# Gene Copy-Number Variations (CNVs) of Complement *C4* and *C4A* Deficiency in Genetic Risk and Pathogenesis of Juvenile Dermatomyositis

Katherine E. Lintner, [1]* Anjali Patwardhan,[1]* Lisa G. Rider,[2] Rabheh Abdul-Aziz,[1] Yee Ling Wu,[1] Emeli Lundström,[3] Leonid Padyukov,[3] Bi Zhou,[1] Alaaedin Alhomosh,[1] David Newsom,[1†] Peter White,[1] Karla B. Jones,[1] Terrance P. O'Hanlon,[2] Frederick W. Miller,[2] Charles H. Spencer,[1] and C. Yung Yu[1¶]

## Supplementary Materials

1. Determination of gene copy-number Variation and protein phenotypes of complement *C4* (Table S1)
2. Plasma protein concentrations of complement C4 and complement factor H in JDM and controls. (Table S2)
3. Global gene expression profiling of PAXgene blood RNA samples from JDM and controls (Table S3 and S4)
4. Figures S1 to S5

## 1. Determination of gene copy-number Variation and protein phenotypes of complement *C4* in JDM.

Figure S1 depicts the structural diversities of human complement *C4* genes and proteins. In the Caucasian population, one to four copies of *C4* genes are present in the class III region of the major histocompatibility (MHC). Segmental duplications of *C4* occur in modular fashion. Each module consists of *RP* (STK19) upstream of *C4*, and *CYP21* and *TNX* downstream of *C4*. Each module is either 32.7 kb or 26.3 kb in size dependent on the size of *C4* genes. Each *C4* gene consists of 41 exons. The long gene consists of an endogenous retrovirus with 6.4 kb in size in intron 9 of the long gene. The isotypic residues for C4A and C4B protein are encoded by exon 26. The Histidine-1106 residue in activated C4B catalyzes nucleophilic attack to target surfaces to form a covalent ester bond.

The RCCX modular variation can be detected by long range mapping of *Pme*I digested genomic DNA resolved by pulsed gel electrophoresis. The specific locations and dosage of the long and short *C4* genes can be demonstrated by *Taq*I restriction fragment length polymorphism. The relative dosage of *C4A* and *C4B* genes can be determined by *Psh*AI-*Pvu*II RFLP. The C4A and C4B protein polymorphisms can be resolved based on gross differences in electric charge by high voltage agarose gel electrophoresis of plasma protein samples, which are digested by neuraminidase and carboxylpeptidase B, and revealed by immunofixation using polyclonal antibodies against human C4. Several previous studies on complement C4 in dermatomyositis were based on a two-locus model for C4A and C4B (references 30-32 of the manuscript), which led to data misinterpretation.

For protein allotypes, difference in band intensities for C4A and C4B protein in a allotyping gel reflect unequal expression levels of C4A and C4B that can be a result of unequal gene copy-number, lower expression of C4 protein from long C4 genes, and a potential gene mutation. In other words, unequal expression levels of C4A and C4B are

not necessarily caused by heterozygous deficiency. It is important to interpret C4 phenotype data with reference to *C4A* and *C4B* gene copy-number.

In genomic Southern blot analyses, in *Taq*I RFLP, the 7.0 kb, 6.4 kb, 6.0 kb and 5.4 kb fragments represent RP1-C4 long, RP1-C4 short, RP2-C4 long and RP2-C4 short, respectively. The *Taq*I RFLPs described above are due to the position of the *C4* genes in the MHC and the presence of the endogenous retrovirus HERV-K(C4). *Taq*I restriction fragments do not yield data on whether a *C4* gene code for a C4A or a C4B protein. The isotypic residues for C4A and C4B are encoded by exon 26, which is physically far from the intron 9 (with or without the endogenous retrovirus). Similarly, *Hind*III-RFLP did not reflect the presence or absence of a *C4A* or *C4B* gene. The 8.5 kb *Hind*III fragment corresponding to the 5' region of a *C4* gene reflects a short *C4* gene present in the first *C4* locus. There are examples of bimodular short-short (S-S) genotypes coding for a C4A protein and a C4B protein.

Figure S2 illustrates laboratory methods used to elucidate the genotypic and phenotypic diversities of complement C4 in eight JDM patients. Briefly, specific *C4* haplotypes were determined by long-range mapping by PFGE following *Pme*I digestion (Figure S2A). One to three copies of *C4* genes in an RCCX module on each Chromosome 6 were demonstrated. The GCNs of *C4* and details on long and short *C4* genes together with their neighboring genes *RP1* and *RP2* were revealed by Southern blot analyses using *Taq*I-digested genomic DNA (Figure S2B). Haplotypes with single short *C4* gene (S), two long *C4* genes (LL), one long and one short *C4* genes (LS), and three long genes (LLL) were observable. The relative copy-number of *C4A* and *C4B* genes in each subject was determined by *Psh*AI-*Pvu*II digested genomic DNA (Figure S2C). C4A and C4B protein polymorphisms were resolved by immunofixation of EDTA-plasma (Figure S2D).

For example, Patient No. 30 (first lane) had homozygous single short (S) *C4* genes with 107 kb *Pme*I fragment, which was confirmed by the presence of the 6.4 kb

*Taq*I fragment (Figure S2B) and the presence of single *C4B* gene and absence of *C4A* gene (Figure S2C). Immunofixation of plasma revealed the presence of C4B1 protein and a deficiency of C4A protein (Figure S2D).

Interpretations of C4 genotypes and phenotypes together with data of *HLA DRB1* genotypes for eight JDM patients are shown in Table S1. Such C4 genotyping and phenotyping strategies were performed for all 41 JDM patients recruited in Columbus. For 34 JDM patients recruited by NIH, alternative genotyping methods by TaqMan-based quantitative real-time PCR were applied.

**Table S1.** RCCX haplotypes (H1 and H2), complement *C4B* and *C4A* GCN and protein polymorphism, and HLA-*DRB1* haplotypes (H1 and H2) in eight JDM patients.

| **DM ID#** | **RCCX H1** | **RCCX H2** | ***C4B* GCN** | ***C4A* GCN** | **C4B protein** | **C4A protein** | ***DRB1** H1** | ***DRB1** H2** |
|---|---|---|---|---|---|---|---|---|
| 30 | S | S | 2 | 0 | B1,B1 | Q | 03 | 03 |
| 18 | LS | S | 2 | 1 | B1,B1 | A3 | 01 | 03 |
| 29 | LL | S | 2 | 1 | B1,B1 | A3 | 15 | 03 |
| 31 | LS | LS | 2 | 2 | B2B2 | A4,A2 | 04 | 08 |
| 03[†] | LS | LS | 1 | 3 | B**3**[¶] | A91[¶]A12[¶]A3 | 03 | 07 |
| 22 | LL | LL | 1 | 3 | B1 | A3,A3,A3 | 04 | 15 |
| 32 | LLL | LS | 0 | 5 | Q | A3,A3,A3,A3,A3 | 04 | 13 |
| 24 | LL | LS | 2 | 2 | B1,B1 | A3,A3 | 03 | 07 |

GCN, gene copy number; [†] DM03 was African American; H1, haplotype 1; H2, haplotype 2

[¶]B3, A91 and A12 had similar electrophoretic mobilities in an agarose typing gel.

## 2. Complement C4 and complement factor H plasma protein concentrations in JDM and controls.

Figure S3 shows a scatter-plot of normalized C4 protein levels plotted against GCN of total *C4* among JDM patients. The means of normalized C4 protein levels positively correlated with GCN of total *C4*. However, it is clear that there are large degrees of variations in protein levels within each GCN group. The median total C4 protein concentration in Columbus JDM patients (N=32) was 28.3 mg/dL (IQR: 22.4-35.2) and did not significantly differ from race-matched healthy controls from the same location (N=368) (30.1 mg/dL; IQR: 23.8-37.4) (Table S2). Likewise, we did not detect differences in C4A and C4B protein levels between JDM and controls.

Microarray data revealed transcript levels in peripheral blood for *CFH* were downregulated by 1.7-fold in JDM. We measured plasma protein levels of CFH in 31 JDM and 16 controls (Table S2). As expected, CFH protein levels were lower in JDM compared to controls [JDM: 55.3 mg/dL (IQR: 50.4-59.5 mg/dL); Control: 73.9 mg/dL (IQR: 68.8-84.9 mg/dL); p<0.0001].

**Table S2. A comparison of plasma protein levels for complement C4 and CFH between JDM and race matched controls**

| | JDM | Control | p-value | OR (95% CI) |
|---|---|---|---|---|
| *D. Median (IQR) plasma protein conc.* | | | | |
| *C4* protein, mg/dL | 28.3 (22.4-35.2) | 30.1 (23.8-37.4) | NS | |
| N | 32 | 368 | | |
| CFH protein, mg/dL | 55.3 (50.4-59.5) | 73.9 (68.8-84.9) | <0.0001 | |
| N | 31 | 16 | | |

**3. Global gene expression profiling of PAXgene blood RNA samples from JDM and controls.**

Tables S3 lists the patient demographics, complement *C4* and RCCX genetic background, *HLA-DRB1* genotypes, state of serum muscle levels and medications of 19 JDM patients subjected to global gene expression analyses using PAXgene peripheral blood RNA samples for Agilent microarrays. Different expression of 56 genes were identified and tabulated in Table S4. Among them are 24 upregulated genes and 32 down-regulated genes. Differential expression of transcripts (RNA) in blood samples reflects gene expression levels and the constituents of leukocyte populations present in samples (Figure S4 for interferon-stimulated genes IFI17 and IFI44 with polymorphs (PMNs), and Figure S5 for CCR5 and lymphocytes S5).

**Table S3. Patient demographics, *C4* CNV, *HLA-DRB1* genotypes, serum muscle enzymes and medications of 19 JDM patients (at recruitment) for microarray gene profiling experiments.**

| Microarray ID | M_ORDER | Age | Age (diagnosis) | Sex | race | Total C4 GCN | C4B GCN | C4A GCN | RCCX-H1 | RCCX-H2 | DRB1-H1 | DRB1-H2 | DR51/52/53 | Elevated Muscle Enzymes | Disease Status | Current Treatment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DM04** | 1 | 18.7 | 17.9 | F | W | 4 | 2 | 2 | LL | LL | 1 | 14 | | AST, CK | Active | MTX |
| **DM12** | 2 | 16.6 | 13.1 | F | W | 4 | 2 | 2 | LL | LS | 7 | 15 | 51 | Aldolase, AST, CK | Active | MTX |
| **DM18** | 3 | 17.3 | 17.3 | F | W | 4 | 2 | 2 | LS | LS | 4 | 8 | | Aldolase, AST | Active | MTX, Pred |
| **DM05** | 4 | 14.6 | 8 | M | W | 4 | 1 | 3 | LL | LL | 4 | 15 | 51 | None | Active | MTX |
| **DM10** | 5 | 9.5 | 6.3 | M | W | 3 | 2 | 1 | LS | S | 3 | 7 | 52 | None | Active | IVIG, MTX, Pred |
| **DM09** | 6 | 10.9 | 8.8 | F | B | 4 | 1 | 3 | LS | LS | 7 | 18 | 52 | None | Active | IVIG, Pred |
| **DM08** | 7 | 14.7 | 10.7 | M | W | 3 | 2 | 1 | LS | S | 3 | 7 | 52 | None | Flare | IVIG, MTX, Pred |
| **DM06** | 8 | 10.5 | 4.1 | F | W | 4 | 2 | 2 | LL | LS | 1 | 4 | | None | Remission | None |
| **DM16** | 9 | 15.7 | 14.1 | F | W | 3 | 1 | 2 | LL | S | 1 | 3 | 52 | n/a | Remission | MTX |
| **DM17** | 10 | 3.1 | 1.7 | F | W | 3 | 2 | 1 | LL | S | 3 | 11 | 52 | None | Remission | MTX |
| **DM03** | 11 | 12.6 | 9.6 | F | B | 4 | 2 | 2 | LL | LS | 9 | 11 | 52 | None | Remission | MTX |
| **DM01** | 12 | 6 | 2.9 | F | B | 4 | 0 | 4 | LL | LS | 13.1 | 18 | 52 | Aldolase | Active | MTX |
| **DM15** | 13 | 9.4 | 4.4 | M | W | 4 | 1 | 3 | LL | LL | 4 | 4 | | None | Remission | None |
| **DM19** | 14 | 12 | 5.6 | M | W | 4 | 3 | 1 | LSS | S | 3 | 7 | 52 | None | Remission | IVIG, Pred |
| **DM14** | 15 | 8.9 | 2.1 | F | W | 3 | 2 | 1 | LS | S | 1 | 3 | 52 | None | Remission | MTX |
| **DM02** | 16 | 13.8 | 6 | F | W | 4 | 2 | 2 | LL | SS | 3 | 13.1 | 52 | None | Active | IVIG, Dexamethasone |
| **DM07** | 17 | 10.6 | 3.7 | F | W | 3 | 2 | 1 | LS | S | 3 | 7 | 52 | None | Active | MTX |
| **DM13** | 18 | 7.7 | 4 | M | W | 3 | 2 | 1 | LL | S | 3 | 15 | 51 | Aldolase | Flare | MTX |
| **DM11** | 19 | 7.2 | 4 | M | W | 5 | 1 | 4 | LLL | LS | 1 | 1 | | none | Remission | MTX, Pred |

F, female; M, male; W, White; B, Black; RCCX: RP-C4-CYP21-TNX; GCN, gene copy number; H1, haplotype 1; H2, haplotype 2

AST, aspartate aminotransferase; CK, creatine kinase;

MTX, methotrexate; Pred, prednisone; IVIG, intravenous immunoglobulin IgG.

**Table S4. Differentially expressed genes in blood samples of JDM patients**

| Gene Symbol | Gene Name<br>*putative functions* | GenBank Accession | Mean Fold-Change* |
|---|---|---|---|
| **Interferon response - antiviral** | | | |
| IFI44 | interferon-induced protein 44 | NM_006417 | **2.9** |
| EPSTI1 | epithelial stromal interaction 1 (breast) | NM_033255 | **2.6** |
| USP18 | ubiquitin specific peptidase 18<br>*downregulates interferon responses* | NM_017414 | **2.1** |
| OAS3 | 2'-5'-oligoadenylate synthetase 3<br>*100kDa; binds and activates RNase L, inhibition of cellular protein synthesis and viral infection resistance* | NM_006187 | **2.2** |
| EIF2AK2 | eukaryotic translation initiation factor 2-alpha kinase 2<br>*inhibition; IFN-stimulated* | NM_001135652 | **2.1** |
| IFITM1 | interferon induced transmembrane protein 1 (9-27) | NM_003641 | **1.6** |
| ZC3HAV1 | zinc finger CCCH-type, antiviral 1<br>*prevents infection by retroviruses* | NM_024625 | **1.5** |
| TRIM5 | tripartite motif containing 5<br>*E3 ubiquitin-ligase and ubiqutinates itself, retroviral restriction* | NM_033092 | **1.5** |
| PYHIN1 | pyrin and HIN domain family, member 1;<br>*HIN-200 family of interferon-inducible proteins* | NM_198930 | **-1.6** |
| | | | |
| **Immune system** | | | |
| **B cell functions** | | | |
| VPREB1 | pre-B lymphocyte 1 | NM_007128 | **1.8** |
| FCRL1 | Fc receptor-like 1 | NM_001159397 | **1.6** |
| IL37 | interleukin 37<br>*IL-1 family, binds IL-1 receptor* | NM_014439 | **1.5** |

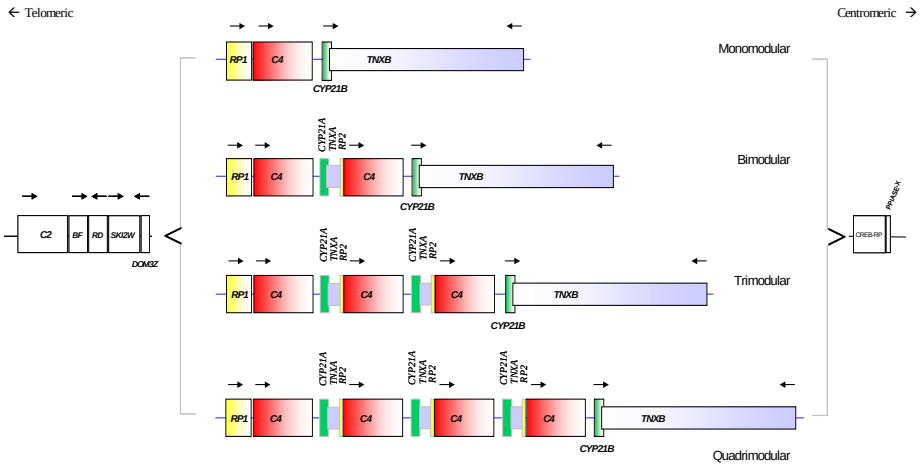
| | | | |
|---|---|---|---|
| **T cells functions** | | | |
| CD2 | CD2 molecule | NM_001767 | **-1.5** |
| TIGIT | T cell immunoreceptor with Ig and ITIM domains | NM_173799 | **-1.6** |
| ANKRD35 | ankyrin repeat domain 35 | NM_144698 | **-1.6** |
| MID2 | midline 2<br>*localizes to microtubular structures in the cytoplasm, TRIM member; antiviral* | NM_012216 | **-1.7** |
| WNT1 | wingless-type MMTV integration site family, member 1<br>*mesencephalon and cerebellum, cell signaling, embryogenesis; T cell development* | NM_005430 | **-1.7** |
| CTSW | cathepsin W<br>*regulation of T-cell cytolytic activity; T cells* | NM_001335 | **-1.7** |
| CTSZ | cathepsin Z<br>*lysosomal cysteine proteinase, carboxyl mono and di-peptidase activities* | NM_001336 | **-1.8** |
| EOMES | eomesodermin<br>*differentiation of effector CD8+ T cells ; essential during trophoblast development and gastrulation; T-box DNA binding protein* | NM_005442 | **-1.9** |
| RORC | RAR-related orphan receptor C<br>*nuclear hormone receptors; DNA binding lymphoid organogenesis ; inhibit the expression of Fas ligand and IL2* | NM_005060 | **-2.4** |
| GZMK | granzyme K<br>*granzyme 3; tryptase II* | NM_002104 | **-2.5** |
| **Innate immunity /chemokine receptors** | | | |
| CFH | complement factor H | NM_001014975 | **-1.7** |
| CCR5 | chemokine (C-C motif) receptor 5<br>*expressed on T cells and macrophages* | NM_000579 | **-2.3** |
| CXCR3 | chemokine (C-X-C motif) receptor 3;<br>*G protein-coupled receptor with selectivity for three chemokines* | NM_001504 | **-2.2** |
| CXCR6 | chemokine (C-X-C motif) receptor 6<br>*homing of T-cells to skin* | NM_006564 | **-2.5** |

| | | | |
|---|---|---|---|
| **Blood vessel functions** | | | |
| PHACTR1 | phosphatase and actin regulator 1<br>*blood vessel, endothelial cells; a key component in the angiogenic process* | ENST00000379350 | **1.7** |
| CYP4A11 | cytochrome P450, family 4, subfamily A, polypeptide 11<br>*coronary artery disease* | NM_000778 | **1.6** |
| ZNF831 | zinc finger protein 831<br>*blood pressure* | NM_178457 | **-1.5** |
| EMILIN1 | elastin microfibril interfacer 1<br>*blood vessels* | NM_007046 | **-1.5** |
| KLKB1 | kallikrein B, plasma (Fletcher factor) 1<br>*surface-dependent activation of blood coagulation, fibrinolysis, kinin generation and inflammation* | NM_000892 | **-1.6** |
| | | | |
| **Signaling pathways, membrane transports** | | | |
| XAF1 | XIAP associated factor 1<br>*apoptosis signaling* | NM_017523 | **2.1** |
| GNG11 | guanine nucleotide binding protein (G protein), gamma 11<br>*guanine nucleotide-binding protein* | NM_004126 | **1.9** |
| CDCA7 | cell division cycle associated 7<br>*c-Myc - phosphorylated by Akt-transformation; apoptosis* | NM_031942 | **1.6** |
| KIAA1958 | KIAA1958 | ENST00000374244 | **1.5** |
| PLEKHG3 | pleckstrin homology domain containing family G (with RhoGef domain) member 3;<br>*rhoGTPase exchange* | NM_015549 | **-1.5** |
| PERP | PERP, TP53 apoptosis effector<br>*p53 effector* | NM_022121 | **-1.7** |
| IGFBP3 | insulin-like growth factor binding protein 3<br>*proapoptotic and antiproliferative; TGF-β signaling -receptor* | NM_001013398 | **-1.8** |
| | | | |
| **Nervous system, synapses, $Ca^{+2}$, muscle functions** | | | |
| SNX22 | sorting nexin 22 | NM_024798 | **1.7** |
| BAALC | brain and acute leukemia, cytoplasmic | ENST00000297574 | **1.7** |
| GABRA3 | gamma-aminobutyric acid (GABA) A receptor, alpha 3 | NM_000808 | **1.7** |
| SAMD9L | sterile alpha motif domain containing 9-like<br>*mutated in familial tumoral calcinosis; cell proliferation, lost in mouse* | NM_152703 | **1.7** |
| ENKUR | enkurin, TRPC channel interacting protein | NM_145010 | **1.6** |

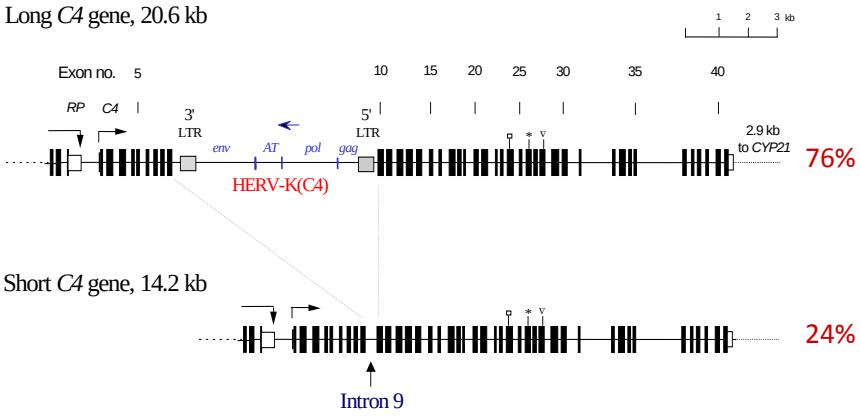
| | | | |
|---|---|---|---|
| | *signal transduction to Ca-channels* | | |
| ABCC3 | ATP-binding cassette, sub-family C (CFTR/MRP), member 3 | NM_001144070 | **1.6** |
| NGFRAP1 | nerve growth factor receptor (TNFRSF16) associated protein 1 | NM_014380 | **1.5** |
| PARD3 | par-3 partitioning defective 3 homolog (C. elegans)<br>*polarized cell growth, neural tube* | NM_001184792 | **1.5** |
| DLG3 | discs, large homolog 3 (Drosophila)<br>*excitoray synapses* | NM_021120 | **-1.5** |
| CAMK2N1 | calcium/calmodulin-dependent protein kinase II inhibitor 1<br>*negative regulation of protein kinase* | NM_018584 | **-1.5** |
| CACNA1I | calcium channel, voltage-dependent, T type, alpha 1I subunit | NM_021096 | **-1.5** |
| CACNA2D2 | calcium channel, voltage-dependent, alpha 2/delta subunit 2<br>*early infantile epileptic encephalopathy* | NM_001005505 | **-1.6** |
| TMED9 | transmembrane emp24 protein transport domain containing 9<br>*Golgi and ER* | NM_017510 | **-1.7** |
| LRFN3 | leucine rich repeat and fibronectin type III domain containing 3<br>*adhesion, synapses* | NM_024509 | **-1.7** |
| CLIC5 | chloride intracellular channel 5<br>*associates with actin-based cytoskeletal structures; hair cell stereocilia formation, myoblast proliferation and glomerular podocyte and endothelial cell maintenance* | NM_016929 | **-1.8** |
| ATXN7L2 | ataxin 7-like 2 | NM_153340 | **-1.9** |
| SLC4A10 | solute carrier family 4, sodium bicarbonate transporter, member 10<br>*$HCO3^-$, $Cl^-$ transport; regulates pH* | NM_022058 | **-2.3** |
| | | | |
| **Other** | | | |
| LOC144571 | uncharacterized LOC144571 | NR_026971 | **-1.8** |

* from 19 JDM patients and 7 healthy controls.

A.




B.




C.

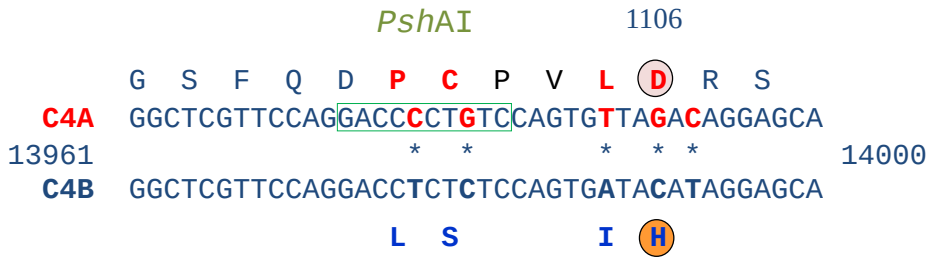



**Figure S1**. Genetic diversity of complement C4. **A**. Copy-number variation (CNV) of *RP-C4-CYP21-TNX* (RCCX) modules in the class III region of the HLA. **B**. Gene size dichotomy of long and short *C4* genes. Exon-intron structures of long and short C4 genes are shown. The long gene is due to endogenous retrovirus integrated into intron 9 of the *C4* gene. **C**. DNA sequences and derived amino acid sequences determining C4A and C4B. C4A and C4B isotypic sequences are present in Exon 26 of the *C4* genes. The long/short *C4* genes and *C4A/C4B* genes are independent structural features of *C4* genes that can be distinguished by *Taq*I RFLP and *Psh*AI RFLP, respectively.

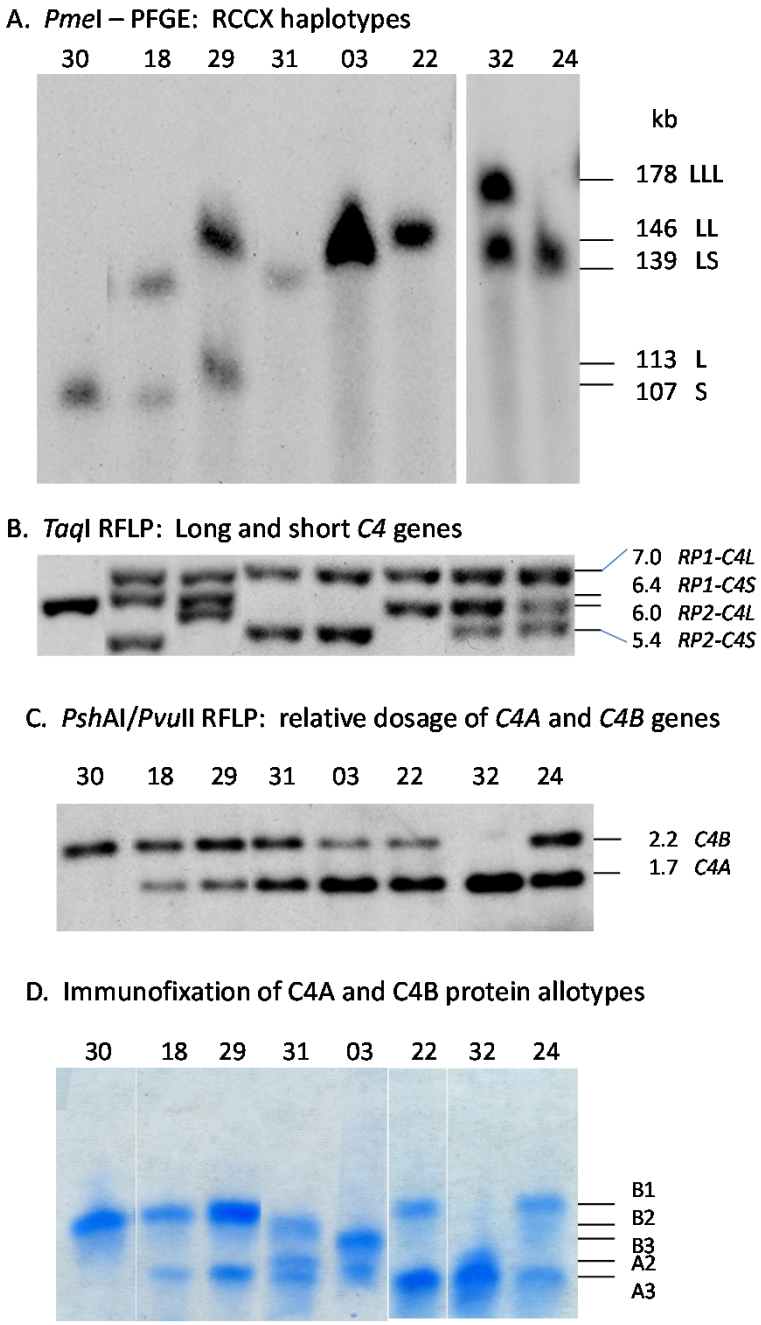



**Figure S2.** Variations of *C4* haplotypes and gene copy-numbers (GCNs) of total *C4*, *C4A* and *C4B*, and C4 protein polymorphisms in eight selected JDM subjects. **A**. *C4* haplotype structures were determined by PFGE *PmeI*-digested genomic DNA and hybridization to a *C4*d-specific probe. **B**. Total *C4* copy-number and gene sizes were determined by *Taq*I genomic RFLP. **C**. Relative dosages of *C4A* and *C4B* were determined by *PshA*I-*Pvu*II RFLP. **D**. C4A and C4B protein polymorphisms were revealed by immunofixation of EDTA-plasma resolved by high-voltage agarose gel electrophoresis.

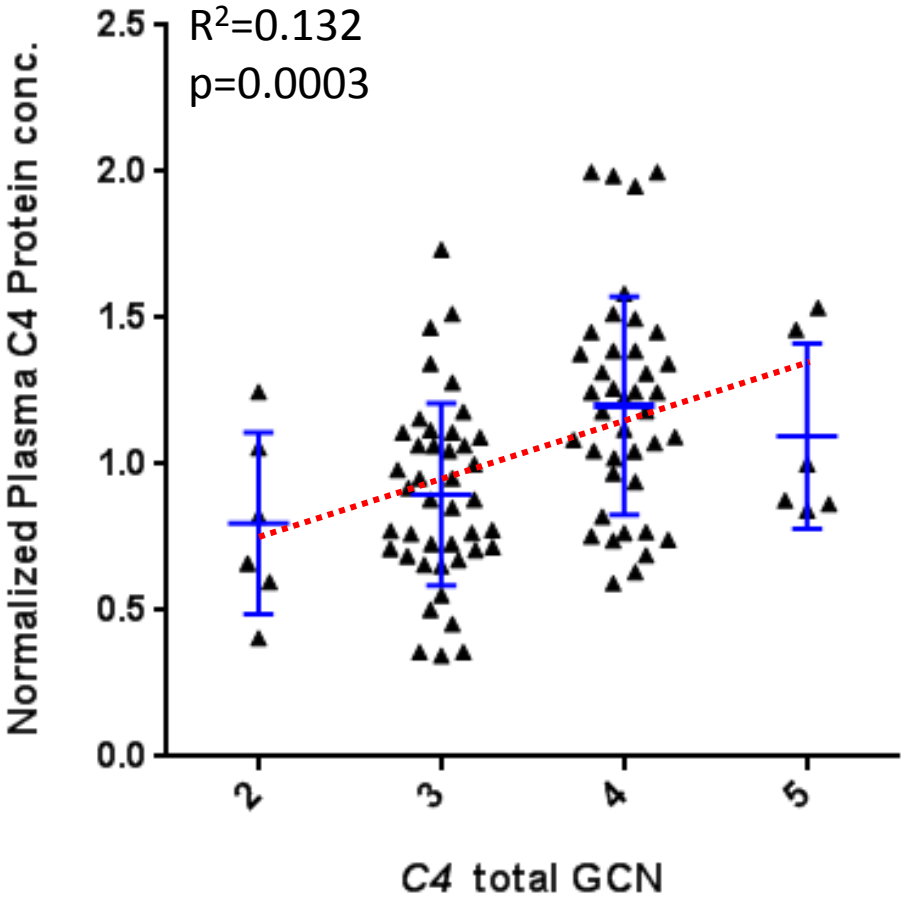



**Figure S3**. Relationship between C4 plasma protein levels plotted versus total *C4* GCN in 94 White JDM patients; C4 plasma levels are normalized and graphed as fold changes, which were calculated as the fold change compared to the mean plasma level of either the Columbus plasma samples or NIH plasma samples GCN, gene copy-number. Red dotted line represents the trend line.

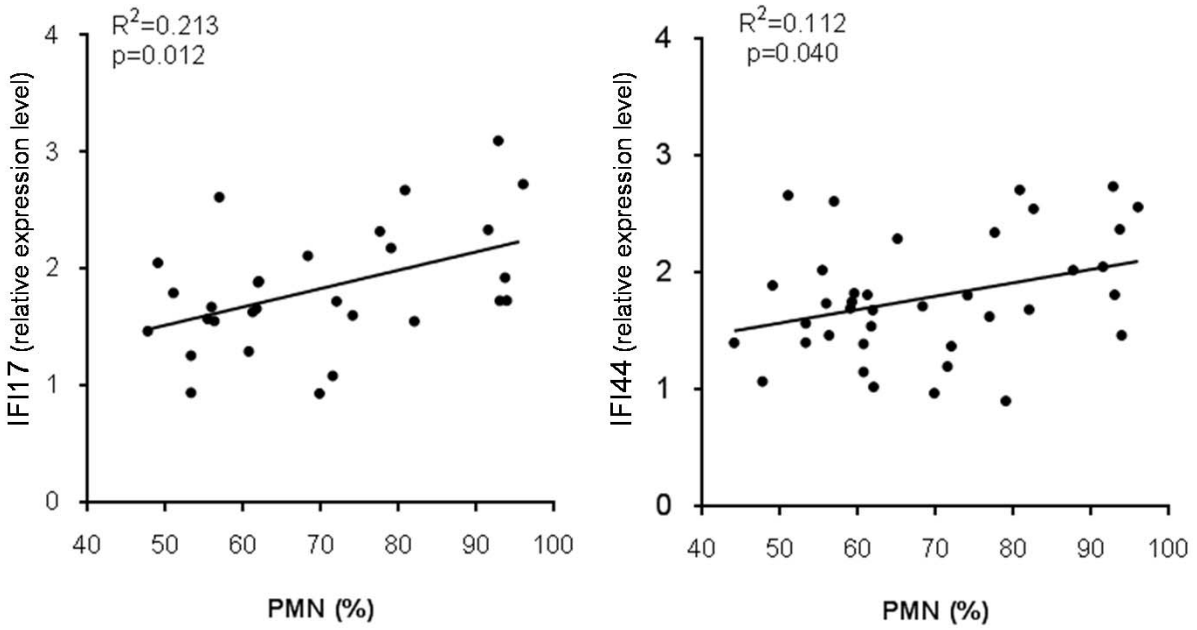



**Figure S4**. Correlations of transcript levels for interferon-stimulated genes IFI17 (left) and IFI44 (right) with polymorphonuclear cells present in leukocytes of JDM blood samples.

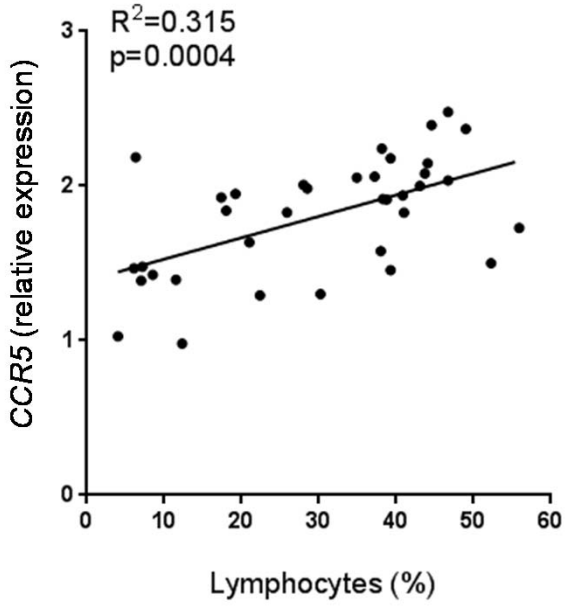



**Figure S5**. Correlation of CCR5 transcripts with lymphocytes present in leukocytes of JDM blood samples