expression vectors of these variants were introduced into cultured cells, the corresponding 98-aa short protein was produced (unpublished data). Since this short isoform lacked a minimal transactivation domain (Friedman et al., 2004), it showed no transcriptional activity as expected. The Leu zipper domain has been reported to interact with a CRX homeodomain (Mitton et al., 2000). The short isoform carrying only the Leu zipper domain may be involved in the regulation of complex formation through this domain.

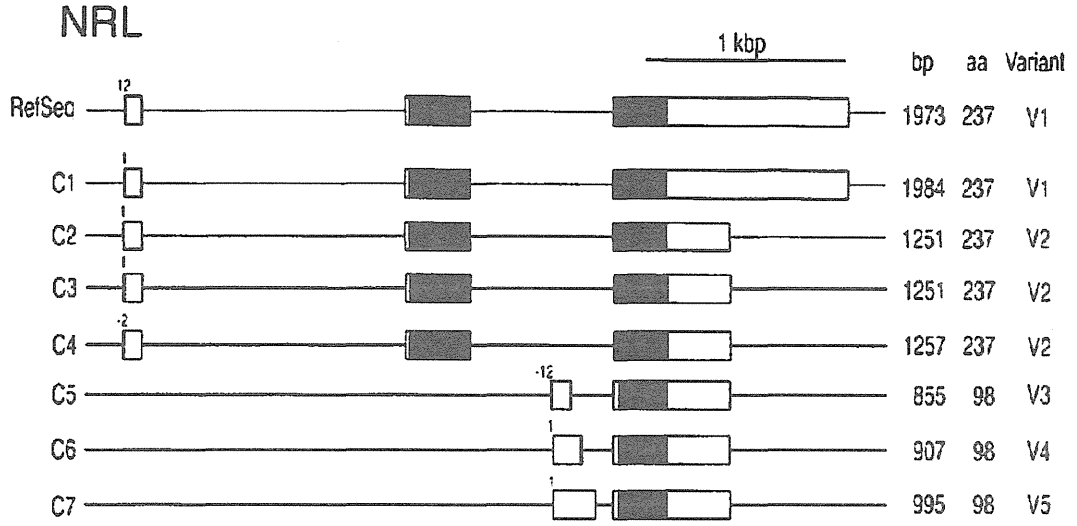


Figure 4. The exon-intron structure of alternative splicing variants for *NRL*. Seven clones (C1-C7) were classified into four variants (V1-V5). The relative position of a transcription start site was indicated on the first exon. RefSeq correspond to GenBank Accession No. NM_006177.3. Our clones correspond to AB593102.1 - AB593104.1.

GenBank contained three mRNAs (one corresponding to V1 and two to V2) except for the six sequences we registered. The dbEST contained six V4 sequences and one V5. Although one research group cloned the cDNA corresponding to V4, the authors could not judge whether it was a full-length or truncated one (Wistow et al., 2002). Like this case, when only one cDNA different from known ones is cloned, it is difficult to judge its intactness. Even in such case, our clone can be identified to be a full-length clone by the presence of an additional dG at the 5' end. Since the V1 sequence had SfiI sites, the Y79 cDNA libraries prepared using the oligo-capping method missed cloning the full-length cDNA for *NRL*.

### *3.2.4. OTX2 Antisense RNA 1 (OTX2-AS1)*

Our libraries contained many novel non-coding RNAs including rare variants. As an example of an eye-specific non-coding RNA, we obtained five clones for *OTX2-AS1* from the Y79 libraries. These clones showed a variety of structures as shown in Figure 5. The length of cDNA varied from 303 bp of V2 to 2900 bp of V3 and the splicing pattern varied from clone to clone. Only a part of exon 1 was shared within all variants. The sequences in dbEST are also rich in variety. *OTX2-AS1* is a gene transcribed in the opposite direction at the upstream region of the locus of orthodenticle homeobox 2 (OTX2) that is a transcription factor involved in the development of brain and sensory organs (Alfano et al., 2005). Our Y79 libraries also contained three clones for *OTX2*, each of which is an AS variant (data not

shown). The exon 1 of a variant of mouse *Otx2-as1* overlapped with the antisense strand of the exon 1 of *Otx2* (Alfano et al., 2005), but there was no human variant whose exon 1 overlapped to *OTX2*. Since all ESTs for *OTX2-AS1* in dbEST were obtained from retina cDNA libraries, this gene might be involved in the development of retina. The presence of diverse AS variants implies the complex regulation system by this gene.

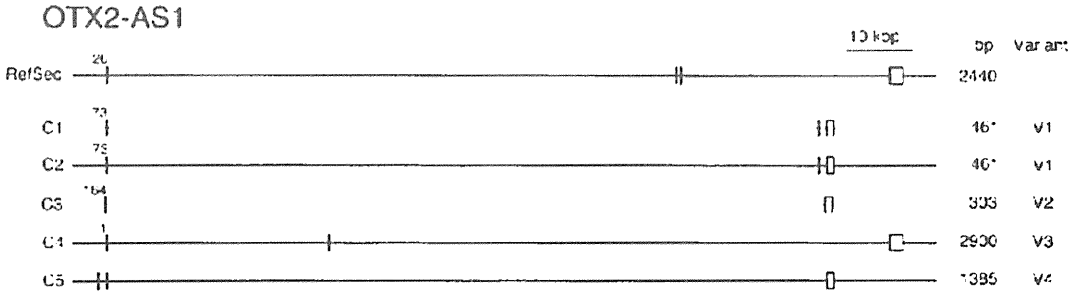


Figure 5. The exon-intron structure of alternative splicing variants for *OTX2-AS1*. Five clones (C1-C5) were classified into four variants (V1-V4). The relative position of a transcription start site was indicated on the first exon. RefSeq correspond to GenBank Accession No. NR_029385.1. Our clones correspond to AB593038.1 - AB593041.1.

## 3.3. Characterization of Long-Sized Genes

### *3.3.1. Very-Long-Sized Genes*

We succeeded to clone 82 full-length cDNAs with >7kbp from the libraries prepared using the vector-capping method (Oshikawa et al., 2008; Oshikawa et al., 2011). The ARPE-19 libraries contained full-length cDNA clones encoding golgin B1 (*GOLGB1*, 11.2kbp), NEDD4 binding protein 2 (*N4BP2*, 9.7 kbp), acetyl-CoA carboxylase alpha (*ACACA*, 9.5 kbp), filamin B, beta (*FLNB*, 8.0-9.4 kbp), filamin C, gamma (*FLNC*, 9.2 kbp), spectrin, beta, non-erythrocytic 1 (*SPTBN1*, 8.4 kbp), filamin A, alpha (*FLNA*, 8.2 kbp), collagen, type V, alpha 1 (*COL5A1*, 8.1 kbp), spectrin, alpha, non-erythrocytic 1 (*SPTAN1*, 7.8 kbp), fibronectin 1 (*FN1*, 7.8 kbp), myosin, heavy chain 9, non-muscle (*MYH9*, 7.4 kbp), and agrin (*AGRN*, 7.3 kbp). The Y79 libraries contained full-length cDNAs encoding Dmx-like 1 (*DMXL1*, 12.8 kbp), GOLGB1 (11.1 kbp), SEC16 homolog A (*SEC16A*, 9.0 kbp), FLNA (8.4 kbp), eyes shut homolog (*EYS*, 8.0 kbp). Out of these genes, four genes having multiple AS variants were selected and their structures were analyzed below.

### *3.3.2. Golgin B1 (GOLGB1)*

GOLGB1 is a huge integral membrane protein located in Golgi, originally named giantin (Linstedt et al., 1993). Two research groups cloned approximately 10-kbp cDNA encoding a protein that reacts with autoantibody contained in sera of patients with chronic rheumatism: mRNA1, 10,295-bp cDNA encoding 3,225-aa protein (Sohda et al., 1994); mRNA2, 10,300-bp cDNA encoding 3,259-aa protein (Seelig et al., 1994). These clones were not derived from a single mRNA. The full sequence was constructed by combining the sequences of cDNA fragments. Thus, it is doubtful whether the sequence reflects the true structure of the AS variant.

Our libraries contained two full-length cDNA clones for *GOLGB1* (V1 from ARPE-19, 11.2 kbp; V2 from Y79, 11.1 kbp). The exon-intron structures of the above four clones were different as shown in Figure 6. In GenBank, RefSeqs seem to be constructed by referring to registered mRNAs including our clones: RefSeq1 to V1; RefSeq2 to mRNA2; RefSeq3 to mRNA1; RefSeq4 to V2. Although exon 1 was shared within all clones, they were all different AS variants encoding the protein with the different number of aa residues. In V1 and mRNA1, the 28-bp downstream shift of the 3' splice site of exon 2 (designated by #1 in Figure 6) caused a frame shift, and thus the initiation codon in exon 3 was used. As a result, the N-terminal sequence was shortened by 39 aa compared with the isoform for V1. Furthermore, V2 lacked a 41-aa sequence corresponding to exon 7 by exon skipping. mRNA2 lacked a 5-aa sequence by the 15-bp downstream shift of the 5' splice site of exon 7 (#2). V1 had 5-aa insertion by the 15-bp downstream shift of the 3' splice site of exon 18 (#3). The dbEST contained ESTs carrying not only these four variations but also other variations including the shift of splice site or skipping of exon 4, 6, 10, 11, 12, 15-21, suggesting the presence of diverse AS variants of *GOLGB1*.

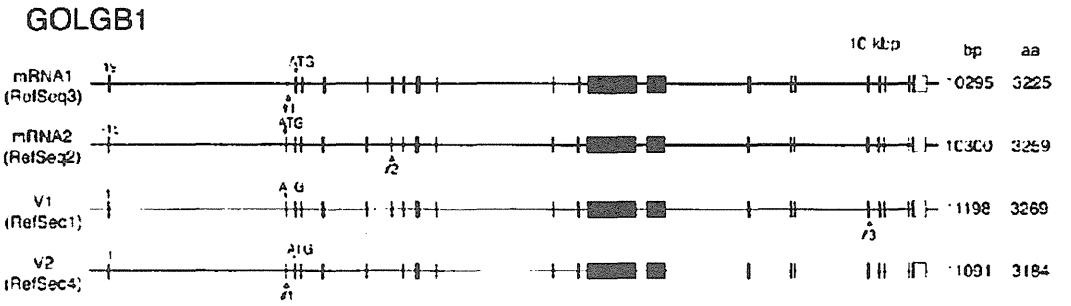


Figure 6. The exon-intron structure of alternative splicing variants for *GOLGB1*. C1 was cloned from the ARPE-19 cDNA library and C2 from the Y79 cDNA library. The relative position of a transcription start site was indicated on exon 1. Arrowheads represent the following sequence variations: #1, the downstream shift of the 3' splice site (28 bp); #2, the downstream shift of the 5' splice site (15bp); #3, the downstream shift of the 5' splice site (15 bp). mRNA1 and mRNA2 correspond to GenBank Accession No. D25542.1 and X75304.1, respectively. Our clones correspond to AB371588.1 and AB593126.1.

GOLGB1 is an integral membrane protein involved in linkage between a Golgi membrane and a COPI vesicle (Sönnichsen et al., 1998). This protein has no N-terminal secretory signal sequence, but has a C-terminal transmembrane domain. Most of the cytoplasmic part is composed of a coiled-coil structure in which the AS variants had deletion or insertion. This structure is thought to be involved in regulation of retrograde trafficking to the endoplasmic reticulum in Golgi apparatus through binding of small GTPase such as Rab6 and Rab1 (Rosing et al., 2007). Thus, each AS variant may play a role in the regulation of this trafficking. To elucidate the detailed mechanism, further investigation is necessary using these AS variants.

### *3.3.3. Filamin A, Alpha (FLNA)*

*FLNA* was most abundantly found in our libraries as a long-sized gene with >7kbp. FLNA is an actin-binding protein involved in change in cell shape and migration through crosslinking of actin filaments and linking actin filaments to membrane glycoproteins. ARPE-

19 and Y79 libraries contained eight clones (7.3 – 8.2 kbp) and one clone (8.4 kbp), respectively. These clones were classified into four variants as shown in Figure 7. V1 and V2 were main components in ARPE-19 cells. V2 lacked exon 30, resulting in deletion of an 8-aa sequence. V3 lacked exon 38-41 because of AS between the middle splice site in exon 37 and the middle splice site in exon 42. Y79-originated V4 started from a 135-bp upstream TSS compared with V1, resulting in the generation of a novel initiation codon that caused 27-aa extension of the N-terminal sequence.

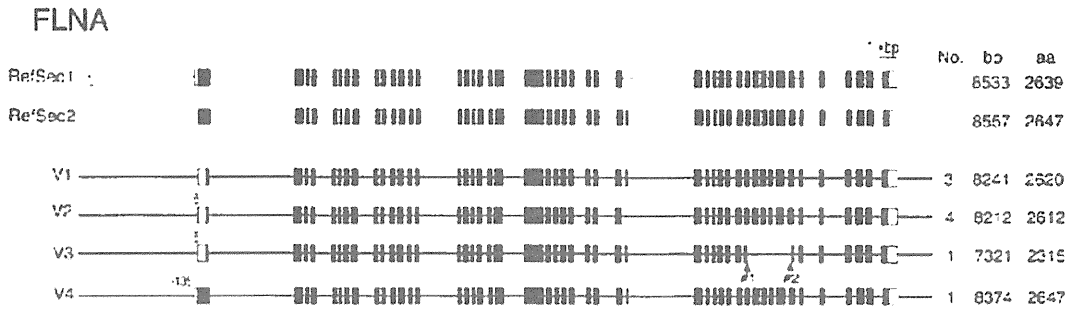


Figure 7. The exon-intron structure of alternative splicing variants for *FLNA*. Eight clones obtained from ARPE-19 were classified into three variants (V1-V3). V4 was obtained from Y79. The relative position of a transcription start site was indicated on the first exon. "No." represents the number of obtained clones. Arrowheads represent the following sequence variations: #1, the 96-bp upstream shift of the 3' splice site of exon 37; #2, the 72-bp downstream shift of the 5' splicing site of exon 42. RefSeq1 and RefSeq2 correspond to GenBank Accession No. NM_001456.3 and NM_001110556.1, respectively. Our clones correspond to AB191259.1 - AB191260.1, AB371574.1- AB371579.1 and AB593010.1.

GenBank contained 3 mRNAs with a full ORF except for the nine clones we registered. The sequence of mRNA1 (X53416) was constructed by combining seven fragments cloned from a human endothelial cell (Gorlin et al., 1990). mRNA2 (AK090427) originated from a single mRNA and had the same initiation codon as V4, but lacked exon 30. Although this clone seems to be near full-length cDNA, the registrant regarded it as a truncated clone maybe because of no evidence for intactness of the 5' end of the cDNA. mRNA3 (GU727643) was synthesized with RT-PCR based on RefSeq. RefSeq1 corresponding to mRNA1 has exon 1, which uses an upstream alternative promoter. Its ORF starts from the same initiation codon with V4. Our nine clones had no exon 1. There were 14 sequences having exon 1 in dbEST. RefSeq2 was constructed based on our clone V1 except for exon 1. The dbEST contained a sequence (CN421698) that has the same deletion as V3.

FLNA has a rod-like structure composed of 24 repeats of the beta-pleated sheet unit: an actin-filament binding domain (Rod1, repeats 1-15), a partner protein binding domain (Rod2, repeats 16-23), a self-assembly domain (repeat 24), and a hinge linking the domains (Hinge-1 and Hinge-2) (Nakamura et al. 2011). V2 lacked the 8-aa residues that were located in the last repeat 15 of Rod1. V3 lacked the 114-aa sequence that was the part of repeats 18 and 19 of Rod2. Since these repeats are known to bind to several partner proteins, these isoforms may lose a function borne by the corresponding repeat.

### 3.3.4. Filamin B, Beta (FLNB)

FLNB and FLNC as well as FLNA are a member of a filamin family. The ARPE-19 libraries contained four *FLNB* clones (8-9 kbp) and one *FLNC* clone (9156 bp). The *FLNC* clone corresponded to RefSeq (data not shown). All clones for FLNB were different AS variants as shown in Figure 8. Four RefSeqs are constructed based on our four clones. They had a similar TSS. V4 had a total of 47 exons and the other three variants lacked exon 26 (93 bp, 31 aa). V2 lacked 11-aa residues due to the 33-bp upstream shift of 3' splice site of exon 31 (designated by #1 in Figure 8). V1 and V4 had a shorter exon 47 due to the use of an alternative polyadenylation signal. GenBank contained two mRNA sequences except for our clones. These two sequences corresponding to V1 were constructed by combining the sequences of cDNA fragments (Takafuta et al., 1998; Xu et al., 1998). Xu et al. (Xu et al., 1998) have reported near full-length cDNA clone (9.5 kb) for V3 derived from a single mRNA. The dbEST contained four clones possessing a shortened exon 31 found in V2 and two clones possessing exon 26 found in V4.

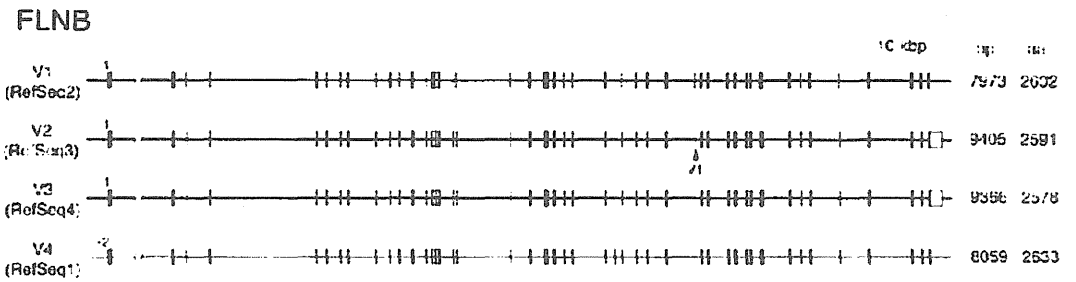


Figure 8. The exon-intron structure of alternative splicing variants for *FLNB*. Four clones were all different variants. Four RefSeqs are constructed based on our four clones as shown in parenthesis. The relative position of a transcription start site was indicated on the first exon. Arrowhead #1 represents the 33-bp upstream shift of the 3' splice site of exon 31. Our clones correspond to AB191258.1 and AB371580.1- AB371582.1.

FLNB has a domain structure similar to FLNA. V4 encoded an isoform possessing 31-aa insertion in the middle part of repeat 31 due to the insertion of exon 26. The isoform encoded by V3 lacked Hinge-1 of 24-aa residues corresponding to exon 31. V2 encoded an isoform lacking the 11-aa C-terminal half of Hinge-1. The deletion of these aa sequences might affect the function of each isoform through the change of binding ability to the partner proteins as well as FLNA. For example, van der Flier et al. showed that an FLNB fragment lacking a part of repeat 19-20 or C-terminal repeat 24 obtained using RT-PCR had a different binding ability to integrin beta subunit (van der Flier, 2002). Furthermore, they showed that the expression pattern of these variants varied from tissue to tissue and during myogenesis. We have to keep in mind that this kind of experiment using RT-PCR shows the expression level of only a partial sequence of transcript and the expression pattern does not reflect the change of the full-length transcript.

### 3.3.5. Eyes Shut Homolog (EYS)

EYS is an extracellular matrix specifically produced in photoreceptor cells (Abd El-Aziz, 2008; Collin et al., 2008). Recently, we showed that one-third of Japanese patients with retinitis pigmentosa had founder mutations in the *EYS* gene. RefSeq1 for the *EYS* gene

comprises 43 exons as shown in Figure 9 and the length of mRNA is 11 kb. In GenBank, there are two short RefSeqs terminating by exon 11. RefSeq2 has a long 3'-UTR. RefSeq3 uses an alternative promoter located between exon 2 and exon 3, resulting in the formation of a new exon 1. The Y79 libraries contained two clones only for short variants. Although V1 was a very-long-sized clone with an insert of 7,898 bp, it terminated by exon 11 containing a long 3'-UTR and encoded the N-terminal 594-aa sequence as well as RefSeq2. The exon 11 of RefSeq2 was split due to splicing. V2 terminated by exon 4 and encoded a short isoform of 318 aa. The EYS protein comprises 27 EGF-like domains and 5 laminin G-like domains. Thus, the isoform encoded by V1 terminated at the middle of the sixth EGF-like domain, and the isoform for V2 at the middle of the third EGF-like domain. The function of these short forms of the EYS protein remains to be solved.

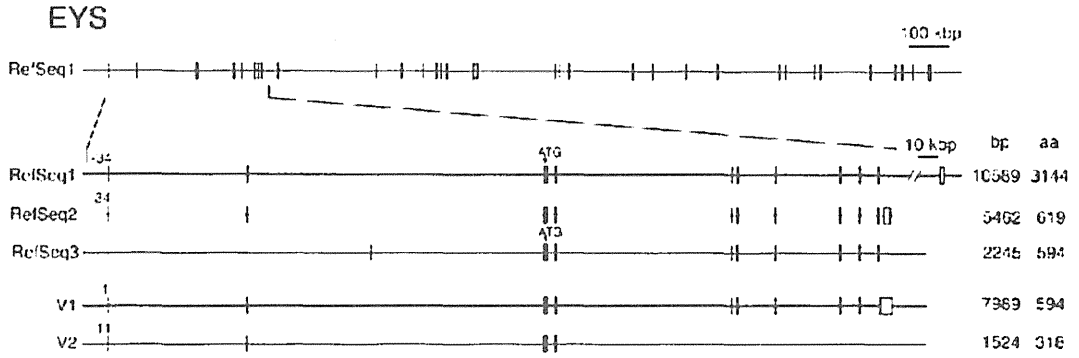


Figure 9. The exon-intron structure of alternative splicing variants for *EYS*. Two variants were cloned from the Y79 cDNA library. The relative position of a transcription start site was indicated on the first exon. RefSeq1, RefSeq2 and RefSeq3 correspond to GenBank Accession No. NM_001142800.1, NM_001142800.2, and NM_198283.1, respectively. Our clones correspond to AB593114.1 and AB593112.1.

## 4. Full-Length cDNA Libraries Derived from Other Species

The power of the vector-capping method was first demonstrated by the transcriptome analysis of budding yeast. Miura et al. performed a large-scale analysis of full-length cDNA libraries prepared from budding yeast cells growing exponentially in a minimal medium and meiotic cells (Miura et al., 2006). They identified 11,575 TSSs associated with 3,638 genes, suggesting that most yeast genes have two or more TSSs. They also identified 45 previously undescribed introns, including those spliced alternatively. Furthermore, they found 667 transcripts in the intergenic region and 367 transcripts derived from antisense strands of known genes. These results suggest that many genes remain unidentified even in an intensively analyzed simple organism such as budding yeast.

Since the vector-capping method was published in 2005 (Kato et al., 2005), it has been adopted by various research projects to construct cDNA libraries from various tissues of various species: plants such as burma mangrove (Miyama et al., 2006), miniature tomato (Aoki et al., 2010), Chinese cabbage (Abe et al., 2011), rubber tree (Suzuki et al., 2012);

mammals such as macaque monkey (Osada et al., 2009), pig (Uenishi et al., 2012), common marmoset (Tatsumoto et al., 2013); parasites such as *Haemaphysalis* (Zhou et al., 2006), *Echinococcus* (Watanabe et al., 2007), *Babesia* (Aboge et al., 2008); hagfish (Uchida et al., 2010); *Bombyx mori* nucleopolyhedrovirus (Katsuma et al., 2011). In many cases, it seems to have been difficult to obtain a large amount of starting material. The vector-capping method requires only several micrograms of total RNA. This seems to be one reason why this method was adopted to prepare the libraries from the above samples. The cloning ability of a long-sized cDNA was confirmed by the cloning of very-long-sized genes (9.1kb and 9.8kb) that encode egg case silk from a wasp spider (Zhao et al., 2006).

# 5. Problems on Identification of AS Variants

## 5.1. AS Variants of Rare or Long-Sized Genes

In the above sections, I have described some problems in identifying AS variants using the conventional methods. A serious problem occurred in the case of a long-sized gene. The combination of alternative promoter usage, multiple alternative splicing sites, and alternative polyadenylation can produce diverse forms of transcripts. The partial sequence analyses using RT-PCR or RNA-seq do not disclose this combination. To determine the precise structure of the AS variant, it is necessary to determine the full sequence of a single mRNA. One solution for this requirement is to determine the full sequence of a full-length cDNA derived from a single mRNA.

Another problem is related to the intactness of the full-length cDNA. In the case of an abundant gene, many cDNA clones can be obtained. If these cDNAs are shown to start at the similar site by comparing their 5'-end sequences, we could regard them as a full-length or near full-length cDNA having a capped site sequence. However, a rare gene may give only one cDNA clone, thus we cannot judge the intactness of this cDNA. The same problem occurs in the case of very short or very long genes. It may be difficult to judge whether the cDNA are derived from intact mRNA or degraded mRNA. The vector-capping method solves these problems. We can judge the intactness of the cDNA by inspecting the presence of the additional dG at the 5' end of the cDNA.

## 5.2. Synthesis of Full-Length cDNA

It has been difficult to obtain a full-length cDNA for a rare or long-sized gene using conventional methods. Here, the problems are shown with regard to each step of cDNA synthesis.

### *(1) Oligo(dT) Priming*

The conventional methods usually use an oligo(dT) primer of ~20 nt to synthesize the first-strand cDNA. When mRNA has a short A stretch, the oligo(dT) primer can accidentally hybridize to this site and be used for cDNA synthesis, resulting in missing the downstream part of mRNA to a poly(A) tail. A good example is an *AIPL1* gene shown in section 3.2.1.

We observed several other examples of such mispriming (data not shown). The vector-capping method uses a vector primer possessing approximately 60-nt dT at one end of the vector. The long dT tail may rarely prime a short A stretch in mRNA.

### *(2) Reaction Conditions*

According to our experience, the amount of template mRNA, reverse transcriptase and substrate nucleotides seem to be essential factors that are related to biases by the expression level or size of mRNA. Usually the first-strand cDNA synthesis is carried out using several micrograms of poly(A)$^+$RNA. In these reaction conditions, most reverse transcriptase and substrate nucleotides seem to be consumed to synthesize cDNA mainly from abundant or short-sized mRNAs, causing biases by the expression-level and size of mRNA. In the vector–capping method, total RNA is used as a template in place of poly(A)$^+$RNA to synthesize the first-strand cDNA under the same reaction conditions. Thus, the amount of enzyme and substrate might be enough to synthesize cDNA from rare or long-sized mRNAs. Omitting mRNA purification steps also may help to reduce these biases.

### *(3) PCR Step*

Some conventional methods including the oligo-capping method contain a PCR step in the procedure for preparing the cDNA library. The amplification step by PCR may cause bias by the expression level and the size of mRNA. In fact, when the full-length cDNA libraries prepared from monkey liver and kidney by the oligo-capping method were compared with those prepared by the vector-capping method, the redundancy of the vector-capped libraries is lower than those of the oligo-capped libraries (Osada et al., 2009). To synthesize rare or long-sized cDNAs, the PCR step should be avoided.

### *(4) Restriction Enzyme Treatment*

The conventional methods contain a linker attachment step, in which a oligonucleotide linker with a restriction enzyme site (e.g. NotI, EcoRI, SalI, XhoI, SfiI et al.) are ligated to the double-stranded cDNA and then after cutting by restriction enzyme the cDNA are introduced into a vector. If the cDNA has the same restriction enzyme site as the linker, it is difficult to obtain full-length cDNA or any cDNA in some cases. The examples are shown in the above sections on *LHX3* and *NRL*. Many clones registered in dbEST seem to be a truncated cDNA that was generated due to this step.

### *(5) Size Fractionation*

Some protocol contains a size fractionation step to remove short cDNA fragments. This step should be avoided because there are many short transcripts having a poly(A) tail. We observed such short full-length cDNAs with < 100 bp (data not shown).

## 5.3. Vector-Capping Method

The vector-capping method solves all the above problems. Thus, this will be the most effective method to synthesize genuine full-length cDNAs at present. However, this has one limitation. It is difficult to obtain full-length cDNA clones from a low-quality RNA sample

containing highly degraded mRNA, because this protocol does not contain a step for experimentally selecting full-length cDNAs, such as a cap-dependent linker ligation in the oligo-capping method. In addition, it requires a lot of labor and cost to search novel AS variants from the vector-capped libraries. This is the case particularly when the target cell expresses genes with low complexity. In that case, we should use a subtraction or normalization protocol together. If the target gene has been decided, we may isolate in advance target cDNA using a probe for the target gene.

## Conclusion

Here I have demonstrated that the vector-capping method provides us with a high-quality cDNA library composed of genuine full-length cDNA clones derived from a single mRNA and that the obtained clones can be used to effectively identify AS variants. This library contains many full-length cDNA clones for rare or long-sized genes whose intactness is guaranteed. By analyzing these clones, we can identify novel AS variants for rare or long-sized genes that have been difficult to obtain using conventional methods. These results suggest that comprehensive, in-depth analysis of full-length cDNA clones isolated from the vector-capped libraries is the most effective way to identify an entire set of AS variants. Furthermore, these full-length cDNA clones can be used as a resource for producing the encoded proteins. I hope that the vector-capping method will be widely used for analyzing full-length AS variants derived from various tissues of various species.

## References

Abd El-Aziz MM, Barragan I, O'Driscoll CA, et al. EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nat Genet.* 2008;40(11):1285-1287.

Abe H, Narusaka Y, Sasaki I, Hatakeyama K, et al. Development of full-length cDNAs from Chinese cabbage (Brassica rapa Subsp. pekinensis) and identification of marker genes for defense response. *DNA Res.* 2011;18(4):277-289.

Aboge GO, Jia H, Terkawi MA, et al. Cloning, expression, and characterization of Babesia gibsoni dihydrofolate reductase-thymidylate synthase: inhibitory effect of antifolates on its catalytic activity and parasite proliferation. *Antimicrob Agents Chemother.* 2008;52(11):4072-4080.

Akey DT, Zhu X, Dyer M, et al. The inherited blindness associated protein AIPL1 interacts with the cell cycle regulator protein NUB1. *Hum Mol Genet.* 2002;11(22):2723-2733.

Alfano G, Vitiello C, Caccioppoli C, et al. Natural antisense transcripts associated with genes involved in eye development. *Hum Mol Genet.* 2005;14(7):913-923.

Aoki K, Yano K, Suzuki A, et al. Large-scale analysis of full-length cDNAs from the tomato (Solanum lycopersicum) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics.* 2010;11:210.

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 2000;10(7):1001-1010.

Bett JS, Kanuga N, Richet E, et al. The inherited blindness protein AIPL1 regulates the ubiquitin-like FAT10 pathway. *PLoS One.* 2012;7(2):e30866.

Clamp M, Fry B, Kamal M, et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A.* 2007;104(49):19428-19433.

Collin RW, Littink KW, Klevering BJ, et al. Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *Am J Hum Genet.* 2008;83(5):594-603.

Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101-108.

ENCODE Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* 2011;9(4):e1001046.

Friedman JS, Khanna H, Swain PK, et al. The minimal transactivation domain of the basic motif-leucine zipper transcription factor NRL interacts with TATA-binding protein. *J Biol Chem.* 2004;279(45):47233-47241.

Gorlin JB, Yamin R, Egan S, et al. Human endothelial actin-binding protein (ABP-280, nonmuscle filamin): a molecular leaf spring. *J Cell Biol.* 1990;111(3):1089-1105.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-945.

Kato S, Ohtoko K, Ohtake H, Kimura T. Vector-capping: a simple method for preparing a high-quality full-length cDNA library. *DNA Res.* 2005;12(1):53-62.

Kato S, Oshikawa M, Ohtoko K. Full-length transcriptome analysis using a bias-free cDNA library prepared with the vector-capping method. *Methods Mol Biol.* 2011;729:53-70.

Kato S, Sekine S, Oh SW, Kim NS, et al. Construction of a human full-length cDNA bank. *Gene.* 1994;150(2):243-250.

Katsuma S, Kang W, Shin-i T, et al. Mass identification of transcriptional units expressed from the Bombyx mori nucleopolyhedrovirus genome. *J Gen Virol.* 2011;92(Pt 1):200-203.

Kimura K, Wakamatsu A, Suzuki Y, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.* 2006;16(1):55-65.

Kolandaivelu S, Huang J, Hurley JB, Ramamurthy V. AIPL1, a protein associated with childhood blindness, interacts with alpha-subunit of rod phosphodiesterase (PDE6) and is essential for its proper assembly. *J Biol Chem.* 2009;284(45):30853-30861.

Li J, Zoldak G, Kriehuber T, et al. Unique Proline-Rich Domain Regulates the Chaperone Function of AIPL1. *Biochemistry.* 2013;52 (12):2089–2096.

Linstedt AD, Hauri HP. Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. *Mol Biol Cell.* 1993;4(7):679-693.

Mitton KP, Swain PK, Chen S, Xu S, Zack DJ, Swaroop A. The leucine zipper of NRL interacts with the CRX homeodomain. A possible mechanism of transcriptional synergy in rhodopsin regulation. *J Biol Chem.* 2000;275(38):29794-29799.

Miura F, Kawaguchi N, Sese J, et al. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci U S A.* 2006;103(47):17846-17851.

Miyama, M., Shimizu H., Sugiyama, M. and Hanagata N. Sequencing and analysis of 14,842 expressed sequence tags of burma mangrove, Bruguiera gymnorrhiza. Plant Science. 2006;171(2):234-241.

Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 2001;29(13):2850-2859.

Nakamura F, Stossel TP, Hartwig JH. The filamins: organizers of cell structure and function. *Cell Adh Migr.* 2011;5(2):160-169.

Netchine I, Sobrier ML, Krude H, et al. Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat Genet.* 2000;25(2):182-186.

Osada N, Hirata M, Tanuma R, et al. Collection of Macaca fascicularis cDNAs derived from bone marrow, kidney, liver, pancreas, spleen, and thymus. *BMC Res Notes.* 2009;2:199.

Oshikawa M, Sugai Y, Usami R, Ohtoko K, Toyama S, Kato S. Fine expression profiling of full-length transcripts using a size-unbiased cDNA library prepared with the vector-capping method. *DNA Res.* 2008;15(3):123-136.

Oshikawa M, Tsutsui C, Ikegami T, et al. Full-length transcriptome analysis of human retina-derived cell lines ARPE-19 and Y79 using the vector-capping method. *Invest Ophthalmol Vis Sci.* 2011;52(9):6662-6670.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008;40(12):1413-1415.

Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 2009;37(Database issue):D32-36.

Rosing M, Ossendorf E, Rak A, Barnekow A. Giantin interacts with both the small GTPase Rab6 and Rab1. *Exp Cell Res.* 2007;313(11):2318-2325.

Seelig HP, Schranz P, Schröter H, Wiemann C, Renz M. Macrogolgin--a new 376 kD Golgi complex outer membrane protein as target of antibodies in patients with rheumatic diseases and HIV infections. *J Autoimmun.* 1994;7(1):67-91.

Sloop KW, Dwyer CJ, Rhodes SJ. An isoform-specific inhibitory domain regulates the LHX3 LIM homeodomain factor holoprotein and the production of a functional alternate translation form. *J Biol Chem.* 2001;276(39):36311-36319.

Sloop KW, Meier BC, Bridwell JL, Parker GE, Schiller AM, Rhodes SJ. Differential activation of pituitary hormone genes by human Lhx3 isoforms with distinct DNA binding properties. *Mol Endocrinol.* 1999;13(12):2212-2225.

Sohda M, Misumi Y, Fujiwara T, Nishioka M, Ikehara Y. Molecular cloning and sequence analysis of a human 372-kDA protein localized in the Golgi complex. *Biochem Biophys Res Commun.* 1994;205(2):1399-1408.

Sohocki MM, Bowne SJ, Sullivan LS, et al. Mutations in a new photoreceptor-pineal gene on 17p cause Leber congenital amaurosis. *Nat Genet.* 2000;24(1):79-83.

Sönnichsen B, Lowe M, Levine T, Jämsä E, Dirac-Svejstrup B, Warren G. A role for giantin in docking COPI vesicles to Golgi membranes. *J Cell Biol.* 1998;140(5):1013-1021.

Suzuki N, Uefuji H, Nishikawa T, et al. Construction and analysis of EST libraries of the trans-polyisoprene producing plant, Eucommia ulmoides Oliver. *Planta.* 2012;236(5):1405-1417.

Suzuki Y, Sugano S. Construction of full-length-enriched cDNA libraries. The oligo-capping method. *Methods Mol Biol.* 2001;175:143-153.

Swaroop A, Xu JZ, Pawar H, Jackson A, Skolnick C, Agarwal N. A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc Natl Acad Sci USA.* 1992;89(1):266-270.

Takafuta T, Wu G, Murphy GF, Shapiro SS. Human beta-filamin is a new protein that interacts with the cytoplasmic tail of glycoprotein Ibalpha. *J Biol Chem.* 1998;273(28):17531-17538.

Takeda J, Suzuki Y, Nakao M, et al. Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res.* 2006;34(14):3917-3928.

Tatsumoto S, Adati N, Tohtoki Y, et al. Development and Characterization of cDNA Resources for the Common Marmoset: One of the Experimental Primate Models. *DNA Res.* 2013;20(3):255-262.

Uchida K, Moriyama S, Chiba H, et al. Evolutionary origin of a functional gonadotropin in the pituitary of the most primitive vertebrate, hagfish. *Proc Natl Acad Sci U S A.* 2010;107(36):15832-15837.

Uenishi H, Morozumi T, Toki D, Eguchi-Ogawa T, Rund LA, Schook LB. Large-scale sequencing based on full-length-enriched cDNA libraries in pigs: contribution to annotation of the pig genome draft sequence. *BMC Genomics.* 2012;13:581.

van der Flier A, Kuikman I, Kramer D, et al. Different splice variants of filamin-B affect myogenesis, subcellular distribution, and determine binding to integrin [beta] subunits. *J Cell Biol.* 2002;156(2):361-376.

Wakamatsu A, Kimura K, Yamamoto J, et al. Identification and functional analyses of 11,769 full-length human cDNAs focused on alternative splicing. *DNA Res.* 2009;16(6):371-383.

Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008;456(7221):470-476.

Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S. Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. *Nucleic Acids Res.* 2007;35(Database issue):D431-438.

Wistow G, Bernstein SL, Wyatt MK, et al. 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;8:205-220.

Xu Wf, Xie Z, Chung DW, Davie EW. A novel human actin-binding protein homologue that binds to platelet glycoprotein Ibalpha. *Blood.* 1998;92(4):1268-1276.

Xu XL, Fang Y, Lee TC, et al. Retinoblastoma has properties of a cone precursor tumor and depends upon cone-specific MDM2 signaling. *Cell.* 2009;137(6):1018-1031.

Zhao AC, Zhao TF, Nakagaki K, et al. Novel molecular and mechanical properties of egg case silk from wasp spider, Argiope bruennichi. *Biochemistry.* 2006;45(10):3348-3356.

Zhou J, Liao M, Hatta T, Tanaka M, Xuan X, Fujisaki K. Identification of a follistatin-related protein from the tick Haemaphysalis longicornis and its effect on tick oviposition. *Gene.* 2006;372:191-198.

眼科医の手引 <814>

学術委員会編

# *EYS*遺伝子変異による網膜色素変性症

国立障害者リハビリテーションセンター病院眼科 岩波将輝

## 1. *Eyes shut homolog*（以下，*EYS*）遺伝子について

網膜色素変性症（Retinitis Pigmentosa, 以下RP）は，網膜の視細胞変性を主体とする遺伝性疾患である。*EYS*遺伝子は2008年に常染色体劣性RP（以下，arRP）の原因遺伝子として同定され，疫学的にarRPの約5-15％を占めることが欧州で報告された[1),2)]。*EYS*遺伝子は，染色体6q12に位置し，2 Mbのゲノム領域にわたり43エクソンからなり，EYS蛋白は3165アミノ酸から構成される。眼で発現する遺伝子群の中では最大のゲノムサイズをもつことがその特徴である。また，*EYS*遺伝子は細胞外基質を構成する蛋白として視細胞に特異的に発現することから，構造蛋白として視細胞を物理的ストレスから保護する作用が推測されている[3)]。

日本人RP患者における*EYS*遺伝子の新規変異が2012年に報告され[4),5)]孤発例を含めたarRPの約30～40％を占めることが推測された。海外の報告とは異なり，日本人で高頻度に検出される理由として，日本人に特異な3つの変異（c.4957 dupA, c.8805 C＞A, c.2528 G＞A）の存在が示された。これら3つの変異は血縁関係をもたない複数の患者家系から同定され，各変異の近傍遺伝子座に対するハプロタイプ解析より，日本人に特異な3つの創始者変異であることが示唆された[4)]。すなわち，それぞれ3つの変異は，独立にある一人の祖先（創始者）の遺伝子に生じた突然変異であり，この変異が代々子孫に引き継がれて集団に拡散していったものと推測された。また，上記以外にも10以上の変異が同定されており，現在，遺伝疫学の詳細が明らかにされつつある[4),5)]。

## 2. *EYS*遺伝子変異とその臨床像

筆者らの報告[6)]において，*EYS*遺伝子変異が両側アレルともに同定された患者23名（平均年齢52.5 ± 8.5歳）について，診療録の病歴・所見をもとに後ろ向き調査を行った結果，自覚症状は20歳代が最も多く，RP診断時年齢は平均39.0 ± 9.7歳（33-61歳）と遅発型を呈する定型RP像であった。良い方の眼の矯正視力0.1以上の眼は14名（60.8％）で観察され，緩徐な進行を示した。また，ゴールドマン視野検査で，Ⅰ／4視標の半径にもとづく求心性視野狭窄5°以下が19名（82.6％）であり，視力が比較的維持されていても，視野障害に伴う日常生活上の困難が中年期以降に強く現れる可能性が高い。そして，良い方の眼の矯正視力0.1未満は9名（39.1％，43-59歳）であり，全例が求心性視野狭窄3°未満に含まれていた。対象の23名のうち20名（86.9％）に白内障がみられ，13名（56.5％）で平均51.1歳（38-67歳）における白内障手術既往があった。これらの結果より，併発白内障の進行時期については中年期（40-60歳）に多いことが推測され，健常者の白内障発症と比較して，より早期に発症し，進行することが示唆された。

## 3. 遺伝子変異型—表現型による臨床重症度の評価

筆者らの報告[4)]より，*EYS*遺伝子変異をもつ患者15名の長期経過（平均7.3年）での解析において，40歳以降の中年期において顕著な視力低下の傾向が観察された。遺伝子型に着目すると，短縮型変異（c.4957 dupA, c.8805 C＞Aなど，蛋白を生成しない機能欠失型変異）をホモまたはヘテロ接合で2つ有する個体群は，短縮型変異とミスセンス変異

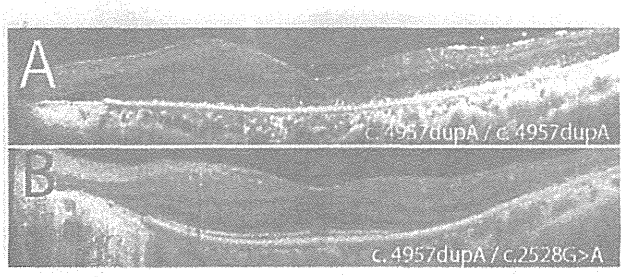


図1

*EYS*遺伝子変異をもつ患者症例（A, Bともに60歳代，女性，左眼），光干渉断層計による網膜断層像を示す。A）短縮型変異c.4957 dupAを有するホモ接合体では，視細胞内節外節接合部（IS/OS line）の消失を認める。矯正視力は0.04，視野（V/4 e視標）は3°であった。B）短縮型変異c.4957 dupAとミスセンス変異c.2528 G＞Aを有するヘテロ接合体では，黄斑部におけるIS/OS lineの描出を認める。矯正視力は0.7，視野は10°であった。

（c.2528 G＞Aなど，異常蛋白を生じる変異）をヘテロ接合で有する個体群と比較して，長期的な予後において，強い視力低下を生じる傾向が確認された。図1A, Bに異なる変異型をもつ患者症例の網膜断層像を示す。今後，これら臨床重症度の予測を可能にするため，より多くの症例の臨床情報を蓄積し，詳細な解析を行う必要がある。

## 4. 今後の展望

*EYS*遺伝子に関連するRPは，日本人arRP患者の約3割～4割を占めることが推測され，現在までに特定された原因遺伝子の中では最も規模が大きい集団と考えられる。これら単一の原因をもつEYS関連RPの患者集団において，薬物・治療の効果の判定を考慮することは，RPにおける治療法の的確な評価と安全な導入につながるものと考えられる。また，RPE65遺伝子変異によるレーベル先天盲で成果の示されている遺伝子治療[7]についても，*EYS*遺伝子における治療に向けた研究が期待される。

［文　献］

1) Abd El-Aziz MM, Barragan I, O'Driscoll CA, et al: *EYS*, encoding an ortholog of *Drosophila spacemaker*, is mutated in autosomal recessive retinitis pigmentosa. Nat Genet 40：1285-1287, 2008.
2) Collin RW, Littink KW, Klevering BJ, et al: Identification of a 2 Mb human ortholog of *Drosophila eyes shut/spacemaker* that is mutated in patients with retinitis pigmentosa. Am J Hum Genet 83：594-603, 2008.
3) Cook B, Hardy RW, McConnaughey WB, et al: Preserving cell shape under environmental stress. Nature 452：361-364, 2008.
4) Iwanami M, Oshikawa M, Nishida T, et al: High prevalence of mutations in the *EYS* gene in Japanese patients with autosomal recessive retinitis pigmentosa. Invest Ophthalmol Vis Sci 53：1033-1040, 2012.
5) Hosono K, Ishigami C, Takahashi M, et al: Two novel mutations in the *EYS* gene are possible major causes of autosomal recessive retinitis pigmentosa in the Japanese population. PLoS One 7：e31036, 2012.
6) 岩波将輝，西田朋美，三輪まり枝，他：日本人網膜色素変性症における*EYS*遺伝子変異とその臨床像．臨床眼科67：275-279, 2013.
7) Maguire AM, High KA, Auricchio A, et al: Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. Lancet 374：1597-1605, 2009.

Genes to Cells

# Derivation of human differential photoreceptor cells from adult human dermal fibroblasts by defined combinations of *CRX, RAX, OTX2* and *NEUROD*

**Yuko Seko[1,2]*, Noriyuki Azuma[3], Toshiyuki Ishii[4], Yukari Komuta[1], Kiyoko Miyamoto[1], Yoshitaka Miyagawa[2], Makoto Kaneda[4] and Akihiro Umezawa[2]**

[1]*Visual Functions Section, Department of Rehabilitation for Sensory Functions, Research Institute, National Rehabilitation Center for Persons with Disabilities, Tokorozawa, Saitama 359-8555, Japan*
[2]*Department of Reproductive Biology, Center for Regenerative Medicine, National Institute for Child Health and Development, Okura, Setagaya, Tokyo, 157-8535, Japan*
[3]*Department of Ophthalmology, National Center for Child Health and Development, Setagaya, Tokyo 157-8535, Japan*
[4]*Department of Physiology, Nippon Medical School, Sendagi, Bunkyo, Tokyo 113-8602, Japan*

**Redirecting differentiation of somatic cells by over-expression of transcription factors is a promising approach for regenerative medicine, elucidation of pathogenesis and development of new therapies. We have previously defined a transcription factor combination, that is, *CRX*, *RAX* and *NEUROD*, that can generate photosensitive photoreceptor cells from human iris cells. Here, we show that human dermal fibroblasts are differentiated to photoreceptor cells by the same transcription factor combination as human iris cells. Transduction of a combination of the *CRX*, *RAX* and *NEUROD* genes up-regulated expression of the photoreceptor-specific genes, recoverin, blue opsin and PDE6C, in all three strains of human dermal fibroblasts that were tested. Additional *OTX2* gene transduction increased up-regulation of the photoreceptor-specific genes blue opsin, recoverin, S-antigen, CNGB3 and PDE6C. Global gene expression data by microarray analysis further showed that photoreceptor-related functional genes were significantly increased in induced photoreceptor cells. Functional analysis, that is, patch-clamp recordings, clearly revealed that induced photoreceptor cells from fibroblasts responded to light. Both the *NRL* gene and the *NR2E3* gene were endogenously up-regulated in induced photoreceptor cells, implying that exogenous *CRX*, *RAX*, *OTX2* and *NEUROD*, but not *NRL*, are sufficient to generate rod photoreceptor cells.**

## Introduction

Redirecting differentiation of somatic cells by over-expression of transcription factors is a promising approach for regenerative medicine, elucidation of pathogenesis and development of new therapies. The process is called 'direct reprogramming' or 'direct conversion' and has been shown in β cells, cardiomyocytes, neurons, platelets and photoreceptors. A specific combination of three transcription factors (Ngn3, Pdx1 and MafA) reprogram differentiated pancreatic exocrine cells in adult mice into cells that closely resemble beta cells (Zhou *et al.* 2008) and a combination of three factors (Gata4, Tbx5 and Baf60c) induce noncardiac mesoderm to differentiate directly into contractile cardiomyocytes (Takeuchi & Bruneau 2009). We recently employed the strategy of 'direct reprogramming' to generate retinal photoreceptor cells from human somatic cells, defining a combination of transcription factors, *CRX, RAX* and *NEUROD*, that induce light responsive photoreceptor cells (Seko *et al.* 2012). In that study, we induced 'iris cells' into photoreceptor cells. During vertebrate eye development, the inner layer of the optic cup differentiates into the

*Communicated by*: Takashi Tada
**Correspondence*: seko-yuko@rehab.go.jp

DOI: 10.1111/gtc.12127

neural retina and iris-pigmented epithelium (IPE). Therefore, the common developmental origin of the iris and the retina may make photoreceptor-induction from iris cells easier than from other types of somatic cells.

The induced pluripotent stem cells (iPS) developed by Takahashi and Yamanaka were the first model for 'direct reprogramming', in which mouse adult fibroblasts were reprogrammed by transduction of four transcription factor genes, Oct3/4, Sox2, c-Myc and Klf4 (Takahashi & Yamanaka 2006). Additionally, functional neurons were generated from mouse fibroblasts by a combination of three factors (Ascl1, Brn2 and Myt1 l) (Vierbuchen *et al.* 2010), and functional platelets were generated from mouse and human fibroblasts by a combination of three factors (p45NF-E2, MafG and MafK) (Ono *et al.* 2012). Because human dermal fibroblasts are less specialized than iris cells, we tested whether human dermal fibroblasts could be converted into photoreceptors by the same defined combination of genes used successfully for human iris cells, *CRX, RAX* and *NEUROD*, to generalize and establish our technology for 'generating photoreceptors'.

In this study, we also investigated an effect of additional transcription factor, *OTX2,* on transdifferentiation of somatic cells into retinal cells. Otx2 is essential for the cell fate determination of retinal photoreceptor cells (Nishida *et al.* 2003), and conditional disruption of the Otx2 gene decreases photoreceptor-associated genes (Omori *et al.* 2011).

Here, we show that the same combination of genes used for human iris cells, that is, *CRX, RAX and NEUROD*, generate human photoreceptor cells from human dermal fibroblasts, and that additional *OTX2* gene transduction further amplifies the expression of retina-specific genes. Our data therefore indicate that human dermal fibroblasts are a superior cell source for reprogramming into photoreceptor cells.

## Results

### Human dermal fibroblasts are induced into a rod- or cone-specific phenotype by defined transcription factors

We selected seven genes, *POU1F1, SOX2, PAX6, RAX, CRX, OTX2* and *NEUROD*, as candidate factors that may contribute to induce photoreceptor-specific phenotypes in human dermal fibroblasts, on the basis that such factors play a role in the development of photoreceptors. *CRX, RAX* and *NEUROD* are essential factors that induce photoreceptor cells from human iris cells (Seko *et al.* 2012) and *POU1f1,* Sox2 and Otx2 bind to the Rx promoter (Martinez-de Luna *et al.* 2010). Human dermal fibroblasts were infected with these genes and were examined for inducible expression of photoreceptor-specific genes. RT-PCR results showed that transduction of *CRX, RAX* and *NEUROD* (CRN) genes up-regulated the expression of the photoreceptor-specific genes recoverin, blue opsin and PDE6C, in all strains of fibroblasts tested (Fig. 1, panel A, B, C). Additionally, CRN-infected fibroblasts became positive for rhodopsin and blue opsin by immunohistochemistry (Fig. 1D). These results suggest that photoreceptor-specific phenotypes are induced by the same combination of transcription factors in human dermal fibroblasts as in human iris cells. However, it appeared that the combination of *CRX, RAX, NEUROD* and *OTX2* (CRNO) up-regulated the photoreceptor-specific blue opsin gene more strongly than the combination of CRN.

### Additional *OTX2* gene transduction increases up-regulation levels of photoreceptor-specific genes

Expression levels of opsin- and phototransduction-related genes in induced- and noninduced fibroblasts were quantitated. Expression levels of S-antigen and recoverin, which are specifically expressed in rod photoreceptors, were much higher in CRNO-infected cells than in CRN-infected cells (S-antigen, $P < 0.01$, recoverin, $P < 0.05$; Welch's *t*-test, Fig. 2). In contrast, expression levels of rhodopsin, blue opsin, green opsin, recoverin, S-antigen, CNGB3 and PDE6C were not increased by additional *PAX6* gene infection (CRNP vs. CRN, in Fig. 2).

### *OTX2* is not an essential factor but an amplifier for induction of photoreceptor cells from human dermal fibroblasts

To investigate whether *OTX2* could be used as an alternative to the essential three genes, that is, *CRX, RAX* and *NEUROD*, we tested the effect of withdrawal of each individual factor from the four genes, that is, *CRX, RAX, NEUROD* and *OTX2*, on expression levels of the opsin- and phototransduction-related genes in induced photoreceptor cells (Fig. 3). Removal of either *CRX, RAX* or *NEUROD* resulted in a marked decrease in blue opsin, S-antigen, PDE6C

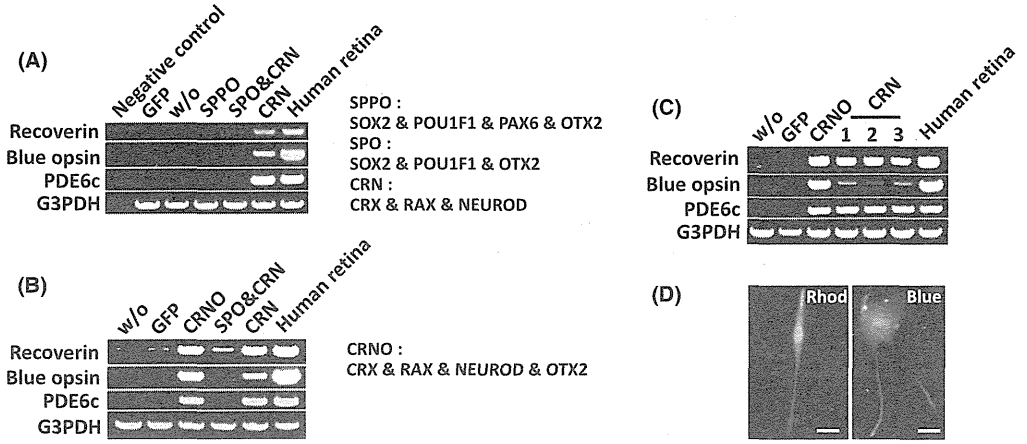


**Figure 1** Induction of retina-specific genes in human dermal fibroblasts by the retroviral infection of genes for defined transcription factors. (A) RT-PCR analysis for photoreceptor-specific genes in cultured human dermal fibroblasts (NHDF) obtained from Lonza after gene transfer of several kinds of transcription factors. Recoverin, blue opsin and PDE6c genes were up-regulated by CRN transduction. 'Negative control': amplified water as a negative control. 'GFP': cultured fibroblasts after retroviral gene transfer of the GFP gene as another negative control. 'w/o': cultured fibroblasts without gene transfer as the other negative control. 'SPPO': *SOX2, POU1F1, PAX6* and *OTX2*. 'SPO': *SOX2, POU1F1* and *OTX2*. 'CRN': *CRX, RAX* and *NEUROD*. 'Human retina': human retinal tissue as a positive control. The amount of cDNA as a template was a half in the positive control. (B) RT-PCR analysis for photoreceptor-specific genes in cultured human dermal fibroblasts (NHDF) obtained from Promo Cell after gene transfer of several transcription factors. Recoverin, blue opsin and PDE6c genes were up-regulated by CRN or CRNO transduction. 'w/o': cultured fibroblasts without gene transfer as a negative control. 'GFP': cultured fibroblasts after retroviral gene transfer of the GFP gene as another negative control. 'CRNO': *CRX, RAX, NEUROD* and *OTX2*. (C) RT-PCR analysis for photoreceptor-specific genes in cultured human dermal fibroblasts (HDF-a) obtained from ScienCell after gene transfer of several transcription factors. Recoverin, blue opsin and PDE6C genes were up-regulated by CRN or CRNO transduction. Expression levels of blue opsin were increased by additional *OTX2* gene transduction. 'w/o': cultured fibroblasts without gene transfer as a negative control. 'GFP': cultured fibroblasts after retroviral gene transfer of the GFP gene as another negative control. 'CRNO': *CRX, RAX, NEUROD* and *OTX2*. '1', '2' and '3' mean independently cultured, transfected and harvested cells by the same combination of CRN genes. (D) Immunocytochemistry using antibodies to rhodopsin and blue opsin (green). Nuclei were stained with DAPI (blue). Experiments were carried out at 2 weeks after infection. The cells in the left panel and the right panel are CRN-infected Fib#2 and Fib#1, respectively. Scale bars represent 10 μm.

and CNGB3 levels; withdrawal of *RAX* resulted in a marked decrease in expression of blue opsin, and withdrawal of *NEUROD* resulted in a striking decrease in expression of PDE6C. Alternatively, withdrawal of *OTX2* alone did not affect the up-regulation of any of the tested photoreceptor-specific genes. These results indicate that *OTX2* is not an essential factor but an amplifier for induction of photoreceptor cells from human dermal fibroblasts, suggesting that additional *OTX2* plays a role in improving the balance and stability of photoreceptor-related gene expression in induced photoreceptor cells. Removal of either *CRX, RAX* or *NEUROD* resulted in a marked decrease in blue opsin, S-antigen, PDE6C and CNGB3 levels, suggesting that each transcription factor plays a role for specific molecular functions along with a role as a constituent of a combination for transdifferentiation to photoreceptor cells.

### Photoreceptor-related functional genes are clearly up-regulated in induced photoreceptor cells from human dermal fibroblasts

To clarify the specific gene expression profile in induced photoreceptor cells, we compared the expression profiles of 50 599 probes in the induced photoreceptor cells (CRN-infected fibroblasts (CRN -Fib), CRNO-infected fibroblasts (CRNO-Fib) and parental cells [fibroblast (Fib)] by microarray analysis (uploaded to GEO accession #GPL16699 at http://www.ncbi.nlm.nih.gov/geo/index.cgi). We first extracted the intersection of the two groups of genes, that is, up-regulated genes by CRN-infection ([CRN-Fib] vs. [Fib]) and those by CRNO-infection ([CRNO-Fib] vs. [Fib]) (signal ratio $\geq +1.5$ for 'up'). According to gene ontology (GO) term annotation, the differentially expressed genes (4124 probes), which were

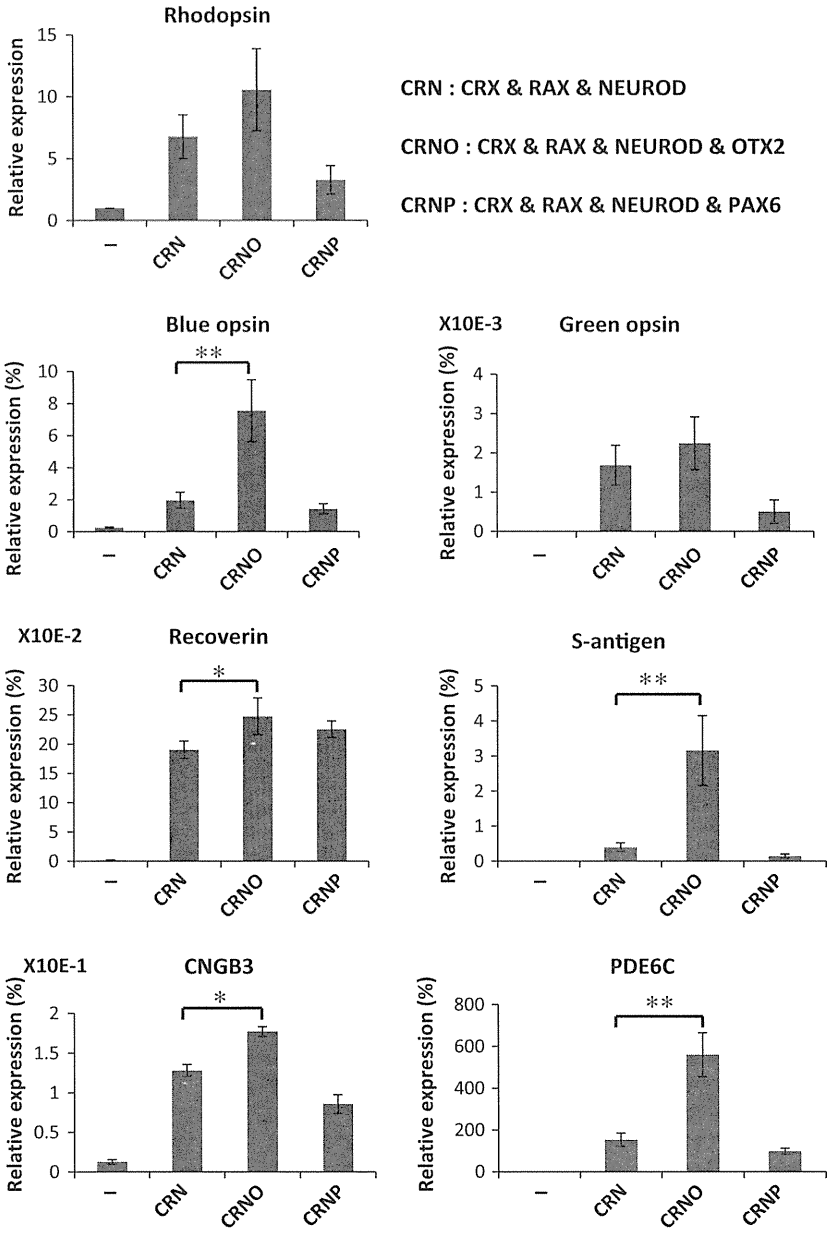


**Figure 2** Effect of additional *OTX2* gene infection. Quantitative RT-PCR results for expression levels of rod- or cone-specific genes in induced photoreceptor cells from human dermal fibroblasts by the defined transcription factors. Quantitative expression levels of rhodopsin, blue opsin, green opsin, recoverin, S-antigen, CNGB3 and PDE6c genes were investigated. The data of green opsin, recoverin and CNGB3 were the results in experiments using Fib#2, and the data of rhodopsin, blue opsin and S-antigen were the results in experiments using Fib#3. The vertical axis indicates expression levels of each gene (%) in the indicated cells, relative to human retinal tissues. For rhodopsin, expression in cultured fibroblasts is regarded as 1.0. Results of statistical analyses for comparison of expression levels between CRNO-infected cells and CRN-infected cells are shown [$*P < 0.05$ and $**P < 0.01$ (Welch's *t*-test)]. '–': cultured fibroblasts without gene transfer as a negative control. 'CRN': *CRX, RAX* and *NEUROD*. 'CRNO': *CRX, RAX, NEUROD* and *OTX2*. 'CRNP': *CRX, RAX, NEUROD* and *PAX6*.

included in the intersection, were categorized into functional groups. Interestingly, when phototransduction-related genes were extracted, they accounted for up to 0.2% of the total (Fig. 4A; Table S1 in Supporting Information). In fact, signals of 16 probes were increased among the 30 phototransduction-related probes.

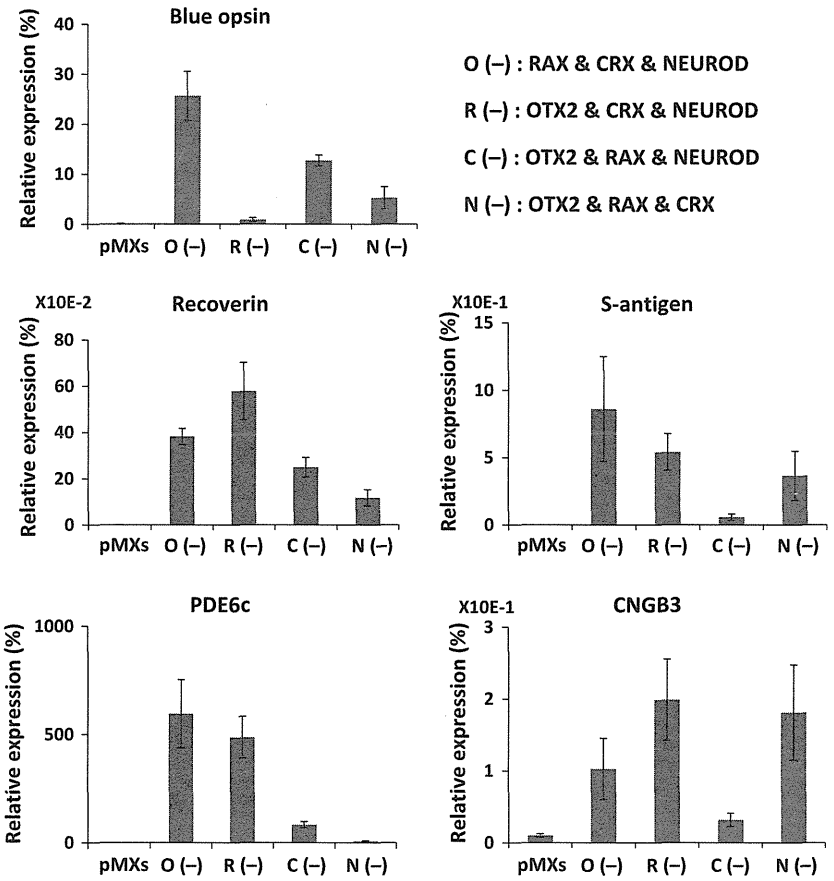


**Figure 3** Effect of individual withdrawal of each gene from the combination of *CRX, RAX, NEUROD* and *OTX2*. Quantitative RT-PCR results for expression levels of rod- or cone-specific genes in induced photoreceptor cells from human dermal fibroblasts by the defined transcription factors. To determine which of the four genes, that is, *CRX, RAX, NEUROD* and *OTX2*, are critical, we examined the effect of withdrawal of individual factors from the pool of the candidate genes on expression of the opsin genes. In this experiment, Fib#3 was used. Quantitative expression levels of blue opsin, recoverin, S-antigen, CNGB3 and PDE6C genes were investigated. Vertical axis indicates expression levels of each gene (%) in the indicated cells, relative to human retinal tissues. Individual withdrawal of *RAX* resulted in a significant decrease in expression of blue opsin and withdrawal of *CRX* resulted in a significant decrease in S-antigen PDE6C and CNGB3. Individual withdrawal of *NEUROD* resulted in a significant decrease in PDE6C. However, withdrawal *OTX2* could up-regulate all of the retina-specific genes tested. 'O(-)': *CRX, RAX* and *NEUROD*. 'R(-)': *CRX, OTX2* and *NEUROD*. 'C(–)': *OTX2, RAX* and *NEUROD*. 'N(-)': *CRX, RAX* and *OTX2*. 'pMXs': cultured fibroblasts after retroviral gene transfer of the pMXs gene as a negative control.

To clarify the difference in gene expression profiles between fibroblast-derived and iris-derived photoreceptor cells, we investigated the expression profiles of default cells (iris cells) and induced cells (CRN-infected iris cells). We carried out GO analysis based on the differentially expressed genes (2585 probes), which were included in the commonly up-regulated genes, that is, ([CRNO-Fib] vs. [Fib]) and ([CRN-Iris] vs. [Iris]) (signal ratio $\geq +1.5$ for 'up'). The phototransduction-related genes were extracted and accounted for up to 4.4% (Fig. 4B; Table S2 in Supporting Information). Although detection/perception, which includes detection of external stimulus, detection of abiotic stimulus and detection of light stimulus, accounted for up to 0.6% of the total in Fig. 4A, the detection/perception accounted for up to 21.1% in Fig. 4B.

## A dermal fibroblast could be a cell source as well as an iris cell

We searched up-regulated genes both in the CRNO-infected fibroblasts and in CRN-infected iris cells (signal ratio $\geq 2.0$ for 'up') and named as 'intersection

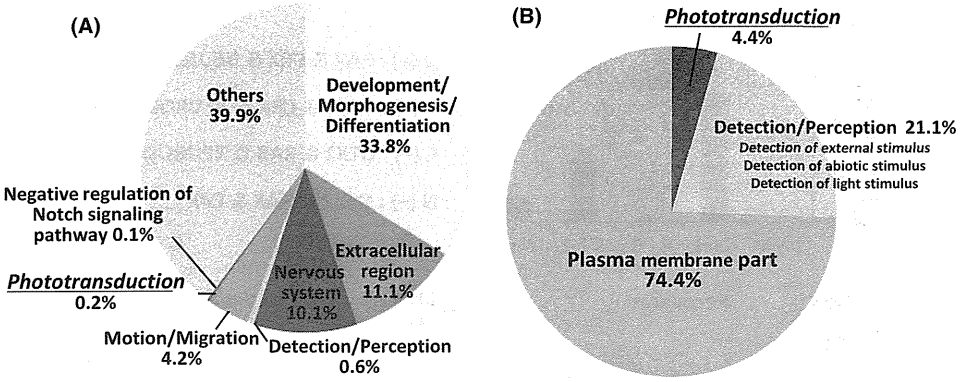


**Figure 4** Categorization of the genes differentially expressed in induced photoreceptor cells from human dermal fibroblasts by the defined transcription factors. (A) Categorization of commonly up-regulated genes in induced photoreceptor cells from human dermal fibroblasts by genes transduction of CRN and CRNO. To clarify the specific gene expression profile in induced photoreceptor cells, we compared the expression levels of 50 599 probes in the induced photoreceptor cells (CRN-infected fibroblasts (CRN-Fib), CRNO-infected fibroblasts (CRNO-Fib) and parental cells [fibroblast (Fib)] by microarray analysis. We searched up-regulated genes in the induced photoreceptor cells by CRN- and CRNO-infection compared with parental cells (signal ratio $\geq +1.5$ for 'up'), respectively. We then extracted the intersection of the two groups of genes, that is, up-regulated genes by CRN-infection and those by CRNO-infection. According to gene ontology (GO) term annotation, the genes differentially expressed in the induced photoreceptor cells by CRN- and CRNO-infection (4124 probes) were categorized into functional group, figuring out the relative importance or significance of the GO-term [corrected *P*-value $< 0.01$)]. After that, we carried out additional categorization into eight groups. Interestingly, phototransduction-related genes were extracted and account for up to 0.15% of the total. (B) Categorization of commonly up-regulated genes (2585 probes) in the CRNO-transfected dermal fibroblasts and CRN-transfected iris cells (signal ratio $\geq 1.5$ for 'up'). According to gene ontology (GO) term annotation, the genes differentially expressed in the induced photoreceptor cells (2585 probes) were categorized into functional groups to figure out the relative importance or significance of the GO term (corrected *P*-value $< 0.01$).

of Fib and Iris'. Then, we extracted retina-related genes from them according to Gene Ontology and a previous paper (Omori *et al.* 2011). We focused on remarkably up-regulated genes ([CRNO-Fib]/[Fib] $> 9.0$) and extracted them (Fig. 5A). We then compared signal ratios between [CRNO-Fib]/[Fib] ($\Delta$Fib) and [CRN-Iris]/[Iris] ($\Delta$Iris). The signal ratios of 18 probes were higher in [CRNO-Fib] ($\Delta$Fib/$\Delta$Iris $\geq 2.0$); however, the signal ratios of 47 signals were higher in [CRN- Iris] ($\Delta$Iris/$\Delta$Fib $\geq 2.0$). As for other 78 probes, the signal ratios were regarded not to be significantly different (Fig. 5B; Table S3 in Supporting Information). To analyze the gene expression data in an unsupervised manner, we carried out principal component analysis (PCA). The gene expression patterns in the CRN-Fib, CRNO-Fib and CRN-Iris were close based on component 2 (PC2) but were apart from the parent cells (Fib and Iris) (Fig. 5C). We investigated the difference in endogenous expression of *CRX*, *RAX* and *NEUROD* between CRN-Fib and CRN-iris by RT-PCR (Fig. 5D). The endogenous *CRX* genes started to be expressed in CRN-Fib, but the expression levels of *RAX* and *NEUROD* were higher in CRN-Iris than in CRN-Fib. Both the *NRL* gene and the *NR2E3* gene were endogenously up-regulated in the induced photoreceptor cells, that is, CRN-Fib, CRNO-Fib and CRN-Iris (Fig. 5E).

### Induced photoreceptor cells from fibroblasts are photoresponsive *in vitro*

Light stimulation was applied to infected or non-infected human fibroblasts because CRN- or CRNO-infected cells showed the photoreceptor-like phenotypes by RT-PCR and global gene expression analyses. Among cells tested, significant light responses were detected in a portion of infected cells (Fig. 5F; Fig. S2 in Supporting Information). An infected cell presented a large outward current when exposed to light (Fig. 5F, upper panel). However, no detectable outward current was evoked when light stimulation was given to a noninfected cell (Fig. 5F, lower panel).

## Discussion

This is the first report that human dermal fibroblasts can differentiate into photoreceptor cells by the same combination of transcription factors, *CRX*, *RAX* and