Purification and biochemical characterization of interchromatin granule clusters

Paul J.Mintz^{1,2}, Scott D.Patterson³,
Andrew F.Neuwald², Chris S.Spahr³ and
David L.Spector^{2,4}

¹Department of Molecular Genetics and Microbiology, S.U.N.Y. Stony Brook, Stony Brook, NY 11794, ²Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724 and ³Amgen Center, MS14-2-E, 1 Amgen Center Drive, Thousand Oaks, CA 91320-1789, USA

⁴Corresponding author
e-mail: spector@cshl.org

Components of the pre-mRNA splicing machinery are localized in interchromatin granule clusters (IGCs) and perichromatin fibrils (PFs). Here we report the biochemical purification of IGCs. Approximately 75 enriched proteins were present in the IGC fraction. Protein identification employing a novel mass spectrometry strategy and peptide microsequencing identified 33 known proteins, many of which have been linked to pre-mRNA splicing, as well as numerous uncharacterized proteins. Thus far, three new protein constituents of the IGCs have been identified. One of these, a 137 kDa protein, has a striking sequence similarity over its entire length to UV-damaged DNA-binding protein, a protein associated with the hereditary disease xeroderma pigmentosum group E, and to the 160 kDa subunit of cleavage polyadenylation specificity factor. Overall, these results provide a key framework that will enable the biological functions associated with the IGCs to be elucidated.

Keywords: interchromatin granule clusters/nuclear structure/RNA processing/speckles/xeroderma pigmentosum

Introduction

The mammalian cell nucleus is a highly organized structure in which DNA replication, transcription, pre-mRNA processing, ribosome biogenesis and RNA transport take place (for reviews, see Spector, 1993; Lamond and Earnshaw, 1998). The biochemical constituents and molecular mechanisms of these processes are reasonably well understood. However, less is known about how each of these processes is coordinated spatially and temporally within the structural framework of the nucleus.

At present, only a few nuclear structures have been isolated by biochemical fractionation, i.e. nucleoli, nuclear pore–lamina complex and heterogeneous nuclear ribonucleoprotein particles (for a review, see Singer and Green, 1997). The isolation of these structures has facilitated the identification and functional characterization of their constituents, which enabled their biological roles to be substantiated. However, there are numerous other sub-

nuclear structures including coiled bodies, the perinuclear compartment (PNC), gems, cleavage bodies, promyelocytic leukemia (PML) bodies and interchromatin granule clusters (IGCs) which are less well characterized at the functional level because thus far it has not been possible to isolate them biochemically (for reviews, see Spector, 1993; Lamond and Earnshaw, 1998).

Pre-mRNA splicing factors have been localized by immunofluorescence microscopy in a speckled nuclear distribution pattern which corresponds to perichromatin fibrils (PFs) and IGCs at the electron microscopic level (for a review, see Spector, 1993). PFs are fibrillar structures measuring 3–5 nm in diameter and are often found at the periphery of IGCs and distributed throughout the nucleoplasm (Monneron and Bernhard, 1969). In contrast to IGCs, PFs are labeled rapidly after short pulses with [³H]uridine, suggesting that they represent nascent transcripts (for a review, see Fakan, 1994). In support of this view, PFs are sensitive to RNase digestion and are reduced in cells treated with drugs that inhibit transcription by RNA polymerase II (Monneron and Bernhard, 1969; Petrov and Sekeris, 1971).

IGCs, first identified by Swift (1959), represent the major part of the speckled staining pattern. IGCs measure 0.3–1.8 µm along their longest dimension (Thiry, 1995) and are composed of granules measuring 20–25 nm in diameter that appear to be connected in places by a 9–10 nm fiber (for reviews, see Fakan and Puvion, 1980; Spector, 1993). IGCs contain little to no DNA and are highly resistant to nuclease, detergent and high salt extractions (Monneron and Bernhard, 1969; Spector *et al.*, 1983). Bromo-UTP and [³H]uridine incorporation studies have shown little to no labeling within IGCs, demonstrating that they are not centers of active transcription (for a review, see Spector, 1993). However, labeling has been found at the periphery of IGCs (Fakan and Nobis, 1978; Spector, 1990). Consistent with these findings, several studies have shown highly active genes and their transcripts to be associated with the periphery of speckled domains (for a review, see Huang and Spector, 1997). However, this association does not appear to be essential, as transcription can occur throughout the nucleoplasm (Zhang *et al.*, 1994).

Splicing factors have been shown to be recruited from IGCs to sites of active transcription (Jiménez-García and Spector, 1993; Huang and Spector, 1996; Misteli *et al.*, 1997). The molecular mechanism(s) for the recruitment of splicing factors to the sites of transcription is not completely understood. However, it has been shown that phosphorylation of the RS domain of SR proteins is involved in releasing these factors from IGCs and that dephosphorylation is at least part of the signal for the return of these factors to the IGCs (Misteli and Spector, 1996; Misteli *et al.*, 1998). Recently, the C-terminal

Fig. 1. (A) Transmission electron micrograph of mouse liver nucleus. Thin section of a nucleus showing interchromatin granule clusters (IGCs) (arrowheads). Bar: 1 μ m. (B) Flow chart of the biochemical purification of IGCs.

domain (CTD) of the largest subunit of RNA polymerase II has been shown to be involved in the intranuclear targeting of splicing factors to transcription sites *in vivo* (Misteli and Spector, 1999). Truncation of the CTD prevented the accumulation of splicing factors at a newly formed site of transcription, resulting in the inhibition of pre-mRNA splicing.

While much has been learned about the nuclear distribution of IGCs (for reviews, see Fakan and Puvion, 1980; Spector, 1993; Thiry, 1995), the complete protein composition of these nuclear structures is not known. Elucidating the protein composition of IGCs is central to revealing the functions associated with this nuclear structure. We developed a biochemical nuclear fractionation protocol that enabled us to separate IGCs from chromatin, nuclear lamins, nucleoli and other nuclear bodies. We employed a novel mass spectrometry strategy and peptide microsequencing to identify rapidly many of the 75 proteins that are enriched in this nuclear structure. These data provide a key framework that will enable the biological functions associated with the IGCs to be elucidated.

Results

Biochemical purification and morphological characterization of IGCs

Adult mouse liver nuclei contain 20–30 irregularly shaped IGCs each measuring ~ 0.3 – 1.8 μ m along their longest axis (Figure 1A). In order to elucidate the biochemical composition of IGCs, we have developed a cell fractionation procedure to isolate IGCs from mouse liver nuclei (Figure 1B). Nuclei were isolated by sucrose gradient sedimentation (Berezney and Coffey, 1977). This procedure produced clean nuclei with little to no cytoplasmic contamination (Figure 2A).

In order to purify IGCs away from other nuclear

structures, we proceeded first to remove the nuclear membrane by using a non-ionic detergent, 1% Triton X-100 (data not shown). Next, DNA was digested with DNase I and chromatin was released by several sequential extractions with 0.5 M NaCl. After the nuclease digestion and salt extractions, three major types of nuclear structures still remained visible: nuclear lamina, nucleoli and IGCs (Figure 2B). The IGCs remained well preserved whereas the nucleoli appeared to have been extracted after the 0.5 M NaCl treatments. In order to separate the IGCs from other nuclear structures and release them from the nuclear framework, the nuclei were disrupted mechanically in the presence of 5 mM dithiothreitol (DTT) and the resultant homogenate was layered onto a 0.25 M cesium sulfate solution followed by low-speed centrifugation. The cesium sulfate pellet contained predominantly nuclear lamina, nucleoli and residual nuclear components (Figure 2C). The IGCs, which were enriched in the cesium sulfate supernatant, were retrieved by high-speed centrifugation (Figure 3A and B). Figure 3A shows a low magnification electron micrograph of a thin section of the IGC pellet revealing a relatively homogenous population of IGCs and little to no apparent contamination by other nuclear structures. The isolated interchromatin granules measured between 18 and 27 nm in diameter (Figure 3B), similar to their *in situ* counterparts. To confirm that the purified IGCs contained splicing factors, thin sections were immunolabeled with 3C5 monoclonal antibody (Turner and Franchi, 1987), which recognizes a family of SR proteins, followed by colloidal gold-conjugated secondary antibody. The colloidal gold particles labeled the IGCs (Figure 3C), whereas the control sections, which were not incubated with primary antibody, showed little to no labeling (Figure 3D). These data demonstrate, at a structural level, the high degree of purity of the IGC fraction.

Fig. 2. Morphological characterization of the IGC fractionation. All samples were post-stained using the EDTA-regressive method which preferentially stains ribonucleoproteins (Bernhard, 1969). (A) The nuclear pellet contains well preserved nuclei. The IGCs, nucleoli (N), nuclear lamina (L) and chromatin (C) are readily observed. (B) Thin section of the nuclear pellet after 1% Triton X-100, DNase I and 0.5 M NaCl treatments. The IGCs are still well preserved, whereas the nucleoli appear partially extracted. Little to no chromatin is present in the extracted nuclei. (C) The cesium sulfate pellet contains nuclear lamina (L), nucleoli (N) and residual nuclear components (asterisk). Bars: 2 μ m.

Biochemical composition of IGCs

To investigate the protein composition of each step in the IGC purification, we performed SDS-PAGE using 4–20% gradient gels. Such analyses showed a complex protein composition in all of the fractions examined (Figure 4A).

Different proteins were enriched in various fractions. For example, a group of proteins in the 10–15 kDa range were enriched in the Triton pellet, several proteins in the 60–70 kDa range were enriched in the cesium sulfate pellet, and a group of three proteins around 220 kDa were enriched in the IGC pellet. Previous studies have used immunopurification to obtain highly specific macromolecular complexes (e.g. hnRNPs) or organelles (Kvalheim *et al.*, 1987; Piñol-Roma *et al.*, 1988; Saucan and Palade, 1994). In order to address whether the protein composition of the purified IGC pellet fraction is specific, and not due to other proteins randomly co-sedimenting with the IGCs, we have purified IGCs further by using magnetic beads coupled with a splicing factor antibody (3C5). The immunopurified IGCs showed a protein composition similar to that of the isolated IGC fraction when examined by SDS-PAGE (Figure 4A and B, lanes 5 and 1, respectively). As shown in the control lanes (Figure 4B, lanes 2 and 3), there was little to no binding of the IGCs to beads that did not contain coupled primary antibody. Since the protein composition of the immunopurified fraction was similar to that of the initial IGC pellet fraction, all further analyses were performed on the IGC pellet fraction.

In order to identify at greater resolution the constituents of the IGCs, purified IGCs were analyzed by two-dimensional gel electrophoresis. Approximately 300 protein spots were identified in the IGC fraction (Figure 4D), 75 of which were enriched when compared with the nuclear fraction (Figure 4C). The degree of enrichment for each spot varied. For example, a series of three spots just above the 20 kDa marker were almost undetectable in the two-dimensional gel of whole nuclei and were significantly enriched in the IGC fraction (Figure 4C and D). Two additional spots, just below the 10 kDa marker, were present in the nuclear fraction and were significantly enriched in the IGCs.

Since IGCs are known to contain pre-mRNA splicing factors, we were interested next in determining their presence in the purified IGC fraction as a means of further assessing its purity. This was accomplished by immunoblotting the purified IGC fraction. Several pre-mRNA splicing and non-splicing factors including SF3a66, SF2/ASF, U2AF⁶⁵ and U2AF³⁵ were enriched in the IGC fraction (Figure 5A, top row). We found that several other pre-mRNA processing-related proteins, U1-70K, U2-B'' and PAB II, while present in the IGC fraction, were not enriched (Figure 5A, bottom row). This may be due to their more loose association with the IGCs. Interestingly, the hyperphosphorylated form of the largest subunit of RNA polymerase II is also present in the IGC fraction, but is not enriched. This result was not unexpected since it has been shown that this form of the polymerase is present in nuclear speckles (Bregman *et al.*, 1995). Several other nuclear proteins that have been shown by immunofluorescence to be localized to other nuclear compartments or to be distributed diffusely throughout the nucleoplasm, such as hnRNP A1, lamins A and C, ribosomal protein S6 and the PML protein show little to no immunoreactivity in the IGC fraction (Figure 5B). It must be pointed out here that it is not possible to quantitate the fold enrichment for the IGCs accurately since thus far all of the proteins that have been localized to this structure are also present diffusely throughout the nucleoplasm. The diffuse nuclear

Fig. 3. Transmission electron micrographs of the purified IGC fraction. (A) Low magnification view of the purified IGCs showing a homogenous population Bar: 0.5 μ m. (B) High magnification view of the IGCs showing intact and well-preserved interchromatin granules. Differences in staining intensity reflect the plane of section. Bar: 200 nm. (C) The IGC fraction was labeled with monoclonal antibody 3C5, which recognizes phosphorylated SR proteins, followed by 5 nm colloidal gold-conjugated secondary antibody Bar: 200 nm; inset: 50 nm. (D) Control section labeled only with the secondary antibody shows little to no immunogold labeling. Bar: 200 nm.

pool represents factors that are at transcription sites as well as those that are in transit between IGCs and transcription sites.

To understand better where in our fractionation scheme proteins associated with different nuclear compartments are enriched preferentially, we immunoblotted all of the pellet fractions from each step in the IGC purification. Figure 6A shows that lamin B1 is enriched predominantly in the cesium sulfate pellet, consistent with our observations of this fraction by transmission electron microscopy (Figure 2C). Interestingly, nucleolar protein B23 (nucleophosmin) (Ochs *et al.*, 1983) was not enriched in the IGC or cesium sulfate pellets (Figure 6B, lanes 5 and 4, respectively). Most of nucleolar protein B23 was removed by the three 0.5 M NaCl extractions (Figure 6B, lane 3). Previous studies have shown that some hnRNPs are also extractable from nuclei at 0.5 M NaCl (Beyer *et al.*, 1977; Piñol-Roma *et al.*, 1988). We examined the distribution of hnRNP C1/C2 proteins across our fractionation and found that most of these proteins were removed by the nuclease digestion and salt extractions (Figure 6C, lane 3). We have also examined the distribution of SR proteins and found that most of the SR proteins were enriched in the IGC fraction when compared with the nuclear pellet (Figure 6D, lanes 5 and 1, respectively). In addition, there was little to no immunoreactivity in the cesium sulfate pellet. These data, along with the SDS-PAGE analysis, strongly support the view that the IGCs were biochemically

purified and demonstrate the complex protein composition of IGCs.

Mass spectrometry analysis

In recent years, analysis by mass spectrometry has been a powerful approach used to identify the protein components of large macromolecular complexes rapidly by subjecting individual gel bands to analysis (for reviews, see Lamond and Mann, 1997; Patterson, 1998). For example, several recent reports have used this approach to identify the protein constituents of *in vitro* assembled spliceosomes (Neubauer *et al.*, 1998), yeast U1 small nuclear ribonucleoprotein particles (Neubauer *et al.*, 1997) and the yeast spindle pole (Wigge *et al.*, 1998). Here we have used a novel mass spectrometry approach to identify many of the IGC proteins rapidly. The entire IGC protein mixture was digested enzymatically and subjected to liquid chromatography electrospray ionization tandem mass spectrometry (LC-MS/MS) in a data-dependent manner followed by uninterpreted fragment ion searching of non-redundant and expressed sequence tag databases (dbEST). This is the first time that a large and complex protein mixture has been digested and subjected to mass spectrometry analysis *en toto*. Purified IGCs (denatured, reduced and alkylated) were first subjected to partial trypsin digestion. The complex peptide mixture was fractionated by HPLC with on-line ion-trap electrospray mass spectrometry. In two separate runs, we have thus far identified 33 known

Fig. 4. One- and two-dimensional gel electrophoresis. (A) All pellet fractions from the IGC purification were separated on a 4–20% SDS gradient gel. Equal amounts (40 μg) of protein were loaded. (B) Magnetic beads coupled with 3C5 monoclonal antibody were used to purify the IGCs further. Immunopurified IGCs (lane 1) show a protein composition similar to that of the biochemically purified IGCs. Control beads (lanes 2 and 3) show no non-specific proteins. (C and D) The IGC or nuclear pellets were first separated by 2.7% polyacrylamide using pH 3.5–10 ampholites in the first dimension followed by 12.5% SDS–PAGE in the second dimension. The proteins were visualized by silver staining. (C) Nuclear pellet (100 μg). (D) IGC pellet (100 μg). Approximately 75 protein spots were enriched in the IGC fraction as indicated by the circled regions.

proteins (Table I) and ESTs encoding at most 16 proteins (to date) after searching a non-redundant protein database or dbEST (NCBI and DDBJ/EMBL/GenBank) with the uninterpreted MS/MS spectra. A group of 19 proteins (identified from 71 spectra), previously shown to be localized to IGCs, is composed largely of splicing factors. One member of this group, a previously characterized RNA/DNA-binding protein (RNPS1) has been identified recently as SF7A, a general activator of pre-mRNA splicing (A.R.Krainer and A.Mayeda, personal communication). In addition to splicing factors, two proteins related to different types of cancer were identified. A peptide

sequence corresponding to a region of the Skip protein was identified. Skip has been shown to be associated with the Ski oncoprotein (Dahl *et al.*, 1998). The Skip protein was also identified when spliceosomes were analyzed by mass spectrometry (Neubauer *et al.*, 1998) and, in addition, it was shown to be localized to nuclear speckles. A second protein, TLS-associated serine-arginine (TASR) protein (Yang *et al.*, 1998), which is associated with the TLS (translocated in liposarcoma) protein, was also identified in the IGC fraction. In addition, our analysis identified a subset of proteins (14 identified from 22 MS/MS spectra) that have been characterized in a different biological

Fig. 5. Immunoblot analysis of the IGC fraction. (A) Immunoblots in the upper row show several pre-mRNA splicing factors that are enriched in the IGC fraction. The lower row shows several other splicing factors and RNA polymerase II to be present, but not enriched, in the IGC fraction. (B) Immunoblots using antibodies that recognize proteins associated with other cellular compartments show little to no immunoreactivity in the IGC fraction. Equal amounts of protein (10 μ g) were loaded.

context and thus far have not been shown to be present in IGCs (Table I). Some of these proteins (e.g. histones and ribosomal proteins) may be contaminants based upon their previous functional associations. However, we cannot

rule out the possibility that they may also be minor constituents of the IGCs. Finally, we have identified several new proteins (Table I) and ESTs (data not shown) associated with the IGCs.

The identified peptide sequences were used to screen protein databases. Two peptide sequences SLLPNSS-QDELMEVEK and LGLIQEDVASSCIPR (matched to Amgen EST sequences) were used to identify a full-length human cDNA clone, KIAA0324, after searching a non-redundant protein database (DDBJ/EMBL/GenBank accession No. AB002322). The open reading frame (ORF) of KIAA0324 contains many RS/SR dipeptide repeats and a putative nuclear localization signal (data not shown). We did not find any potential homologs of KIAA0324 in the databases. The predicted molecular mass of this protein is ~125.6 kDa and it is a highly basic protein ($pI = 12$). To verify that this protein is a bona fide constituent of the IGCs, a fusion protein was made with yellow fluorescent protein (YFP) and expressed in HeLa cells by transient transfection (Figure 7A). The transfected cells were also immunostained with anti-SC35 monoclonal antibody which diagnostically identifies nuclear speckles (Figure 7B). The expressed fusion protein co-localizes with the endogenous SC35 splicing factor (Figure 7C). In addition, the fusion protein was also present in the cytoplasm, suggesting that it may shuttle.

A third peptide (AVTIATPATAAPAAVSAATTTSAQE-EPAAPEPR) was matched to a full-length mouse cDNA clone called Plenty-of-Proline (POP) (DDBJ/EMBL/GenBank accession No. AF062655). The POP protein is highly basic ($pI = 11.9$) and has a predicted molecular mass of ~101 kDa. The ORF of POP contains many RS/SR dipeptide repeats, several proline-rich regions and a putative NLS (data not shown). A homolog to POP, SRm160, was identified (DDBJ/EMBL/GenBank accession No. AF048977), which is a human splicing co-activator previously localized to speckles (Blencowe *et al.*, 1998). Using a sequence alignment program (NCBI, BLAST 2 Sequences), POP shares 90% identity and 92% similarity with SRm160 (data not shown). We have also made a POP-YFP fusion protein and have confirmed that POP is a bona fide component of the IGCs (data not shown). The remaining 16 ESTs currently do not match to any identified full-length cDNAs.

Microsequence analysis of the 140 kDa protein

In addition to mass spectrometry, peptide microsequencing of gel bands was used to identify new proteins of the IGC fraction. An enriched 140 kDa gel band of the IGC fraction was excised from a Coomassie-stained SDS-polyacrylamide gradient gel for microsequencing. The peptide sequences obtained (KFVIHPESNNLIIEETD and KNVSEELDRTPPEVSK) were used to search a non-redundant protein database, and a full-length cDNA clone, KIAA0017, was identified (DDBJ/EMBL/GenBank accession No. D13642). The protein is acidic ($pI = 5.1$) and has a predicted molecular mass of ~136.6 kDa. The ORF of KIAA0017 contains a putative NLS (data not shown); however, no other known motifs were identified.

Database searching using the KIAA0017 sequence has allowed us to identify closely related proteins in other metazoans (Figure 8) (*Caenorhabditis elegans* and *Drosophila melanogaster*), in a plant (*Arabidopsis thaliana*) and

Fig. 6. Further characterization of purified IGCs. Immunoblots were probed with antibodies against (A) lamin B1, (B) nucleolar protein B23, (C) hnRNP C1/C2 or (D) 3C5 antibody to a family of SR proteins. Proteins associated with different nuclear structures were enriched in different fractions.

in a protozoan (*Plasmodium falciparum*), suggesting that human protein KIAA0017 performs an important, perhaps essential, biological function that arose early during eukaryotic evolution. Given their high degree of sequence identity to KIAA0017 (43–61%), these proteins are likely to perform the same function in these diverse organisms as KIAA0017 does in humans. Notably, even budding yeast, which lacks IGCs, encodes a member of this family. The function of this yeast protein may be somewhat distinct from that of these other proteins, however, given its lower degree of sequence identity to KIAA0017 (25% identity).

A biological function for the KIAA0017 protein family is suggested by weaker, yet clearly significant, sequence similarity to two other protein families: a family of known and putative UV-damaged DNA-binding factors (UV-DDB) (E-value = 10^{-27}) and a family of RNA cleavage and polyadenylation specificity factors (CPSF, 160 kDa subunit) (E-value = 10^{-10}). Of these two families, KIAA0017 is more closely related to UV-DDB proteins which are associated with the hereditary disease xeroderma pigmentosum group E (Keeney *et al.*, 1994).

The genomes of higher eukaryotes appear to encode one member from each of the KIAA0017, UV-DDB and CPSF (160 kDa subunit) protein families. This is seen, for example, for three metazoans, human, *C.elegans* and *Drosophila*, whose genomes are close to being sequenced completely. Thus, all three proteins may be present in most, if not all, metazoans. Similarly, the vascular plant *Arabidopsis* also encodes all three proteins, although the *Arabidopsis* proteins closely related to KIAA0017 and CPSF currently are available only as short, partially sequenced fragments. By contrast, budding yeast, whose entire genome has been sequenced, lacks an UV-DDB-like protein; the lack of this protein in yeast suggests a potential functional overlap between KIAA0017 and UV-DDB. Based on available complete genomic sequences, all three families appear to be absent from archaea and

eubacteria, which is consistent with a distinctly nuclear role for these proteins.

Remarkably, the similarity between these three families, though characterized by scattered regions of sequence conservation, extends essentially over the entire length of the proteins (Figure 8). This is demonstrated through PSI-BLAST searches using profiles derived from four distinct, adjacent ~300 residue regions of the KIAA0017 family alignment; each of these searches easily detects the UV-DDB and CPSF families (E-values = 10^{-6}), with the exception of one region that is absent in CPSF proteins. This region, which maps to residues 575–705 of KIAA0017 (Figure 8, bracket), corresponds to an unrelated region within CPSF proteins; surprisingly, even CPSF proteins from different organisms appear unrelated in this region. By contrast, the 575–705 region of KIAA0017 is conserved, albeit weakly (E-value = 0.0002), in UV-DDB proteins, again suggesting greater functional similarity between these two families. As a whole, the co-linear motifs conserved in these proteins, which are very likely to correspond to multiple domains, appear to be present only in these three families and are conspicuously absent from other currently sequenced proteins. This suggests an unusual degree of functional dependency between these conserved domains.

To confirm that the KIAA0017 protein is a constituent of the IGCs, the protein was tagged with YFP and expressed in BHK cells by transient transfection (Figure 7D). The expressed fusion protein (Figure 7D) co-localizes with the endogenous U2 snRNP B' protein (Figure 7F), confirming the relevant localization of the expressed fusion protein. In addition to being localized in the nucleus, the fusion protein was also present in the cytoplasm.

Discussion

Previous studies have used immunofluorescence, *in situ* hybridization and electron microscopy to assess the bio-

Table I. Proteins identified in purified IGC fraction

	Peptide sequence	No. of peptides identified	No. of times identified	Accession No.
Previously characterized proteins				
Splicing factor SF2/ASF	IYVGNLPPDIR	5	7	Q07955
Splicing factor SC35	ELRVQMAR	2	3	Q01130
Splicing factor SRp20	VYVGNLGNNGNKTELER	4	7	X91656
Splicing factor SRp55	ALDKLDGTEINGR	1	2	U30829
Splicing factor SRp75	LIVENLSSR	3	5	Q08170
U5 snRNP 116 kDa protein	IAVEPVNPSELPK	1	1	U97079
snRNP Sm D2 protein	NNTQVLINCR	1	1	P43330
snRNP Sm D3 protein	VAQLEQVYIR	1	1	P43331
snRNP-associated protein	RVLGLVLLR	4	4	P27048
RNA/DNA-binding protein (RNPS1)	GYAYVEFENPDEAEK	1	1	S43417
hnRNP A2/B1	QEMQEVQSSR	1	1	P22626
hnRNP A3	SSGSPYGGGYGSGGGSGGYGSR	1	2	P51991
hnRNP M	MGPAMGPALGAGIER	10	17	P52272
hnRNP G	DYAPPPRDYTYR	4	6	P38159
hnRNP D	IFVGGLSPDTPEEK	1	1	D55671
hnRNP E1	IANPVEGSSGR	1	2	S42472
Nuclear protein Skip	YTPSQQGVAFNSG	1	2	Q13573
TLS-ass. with SR repeats (TASR)	YLRPPNTSLFVR	3	5	AF042383
Heat shock protein 70	TTPSYVAFDTER	3	3	P08109
Lamin A	SGAQASSTPLSPTR	1	1	P48678
Lamin B1	ASAPATPLSPTR	1	1	P14733
Lamins C1/C2	SLETENAGLR	3	3	P11516
nhp2/rs6 family protein	LLDLVQQSCNYK	3	3	P55770
CBF5	VAKLDTSQWPLLLK	1	1	U59151
Histone H2a	AGLQFPVGR	3	6	S04152
Histone H2b	AMGIMNSFVNDIFER	3	6	P10853
Histone H3	DIQLAR	1	2	P06351
Histone H4	DNIQGITKPAIR	3	4	S03427
Importin β -1 subunit	VLANPGNSQVAR	1	1	P70168
Ribosomal protein S11	YYKNISSR	1	1	C76740
60S ribosomal protein L7a	AGVNTVTTLVENK	1	1	P12970
60S ribosomal protein L6	AVDLQILPK	1	1	P47911
60S ribosomal protein L18	ILTFDQLALESPK	1	1	P35980
New IGC proteins				
Plenty-of-proline (POP) ^a	AVTIATPATAAPAAVSAATTTSA....	1	3	W13917
KIAA0324 ^a	SLLPNSSQDELMEVEK	2	4	AB002322
KIAA0017 ^{a,b}	KFVIHPESNNLIHETD	2	2	D13642

^aCo-localization of YFP fusion proteins with an endogenous splicing factor *in vivo*.

^bProtein identified by peptide microsequencing.

logical function of nuclear speckles (IGCs) in the mammalian cell nucleus (for a review, see Misteli and Spector, 1998). These studies have suggested that nuclear speckles play an important role in RNA metabolism based upon the localization of pre-mRNA splicing factors to this nuclear domain. Recently, nuclear speckles were shown to be dynamic in living cells and to provide splicing factors to sites of transcription which are located on their periphery or some distance from a speckle (Misteli *et al.*, 1997). In order to advance our understanding of nuclear speckles (IGCs) significantly, we have taken a biochemical cell fractionation approach to isolate IGCs from mouse liver nuclei. Here we report the first successful biochemical purification and characterization of IGCs and the identification of three new IGC proteins.

Enormous progress has been made in our understanding of the functional and structural organization of eukaryotic cells. In particular, the highly compartmentalized nature of the cytoplasm and the functions associated with both membrane-bound (mitochondria, Golgi apparatus, endoplasmic reticulum, microbodies, etc.) and non-membrane-bound (ribosomes, microtubules, microfilaments, intermediate filaments, centrioles, etc.) organelles have

been identified readily using standard microscopy approaches and fractionation procedures developed early on to purify and characterize these structures (Novikoff and Holtzman, 1970). However, the functional organization of the nucleus remains one of the black boxes in cell biology, primarily because it has been difficult to isolate and characterize biochemically many of the identified nuclear structures such as coiled bodies, gems, cleavage bodies, the PNC, PML bodies, IGCs, PFs, etc. Recently, Singer and Green (1997) argued that nucleoli and the NPC, both of which are non-membrane-bound organelles, are bona fide nuclear compartments because they can be identified microscopically and purified biochemically. According to these criteria, IGCs now also qualify as a bona fide nuclear organelle.

Since the identification of interchromatin granule clusters 40 years ago by Swift (1959), a single published attempt has been made to isolate this nuclear structure (Zborek *et al.*, 1990). However, the isolated IGCs could not be fractionated away completely from a large number of nuclear structures including nucleoli, nuclear lamins and other residual nuclear components. We have employed a series of approaches in an effort to characterize purified

Fig. 7. Characterization of KIAA0324 and KIAA0017. The human full-length KIAA0324 (A–C) or KIAA0017 (D–F) cDNAs were fused to yellow fluorescent protein (YFP). HeLa (A–C) or BHK (D–F) cells were transiently transfected and immunostained with the anti-SC35 (B) or anti-B'' (E) antibodies. The fusion proteins localized in nuclear speckles and diffusely in the cytoplasm. (C) and (F) are the respective merged images of (A) and (B) and (D) and (E). Bars: 10 μ m; 5 μ m.

IGCs morphologically and biochemically. We have isolated IGCs successfully from other nuclear components (lamins, nucleoli and chromatin) without compromising the structural integrity of the granules. Approximately 300 protein spots were identified in the IGC fraction using two-dimensional gel electrophoresis. We have used a novel mass spectrometry methodology and protein microsequencing to identify the constituents of IGCs. Thus far, 33 known proteins have been identified as well as several uncharacterized ESTs. Of the 33 known proteins, 19 have been shown to be present in the speckles and, as expected, many of them are splicing factors. These data strongly demonstrate that we have purified a highly enriched IGC fraction. Recently, Neubauer *et al.* (1998) have employed mass spectrometry and have identified 19 new proteins associated with an *in vitro* assembled spliceosomal complex. Several of the proteins we have identified in the IGC fraction, including Skip, U2AF³⁵, U2AF⁶⁵, hnRNP C and U2-snRNP B'', were also found in the purified spliceosome complex. However, members of the SR family that are enriched in the IGCs were not identified in the purified *in vitro* assembled spliceosome fraction (Neubauer *et al.*, 1998). One interpretation of these data is that the SR proteins are more tightly associated with the IGCs than they are in the *in vitro* assembled spliceosome.

Most interestingly, analysis of the IGC fraction has allowed us to identify 17 proteins with unknown functions. Thus far, we have investigated three of these unknown proteins (KIAA0324, POP and KIAA0017) and found that they localize *in vivo* in speckles, in addition to being distributed diffusely throughout the nucleoplasm. Interestingly, the KIAA0324 and KIAA0017 proteins were also localized in the cytoplasm, consistent with the possibility that they may shuttle between the nucleus and cytoplasm. A growing number of nuclear proteins have been shown

to shuttle between the nucleus and cytoplasm (for a review, see Nakielnny and Dreyfuss, 1997). After searching a non-redundant protein database, we found that the KIAA0017 protein has sequence similarity throughout its ORF with a UV-damaged DNA-binding (UV-DDB) factor in human, slime mold, *C.elegans*, *Schizosaccharomyces pombe* and *A.thaliana*, suggesting that all of these proteins are related. The UV-DDB protein (127 kDa) is involved in the damage recognition step of the nucleotide excision repair (NER) pathway (Hirschfeld *et al.*, 1990; Treiber *et al.*, 1992; Otrin *et al.*, 1997). Our finding of an additional cytoplasmic pool of KIAA0017 is consistent with a recent report by Watanabe *et al.* (1999) showing that the UV-DDB protein also interacts with the cytoplasmic domain of the Alzheimer's amyloid precursor protein (APP) in co-immunoprecipitation experiments. At present, the precise biological role of UV-DDB is unknown. In addition to the UV-DDB protein, the 160 kDa subunit of CPSF from *Saccharomyces cerevisiae*, *S.pombe* and human cells was also found to have conserved sequence similarity to the KIAA0017 ORF. The sequence similarity over their entire length suggests that all three families share common structural/functional constraints that have been evolutionarily conserved. Taken together, this sequence similarity between several distant species suggests that KIAA0017 is involved in one or more important biological functions, possibly associated with DNA repair and RNA processing.

Although the biological relevance of IGCs has been well established during the past 40 years, questions have remained concerning the biological nature of nuclear speckles, giving rise to the suggestion that speckles are the result of exaggerated antibody labeling conditions or artificial threshold effects during digital image acquisition (Fay *et al.*, 1997; Neugebauer and Roth, 1997). Numerous

studies have presented conclusive evidence that speckles are bona fide non-membrane-bound organelles. First, nuclear speckles (IGCs) can be visualized as a unique structure

in the electron microscope without the need for antibody labeling (Swift, 1959; Monneron and Bernard, 1969). Secondly, nuclear speckles can be visualized in living

[illegible]

cells using splicing factor–GFP fusion proteins (Misteli *et al.*, 1997). Finally, as we now show, nuclear speckles can be purified biochemically, and new protein members of the isolated speckles are localized to these structures when their cDNAs are expressed in living cells. Our biochemical strategy will enable the elucidation of the functions associated with this nuclear organelle and advance our overall understanding of the functional/structural organization of the nucleus.

Materials and methods

IGC purification

Livers (29–33 g) from 20 five-week-old Swiss Webster female mice were washed once in ice-cold TM5 buffer (10 mM Tris–HCl, pH 7.4 and 5 mM MgCl₂, pH 7.4) containing protease inhibitors [1 mM phenylmethylsulfonyl fluoride (PMSF), 5 µg/ml leupeptin and aprotinin (Sigma, St Louis, MO)] and minced. A 6 ml aliquot of 0.25 M sucrose in TM5 buffer was added to 2 ml of minced liver in a 55 ml Dounce homogenizer (Wheaton) and homogenized using the 'A' pestle. The homogenate was transferred to a 50 ml tube (Corning, NY) and was filtered twice through two layers of cheesecloth. An equal volume of 2.2 M sucrose in TM5 buffer containing 0.2 mM spermine and 0.5 mM spermidine was added to the homogenate and mixed vigorously. The homogenate was transferred to 30 ml conical centrifuge tubes (Oak Ridge, TN) containing 5 ml each of 2.2 M sucrose (Berezney and Coffey, 1977) and was centrifuged at 39 000 g for 1.5 h at 4°C. The nuclear pellet was washed once with TM5 buffer. On average, 4–6 × 10⁸ nuclei/ml were recovered. Small samples (3 µl) of the nuclear pellet were taken for examination by phase-contrast microscopy and transmission electron microscopy.

To isolate IGCs, 4 ml of a nuclear suspension (5 × 10⁸ nuclei/ml) were treated with Triton X-100 to a final concentration of 1% in the presence of 2 mM vanadium ribonucleoside complex (Gibco-BRL, Frederick, MD) for 5 min on ice and centrifuged at 780 g for 5 min. The supernatant (and all subsequent supernatants) was removed and analyzed by SDS–PAGE. The pellet was resuspended in a final volume of 4 ml with TM5 buffer containing protease inhibitors (5 µg/ml leupeptin and aprotinin) and 3 µl of the pellet was set aside for analysis by transmission electron microscopy (Triton pellet). A 880 µl aliquot of the resuspended Triton X-100 extracted nuclei was placed into four microcentrifuge tubes and 20 µl of RNase-free DNase I (10 U/µl) (Boehringer Mannheim) was added and incubated for 1 h at 4°C on a slow rotator. The digested chromatin was removed by addition of 5 M NaCl, to a final concentration of 0.5 M, and incubated on ice for 5 min, then centrifuged at 770 g for 5 min at 4°C. Each pellet was resuspended in 500 µl of 0.5 M NaCl in TM5 buffer and treated as above twice. Three microlitres of the final pellet (DNase I/0.5 M NaCl pellet) was analyzed by transmission electron microscopy. Each pellet was resuspended in 450 µl of 0.5 M NaCl in TM5 buffer and divided into two microcentrifuge tubes. The volume was adjusted to 955 µl with 0.5 M NaCl in TM5 buffer, and 5 µl of 1 M DTT was added and incubated for 5 min on ice. The homogenates were passed 10 times through a 27-gauge needle, pooled and homogenized 20 times using a tight pestle (1 ml Dounce homogenizer, Wheaton). Aliquots (100 µl) of the homogenate were transferred to 16 microcentrifuge tubes containing 500 µl of 0.25 M Cs₂SO₄ in TM5 buffer and centrifuged at 20 800 g for 2 min, and the supernatant containing the IGCs was transferred to new tubes. Each pellet was resuspended with 20 µl of TM5 buffer and analyzed by transmission electron microscopy (Cs₂SO₄ pellet). To each supernatant, 200 µl of TM5 buffer was added and divided equally into six 3 ml polycarbonate tubes (Beckman) and centrifuged at 157 000 g in a Beckman Optima TLX ultracentrifuge (TLA rotor 100.3) for 1 h at 4°C. The supernatant was removed and

discarded. Each pellet was resuspended in 30 µl of TM5 buffer and transferred to a single microcentrifuge tube. One of the pellets was used for analysis by transmission electron microscopy (IGC pellet). All of the remaining pellets were sonicated for 1–3 min at a constant cycle using a water bath sonicator with a Branson sonifier model 450.

Immunopurification

Two-step purification of antibody 3C5 was done by ammonium sulfate precipitation and DE52 anion exchange chromatography (Harlow and Lane, 1988). The purified 3C5 antibody was coupled to Dynal M-450 rat anti-mouse IgM magnetic beads (Dynal, Lake Success, NY) in phosphate-buffered saline (PBS) (pH 7.4) for 18 h at room temperature, after which coupled beads were washed four times in PBS. In addition, two controls were performed as follows: 190 µg of purified IgMλ and IgMκ was coupled to the magnetic beads or PBS without the primary antibody. Then 250 µg of the IGC fraction was added to the coupled magnetic beads and the final volume was adjusted to 300 µl with PBS (containing 5 µg/ml leupeptin and aprotinin) followed by a 24 h incubation at 4°C on a slow rotator. The beads were washed five times in PBS with protease inhibitors (5 µg/ml leupeptin and aprotinin) and resuspended in 250 µl of PBS. The samples (70 µl) were dissolved in SDS sample buffer (Spector *et al.*, 1998) and the beads were removed by brief centrifugation prior to loading onto a 4–20% SDS gradient gel.

Transmission electron microscopy

All of the pellets from the IGC fractionation were microcentrifuged quickly and fixed with freshly prepared 2% formaldehyde and 0.5% glutaraldehyde in PBS (pH 7.4) (Spector *et al.*, 1998). LR-White sections (170–200 nm) were cut on a Reichert-Jung Ultracut E ultramicrotome with a Diatome diamond knife and stained using the EDTA-regressive method (Bernhard, 1969). Immunogold labeling was performed according to previously published procedures (Spector *et al.*, 1998). The samples were examined at 75 kV in a Hitachi H-7000 transmission electron microscope.

One- and two-dimensional gel electrophoresis and immunoblotting

Prior to measuring the protein concentration and loading the samples on gels, all pellets were sonicated for 10 s at a constant cycle using a water bath sonicator with a Branson sonifier model 450. The protein concentration was determined by the BCA Protein Assay (Pierce, Rockford, IL), and γ-globulin (Bio-Rad, Hercules, CA) was used as a standard. Proteins in the pellet and supernatant fractions from the IGC purification were separated on 4–20% SDS gradient gels (Jule Inc., New Haven, CT) and stained with Coomassie Blue R-250 or transferred to nitrocellulose as described (Misteli and Spector, 1996). Immunoreactive bands were detected by chemiluminescence using the Chemiluminescence Plus system (NEN, Boston, MA). Equilibrium two-dimensional gel electrophoresis was performed by the 2-D Gel Core Facility at Cold Spring Harbor Laboratory.

Mass spectrometry and peptide microsequencing

A 850 µg aliquot of IGC sample was thawed, denatured in 6 M urea/45 mM Tris–HCl, pH 7.0, reduced with 10 mM DTT for 30 min at 37°C, and then alkylated with 20 mM iodoacetamide for 30 min at room temperature. The sample was diluted to give a final concentration of 2 M urea, then 10 µg of sequencing grade trypsin (Boehringer Mannheim, IN) was added and the sample incubated overnight at 37°C. The peptide digest was acidified with 1% formic acid and subjected to solid-phase extraction using an Oasis cartridge (Waters Corp., MA). The cartridge was washed with 100% MeOH, equilibrated with 0.1% formic acid, and the peptide digest loaded. The cartridge was washed using 0.1% formic acid/7% MeOH, then peptides were eluted using 100% MeOH. The sample was dried, then resuspended in 0.1% formic acid. Aliquots corresponding to ~220 µg (original estimate) were separated by capillary

Fig. 8. Alignment of protein sequences related to the human hypothetical protein KIAA0017. The alignment was generated using PSI-BLAST (Altschul *et al.*, 1997) and conserved residues were highlighted using an automated procedure (Neuwald *et al.*, 1999); minor adjustments in the alignment were made to improve readability. Protein designations are color coded by family using the following scheme: KIAA0017, red; UV–DDB, blue; CPSF 160 kDa subunit, black. Hydrophobic conserved residues are highlighted in yellow; other conserved residues in red; poorly conserved or unconserved regions in gray. Gaps within the alignment are indicated as integer lengths. DDBJ/EMBL/GenBank identifiers (and organisms) corresponding to sequence designations are: IGC_human, 3540219; IGC_worm, 2804455 (*C.elegans*); yme9_yeast, 2497090 (*S.cerevisiae*); uv_arab, 2911067 (*A.thaliana*); uv_dicdi, 2130171 (*D.discoideum*); uv_human, 1136228; uv_pombe (*S.pombe*), 2330717; uv_worm, 3878704 cpsf_human, 1706102; cpsf_pombe, 3738146; cpsf_yeast, 2132494.

reverse-phase HPLC, using a YMC 0.5 mm×150 mm C18 column (Western Analytical, Temecula, CA) with an LCQ ion-trap mass spectrometer (Finnigan, San Jose, CA) connected on-line. The mass spectrometer was operated in the data-dependent manner (triple-play), whereby the most intense ion in each MS scan was selected for high-resolution zoom scan analysis to determine the charge state, followed by MS/MS of this selected ion. The dynamic exclusion option was utilized with a mass width of 5 Da for 1.5 min, to reduce the possibility of fragmenting the same ion repeatedly. The mode of operation, instrument settings and data handling are detailed elsewhere (Courchesne *et al.*, 1998). All MS/MS spectra were searched using an uninterpreted fragment ion search program against an in-house non-redundant database and a six-way translation of dbEST (Courchesne *et al.*, 1997).

For microsequencing gel bands, the IGC fraction was separated by 4–20% SDS gradient gels and stained with Coomassie G. The enriched 140 kDa gel band was excised and subjected to Lys-C endopeptidase digestion, and the resulting peptides were separated by HPLC and automated Edman degradation (Protein Chemistry Core Facility, Cold Spring Harbor Laboratory).

Sequence analysis

The BLAST-2 and PSI-BLAST programs (Altschul *et al.*, 1997) were used for searches of the NCBI non-redundant protein database and the TBLASTN program was used for searches of dbEST and DDBJ/EMBL/GenBank. The command line version of blastpgp with checkpointing (Altschul *et al.*, 1997) was used to construct PSI-BLAST profiles of specific protein families. PSI-BLAST profiles of one family were used to assess the statistical significance of sequence similarity to members of another family.

YFP fusion protein constructs

The ORFs of KIAA0017 and KIAA0324 in pBluescript SK⁺ were excised with *Sac*II and *Stu*I or with *Sa*I and *Sac*II, respectively, and inserted in-frame into pEYFP-C1 (Clontech Laboratories, Inc., Palo Alto, CA). Each fusion protein contained YFP at its C-terminus.

Cell culture and transient transfection assay

HeLa and BHK cells were grown on 100 mm culture dishes in Dulbecco's modified Eagle's medium (Gibco-BRL, Gaithersburg, MD) containing 10% fetal bovine serum, 50 U/ml penicillin and 50 U/ml streptomycin at 37°C in 10% CO₂. Confluent HeLa cells were collected and the KIAA0324-YFP construct was transfected by electroporation of 5 µg of DNA and 15 µg of salmon sperm DNA at 230 V using a Bio-Rad Gene Pulser II (Hercules, CA). Cells subsequently were seeded onto glass coverslips in 35 mm Petri dishes and grown for 14 h. For the KIAA0017-YFP construct, BHK cells were transfected and grown for 14 h.

Immunofluorescence and light microscopy

The transfected cells on coverslips were fixed for 15 min at room temperature in freshly prepared 2% formaldehyde in PBS (pH 7.4) according to previously published procedures (Spector *et al.*, 1998). Anti-SC35 monoclonal antibody was used at a dilution of 1:1000 (Fu and Maniatis, 1990) and anti-B' monoclonal antibody was used at a dilution of 1:70 (Habets *et al.*, 1989). Texas red-conjugated goat anti-mouse secondary antibody (Cappel Laboratories, Cochranville, PA) was used at a dilution of 1:100. Images were acquired using a Zeiss LSM 510 confocal laser scanning microscope with a Kr/Ar laser using simultaneous dual scans.

Acknowledgements

We are grateful to Tamara Howard for her excellent electron microscopy assistance. We thank Paula Bubulya, Adrian Krainer, Thoru Pederson and Noriko Saitoh for critical reading of the manuscript, Thoru Pederson and Mona Spector for critical suggestions, and Laura Mintz for editorial assistance. We thank the following individuals for providing antibodies: Gideon Dreyfuss (hnRNP C1/C2, hnRNP A1), Adrian Krainer (SF2/ASF), Tom Maniatis (U2AF65, U2AF35), Angela Kramer (SF3a66), Steve Warren (H5), Elmar Wahle (PAB II), Roel van Driel (PML) and Brian Turner (3C5). We thank Ryuji Kobayashi for peptide microsequencing and Aigoul Nourjanova for two-dimensional gel electrophoresis. We thank Jacques Camonis for providing the POP cDNA. We are grateful to Takahiro Nagase (Kazusa DNA Research Institute, Japan) for providing the KIAA0017 and KIAA0324 cDNAs. We thank Toshiro Tsukamoto for helpful advice in cloning the YFP fusion constructs. D.L.S. is funded

by a grant from the National Institute of General Medical Sciences (GM42694).

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Berezney, R. and Coffey, D.S. (1977) Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. *J. Cell Biol.*, **73**, 616–637.
- Bernhard, W. (1969) A new staining procedure for electron microscopical cytology. *J. Ultrastruct. Res.*, **27**, 250–265.
- Beyer, A.L., Christensen, M.E., Walker, B.W. and LeSturgeon, W.M. (1977) Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell*, **11**, 127–138.
- Blencowe, B.J., Issner, R., Nickerson, J.A. and Sharp, P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev.*, **12**, 996–1009.
- Bregman, D.B., Du, L., van der Zee, S. and Warren, S.L. (1995) Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J. Cell Biol.*, **129**, 287–298.
- Courchesne, P.L., Luethy, R. and Patterson, S.D. (1997) Comparison of in-gel and on-membrane digestion methods at low to sub pmol level for subsequent peptide and fragment-ion mass analysis using matrix-assisted laser-desorption/ionization mass spectrometry. *Electrophoresis*, **18**, 369–381.
- Courchesne, P.L., Jones, M.D., Robinson, J.H., Spahr, C.S., McCracken, S., Bentley, D.L., Luethy, R. and Patterson, S.D. (1998) Optimization of capillary chromatography ion-trap mass spectrometry for identification of gel-separated proteins. *Electrophoresis*, **19**, 956–967.
- Dahl, R., Wani, B. and Hayman, M.J. (1998) The Ski oncoprotein interacts with Skip, the human homolog of *Drosophila* Bx42. *Oncogene*, **16**, 1579–1586.
- Fakan, S. (1994) Perichromatin fibrils are *in situ* forms of nascent transcripts. *Trends Cell Biol.*, **4**, 86–90.
- Fakan, S. and Nobis, P. (1978) Ultrastructural localization of transcription sites and of RNA distribution during the cell cycle of synchronized CHO cells. *Exp. Cell Res.*, **113**, 327–337.
- Fakan, S. and Puvion, E. (1980) The ultrastructural visualization of nucleolar and extranucleolar RNA synthesis and distribution. *Int. Rev. Cytol.*, **65**, 255–299.
- Fay, F.S., Taneja, K.L., Shenoy, S., Lifshitz, L. and Singer, R.H. (1997) Quantitative digital analysis of diffuse and concentrated nuclear distributions of nascent transcripts, SC35 and poly(A). *Exp. Cell Res.*, **231**, 27–37.
- Fu, X.-D. and Maniatis, T. (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature*, **343**, 437–441.
- Habets, W.J., Hoet, M.H., De Jong, B.A.W., Van der Kemp, A. and van Venrooij, W.J. (1989) Mapping of B cell epitopes on small nuclear ribonucleoproteins that react with human autoantibodies as well as with experimentally induced mouse monoclonal antibodies. *J. Immunol.*, **143**, 2560–2566.
- Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Hirschfeld, S., Levine, A.S., Ozato, K. and Protic, M. (1990) A constitutive damage-specific DNA-binding protein is synthesized at higher levels in UV-irradiated primate cells. *Mol. Cell Biol.*, **10**, 2041–2048.
- Huang, S. and Spector, D.L. (1996) Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J. Cell Biol.*, **133**, 719–732.
- Huang, S. and Spector, D.L. (1997) Pre-mRNA splicing: nuclear organization of factors and substrates. In Krainer, A.R. (ed.), *Eucaryotic Messenger RNA Processing: Frontiers in Molecular Biology*. pp. 37–67. IRL Press, Oxford, UK.
- Jiménez-García, L.F. and Spector, D.L. (1993) *In vivo* evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell*, **73**, 47–59.
- Keeney, S., Eker, A.P., Brody, T., Vermeulen, W., Bootsma, D., Hoeijmakers, J.H. and Linn, S. (1994) Correction of the DNA repair defect in xeroderma pigmentosum group E by injection of a DNA damage-binding protein. *Proc. Natl Acad. Sci. USA*, **91**, 4053–4056.
- Kvalheim, G., Fodstad, O., Pihl, A., Nustad, K., Pharo, A., Ugelstad, J. and Funderud, S. (1987) Elimination of B-lymphoma cells from human bone marrow: model experiments using monodisperse magnetic

- particles coated with primary monoclonal antibodies. *Cancer Res.*, **47**, 846–851.
- Lamond, A. and Mann, M. (1997) Cell biology and the genome projects—a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol.*, **7**, 139–142.
- Lamond, A.I. and Earnshaw, W.C. (1998) Structure and function in the nucleus. *Science*, **280**, 547–553.
- Misteli, T. and Spector, D.L. (1996) Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol. Biol. Cell*, **7**, 1559–1572.
- Misteli, T. and Spector, D.L. (1998) The cellular organization of gene expression. *Curr. Opin. Cell Biol.*, **10**, 323–331.
- Misteli, T. and Spector, D.L. (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell*, **3**, 697–705.
- Misteli, T., Cáceres, J.F. and Spector, D.L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, **387**, 523–527.
- Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F. and Spector, D.L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription *in vivo*. *J. Cell Biol.*, **143**, 297–307.
- Monneron, A. and Bernhard, W. (1969) Fine structural organization of the interphase nucleus in some mammalian cells. *J. Ultrastruct. Res.*, **27**, 266–288.
- Nakielnny, S. and Dreyfuss, G. (1997) Nuclear export of proteins and RNAs. *Curr. Opin. Cell Biol.*, **9**, 420–429.
- Neubauer, G., Gottschalk, A., Fabrizio, P., Séraphin, B., Lührmann, R. and Mann, M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA*, **94**, 385–390.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.*, **20**, 46–50.
- Neugebauer, K.M. and Roth, M.B. (1997) Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.*, **11**, 1148–1159.
- Neuwald, A.F., Aravind, L., Spouge, J.L. and Koonin, E.V. (1999) AAA⁺: a class of chaperone-like ATPases associated with the assembly, operation and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
- Novikoff, A.B. and Holtzman, E. (1970) Cell organelles. In Ebert, J.D., Loewy, A.G., Miller, R.S. and Schneiderman, H.A. (eds), *Cells and Organelles*. Holt, Rinehart and Winston, Inc., New York, NY, pp. 39–160.
- Ochs, R., Lischwe, M., O'Leary, P. and Busch, H. (1983) Localization of nucleolar phosphoproteins B23 and C23 during mitosis. *Exp. Cell Res.*, **146**, 139–149.
- Otrin, V.R., Mclenigan, M., Takao, M., Levine, A.S. and Protic, M. (1997) Translocation of a UV-damaged DNA binding protein into a tight association with chromatin after treatment of mammalian cells with UV light. *J. Cell Sci.*, **110**, 1159–1168.
- Patterson, S.D. (1998) Protein identification and characterization by mass spectrometry. In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds), *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., New York, pp. 10.22.1–10.22.24.
- Petrov, P. and Sekeris, C.E. (1971) Early action of α -amanitin on extranucleolar ribonucleoproteins, as revealed by electron microscopic observation. *Exp. Cell Res.*, **69**, 393–401.
- Piñol-Roma, S., Choi, Y.D., Matunis, M.J. and Dreyfuss, G. (1988) Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.*, **2**, 215–227.
- Saucan, L. and Palade, G.E. (1994) Membrane and secretory proteins are transported from the Golgi complex to the sinusoidal plasmalemma of hepatocytes by distinct vesicular carriers. *J. Cell Biol.*, **125**, 733–741.
- Singer, R.H. and Green, M.R. (1997) Compartmentalization of eucaryotic gene expression: causes and effects. *Cell*, **91**, 291–294.
- Spector, D.L. (1990) Higher order nuclear organization: three-dimensional distribution of small nuclear ribonucleoprotein particles. *Proc. Natl Acad. Sci. USA*, **87**, 147–151.
- Spector, D.L. (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.*, **9**, 265–315.
- Spector, D.L., Schrier, W.H. and Busch, H. (1983) Immunoelectron microscopic localization of snRNPs. *Biol. Cell*, **49**, 1–10.
- Spector, D.L., Goldman, R.D. and Leinwand, L.A. (1998) *Cells: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Swift, H. (1959) Studies on nuclear fine structure. *Brookhaven Symp. Biol.*, **12**, 134–152.
- Thiry, M. (1995) Behavior of interchromatin granules during the cell cycle. *Eur. J. Cell Biol.*, **68**, 14–24.
- Treiber, D.K., Chen, Z. and Essigmann, J.M. (1992) An UV light-damaged DNA recognition protein absent in xeroderma pigmentosum group E cells binds selectively to pyrimidine (6–4) pyrimidone photoproducts. *Nucleic Acids Res.*, **20**, 5805–5810.
- Turner, B.M. and Franchi, L. (1987) Identification of protein antigens associated with the nuclear matrix and with clusters of interchromatin granules in both interphase and mitotic cells. *J. Cell. Sci.*, **87**, 269–282.
- Watanabe, T., Sukegawa, J., Sukegawa, I., Tomita, S., Iijima, K., Oguchi, S., Suzuki, T., Nairn, A.C. and Greengard, P. (1999) A 127-KDa protein (UV-DDB) binds to the cytoplasmic domain of the Alzheimer's amyloid precursor protein. *J. Neurochem.*, **72**, 549–556.
- Wigge, P.A., Jensen, O.N., Holmes, S., Soues, S., Mann, M. and Kilmartin, J.V. (1998) Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.*, **141**, 967–977.
- Yang, L., Embree, L.J., Tsai, S. and Hickstein, D.D. (1998) Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *J. Biol. Chem.*, **273**, 27761–27764.
- Zborek, A., Krzyzowska-Gruca, S. and Gruca, S. (1990) Nuclear fraction enriched in interchromatin granules. *Cell Biol. Int. Rep.*, **14**, 79–88.
- Zhang, G., Taneja, K.L., Singer, R.H. and Green, M.R. (1994) Localization of pre-mRNA splicing in mammalian nuclei. *Nature*, **372**, 809–812.

Received April 8, 1999; revised and accepted June 15, 1999

Note added in proof

Recently, Casparly *et al.* [*EMBO J.*, **18**, 3463–3474 (1999)] identified the yeast protein Rse1p as an essential pre-mRNA splicing factor required for pre-spliceosome assembly. This protein is the yeast homolog of KIAA0017.