[Frontiers in Bioscience 6, d927-935, August 1, 2001]

FEATURES OF THE TWO GENE PAIRS *RD-SKI2W* AND *DOM3Z-RP1* LOCATED BETWEEN COMPLEMENT COMPONENT GENES FACTOR *B* AND *C4* AT THE MHC CLASS III REGION

Zhenyu Yang, Xiaodong Qu and C. Yung Yu

Children's Research Institute, 700 Children's Drive, Columbus, Ohio 43205-2664, and Department of Pediatrics, and The Molecular, Cellular and Developmental Biology Graduate Program, The Ohio State University, Columbus Ohio 43210

TABLE OF CONTENTS

1. ABSTRACT

Located at the 30 kb genomic region between complement factor *B* and component *C4* are four ubiquitously expressed genes *RD*, *SKI2W*, *DOM3Z* and *RP1*. Besides RP1, the protein products of the other three genes each has highly conserved homologues or related proteins in lower eukaryotes, contains leucine zipper motifs for protein interaction, and plays important roles related to RNA metabolism. RD is a subunit of the negative transcription elongation factor, critical for the regulation of gene expression. It has an RNA recognition motif and 24 copies of Arg-Asp (RD) repeats. Ski2w is a nucleolar and cytoplasmic protein that has a putative RNA helicase domain. Fusion proteins of human Ski2w expressed in insect cells and bacteria have ATPase activity. The cytoplasmic protein of human Ski2w is associated with the polysomes and probably the 40S subunit of ribosomes. Ski2w is probably involved in the regulation of translation and RNA turnover. Dom3z is a nuclear protein whose yeast homologue forms a complex with an exoribonuclease. RP1 (or STK19) is a Ser/Thr nuclear protein kinase. No homologues of RP1 in lower eukaryotes have been discovered. Six polymorphic residues are present in human Ski2w and two in Dom3z. The potential roles of Ski2w and Dom3z on the clearance of degraded nuclear and cytoplasmic RNA raised their possibilities as susceptibility genes of systemic lupus erythematosus that is a disease with flawed processes in the removal of apoptotic materials.

2. INTRODUCTION

In October 1999, the complete sequence and gene map of a human major histocompatibility complex (MHC) was reported. This sequence is 3,838,986 bp in size. It contains 224 gene loci, of which 128 are predicted to be expressed (1). Many autoimmune, genetic, malignant and multifactorial disorders are associated with the MHC (2-4). About 60% of the expressed genes in the human MHC have *no* obvious function related to the immune system. Most of these so-called "non-HLA" genes are located in the class III region.

The MHC complement gene cluster (MCGC, Figure 1) is located 115 kb downstream of the *NOTCH4* gene, the centromeric outpost of the class III region. The defining feature of the MCGC is the four complement component genes, *C2*, *BF*, *C4A* and *C4B*, that code for subunit proteins of the complement C3 and C5 convertases. These proteins basically form the engines of the complement system and drive the three activation pathways. *C2-BF* and *C4A-C4B* are the results of two gene duplications from different ancestral genes. In both human and mouse, *BF* and *C4* are separated by a ~30 kb genomic region. This region contains a complex of four genes, *RD*, *SKI2W*, *DOM3Z* and *RP1*. These four genes are ubiquitously expressed and organized as two head-to-head gene pairs with minimal intergenic regions (1, 5-7). They are involved in the regulation of gene expression or nuclear protein phosphorylation. Characterization of these four genes are of significance because of their intriguing gene

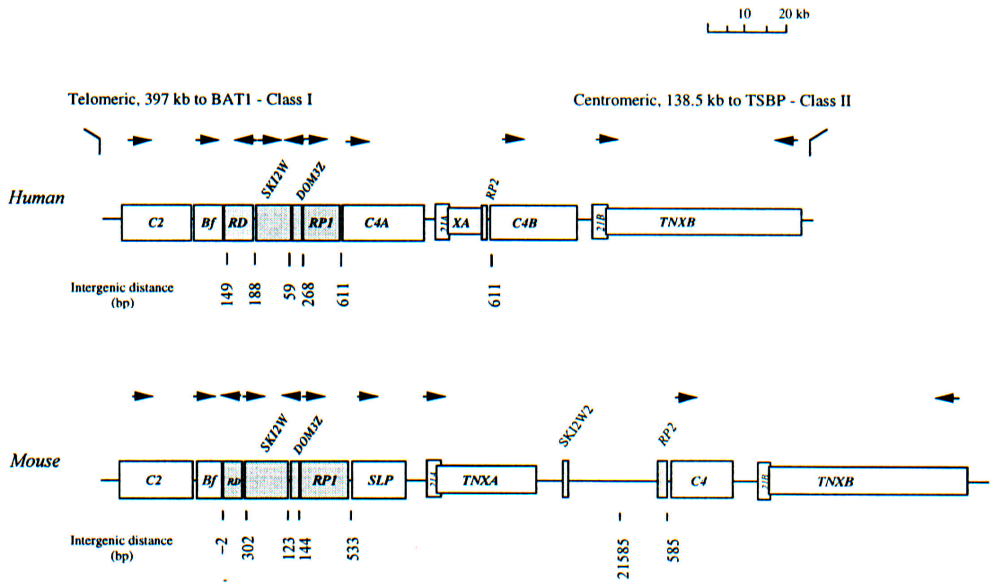

Figure 1. Gene organizations of the human and mouse MHC complement gene clusters. The genes for *RD*, *SKI2W*, *DOM3Z* and *RP1* are shaded. Arrows represents transcriptional orientations of structural genes (Modified from Ref. 20).

organizations, essential cellular functions, and their potential associations with many diseases linked to complement *C2* and *C4*. Here we update our knowledge of these four recently discovered genes.

3. SRUCTURAL AND GENETIC PROPERTIES OF RD, SKI2W, DOM3Z AND RP1

3.1. RD

The RD gene is 6.7 kb in size. It consists of 11 exons and is organized in a tail-to-tail configuration with BF (Figure 2). The intergenic distance between the 3' ends of RD and BF is 205 nucleotides. The RD protein has 380 amino acids and is 46 kDa in size (8-11). RD protein contains an RNA recognition motif (RRM) of the ribonucleoprotein family that is also present in poly(A) binding proteins and in nucleolysin (12-14). This RRM is present between residues 264-327 and is encoded by exons 8-10. The signature feature of RD, however, is the presence of 24 tandem repeats of positively and negatively charged amino acids Arg-Asp (RD) (8). This Arg-Asp sequence is encoded by exon 7, which is a symmetric phase 1-1 exon. (A symmetric exon may be acquired during evolution without shifting the gene's reading frame.) Arg-rich tract is another feature of nuclear RNA-binding proteins. This central region with basic-acidic repeating residues is similar to, and more strictly alternating than those seen in the 70 kDa protein of the U1 small nuclear ribonucleoprotein. The structural motifs suggest a possible function related to RNA splicing. However, this postulation was challenged as it was demonstrated that RD is a subunit of the negative elongation factor (NELF), which represses the transcriptional elongation by RNA polymerase II (RNAPII) (15). The elongation step of transcription is one of the critical processes for regulation of gene expression (16). During the transition from transcriptional initiation to elongation, the carboxyl terminal domain (CTD) of the RNAPII becomes extensively phosphorylated and remains so for successful completion of transcription. Phosphorylation of the CTD facilitates the release of negative elongation factors NELF and DSIF (17) from RNAPII. RD is the smallest of the 5-polypeptide complex constituting the NELF. RD alone does not suffice for NELF activity, suggesting that other subunits are essential for its function. It is possible that through the RRM and/or the Arg-rich tract, RD binds to RNA; while through the leucine-zipper motif at the N-terminal region, RD interacts with other subunits of the NELF.

The high sequence repetition of RD may predispose the gene to frequent mutations. Indeed, it was reported that in a group of 107 subjects, 3.3% carried 22 or 23 copies of RD dipeptide repeats. However, no difference was found in the frequency of such polymorphism between normal individuals and systemic lupus erythematosus patients (18).

It is worthwhile to note that in mouse, a targeted disruption of BF resulted in the loss of expression of the downstream RD gene (19). In the mouse MHC, the 3' ends of BF and RD overlap by 2 bp (20). It is possible that some of the regulatory sequences for RD gene expression may reside in its neighboring genes BF and SKI2W.

In Drosophila a protein termed Anon (21) has extensive sequence similarity (49-54%) to human RD. However, Anon is 100 amino acids smaller than human RD. It appears that the sequence encoded by exon 7 of the human gene that include the distinct RD sequence is absent in the Drosophila protein, suggesting that the mammalian gene might have gained the RD-feature by acquisition of exon 7. The function of Anon is unknown but insertion of the transposon P-element to the joint promoter region of Anon and ribosomal protein gene RPL14 led to a "Minute" phenotype (21).

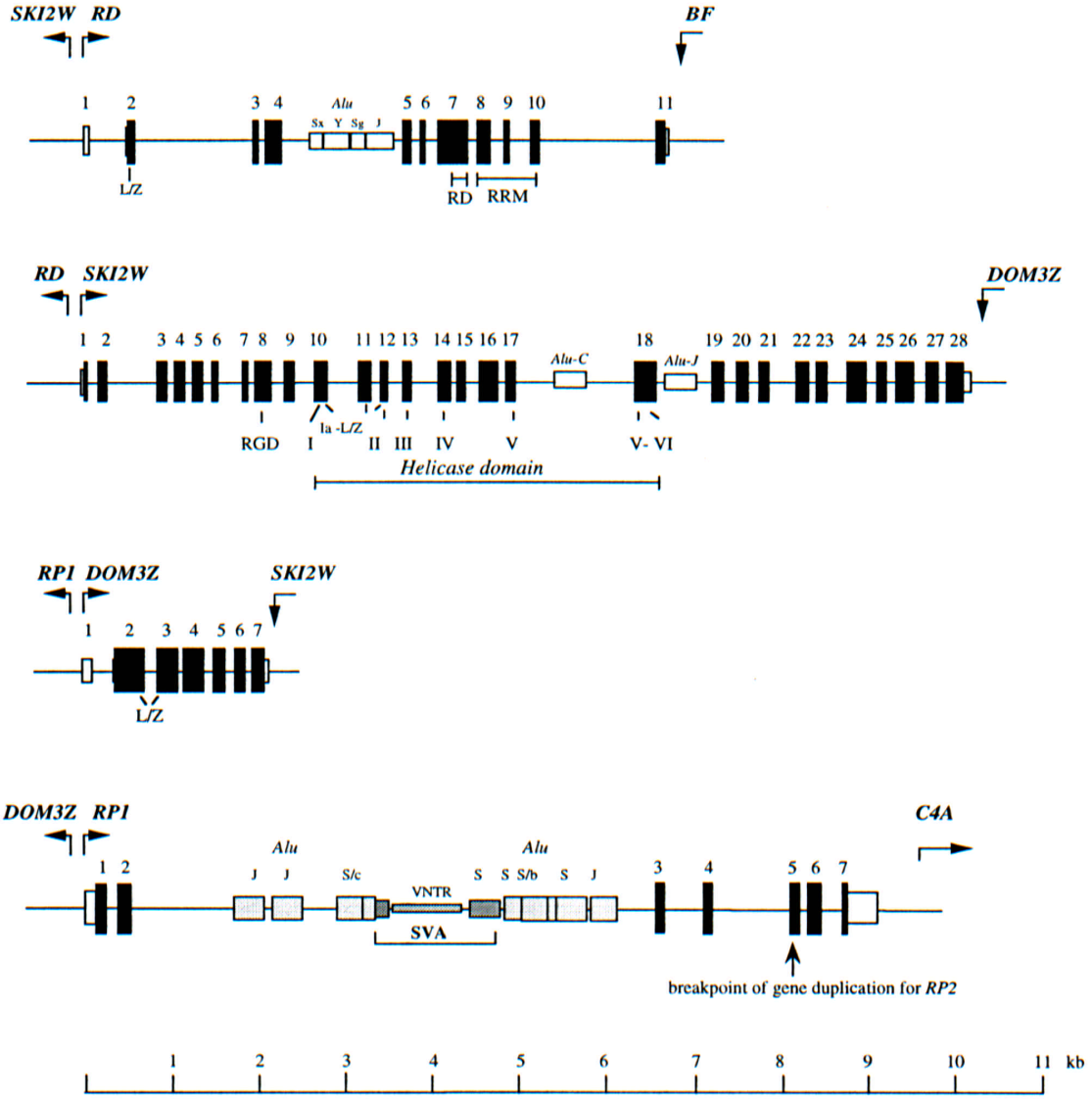

Figure 2. The exon-intron structures of *RD*, *SKI2W*, *DOM3Z* and *RP1*. Coding exons are in solid boxes; 5' and 3' untranslated exons are in empty boxes. The locations for repetitive elements Alu and retrotransposon SVA are shown. Horizontal arrows show the transcriptional orientations. A vertical, downward arrow represents the transcriptional termination site of the neighboring gene. L/Z, leucine zipper; RD, Arg-Asp motif; RRM, RNA recognition motif; RGD, cell adhesion motif. Modified from Refs. 9, 18, 20, 29, 30, 36 and 37.

3.2. SKI2W

Only 171 bp away from the 5' end of *RD* is *SKI2W* (Figure 2). These two genes are arranged in a head-to-head configuration. The human *SKI2W* gene spans 11 kb and contains 28 exons. The 4-kb transcript encodes a polypeptide of 1,246 amino acids. The Ski2w protein is a putative RNA helicase with a DEVH-box. For many helicases, hydrolysis of ATP is required as an energy source to unwind helical structures of nucleic acids (22). Indeed, ATPase/GTPase activity was demonstrated from human Ski2w fusion proteins produced in bacteria and in insect cells (23). The putative helicase activity in human Ski2w, however, has not been shown.

The human Ski2w contains two consecutive leucine zipper motifs that might be involved in protein-protein interaction. There is also an RGD motif that could be a ligand for cell adhesion molecules. The human Ski2w protein shares a striking and extensive sequence similarity with the yeast antiviral protein Ski2p. The yeast *ski2* mutants manifest a superkilling phenotype by RNA viruses that is characterized by increased copy number of the viruses and inability of the yeast to proliferate at a temperature below 8^0C. Genetic experiments revealed that *ski2* mutants were unable to repress the translation of poly(A)$^-$ and cap$^-$ (or decapped) cytoplasmic and viral RNAs (24-26). Mutants of *SKI3*, *SKI8* and *SKI6* in

yeast also lead to the superkilling phenotype by yeast RNA viruses. Ski2p, Ski3p and Ski8p are required for the normal 3' to 5' mRNA decay and for the suppression of poly(A)$^-$ RNA translation. Recently it is shown that Ski2p forms a heterotrimeric complex with Ski3p and Ski8p in an equimolar stoichiometry (27). Ski6p is one of the exoribonucleases of the RNA turnover complex termed the exosome (28). Shortening of poly(A) and decapping of mRNA are important processes for the degradation of cellular RNAs. Therefore, it appears that the yeast Ski proteins are involved in a pathway of RNA turnover and/or regulation of translation. It will be of interest to determine if the similar pathway is present in humans.

Immunoblot and indirect immunofluorescence experiments of HeLa cells using polyclonal antisera showed that the endogenous human Ski2w is approximately 140 kDa in size and present in the nucleolus and in the cytoplasm. This result is confirmed by transient expression experiments of CHO cell transfectants of human Ski2w fusion protein tagged with an epitope recognizable by a commercial monoclonal antibody. Ribosomal profile experiments coupled with immunoblot analysis further revealed that in the cytoplasm of HeLa cell extracts, Ski2w is associated with polysomes and probably the 40S subunit of ribosomes. The association of human Ski2w with ribosomes is not mRNA dependent because it is not abrogated by RNase A treatment. Since this association is not disrupted by 0.5M KCl treatment, it is suggested that Ski2w has a high affinity to the ribosomes. In essence, Ski2w is present at the machinery of protein synthesis and ribosome biogenesis, suggesting that human Ski2w shares similar functional properties with its yeast homolog. It is proposed that human Ski2w would play a role in linking the machinery for protein synthesis (i.e. polysomes) and RNA turnover (i.e. exosomes). This would fulfil a crucial step in the temporal control of gene expression, i.e, to facilitate the degradation of mRNAs and to prohibit the degrading/degraded mRNAs from being translated (29).

Analysis of the derived amino acid sequences from a cDNA and three independent genomic DNA sequences of Ski2w reveals six polymorphic residues in the protein with 1246 amino acids (30, 31). The polymorphic sites are A5E, R151Q, L214M, V917M, F1052L, A1071V. Among them the first two substitutions are non-conservative changes and might affect the protein function or structure. It is also of interest to point out that these changes are located at regions flanking the helicase domain.

In the mouse MHC class III region, *SKI2W* is partially duplicated (Figure 1), which is not observed in the human MCGC (20), suggesting that the human and mouse MCGC might have undergone independent secondary gene duplication and rearrangement events.

Homologues of Ski2w are probably present in all eukaryotes. An alignment of Ski2w amino acid sequences from *Homo sapiens (*Hosa*), Schizosaccharomyces pombe (*Scpo*), Saccharomyces cerevisiae (*Sace*), Caenorhabditis elegans* (Cael) and *Arabidopsis thaliana* (Arth) revealed that the most conserved sequences are located around the putative RNA helicase domains and at the carboxyl termini (Figure 3).

3.3. DOM3Z

Downstream of *SKI2W* is *DOM3Z*. These two genes are arranged in tail-to-tail configuration and the intergenic region between them is only 59 bp (Figure 2). The full-length *DOM3Z* cDNA is 1386 bp in size and consists of 7 exons (30). The 5' RACE results and sequence analyses revealed that *DOM3Z* transcripts exhibit three groups of splice variants at the 5' region. The reading frames for all transcripts remain identical because of an in-frame stop codon (20). Sequence comparisons revealed that human and mouse *DOM3Z* have identical gene structures with similar exon and intron sizes. The coding regions of human and mouse *DOM3Z* cDNAs are 86.6% identical. Like human *DOM3Z*, the 5' region of the mouse cDNA is heterogeneous with three groups of variants, all of which share the same initiation codon. Therefore, these transcripts encode for an identical protein. However, differences in the 5' untranslated regions of the mRNAs may have variable translational efficiency for the Dom3z protein. The Dom3z protein contains 396 amino acids, of which 11.6% are proline. Close to the amino terminus is a leucine zipper motif that might be involved in protein interaction. Two polymorphic residues have been detected in human Dom3z, which are S28T and H261Q.

The function of human Dom3z is unknown. Similar to Ski2w, related proteins in *Schizosaccharomyces pombe* (fission yeast), *Saccharomyces cerevisiae* (baker's yeast), *Caenorhabditis elegans* (a worm) and *Arabidopsis thaliana* (a flowering plant) share significant sequence similarities with human Dom3z. It is worthwhile to mention that the sequence similarity of the human Dom3z protein to the related protein in fission yeast (52.4%) is as close as that between the two yeast species (53.3%). The conservation of these proteins implies a fundamental function in many organisms (30). The yeast homolog of Dom3z, Rai1p, interacts with Rat1p. Rat1p is a nuclear 5' to 3' exoribonuclease required for RNA turnover (32). Human Dom3z could have a similar function. It is therefore of particular interest to note that the human Dom3z has an *E. coli* RNase PH signature motif close to its N-terminal region (i.e, kTevAepRNKLpRpApt, from residues 22-39 of Dom3z, conserved amino acid residues in the motif in upper case) (33). In *C. elegans*, Dom-3 is located in the same operon together with Mes-3. Mes-3 encodes a maternally supplied product that is required for proliferation of germ cells and for maintenance of viable germ cells that are competent to differentiate into gametes (34). Dom-3 and Mes-3 show parallel expression patterns and it is suggested that Dom-3 could have related function to Mes-3. In *C. elegans* mechanisms that repress the production of mRNA appear to be essential to maintain germ cell fate and viability (35). The predicted role of Dom3z in exoribonuclease function is consistent with this scheme. The higher levels of human *DOM3Z* transcripts detectable in the testis and ovary (Figure 4) would suggest a possible role in reproduction.

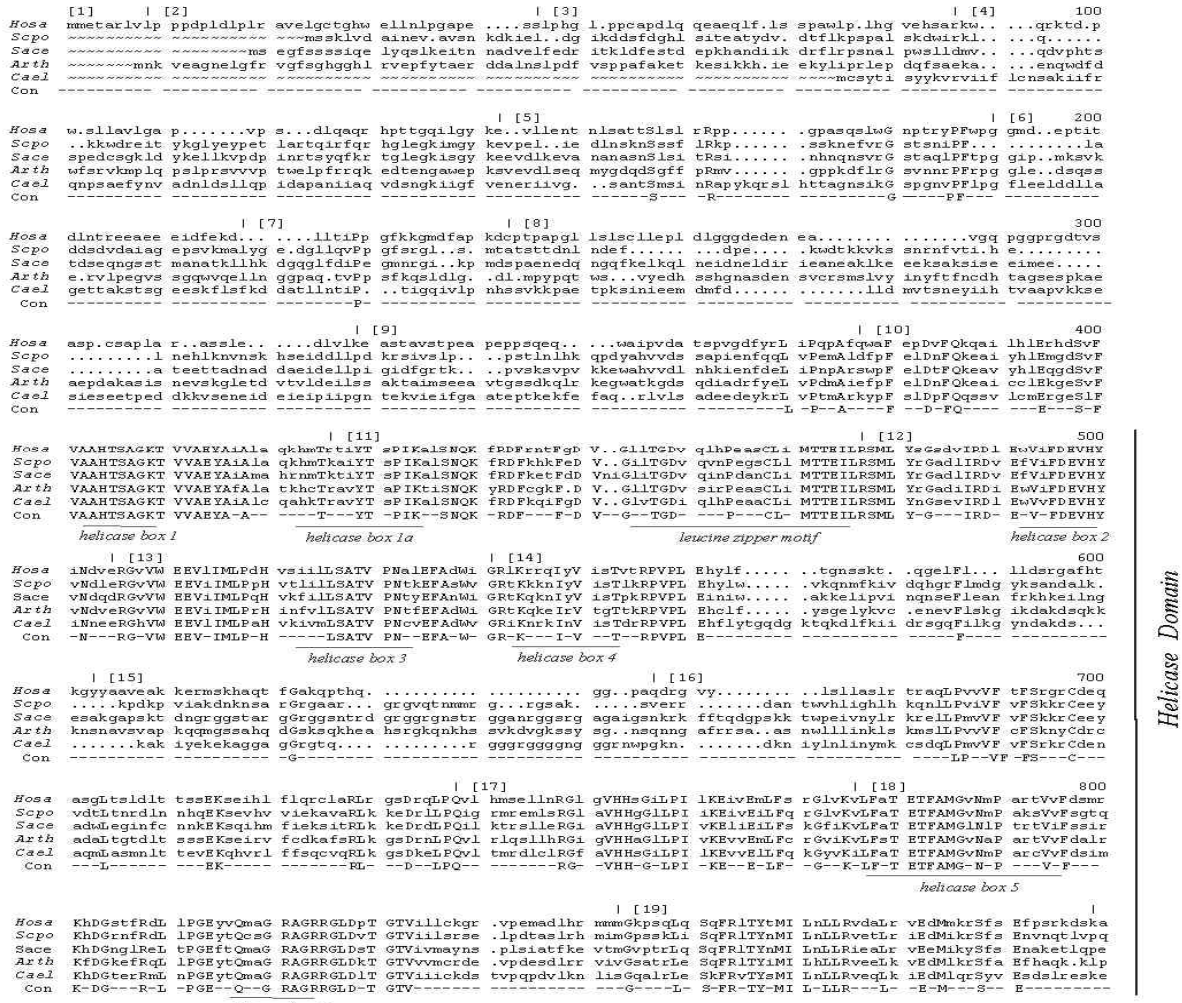

Figure 3. An alignment of the Ski2w protein sequences from *Homo sapiens* (Hosa, accession no. S56752), Schizzosaccharomyces *pombe* (Scpo, accession no. T41378), *Saccharomyces cerevisae* (Sace, accession no. S55954), *Arabdiopsis thaliana* (Arth, accession no. CAB61942) and *Caenorhabditis elegans* (Cael, accession no. T16755). The amino acid sequences conserved among all species are in upper case and shown in the consensus (Con). Sequences not conserved among all species are shown as a dash (-). A dot (.) indicates a gap in sequence. The locations of the exon-intron boundaries for human *SKI2W* gene are marked vertical strokes and the exon numbers indicated in square brackets.

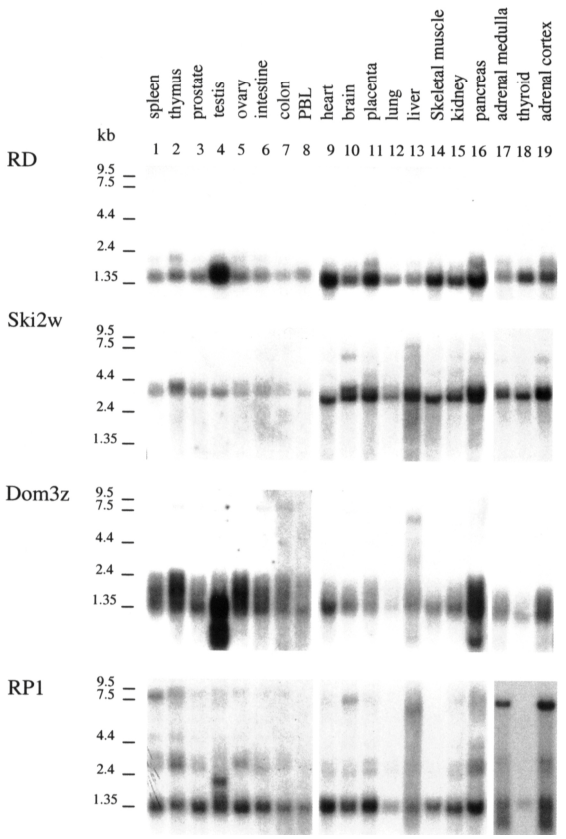

Figure 4. Northern blot analyses of human *RD, SKI2W, DOM3Z* and *RP1*. Multiple tissue blots (Clontech, CA) with poly(A)$^{?}$RNA were used (modified from Refs. 29 and 30).

3.4. RP1 (G11/STK19)

Located at the 5' region of the *DOM3Z* gene is the *RP1* gene. *DOM3Z* and *RP1* are arranged in head-to-head configuration (Figure 2). The *RP1* gene contains 7 or 9 exons (20, 36). The 9-exon *RP1* gene codes for 364 amino acids and is 11 kb in size, which overlaps to the 5' region of the *DOM3Z* gene. The calculated molecular weight for this larger form is 41.5 kDa. The 7-exon gene is 9.1 kb in size that has a 267 bp intergenic distance with *DOM3Z*. It codes for 254 amino acids. The calculated molecular weight for this smaller form is 30 kDa. The first 110 amino acid residues encoded by the first two exons are absent in the shorter *RP1* transcripts (generated from the 7-exon gene). Moreover, these two 5' exons are not conserved in the mouse *RP1*. Therefore, it is presumed that the 7-exon *RP1* gene is the prevalent or "ancestral" form and therefore its numbering of exons and amino acids are adopted, and referred to hereafter.

A splice variant at the 3' end of exon 3 has been described. This splice variant would result in an extension of 4 amino acid residues after amino acid residue 112, Cys-Asp-Cys-Val (37). No homologues of RP1 have been found in lower eukaryotes including the yeast, worm and fly (36). There is a bipartite nuclear localization signal close to the amino terminal of the human protein sequence. Transient expression of the larger form of the RP1 (with a tag) in monkey kidney cell COS7 showed the localized expression in the nucleus (38).

RP1 shares limited sequence similarity with the tyrosine kinase transforming protein from Fujinami virus (37). The human RP1 protein was produced in insect cells via a recombinant baculovirus vector and purified by immunoprecipitation. The recombinant protein could be covalently modified by the reactive ATP analogue 5'p-fluorosulfonylbenoyladenosine (FSBA). It could also phosphorylate the serine and threonine residues of α-casein, and serine residues of histone (38). Site directed mutagenesis of the lysine residues close to the carboxyl terminal region of RP1 either completely abrogated or greatly reduced the kinase activity. Therefore, it was concluded that RP1 is a novel nuclear Ser/Thr protein kinase. The optimal kinase activity was suggested to be at 30^{0}C in the presence of 20 mM Mn^{+2} (38). However, the physiological substrates and properties of the endogenous RP1 remain to be shown. It is of particular interest to determine if RP1 would interact/cooperate with RD to control the transcription elongation process through phosphorylation of the CTD of RNAPII, and if RP1 would modify the properties of Dom3z or Ski2w in the RNA turnover process.

In humans, *RP* gene is duplicated in a modular fashion with complement *C4*, steroid *CYP21* and tenascin *TNX* to form the RCCX module (Figure 1). Similar to the human gene, mouse *RP* is also partially duplicated and *RP2* is unlikely to code for a functional protein product. However, the mouse *RP1* significantly differs from the human gene in intron 4, where an endogenous retrovirus termed the "imposon" (IMP) is present. This endogenous retrovirus was discovered while the mechanism leading to androgen-dependent expression of the sex-limited protein (Slp) was investigated (39). The 5' LTRs of the IMP are shown to contain glucocorticoid responsive elements that confer the sex limited expression of SLP in male mice (40, 41). Whether the LTRs would bestow a differential expression of *RP1* and/or *DOM3Z* in male and female mice has yet to be determined. There is no trace of IMP integration in the partially duplicated gene *RP2*, suggesting that the *RP* gene duplication in mouse occurred prior to the integration of IMP to *RP1*. The IMP is configured in the reverse transcriptional orientation with respect to that of mouse *RP1*.

4. COMMON FEATURES OF THE TWO GENE PAIRS

The two tandem gene pairs *RD-SKI2W* and *DOM3Z-RP1* are both organized in head-to-head configurations. While GC dinucleotides are generally under-represented in the mammalian genome, there are numerous copies of GC dinucleotides at the intergenic regions between the *RP1/DOM3Z* genes and between the *RD/SKI2W* genes. Moreover, pulsed field gel electrophoresis of genomic DNA digested with *Bss*H II and Southern blot analysis indicated that the GC sequences at the promoter regions of these gene pairs are probably unmethylated or hypomethylated (30). Hypomethylation of GC sequences is an indication of active transcription (42).

Indeed, Northern blot analyses revealed that *RD-SKI2W-DOM3Z-RP1* are ubiquitously expressed. All four genes are expressed at significantly higher levels in testis and pancreas (Figure 4). Relatively lower expression levels are observed in the lung. Detection of multiple transcripts in most tissues is a common feature for the *DOM3Z-RP1* gene pair. Instead of the consensus TATA boxes, there are multiple SP1 and AP1 sites for bindings of ubiquitous gene expression factors at the 5' regulatory regions of these two gene pairs. The structural features and ubiquitous expression patterns suggest that these four genes are probably housekeeping genes. These four genes are driven by two sets of bi-directional promoters. Indeed it has been shown that a 262 bp sequence proximal to human *SKI2W* (DDX13) is sufficient for concurrent expression in both directions in a transient expression assay using chloramphenicol acetyltransferase as a reporter gene (43).

The very short intergenic distances among these four genes can again reflect the extremely high gene density in the MHC class III region. The very close gene placements would imply the presence of regulatory sequences of a gene in its neighboring genes. Therefore, the disruption of one gene might affect the expression its neighbors. It would be of interest to investigate whether the expression of these four genes are coordinated or controlled by similar *trans*-acting transcriptional factors. Investigation on the bi-directionality of the 5' regulatory regions for *RD-SKI2W*, and for *DOM3Z-RP1* would yield relevant information on the control of expression of these genes.

The protein sequences for RD, Ski2w and Dom3z all contain leucine zipper motifs involved in protein interactions. The tight linkage of *RD*, *SKI2W*, *DOM3Z* and *RP*, the ubiquitous gene expression patterns imply that these proteins may have concerted functions. For example, it appears that Dom3z and Ski2w both would play critical roles in RNA turnover. It is plausible that Ski2w is involved in a cytoplasmic exoribonuclease pathway (29), while Dom3z is involved in a nuclear exoribonuclease pathway. It cannot be overemphasized that RNA turnover is one of the most important processes to achieve the regulation of gene expression.

5. THE POTENTIAL ASSOCIATION WITH AUTOIMMUNE DISEASE

It has been inferred that systemic lupus erythematosus (SLE) susceptibility genes may be present in linkage disequilibrium with HLA class II or complement *C4* genes (44, 45). Defects in the physiological mechanisms for the removal of dying cells may promote the disease susceptibility to SLE (46). In keeping with this theory, it has been shown that ablation of the *DNase I* gene in mouse results in the development of anti-chromatin autoimmunity and glomerulonephritis (47). This is particularly illuminating because increased liberation or disturbed clearance of DNA-protein and ribonucleoprotein complexes after cell-death may initiate and propagate lupus (48, 49). The novel genes *SKI2W* and *DOM3Z* located at the MCGC are particularly attractive candidate genes not only because of their physical location but also of their basic cellular functions related to the degradation of RNA in the nucleus and in the cytoplasm. These two proteins could have a role on the clearance of materials from the apoptotic and necrotic cells similar to the DNase I (46, 47). Interestingly, Ski2w, Dom3z and DNase I are all highly expressed in the pancreas.

6. CONCLUSION AND PERSPECTIVE

Sequencing data and structural characterizations have yielded clues for functional and genetic studies of novel genes. In depth biochemical characterizations of RD, Ski2w, Dom3z and RP proteins, determination of the gene expression patterns in healthy and in disease stages, mutational analysis of these genes in normal individuals and in MHC-associated disease patients, temporal or spatial knockout of these proteins in model organisms and in cell lines, will help elucidating the physiological roles of these and many other novel genes discovered by the Genome Projects.

7. ACKNOWLEDGMENTS

This work was supported by grants from NIAMS, National Institutes of Health (R01 AR43969), the Central Ohio Diabetes Association, the Children's Research Institute, Columbus, Ohio (219999), and the Pittsburgh Supercomputing Center through the NIH Center for Research Resources Cooperative Agreement (1P41 RR06009). ZY was a recipient of an Ohio State University Presidential Fellowship.

8. REFERENCES

1. The MHC Sequencing Consortium: Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401, 921-923 (1999)

2. JL Tiwari & PI Terasaki: HLA and Disease Associations. Springer-Verlag, New York (1985)

3. Porter R. R: Complement polymorphism, the major histocompatibility complex and associated diseases: a speculation. *Mol Biol Med* 1, 161-168 (1983)

4. Porter R. R: The complement components coded in the major histocompatibility complexes and their biological function. *Immunol Rev* 87, 7-17 (1985)

5. Yu C. Y, Z. Yang, C. A. Blanchong & W. Miller: The human and mouse MHC class III region: a parade of the centromeric segment with 21 genes. *Immunol Today* 21, 320-328 (2000)

6. Yu C. Y: Molecular genetics of the human MHC complement gene cluster. *Exp Clin Immunogenet* 15, 213-230 (1998)

7. Blanchong C. A, B. Zhou, K. L. Rupert, E. K. Chung, K. N. Jones, J. F. Sotos, R. M. Rennebohm & C. Y. Yu: 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 (2000)

8. Levi-Strauss M, M. C. Carroll, M. Steinmetz & T. Meo: A previously undetected MHC gene with an unusual periodic structure. *Science* 240, 201-204 (1988)

9. Speiser P. W & P.C. White: Structure of the human RD gene: a highly conserved gene in the class III region of the major histocompatibility complex. *DNA* 8, 745-751 (1989)

10. Surowy C. S, G. Hoganson, J. Gosink, K. Stunk & R. A. Spritz: The human RD protein is closely related to nuclear RNA-binding proteins and has been highly conserved. *Gene* 90, 299-302 (1990)

11. Cheng J, K. J. Macon & J. E. Volanakis: cDNA cloning and characterization of the protein encoded by RD, a gene located in the class III region of the human major histocompatibility complex. *Biochem J* 294, 589-593 (1993)

12. Birney E, S. Kuman & A. R. Krainer: Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 25, 5803-5816 (1993)

13. Hornstein E, A. Abramzon-Talianker, I. Wiesel & J. P. Michalski: The human poly(A)-binding protein (PABP) gene: structural and functional analysis. *GenBank,* Accession number 1562511 (1997)

14. Kawakami A, Q. Tian, X. Duan, T. Anderson, S. F. Schlossman & P. Anderson: Identification and functional characterization of a TIA-1-related nucleolysin. *Proc Natl Acad Sci USA* 89, 8681-8685 (1992)

15. Yamaguchi Y, T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa & H. Handa: NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97, 41-51 (1999)

16. Shilatifard A: Factors regulating the transcriptional elongation activity of RNA polymerase II. *FASEB J* 12, 1437-1446 (1998)

17. Yamaguchi Y, T. Wada, D. Watanabe, T. Takagi, J. Hasegawa & H. Hands: Structure and Function of the Human Transcription Elongation Factor DSIF. *J Bio Chem* 274, 8085-8092 (1999)

18. White P. C, J. Vitek, R. G. Lahita & P. W. Speiser: Polymorphism in the RD (D6S45) gene. *Hum Genet* 89, 243-244 (1992)

19. Taylor P. R, J. T. Nash, E. Theodoridis, A. E. Bygrave, M. J. Walport & M. Botto: A targeted disruption of the murine complement factor B gene resulting in loss of expression of three genes in close proximity, factor B, C2 and D17H6S45. *J Biol Chem* 273, 1699-1704 (1998)

20. Yang Z & C. Y. Yu: Organization and gene duplications of the human and mouse MHC complement gene clusters. *Exp Clin Immunogenet* 17, 1-17 (2000)

21. Saeboe-Larrsen S, Urbeanczk Mohebi B. & A. Lambertsson: The Drosophila ribosomal protein L14-encoding gene, identified by a novel Minute mutation in a dense cluster of previously undescribed genes in cytogenetic region 66D. *Mol Gen Genet* 255, 141-151 (1997)

22. Fuller-Pace F. V: RNA helicase: modulators of RNA structure. *Trends Cell Biol* 4, 271-274 (1994)

23. Dangel A. W, L. Shen, A. R. Mendoza, L.-C. Wu & C. Y. Yu: Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family. *Nucleic Acids Res* 23, 2120-2126 (1995)

24. Widner W. R & R. B. Wickner: Evidence that the SKI antiviral system of S*accharomyces cerevisiae* acts by blocking expression of viral mRNA. *Mol Cell Biol* 13, 4331-4341 (1993)

25. Masison D, A. Blanc, J. C. Ribas, K. Carroll, N. Sonenberg & R. B. Wickner: Decoying the cap- mRNA degradation system by a double stranded RNA virus and poly(A)- mRNA surveillance by a yeast antiviral system. *Mol Cell Biol* 15, 2763-2771 (1995)

26. Jacobs Anderson J & R. Parker: The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SK12 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* 17, 1497-1506 (1998)

27. Benard L, K. Carroll, R. C. P. Valle & R. B. Wickner: Ski6p is a homolog of RNA-processing enzymes that affects translation of non-poly(A) mRNAs and 60S ribosomal subunit biogenesis. *Mol Cell Biol* 18, 2688-2696 (1998)

28. Brown J. T, X. Bai & A. W. Johnson: The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex *in vivo*. *RNA* 6, 449-457 (2000)

29. Qu X, Z. Yang, S. Zhang, L. Shen, A. W. Dangel, J. H. Hughes, K. L. Redman, L. C. Wu & C. Y. Yu: The human DEVH-box *protein* Ski2w from the HLA is located in nucleoli and ribosomes. *Nucleic Acids Res* 26, 4068-4077 (1998)

30. Yang Z, L. Shen, A. W. Dangel, L. C. Wu & C. Y. Yu: 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 (1998)

31. Rowen L, C. Dankers, D. Baskin, J. Faust, C. Loretz, M. E. Ahearn, A. Banta, S. Schwartzell, T. M. Smith, T. Spies & L. Hood: Homo sapiens HLA class III region containing tenascin X gene, partial cds, cytochrome P450 21-hydroxylase (CYP21B), complement C4 (C4B), G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes, complete cds. *GenBank*, Accession no. AF019413 (1997)

32. Xue Y, X. Bai, I. Lee, G. Kallstrom, J. Ho, J. Brown, A. Stevens & A. W. Johnson: *Saccharomyces cerevisiae RAI1* (YGL246c) is homologous to DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Mol Cell Biol* 20, 4006-4015 (2000)

33. Kelly K. O. and M. P. Deutscher: Characterization of *Escherichia coli* RNase PH. *J Biol Chem* 267, 17153-17158 (1992)

34. Paulsen J. E, E. E. Capowski & S. Strome: Phenotypic and molecular analysis of *mes-3*, a maternal-effect gene required for proliferation and viability of the germ line in *C. elegans*. *Genetics* 141, 1383-1398 (1995)

35. Seydoux G. and S. Strome: Launching the germline in Caenorhabditis elegans: regulation of gene expression in early germ cells. *Development* 126, 3275-3283 (1999)

36. Shen L, L.-C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. W. Dangel, M. C. Carroll, W. B. Zipf & C. Y. Yu: Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the 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 (1994)

37. Sargent C. A, M. J. Anderson, A.-L. Hsieh, E. Kendall, N. Gomez-Escobar & R. D. Campbell: Characterisation of the novel gene G11 lying adjacent to the complement C4A gene in the human major histocompatibility complex. *Hum Mol Genet* 3, 481-488 (1994)

38. Gomez-Escobar N, Chou C.F, W.-W. Lin, S.-L. Hsieh & R. D. Campbell: The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase. *J Biol Chem* 273, 30954-30960 (1998)

39. Stavenhagen J. B & D. M. Robins: An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 55, 247-254 (1988)

40. Adler A. J, M. Manielsen & D. M. Robins: Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors. *Proc Natl Acad Sci USA* 89, 11660-11663 (1992)

41. Ramakrisman C. and D. M. Robins: Steroid hormone responsiveness of a family of closely related mouse proviral elements. *Mamm Genome* 8, 811-817 (1998)

42. Bird A. P: CpG islands as gene markers in the vertebrate nucleus. *Trends Genet* 3, 342-347 (1987)

43. Lee S & Song KY: Identification and characterization of a bidirectional promoter from the intergenic region between human DDX13 and RD genes. *Mol Cell* 10, 46-53 (2000)

44. Hartung K, M. P. Baur, R. Coldewey, M. Fricke, J. R. Kalden, H. J. Lakomek, H. H. Peter, D. Schendel, P. M. Schneider, S. A. Seuchter, W. Stangel & H. R. G. Deicher: Major histocompatibility complex haplotypes and complement C4 alleles in systemic lupus erythematosus. *J Clin Invest* 90, 1346-1351 (1992)

45. Reveille J. D: Major histocompatibility complex class II and non-major histocompatibility complex genes in the pathogenesis of systemic lupus erythematosus. In: Systemic Lupus Erythematosus. Eds: Lahita R G, Academic Press, San Diago, 67-90 (1999)

46. Walport M. J: Lupus, DNase and defective disposal of cellular debris. *Nature Genetics* 25, 135-136 (2000)

47. Napirei M, H. Karsunky, B. Zevnik, H. Stephan, Mannherz H.G. & T. Moroy: Features of systemic lupus erythematosus in Dnase 1-deficient mice. *Nat Genet* 25, 177-181 (2000)

48. Eilat D &Y. Naparstek: Anti-DNA autoantibodies: a puzzle of autoimmune phenomenon. *Immunol Today* 20, 339-342 (1999)

49. Lachmann P. J: An attempt to characterize the lupus erythematosus cell antigen. *Immunology* 4, 153-163 (1961)

Key Words: Exoribonuclease, Nucleoli, Ribosomes, RNA Elongation, RNA Turnover, Nuclear Protein Kinase, Sequence Conservation, Review

Send correspondence to: Zhenyu Yang, Ph.D, Scientist, PPD Discovery, 1505 O'Brien Drive, Suite B, Menlo Park, CA 94025, Tel: 650-617-2033 ext. 288, Fax: 650-617-9574, E-mail: zhenyu.yang@menlo.ppdi.com