An Expression Profile of Genes in Human Retina and Isolation of a Complementary DNA for a Novel Rod Photoreceptor Protein

Akiyo Shimizu-Matsumoto,† Wakako Adachi,*† Katsuya Mizuno,* Johji Inazawa,‡ Kohji Nishida,† Shigeru Kinoshita,† Kenichi Matsubara,* and Kousaku Okubo**

Purpose. To characterize expression patterns of active genes in human retina, and to isolate novel genes that are uniquely expressed in this tissue.

Methods. A 3′-directed complementary DNA (cDNA) library that faithfully represents the composition of messenger RNA (mRNA) was constructed with an mRNA preparation from a cadaveric human retina. A total of 925 3′ terminal sequences were collected by sequencing randomly selected clones, of which 789 were regarded as representing chromosomally coded genes (gene signatures [GS]). GS were compared with each other and searched against GenBank. The resulting expression profile, listing gene species and their frequency, represents the composition of mRNA in the retina. By comparing this expression profile with those obtained from 10 other source cells or tissues, genes uniquely active in the retina were discovered, including some not previously described. A full-sized cDNA corresponding to one of these was isolated and sequenced. Its expression was analyzed by multitissue Northern hybridization and in situ hybridization to the retina specimen. It was then mapped on human chromosomes.

Results. In the expression profile, 108 genes were detected recurrently, suggesting that they are very active. Fifty-five of them were identified in GenBank, including the most abundant opsin gene and several other genes for phototransduction. Among the remaining novel and active genes, 19 were considered unique to retina on the basis of their representation status in other expression profiles and in dbEST. One of these was identified as a gene that encodes a novel secretory protein expressed in a rod photoreceptor that maps to chromosome 18p11.3.

Conclusions. The expression profile of active genes in the retina represents the composition of mRNA, which reflects the relative activities of genes in this tissue. A comparison of this expression profile with those obtained with other tissues resulted in isolation of a novel cDNA specifically expressed in the rod photoreceptor. It is anticipated that additional novel genes that are uniquely active in the neural retina may be obtained with the same strategy, leading to further clarification of the biologic or physiological characteristics of this tissue. Invest Ophthalmol Vis Sci. 1997;38: 2576–2585.

The neural retina contains highly specialized photoreceptor cells that perform phototransduction uniquely in the human body. Recent attempts at isolating the genes expressed only in this tissue have made a significant contribution to clarifying its specialized functions. Examples are arrestin,[1] recoverin,[2] RDS/peripherin,[3] ROM1,[4] interphotoreceptor retinoid-binding protein,[5] and cellular retinaldehyde-binding protein.[6] Some of these genes have been identified by the so-called candidate gene approach as causative genes for inherited diseases of the retina.[7–9] Accordingly, systematic isolation of genes uniquely active in this tissue would help clarify its unique function and may lead to understanding of the pathogenesis of retinal disease.

We have developed a method called expression profiling of active genes in which the messenger RNA (mRNA) population in a given tissue is de-

*From the *Institute for Molecular and Cellular Biology, Osaka University, Japan, and the Departments of †Ophthalmology and ‡Hygiene, Kyoto Prefectural University of Medicine, Japan.*
Supported in part by a grant-in-aid for the Human Genome Program, research grants 06454500 and 06771529 from the Japanese Ministry of Education, Science and Culture of Japan, and a research fund from the Kyoto Foundation for the Promotion of Medical Science.
Submitted for publication October 2, 1996; revised April 1, 1997, and July 14, 1997; accepted July 14, 1997.
Proprietary interest category: N.
Reprint requests: Kousaku Okubo, Institute for Molecular and Cellular Biology, Osaka University, 1-3, Yamada-oka, Suita, Osaka 565, Japan. E-mail: kousaku@imcb.osaka-u.ac.jp

Investigative Ophthalmology & Visual Science, November 1997, Vol. 38, No. 12

scribed quantitatively by sequencing randomly selected clones from a 3′-directed complementary DNA (cDNA) library that faithfully represents the mRNA composition.[10,11] The proportional relation between recurrence of the same sequence in the library and the abundance of the corresponding transcript has been tested on several occasions, and the implications of the profiles have been discussed.[11,12] Expression profiling may be categorized in random cDNA sequencing, or expressed sequence tag (EST) collection,[13] which has been successfully and widely employed in finding new transcripts from various tissues,[14–16] including fovea.[17] However, mRNA composition cannot be represented in these EST collections because cDNA libraries have not been made to represent mRNA composition, or sequencing has been directed to 5′-ends that represent arbitrary positions of each transcript. Knowing the mRNA composition of various tissues enables us to isolate genes uniquely active in a given tissue by comparing them with each other.[18–20] This article describes the expression profile of active genes in the retina and the isolation of a novel retina-specific cDNA (rod photoreceptor protein [RPP]) by comparing the profiles.

MATERIALS AND METHODS

All human studies were performed in accordance with the guidelines of the Declaration of Helsinki and were approved by the institutional review board. Informed consent was obtained after the nature of the study was explained.

Preparation of Retinal RNA, Construction of 3′-Directed cDNA Library, and Sequencing

A human retinal tissue (neural retina) was obtained at autopsy from a 50-year-old male 3 hours after death. Poly(A) RNA was isolated with the QuickPrep mRNA purification kit (Pharmacia, Upsala, Sweden). cDNA was synthesized by using pUC19-based vector primer, trimmed by digestion with MboI and circularized and transformed into *E. coli* DH5 as described previously.[10,11] The transformant colonies were randomly selected and lysed, and the cDNA insert, connecting the MboI site (GATC) to poly(A), was amplified by polymerase chain reaction. The products were subjected to cycle sequencing reactions using dye-labeled primers according to the manufacturer's protocol (Perkin-Elmer Applied Biosystems, Foster City, CA) and analyzed by autosequencer 373A (Perkin-Elmer Applied Biosystems). From the resulting sequences, we eliminated those shorter than 20 base pairs (bp) or those with more than 5% ambiguous nucleotides. Mitochondrial gene sequences, ribosomal RNA sequences, and those containing only repetitive sequences such as Alu were also eliminated from the analysis. Sequences thus selected were regarded as representing chromosomal gene transcripts and were termed gene signatures (GS). They were then compared with each other to obtain the frequency of appearance and then compared with GS previously collected from other sources. A GS number of five digits (e.g., GS12345) was assigned to new GS sequences. In this study, 365 new GS were identified and submitted to DDBJ under accession numbers c20544 through c20908. All the GS were then compared with sequences in the primate division of GenBank (Re94) for gene identification using the FastA program.[21] The BLAST program[22] was used to compare the recurrently appearing sequences with the dbEST division of GenBank (Re94). With every comparison, a similarity spanning more than 90% of query sequence and having more than 95% identity was regarded as identical.

Northern Blot Hybridization

0.5 μg of Poly(A) RNA from human retina and 10 other tissues (Clontech Laboratories, Palo Alto, CA) were electrophoresed in formalin agarose gel and transferred to a Hybond N+ membrane (Amersham International, Buckinghamshire, United Kingdom) according to the manufacturer's protocol. Hybridization was performed for 18 hours at 42°C in a buffer containing 6 × SSPE, 1% sodium dodecyl sulfate (SDS), 1 × Denhardt's solution, 50% formamide, 100 μg/ml of denatured salmon sperm DNA, and ^{32}P-labeled GS4642 fragment. The filter was then washed in a solution containing 2 × SSC and 0.1% SDS for 10 minutes at 65°C and analyzed with a Fuji BAS-2000 imaging analyzer (Fuji Medical Systems, New Haven, CT).

Full-Length cDNA Library Screening and Sequencing

A full-length cDNA library was constructed from 5 μg of retina poly(A) RNA (Clontech) using the ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA). Plaque hybridization was performed by using 3′-directed cDNA clones as probes. Duplicated filter (Dupont-New England Nuclear, Boston, MA) lifts were hybridized at 65°C for 20 hours in solution containing 6 × SSC, 1% SDS, and 1 × Denhardt's solution. The filters were then washed in 1 × SSC and 0.1% SDS at 65°C for 15 minutes and exposed to Kodak (New York, NY) XAR-5 film at −70°C overnight with an intensifying screen. For full-length sequencing, the shotgun method[23] was adapted: the cDNA inserts from the Lambda ZAP clones were amplified by polymerase chain reaction, ligated, sonicated, and subcloned into the pUC19 vector. Inserted fragments were sequenced with Taq Dye Primer Cycle Sequencing Kit (Perkin-

GS#	GeneSignatures: retina	HL60	HL/DMSO	HL/TPA	granulocyte	HepG2	lung	colon	liver	keratinocyte	cornea	GenBank (Re94): Acc#	gene_name	dbEST: retina	brains	others
4783	12	-	-	-	-	-	-	-	-	-	-	K02281	opsin	4	-	-
4895	11	-	-	-	-	-	-	-	-	-	-	D87330	*analyzed in this study (1)*	-	1	1
418	9	22	1	10	1	5	-	-	-	-	7	D14531	ribosomal protein L9	-	1	55
4754	7	-	-	-	-	-	-	-	-	-	-	U40998	retinal gene4	-	1	3
4888	6	-	-	-	-	-	-	-	-	-	-	X05196	aldolase C	-	9	4*
273	5	6	4	11	3	9	1	5	2	3	3	X16064	translationally controlled tumor prot	1	8	84
1608	5	-	2	-	1	-	-	-	-	1	-	S60099	amyloid precursor protein homolog	-	1	20
290	4	9	3	8	-	5	1	1	-	5	2	X66699	ribosomal protein L37a	-	27	19
285	4	14	1	8	-	6	1	5	3	1	2	X89401	ribosomal protein L21	-	1	9
583	4	2	5	3	-	-	3	2	1	1	-	X73460	ribosomal protein L3	-	12	60
4763	4	-	-	-	-	-	-	-	-	-	-	K02281	opsin	2	-	-
4626	4	-	-	-	-	-	-	-	-	-	-	S62027	transducin-γ subunit	-	-	-
543	3	9	4	5	1	-	-	2	-	2	2	X56932	23 kD highly basic protein	-	1	18
211	3	24	2	10	-	8	5	3	2	1	4	X67247	ribosomal protein S8	-	-	11
19	3	8	2	23	-	17	1	4	1	5	5	X16869	elongation factor 1-alpha	-	48	53
917	3	3	1	2	-	-	-	4	1	3	1	D19955	-	-	-	9
795	3	1	1	2	1	-	-	-	2	-	6	D19839	-	-	1	13
743	3	6	1	5	-	-	-	-	1	2	3	U14973	ribosomal protein S29	-	35	25
675	3	4	1	4	-	-	2	-	-	-	1	L01124	ribosomal protein S13	1	5	32
380	3	12	-	1	-	1	-	-	3	1	2	M13934	ribosomal protein S14	-	4	20
861	3	6	-	1	-	-	-	-	-	-	3	M31520	ribosomal protein S24	-	1	10
50	3	-	2	-	-	1	1	-	-	1	-	L28809	DNA-binding protein B (dbpB)	-	10	9
270	3	-	1	-	-	1	-	1	-	-	-	D11980	-	-	5	3
4711	3	-	-	-	1	-	-	-	-	-	-	C10293	-	-	7	5
4008	3	-	-	-	-	-	-	-	-	-	2	U20998	signal recognition particle 9	-	10	-
1402	3	-	1	-	-	-	-	-	-	-	-	U28387	hexokinase II	-	-	1
172	3	-	-	-	-	1	-	1	-	-	-	M12623	HMG-17	-	2	11
813	3	1	-	-	-	-	-	-	-	-	-	D19857	-	-	2	6
2155	3	-	-	-	-	-	-	-	10	-	-	M12530	transferrin	-	-	53*
4772	3	-	-	-	-	-	-	-	-	-	-	C20742	-	-	-	4
4786	3	-	-	-	-	-	-	-	-	-	-	C20755	-	-	8	-
4642	3	-	-	-	-	-	-	-	-	-	-	D63813	*analyzed in this study(2)*	1	1	-
4629	3	-	-	-	-	-	-	-	-	-	-	D45399	cone cGMP PDE-γ	-	-	-
4597	3	-	-	-	-	-	-	-	-	-	-	C20603	-	-	-	-
4464	3	-	-	-	-	-	-	-	-	-	-	C20601	-	-	-	-
650	2	14	2	5	1	-	1	6	-	1	2	L06432	ribosomal protein S18	-	8	27
689	2	1	1	2	-	-	3	2	3	1	1	D19733	-	-	4	2
293	2	13	1	6	-	6	3	1	2	2	5	U14968	ribosomal protein L27a	-	1	22
1657	2	-	1	1	-	-	4	4	2	1	2	L11566	ribosomal protein L18	-	3	22
1321	2	-	3	2	2	-	1	1	-	-	2	D20347	-	-	3	7
704	2	4	1	-	-	-	1	2	-	2	3	M73791	anonymous mRNA	-	19	31
504	2	5	1	-	-	1	1	1	1	2	12	X16295	non-neuronal enolase	-	7	5
314	2	3	1	3	2	1	5	-	1	-	3	X06617	ribosomal protein S11	-	3	11
1574	2	-	2	-	1	-	1	-	-	1	1	D20599	-	-	6	1
1244	2	-	3	1	-	-	-	2	-	1	1	M22348	ubiquinone-binding protein	1	-	9
818	2	4	1	1	1	-	-	2	1	-	-	U14966	ribosomal protein L5	-	6	22
707	2	4	-	1	-	-	-	1	1	1	3	J02984	insulinoma rig-analog	-	3	9
759	2	4	1	1	-	-	-	-	1	1	4	M13932	ribosomal protein S17	-	2	9
437	2	4	2	2	-	5	1	-	2	-	5	X57959	ribosomal protein L7	-	5	17
644	2	5	-	2	-	-	-	-	1	1	2	U12465	ribosomal protein L35	-	3	13
463	2	2	-	1	-	2	-	-	-	2	3	X02152	lactate dehydrogenase-A	-	3	5
934	2	3	1	1	1	-	-	-	3	-	-	X13923	cox VIb	-	2	6
1674	2	-	1	8	-	-	-	-	1	-	1	L43509	methionine adenosyltransferase	-	4	3
1671	2	-	2	-	-	-	4	-	-	-	1	D20695	-	-	-	3

GS#	GeneSignatures: retina	HL60	HL/DMSO	HL/TPA	granulocyte	HepG2	lung	colon	liver	keratinocyte	cornea	GenBank (Re94): Acc#	gene_name	dbEST: retina	brains	others
827	2	1	1	-	1	-	-	-	-	-	-	M89913	dUTP pyrophosphatase	-	1	6
679	2	1	1	-	-	-	-	-	-	1	-	J04173	phosphoglycerate mutase	-	14	11
422	2	3	-	-	-	1	-	1	-	-	1	M10036	triosephosphate isomerase	-	6	12
351	2	-	-	-	-	1	1	1	-	2	-	D11887	-	-	8	2
683	2	1	2	-	-	-	-	1	-	-	-	X13238	cytochrome c oxidase subunit VIc	-	5	12
2847	2	-	-	-	-	-	1	-	-	1	-	M86842	cAMP response element regulatory prot.	-	6	8
2813	2	-	-	-	1	-	1	-	-	-	-	D45619	-	-	1	4
2309	2	-	-	-	-	-	-	-	-	2	1	X15183	90-kDa heat-shock protein	-	11	12
2073	2	-	-	-	-	-	1	3	-	-	-	D25550	-	-	2	2
1966	2	-	-	1	-	-	-	1	-	-	-	D20984	-	-	1	2
481	2	1	-	-	-	1	-	-	2	-	1	D12109	-	-	1	14
4914	2	-	-	-	-	-	-	-	-	-	1	C00400	-	-	1	1
4721	2	-	-	-	1	-	-	-	-	-	-	C01294	-	-	2	5
2812	2	-	-	-	-	-	1	-	-	-	-	C20779	-	1	18	12
2316	2	-	-	-	-	-	-	-	-	-	1	M38188	unknown protein from clone pHGR74	-	3	8
2186	2	-	-	-	-	-	-	-	-	-	35	M74816	Apo J / SP_40,40 /clusterin	1	21	47
2034	2	-	-	2	-	-	-	-	-	-	-	X64364	M6 antigen	-	4	2
714	2	3	-	-	-	-	-	-	-	-	-	X17621	Nm23 protein	-	10	4
217	2	-	-	-	-	1	1	-	1	-	-	D11841	-	-	9	6
4665	2	-	-	-	-	-	-	-	-	-	1	U20352	malate dehydrogenase (MDHA)	-	10	21
1296	2	-	1	-	-	-	-	-	-	-	-	D20322	-	-	9	10
959	2	1	-	-	-	-	-	-	-	-	-	U28946	G/T mismatch binding protein (GTBP)	-	-	3
2613	2	-	-	-	-	-	-	-	1	-	-	C20554	-	-	7	-
4773	2	-	-	-	-	-	-	-	-	-	-	C20743	-	-	-	6
4603	2	-	-	-	-	-	-	-	-	-	-	C20609	-	-	-	4*
4639	2	-	-	-	-	-	-	-	-	-	-	C20637	-	-	-	4
4605	2	-	-	-	-	-	-	-	-	-	-	C20611	-	-	-	2*
4735	2	-	-	-	-	-	-	-	-	-	-	J02966	mitochondrial ADP/ADT translocator	-	6	2
4613	2	-	-	-	-	-	-	-	-	-	-	C20617	-	-	1	2
4655	2	-	-	-	-	-	-	-	-	-	-	M22349	neuron-specific γ-2 enolase	2	15	1
4889	2	-	-	-	-	-	-	-	-	-	-	M54968	K-ras oncogene	-	7	1
4905	2	-	-	-	-	-	-	-	-	-	-	C20849	-	-	6	1
4819	2	-	-	-	-	-	-	-	-	-	-	C20784	-	-	2	1
4973	2	-	-	-	-	-	-	-	-	-	-	C20899	-	-	1	1
4767	2	-	-	-	-	-	-	-	-	-	-	C20737	-	-	2	-
3901	2	-	-	-	-	-	-	-	-	-	-	C01907	-	-	3	-
4829	2	-	-	-	-	-	-	-	-	-	-	C20791	-	-	8	-
4792	2	-	-	-	-	-	-	-	-	-	-	C20761	-	-	1	-
3869	2	-	-	-	-	-	-	-	-	-	-	C01001	-	-	-	1
4659	2	-	-	-	-	-	-	-	-	-	-	C20654	-	-	-	1
4791	2	-	-	-	-	-	-	-	-	-	-	C20760	-	-	-	1
4859	2	-	-	-	-	-	-	-	-	-	-	C20812	-	-	-	1
4894	2	-	-	-	-	-	-	-	-	-	-	C20843	-	-	-	1
4898	2	-	-	-	-	-	-	-	-	-	-	C20846	-	-	-	1
4654	2	-	-	-	-	-	-	-	-	-	-	C20651	-	-	-	-
4664	2	-	-	-	-	-	-	-	-	-	-	C20659	-	-	-	-
4679	2	-	-	-	-	-	-	-	-	-	-	L36861	guanylate cyclase activating protein	1	-	-
4802	2	-	-	-	-	-	-	-	-	-	-	C20770	-	-	-	-
4805	2	-	-	-	-	-	-	-	-	-	-	C20773	-	-	-	-
4836	2	-	-	-	-	-	-	-	-	-	-	U20325	cocaine amphetamine regulated transcript	-	-	-
4873	2	-	-	-	-	-	-	-	-	-	-	C20824	-	-	-	-
4874	2	-	-	-	-	-	-	-	-	-	-	C20825	-	-	-	-
4875	2	-	-	-	-	-	-	-	-	-	-	C20826	-	-	-	-
4915	2	-	-	-	-	-	-	-	-	-	-	C20855	-	-	-	-

Recurrently appearing Gene Gignatures(GS) are listed in descending order of appearance. For each GS, its occurence in the expression profiles of promyelocytic cell line HL60, HL60 treated with DMSO (HL/DMSO), and with TPA (HL/TPA), peripheral granulocyte, hepatoblastoma cell line HepG2, lung, colon mucosa (colon), adult liver(liver), primary cultured keratinocyte (keratinocyte) and corneal epithelium(cornea) is shown. The total number of GS in each profile is: HL60:861; HL/DMSO:1064; HL/TPA:894; granulocyte:1182; HepG2:746; lung:876; colon:931; liver:959 keratinocyte:851; cornea:1159. Shaded rows represent GS commonly found in some of these libraries. For GS having an identical sequence in human entries in the GenBank (Re94), the gene names are given with their accession numbers. For those matched to 3'-ESTs in the dbEST division of GenBank, the recurrence is also given with respect to the libraries from which they originated. The number of 3 ESTs in dbEST (Re94) is 2,960 from the retina and 155,995 from the brain and other sources. Asterisks(*) denote multiple ESTs found in a single source tissue. Rows boxed by bold lines represent GS regarded as unique to the retina, on the basis of representation patterns in dbEST.

FIGURE 1. An expression profile of the neural retina.

TABLE 1. Representation of Genes Unique to the Retina

Accession Number	Genes	GS		3'-EST	
		Retina	Others	Retina*	Others
K02281	Opsin	16	—	6	—
K03494	Green cone photoreceptor pigment	—	—	—	—
M13305	Red cone photoreceptor pigment	—	—	—	—
M13299	Blue cone photoreceptor pigment	—	—	—	—
J03260	Transducin α	1	—	—	—
X04526	Transducin β	1	—	—	4
S62027	Transducin γ	4	—	1	—
D10384	Cone transducin α	—	—	—	—
M26061	Rod cGMP phosphodiesterase α†	—	—	1	—
S41458	Rod cGMP phosphodiesterase β†	—	—	kl	—
M36476	Rod cGMP phosphodiesterase γ†	—	—	18	1
U20212	Cone cGMP phosphodiesterase α	—	—	—	—
D45399	Cone cGMP phosphodiesterase γ	3	—	—	—
S42457	Rod photoreceptor cGMP-gated channel	1	—	—	1
U18945	cGMP gated channel γ	—	—	—	—
M92432	Retinal guanylyl cyclase (retGC)	—	—	—	—
L37378	Retinal guanylyl cyclase (retGC-2)	—	—	1	—
L36861	Guanylate cyclase activating protein (GCAP)	2	—	2	—
M33478	33-kDa phototransducing protein/ phosducin	—	—	—	—
S43855	Recoverin/Ca-sensitive regulator	—	—	18	—
X12453	Retinal S-antigen	1	—	4	—
U03626	C-arrestin	—	—	1	—
M73531	Retinal degeneration slow (PERIPHERIN)	1	—	—	1
L07894	Rod outer segment membrane protein 1 (ROM1)	—	—	4	2
M22453	Interphotoreceptor retinoid-binding protein (IRBP)	—	—	2	—
J04213	Cellular retinaldehyde-binding protein	—	—	—	—
S74445	Cellular retinoic acid-binding protein	1	—	2	1
M81840	NRL gene product (retina specific leucine zipper)	—	—	—	—
U40998	Retinal protein (HRG4) novel photoreceptor protein	7	—	1	5
U14910	RPE-retinal G protein-coupled receptor	—	—	—	—
	Total number of clones in analysis	789	9523	2960	155,995

Human genes whose expression have been proven to be limited or highly preferential to retina were collected from GenBank (Re94). From the entries retrieved by word search with key words human AND (retina or phototransduction or opsin) NOT EST, all those with published proof in expression patterns are listed. Using their mRNA or gene sequences as probes, dbEST and GS database were searched and the recurrence of corresponding entries is shown. For ESTs only the 3'-EST were counted to prevent multiple counting of the same clone.

* ESTs from retina are all from normalized libraries where recurrence of the same clone does not correlate with the abundance of mRNA.

† cDNA for the γ-subunit has no GATC sequence that is essential to be cloned in the 3'-directed library. Mesasges for α and β subunits terminate within repetitive sequences, and were eliminated from analysis at the qualification step in the data analysis. Such cases are not common but unfortunately occurred in genes for 3 subunits of the same enzyme.

Elmer Applied Biosystems, Foster City, CA) and assembled using the DNASIS-MAC volume 3.2 program (Hitachi Software, Yokohama, Japan).

Tissue Preparation and RNA Probe Synthesis

An eye from a 29-year-old man was obtained at autopsy 4 hours after death, was immediately fixed in 4% paraformaldehyde in 0.1 M phosphate buffer for 16 hours, and was dehydrated with graded ethanol sets and finally with chloroform. After embedding in paraffin, the eye was cut into 8-μm sections that were floated on diethyl pyrocarbonate-treated water, placed on slides coated with 3-aminopropyltriethoxysilane (Aldrich Chemical, Milwaukee, WI), and dried on a 42°C warming plate.

The full-length clone for GS4642 (RPP) was subcloned into the XhoI/SmaI site of pSP72 vector and pBluescript SK−vector. Both plasmids were linearized by XhoI and SmaI, and single-strand sense RNA probe and antisense RNA probe were prepared by T7 RNA polymerase from pSP72 and pBS, respectively, using DIG RNA Labeling Kit (Boehringer-Mannheim, Mannheim, Germany) according to the manufacturer's instructions.

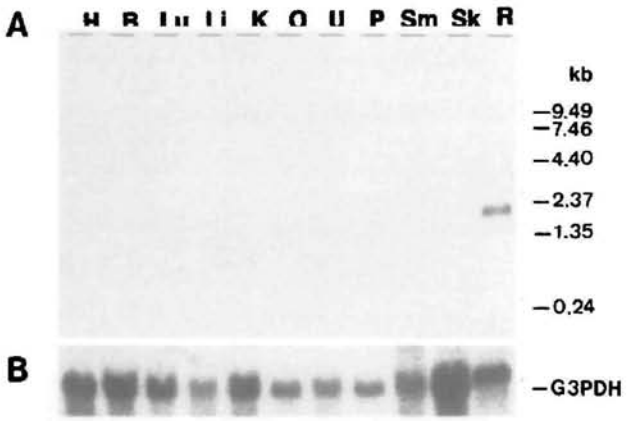

FIGURE 2. Expression pattern of rod photoreceptor protein gene (GS4642). (**A**) 0.5 μg each of poly(A) RNAs from retina (R) and 10 other tissues (H: heart; B: brain; Lu: lung; Li: liver; K: kidney; O: ovary; U: uterus; P: placenta; Sm: small intestine; Sk: skeletal muscle) were hybridized with ^{32}P-labeled GS4642 fragment. (**B**) The same filter was rehybridized with glyceraldehyde 3-phosphate dehydrogenase (G3PDH) cDNA probe.

In Situ Hybridization

Paraffin was removed from sections with three 10-minute changes of xylene and hydrated through graded ethanol. Slides were dipped in 0.1 M triethanolamine containing 0.25% acetic anhydride. The hybridization solution, in a final volume of 200 μl per section, contained 50% formamide, 0.3 M NaCl, 10% SDS, 20 mM Tris-HCl (pH 8.0), 0.2% sarcosyl, 0.02% salmon sperm DNA, 1 × Denhardt's solution, and 50 ng digoxigenin-labeled probe. The slides were placed in a humidified chamber and incubated at 50°C for 20 hours. They were first washed in 5 × SSC at 50°C, then in a buffer containing 50% formamide and 2 × SSC at 50°C for 30 minutes. The sections were then incubated in 10 μg/ml RNase A (Sigma Chemical Co., St. Louis, MO) in 10 mM Tris-HCl (pH 7.5), 1 mM ethylenediaminetetraacetic acid, and 0.5 M NaCl at 37°C for 30 minutes, and again washed at high stringency. RNA hybrids were detected by incubation for 1 hour with an alkaline-phosphatase–conjugated antidigoxigenin antibody (Boehringer-Mannheim) diluted 1:500 in 100 nM Tris-HCl (pH 7.5) and 150 mM NaCl, and then for 16 hours with the chromogenic substrates 4-nitroblue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (both Boehringer-Mannheim).

Fluorescence In Situ Hybridization

Fluorescence in situ hybridization was carried out as previously reported.[24] In brief, metaphase chromosomes were prepared with the thymidine synchronization, bromodeoxyuridine release technique for delineation of replication G-bands. Before hybridization, chromosomes were stained in Hoechst 33258 and irradiated with ultraviolet light. A 1737-bp RPP cDNA was labeled with biotin-16-uridine triphosphate (Boehringer-Mannheim) by nick-translation and hybridized to the denatured chromosomes at a final concentration of 25 ng/ml in a mixture of 50% formamide, 10% dextran sulfate (Sigma), 2 × SSC, sonicated salmon sperm DNA (2 μg/μl), and *E. coli* tRNA (2 μg/μl). The hybridization signals were detected with fluorescein isothiocyanate–avidin (Boehringer-Mannheim), and chromosomes were counterstained with propidium iodide (1 μg/ml). The precise signal position was determined by the delineation of G-banding patterns.

RESULTS

Expression Profile of Active Genes in the Retina

A 3′-directed library, harboring cDNA fragments connecting poly(A) to the nearest MboI site (GATC), was made from the mRNA of a neural retina of a male cadaver. Inserted sequences of 925 randomly selected clones were determined by single-run sequences, and 789 of them were regarded as representing chromosomally coded genes (see above for details of qualification). These 789 sequences were composed of 607 different sequences. These sequence species are referred to as GS because they generally have a one-to-one correspondence to a gene. Because the library represents the mRNA composition, the variety of GS and their recurrence represent gene expression status in the retina (expression profile). In this expression profile, 108 recurrently appearing GS were assumed to represent very active genes (Fig. 1). The most active gene in the retina was shown to be the opsin gene. It appeared 16 times in total; there are two opsin gene transcripts (GS4783 and GS4763) having different 3′-termini. Among sequences identified in GenBank, 17 species are for protein synthesis machinery, such as ribosomal components or elongation factors. The 17 species comprised 52 clones altogether. Such a high content of mRNA for the protein synthesis machinery is an inherent characteristic of cells with a very high turnover of proteins. Six of the genes identified are for energy production, five for signal transduction and nuclear protein. In addition to the opsin gene, genes for components of phototransduction machinery, namely transducin (GS4626), cone cyclic guanosine monophosphate phosphodiesterase (GS4629), and G-cyclase activating protein (GS4679), were also found to be very active in the retina. The transcript for transferrin (GS2155), which has been regarded as the "serum" component of vitreous fluid, should also be mentioned.

Identification of Genes Uniquely Active in the Retina

The expression pattern of genes represented by recurring Gene signatures were monitored by means of their representation in the expression profiles of four types of blood cells[12] and six epithelial tissues,[11,18,25–28] each

```
AGAAGCTGGT GGCAACTTCA CTGGGGAGAT ATTGCAAATA ACAGCGGGAA CATGAAGCCG    60
                                                          M  K  P
CCACTCTTGG TGTTTATTGT GTGTCTGCTG TGGTTGAAAG ACAGTCACTG CGCACCCACT   120
 P  L  L  V  F  I  V  C  L  L  W  L  K  D  S  H  C  A  P  T
TGGAAGGACA AAACTGCTAT CAGTGAAAAC CTGAAGAGTT TTTCTGAGGT GGGGGAGATA   180
 W  K  D  K  T  A  I  S  E  N  L  K  S  F  S  E  V  G  E  I
GATGCAGATG AAGAGGTGAA GAAGGCTTTG ACTGGTATTA AGCAAATGAA AATCATGATG   240
 D  A  D  E  E  V  K  K  A  L  T  G  I  K  Q  M  K  I  M  M
GAAAGAAAAG AGAAGGAACA CACCAATCTA ATGAGCACCC TGAAGAAATG CAGAGAAGAA   300
 E  R  K  E  K  E  H  T  N  L  M  S  T  L  K  K  C  R  E  E
AAGCAGGAGG CCCTGAAACT TCTGAATGAA GTTCAAGAAC ATCTGGAGGA AGAAGAAAGG   360
 K  Q  E  A  L  K  L  L  N  E  V  Q  E  H  L  E  E  E  E  R
CTATGCCGGG AGTCTTTGGC AGATTCCTGG GGTGAATGCA GGTCTTGCCT GGAAAATAAC   420
 L  C  R  E  S  L  A  D  S  W  G  E  C  R  S  C  L  E  N  N
TGCATGAGAA TTTATACAAC CTGCCAACCT AGCTGGTCCT CTGTGAAAAA TAAGATTGAA   480
 C  M  R  I  Y  T  T  C  Q  P  S  W  S  S  V  K  N  K  I  E
CGGTTTTTCA GGAAGATATA TCAATTTCTA TTTCCTTTCC ATGAAGATAA TGAAAAAGAT   540
 R  F  F  R  K  I  Y  Q  F  L  F  P  F  H  E  D  N  E  K  D
CTCCCCATCA GTGAAAAGCT CATTGAGGAA GATGCACAAT TGACCCAAAT GGAGGATGTG   600
 L  P  I  S  E  K  L  I  E  E  D  A  Q  L  T  Q  M  E  D  V
TTCAGCCAGT TGACTGTGGA TGTGAATTCT CTCTTTAACA GGAGTTTTAA CGTCTTCAGA   660
 F  S  Q  L  T  V  D  V  N  S  L  F  N  R  S  F  N  V  F  R
CAGATGCAGC AAGAGTTTGA CCAGACTTTT CAATCACATT TCATATCAGA TACAGACCTA   720
 Q  M  Q  Q  E  F  D  Q  T  F  Q  S  H  F  I  S  D  T  D  L
ACTGAGCCTT ACTTTTTTCC AGCTTTCTCT AAAGAGCCGA TGACAAAAGC AGATCTTGAG   780
 T  E  P  Y  F  F  P  A  F  S  K  E  P  M  T  K  A  D  L  E
CAATGTTGGG ACATTCCCAA CTTCTTCCAG CTGTTTTGTA ATTTCAGTGT CTCTATTTAT   840
 Q  C  W  D  I  P  N  F  F  Q  L  F  C  N  F  S  V  S  I  Y
GAAAGTGTCA GTGAAACAAT TACTAAGATG CTGAAGGCAA TAGAAGATTT ACCAAAACAA   900
 E  S  V  S  E  T  I  T  K  M  L  K  A  I  E  D  L  P  K  Q
GACAAAGCTC CTGACCACGG AGGCCTGATT TCAAAGATGT TACCTGGGCA GGACAGAGGA   960
 D  K  A  P  D  H  G  G  L  I  S  K  M  L  P  G  Q  D  R  G
CTGTGTGGGG AACTTGACCA GAATTTGTCA AGATGTTTCA AATTTCATGA AAAATGCCAA  1020
 L  C  G  E  L  D  Q  N  L  S  R  C  F  K  F  H  E  K  C  Q
AAATGTCAGG CTCACCTATC TGAAGACTGT CCTGATGTAC CTGCTCTGCA CACAGAATTA  1080
 K  C  Q  A  H  L  S  E  D  C  P  D  V  P  A  L  H  T  E  L
GACGAGGCGA TCAGGTTGGT CAATGTATCC AATCAGCAGT ATGGCCAGAT TCTCCAGATG  1140
 D  E  A  I  R  L  V  N  V  S  N  Q  Q  Y  G  Q  I  L  Q  M
ACCCGGAAGC ACTTGGAGGA CACCGCCTAT CTGGTGGAGA AGATGAGAGG GCAATTTGGC  1200
 T  R  K  H  L  E  D  T  A  Y  L  V  E  K  M  R  G  Q  F  G
TGGGTGTCTG AACTGGCAAA CCAGGCCCCA GAAACAGAGA TCATCTTTAA TTCAATACAG  1260
 W  V  S  E  L  A  N  Q  A  P  E  T  E  I  I  F  N  S  I  Q
GTAGTTCCAA GGATTCATGA AGGAAATATT TCCAAACAAG ATGAAACAAT GATGACAGAC  1320
 V  V  P  R  I  H  E  G  N  I  S  K  Q  D  E  T  M  M  T  D
TTAAGCATTC TGCCTTCCTC TAATTTCACA CTCAAGATCC CTCTTGAAGA AAGTGCTGAG  1380
 L  S  I  L  P  S  S  N  F  T  L  K  I  P  L  E  E  S  A  E
AGTTCTAACT TCATTGGCTA CGTAGTGGCA AAAGCTCTAC AGCATTTTAA GGAACATTTT  1440
 S  S  N  F  I  G  Y  V  V  A  K  A  L  Q  H  F  K  E  H  F
AAAACCTGGT AAGAAGATCT AATGCATCCT ATATCCAGTA AGTAGAATTA TCTCTTCATC  1500
 K  T  W  *
TGGGACCTGG AAATCCTGAA ATAAAAAAGG ATAATGCAAT AAACACAGTT GCAGGAAAGT  1560
ATGTTAGCTA TATACTATGA AGTACTCTTA GTTTACTTAT GTTGAATGGC TTAGCTATTA  1620
ATACTCAAAT TGAGTTAAAA TGAAAATTCC TCCTTAAAAA ATCAAACGTA ATATGTATTA  1680
CATTTCATGG TACATTAGTA GTTCTTTGTA TATTGAATAA ATACTAAATC ACCTAAA     1737
```

FIGURE 3. Nucleotide sequence of the full-length complementary DNA (cDNA) for GS4642 and the amino acid sequence of hypothetical protein (rod photoreceptor protein) encoded in it. The putative signal peptide is underlined. Among 3 of the 7 independent cDNA clones analyzed, a 7-base pair insertion (GTTAGTG) was found at the position between nt 38 and 39 (*arrowhead*), probably because of polymorphism among the pooled samples. The gene signature representing this gene (GS4642) is boxed in. This sequence will appear in the DDBJ, EMBL, and GenBank nucleotide sequence data bank under accession number D63813.

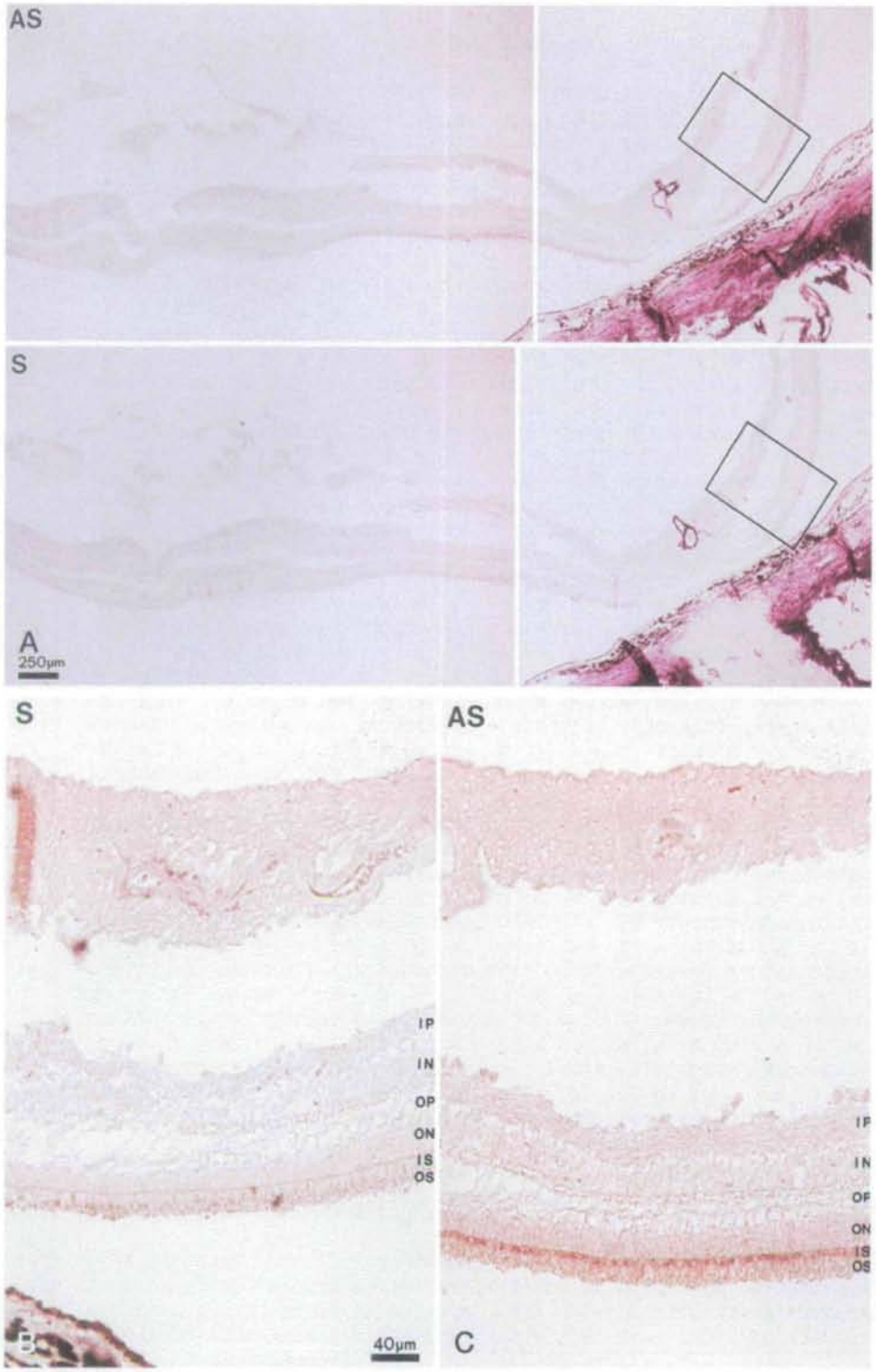

FIGURE 4. In situ hybridization of rod photoreceptor protein (RPP; GS4642) fragment. (**A**) Sections of human retina through the fovea, hybridized with sense (S) and antisense (AS) RPP probes. (**B**) Higher magnification of each section at the peripheral area (boxed area in **A**). OS: outer segment; IS: inner segment; ON: outer nuclear layer; OP: outer plexiform layer; IN: inner nuclear layer; IP: inner plexiform layer.

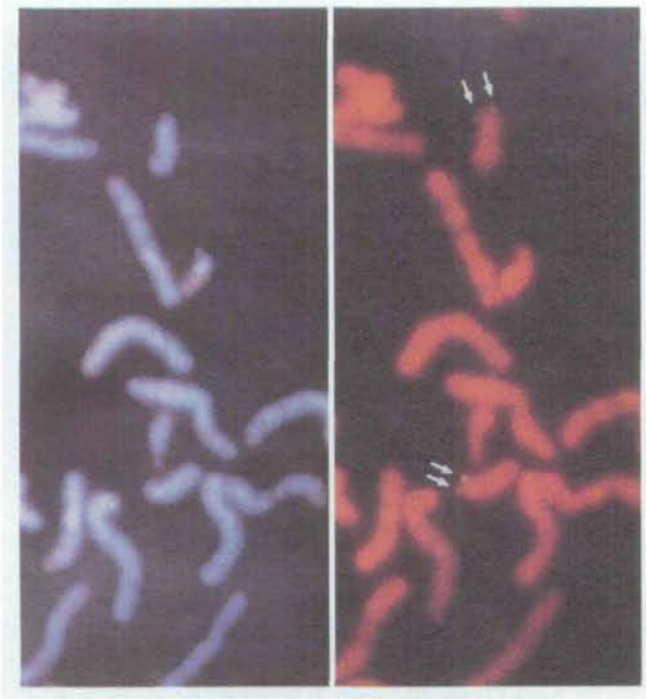

FIGURE 5. Fluorescence in situ hybridization of rod photoreceptor protein (RPP) complementary DNA. (*Right*) Arrows indicate double-spot signals specific for RPP at 18p on both homologous chromosomes 18. Of the 41 metaphase cells examined, 4 exhibited double-spot signals at 18p11.3 on both of the sister chromatids and 13 cells showed double-spots only on one of the homologous chromosomes at 18p11.3. (*Left*) The G-banding pattern of the same chromosomes through a ultraviolet-2A filter (Nikon Inc., Melville, NY) indicates that the precise position of the signals is at 18p11.3.

of which was obtained by the same procedure and assumed to represent the mRNA composition of the material (see Figure 1). Gene signatures also detected in some of these tissues (e.g., GS0917, GS0795) are regarded as those for housekeeping functions, such as protein synthesis or energy production. However, those that are not detected elsewhere are thought to feature preferential expression patterns for the retina. In Table 1, 43 of the listed GS (nonshaded rows) are categorized as such. Because only 10 expression profiles were compared, and they were all nonneural tissues, these genes may include those having additional tissue with high-level expression or those commonly expressed among neural cells. Of the 43 GS, 11 were identified in GenBank (Re94). As anticipated, in addition to all the above-mentioned genes for phototransduction and retinal gene 4 (GS4754),[29] they included genes for aldolase C (GS4888), K-ras oncogene (GS4889), neuron-specific gamma-2 enolase (GS4655), cocaine- and amphetamine-regulated transcript (GS4836), and mitochondrial adenonsine triphosphate–adenosine diphosphate translocator (GS4735). Such limitations originating from a shortage of tissue coverage can be ameliorated by using dbEST.[30] dbEST is a collection of partial cDNA sequences (EST) with >400,000 entries at present, including 2960 3′-ESTs from retina. Because most of the data deposited in dbEST were derived from normalized libraries,[16] which were designed to represent the rarest transcript as frequently as the most abundant one, they do not always represent biologically significant expression. Nevertheless, the large number of entries from brain and some other tissues not covered in our data can be used to eliminate those common to neural cells or those with additional activity sites. Taking only recurrent representation in dbEST as significant, 18 of the 43 GS, unique to the retina among expression profiles, were regarded as common to neural cells or as having additional sites for active expression. The remaining 25 GS, 6 for known and 19 for novel genes, are tentatively categorized as GS representing genes relating unique function or structure of retina (boxed area, Figure 1). All the GS for retina-unique genes identified in GenBank are included among these.

Isolation of a Retina-Specific cDNA

From these final candidates, GS4895 and GS4642, which appeared 11 and 3 times, respectively, and had highly preferential expression patterns for the retina, were selected for further study.

The expression of the genes for GS4642 in a wider variety of tissues was tested by Northern hybridization. Gene signature fragment hybridized with a 1.7-kilobase message in the retina RNA but not in the RNAs from 10 other tissues (Fig. 2). Full-length cDNA clones for this GS were isolated from a conventional cDNA library constructed with poly(A) RNA from pooled retinas. The 1737-bp cDNA contains an open reading frame coding for 466 amino acids (Fig. 3). It shows a significant similarity only to apolipoprotein J/SP-40,40/clusterin,[31] to which 26% (121 of 462) of amino acids were identical when aligned. The cell type producing this transcript was identified by in situ hybridization using human retina specimens. The hybridizing signal apparently distributed over photoreceptor inner segments with density gradually declining from periphery toward fovea, suggesting that this transcript is unique to rod photoreceptor cells (Fig. 4). The corresponding gene was mapped to chromosome 18 at p11.3 (Fig. 5) by fluorescence in situ hybridization.

The analysis of a full-length clone for the other GS, GS4895, revealed that this GS represents a transcript of human sodium-potassium ATPase beta 2 subunit, which has a 387-bp longer 3′ untranslated region than the one in GenBank (Accession: M81181[32])

DISCUSSION

Expression Profile of Retina

In random isolation from a cDNA library representing an mRNA composition, the chance of isolating a clone is proportional to the abundance of the corresponding

transcript in the mRNA, which mainly reflects the rate of synthesis of the encoded protein. In neural retina, disks in the outer segments are renewed constantly among other static components. This is reflected in the expression profile, where the two most abundant transcripts are both for the components of the outer disks, namely opsin and sodium-potassium–transporting ATPase. Transcripts for other components of phototransduction machinery are also found repeatedly in the profile. In addition to these, transcripts for secretory proteins, which also have a rapid turnover, such as transferrin and SP-40,40, are found recurrently. Accordingly, new components for outer segments and novel secretory proteins are expected to be included in the list of very active genes (see Figure 1).

In this study, a total of 607 species of transcripts were detected in retina mRNA. This number might be trivial when compared with the number of genes expressed in the retina, which is composed of 8 different cell types, each of which is thought to express 10,000 to 30,000 genes. Nevertheless, among the GenBank entries for genes that feature unique functions of the retina, about one third (11 of 30) are included in these 607 (Table 1). This unexpectedly high incidence may be partly attributable to the fact that most of the unique functions in the retina reside in parts undergoing constant renewal, which makes the relevant mRNA abundant. Thus, it is expected that more genes related to unique functions of the retina will be discovered by further expansion of the GS collection from the retina. For this purpose, more efficient collection of much simpler "signatures" by short tag concatenation (serial analysis of gene expression[33]) or by cDNA fingerprinting[34] may be used.

By comparing the representation of the retina-unique genes in GS data and dbEST, the differences between these two sets of data are clearly illustrated (see Table 1). In GS data, the representations of all the experimentally proven retina-specific genes are limited to the retina library. However, half of the "hits" are also found in other tissues in dbEST. These representations could reflect additional expression sites. Alternatively, they may represent leaky expressions with no significant physiological meaning.

Isolation of a Novel Rod-Specific Secretory Protein cDNA

Nineteen novel genes whose expression is likely to be unique to the retina were selected by comparing expression profiles. Subtractive hybridization[35] and differential displays[36] are widely used techniques for discovering tissue-specific or disease-specific genes, but our system using "mRNA subtraction in silico" has been shown to be quite powerful as well. Most importantly, this method allows us to compare many tissues at one time. In principle, moreover, no uniquely active genes will be overlooked by this method, as long as their transcripts appear above a certain level (in the example of Figure 1, twice among 789 collections).

One of the retina-specific GS (GS4629) identified by this approach was the γ subunit of human cone cyclic guanosine monophosphate phosphodiesterase.[19] Another unique gene candidate, GS4642, was selected in this study. Although the sequence for GS4642 was represented in dbEST (GenBank, R61493) as a transcript in brain, our study has shown that its expression, at a level detectable by Northern hybridization, is limited to the retina (Fig. 2). In situ hybridization located the corresponding gene in the rod photoreceptor. We thus named the hypothetical protein encoded by this transcript "rod photoreceptor protein" (RPP). The predicted amino acid sequence of RPP has a signal peptide sequence of 21 amino acids, detected by the protein sorting prediction program PSORT (available on the World Wide Web at http://www.imcb.osaka-u.ac.jp),[37] and the most likely space for secretion is the interphotoreceptor matrix, because most of the secretion of photoreceptor cells is directed to this space. Immunohistochemistry using the antibody to RPP should lead to further confirmation. A significant similarity to RPP has been noted only in apolipoprotein J/SP-40,40/clusterin. Although this lipid-binding protein has been detected in diverse tissues and body fluids, its function remains elusive. In retina it is speculated to be a lipid scavenger transporter.[38,39] The RPP gene was mapped to chromosome 18p11.3, where no retinal disease had been assigned thus far.[40]

Thus, the function of RPP in retinal tissue remains unclear, and further studies on its protein products are needed. The rest of the unique candidates will be subjected to similar studies. Even if their functions cannot be readily determined, they can be used as probes for studying the physiology and pathology of the retina.

Acknowledgments

The authors thank Dr. Kohichi Itoh, Dr. Yasuhide Miyamoto, and Dr. Yasuhiro Matsumoto for technical help and Dr. Kenta Nakai for constructing the World Wide Web Body Map server.

Key Words

cloning, gene expression, gene signatures, retina, rod photoreceptor protein

References

1. Yamaki K, Tsuda M, Shinohara T. The sequence of human retinal S-antigen reveals similarities with α-transducin. *FEBS Lett.* 1988;234:39–43.
2. Murakami A, Yajima T, Inana G. Isolation of human retinal genes: Recoverin cDNA and gene. *Biochem Biophys Res Comm.* 1992;187:234–244.
3. Travis GH, Christerson L, Danielson E, et al. The human retinal degeneration slow (RDS) gene: Chromosomal assignment and structure of the mRNA. *Genomics.* 1991;10:733–739.
4. Bascom RA, Manara S, Collins L, Molday RS, Kalnins V, McInnes RR. Cloning of the cDNA for a novel pho-

toreceptor membrane protein (rom-1) identifies a disk rim protein family implicated in human retinopathies. *Neuron.* 1992;8:1171–1184.
5. Si J-S, Brost DE, Redmond TM, Nickerson JM. Cloning of cDNAs encoding human interphotoreceptor retinoid-binding protein (IRBP) and comparison with bovine IRBP sequences. *Gene.* 1989;80:99–108.
6. Intres R, Goldflam S, Cook JR, Crabb JW. Molecular cloning and structural analysis of the human gene encoding cellular retinaldehyde-binding protein. *J Biol Chem.* 1994;269:25411–25418.
7. Fuchs S, Nakazawa M, Maw M, Tamai M, Oguchi Y, Gal A. A homologous 1-base pair deletion in the arrestin gene is a frequent cause of Oguchi disease in Japanese. *Nature Genet.* 1995;10:360–362.
8. Farrar GJ, Kenna P, Jordan SA, et al. A three-base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature.* 1991;354:478–480.
9. Kajiwara K, Berson EL, Dryja TP. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science.* 1994;264:1604–1607.
10. Okubo K, Hori N, Matoba R, Niyama T, Matsubara K. A novel system for large-scale sequencing of cDNA by PCR amplification. *DNA Seq.* 1991;2:137–144.
11. Okubo K, Hori N, Matoba R, et al. Large-scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genet.* 1992;2:173–179.
12. Okubo K, Itoh K, Fukushima A, Yoshii J, Matsubara K. Monitoring cell physiology by expression profiles and discovering cell type-specific genes by compiled expression profiles. *Genomics.* 1995;30:178–186.
13. Adams MD, Kelley JM, Gocayne JD, et al. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science.* 1991;252:1651–1656.
14. Adams MD, Dubnick M, Kerlavage AR, et al. Sequence identification of 2,375 human brain genes. *Nature.* 1992;355:632–634.
15. Bell GI, Takeda J. A molecular inventory of human pancreatic islets: sequence analysis of 1000 cDNA clones. *Human Mol Genet.* 1993;2:1793–1798.
16. Hiller L, Lennon G, Becker M, et al. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* 1996;6:807–828.
17. Bernstein SL, Brost DE, Neuder ME, Wong P. Characterization of a human fovea cDNA library and regional differential gene expression in the human retina. *Genomics.* 1996;32:301–308.
18. Nishida K, Adachi W, Shimizu-Matsumoto A, et al. An expression profile of active genes in human corneal epithelium and the isolation of human keratin 12 cDNA. *Invest Ophthalmol Vis Sci.* 1996;37:1800–1809.
19. Shimizu-Matsumoto A, Itoh K, Inazawa J, et al. Isolation and chromosomal localization of human cone cGMP phosphodiesterase γ cDNA. *Genomics.* 1996;32:121–124.
20. Maeda K, Okubo K, Shimomura I, Funahashi T, Matsuzawa Y, Matsubara K. cDNA cloning and expression of a novel adipose specific collagen-like factor, apM1 (Adipose Most Abundant Gene Transcript 1). *Biochem Biophys Res Comm.* 1996;221:286–289.
21. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA.* 1988;85:2444–2448.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–410.
23. Deininger PL. Random subcloning of sonicated DNA: Application to shotgun DNA sequencing analysis. *Anal Biochem.* 1981;129:216.
24. Inazawa J, Saito H, Ariyama T, Abe T, Nakamura Y. High-resolution cytogenetic mapping of 432 new cosmid markers including 43 RFLP markers on human chromosome 17 by fluorescence in situ hybridization. *Genomics.* 1993;17:153–162.
25. Okubo K, Yoshii J, Yokouchi H, Kameyama M, Matsubara K. An expression profile of active genes in human colonic mucosa. *DNA Res.* 1994;1:37–45.
26. Itoh K, Okubo K, Yoshii J, Yokouchi H, Matsubara K. An expression profile of active genes in human lung. *DNA Res.* 1994;1:278–287.
27. Kawamoto S, Matsumoto Y, Mizuno K, Okubo K, Matsubara K. Expression profiles of active genes in human and mouse livers. *Gene.* 1996;174:151–158.
28. Kita H, Okubo K, Matsubara K. An expression profile of active genes in cultured human keratinocytes. *DNA Res.* 1996;3:1–7.
29. Higashide T, Murakami A, McLaren JM, Inana G. Cloning of the cDNA for a novel photoreceptor protein. *J Biol Chem.* 1996;271:1797–1804.
30. Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for "expressed sequence tags" [letter]. *Nat Genet.* 1993;4:332–333.
31. Kirszbaum L, Sharpe JA, Murphy B, et al. Molecular cloning and characterization of the novel human complement-associated protein, SP-40,40: A link between the complement and reproductive systems. *EMBO J.* 1989;8:711–718.
32. Hernando N, Martin-Vasallo P, Ghosh S, Ghosh PK, Swaroop A, Coca-Prados M. Nucleotide sequence of a cDNA for the beta 2 subunit isoform of Na^+, K(+)-ATPase from human retina. *Biochim Biophys Acta.* 1994;1189:109–111.
33. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science.* 1995;270484–487.
34. Kato K. Description of the entire mRNA population by a 3′ end cDNA fragment generated by class IIS restriction enzymes. *Nucleic Acids Res.* 1995;23:3685–3690.
35. Duguid JR, Rohwer RG, Seed B. Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc Natl Acad Sci USA.* 1988;85:5738–5742.
36. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* 1992;257:967–961.
37. Nakai K, Kanehisa M. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics.* 1992;14:897–911.
38. Jones SE, Meerabux JM, Yeats DA, Neal MJ. Analysis of differentially expressed genes in retinitis pigmentosa retinas. Altered expression of clusterin mRNA. *FEBS Lett.* 1992;300:279–282.
39. Jomary C, Ahir A, Agarwal N, Neal MJ, Jones SE. Spatiotemporal pattern of ocular clusterin mRNA expression in the *rd* mouse. *Mol Brain Res.* 1995;29:172–176.
40. Bird AC. Retinal photoreceptor dystrophies. *Am J Ophthalmol.* 1995;119:543–562.