. Sci. USA* 80: 259–263.
27. de Bakker, P. I., G. McVean, P. C. Sabeti, M. M. Miretti, T. Green, J. Marchini, X. Ke, A. J. Monsuur, P. Whittaker, M. Delgado, et al. 2006. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* 38: 1166–1172.
28. Graham, R. R., W. A. Ortmann, C. D. Langefeld, D. Jawaheer, S. A. Selby, P. R. Rodine, E. C. Baechler, K. E. Rohlf, K. B. Shark, K. J. Espe, et al. 2002. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am. J. Hum. Genet.* 71: 543–553.
29. Tiwari, J. L., and P. I. Terasaki. 1985. *HLA and Disease Associations.* Springer-Verlag, New York.
30. Thorsby, E. 1997. HLA associated diseases. *Hum. Immunol.* 53: 1–11.
31. Horton, R., L. Wilming, V. Rand, R. C. Lovering, E. A. Bruford, V. K. Khodiyar, M. J. Lush, S. Povey, C. C. Talbot, Jr., M. W. Wright, et al. 2004. Gene map of the extended human MHC. *Nat. Rev. Genet.* 5: 889–899.
32. Stewart, C. A., R. Horton, R. J. Allcock, J. L. Ashurst, A. M. Atrazhev, P. Coggill, I. Dunham, S. Forbes, K. Halls, J. M. Howson, et al. 2004. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* 14: 1176–1187.
33. Traherne, J. A., R. Horton, A. N. Roberts, M. M. Miretti, M. E. Hurles, C. A. Stewart, J. L. Ashurst, A. M. Atrazhev, P. Coggill, S. Palmer, et al. 2006. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet.* 2: e9.
34. Schneider, P. M., M. C. Carroll, C. A. Alper, C. Rittner, A. S. Whitehead, E. J. Yunis, and H. R. Colten. 1986. Polymorphism of human complement C4 and

steroid 21-hydroxylase genes: restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants. *J. Clin. Invest.* 78: 650–657.

35. Chung, E. K., Y. Yang, K. L. Rupert, K. N. Jones, R. M. Rennebohm, C. A. Blanchong, and C. Y. Yu. 2002. 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.* 71: 810–822.
36. Chung, E. K., Y. L. Wu, Y. Yang, B. Zhou, and C. Y. Yu. 2005. Human complement components C4A and C4B genetic diversities: complex genotypes and phenotypes. In *Current Protocols in Immunology*. J. E. Coligan, B. E. Bierer, D. H. Margulis, E. M. Shevach, and W. Strober, eds. John Wiley & Sons, Edison, NJ, pp. 13.8.1–13.8.36.
37. Yu, C. Y., C. A. Blanchong, E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou, and J. M. Moulds. 2002. Molecular genetic analyses of human complement components C4A and C4B. In *Manuals of Clinical Laboratory Immunology.* N. R. Rose, R. G. Hamilton, and B. Detrick, eds. ASM Press, Washington, DC, pp. 117–131.
38. Uejima, H., M. P. Lee, H. Cui, and A. P. Feinberg. 2000. Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. *Nat. Genet.* 25: 375–376.
39. Carroll, M. C., R. D. Campbell, D. R. Bentley, and R. R. Porter. 1984. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature* 307: 237–241.
40. Yang, Z., L. Shen, A. W. Dangel, L. C. Wu, and C. Y. Yu. 1998. Four ubiquitously expressed genes. *RD* (D6S45)-*SKI2W* (SKIV2L)-*DOM3Z-RP1*(D6S60E), are present between complement component genes factor B and C4 in the class III region of the HLA. *Genomics* 53: 338–347.
41. Carroll, M. C., A. Palsdottir, K. T. Belt, and R. R. Porter. 1985. Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. *EMBO J.* 4: 2547–2552.
42. Yu, C. Y., and C. Milstein. 1989. A physical map linking the five CD1 human thymocyte differentiation antigen genes. *EMBO J.* 8: 3727–3732.
43. Yu, C. Y., L. C. Wu, L. Buluwela, and C. Milstein. 1993. Cosmid cloning and walking to map human CD1 leucocyte differentiation genes. *Methods Enzymol.* 217: 378–398.
44. Lo, A. K., K. W. Lo, S. W. Tsao, H. L. Wong, J. W. Hui, K. F. To, D. S. Hayward, Y. L. Chui, Y. L. Lau, K. Takada, and D. P. Huang. 2006. Epstein-Barr virus infection alters cellular signal cascades in human nasopharyngeal epithelial cells. *Neoplasia* 8: 173–180.
45. Awdeh, Z. L., D. Raum, and C. A. Alper. 1979. Genetic polymorphism of human complement C4 and detection of heterozygotes. *Nature* 282: 205–208.
46. O'Neill, G. J., S. Y. Yang, and B. DuPont. 1978. Two HLA-linked loci controlling the fourth component of human complement. *Proc. Natl. Acad. Sci. USA* 75: 5165–5169.
47. Roos, M. H., E. Mollenhauer, P. Demant, and C. Rittner. 1982. A molecular basis for the two locus model of human complement component C4. *Nature* 298: 854–855.
48. Awdeh, Z. L., D. Raum, E. J. Yunis, and C. A. Alper. 1983. Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc. Natl. Acad. Sci. USA* 80: 259.
49. Illei, G. G., E. Tackey, L. Lapteva, and P. E. Lipsky. 2004. Biomarkers in systemic lupus erythematosus, I: general overview of biomarkers and their applicability. *Arthritis Rheum.* 50: 1709–1720.
50. Illei, G. G., E. Tackey, L. Lapteva, and P. E. Lipsky. 2004. Biomarkers in systemic lupus erythematosus, II: markers of disease activity. *Arthritis Rheum.* 50: 2048–2065.
51. Liu, C. C., S. Manzi, and J. M. Ahearn. 2005. Biomarkers for systemic lupus erythematosus: a review and perspective. *Curr. Opin. Rheumatol.* 17: 543–549.
52. Szilagyi, A., B. Blasko, D. Szilassy, G. Fust, M. Sasvari-Szekely, and Z. Ronai. 2006. Real-time PCR quantification of human complement C4A and C4B genes. *BMC Genet.* 7: 1–9.
53. Falus, A., J. Kramer, E. Walcz, Z. Varga, J. Setalo, K. Jobst, T. Lakatos, and K. Meretey. 1989. Unequal expression of complement C4A and C4B genes in rheumatoid synovial cells, human monocytoid and hepatoma-derived cell lines. *Immunology* 68: 133–136.
54. Tsukamoto, H., K. Nagasawa, S. Yoshizawa, Y. Tada, A. Ueda, Y. Ueda, and Y. Niho. 1992. Synthesis and regulation of the fourth component of complement (C4) in the human momocytic cell line U937: comparison with that of the third component of complement. *Immunology* 75: 565–569.
55. Mauff, G., B. Luther, P. M. Schneider, C. Rittner, B. Strandmann-Bellinghausen, R. Dawkins, and J. M. Moulds. 1998. Reference typing report for complement component C4. *Exp. Clin. Immunogenet.* 15: 249–260.
56. Isenman, D. E., and J. R. Young. 1984. The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component C4. *J. Immunol.* 132: 3019–3027.
57. Dodds, A. W., S. K. Law, and R. R. Porter. 1985. The origin of the very variable haemolytic activities of the common human complement component C4 allotypes including C4-A6. *EMBO J.* 4: 2239–2244.
58. Dodds, A. W., X.-D. Ren, A. C. Willis, and S. K. A. Law. 1996. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* 379: 177–179.
59. Yu, C. Y., R. D. Campbell, and R. R. Porter. 1988. 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* 27: 399–405.
60. Wu, Y. L., G. C. Higgins, R. M. Rennebohm, E. K. Chung, Y. Yang, B. Zhou, N. Nagaraja, D. J. Birmingham, B. H. Rovin, L. A. Hebert, and C. Y. Yu. 2006. Three distinct profiles of serum complement C4 proteins in pediatric systemic lupus erythematosus (SLE) patients: tight associations of complement C4 and C3 protein levels in SLE but not in healthy subjects. *Adv. Exp. Med. Biol.* 586: 227–247.
61. Blanchong, C. A., E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou, and C. Y. Yu. 2001. Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologs, Slp and C4. *Int. Immunopharmacol.* 1: 365–392.
62. White, P. C., A. Vitek, B. DuPont, and M. I. New. 1988. Characterization of frequent deletions causing steroid 21-hydroxylase deficiency. *Proc. Natl. Acad. Sci. USA* 85: 4436–4440.
63. Yu, C. Y., and C. C. Whitacre. 2004. Sex, MHC and complement C4 in autoimmune diseases. *Trends Immunol.* 25: 694–699.
64. Price, P., C. Witt, R. Allcock, D. Sayer, M. Garlepp, C. C. Kok, M. French, S. Mallal, and F. Christiansen. 1999. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol. Rev.* 167: 257–274.
65. Jahn, I., J. E. Mejia, M. Thomas, C. Darke, H. Schroder, G. Geserick, and G. Hauptmann. 1994. Genomic analysis of the F subtypes of human complement factor B. *Eur. J. Immunogenet.* 21: 415–423.
66. Jahn, I., B. Uring-Lambert, D. Arnold, S. Clemenceau, and G. Hauptmann. 1990. C2 reference typing report. *Complement Inflamm.* 7: 175–182.