Mammalian Genome 11, 292–299 (2000).

Mammalian Genome
Incorporating Mouse Genome

High-resolution FISH mapping of the rat α_{2u}-globulin multigene family

D.A. McFadyen,* J. Locke

Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada

Received: 26 July, 1999 / Accepted: 8 December 1999

Abstract. The rat α_{2u}-globulins are a group of similar proteins that are encoded by a family of approximately 20 genes located a single locus of ≤880 kbp on Chromosome (Chr) 5q. Individual members of this gene family demonstrate complex tissue, hormonal, and developmental expression patterns despite a high degree of sequence similarity among the members and consequently provide an interesting system for studying the evolution of differential gene expression. Hybridization analysis indicated that gene classes, similar to those identified at the homologous MUP locus in the mouse, do not exist within the rat α_{2u}-globulin locus. Furthermore, cross-hybridization analysis revealed the presence of conserved sequences in the 5′ and 3′ regions flanking the α_{2u}-globulin genes, some of which were present in an inverted orientation. We have used high-resolution fiber FISH to examine the structural organization of the α_{2u}-globulin locus, and found the genes to be arranged as an array of both direct and inverted repeats. The organization of the rat α_{2u}-globulin genes differs from the MUP genes and suggests different evolutionary events have reorganized these homologous sets of genes.

Introduction

The rat α_{2u}-globulins are a model system for studying the mechanisms and evolution of differential gene expression. The α_{2u}-globulins are a set of small proteins that are abundant in adult male rat urine and account for approximately half of the urinary protein in mature male rats (Roy 1979). The secreted urinary α_{2u}-globulins are synthesized in the liver (Roy and Neuhaus 1966; Roy and Raber 1972) and regulated in a complex developmental and multihormonal manner. They are completely absent from the urine of female rats and immature male rats. Several other tissues, in both males and females, also express α_{2u}-globulins. These tissues are primarily secretory in nature and include preputial, lachrymal, meibomian, submaxillary, perianal, and mammary glands, kidney, and brain (Laperche et al. 1983; Gubits et al. 1984; MacInnes et al. 1986; Mancini et al. 1989; Wang et al. 1997). As in the liver, expression of α_{2u}-globulin in these tissues is regulated by developmental and hormonal cues.

The α_{2u}-globulin proteins are encoded by a family of approximately 20 similar genes clustered within a region of ≤880 kbp at chromosomal band 5q22-q24 (Kurtz 1981; McFadyen et al. 1999). The sequence similarity among family members is close enough to make identifying the expression patterns of the different nonallelic copies troublesome (Dolan et al. 1982), but divergent enough to permit family members to be expressed differently in different tissues (Wang et al. 1997). To characterize the genetic basis for these expression differences, a proper, detailed comparative study of the genes within the family is required. We previously reported on the isolation and characterization of P1 clones from the α_{2u}-globulin locus (McFadyen et al. 1999) and showed that clones containing multiple α_{2u}-globulin genes indicated that the α_{2u}-globulin locus consists of single genes separated by 13–30 kbp. This arrangement differs from the organization of genes at the homologous MUP (major urinary protein) locus in the mouse. The MUP genes are arranged as a 45-kbp palindrome consisting of one Group 1 (transcribed gene) and one Group 2 (pseudogene) gene, linked in a head-to-head fashion (Bishop et al. 1982, 1985; Clark et al. 1984). Group 1 and 2 genes can be distinguished from each other by cross-hybridization analysis (Clark et al. 1985). Currently, too little is known about the organization of the rat α_{2u}-globulin gene cluster (McFadyen et al. 1999) to be able to sort out the order of the different non-allelic copies. Furthermore, the high degree of sequence similarity among α_{2u}-globulin gene coding and noncoding regions has made the conventional large-scale clone mapping and contig construction impractical, and an alternative approach is needed.

Fluorescence in situ hybridization (FISH) technology has been very useful in the large-scale analysis of genes and genomes (Pinkel et al. 1986). This technique has permitted the rapid and routine assignment of a large number of genes to specific chromosomal regions. The resolution of conventional FISH methods is in the range of 1–2 Mb on metaphase chromosomes (Buckle and Kearney 1994), but improvements have applied FISH to interphase cell nuclei where the DNA is less condensed, allowing for the distinction of probes separated by 100–500 kb (Trask et al. 1989; Lawrence et al. 1990; Van den Engh et al. 1992). New high-resolution FISH mapping protocols that utilize decondensed chromatin and naked DNA fiber targets have been described that permit visualization of overlapping probes as small as 5 kb (Heng et al. 1992; Wiegant et al. 1992; Parra and Windle 1993; Houseal et al. 1994; Fidlerova et al. 1994; Heiskanen et al. 1994; Weier et al. 1995).

These new cytogenetic mapping methods provide a direct visual means to determine the arrangement of specific sequences (genes) along a stretched single DNA fiber. Application of these techniques has already proven useful in the orientation of clones, assessing the degree of clone overlap, estimating the size of uncloned gaps between adjacent contigs, and the examination of intragenic organization and rearrangements (Bengtsson et al. 1994; Tocharoentanaphol et al. 1994; Heiskanen et al. 1995; Pizzuti et al. 1996). Furthermore, these high-resolution FISH methods are ideally suited to studying the structure of amplified sequences. DIRVISH mapping has been used to determine the structural organization of an amplified DHFR locus in a hamster cell line (Parra and Windle 1993).

Here, we report our further characterization of the rat α_{2u}-globulin locus by applying standard molecular techniques and fiber FISH methods to analyze lambda and P1 genomic clones as well as rat genomic DNA. Together they provide a large-scale view of the organization of the entire α_{2u}-globulin locus. The

* *Present address:* Department of Medical Genetics, University of Alberta, Edmonton, Alberta, T6G 2S2, Canada.

Correspondence to: D.A. McFadyen; e-mail:dmcfadye@gpu.srv.ualberta.ca

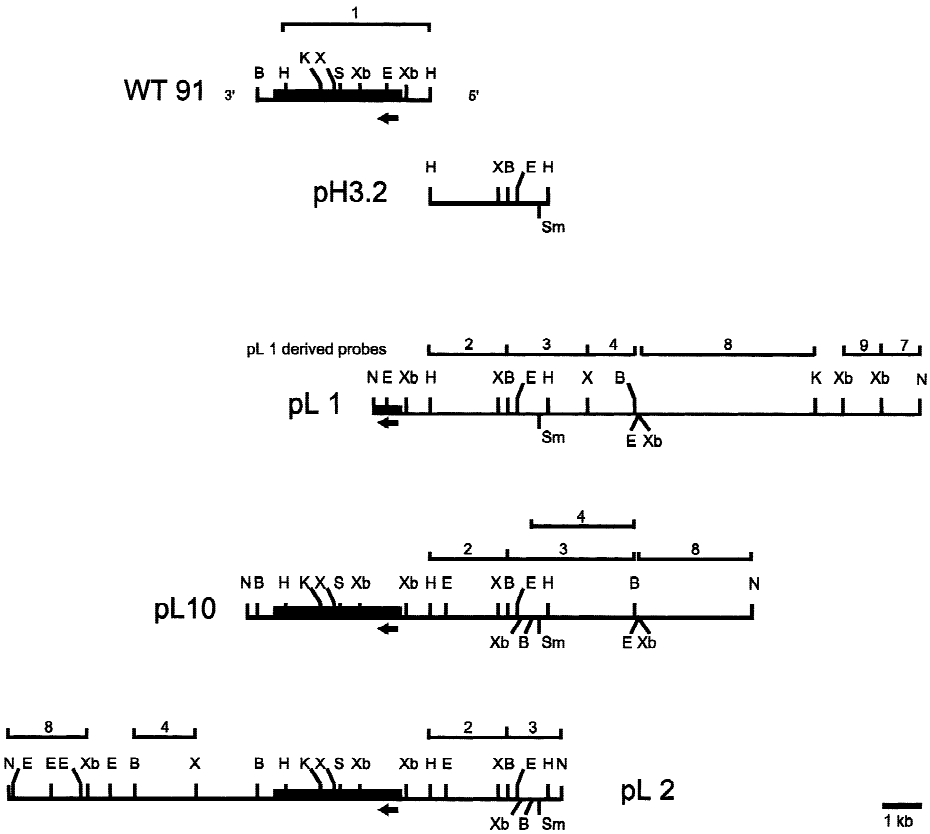

Fig. 1. The α_{2u}-globulin gene containing genomic clones. Plasmids WT91 (Addison and Kurtz 1986), pH 3.2, and three lambda clones with varying amounts of flanking sequences are shown. Transcribed regions are shown as black boxes and the orientation is indicated by the arrows. Hybridization probes were derived from lambda clone pL1 as indicated. The corresponding hybridization position of each probe is indicated above each lambda clone map. WT91 was used in fiber FISH, as was pH 3.2, which contains the 3.2-kb *Hin*dIII of pL1. Restriction sites: B, *Bam*HI; E. *Eco*RI; H, *Hin*dIII; K, *Kpn*I; N. *Not*I; S, *Sal*I; Sm, *Sma*I; X, *Xho*I; Xb, *Xba*I. The *Not*I sites are those of the vector, pBluescript.

results indicate that there is a substantial difference in the organization of the rat α_{2u}-globulin locus and the mouse MUP locus and provide insight into the evolution of these gene families.

Materials and methods

Probes. Plasmid WT91 (Fig. 1) was used as a source of α_{2u}-globulin gene probes. It contains the entire coding region of α_{2u}-globulin gene 91, approximately 750 bp of upstream sequences, and 400 bp of 3′ flanking sequences. Probes were also prepared from clone pL1 (Fig. 1), which contains sequences of α_{2u}-globulin gene 91 (Wang and Hodgetts 1998). Plasmid pH3.2, derived from clone pL1, was used as a FISH probe and contains sequences immediately upstream from WT91 (Fig. 1). Rat α_{2u}-globulin containing lambda genomic clones pL2 and pL10 (Wang and Hodgetts 1998) was used in the mapping of the upstream region of α_{2u}-globulin genes.

Library screening. Duplicate plaque lifts were made from each of six plates, containing approximately 50,000 bacteriophage per plate, from a rat genomic library in λ DASH II according to manufacturer's directions (Stratagene, LaJolla, Calif.). For differential hybridization experiments, the filters were prehybridized in 250 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA, 7% SDS at 50°C for 1 h. Radiolabeled probes were prepared from gel-purified restriction fragments labeled with the Quick Prime labeling kit (Pharmacia, Baie d'Urfé, PQ) with α^{32}P-dCTP (3000 Ci/mmole: 10 μCi/μl, Amersham, Baie d'Urfé, PQ) according to the manufacturer's directions. Hybridization was performed overnight with 10^5–10^6 cpm/ml of radioactively labeled probe. Following hybridization, all 12 filters were washed twice for 15 min each in 250 mM Na_2HPO_4 (pH 7.2), 1 mM EDTA, 1% SDS at room temperature with shaking. After the first set of washes, the duplicate membranes were differentially washed: one set of six filters was washed at low stringency—three washes at 42°C in 4 × SSC/0.1% SDS for 15 min each with shaking. The remaining set of six filters was washed three times in 0.1 × SSC/0.1% SDS for 15 min each at 65°C. The same filters were also probed with additional probes derived from WT91 and pL1, as described, except that the most stringent wash was performed in 2 × SSC/0.1% SDS at 65°C.

Preparation of target fibers for fiber FISH. The method outlined by Heiskanen et al. (1994) was used to prepare target DNA fibers for FISH. Briefly, for genomic DNA targets, a section equivalent to approximately 1/9 of a 200-μl pulsed field gel electrophoresis block was placed near the end of a treated microscope slide (Probe On Plus, Fisher Scientific, Edmonton, AB) in 15 μl of sterile distilled water. The slide was heated over a hot plate until the agarose block had melted completely. The resulting liquid drop was extended over the surface of the slide by use of the end of another slide. Slides were air-dried for at least 30 min at room temperature before their use in an in situ hybridization experiment.

To prepare bacteriophage P1 targets, we linearized 10 μg of each of the P1 clones p2860, p2861, and p2862 (McFadyen et al. 1999) with *Not*I (p2860 and p2862), or *Sfi*I (p2861) according to supplier's directions (New England Biolabs, Mississauga, ON). Digested DNAs were mixed with an equal volume 1% low melt agarose (BRL) prepared in 0.5× TBE and loaded into the wells of a 1% low-melt agarose pulsed field gel. DNAs were resolved in a CHEF DR III (Biorad, Mississauga, ON) apparatus in 0.5× TBE running buffer at 14°C, 6 V/cm, a reorientation angle of 120°, and a linearly ramped switch time from 1 to 6 s for 24 h. Following electrophoresis, gels were stained with ethidium bromide, and the region containing linearized P1 DNAs was excised from the gel. Target fibers for FISH were then prepared as described above.

In situ hybridization. Probes (1 μg) for fiber-FISH were labeled by nick translation according to standard methods with either 25 μM digoxigenin-11-dUTP or 25 μM biotin-16-dUTP (Boehringer, Laval, PQ). Probes were

purified by passing them over a Biogel P60 (Biorad) column and ethanol precipitation. Probes were resuspended in hybridization buffer [50% formamide, 2× SSC, 10% dextran sulfate, 40 mM sodium phosphate, 0.1% SDS, 1× Denhardt's, and 100 μg/ml sonicated salmon sperm DNA (pH 7.0)] at a final concentration of 2 ng/μl. Probes were denatured at 70°C for 5 min and cooled on ice before use.

Slides containing genomic DNA or P1 DNA fibers were treated with RNase A (100 μg/ml; Sigma, Oakville, ON) in 2× SSC (pH 7.0) at 37°C for 1 h in a moist chamber. The DNA was subsequently denatured in 70% formamide/2× SSC, pH 7.0 at 70°C for 3 min and then immediately passed through an ice-cold ethanol series (70%, 70%, 95%), and air-dried completely. A 100-μl aliquot of the denatured probe solution was applied to each slide, and hybridization was allowed to proceed overnight at 37°C in a moist chamber.

Post-hybridization washes were carried out in 50% formamide/2x SSC, pH 7.0 and 2x SSC, pH 7.0, twice each for 4 min at 45°C. Biotinylated probes were detected using FITC-conjugated avidin DCS, and the signal was amplified by biotinylated goat anti-avidin and another layer of FITC-conjugated avidin DCS (all from Vector, Burlington, ON) according to Raap and Wiegant (1994). Digoxigenin-labeled probes were detected according to Raap and Wiegant (1994) with mouse monoclonal anti-digoxigenin, digoxigeninylated sheep anti-mouse Ig-F(ab′)$_2$-fragment, and TRITC-conjugated sheep anti-digoxigenin (all from Boehringer). Slides were examined with a double band pass filter set (Chroma) for the simultaneous detection of FITC and TRITC labeled conjugates. Photos were taken with Fugi Super HG 1600 ASA 35mm film. Negatives were scanned with a Polaroid SprintScan 35 slide scanner. Images were processed with Adobe Photoshop 4.0 to generate the composite images in Figs. 2 and 3 and to remove extraneous background commonly associated with these methods (Heng et al. 1996).

Image analysis of hybridizing regions involved enlarging suitable negatives with a slide projector and tracing regions of fluorescence. Measurements for calculating the position of α_{2u}-globulin genes on each of the P1 clones were made from the projected image.

Results

Rats have only one class of α_{2u}-globulin genes. The α_{2u}-globulin homologs in the mouse are called the mouse urinary protein (MUP) genes, which are comprised of two major gene classes (transcribed genes and pseudo-genes) that are organized in repeating head-to-head pairs. These two major classes can be distinguished by differential hybridization (a mismatch of ~20%). The possible existence of two major α_{2u}-globulin gene classes in the rat was investigated with a differential stringency screen of clones from a rat lambda genomic library. A duplicate set of six filters representing 1.5 rat genome equivalents was hybridized to the 4-kb *Hin*dIII fragment from WT91 containing the α_{2u}-globulin gene coding region. One set was washed under conditions of low stringency, and a total of 37 plaques was found to hybridize (Table 1). The low-stringency wash conditions were such that it should permit detection of hybrids with sequence mismatch of ~30%. The other set of filters, hybridized with the same probe, but washed under high-stringency conditions, allowed the detection of sequences with only ~5% mismatch. This filter set was positive for the same 37 plaques identified under the low-stringency washing conditions. The lack of clones that hybridize with low stringency but not high indicates that rat α_{2u}-globulin genes that hybridize differ from each other by 5% or less. These differential hybridization conditions, which would reveal the major classes of MUP genes, do not distinguish any different classes of α_{2u}-globulin genes and make a locus organization similar to the mouse MUP genes unlikely.

Regions of similar sequence extend upstream from the α_{2u}-globulin genes themselves. Given the substantial similarity among α_{2u}-globulin gene transcribed regions, we investigated the extent to which this similarity extended into the adjacent untranscribed regions. We used Southern blot analysis of three lambda clones (pL1, pL2, and pL10) representing two different α_{2u}-globulin genes. Probes from clone pL1, which contains the upstream region of α_{2u}-globulin gene 91, were hybridized to clones pL12 and pL10, which contain the downstream and upstream regions of another α_{2u}-globulin gene respectively (Wang and Hodgetts 1998).

The results of the cross-hybridization analysis between these two genes are schematically represented in Fig. 1. Probes 2 and 3, derived from sequences between +0.7 to +4.3 kbp of subclone pL1, were found to cross-hybridize to corresponding regions of pL2 and pL10. Probe 4, derived from +4.3 to +5.7 kbp of subclone pL1, hybridized to sequences at the corresponding upstream position of clone pL10; clone pL2 does not extend that far in the 5′ direction, explaining the failure of probe 4 to hybridize to any fragment of this clone. Probe 8, which extends from +6.6 to +11.4 kbp of pL1, was found to hybridize with the corresponding upstream region of pL10. Probes 9 and 7, which extend from +12.1 to +13.3 and +13.3 to +15.6 of pL1 respectively, did not hybridize to pL2 or pL10. These cross-hybridization results indicate that, in addition to the sequence similarity found in the coding and intron regions of α_{2u}-globulin genes, there are similar sequences that extend at least 6.6 kb upstream, too.

To further examine the possibility that extensive 5′ homology is a common feature of the α_{2u}-globulins, the duplicate filter sets previously screened with gene coding probes (WT91) were re-screened with probes 3 and 8, derived from pL1. The results of this screen are summarized in Table 1. Of the 37 plaques previously identified with the 4-kb *Hin*dIII fragment of WT91 (Probe 1), 14 were also positive for both probes 3 and 8, 15 were positive for probe 3, and 1 was positive for probe 8 alone. The remaining seven plaques that hybridized to the gene-coding probe (WT91) alone were not positive for either probe 3 or 8. The estimated maximum insert size for a clone positive only for the gene-coding region of WT91 would be about 11.5 kb. A fragment of this size would easily be accommodated within the lambda DASH II cloning vector used in the construction of this library (size: 9 to 23 kb). Additionally, seven plaques were found to be positive for probes 3 and 8, but not the gene-coding region. An additional 19 plaques hybridized to probe 8 only, and five plaques were found to hybridize to probe 3 only. This latter class is unexpected given the restriction maps of the lambda subclones (Fig. 1) and the reported size range of inserts for this particular library. On the basis of these maps, a phage positive for only probe 3 would have a maximum insert size of approximately 6.5 kb, making it smaller than the 9 kb reported to be the smallest fragment size accepted by the vector. A possible explanation for this class of plaques could be that there are α_{2u}-globulin genes with insertions in the region between the gene and probe 3 sequences, and/or insertions in the region between Probe 3 and Probe 8, thus generating fragments that could be accommodated by the cloning vector containing Probe 3.

The number of positive plaques observed when each of the three probes (3, 8, and the coding region of WT91) was used separately to screen the same filter sets is 41, 41, and 37 respectively. These results are consistent with these sequences, obtained from the 5′ flanking sequences of the α_{2u}-globulin gene represented in pL1, being common to most if not all of the α_{2u}-globulin genes in the rat genome.

In the course of the above analysis, we observed that probes 4 and 8, which were isolated from the upstream sequences of gene 91 in clone pL1, hybridized to sequences downstream of the α_{2u}-globulin gene in clone pL2 (Fig. 1). The order of probes 4 and 8 in the downstream region is inverted with respect to the order in which they appear in the upstream region. These cross-hybridizing regions were examined in more detail by restriction endonuclease analysis. The positions of restriction sites demonstrated that the two cross-hybridizing segments represent similar sequences but in an inverted orientation. This makes the organization of these sequences inconsistent with a simple tandem duplication where the sequences present in the downstream region of clone pL2 represent the upstream flanking sequences of an adjacent α_{2u}-globulin gene.

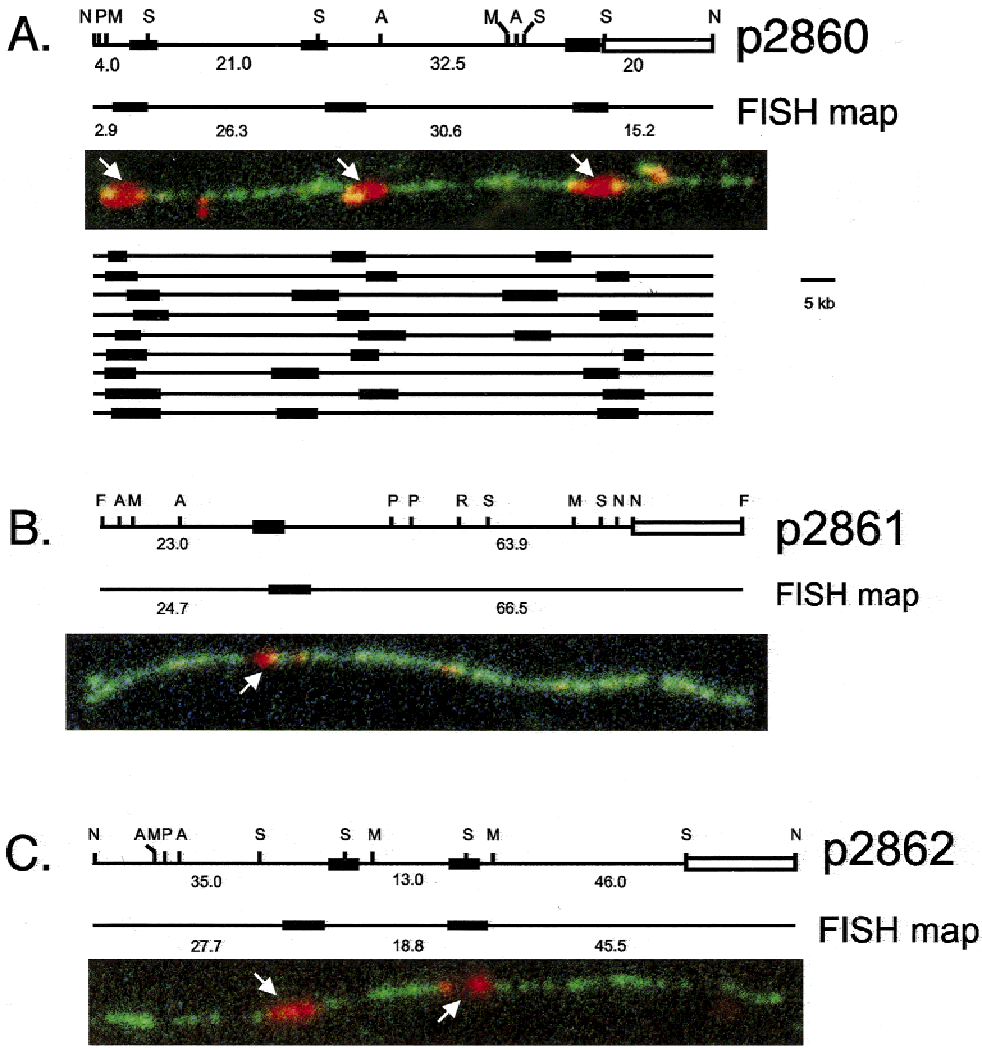

Fig. 2. Comparison of α_{2u}-globulin gene location on P1 clones determined by restriction mapping and fiber FISH. The top line of each part shows the restriction map of a P1 clone containing α_{2u}-globulin sequences. **(A)** p2860, **(B)** p2861, **(C)** p2862 (McFadyen et al. 1999). Vector sequences are indicated by the open boxes on the right. The second line in each part depicts the map obtained from FISH analysis of each P1 clone. Intergene distances as well as the distances between α_{2u}-globulin genes and the termini of the clones are indicated (in kbp) for both the restriction map and FISH map. The size of each α_{2u}-globulin gene is shown as 4.5 kb on the restriction maps. The photograph in each part shows an example of a fiber FISH image of the corresponding P1 clone hybridized with both WT91, to indicate the size and position of α_{2u}-globulin genes (red), and P1 DNA, to delineate the clone itself (green). Part (A) contains a schematic representing a sample of nine different P1 molecules used to derive the FISH map for p2860 above.

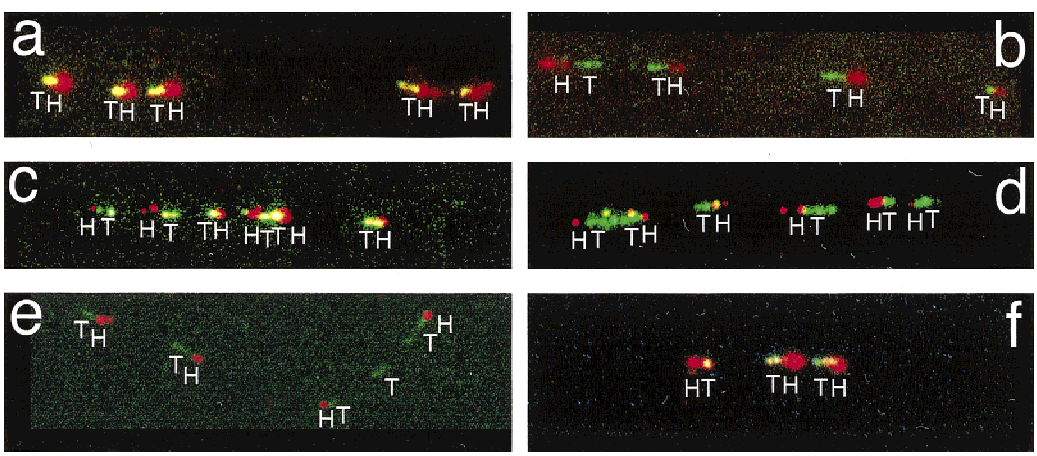

Fig. 3. Photographs of six selected genomic fiber FISH profiles that together span the α_{2u}-globulin gene locus. Genomic DNA was probed with two probes, WT91 (containing the gene and 3′ end—T) and pH3.2 (the 5′ end sequences—H), and counterstaining with DAPI. Biotinylated WT91 DNA was detected with FITC (green) and digoxigenin-labeled pH 3.2 DNA was detected with rhodamine (red). Photographs were taken using a double-band pass filter for FITC and rhodamine, thereby preventing the DAPI staining from interfering with signal positioning. Part **(a)** profile #36, **(b)** profile #20, **(c)** profile #47, **(d)** profile #18, **(e)** profile #13, and **(f)** profile #41.

Table 1. The number of signals detected in a lambda library with various α_{2u}-globulin gene probes.

Probe	Total Number	Alone	With Probe 1 (gene)	With Probe 3 (proximal 5′)	With Probe 8 (distal 5′)	With Probe 3 and Probe 8
Probe 1	37	7	–	15	1	14
Probe 3	41	5	15	–	7	–
Probe 8	41	19	1	7	–	–

Note: Probe locations are defined in Fig. 1.

This degree of sequence similarity, both upstream and downstream, among the repeated α_{2u}-globulin genes prevents any large-scale mapping of the locus with typical procedures such as cross-hybridization or restriction site overlapping. To overcome this obstacle, we have employed FISH to examine the arrangement of repeated genes at the α_{2u}-globulin locus.

Mapping the α_{2u}-globulin genes of P1 clones with fiber FISH. To begin an analysis of the organization of the genes in this cluster, we have mapped the number and location of each gene on three P1 clones, using two-color fiber FISH. The α_{2u}-globulin gene within each P1 was localized with digoxigenin-labeled WT91 and visualized with rhodamine (red signals). The DNA from each P1 DNA was "stained" by hybridization with the biotinylated P1 DNA and visualized with FITC (green signal). Thus, the red signals of the α_{2u}-globulin genes could be mapped at specific locations along the green *Not*I or *Sfi*I linearized P1. Initial observations showed clone p2860 has three genes: p2861, a single gene; and p2862, two α_{2u}-globulin genes (Fig. 2).

Quantitative analysis of the images permitted the signal locations to be mapped. Both the distance between genes on the same molecule and their distance from the ends of the linearized P1 DNAs were determined for many molecules. An example is shown for p2860, where nine molecules were selected and total length measurements indicated the average degree to which the DNA was stretched was 0.266 μm/kb (range = 0.152–0.389 μm/kbp), which is approximately 80% of that expected for relaxed DNA (0.34 μm/kbp). The orientation of each molecule was decided by the proximity of the closest gene to one terminus. The location of each α_{2u}-globulin gene on each of the nine molecules examined is shown schematically in Fig. 2. In this case, the genes adjacent to the termini have been positioned approximately 15.2 kbp from the end containing vector sequences following linearization with *Not*I, and approximately 2.9 kbp from the other end. These compare (Fig. 2) with values of approximately 20 kbp and 5.5 kbp respectively, as determined by restriction endonuclease and Southern hybridization analyses (McFadyen et al. 1999). The intergene distances were also calculated from the fiber FISH results. The distance between the α_{2u}-globulin gene adjacent to the vector sequences in the linearized clone and the next gene along the P1 was estimated to be 30.6 kbp, while the distance between the second gene and the gene located at the other terminus was determined to be about 26.3 kbp. These values correspond to the estimates of 32.5 kbp and 21 kbp for the same two intervals based on restriction enzyme digestion and Southern hybridization (McFadyen et al. 1999). The analysis of P1 clone p2861 (10 molecules) and P2862 (11 molecules) is shown in Fig. 2b and c.

From these measurements we conclude the relative distance estimates of our fiber FISH are good, in that they agree with the previous values (McFadyen et al. 1999) while absolute distance determinations are poor owing to local and global variation in the extent of DNA stretching. Consequently, absolute measurements are probably not good enough to obtain a reliable distance measure of gene-to-gene distances, but should be more than acceptable for determining gene orientation with two-color fiber FISH.

Mapping the α_{2u}-globulin genes in genomic DNA using fiber FISH. Our molecular and cytogenetic analysis of the three P1 clones has defined three small segments of the α_{2u}-globulin locus and obtained three intergene distances. This analysis of small sections provided a basis for applying the fiber FISH methodology to determine an organization for the entire α_{2u}-globulin locus.

Rat genomic DNA was spread to form fibers on slides. The location and orientation of each gene was determined by using two probes, one 3′ and the other 5′, and counterstaining the DNA with DAPI. Biotinylated WT91 DNA detected with FITC (green) identified the gene and 3′ end, and digoxigenin-labeled pH3.2 DNA detected with rhodamine (red) identified the 5′ end (Fig. 1). Only fibers showing hybridization to both probes were analyzed. These signals showed most genes are present in tandem (Fig. 3a–f). However, there are examples of genes demonstrating head-to-head linkage (Fig. 3d) and tail-to-tail linkage (Fig. 3b–f). Thus, the array is not homogeneous with respect to gene orientation. Since these deviations are distributed at several locations throughout the cluster, they permitted us to assemble the patterns into an overall array.

Fifty independent signal profiles obtained from fiber FISH images were analyzed. These profiles were assembled into a single large gene array constituting the ~21 genes present in the α_{2u}-globulin locus (Fig. 4). This assembly was achieved by taking several of the longer profiles and aligning them according to overlapping gene orientation patterns in a manner analogous to the assembly of overlapping clones based on similarly positioned restriction sites. Six example rat genomic fiber FISH profiles that, taken together, account for the entire α_{2u}-globulin array proposed are shown in Fig. 3. The simplicity of these orientation patterns allowed us to assemble the array by hand. Shorter profiles were then added to the longer profile's backbone at the appropriate positions. These shorter profile segments support the arrangement based on the larger arrays. It should be noted that many of the shorter profiles could be placed at several alternative positions in the array, although only a single position is indicated.

The organization for the cluster presented in Fig. 4 represents the simplest necessary to align all 50 profiles. The size of this array, as determined by the 50 profiles reported here, is 21 genes, which is consistent with the number of genes in the cluster based on solution hybridization kinetics (Kurtz 1981) and slot blot analysis (McFadyen et al. 1999). Formally, there is the possibility that the array may extend further in either direction, especially at the left end of the array where the genes are all tandemly repeated; however, such an extension would conflict with the estimated total gene number.

Discussion

Our differential stringency screening failed to distinguish classes of α_{2u}-globulin genes that differ from one another by 5% or more. These results are consistent with α_{2u}-globulin sequences found in Genbank where all the non-allelic copies display a sequence similarity of >95%. Furthermore, our Southern hybridization analysis and restriction endonuclease mapping indicate sequence similarity extends more than 6 kbp upstream and similar conservation of sequence was observed downstream, too. Such extensive homology of flanking sequences is a common feature of multigene families. For example, the members of the human U1 RNA gene family share a high degree of sequence homology at distances of over 24 kb upstream of the gene and 20 kb downstream of the gene (Bernstein et al. 1985).

The similarity in coding and adjacent sequences, coupled with their close proximity, suggests that the non-allelic copies of the α_{2u}-globulin gene originated via gene duplication events, and/or the genes have been acted upon in a fashion that reduces variation between family members (for example, gene conversion or unequal crossing over). The high degree of similarity in coding se-

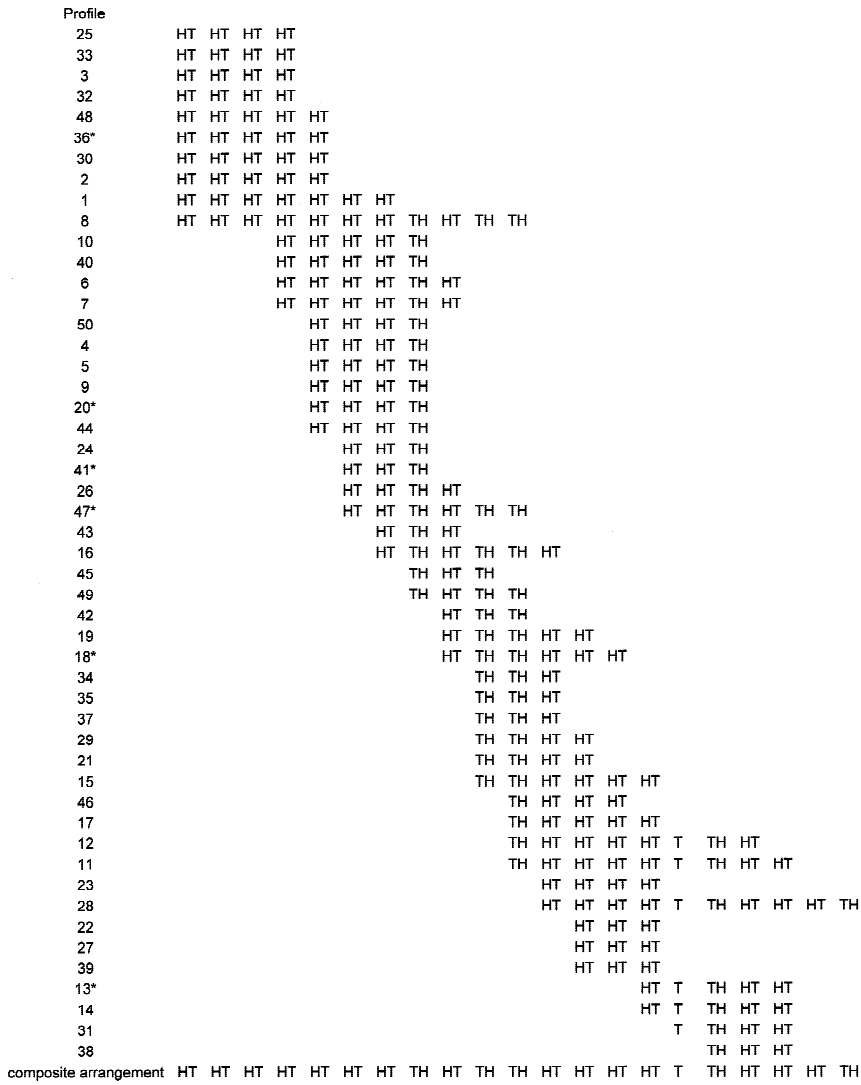

Fig. 4. Proposed arrangement of 21 genes at the α_{2u}-globulin locus. Fifty independent gene orientation profiles, like those shown in Fig. 3, were arranged in an overlapping manner according to the fiber FISH hybridization signal patterns. Symbols are: H, head (5′ end); T, tail (3′ end). A composite arrangement is shown at the bottom. The six profiles shown in Fig. 3 are marked with "*".

quence implies amplification has occurred in recent evolutionary time. The extensive homologies between α_{2u}-globulin genes, extending perhaps more than 12 kb in both the 5′ and 3′ flanking regions, suggest that the original duplication event, regardless of its mechanism, must have involved a DNA segment much larger than the gene itself. Furthermore, a relatively small number of specific changes in nucleotide sequences in the upstream regions of different members are probably responsible for the developmental, hormonal, and tissue-specific expression patterns observed for various subsets of α_{2u}-globulin genes. This evolving gene set makes for an excellent system for identifying the finer details of regulating gene expression.

We have found a large duplicated region in inverted orientation flanking two α_{2u}-globulin genes. This observation has interesting implications for the organization of the α_{2u}-globulin genes which are clustered at a single locus on rat Chr 5 (Kurtz 1981; McFadyen et al. 1999). Bacteriophage lambda clones containing multiple α_{2u}-globulin genes have not been reported; however, we have previously characterized a set of P1 clones containing multiple α_{2u}-globulin genes (McFadyen et al. 1999). Examination of this limited set of clones revealed that the α_{2u}-globulin genes on these clones were directly repeated at distances varying from 13 to 30 kb. We have expanded our analysis of the α_{2u}-globulin locus here by using fiber FISH analysis of rat genomic DNA to determine the structure of the entire locus. Our analysis of 50 independent hybridization signals permitted the assembly of α_{2u}-globulin genes into an array containing 21 genes, consistent with previous reports (Kurtz 1981; McFadyen et al. 1999). The vast majority of the α_{2u}-globulin genes are present as direct tandem repeats; however, several genes are present in an inverted orientation.

Fiber FISH analysis suggests possible gene clustering within the α_{2u}-globulin locus. We observed only small groups (three to five members) of α_{2u}-globulin genes in close proximity to each other. Although such smaller groupings within the locus would be interesting, their presence might result from DNA breakage during the fiber FISH protocol, giving only the appearance of groups within a contiguous array. The variable spacing between α_{2u}-globulin genes precludes using gene-to-gene distances to map these groups more precisely. If smaller groups of genes within the cluster do exist, it will be of interest to determine if there is any functional significance to the groupings; for example, do grouped genes share a similar tissue-specific expression pattern? Three classes of α_{2u}-globulin genes, identified by restriction enzyme site differences, have been shown to be expressed in different overlapping tissue and developmental patterns (Wang et al. 1997; McFadyen et al. 1999). The relationship, if any, between these different expression classes and their position in the locus are not yet known.

Since α_{2u}-globulin and MUP are each present in multiple copies in the rat and mouse genomes respectively, we expect that these two gene clusters (MUP and α_{2u}-globulin) are homologous and presume they were derived from a pre-existing cluster in their

common ancestor (although we can not exclude the possibility of a single gene). We speculate the current differences in number and arrangement are derived from an ancestral repeated array where both tandem and inverted arrangements existed, much like the rat cluster today. This permitted divergence, through unequal crossing-over or gene conversion events, whereby in the lineage leading to the mouse a pre-existing inversion underwent the repeated cycles of unequal crossing-over, resulting in the replacement of the ancestral array with a more derived array comprised mostly of divergently oriented gene pairs (Clark et al. 1984). Within the present-day rat α_{2u}-globulin locus, gene arrangements exist that could have been similar to the source of the gene array found in the present-day MUP gene cluster. Alternatively, production of a localized inverted repeat may have post-dated the divergence of the mouse and rat and served as the amplification substrate in the mouse.

The presence of oppositely oriented short repeats could be responsible for the localized gene inversions that disrupt the otherwise tandem organization of α_{2u}-globulin genes (Jeffreys and Harris 1982). Similarly, these types of sequence elements could have been used in the replacement of a tandem array by the inverted gene pair in the mouse lineage. The mouse repetitive elements R and B1 have been demonstrated to be part of the 45-kb palindrome which forms the predominant organizational unit at the MUP locus (Bishop et al. 1985). The potential involvement of these elements in the evolution of the MUP locus has not been investigated. The ~21 copies of the α_{2u}-globulin gene in the rat (Kurtz 1981; McFadyen et al. 1999) are only half the total number in the mouse, but approximately equivalent to the number of functional MUP genes (Clark et al. 1984). The extra, non-functional MUP copies may reflect its serendipitous presence within the amplified segment.

The evolution of mice and rats, like any speciation, depends on the evolution of effective prezygotic and/or postzygotic isolation mechanisms. Olfactory communication through pheromones plays a major role in many aspects of mammalian speciation (Menzies et al. 1992; Nevo et al. 1976; Yanai and McClearn 1973). We have previously argued that the rapid evolution of the MUP and α_{2u}-globulin genes maybe reflective of a role, through their pheromone transport capabilities, in reproductive isolation (McFadyen et al. 1999). This hypothesis would help explain the apparently rapid evolution of the mouse MUP and rat α_{2u}-globulin genes at both the sequence and gene organization levels and is consistent with observations that sexual isolation between species may have a simple genetic basis (Coyne et al. 1994; Tumlinson et al. 1974).

Acknowledgments. We are grateful to Dr. C.C. Lin and Dr. C. Lee for assistance with the FISH analysis. We thank Dr. K.S. Wang for the genomic clones pL1, pL2, and pL10. This work was supported by the University of Alberta Central Research Fund and a studentship to D.A. McFadyen from the Natural Sciences and Engineering Research Council of Canada.

References

Addison WR, Kurtz DT (1986) Nucleotide sequences required for the regulation of a rat α_{2u}-globulin gene by glucocorticoids. Mol Cell Biol 6, 2334–2346

Bengtsson U, Altherr MR, Wasmuth JJ, Winokur ST (1994) High resolution fluorescence in situ hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. Hum Mol Genet 3, 1801–1805

Bernstein LB, Manser T, Weiner AM (1985) Human U1 small nuclear RNA genes: extensive conservation of flanking sequences suggests cycles of gene amplification and transposition. Mol Cell Biol 5, 2159–2171

Bishop JO, Clark AJ, Clissold PM, Hainey S, Francke U (1982) Two main groups of mouse major urinary protein genes, both largely located on chromosome 4. EMBO J 1, 615–620

Bishop JO, Selman GG, Hickman J, Black L, Saunders RDP et al. (1985) The 45 kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. Mol Cell Biol 5, 1591–1600

Buckle VJ, Kearney L (1994) New methods in cytogenetics. Curr Opin Genet Dev 4, 374–382

Clark JA, Hickman J, Bishop J (1984) A 45-kb DNA domain with two divergently oriented genes is the unit of organisation of the murine major urinary protein genes. EMBO J 3, 2055–2064

Clark JA, Ghazal P, Bingham RW, Barrett D, Bishop JO (1985) Sequence McFadyen and Locke structures of a mouse major urinary protein gene and pseudogene compared. EMBO J 4, 3159–3165

Coyne JA, Crittenden AP, Mah K (1994) Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. Science 265, 1461–1464

Dolan KP, Unterman R, McLaughlin M, Nakhasi HL, Lynch KR et al. (1982) The structure and expression of very closely related members of the α 2u globulin gene family. J Biol Chem 257, 13527–13534

Fidlerova H, Senger G, Kost M, Sanseau P, Sheer D (1994) Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence *in situ* hybridization. Cytogenet Cell Genet 65, 203–205

Gubits RM, Lynch KR, Kulkarni AB, Dolan KP, Gresik EW et al. (1984) Differential regulation of α 2u globulin gene expression in liver, lachrymal gland, and salivary gland. J Biol Chem 259, 12803–12809

Heiskanen M, Karhu R, Hellsten E, Peltonen L, Kallioniemi OP et al. (1994) High resolution mapping using fluorescence *in situ* hybridization to extended DNA fibers prepared from agarose-embedded cells. BioTechniques 17, 928–933

Heiskanen M, Hellsten E, Kallioniemi O, Makela TP, Alitalo K et al. (1995) Visual mapping by Fiber-FISH. Genomics 30, 31–36

Heng HHQ, Squire J, Tsui LC (1992) High resolution mapping of mammalian genes by *in situ* hybridization to free chromatin. Proc Natl Acad Sci USA 89, 9509–9513

Heng HHQ, Tsui L-C, Windle B, Parra I (1996) High-resolution FISH analysis. In *Current Protocols in Human Genetics,* NC Dracopoli, JL Haines, BR Korf, DT Moir, CC Morton et al. eds. (New York, NY.: John Wiley & Sons Inc.), pp 4.5.1–4.5.26

Houseal TW, Dackowski WR, Landes GM, Klinger KW (1994) High resolution mapping of overlapping cosmids by fluorescence *in situ* hybridization. Cytometry 15, 193–198

Jeffreys AJ, Harris S (1982) Processes of gene duplication. Nature 296, 9–10

Kurtz DT (1981) Rat α_{2u} globulin is encoded by a multigene family. J Mol Appl Genet 1, 29–38

Laperche Y, Lynch KR, Dolan KP, Feigelson P (1983) Tissue-specific control of α_{2u} globulin gene expression: constitutive synthesis in the submaxillary gland. Cell 32, 453–460

Lawrence JB, Singer RH, McNeil JA (1990) Interphase and metaphase resolution of different distances within the human dystrophin gene. Science 249, 928–932

MacInnes JI, Nozik ES, Kurtz DT (1986) Tissue-specific expression of the rat alpha$_{2u}$ globulin gene family. Mol Cell Biol 6, 3563–3567

Mancini M, Majumdar D, Chatterjee B, Roy AK (1989) α_{2u}-globulin in modified sebaceous glands with pheromonal functions: localization of the protein and its mRNA in preputial, meibomian, and perianal glands. J Histochem Cytochem 37, 149–157

McFadyen DA, Addison W, Locke J (1999) Genomic organization of the rat α_{2u}-globulin gene cluster. Mamm Genome 10, 463–470

Menzies RA, Heth G, Ikan R, Weinstein V, Nevo E (1992) Sexual pheromones in lipids and other fractions from urine of the male mole rat, *Spalax ehrenbergi.* Physiol Behav 52, 741–747

Nevo E, Bodmer M, Heth G (1976) Olfactory discrimination as an isolating mechanism in speciating mole rats. Experientia 32, 1511–1513

Parra I, Windle B (1993) High resolution visual mapping of stretched DNA by fluorescent hybridization. Nat Genet 5, 17–21

Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83, 2934–2938

Pizzuti A, Novelli G, Mari A, Ratti A, Colosimo A et al. (1996) Human homologue sequences to the *Drosophila* dishevelled segment-polarity gene are deleted in the DiGeorge syndrome. Am J Hum Genet 58, 722–729

Raap AK, Wiegant J (1994) Use of DNA-halo preparations for high resolution DNA in situ hybridization. Methods Mol Biol 33, 123–130

Roy AK (1979) Hormonal regulation of α_{2u} globulin in rat liver. In *Bio-*

chemical Actions of Hormones VI. G Litwack, ed. (New York, NY.: Academic Press Inc.) pp. 481–516

Roy AK, Neuhaus OW (1966) Identification of rat urinary protein by zone and immunoelectrophoresis. Proc Soc Exp Biol Med 121, 894–899

Roy AK, Raber DL (1972) Immunofluorescent localization of α_{2u} globulin in the hepatic and renal tissues of rat. J Histochem Cytochem 20, 89–96

Tocharoentanaphol C, Cremer M, Schrock E, Blonden L, Kilian K et al. (1994) Multicolor fluorescence in situ hybridization on metaphase chromosomes and interphase Halo-preparations using cosmid and YAC clones for the simultaneous high resolution mapping of deletions in the dystrophin gene. Hum Genet 93, 229–235

Trask B, Pinkel D, van den Engh G (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. Genomics 5, 710–717

Tumlinson JH, Yonce CE, Doolittle RE, Heath RR, Gentry CR et al. (1974) Sex pheromones and reproductive isolation of the lesser peachtree borer and the peachtree borer. Science 185, 614–616

Van den Engh G, Sachs R, Trask BJ (1992) Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. Science 257, 1410–1412

Wang KS, Hodgetts RB (1998) The analysis of regulatory sequences required for high level induction of rat alpha 2u-globulin gene expression by dexamethasone. J Mol Endocrinol 20, 129–139

Wang KS, McFadyen DA, Locke J, Hodgetts RB (1997) Three subsets of genes whose tissue specific expression is sex and age-dependent can be identified within the rat α_{2u}-globulin family. Dev Genet 21, 234–244

Weier H-UG, Wang M, Mullikin JC, Zhu Y, Cheng J-F et al. (1995) Quantitative DNA fiber mapping. Hum Mol Genet 4, 1903–1910

Wiegant J, Kalle W, Mullenders L, Brookes S, Hoovers JMN et al. (1992) High resolution in situ hybridization using DNA halo preparations. Hum Mol Genet 1, 587–591

Yanai J, McClearn GE (1973) Assortive mating in mice 3. Genetic determination of female mating preference. Behav Genet 3, 75–84