Molecular Vision 2002; 8:185-95 <http://www.molvis.org/molvis/v8/a25>
Received 31 August 2001 | Accepted 14 December 2001 | Published 15 June 2002

Expressed sequence tag analysis of adult human iris for the NEIBank Project: Steroid-response factors and similarities with retinal pigment epithelium

Graeme Wistow,[1]Steven L. Bernstein,[2]Sugata Ray,[1]M. Keith Wyatt,[1]Amita Behal,[1]Jeffrey W. Touchman,[3]Gerard Bouffard,[3]Don Smith,[1]Katherine Peterson[1]

[1]Section on Molecular Structure and Function, National Eye Institute, National Institutes of Health, Bethesda, MD; [2]Departments of Ophthalmology and Neurobiology & Genetics, University of Maryland School of Medicine, Baltimore, MD; [3]NIH Intramural Sequencing Center, Gaithersburg, MD

Purpose: The iris is a specialized tissue with important roles in the development and function of the eye. It is involved in diseases, including glaucoma and ocular melanoma, and its pigmented cells share an origin with the retinal pigment epithelium (RPE). Expressed sequence tag (EST) analysis of human iris has been performed to explore the repertoire of genes expressed in this tissue.

Methods: An unamplified, un-normalized cDNA library (designated *bx*) was constructed from pooled (4-80 years old) human iris tissue. Over 2000 clones were picked and sequenced. Sequences were analyzed and clustered using GRIST (GRouping and Identification of Sequence Tags). The library was then normalized (and designated *fg*) and a further 2200 clones were sequenced for deeper examination of rarer sequence. Some sequences of interest were investigated further by standard methods.

Results: From *bx* and *fg* libraries respectively, 1263 and 1604 clusters of expressed genes have been identified, giving a combined total of almost 2700 potentially unique genes. The most abundant novel transcript in *bx* is oculoglycan/opticin. Others include glucocorticoid induced leucine zipper protein (GILZ), Ris, a novel member of the Ras family, Iris Ring Finger (IRF), a member of the midline family, melastatin 2 (MLSN2), a member of the transient receptor potential calcium channel family, and iris expressed growth factor (IEGF), a member of the VEGF/PDGF family. Several factors involved in steroid responses are also represented.

Conclusions: The iris libraries are a rich source of novel as well as known genes, including molecular markers for pigmented cells that are also shared with RPE. A number of transcripts code for proteins involved in steroid response, with interesting implications for control of intraocular pressure. These sequence verified clones provide a nonredundant set for micro-array construction.

The iris is a distinctive and familiar feature of human appearance. It has a major mechanistic role in the eye, controlling the amount of light that reaches the retina by contracting or relaxing in response to ambient light levels. In addition, pigment in the iris helps to reduce glare in the eye. The iris also serves a structural role in the eye as a physical delineator of the anterior segment. Anteriorly, it provides a large surface area in contact with the aqueous humor, and as such it is ideally located to affect intraocular pressure. Indeed, recent results from a mouse model have linked stromal atrophy and pigment dispersion in the iris with glaucoma [1]. Posteriorly, the iris epithelium contributes in a similar way to the anterior chamber, a space enclosing the proliferative zone of the epithelial cells of the lens. Growth factors secreted into the anterior chamber are thought to contribute to the processes of lens growth by triggering the processes that lead to the proliferation, migration and eventual differentiation of lens epithelial cells into fibers [2]. The identity of factors involved in this signaling is not yet known.

The iris is a pigmented tissue and its characteristic colors derive from the degree of pigmentation. Iris pigment epithelium cells (IPE) share a developmental origin with the better known retinal pigment epithelium (RPE) cells and it has been proposed that IPE cells may be able to serve as autologous replacements for RPE in cell transplants aimed at treating retinal degenerative diseases [3]. The iris also contains melanocytes of neural crest derivation and is one of the ocular tissues afflicted by melanoma [4].

The iris contains several other cell types, including epithelial and muscle cells and nerves of parasympathetic origin. Relatively little is known about the molecular markers associated with any of these cell types in the iris. The iris shares developmental and anatomical features with the neighboring ciliary body. The iris epithelium is a continuation of the ciliary epithelium while the non-pigmented layer of the ciliary epithelium becomes pigmented in the iris. The pigmented ciliary epithelium, a continuation of the RPE, is continuous with the iris epithelium. Some molecular analysis of human ciliary

Correspondence to: Graeme Wistow, Ph.D., Chief, Section on Molecular Structure and Function, National Eye Institute, Building 6, Room 331,National Institutes of Health, Bethesda, MD, 20892-2740; Phone: (301) 402-3452; FAX: (301) 496-0078; email: graeme@helix.nih.gov

body has been performed and 289 ESTs derived from a subtracted human ciliary body cDNA library [5,6] have been deposited in the dbEST section of GenBank.

Expressed sequence tag (EST) analysis is a powerful tool for probing the molecular repertoire of tissues and cell type. NEIBank is a project aimed at expanding the database of genes expressed in regions of the eye [7]. Here we describe the initial results of EST analysis of a cDNA library made from pooled human iris tissue. The unamplified, un-normalized library (*bx*) gives an estimation of abundantly expressed genes. It also tends to reveal cDNAs that may be selectively lost through library manipulations such as amplification because of poor growth characteristics in bacterial hosts. The normalized version of the same library (*fg*) loses abundance information, but has the potential for deeper penetration of the rarer transcripts.

This paper will summarize some of the basic findings from the ongoing analyses of these libraries. Additional sequences and annotation will be added to NEIBank as the project continues. Clones from all the NEIBank libraries are being compiled in a large nonredundant set for construction of human eye cDNA microarrays.

```
   L  R  G  L  L  L  P  V  P  R  I  S  A  Q  A  Q
TGCTCCGGGGACTGCTTTTGCCAGTACCCAGAATCAGTGCTCAGGCTCAG   50
 K  S  W  I  E  R  A  F  Y  K  R  E  C  V  H  I  I
AAATCCTGGATAGAAAGAGCATTTTATAAAAGAGAATGTGTCCACATCAT  100
   P  S  T  K  D  P  H  R  C  C  C  G  R  L  I  G
ACCCAGCACCAAAGACCCCCATAGGTGTTGCTGTGGGCGTCTGATAGGCC  150
Q  H  V  G  L  T  P  S  I  S  V  L  Q  N  E  K  N
AGCATGTTGGCCTCACCCCCAGTATCTCCGTGCTTCAGAATGAGAAAAAT  200
 E  S  R  L  S  R  N  D  I  Q  S  E  K  W  S  I  S
GAAAGTCGCCTCTCCCGAAATGACATCCAGTCTGAAAAGTGGTCCATCAG  250
   K  H  T  Q  L  S  P  T  D  A  F  G  T  I  E  F
CAAACACACTCAACTCAGCCCTACGGATGCTTTTGGGACCATTGAGTTCC  300
Q  G  G  G  H  S  N  K  A  M  Y  V  R  V  S  F  D
AAGGAGGTGGCCATTCCAACAAAGCCATGTATGTGCGAGTATCTTTTGAT  350
 T  K  P  D  L  L  L  H  L  M  T  K  E  W  Q  L  E
ACAAAACCTGATCTCCTCTTACACCTGATGACCAAGGAATGGCAGTTGGA  400
   L  P  K  L  L  I  S  V  H  G  G  L  Q  N  F  E
GCTTCCCAAGCTTCTCATCTCTGTCCATGGGGGCCTGCAGAACTTTGAAC  450
L  Q  P  K  L  K  Q  V  F  G  K  G  L  I  K  A  A
TCCAGCCAAAACTCAAGCAAGTCTTTGGGAAAGGGCTCATCAAAGCAGCT  500
 M  T  T  G  A  W  I  F  T  G  G  V  N  T  G  V  I
ATGACAACTGGAGCGTGGATATTCACTGGAGGGGTTAACACAGGTGTTAT  550
   R  H  V  G  D  A  L  K  D  H  A  S  K  S  R  G
TCGTCATGTTGGCGATGCCTTGAAGGATCATGCCTCTAAGTCTCGAGGAA  600
K  I  C  T  I  G  I  A  P  W  G  I  V  E  N  Q  E
AGATATGCACCATAGGTATTGCCCCCTGGGGAATTGTGGAAAACCAGGAG  650
D  L  I  G  R  D  V  V  R  P  Y  Q  T  M  S  N  P
GACCTCATTGGAAGAGATGTTGTCCGGCCATACCAGACCATGTCCAATCC  700
   M  S  K  L  T  V  L  N  S  M  H  S  H  F  I  L
CATGAGCAAGCTCACTGTTCTCAACAGCATGCATTCCCACTTCATTCTGG  750
A  D  N  G  T  T  G  K  Y  G  A  E  V  K  L  R  R
CTGACAACGGGACCACTGGAAAATATGGAGCAGAGGTGAAACTTCGAAGA  800
 Q  L  E  K  H  I  S  L  Q  K  I  N  T  R  I  G  Q
CAACTGGAAAAGCATATTTCACTCCAGAAGATAAACACAAGAATCGGTCA  850
   G  V  P  V  V  A  L  I  V  E  G  G  P  N  V  I
AGGTGTTCCTGTGGTGGCACTCATAGTGGAAGGAGGACCCAATGTGATCT  900
S  I  V  L  E  Y  L  R  D  T  P  P  V  P  V  V  V
CGATTGTTTTGGAGTACCTTCGAGACACCCCTCCCGTGCCAGTGGTTGTC  950
 C  D  G  S  G  R  A  S  D  I  L  A  F  G  H  K  Y
TGTGATGGGAGTGGACGGGCATCGGACATCCTGGCCTTTGGGCATAAATA 1000
   S  E  E  G  G  *
CTCAGAAGAAGGCGGGTAGGTAACTTTCCAGGCCCCATGGAAGAACCCTA 1050
AAGCCTGTTTGGAAACGAGGGTATGAGTGGATTATGTTTTCAGTAGCTCA 1100
ACCAAGACCTCAAATCAAAACAAGCTATGAACAAATTGTCTAAAAAATGT 1150
CTGTCATGGGAGGGCTGTGGTGAAGAACAGAGAAACATATTCTAAATGTC 1200
CTGTGAAGTGGGAAATTCTATGAAAGCTACACGGATAATAAAAGGGTGA 1250
GGAAAAGAAAAAAAAAAAAAAAAA                           1274
```

Figure 1. Sequence of clone bx18g11 for Melastatin 2 (MLSN2). The full sequence of the clone is shown, together with the ORF (which extends further 5' than that originally proposed for MLSN1). The clone is probably not full length at the 5' end. The polyadenylation signal is shown in red.

METHODS

Tissue and RNA preparation: Human post-mortem irises were obtained under University of Maryland School of Medicine IRB exemptions SB-019701 and SB-129901. Total RNA was extracted using RNAzol or TRIzol (Tel-Test Inc., Friendswood, TX). PolyA RNA was prepared using an oligo(dT) cellulose affinity column [8]. Tissue was pooled from 10 individuals ranging in age from 4-80 years and RNA was extracted. From this pooled sample an aliquot of 60 μg of total RNA yielded 2.17 μg of mRNA.

cDNA library construction, sequencing, and bioinformatics: A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies (Rockville, MD; now part of Invitrogen Corp.) [9]. For this library, cDNA inserts were cloned into the NotI/EcoRV sites of the vector. Normalization was also performed using a self subtraction procedure based on hybridization of biotinylated single strand RNA copies of the clones with single strand DNA copies [9].

Sequencing was performed at the NIH Intramural Sequencing Center (NISC). Sequences were grouped and identified using **GR**ouping and **I**dentification of **S**equence **T**ags (GRIST; [10]). Sequences were also analyzed using public genome resources at the National Center for Biotechnology Information (NCBI) and the Human Genome Project. Protein motifs were searched using GenomeNet, the NCBI, and the Swiss Institute of Bioinformatics, using the Pfam [11] and Smart [12] databases of motif sequences. Transmembrane helix and signal peptide prediction was performed using the

Table 1. The "top 25" clones in the un-normalized iris cDNA library *bx*

Rank	Description	Gene ID	UniGene	Chromosome	#
1	clusterin (apolipoprotein J)	M64722	75106	8p21-p12	38
2	elongation factor 1 α 1	BC001412	181165	6q14	36
3	glutathione peroxidase 3 (plasma)	D00632	172153	5q23	36
4	ribosomal protein L3 (RPL3)	XM_009998	119598	22q12.1-q12.3	14
5	GADPH	AF261085	169476	12p13	13
6	opticin/oculoglycan	XM_001780	241413	1q31	13
7	prostaglandin D2 synthase	XM_005431	8272	9q34.2-q34.3	12
8	vimentin	BC000163	2064	10p13	12
9	Serpin G1	XM_006339	151242	11q12-q13.1	10
10	aldolase A	X05236	273415	16q22-q24	10
11	actin β	BC001301	288061	7p15-p12	9
12	myosin regulatory light chain 2	XM_009501	9615	20	9
13	ribosomal protein L4	BC001365	286	15q	9
14	Pmel (silver homolog)	BC001414	95972	12q13-q14	8
15	cofilin 1 (non-muscle)	NM_005507	180370	11q13	8
16	α-tubulin	AK026516	272897	2q	8
17	gelsolin	NM_000177	290070	9q33	7
18	ribosomal protein L13a	BC000514			7
19	insulin-like growth factor binding protein 7 (IGFBP7)	NM_001553	119206	4q12	7
20	tissue inhibitor of metalloproteinase 3 (TIMP3)	XM_009943	245188	22q12.3	6
21	prothymosin α	BC003510	250655	2q35-q36	6
22	MHC class II antigen γ chain	K01144	84298	5q32	6
23	MHC class I lymphocyte antigen (HLA-A 2402)	AK027084	181244	6p21.3	6
24	ATPase, H+ transporting, lysosomal (vacuolar proton pump) 16kD	M62762	76159	16p13.3	6
25	ubiquitin C	AB009010	183704	12q24.3	6

Clusters are arranged by abundance. This list is an edited detail from the output of the clustering procedure GRIST. Representative GenBank entries are shown, along with the appropriate Unigene, chromosomal location, and the number of ESTs identified.

TMHMM server. Cladistic analysis used the program MEGA2 [13]. For this procedure, gaps and missing data were handled by pair-wise deletion, Poisson correction was applied and the UPGMA option was used to generate trees of related sequences.

RH mapping: For some clones, chromosomal location was not immediately apparent from the databases. Radiation hybrid mapping for selected genes was performed at Research Genetics (Huntsville AL), using the Stanford G3 panel (Stanford Human Genome Center, Stanford, CA). PCR primers were designed from the 3' UTR of each sequence, a region unlikely to be interrupted by introns in the genome. PCR was used to amplify unique marker sequences from a total of 83 clones and two controls. An email server operated by the Stanford Human Genome Center was used to link the marker to more than 15000 framework markers. For clone bx2h05 (IEGF) primers used were CTT CTC CCT TGG TTA CCT GTT and TAA TTT GTA CGG CCA GCT TTT. For clone bx05f09 (Ris), primers used were AGC AAG AAA AGG CTG TCC AA and TCT GAC CTC AGC TCC TCC AT.

TABLE 2. TRANSCRIPTION FACTORS

Description	GenBank	UniGene	Chromosome	#
transforming growth factor β-stimulated protein TSC-22	BC000456	114360	13q14	4
activating transcription factor 4 (ATF4)	XM_010004	181243	22q13.1	2
c-jun	BC002646	78465	1p32-p31	2
glucocorticoid-induced leucine zipper (GILZ)	AF183393	75450	4	2
inhibitor of DNA binding 3	BC003107	76884	1p36.13-p36.12	2
sterol regulatory element-binding protein (CNBP)	M28372	2110	3q13.3-q24	2
sterol regulatory element binding transcription factor 1 (SREBF1)	NM_004176	166	17p11.2	2
ELK1	AB016194	181128	Xp11.2	1
AT-binding transcription factor 1 (ATBF1)	NM_006885	101842	16q22.3-q23.1	1
chromobox homolog 6 (CBX6)	NM_014292	107374	22q13.1	1
fos	NM_005252	25647	14q24.3	1
glucocorticoid receptor repression factor 1 (GRF-1)	M73077	102548	19q13.3	1
high-mobility group (nonhistone chromosomal) protein 1	NM_002128	144321	13q12	1
homeo box B2 (HOXB2)	NM_002145	2733	17q21-q22	1
inhibitor of DNA binding	BC000613	75424	20q11	1
jun B	NM_002229	198951	19p13.2	1
nuclear factor (erythroid-derived 2)-like 1 (NFE2L1)	NM_003204	83469	17q21.3	1
TAF2G	BC003400	60679	5q11.2-q13.1	1
TAF2N	AB010060	66772	17q11.1-q11.2	1
transcription factor NRF	XM_010360	119018	X	1
upstream transcription factor 2, c-fos interacting (USF2)	AY007087	93649	19q13	1
X-box binding protein 1	BC000938	149923	22q12.1	1
YL-1	D43642	2430	1q21	1
79% protein similarity to: procKr2 - chicken			19q13.42	1
sterol regulatory element binding transcription factor 1 (SREBF1)	NM_004176	166	17p11.2	3
albumin D-box binding protein mRNA	U79283	155402	19q13.3	2
CCAAT/enhancer binding protein δ(C/EBPD)	XM_005023	76722	8p11.2-p11.1	2
high mobility group protein (HMG-I(Y))	L17131	139800	6p21	2
interferon regulatory factor 2 (IRF2)	XM_003522	83795	4q34.1-q35.1	2
interferon regulatory factor 3 (IRF3)	NM_001571	75254	19q13.3-q13.4	2
leucine-zipper-like transcriptional regulator, 1 (LZTR1)	XM_009853	301415		2
activating transcription factor 1		36908	12q13	1
AP-2 α	M36711	18387	6p24	1
BANP homolog, SMAR1 homolog	AK001039	194637	16q24	1
butyrate response factor 2 (EGF-response factor 2)	NM_006887	78909	2	1
CAAT-box binding transcription factor CTF-1	X12492	184771	19p13.3	1
COUP-TFII	M62760	288869	15	1
CREB3 (luman)	AF211848	287921	9	1
FEZ1 (FEZ1) mRNA, complete cds	AF123659	93605	8p22	1
forkhead box O1A (FKHR)	AF032885	170133	13q14.1	1
FOXJ2 forkhead factor	XM_006928	120844	12	1
HMG box 20B	AF146223	32317	19p13.3	1
HMG box (SSRP1) mRNA, complete cds	M86737	79162	11q12	1
HMG domain protein HMGX2 (HMGX2)	AF146223	32317	19p13.3	1
homeobox protein (HLX1)	U14325	74870	1q41-q42.1	1
inhibitor of DNA binding 3	BC003107	76884	1p36.13-p36.12	1
interleukin enhancer binding factor 3 (NFAT subunit)	AJ271745	256583	19	1
modulator recognition factor I (MRF-1)	M62324	920	2	1
myoneurin	NM_018657	22934	3	1
NFKBIB	XM_008901	9731	19q13.1	1
nuclear factor I/C (CCAAT-binding transcription factor)		184771	19p13.3	1
nuclear factor, interleukin 3 regulated (NFIL3)	XM_005337	79334	9q22	1
nuclear transcription factor Y, γ (NFYC)	NM_014223	168157	1p32	1
orphan homeobox protein (PRH)	L16499	118651	10	1
p65 subunit of transcription factor NF-kappaB	Z22951	75569	11q13	1
PCAF associated factor 65 α	AF069735	131846	11	1
runt-related transcription factor 1 (AML1)	D10570	129914	21q22.3	1
serum response factor (SRF)	J03161	155321	6	1
sine oculis homeobox homolog 3 (SIX3), mRNA	XM_002400	227277	2p16-p21	1
SRY (sex determining region Y)-box 10		44317	22q13.1	1
TAF2G		60679	5q11.2-q13.1	1
TAF2J	NM_005644	82037		1
Weakly similar to A57717 transcription factor EC2		270334		1
zinc finger homeobox protein ZHX1	AF106862	12940	8q	1
98% protein similarity to HES-5 [mouse]		57971		1

Entries in black are from the un-normalized *bx* collection; those in red are from the normalized *fg* collection. Representative GenBank entries are shown, along with the appropriate Unigene, chromosomal location and the number of ESTs identified using GRIST.

RESULTS & DISCUSSION

Libraries: The *bx* unamplified, un-normalized library in pCMVSPORT6 contains 6.4x10^6 primary recombinants with an average insert size of 1.5 kbp. The average length of high quality sequence reads is 480 bp. The content of empty vector clones is 0%, while clones containing mitochondrial genome represent 7% and rRNA 0.5% of the total. For the normalized version of the library, *fg*, the insert size is unchanged while rRNA content is 0.4%. As one measure of the effect of normalization, the content of cDNA for elongation factor 1α(EF1α) is 93 fold lower in *fg* than in *bx*. The full listing of clones obtained from these iris libraries can be found at NEIBank.

Abundant cDNAs: similarities with RPE: While cDNAs from lens are dominated by the characteristic lens crystallins [9], the most abundant clones in the *bx* unamplified iris library are typical of other tissues, including the stress protein apolipoprotein J, EF1α, ribosomal protein L3, and glyceral-

```
TGGCGGCCGTGGAGTTTGTGACATACGAGGTGACACCCCTCGAGTCACTTCCCTTCAACT    60
CCAGCTGGAGCGCCTGCTTGGCTTTGGGTTCGTTCTGCAGCCTTCGCCCCGCTCCTAGCC  120
TCAGGGCCGGACTCCAGCGCAGAGCCCAGCCCAGCGCAGCCTGCCAGCAGCCACCCAGCC  180
                                                          M       1
GCCCAGCCGCCCAGCCCCGCACGAAACCCGGCCAGAGCTTCCTAGCAGCCCGAGCCATGA  240
N  T  E  M  Y  Q  T  P  M  E  V  A  V  Y  Q  L  H  N  F  S      21
ACACCGAAATGTATCAGACCCCCATGGAGGTGGCGGTCTACCAGCTGCACAATTTCTCCA  300
I  S  F  F  S  S  L  L  G  G  D  V  V  S  V  K  L  D  N  S      41
TCTCCTTCTTCTCTTCTCTGCTTGGAGGGGATGTGGTTTCCGTTAAGCTGGACAACAGTG  360
A  S  G  A  S  V  V  A  I  D  N  K  I  E  Q  A  M  D  L  V      61
CCTCCGGAGCCAGCGTGGTGGCCATAGACAACAAGATCGAACAGGCCATGGATCTGGTGA  420
K  N  H  L  M  Y  A  V  R  E  E  V  E  I  L  K  E  Q  I  R      81
AGAATCATCTGATGTATGCTGTGAGAGAGGAGGTGGAGATCCTGAAGGAGCAGATCCGAG  480
E  L  V  E  K  N  S  Q  L  E  R  E  N  T  L  L  K  T  L  A     101
AGCTGGTGGAGAAGAACTCCCAGCTAGAGCGTGAGAACACCCTGTTGAAGACCCTGGCAA  540
S  P  E  Q  L  E  K  F  Q  S  C  L  S  P  E  E  P  A  P  E     121
GCCCAGAGCAGCTGGAGAAGTTCCAGTCCTGTCTGAGCCCTGAAGAGCCAGCTCCCGAAT  600
S  P  Q  V  P  E  A  P  G  G  S  A  V  *                       134
CCCCACAAGTGCCCGAGGCCCCTGGTGGTTCTGCGGTGTAAGTGGCTCTGTCCTCAGGGT  660
GGGCAGAGCCACTAAACTTGTTTTACCTAGTTCTTTCCAGTTTGTTTTTGGCTCCCCAAG  720
CATCATCTCACGAGGAGAACTTTACACCTAGCACAGCTGGTGCCAAGAGATGTCCTAAGG  780
ACATGGCCACCTGGGTCCACTCCAGCGACAGACCCCTGACAAGAGCAGGTCTCTGGAGGC  840
TGAGTTGCATGGGGCCTAGTAACACCAAGCCAGTGAGCCTCTAATGCTACTGCGCCCTGG  900
GGGCTCCCAGGGCCTGGGCAACTTAGCTGCAACTGGCAAAGGAGAAGGGTAGTTTGAGGT  960
GTGACACCAGTTTGCTCCAGAAAGTTTAAGGGGTCTGTTTCTCATCTCCATGGACATCTT 1020
CAACAGCTTCACCTGACAACGACTGTTCCTATGAAGAAGCCACTTGTGTTTTAAGCAGAG 1080
GCAACCTCTCTCTTCTCCTCTGTTTCGTGAAGGCAGGGGACACAGATGGGAGAGATTGAG 1140
CCAAGTCAGCCTTCTGTTGGTTAATATGGTATAATGCATGGCTTTGTGCACAGCCCAGTG 1200
TGGGATTACAGCTTTGGGATGACCGCTTACAAAGTTCTGTTTGGTTAGTATTGGCATAGT 1260
TTTTCTATATAGCCATAAATGCGTATATATACCCATAGGGCTAGATCTGTATCTTAGTGT 1320
AGCGATGTATACATATACACATCCACCTACATGTTGAAGGGCCTAACCAGCCTTGGGAGT 1380
ATTGACTGTTCCCTTACCTCTTATGGCTAAGTCTTTGACTGTGTTCATTTACCAAGTTGA 1440
CCCAGTTTGTCTTTTAGGTTAAGTAAGACTCGAGAGTAAAGGCAAGGAGGGGGGCCAGCC 1500
TCTGAATGCGGCCACGGATGCCTTGCTGCTGCAACCCTTTCCCCAGCTGTCCACTGAAAC 1560
GTGAAGTCCTGTTTTGAATGCCAAACCCACCATTCACTGGTGCTGACTACATAGAATGGG 1620
GTTGAGAGAAGATCAGTTTGGGCTTCACAGTGTCATTTGAAAACGTTTTTTGTTTTGTTT 1680
TGTAATTATTGTGGAAAACTTTCAAGTGAACAGAAGGATGGTGTCCTACTGTGGATGAGG 1740
GATGAACAAGGGGATGGCTTTGATCCATGGAGCCTGGGAGGTGTGCCCAGAAAGCTTGTC 1800
TGTAGCGGGTTTTGTGAGAGTGAACACTTTCCACTTTTTGACACCTTATCCTGATGTATG 1860
GTTCCAGGATTTGGATTTTGATTTTCCAAATGTAGCTTGAAATTTCAATAAACTTTGCTC 1920
TGTTTTTCTAAAA                                                 1933
```

Figure 2. Human homolog of mouse glucocorticoid induced leucine zipper protein (GILZ). Sequence of clone bx18e04 (GenBank accession number AF183393) and predicted translation. The protein domain related to TSC-22 and *shortsighted* proteins is shown in blue with the leucines of the zipper motif in red. The polyadenylation signal is shown in green.

dehyde-3-phosphate dehydrogenase (Table 1). The anti-oxidation enzyme plasma glutathione peroxidase 3 (GP3) is ranked third in abundance with 36 copies. The iris is subject to variety of oxidative stresses; the tissue is exposed to sunlight and there are high levels of hydrogen peroxide in the aqueous humor [14,15]. GP3 may be part of a defense mechanism to protect the iris from oxidative damage. The high level of GP3 expression may also be related to the pigment epithelial cells of the iris, since the enzyme is also highly abundant (the second ranking transcript) in the NEIBankRPE/choroid sequence dataset [16].

Transcripts for another enzyme, prostaglandin D2 synthetase (PDS), are represented by 12 ESTs in the *bx* sequence collection. This is consistent with animal studies that have shown high levels of this enzyme in iris and have implicated prostaglandin D in control of intraocular pressure [17,18], and suggests that PDS is similarly abundant in human iris. PDS is also abundant in RPE/choroid [16] and in lens [9], where it could also be contributing to prostaglandin D levels in the aqueous. Both GP3 and PDS are also expressed in ciliary body [5].

The most abundant novel cDNA discovered in the unamplified iris library ranks at equal fifth position. This codes for a protein known as oculoglycan or opticin (Optc), a member of the small leucine rich repeat proteoglycan (SLRP) superfamily [19]. This has been described in detail elsewhere [20] and has been independently observed by two other groups [21,22]. Optc is a secreted protein, found at high levels in the vitreous [21] and, like other members of its superfamily, probably has a specific binding function. However, although molecular modeling of Optc has been performed [20], no ligand has yet been identified. In yet another similarity between iris and RPE, Optc is also abundant in the RPE/choroid cDNA library [16]. The OPTC gene maps to human chromosome 1q31 in the region of two inherited retinal degenerations disorders (ARMD1 and AXPC1) [20], but no connection with either of these diseases has yet been demonstrated.

The *bx* library also contains clones for a number of genes that have been positively identified as the loci of inherited eye diseases. These include myocilin, which is mutated in inherited open angle glaucoma and is thought to play a key role in the trabecular meshwork [23,24], TIMP3, which is mutated in Sorsby fundus dystrophy [25], and CYP1B1, the locus of infantile glaucoma 3 [26]. Again, all three of these genes are also expressed in the RPE/choroid [16].

Pigment cell markers and a novel gene, MLSN2: As expected, markers for pigment cells are prominent in the library. The most abundant of these is Pmel17, the homolog of the locus for the mouse *silver* mutation [27,28], which is represented by eight ESTs. Other markers include the enzymes tyrosinase related protein 1 and dopachrome delta-isomerase, tyrosine related protein 2 and two antigens associated with melanoma or melanocytes, melan-A and CD63 antigen (melanoma 1 antigen).

In addition to these known genes, the *bx* sequence collection contains a novel transcript related to a gene associated with suppression of melanoma. This cDNA (bx18g11) has weak but significant similarity in predicted protein sequence

```
...PASAERPYKCAECGKSFKGSSGLRYHLRDHTGERPYQCGECGKAFKRSSLLAIHQRVH Human
      :::: * :*:: ::::::::::* ::*:::::: * :* :::::::::::*:::* 
...SVPSERPYQCTECGKAFKGSSGLRYHMRDHTGERPYKCSECPKAFKRSSLLAIHQRVH Chick
   | 305

TGLRAFTCGQCGLTFKWSSHYQYHLRLHSGERPYACGECGKAFRNTSCLRRHRHVHTGERP Human
:::::  *  *::::::::::* ::*:::* ::::: *  * ::: :::*:  *:  *:::::
TGLRAYKCPSCGLTFKWSSHLQYHLRLHTGERPYRCPDCPKAFKNTSCLGPHRQLHTGERP Chick

HACGVCGKSFA...                                                 Human
*:*  *:: :
HACPICGKAFT...                                                 Chick
         | 434
```

Figure 3. A human homolog of the chicken neural crest zinc-finger transcription factor cKr2. Predicted protein sequence of hKr2 (human), derived from clone bx06e07, compared with partial sequence of cKr2 (chicken) taken from GenBank accession number X56805. Numbering refers to chicken protein sequence. The positions of the C2H2 zinc-fingers are shown in alternating red and blue. Conserved cysteines and histidines of the zinc-fingers are indicated by asterisks (*).

Table 3. Cytoskeleton related transcripts from the combined iris libraries

Description	GenBank	UniGene	Chromosome	#
vimentin	BC000163	2064	10p13	12
actin, β	BC001301	288061	7p15-p12	9
myosin regulatory light chain 2	XM_009501	9615	20	9
tubulin α	AK026516	272897	2q	8
cofilin 1	NM_005507	180370	11q13	8
gelsolin	NM_000177	290070	9q33	7
actin, γ 1	NM_001614	14376	17q25	5
myosin, heavy polypeptide 11	AB020673	251664	16p13.13-p13.12	4
actin, α 2	NM_001613	195851	10q23.3	2
tubulin β	AK026594			2
destrin	NM_006870	82306	20	2
fibulin-4	AK000980	6059	11q13	2
plakophilin 4	XM_002743	152151	2q23-q31	2
tropomyosin 1 α	AL050179	77899	15q22.1	2
actin related protein 2/3 complex, subunit 1A	XM_004844	90370	7	1
actin related protein 2/3 complex, subunit 3	AF006086	6895	12q24	1
actin related protein 2/3 complex, subunit 5	AF006088	82425	1	1
actinin, α 4	D89980	182485	19q13	1
actin-related protein 3-b	AF023453	12887	7	1
centractin α	BC000693	153961	10	1
centractin β	NM_005735	2477	2q11.1-q11.2	1
dynactin 2 (p50)	BC000718	84153	12	1
dynactin 3 (p22) (DCTN3)	XM_011811			1
dysferlin	AF075575	143897	2p13.3-p13.1	1
flightless I homolog		83849	17p11.2	1
lamin A/C	BC000511	77886	1q21.2-q21.3	1
myosin class I, myh-1c	AJ001382	121576	2	1
myosin IB	X98507	286226	17p13.2-p13.3	1
myosin, light polypeptide 6	AK026164	77385	12	1
rhoC	L25081	179735	1p21-p13	1
rho guanine nucleotide exchange factor (GEF) 1	NM_004706	252280	19q13.13	1
spectrin b	AK023762			1
titin-cap (telethonin)	AJ000491	111110	17q12	1
tropomyosin 2 (b)	XM_005391	180266	9p13.2-p13.1	1
tropomyosin 4	BC000771	250641	19p13.1	1
villin 2 (ezrin)	AL162086	155191	6q25-q26	1
weakly similar to ANKYRIN	AK023196	99603	10	1
moesin	XM_013042	170328	Xq11.2-q12	1
talin	AB028950	18420	9p22-p21	1
catenin β 1	NM_001904	171271	3p21	1
catenin α 1	BC000385	178452	5q31	1
titin-cap (telethonin)	AJ000491	111110	17q12	3
dynactin 1	NM_004082	74617	2p13	2
myosin, light polypeptide 5, regulatory	AK025922	170482	4p16.3	2
plakophilin 4	NM_003628	152151	2q23-q31	2
Rho guanine nucleotide exchange factor (GEF) 4	NM_015320	6066	2q22	2
rho/rac guanine nucleotide exchange factor (GEF) 2	AB014551	155120	1q21-q22	2
actin related protein 2/3 complex, subunit 4		158196	3	1
actinin, a 1	M95178	119000	14q24	1
actinin, a 4	D89980	182485	19q13	1
capping protein (actin filament)	BC000728	82422	2cen-q24	1
Cdc42-interacting protein 4	AJ000414	73999	19	1
centractin a		153961	10	1
coronin 1A	XM_008114	109606	16	1
desmoplakin (DPI, DPII)	M77830	74316	6p24	1
myoferlin	XM_011919	234680	10	1
myosin IE	X98411	121555	19	1
profilin 1	J03191	75721	17p13.3	1
Rac1	M31467	173737	7	1
rho G	NM_001665	75082	11p15.5-p15.4	1
rho guanine nucleotide exchange factor (GEF) 1	NM_004706	252280	19q13.13	1
rho4 GTPase	NM_001666	3109	Xq28	1
supervillin	NM_003174	154567	10p11.2	1

Entries in black are from the un-normalized *bx* collection; those in red are from the normalized *fg* collection. Representative GenBank entries are shown, along with the appropriate Unigene, chromosomal location and the number of ESTs identified using GRIST.

with melastatin (MLSN1), a member of the family of transient receptor potential calcium channel proteins that is expressed in melanocytes and whose expression is inversely correlated with risk for metastasis in melanoma [29,30]. From the family relationship, the new transcript from iris was designated melastatin 2 (MLSN2).

The clone for MLSN2 was completely sequenced (GenBank Accession: AF325212). It is 1274 bp in length, with a 5' open reading frames (ORF) of 1019 bp (Figure 1). Compared with the original published sequence of MLSN1, MLSN2 initially appeared to be full length at the 5' end. However it now seems that both sequences are incomplete since a much longer ORF for MLSN1 has now been identified (GenBank Accession: XM_007662). At the time of writing, there are four other anonymous ESTs for MLSN2 in Unigene cluster Hs.287445, three derived from embryonic head and one from a CGAP kidney library. Comparisons of MLSN1, MLSN2, and the related transcripts in Unigene and GenBank suggest that members of the family may undergo extensive alternative splicing. Search of current builds of the human genome locates the MLSN2 gene on human chromosome 9p13, close to PAX5, and distinct from the location of MLSN1 on chromosome 15. Further analyses of MLSN2 and its expression in eye and uveal melanoma are in progress. Interestingly, a partial cDNA for MLSN2 is also present in the EST collection from adult lens [9], suggesting that the function of this gene may not be limited to pigmented tissues.

Transcription factors: a steroid response connection?: Table 2 lists the transcription factors identified so far from the combined *bx* and *fg* library datasets. At the cDNA level, the most abundant transcription factor in the *bx* dataset is transforming growth factor β-stimulated protein (TSC-22), a leucine zipper protein [31], which is represented by four ESTs. Interestingly, TSC-22 is related to a protein in *Drosophila*, called *shortsighted* or *bunched*, that is involved in eye development [32,33]. In addition, the *bx* collection also contains a single EST (bx18e04) that apparently encodes a novel relative of TSC-22, the homologue of mouse glucocorticoid induced leucine zipper protein (GILZ) [34]. This clone was completely sequenced (GenBank Accession: AF183393) and was found to contain a 2023 bp full length sequence (Figure 2).

```
                              RING Domain
MDYKSSLIQDGNPMENLEKQLICPICLEMFTKPVVILPCQHNLCRKCANDIFQAANPYWT  60

SRGSSVSMSGGRFRCPTCRHEVIMDRHGVYGLQRNLLVENIIDIYKQECSSRPLQKGSHP 120

               B-box
MCKEHEDEKINIYCLTCEVPTCSMCKVFGIHKACEVAPLQSVFQGQKTELNNCISMLVAG 180

                          Coiled-coil regions
NDRVQTIITQLEDSRRVTKENSHQVKEELSQKFDTLYAILDEKKSELLQRITQEQEKKLS 240

FIEALIQQYQEQLDKSTKLVETAIQSLDEPGGATFLLTAKQLIKSIVEASKGCQLGKTEQ 300

                                  Poly-Glu
GFENMDFFTLDLEHIADALRAIDFGTDEEEEEFIEEEDQEEEESTEGKEEGHQ        353
```

Figure 4. Iris Ring Finger (IRF/MURF-1). Predicted protein sequence is shown for human Iris Ring Finger (IRF). RING finger, B-box, coiled-coil, and poly-glutamate regions are indicated with colored text.

The gene is located on human chromosome Xq22.3 in current human genome assemblies. Both TSC-22 and human GILZ are also expressed in human RPE/choroid [16].

Two cDNAs for another steroid related transcription factor, sterol regulatory element binding transcription factor 1 (SREBF1), were also found in the *bx* library, while C/EBPδ, a transcription factor that is induced by and mediates the effects of glucocorticoids [35] is represented by two ESTs in the normalized iris (*fg*) sequence collection. Glucocorticoids and other steroids have major effects on intraocular pressure and a connection between open angle glaucoma and glucocorticoids has been suggested [36,37], so the presence in iris of transcription factors responsive to these agents could have clinical significance. SREBF1 is also found in lens [9], but interestingly is not represented in the large collection of ESTs from RPE/choroid [16]. This could reflect an anterior segment role for this gene and mark a distinction between iris and RPE. Other steroid related transcription regulators found in iris include glucocorticoid receptor repression factor 1 (GRF-1) and sterol regulatory element binding protein (Table 2).

Some of the transcription-related clones may reflect the neural ectoderm and neural crest origins of different populations of iris cells. One novel EST, bx06e07, contains a partial ORF that is 79% identical in protein sequence to cKr2, a C2H2 zinc-finger transcription factor from chicken that is associated with developing neural crest cells [38] (Figure 3). The similarity is sufficient to suggest that bx06e07 represents the human homolog of the chicken gene (hKr2). Given the neural crest origins of IPE cells, this gene clearly merits further study. The gene for hKr2 can be detected in human genome sequence on chromosome 19q13.42. Interestingly, a retinitis pigmentosa locus, RP11, is also mapped to 19q13.4 [39].

Another clone that may encode a neural lineage transcription factor is found in the normalized iris *fg* sequence collection. Clone fg10a12 contains an ORF which has 98% identity in protein sequence with HES-5, a basic region helix-loop-helix (bHLH) transcription factor that shows preferred expression in neural cell types in the developing mouse [40] and appears to be part of the Notch signaling pathway [41]. It seems likely that fg10a12 is a full length clone for human HES-5.

Cytoskeleton and muscle: Many cell types in the eye have highly organized structures and are, consequently, highly dependent on cytoskeletal proteins. Iris, in addition, is a contractile tissue and contains muscle cells. As shown in Table 3, the iris expresses a wide repertoire of cytoskeleton and muscle related proteins, including regulatory factors (such as GTPases), filament and muscle fiber proteins (actins, myosins, tropomyosins etc) and proteins that connect cytoskeleton to junctions and other cell components (such as catenins, coronin, talin, etc.).

IRF: a novel ring finger protein: During the analysis of potentially novel clones from the *bx*collection, one (bx04e11) was found to contain an ORF with significant similarity to midline cerebellar isoform protein 1 (MID1), the locus of Optiz syndrome [42]. MID1 is a RING finger protein with an important role in midline development and has been shown to be associated with microtubules [43]. The new MID1 related

cDNA was completely sequenced and given the name Iris RING Finger (IRF; GenBank accession number AF353673). A sequence search of the human genome locates this gene on chromosome 1p36.11, divided into eight exons. Analysis of the predicted protein sequence of IRF for conserved motifs shows, as expected, that it contains an N-terminal RING finger domain (a C3HC4 type zinc finger), a B-box zinc finger/coiled-coil (BBCC) region, which is commonly associated with the RING domain [44] (Figure 4). It also contains a C-terminal glutamate rich region of unknown function. RING fingers and BBCC domains are thought be involved in protein interactions and are found in a variety of proteins, including several transcription factors [44]. While this work was in preparation, MURF-1, a protein identical to IRF, was independently described and shown to bind the giant myofibrillar protein titin [45], suggesting that IRF/MURF-1 may have a role in the contractile properties of the iris.

GTP binding proteins: Ris, a novel Ras-like cDNA: Small GTP binding proteins, including Ras, Rho, and other families, act as switches in many of the most important cellular processes, including signal transduction and organization of intracellular structures [46]. These proteins in turn are regulated by guanine exchange factors (GEF) that catalyze the removal of GDP and the regeneration of the active GTPase [47]. Both classes of protein are represented in the iris cDNA collection, and some of those that are associated with actin assembly are listed in Table 3. One EST (bx05f09) contains a complete ORF for a novel member of the Ras family (Pfam accession number 00071). From its origin in iris, this new family member has been given the name Ris. The clone has been completely sequenced (Figure 5). It contains a cDNA insert of 2.5 kbp with an ORF of 801 bp and a long 3' UTR of over 1.5 kbp.

RH mapping was used to map Ris in the human genome.

A

```
GTCCGGGGCCGCGCGCTCCAGGGAAGCCTGCATCCCGCTCCCCTGGGCCCCGACGGGCGG   60
                                                          M     1
GCGCACTGCGCGGGGCCCACCGGCCGCAGACCTGCGTGGCCGTCCAGGGCGTCGCCATGT  120
 S  S  V  F  G  K  P  R  A  G  S  G  P  Q  S  A  P  L  E  V    21
CCTCGGTGTTTGGAAAACCCCGCGCGGGCAGCGGGCCTCAGAGCGCGCCCCTCGAGGTCA  180
N  L  A  I  L  G  R  R  G  A  G  K  S  A  L  T  V  K  F  L    41
 ACCTGGCCATCCTGGGGCGCCGCGGGGCTGGCAAGTCTGCCCTGACCGTGAAGTTTCTGA  240
T  K  R  F  I  S  E  Y  D  P  N  L  E  D  T  Y  S  S  E  E    61
CCAAGAGGTTTATCAGTGAATATGACCCCAACTTGGAGGACACCTACAGCTCCGAGGAGA  300
T  V  D  H  Q  P  V  H  L  R  V  M  D  T  A  D  L  D  T  P    81
CTGTGGACCACCAGCCTGTCCACCTGAGGGTCATGGACACTGCAGACCTGGACACCCCCA  360
R  N  C  E  R  Y  L  N  W  A  H  A  F  L  V  V  Y  S  V  D   101
GGAACTGCGAGCGCTACCTGAACTGGGCCCATGCCTTCCTGGTGGTGTACAGCGTCGACA  420
S  R  Q  S  F  D  S  S  S  S  Y  L  E  L  L  A  L  H  A  K   121
GCCGCCAGAGCTTTGATAGCAGCAGCAGCTACCTGGAGCTGCTTGCCTTGCACGCGAAGG  480
E  T  Q  R  S  I  P  A  L  L  L  G  N  K  L  D  M  A  Q  Y   141
AGACACAGCGCAGCATCCCTGCCCTGCTGCTGGGCAACAAGCTGGACATGGCTCAGTACA  540
R  Q  V  T  K  A  E  G  V  A  L  A  G  R  F  G  C  L  F  F   161
GGCAAGTCACCAAGGCAGAGGGTGTGGCTTTGGCAGGCAGGTTTGGGTGCCTGTTTTTCG  600
E  V  S  A  C  L  D  F  E  H  V  Q  H  V  F  H  E  A  V  R   181
AGGTCTCTGCCTGTCTGGACTTTGAGCACGTGCAGCATGTCTTCCACGAGGCAGTGCGAG  660
E  A  R  R  E  L  E  K  S  P  L  T  R  P  L  F  I  S  E  E   201
AGGCACGGCGGGAGCTGGAGAAGAGCCCCCTGACCCGGCCCCTCTTCATCTCCGAGGAGA  720
R  A  L  P  H  Q  A  P  L  T  A  R  H  G  L  A  S  C  T  F   221
GGGCCCTGCCCCACCAGGCCCCGCTCACCGCGCGGCATGGGCTGGCCAGCTGCACCTTCA  780
N  T  L  S  T  I  N  L  K  E  M  P  T  V  A  Q  A  K  L  V   241
ACACGCTCTCCACCATCAACCTGAAGGAGATGCCCACTGTGGCCCAGGCCAAGCTGGTCA  840
T  V  K  S  S  R  A  Q  S  K  R  K  A  P  T  L  T  L  L  K   261
CCGTGAAGTCATCCCGGGCCCAGAGCAAGCGCAAGGCGCCTACCCTGACTCTCCTGAAGG  900
G  F  K  I  F  *                                             266
GCTTCAAGATCTTCTGAGGCCCCCTCCCCAGGAAGCCTAGGCTCGGTGGCTGGACAGGAC  960
TGCAGCAGGACAGGGACTGGCTTCTCACCACCAGCCTTTCCGTCTGATGGACAGCAGACC 1020
CCGCCTCCAGCACCAAGCAGTGTCCCTACACTCAATTCACTGTGTGAGCTGGAGTGGCAG 1080
GCAGGGATGCTGCTTCTGTTCCCTCCAGGCCTTGGTCTTCATGGTAACCACCGCGTCTGT 1140
GGCCGTAGAGGGAAACAGCCACTGTGACAACCCAGAGACCTGGGGTCAGCCTTAAGTAGG 1200
AAGAAATTGCAGTGTCTATCCTGCTGCCATCAGCCCTTCCTGCTCCAGCTACAGTTGGGC 1260
CAGCTGTGGCCTTAGTGGATGGTCTGGGTGGATGGCTCTGTCAATGTCAGGCCCGAGAAC 1320
AGCCTGCGGAATGACAGCCTGGGGTCCTTGTGCCACCTGGTGGACGGGAAAGGCAGGCGG 1380
CTGCAGTTTCAGAGGCCAGAGCCTGGCCCTCCTGTCATCTGCTGACTTCACAGAAGGCCC 1440
AGGCTACAAACCCCACAGTTTATCAAGGAAAATCTAGTTAAAATAATCCAATGGACAGAA 1500
GCCCCAGGAAGGGGTATTTCCAACACTTGGCGTCTCAATGAGGATAAAAGGGAGGGACTG 1560
TAGGTTGTCCAGCATAAGGAGCCCCAGAGTTCTTTCACCAGCAGACAGATTCTCCATGTG 1620
TTTTCCCACTGTTAAGGACTCTGTCCAAAGAACCTCAGCATCAGCTCCCTCAAGCTTAAT 1680
GATCCAGGCTGCAGAAGAAAGTTGTCACTCCAGTGAGTTACCCAGGGCTCCAGGGAGCCC 1740
ACAGCAGCATCAGACCCATCCAGGGGCTACAGGGCCTGCCCCCCTGCAGGTAGAAACACA 1800
GTATACCCGTTTGTCATCTGGTTTGGCAGGACCCAGGACCCTCAGCAATGAGTCTGAGTG 1860
TTCAAAGCCCACTTTAGAACCTAAAGCATTTTGGGTTAATAAAGCAAAATTTTGCTTATG 1920
TTATTTGGTGGTGTCACCGCAGCTTATACCAAGACACCATGGCTCAGACATGGGATCAGG 1980
ACACACCTTTGAGATCCCACTGTTAGCAAGAAAAGGCTGTCCAAGGCAGGCAGTGTGGTA 2040
GTTGTCTGATTTTGTTTACCACACGCCAGGTCCCAGGGAAGTCTTGCACAGACTCTTAGC 2100
AGACAAGCAGGGAAAAGAGAGGGTCCTGGTGGGTGGAAGGATTGGGAAGGGGATTCTGGG 2160
CCTGGAAGCCTGCTCCTTCCCCTCTCAGCCGTCTCAGGACTACTGTTGGATTGCAAACAG 2220
CATCTCTGACCCAGTCTTCGGACTCATTCTGTAAATGGAGGAGCTGAGGTCAGAGGGGGG 2280
CTGTGAACTACTCAAGACCACATGCTTGCAGTCCTACAGGCTTCCCGACCCACAGGCCAG 2340
CTACGGATTTTGCACTTTTGCTCCAGGCCACTTGTGTGTGTATGTGTGTGTGTACCACTG 2400
CCATCTGAACAAACTATGCAATACCACCTGCTCCCTTCCAGGAGCATTTTAATAAAAGAT 2460
TTTTTTTAATAACAGAAAAAAAAAAAAAAAAAAAAAAAAA                    2500
```

B

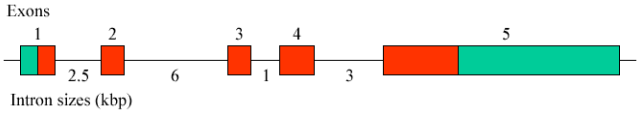


Figure 5. Sequence of human Ris. **A**: The complete sequence of clone bx05f09. The predicted amino acid sequence is shown. The GTP binding motif is shown in red and the rest of the Ras-family signature in blue. The polyadenylation signal is shown in green. **B**: Cartoon of Ris gene structure. Exons are boxed and coding sequence is shown in red.

```
GCTGCTCCCGACGCGGAGCCCGGAGCCCGCGCCGAGCCCCTGGCCTCGCGGTGCCATGCT   60
                        M  A  T  P  L  P  P  P  S  P  R  H  L   13
GCCCCGGCGGCGGCGCTGAAGGATGGCGACGCCGCTGCCTCCGCCCTCCCCGCGGCACCT  120
  R  L  L  R  L  L  L  S  G  L  V  L  G  A  A  L  R  G  A  A   33
GCGGCTGCTGCGGCTGCTGCTCTCCGGCCTCGTCCTCGGCGCCGCCCTGCGTGGAGCCGC  180
  A  G  H  P  D  V  A  A  C  P  G  S  L  D  C  A  L  K  R  R   53
CGCCGGCCACCCGGATGTAGCCGCCTGTCCCGGGAGCCTGGACTGTGCCCTGAAGAGGCG  240
  A  R  C  P  P  G  A  H  A  C  G  P  C  L  Q  P  F  Q  E  D   73
GGCAAGGTGTCCTCCTGGTGCACATGCCTGTGGGCCCTGCCTTCAGCCCTTCCAGGAGGA  300
  Q  Q  G  L  C  V  P  R  M  R  R  P  P  G  G  G  R  P  Q  P   93
CCAGCAAGGGCTCTGTGTGCCCAGGATGCGCCGGCCTCCAGGCGGGGGCCGGCCCCAGCC  360
  R  L  E  D  E  I  D  F  L  A  Q  E  L  A  R  K  E  S  G  Q  113
CAGACTGGAAGATGAGATTGACTTCCTGGCCCAGGAGCTTGCCCGGAAGGAGTCTGGACA  420
  S  T  P  P  L  P  K  D  R  Q  R  L  P  E  P  A  T  L  G  F  133
ATCAACTCCGCCCCTACCCAAGGACCGACAGCGGCTCCCGGAGCCTGCCACCCTGGGCTT  480
  S  A  R  G  Q  G  L  E  L  G  L  P  S  T  P  G  T  P  T  P  153
CTCGGCACGGGGGCAGGGGCTGGAGCTGGGCCTCCCCTCCACTCCAGGAACCCCCACGCC  540
  T  P  H  T  S  L  G  S  P  V  S  S  D  P  V  H  M  S  P  L  173
CACGCCCCACACCTCCCTGGGCTCCCCTGTGTCATCCGACCCGGTGCACATGTCGCCCCT  600
  E  P  R  G  G  Q  G  D  G  L  A  L  V  L  I  L  A  F  C  V  193
GGAGCCCCGGGGAGGGCAAGGCGACGGCCTCGCCCTTGTGCTGATCCTGGCGTTCTGTGT  660
  A  G  A  A  A  L  S  V  A  S  L  C  W  C  R  L  Q  R  E  I  213
GGCCGGTGCAGCCGCCCTCTCCGTAGCCTCCCTCTGCTGGTGCAGGCTGCAGCGTGAGAT  720
  R  L  T  Q  K  A  D  Y  A  T  A  K  A  P  G  S  P  A  A  P  233
CCGCCTGACTCAGAAGGCCGACTACGCCACTGCGAAGGCCCCTGGCTCACCTGCAGCTCC  780
  R  I  S  P  G  D  Q  R  L  A  Q  S  A  E  M  Y  H  Y  Q  H  253
CCGGATCTCGCCTGGGGACCAGCGGCTGGCACAGAGCGCGGAGATGTACCACTACCAGCA  840
  Q  R  Q  Q  M  L  C  L  E  R  H  K  E  P  P  K  E  L  D  T  273
CCAACGGCAACAGATGCTGTGCCTGGAGCGGCATAAAGAGCCACCCAAGGAGCTGGACAC  900
  A  S  S  D  E  E  N  E  D  G  D  F  T  V  Y  E  C  P  G  L  293
GGCCTCCTCGGATGAGGAGAATGAGGACGGAGACTTCACGGTGTACGAGTGCCCGGGCCT  960
  A  P  T  G  E  M  E  V  R  N  P  L  F  D  H  A  A  L  S  A  313
GGCCCCGACCGGGGAAATGGAGGTGCGCAACCCTCTGTTCGACCACGCCGCACTGTCCGC 1020
  P  L  P  A  P  S  S  P  P  A  L  P  *                      325
GCCCCTGCCGGCCCCCAGCTCACCGCCTGCACTGCCATGACCTGGAGGCAGACAGACGCC 1080
CACCTGCTCCCCGACCTCGAGGCCCCCGGGGAGGGGCAGGGCCTGGAGCTTCCCACTAAA 1140
AACATGTTTTGATGCTGTGTGCTTTTGGCTGGGCCTCGGGCTCCAGGCCCTGGGACCCCT 1200
TGCCAGGGAGACCCCCGAACCTTTGTGCCAGGACACCTCCTGGTCCCCTGCACCTCTCCT 1260
GTTCGGTTTAGACCCCCAAACTGGAGGGGGCATGGAGAACCGTAGAGCGCAGGAACGGGT 1320
GGGTAATTCTAGAGACAAAAGCCAATTAAAGTCCATTTCAGAAAAAAAAAAAAAAA     1376
```

Figure 6. Human NPDC-1-like protein. Complete sequence of clone bx14e01 (GenBank accession number NM_015392). The predicted signal peptide is shown in red and the predicted transmembrane helix in blue. The NCH domain region of similarity with nematode and insect proteins is in green. The polyadenylation signal is shown is shown in purple.

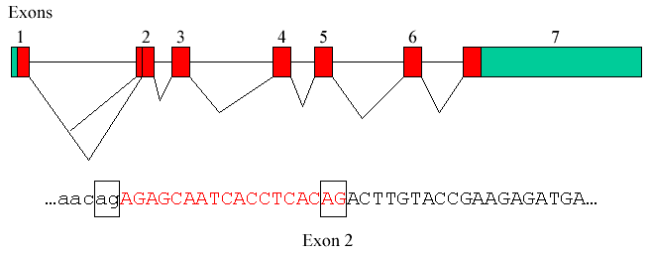
Using the Stanford G3 panel, this showed tight linkage to locus SHGC-33583 (LOD score 16), corresponding to human chromosome 15q22.3. Subsequently, human genome sequence data confirmed this location and shows that the gene is divided into five exons, with the 2 kbp fifth exon accounting for most of the transcribed sequence. ESTs for Ris, originating from several different cDNA libraries, are found in Unigene cluster Hs.27018. Several of these ESTs come from brain and kidney, while most others are derived from tumor cell libraries, particularly from prostate. Considering the apparent expression pattern of Ris, it is interesting that the location of the gene for Ris is similar to that of Bardet-Biedl syndrome (BBS) type 4 (OMIM: 600374), a condition that involves pigmentary retinopathy of the eye in addition to mental retardation and renal failure.

A novel protein potentially related to another part of the small GTP binding protein switch system was also identified. Clone bx14e01 contains an ORF closely related to the mouse

A

```
GGGGGCGAGAGCGCAGGGCGGCGCGGCGTCGGTCCCGGGAGCAGAACCCGGCTTTTTCTT   60
                          M  H  R  L  I  F  V  Y  T              9
GGAGCGACGCTGTCTCTAGTCGCTGATCCCAAATGCACCGGCTCATCTTTGTCTACACTC  120
L  I  C  A  N  F  C  S  C  R  D  T  S  A  T  P  Q  S  A  S     29
TAATCTGCGCAAACTTTTGCAGCTGTCGGGACACTTCTGCAACCCCGCAGAGCGCATCCA  180
I  K  A  L  R  N  A  N  L  R  R  D  E  S  N  H  L  T  D  L     49
TCAAAGCTTGCGCAACGCCAACCTCAGGCGAGATGagagcaatcacctcacagACTTGT  240
Y  R  R  D  E  T  I  Q  V  K  G  N  G  Y  V  Q  S  P  R  F     69
ACCGAAGAGATGAGACCATCCAGGTGAAAGGAAACGGCTACGTGCAGAGTCCTAGATTCC  300
P  N  S  Y  P  R  N  L  L  L  T  W  R  L  H  S  Q  E  N  T     89
CGAACAGCTACCCCAGGAACCTGCTCCTGACATGGCGGCTTCACTCTCAGGAGAATACAC  360
R  I  Q  L  V  F  D  N  Q  F  G  L  E  E  A  E  N  D  I  C    109
GGATACAGCTAGTGTTTGACAATCAGTTTGGATTAGAGGAAGCAGAAAATGATATCTGTA  420
R  Y  D  F  V  E  V  E  D  I  S  E  T  S  T  I  I  R  G  R    129
GGTATGATTTTGTGGAAGTTGAAGATATATCCGAAACCAGTACCATTATTAGAGGACGAT  480
W  C  G  H  K  E  V  P  P  R  I  K  S  R  T  N  Q  I  K  I    149
GGTGTGGACACAAGGAAGTTCCTCCAAGGATAAAATCAAGAACGAACCAAATTAAAATCA  540
T  F  K  S  D  D  Y  F  V  A  K  P  G  F  K  I  Y  Y  S  L    169
CATTCAAGTCCGATGACTACTTTGTGGCTAAACCTGGATTCAAGATTTATTATTCTTTGC  600
L  E  D  F  Q  P  A  A  A  S  E  T  N  W  E  S  V  T  S  S    189
TGGAAGATTTCCAACCCGCAGCAGCTTCAGAGACCAACTGGGAATCTGTCACAAGCTCTA  660
I  S  G  V  S  Y  N  S  P  S  V  T  D  P  T  L  I  A  D  A    209
TTTCAGGGGTATCCTATAACTCTCCATCAGTAACGGATCCCACTCTGATTGCGGATGCTC  720
L  D  K  K  I  A  E  F  D  T  V  E  D  L  L  K  Y  F  N  P    229
TGGACAAAAAAATTGCAGAATTTGATACAGTGGAAGATCTGCTCAAGTACTTCAATCCAG  780
E  S  W  Q  E  D  L  E  N  M  Y  L  D  T  P  R  Y  R  G  R    249
AGTCATGGCAAGAAGATCTTGAGAATATGTATCTGGACACCCCTCGGTATCGAGGCAGGT  840
S  Y  H  D  R  K  S  K  V  D  L  D  R  L  N  D  D  A  K  R    269
CATACCATGACCGGAAGTCAAAAGTTGACCTGGATAGGCTCAATGATGATGCCAAGCGTT  900
Y  S  C  T  P  R  N  Y  S  V  N  I  R  E  E  L  K  L  A  N    289
ACAGTTGCACTCCCAGGAATTACTCGGTCAATATAAGAGAAGAGCTGAAGTTGGCCAATG  960
V  V  F  F  P  R  C  L  L  V  Q  R  C  G  G  N  C  G  C  G    309
TGGTCTTCTTTCCACGTTGCCTCCTCGTGCAGCGCTGTGGAGGAAATTGTGGCTGTGGAA 1020
T  V  N  W  R  S  C  T  C  N  S  G  K  T  V  K  K  Y  H  E    329
CTGTCAACTGGAGGTCCTGCACATGCAATTCAGGGAAAACCGTGAAAAAGTATCATGAGG 1080
V  L  Q  F  E  P  G  H  I  K  R  R  G  R  A  K  T  M  A  L    349
TATTACAGTTTGAGCCTGGCCACATCAAGAGGAGGGGTAGAGCTAAGACCATGGCTCTAG 1140
V  D  I  Q  L  D  H  H  E  R  C  D  C  I  C  S  S  R  P  P    369
TTGACATCCAGTTGGATCACCATGAACGATGTGATTGTATCTGCAGCTCAAGACCACCTC 1200
R  *                                                           370
GATAAGAGAATGTGCACATCCTTACATTAAGCCTGAAAGAACCTTTAGTTTAAGGAGGGT 1260
GAGATAAGAGACCCTTTTCCTACCAGCAACCAAACTTACTACTAGCCTGCAATGCAATGA 1320
ACACAAGTGGTTGCTGAGTCTCAGCCTTGCTTTGTTAAAGCCATGGCAAGTAGAAAGGTA 1380
TATCATCAACTTCTATACCTAAGAATATAGGATTGCATTTAATAATAGTGTTTGAGGTTA 1440
TATATGCACAAACACACACAGAAATATATTCATGTCTATGTGTATATAGATCAAATGTTT 1500
TTTTTGGTATATATACCCAGGTACACCAGAGCTTACATATGTTTGAGTTAGACTCTTAAA 1560
ATCCTTTGCCAAAATAAGGGATGGTCAAATATATGAAACATGTCTTTAGAAAATTTAGGA 1620
GATAAATTTATTTTTAAATTTTGAAACACAAAACAATTTTGAATCTTGCTCTCTTAAAGA 1680
AAGCATCTTGTATATTAAAAATCAAAAGATGAGGCTTTCTTACATATACATCTTAGTTGA 1740
TTATTAAAAAAGGAAAAATATGGTTTCCAGAGAAAAGGCCAATACCTAAGCATTTTTTCC 1800
ATGAGAAACACTGCATACTTACCTATGTGGACTATAATAACCTGTCTCCAAAACCATGCC 1860
ATAATAATATAAGTGCTTTAGAAATTAAATCATTGTGTTTTTTATGCATTTTGCTGAGGC 1920
ATGCTTATTCATTTAACACCTATCTCAAAAACTTACTTAGAAGGTTTTTTATTATAGTCC 1980
TACAAAAGACAATGTATAAGCTGTAACAGAATTTTGAATTGTTTTTCTTTGCAAAACCCC 2040
TCCACAAAAGCAAATCCTTTCAAGAATGGCATGGGCATTCTGTATGAACCTTTCCAGATG 2100
GTGTTCAGTGAAAGATGTGGGTAGTTGAGAACTTAAAAAGTGAACATTGAAACATCGACG 2160
TAACTGGAAATTAGGTGGGATATTTGATAGGATCCATATCTAATAATGGATTCGAACTCT 2220
CCAAACTACACCAATTAATTTAATGTATCTTGCTTTTGTGTTCCCGTCTTTTTGAAATAT 2280
AGACATGGATTTATAATGGCATTTTATATTTGGCAGGCCATCATAGATTATTTACAACCT 2340
AAAAGCTTTTGTGTATCAAAAAAATCACATTTTATTAATGTAAATTTCTAATCGTATACT 2400
TGCTCACTGTTCTGATTTCCTGTTTCTGAACCAAGTAAAATCAGTCCTAGAGGCTATGGT 2460
TCTTAATCTATGGAGCTTGCTTTAAGAAGCCAGTTGTCAATTGTGGTAACACAAGTTTGG 2520
CCCTGCTGTCCTACTGTTAAATAGAAAACTGTTTTACATTGGTTAATGGTATTTAGAGTA 2580
ATTTTTTCTCTCTGCCTCCTTTGTGTCTGTTTTAAAGGAGACTAACTCCAGGAGTAGGAA 2640
ATGATTCATCATCCTCCAAAGCAAGAGGCTTAAGAGAGAAACACCGAAATTCAGATAGCT 2700
CAGGGACTGCTAACAGAGAACTACATTTTTCTTATTGCCTTGAAAGTTAAAAGGAAAGCA 2760
GATTTCTTCAGTGACTTTGTGGTCCTACTAACTACAACCAGTTTGGGTGACAGGGCTGGT 2820
AAAGTCCCAGTGTTAGATGAGTGACCTAAATATACTTAGATTTCTAAGTATGGTGCTCTC 2880
AGGTCCAAGTTCAACTATTCTTAAGCAGTGCAATTCTTCCCAGTTATTTGAGATGAAAGA 2940
TCTCTGCTTATTGAAGATGTACCTTCTAAAACTTTCCTAAAAGTGTCTGATGTTTTTACT 3000
CAAGAGGGGAGTGGTAAAATTAAATACTCTATTGTTCAATTCTCTAAAATCCCAGAACAC 3060
AATCAGAAATAGCTCAGGCAGACACTAATAATTAAGAACGCTCTTCCTCTTCATAACTGC 3120
TTTGCAAGTTTCCTGTGAAAACATCAGTTTCCTGTACCAAAGTCAAAATGAACGTTACAT 3180
CACTCTAACCTGAACAGCTCACAATGTAGCTGTAAATATAAAAAATGAGAGTGTTCTACC 3240
CAGTTTTCAATAAACCTTCCAGGCTGCAATAACCAGCAAGGTTTTCAGTTAAAGCCCTAT 3300
CTGCACTTTTTATTTATTAGCTGAAATGTAAGCAGGCATATTCACTCACTTTTCTTTGCC 3360
TTTCCTGAGAGTTTTATTAAAACTTCTCCCTTGGTTACCTGTTATCTTTTGCACTTCTAA 3420
CATGTAGCCAATAAATCTATTTGATAGCCATCAAAGGAATAAAAAGCTGGCCGTACAAAT 3480
TACATTTCAAAACAAACCCTAATAAATCCACATTTCCGCATGGCTCATTCACCTGGAATA 3540
ATGCCTTTTATTGAATATGTTCTTATAGGGCAAAACACTTTCATAAGTACAGTTTTTTAT 3600
GTTTTTTGTCATATCGGTAACATGCAGCTTTTTCCTCTCATAGCATTTTCTATAGCGAAT 3660
GTAATATGCCTCTTATCTTCATGAAAAATAAATATTGCTTTTGGCCAAAAAAAAAAAAAA 3720
AAAAAAAAA                                                      3729
```

B



C

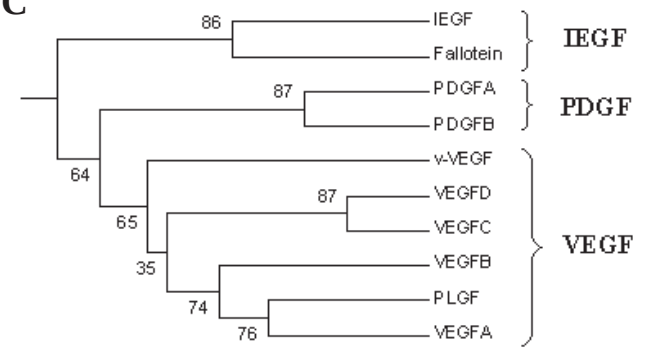


Figure 7. Iris expressed growth factor (IEGF). A: cDNA and predicted protein sequence. Insert sequence for "long form" is shown in red. The CUB domain is shown in green and the PDGF/VEGF domain in blue. The polyadenylation signal is shown in purple. B: Alternative splicing of IEGF. Above: Cartoon of IEGF gene structure with exons shown as boxes, colored red for coding sequence and green for noncoding. Two possible splice events between exons 1 and 2 are indicated. Below: Sequence of the 5' end of exon 2. Intron sequence (taken from human genome sequence) is in lower case and two alternative splice junctions are boxed. C: A cladogram of the PDGF/VEGF domain of IEGF compared with those of some closely related proteins. The tree was drawn using the program MEGA2. Gaps and missing data were handled by pair-wise deletion, Poisson correction was applied and the UPGMA option was used. 500 iterations were used to calculate bootstrap values (shown on each branchpoint) as a measure of confidence for each branch. The related sequences were aligned with the NBCI Conserved Domain browser. The GenBank accession numbers for the following molecules are listed in parentheses: IEGF (AY027517); PDGF-A (P20033); PDGF-B (1PDGA); fallotein (AF091434); VEGFA (1FLTW); VEGFB (1718152); VEGFC (1718154); VEGFD (c-fos induced growth factor; 6753874); PLGF (human placental growth factor; 1709656).

neural protein NPDC-1 (neural proliferation, differentiation, and control gene 1) [48]. The complete insert of the human NPDC-1 clone is 1376 bp in length and contains an ORF of 978 bp encoding a protein of 34 kDa (Figure 6). Human NPDC-1 protein contains putative signal sequence and transmembrane regions (Figure 6). It also contains an NCH domain, a conserved structure common to CAB-1, a protein in *C. elegans* that regulates activity in neural cells through interaction with a GEF protein, and to an unnamed *Drosophila* protein of unknown function [49]. It is possible that proteins of this class provide a link between GTPase switch pathways and other cellular processes [49]. The NPDC-1 gene can be found in the human genome assemblies on chromosome 9q34.3, divided into nine exons.

Growth factors: Growth factors in the iris and ciliary body have the potential to influence the development and health of other tissues in the anterior compartment of the eye. For example, the growth of the eye lens is broadly under the control of two growth factor related mechanisms [2]. Factors in the vitreous, probably secreted from the retina, control the differentiation of lens epithelial cells into elongated fiber cells. Members of the FGF family and other factors have been implicated in this process [2]. A second mechanism controls the proliferation and migration of lens epithelial cells. This is thought to involve a factor, so far unidentified, that is expressed in iris/ciliary body and secreted in the anterior chamber. Growth factors in the anterior segment of the eye may also influence wound healing, angiogenesis, and other important processes in surrounding tissues, including the cornea.

Vascular endothelial growth factor (VEGF), platelet derived growth factor, and related proteins, form an important superfamily of mitogenic factors with specific effects on different cell types. The best understood role for VEGF is in angiogenesis, a critical process in the eye, where disease, injury, or age related neovascularizations of retina, choroid, and cornea are major causes of blindness [50,51]. Two members of the VEGF family, VEGF-A and VEGF-B, are represented in the *bx* and *fg* sequence collections (Table 4), as is an inhibitor of VEGF related angiogenesis, pigment epithelium derived factor [52]. No cDNA sequences for PDGF proteins are in the current collection, although PDGF receptor polypeptides are represented.

IEGF, a novel member of the VEGF/PDGF family: The 5' sequence of one novel EST from the human iris library (clone bx02h05) reveals a weak but significant match in predicted protein sequence to proteins of the VEGF family, particularly fallotein or PDGF-C, a protein of unknown function cloned from ovary [53]. From its expression in iris and its sequence relationships, the new family member has been given the name iris expressed growth factor (IEGF). The complete sequence of IEGF was obtained from the EST clone and has been sequenced in its entirety and contains a long 3' UTR of 2.5 kbp but appears to be incomplete at the 5' end. When the iris cDNA was completely sequenced, its 3' UTR revealed identity with members of a Unigene cluster (Hs.112885) that contains several partial cDNA clones from various tissues, including kidney. Using RNA from human kidney as template, the complete 5' end of the IEGF transcript was determined by 5' RACE (rapid amplification of cDNA ends).

Several 5' end clones were sequenced and, when aligned, revealed an in frame alternative splice in the coding sequence. The complete cDNA sequence of IEGF-l is 3729 bp, while IEGF-s is 3711 bp (GenBank Accession numbers AY027517 and AY027518). Both long and short versions of the IEGF mRNA contain ORF differing only by in frame insertion/deletion of 18 bp or 6 codons (Figure 7A). The predicted sizes of the two versions of the protein are 42.7 kDa and 42 kDa.

Initially, the location of the IEGF gene was unknown, so PCR primers designed from the 3' UTR of bx02h05 were used for RH mapping to localize the gene to human chromosome 11q14-21, linked with a LOD score of 13.07 to marker SHGC-8453. Now the position of the gene can be confirmed by human genome project sequence. The IEGF transcript is divided into seven exons on chromosome 11 and it is apparent that the two mRNA splice variants arise from use of alternative 5' splice sites for exon 2 (Figure 7B). After this work had been completed, the ORF sequence for IEGF was independently described under the name spinal cord derived growth factor B [54] and more recently has been described as PDGF-D [55,56].

The predicted amino acid sequence for IEGF contains an N-terminal CUB domain (Pfam accession number 00431), and a C-terminal PDGF-like growth factor domain (Smart:PDGF; Figure 7A). The CUB domain is an immunoglobulin-like structure involved in protein interactions [57]. The small insertion in the long form of IEGF is positioned just N-terminal to the CUB domain in a region with no strong similarity to other proteins. It is unclear what the effect of the insertion might be at the protein level. Predictions of overall charge for the two versions of the protein show little difference, with both of them predicted to have pIs of about 8. With the complete sequence of IEGF in hand, its evolutionary relationships to other growth factors can be examined. Using the phylogeny program MEGA [13] to align the CUB and growth factor domains with those of related proteins, no significant clustering is apparent using a variety of CUB domains from widely diverged families (not shown), but the better conserved growth factor domain gives a robust grouping (Figure 7C). Comparing IEGF with several closely related members of the VEGF and PDGF families, a tree with three branches is obtained, with distinct VEGF and PDGF families and a third family for IEGF and fallotein.

Other growth factor related proteins: Another newly described protein, which like IEGF contains a growth factor related domain and a CUB domain, is also represented in the *bx* library. Two clones, bx03h02 and bx11a05, encode the apparent human ortholog of the mouse protein SCUBE-1, a factor which contains an epidermal growth factor domain and a CUB domain [58]. Human SCUBE-1 can be localized in human genome sequence on chromosome 22q13.31. Cat-eye syndrome (CES, OMIM: 115470) involves various eye anomalies including coloboma of the iris and aniridia, however this disorder is mapped to 22q11, which would appear to exclude SCUBE-1 as a likely candidate. The iris also contains transcripts for neurotrophin 3, a mitogen for cultured neural crest cells [59], and for two other neurotrophic factors, pleiotrophin

and midkine [60].

Comparison with ciliary body ESTs: Given the close relationship between iris and ciliary body, it might be expected that the tissues have similar patterns of gene expression. So, although the sizes of the databases are very different, it is of interest to compare the clones obtained so far from the subtracted ciliary body cDNA library [5] with those of the combined NEIBank (*byfg*) collection. Clones from the ciliary body collection (CB) have been described in several papers. They include clones coding for several plasma proteins, suggesting that such proteins in the aqueous humor may have an ocular origin [5] and clones for genes involved in neuropeptide synthesis [61]. Some of the iris expressed genes described above, including PDS, GP3, and myocilin have also been detected in CB [5,62].

Surprisingly, when the 289 CB clones are processed through GRIST with the total of over 4000 clones from *bx* and *fg*, only 66 of the CB clones (23%) group with iris clones. The remaining 223 clones are not represented in the large unnormalized and normalized collection. Few of the CB sequences are quite short and may not represent true transcripts, but most of the clones have significant matches to GenBank or Unigene sequences and are therefore identifiable genes.

The CB clones are the result of subtraction, which precludes a direct comparison with the iris data, but it is interesting to see how few of even this relatively small set match clones from the iris collection. This may mean that, in spite of the proximity and related origin of the two tissues, there could indeed be significant differences in gene expression between them. However it is more likely that we need larger datasets to allow a full comparison.

Conclusion: The iris cDNA libraries have provided a large and varied set of clones for future studies, including the construction of human eye cDNA microarrays. By itself, the EST analysis has revealed some interesting parallels with gene expression in human RPE [16] and several transcripts that are associated with neural lineages or with the contractile phenotype. Such markers may be useful for work on IPE cells that focuses on their potential for autologous cell transplant in retinal degenerative disease [3].

The sequence collection also contains a number of transcripts that have interesting implications for control of intraocular pressure, including known glaucoma genes, such as myocilin and CYP1B1, and PDS, an enzyme of prostaglandin synthesis. In addition, there are several cDNAs for enzymes and transcription factors that are involved in responses to steroids. A variety of clinical observations suggest links between glucocorticoid induced ocular hypertension and primary open angle glaucoma [37]. Glucocorticoids and other steroids have wide ranging effects, but foremost among their biological functions is their ability to regulate gene transcription to effect cellular process from inflammation to apoptosis [34,63,64]. Among the transcription factors represented in the EST sampling are several which could play a role in this kind of response, including the human homolog of mouse, GILZ (glucocorticoid induced leucine zipper) [34]. Interestingly, GILZ is related in sequence to TSC-22, a TGFβ-induced factor [31], which happens to be the most abundant transcription factor represented in the iris cDNAs. Both these factors are related to *shortsighted*/*bunched*, a *Drosophila* gene that is itself involved in eye development [32,33]. Given the surprising parallels between vertebrate and invertebrate eye developmental pathways [65], this is intriguing and worthy of further investigation.

ACKNOWLEDGEMENTS

SLB is supported by the V. Kann Rasmussen Foundation (Denmark) and is a Career Development Awardee of Research to Prevent Blindness (RPB). We thank Dr. Weinu Gan for cDNA sequencing and Ray Tabios for technical assistance.

REFERENCES

1. Chang B, Smith RS, Hawes NL, Anderson MG, Zabaleta A, Savinova O, Roderick TH, Heckenlively JR, Davisson MT, John SW. Interacting loci cause severe iris atrophy and glaucoma in DBA/2J mice. Nat Genet 1999; 21:405-9.
2. McAvoy JW, Chamberlain CG, de Iongh RU, Hales AM, Lovicu FJ. Lens development. Eye 1999; 13:425-37.
3. Thumann G. Development and cellular functions of the iris pigment epithelium. Surv Ophthalmol 2001; 45:345-54.
4. Shields JA, Shields CL, De Potter P, Singh AD. Diagnosis and treatment of uveal melanoma. Semin Oncol 1996; 23:763-7.
5. Escribano J, Ortego J, Coca-Prados M. Isolation and characterization of cell-specific cDNA clones from a subtractive library of the ocular ciliary body of a single normal human donor: transcription and synthesis of plasma proteins. J Biochem (Tokyo) 1995; 118:921-31.
6. Coca-Prados M, Escribano J, Ortego J. Differential gene expression in the human ciliary epithelium. Prog Retin Eye Res 1999; 18:403-29.
7. Wistow G. A project for ocular bioinformatics: NEIBank. Mol Vis 2002; 8:161-3.
8. Simms D. mRNA isolation for high quality cDNA. Focus 1995; 17:39-42.
9. Wistow G, Bernstein SL, Wyatt MK, Behal A, Touchman JW, Bouffard G, Smith D, Peterson K. Expressed sequence tag analysis of adult human lens for the NEIBank Project: Over 2000 non-redundant transcripts, novel genes and splice variants. Mol Vis 2002; 8:171-84.
10. Wistow G, Bernstein SL, Touchman JW, Bouffard G, Wyatt MK, Peterson K, Gao J, Buchoff P, Smith D. Grouping and identification of sequence tags (GRIST): Bioinformatics tools for the NEIBank database. Mol Vis 2002; 8:164-70.
11. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2000; 28:263-6.
12. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 2000; 28:231-4.
13. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 1994; 10:189-91.
14. Reddy VN. Glutathione and its function in the lens—an overview. Exp Eye Res 1990; 50:771-8.
15. Spector A, Ma W, Wang RR. The aqueous humor is capable of generating and degrading H2O2. Invest Ophthalmol Vis Sci 1998; 39:1188-97.
16. Wistow G, Bernstein SL, Wyatt MK, Fariss RN, Behal A,

Touchman JW, Bouffard G, Smith D, Peterson K. Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: Over 6000 non-redundant transcripts, novel genes and splice variants. Mol Vis 2002; 8:205-20.
17. Gerashchenko DY, Beuckmann CT, Marcheselli VL, Gordon WC, Kanaoka Y, Eguchi N, Urade Y, Hayaishi O, Bazan NG. Localization of lipocalin-type prostaglandin D synthase (beta-trace) in iris, ciliary body, and eye fluids. Invest Ophthalmol Vis Sci 1998; 39:198-203.
18. Goh Y, Nakajima M, Azuma I, Hayaishi O. Prostaglandin D2 reduces intraocular pressure. Br J Ophthalmol 1988; 72:461-4.
19. Iozzo RV. The biology of the small leucine-rich proteoglycans. Functional network of interactive proteins. J Biol Chem 1999; 274:18843-6.
20. Hobby P, Wyatt MK, Gan W, Bernstein S, Tomarev S, Slingsby C, Wistow G. Cloning, modeling, and chromosomal localization for a small leucine-rich repeat proteoglycan (SLRP) family member expressed in human eye. Mol Vis 2000; 6:72-8.
21. Reardon AJ, Le Goff M, Briggs MD, McLeod D, Sheehan JK, Thornton DJ, Bishop PN. Identification in vitreous and molecular cloning of opticin, a novel member of the family of leucine-rich repeat proteins of the extracellular matrix. J Biol Chem 2000; 275:2123-9.
22. Friedman JS, Ducharme R, Raymond V, Walter MA. Isolation of a novel iris-specific and leucine-rich repeat protein (oculoglycan) using differential selection. Invest Ophthalmol Vis Sci 2000; 41:2059-66.
23. Stone EM, Fingert JH, Alward WL, Nguyen TD, Polansky JR, Sunden SL, Nishimura D, Clark AF, Nystuen A, Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC. Identification of a gene that causes primary open angle glaucoma. Science 1997; 275:668-70.
24. Johnson DH. Myocilin and glaucoma: a TIGR by the tail? Arch Ophthalmol 2000; 118:974-8.
25. Weber BH, Vogt G, Pruett RC, Stohr H, Felbor U. Mutations in the tissue inhibitor of metalloproteinases-3 (TIMP3) in patients with Sorsby's fundus dystrophy. Nat Genet 1994; 8:352-6.
26. Stoilov I, Akarsu AN, Sarfarazi M. Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. Hum Mol Genet 1997; 6:641-7.
27. Adema GJ, de Boer AJ, Vogel AM, Loenen WA, Figdor CG. Molecular characterization of the melanocyte lineage-specific antigen gp100. J Biol Chem 1994; 269:20126-33.
28. Solano F, Martinez-Esparza M, Jimenez-Cervantes C, Hill SP, Lozano JA, Garcia-Borron JC. New insights on the structure of the mouse silver locus and on the function of the silver protein. Pigment Cell Res 2000; 13 Suppl 8:118-24.
29. Duncan LM, Deeds J, Hunter J, Shao J, Holmgren LM, Woolf EA, Tepper RI, Shyjan AW. Down-regulation of the novel gene melastatin correlates with potential for melanoma metastasis. Cancer Res 1998; 58:1515-20.
30. Hofmann T, Schaefer M, Schultz G, Gudermann T. Transient receptor potential channels as molecular substrates of receptor-mediated cation entry. J Mol Med 2000; 78:14-25.
31. Kester HA, Blanchetot C, den Hertog J, van der Saag PT, van der Burg B. Transforming growth factor-beta-stimulated clone-22 is a member of a family of leucine zipper proteins that can homo- and heterodimerize and has transcriptional repressor activity. J Biol Chem 1999; 274:27439-47.
32. Treisman JE, Lai ZC, Rubin GM. Shortsighted acts in the decapentaplegic pathway in Drosophila eye development and has homology to a mouse TGF-beta-responsive gene. Development 1995; 121:2835-45.
33. Dobens LL, Hsu T, Twombly V, Gelbart WM, Raftery LA, Kafatos FC. The Drosophila bunched gene is a homologue of the growth factor stimulated mammalian TSC-22 sequence and is required during oogenesis. Mech Dev 1997; 65:197-208.
34. Riccardi C, Cifone MG, Migliorati G. Glucocorticoid hormone-induced modulation of gene expression and regulation of T-cell death: role of GITR and GILZ, two dexamethasone-induced genes. Cell Death Differ 1999; 6:1182-9.
35. Yeh WC, Cao Z, Classon M, McKnight SL. Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. Genes Dev 1995; 9:168-81.
36. Wordinger RJ, Clark AF. Effects of glucocorticoids on the trabecular meshwork: towards a better understanding of glaucoma. Prog Retin Eye Res 1999; 18:629-67.
37. Fingert JH, Clark AF, Craig JE, Alward WL, Snibson GR, McLaughlin M, Tuttle L, Mackey DA, Sheffield VC, Stone EM. Evaluation of the myocilin (MYOC) glaucoma gene in monkey and human steroid-induced ocular hypertension. Invest Ophthalmol Vis Sci 2001; 42:145-52.
38. Schutz B, Niessing J. Cloning and structure of a chicken zinc finger cDNA: restricted expression in developing neural crest cells. Gene 1994; 148:227-36.
39. Al-Maghtheh M, Vithana E, Tarttelin E, Jay M, Evans K, Moore T, Bhattacharya S, Inglehearn CF. Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. Am J Hum Genet 1996; 59:864-71.
40. Takebayashi K, Akazawa C, Nakanishi S, Kageyama R. Structure and promoter analysis of the gene encoding the mouse helix-loop-helix factor HES-5. Identification of the neural precursor cell-specific promoter element. J Biol Chem 1995; 270:1342-9.
41. Kageyama R, Ohtsuka T. The Notch-Hes pathway in mammalian neural development. Cell Res 1999; 9:179-88.
42. Quaderi NA, Schweiger S, Gaudenz K, Franco B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S, Marion RW, Hennekam RC, Opitz JM, Muenke M, Ropers HH, Ballabio A. Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. Nat Genet 1997; 17:285-91.
43. Schweiger S, Foerster J, Lehmann T, Suckow V, Muller YA, Walter G, Davies T, Porter H, van Bokhoven H, Lunt PW, Traub P, Ropers HH. The Opitz syndrome gene product, MID1, associates with microtubules. Proc Natl Acad Sci U S A 1999; 96:2794-9.
44. Borden KL. RING fingers and B-boxes: zinc-binding protein-protein interaction domains. Biochem Cell Biol 1998; 76:351-8.
45. Centner T, Yano J, Kimura E, McElhinny AS, Pelin K, Witt CC, Bang ML, Trombitas K, Granzier H, Gregorio CC, Sorimachi H, Labeit S. Identification of muscle specific ring finger proteins as potential regulators of the titin kinase domain. J Mol Biol 2001; 306:717-26.
46. Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. Physiol Rev 2001; 81:153-208.
47. Cherfils J, Chardin P. GEFs: structural basis for their activation of small GTP-binding proteins. Trends Biochem Sci 1999; 24:306-11.
48. Galiana E, Vernier P, Dupont E, Evrard C, Rouget P. Identification of a neural-specific cDNA, NPDC-1, able to down-regu-

late cell proliferation and to suppress transformation. Proc Natl Acad Sci U S A 1995; 92:1560-4.
49. Iwasaki K, Toyonaga R. The Rab3 GDP/GTP exchange factor homolog AEX-3 has a dual function in synaptic transmission. EMBO J 2000; 19:4806-16.
50. Campochiaro PA. Retinal and choroidal neovascularization. J Cell Physiol 2000; 184:301-10.
51. Moromizato Y, Stechschulte S, Miyamoto K, Murata T, Tsujikawa A, Joussen AM, Adamis AP. CD18 and ICAM-1-dependent corneal neovascularization and inflammation after limbal injury. Am J Pathol 2000; 157:1277-81.
52. Dawson DW, Volpert OV, Gillis P, Crawford SE, Xu H, Benedict W, Bouck NP. Pigment epithelium-derived factor: a potent inhibitor of angiogenesis. Science 1999; 285:245-8.
53. Tsai YJ, Lee RK, Lin SP, Chen YH. Identification of a novel platelet-derived growth factor-like gene, fallotein, in the human reproductive tract. Biochim Biophys Acta 2000; 1492:196-202.
54. Hamada T, Ui-Tei K, Imaki J, Miyata Y. Molecular cloning of SCDGF-B, a novel growth factor homologous to SCDGF/PDGF-C/fallotein. Biochem Biophys Res Commun 2001; 280:733-7.
55. Bergsten E, Uutela M, Li X, Pietras K, Ostman A, Heldin CH, Alitalo K, Eriksson U. PDGF-D is a specific, protease-activated ligand for the PDGF beta-receptor. Nat Cell Biol 2001; 3:512-6.
56. LaRochelle WJ, Jeffers M, McDonald WF, Chillakuru RA, Giese NA, Lokker NA, Sullivan C, Boldog FL, Yang M, Vernet C, Burgess CE, Fernandes E, Deegler LL, Rittman B, Shimkets J, Shimkets RA, Rothberg JM, Lichenstein HS. PDGF-D, a new protease-activated growth factor. Nat Cell Biol 2001; 3:517-21.
57. Bork P, Beckmann G. The CUB domain. A widespread module in developmentally regulated proteins. J Mol Biol 1993; 231:539-45.
58. Grimmond S, Larder R, Van Hateren N, Siggers P, Hulsebos TJ, Arkell R, Greenfield A. Cloning, mapping, and expression analysis of a gene encoding a novel mammalian EGF-related protein (SCUBE1). Genomics 2000; 70:74-81.
59. Kalcheim C, Carmeli C, Rosenthal A. Neurotrophin 3 is a mitogen for cultured neural crest cells. Proc Natl Acad Sci U S A 1992; 89:1661-5.
60. Nakagawara A, Milbrandt J, Muramatsu T, Deuel TF, Zhao H, Cnaan A, Brodeur GM. Differential expression of pleiotrophin and midkine in advanced neuroblastomas. Cancer Res 1995; 55:1792-7.
61. Ortego J, Escribano J, Crabb J, Coca-Prados M. Identification of a neuropeptide and neuropeptide-processing enzymes in aqueous humor confers neuroendocrine features to the human ocular ciliary epithelium. J Neurochem 1996; 66:787-96.
62. Ortego J, Escribano J, Coca-Prados M. Cloning and characterization of subtracted cDNAs from a human ciliary body library encoding TIGR, a protein involved in juvenile open angle glaucoma with homology to myosin and olfactomedin. FEBS Lett 1997; 413:349-53.
63. Barnes PJ, Adcock I. Anti-inflammatory actions of steroids: molecular mechanisms. Trends Pharmacol Sci 1993; 14:436-41.
64. Planey SL, Litwack G. Glucocorticoid-induced apoptosis in lymphocytes. Biochem Biophys Res Commun 2000; 279:307-12.
65. Treisman JE. A conserved blueprint for the eye? Bioessays 1999; 21:843-50.

The print version of this article was created on 24 Mar 2005. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.