

厚生労働科学研究費補助金（障害者対策総合研究事業 感覚器障害分野）

分担研究報告書

変異 *EYS* を発現している変性視細胞モデルの作出に関する研究 その2

研究分担者 世古裕子 国立障害者リハビリテーションセンター研究所  
感覚機能系障害研究部 視覚機能障害研究室長

研究要旨

本研究の目的は、網膜色素変性症（以下、RP）患者の体細胞を視細胞に分化誘導し、変性視細胞モデルを作出し解析することである。平成24年度から平成25年度にかけて、市販のヒト皮膚線維芽細胞から直接リプログラミングによって網膜細胞へと分化誘導する実験を行った。その結果、ヒト皮膚線維芽細胞から、視細胞特異的な *EYS* 遺伝子や光トランスダクションに関わる遺伝子が発現し光応答のある視細胞様細胞に分化することがわかった。さらに、倫理審査委員会の承認を得て、平成24年度には正常ボランティアの、平成25年度には3名、平成26年度には2名のRP患者の皮膚採取を実施した。正常ボランティアとRP患者由来皮膚線維芽細胞から視細胞に分化誘導し、それぞれに発現する遺伝子と蛋白の解析を行った。RP患者が保有する *EYS* 遺伝子の変異の種類によって、変性視細胞モデルにおける *EYS* 遺伝子の転写産物に違いが見られた。この成果は、RP患者の予後予測や網膜の変性機構の解明、さらには診断法・治療法の開発に役立つことが期待される。

研究協力者：

青井 則之（国立障害者リハビリテーションセンター病院 非常勤医師）

宮本 潔子（国立障害者リハビリテーションセンター研究所 感覚機能系障害研究部 派遣研究員）

小牟田 縁（国立障害者リハビリテーションセンター研究所 感覚機能系障害研究部 流動研究員）

抜本的治療法が望まれている。一方で2008年には、非定型網膜色素変性症の一つであるLeber先天性黒内障において遺伝子治療による視機能の回復が報告され、将来的にRPの治療法として遺伝子治療を含めた個別化医療が行われることが期待されている。

RPの原因となる遺伝子変異は患者毎に様々であるが、いずれの場合も最終的には視細胞の変性を引き起こす。視力を失う前に視細胞の変性を抑え、病気の進行自体を止めることが望まれている。遺伝性疾患における個別化医療では、根本的な治療を行うためにはまず原因となる遺伝子の変異が判っていることが前提条件であり、さらにその変異によって発症へと至るメカニズムが明らかになっ

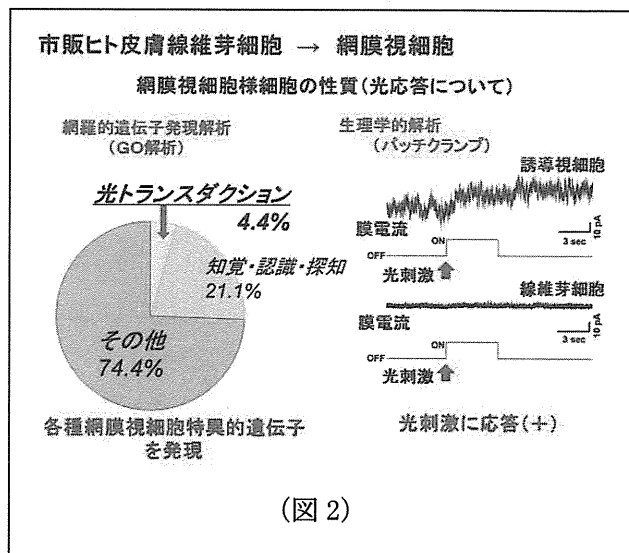
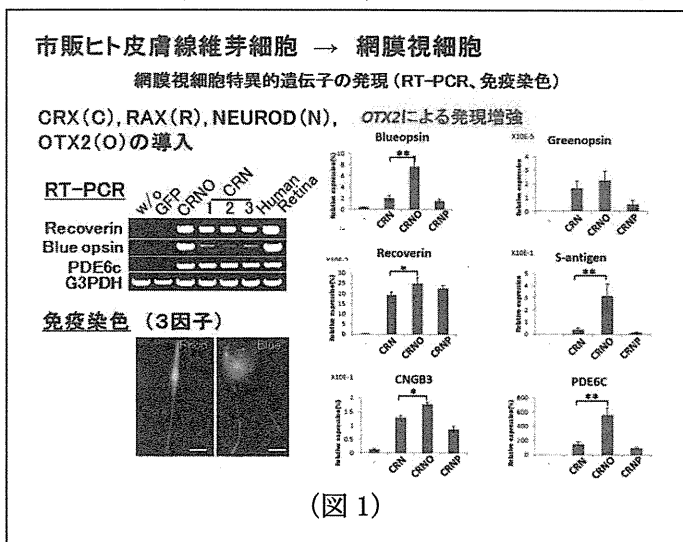
A. 研究の背景と目的

網膜色素変性症（以下、RP）は、夜盲に始まり、徐々に視野狭窄、視力低下が起こる進行性の遺伝子疾患であり、現在有効な治療法はない。中途失明により就労が困難となるケースが多く、RP患者の社会生活を支援するためにも、予後の予測法や

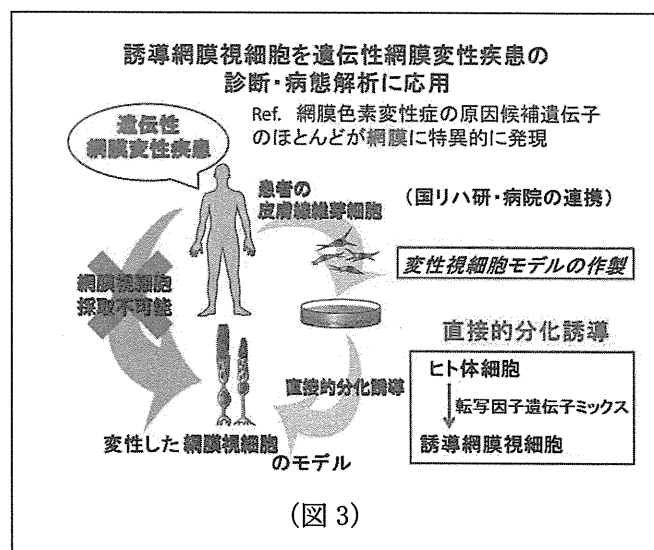
ていることが必須となる。

*EYS* 遺伝子は 2008 年に常染色体劣性 (ar) RP の原因遺伝子として同定され、arRP の約 5-15% を占めることが海外で報告された。当研究所において、先行研究として日本人 RP 患者 68 名における遺伝学的解析を行い、複数の新規変異を同定した。そして、これら変異を保有する患者は arRP の約 30% を占める高い頻度であった (Iwanami et al. IOVS, 2012)。*EYS* 遺伝子に変異を保有する RP 患者について、網膜変性機序を明らかにして治療法を見出すことができれば、日本人 RP の少なくとも 3 割の QOL (quality of life) の向上につながる。さらに残り 7 割については、新規診断法の開発が望まれる。

*EYS* 遺伝子を含めて、RP の原因候補遺伝子の多くは網膜に特異的に発現している。そのため、網膜に発現する遺伝子 (mRNA) を解析できれば、全ゲノムを解析するよりも効率がよい。しかし、患者網膜に発現する遺伝子を直接解析することは困難である。一方で、世古らは、ヒト体細胞 (虹彩細胞・皮膚線維芽細胞) から直接リプログラミングによってヒト網膜視細胞に分化誘導することに成功している (Seko et al., PLoS ONE, 2012, Seko et al., Genes Cells, 2014) (図 1、2)。



この分化誘導の方法を応用し、患者から採取可能な皮膚や血液由来の細胞から網膜細胞へと分化誘導できれば、*EYS* 遺伝子を初めとした網膜のみに発現する RP の原因候補遺伝子の mRNA を解析することが可能となる。これまでの遺伝子解析の中心であったゲノム解析ではなく、より効率が高いと考えられている cDNA をもとにした解析が可能となる (図 3)。



さらに、現時点では、診断がついたとしても治療法が確立していないという問題があり、これを解決するために、変性視細胞モデルを用いた薬剤ス

クリーニングの方法も検討した。本研究では、患者由来の誘導視細胞を用い、日本人の arRP の主要原因遺伝子であることが明らかとなった *EYS* 遺伝子について、遺伝子診断法を確立し、予後予測や治療法の手がかりを得ることが最終目的である。

## B. 研究方法

分担研究者の世古らが開発したヒト虹彩細胞あるいは皮膚線維芽細胞を視細胞に分化誘導する方法 (Seko et al., PLoS ONE, 2012, Seko et al., Genes Cells, 2014) を応用し、RP 患者より採取した皮膚線維芽細胞を単離・培養・凍結保存し、これを視細胞に分化誘導し変性視細胞モデルとした。これと正常皮膚線維芽細胞から分化誘導した誘導視細胞との間で、遺伝子発現を比較した。

具体的には、(1) ヒト皮膚線維芽細胞を培養、(2) 網膜視細胞への分化誘導を高効率で行うことができる転写因子遺伝子 (*CRX*, *NEUROD*, *RAX*, *OTX2* 遺伝子) の混合物をレトロウイルスベクターを用いて導入し、遺伝子導入した細胞 (誘導視細胞様細胞) と導入していない細胞 (コントロール細胞) それぞれから mRNA を抽出、RT-PCR、*EYS* 遺伝子の塩基配列の解析を行った。また、それぞれの細胞から蛋白を抽出し、抗 *EYS* 抗体を用いてウェスタンブロットを行った。

変性視細胞モデルの作製については、平成24年度に倫理申請が受理され、平成24年度には3名の正常ボランティアの皮膚採取を、平成25年度には3名の平成26年度には2名のRPの皮膚採取を実施し、それぞれの皮膚線維芽細胞の単離・培養・凍結を行った。いずれの患者も *EYS* 遺伝子変異を保有する患者であり、5名の患者はそれぞれ変異の種類が異なっていた。倫理面には十分配慮して

実験を行った。

(倫理面への配慮)

RP患者のゲノム解析ならび網膜色素変性症患者の非眼球組織から網膜細胞を分化誘導する研究計画 (「患者由来分化誘導細胞を用いた網膜変性疾患の新規診断法の開発」) については、国立障害者リハビリテーションセンターヒトゲノム・遺伝子解析研究倫理審査委員会において承認を得ている。

## C. 研究結果

### ①誘導視細胞における *EYS* 遺伝子の発現

正常ボランティアのヒト皮膚線維芽細胞を用いて、直接リプログラミングによって網膜細胞へ分化誘導した結果、*EYS* 遺伝子の発現がみられた。

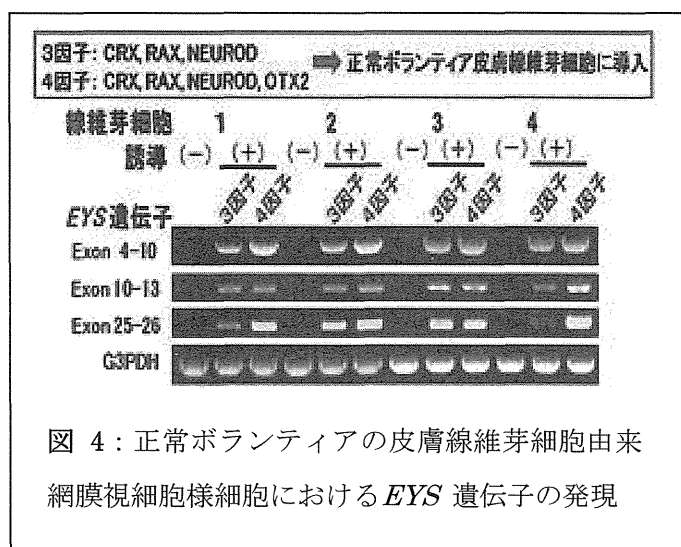


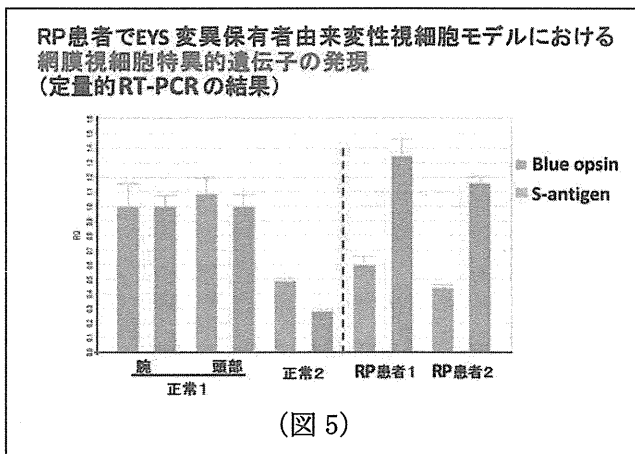
図 4: 正常ボランティアの皮膚線維芽細胞由来網膜視細胞様細胞における *EYS* 遺伝子の発現

図 4は、3名のボランティアから採取した4種類の線維芽細胞の結果である。線維芽細胞1は、ボランティア1の腕から採取、線維芽細胞2は、ボランティア1の頭部から採取、線維芽細胞3&4は、ボランティア2&3の腕から採取した。4因子の導入によって、安定的に *EYS* 遺伝子の発現誘導がみられた。この結果から、皮膚線維芽細胞由来網膜視細胞様細胞を用いる本実験系が、*EYS*

遺伝子変異を保有するRP患者の診断に応用できる可能性が示唆された。

②変性モデル細胞における網膜特異的遺伝子の発現

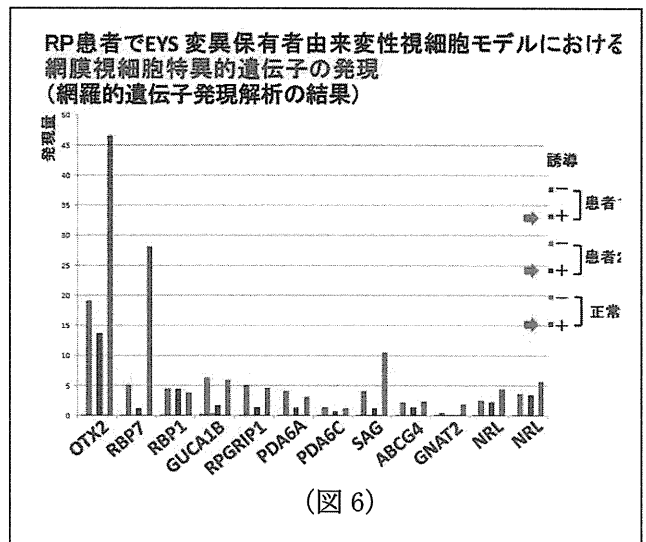
RP患者と正常ボランティア由来の皮膚線維芽細胞それぞれから分化誘導した誘導視細胞に発現する網膜特異的遺伝子の量を定量RT-PCRで比較した。その結果、ブルーオプシンとS-antigenに関しては、正常細胞と変性モデル細胞との間で有意な差は認められなかった。また、皮膚線維芽細胞のソースに関し、同一ボランティアにおいて腕あるいは頭部から採取し比較した結果、場所による有意な差も認められなかった（図 5）。



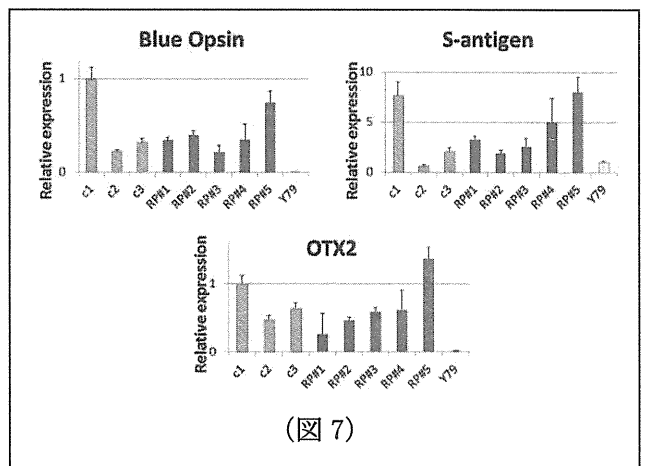
この結果から、多くの網膜視細胞特異的遺伝子の発現量は、EYS 遺伝子変異の有無には依存しない可能性が示唆された。

さらに多数の網膜特異的遺伝子の発現について、網羅的遺伝子発現解析によって検討を行った（図 6）。この結果から、多くの網膜視細胞特異的遺伝子の発現量が、EYS 遺伝子変異の有無には依存しないことが確認された。また、この実験系において、誘導視細胞における網膜特異的遺伝子の発現が、EYS 遺伝子の変異の有無ではなく、導入した転写因子の発現量に依存する可能性が示唆された。

網膜特異的に発現する遺伝子の発現量と EYS 遺



伝子変異との関連をさらに詳細に検討するため、3名の正常ボランティアと5名のRP患者由来の誘導視細胞におけるブルーオプシンとS-antigenの発現量を、定量RT-PCRによって検討した（図 7）。



その結果、正常ボランティア（C1, C2, C3）由来の誘導視細胞の間でも、ブルーオプシンとS-antigenの発現量にはばらつきが見られ、そのばらつきが導入した転写因子のひとつである OTX2 遺伝子の発現量にある程度依存することが確認された。しかしながら、患者細胞由来誘導視細胞（変性視細胞モデル）における発現量のばらつきについては、導入した OTX2 遺伝子の発現量の影響の

他に、個体差の影響も否定できない。従って、さらに多数例での検討が必要ではあるが、*EYS* 遺伝子変異を保有するRP患者由来誘導視細胞において、正常細胞由来誘導視細胞とほぼ同等の網膜視細胞特異的遺伝子の発現がみられることがわかった。

### ③変性モデル細胞における *EYS* 遺伝子の発現

RP患者由来の皮膚線維芽細胞を分化誘導した結果、視細胞特異的な遺伝子である *EYS* 遺伝子の発現増加が見られた。しかし、*EYS* 遺伝子の発現量は、正常細胞とRP患者細胞それぞれから誘導した誘導視細胞との間で有意な差は見られなかった (図8)。

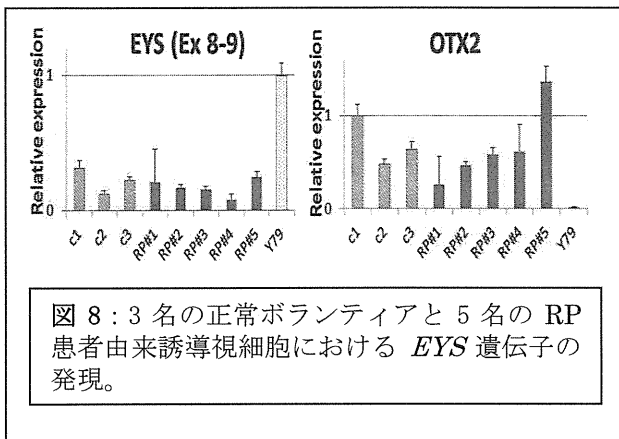
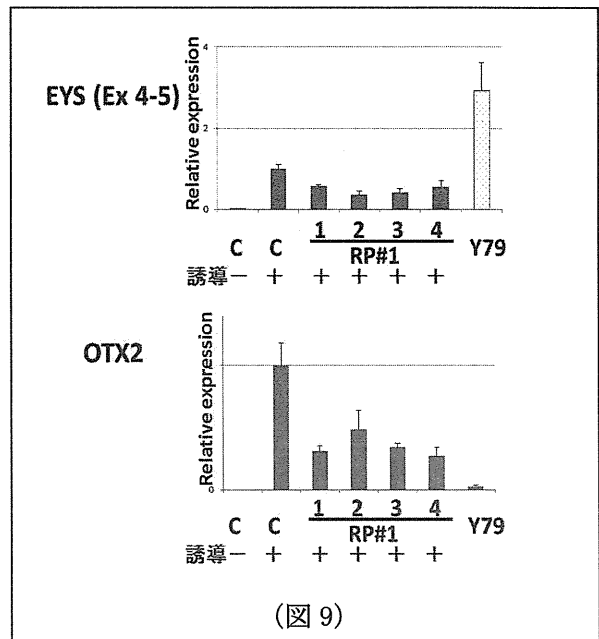


図8: 3名の正常ボランティアと5名のRP患者由来誘導視細胞における *EYS* 遺伝子の発現。

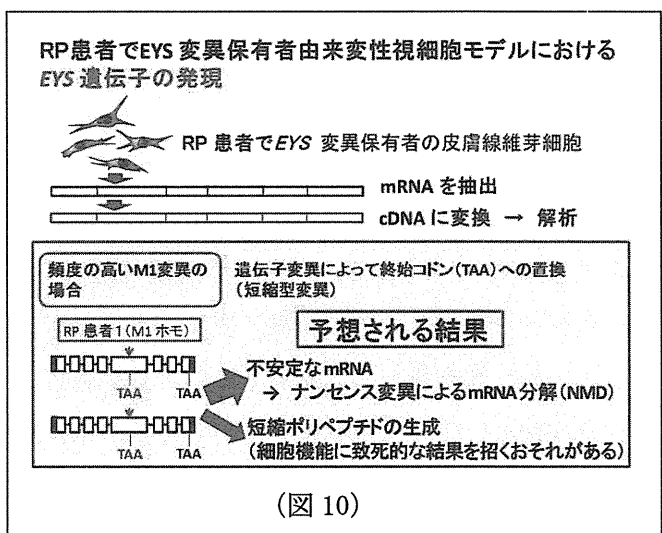
本実験系では、転写因子遺伝子ミックスの導入による導入効率が実験毎に異なることを考慮し、同じ患者1 (RP#1) 由来の誘導視細胞で、4回の、独立して誘導された細胞について、同じ *EYS* 遺伝子の中で異なるエクソンを増幅し、定量的に発現量を調べた (図9)。その結果、やはり、正常細胞と比べ、RP患者由来細胞で *EYS* 遺伝子の発現量の低下は認められなかった (図9)。



(図9)

④網膜色素変性症患者由来の皮膚線維芽細胞から誘導された視細胞様細胞に発現された *EYS* 遺伝子 (cDNA) の遺伝子配列解析の結果。

### 【M1 変異について】

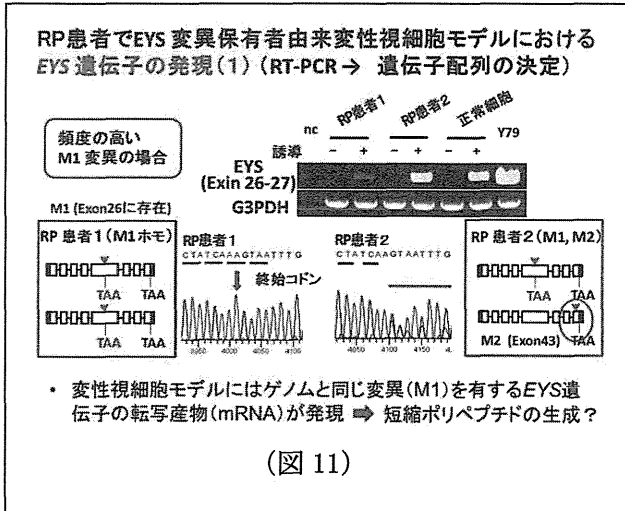


(図10)

患者1では、ゲノム解析では短縮型変異 (M1) が見出されている (両側アレルにM1変異を保有: c.4957dupA, p.S1653Kfs\*2, エクソン26)。この場合、一般的には上図に記載したとおり、ナンセンス変異依存 mRNA 分解機構 (NMD) が働き、*EYS* 遺伝子の転写産物が検出できないことも予想された (図10)。しかし、RT-PCRの結果、正常とほ

とんどかわらない量の *EYS* 遺伝子の転写産物が存在することが前述の実験から示唆された。そこで、この転写産物における *EYS* 遺伝子の遺伝子配列を解析した。

解析結果を以下に示す (図 11)。



患者 1 由来の誘導視細胞における *EYS* 遺伝子の転写産物には、ゲノムに見られた変異と同様の変異 (ナンセンス変異) が見られることが明らかになった。

さらに、患者 2 (片側アレルにM1変異、もう片側のアレルにM2変異をもつ) でも同様に、M1変異を含む領域 (エキソン26-27) について増幅を行い、塩基配列を調べた。その結果、変異を有するアレル由来の転写産物と変異を有しないアレル由来の転写産物のいずれもが発現していることが示された。

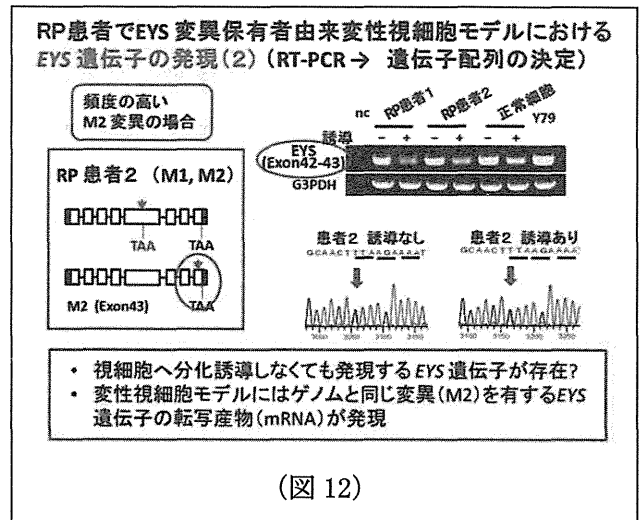
以上の結果は、NMDが働くことなくM1変異を保有する転写産物が存在し、短縮ポリペプチドが生成されていることを示唆している。

#### 【M2変異について】

患者 2 (片側アレルに M1 変異、もう片側のアレルに M2 変異をもつ:c. 4957dupA,p. S1653Kfs\*2、

エキソン 26 ; c. 8805C>A, p. Y2935\*, エキソン 43) について、M2 変異を含む領域 (エキソン 42-43) を増幅し、塩基配列を調べた。その結果、変異を有するアレル由来と考えられる転写産物と変異を有しないアレル由来の転写産物のいずれもが発現していることが示され、ゲノムに見られる変異と同一の変異の存在が確認された。

解析結果を以下に示す (図 12)。

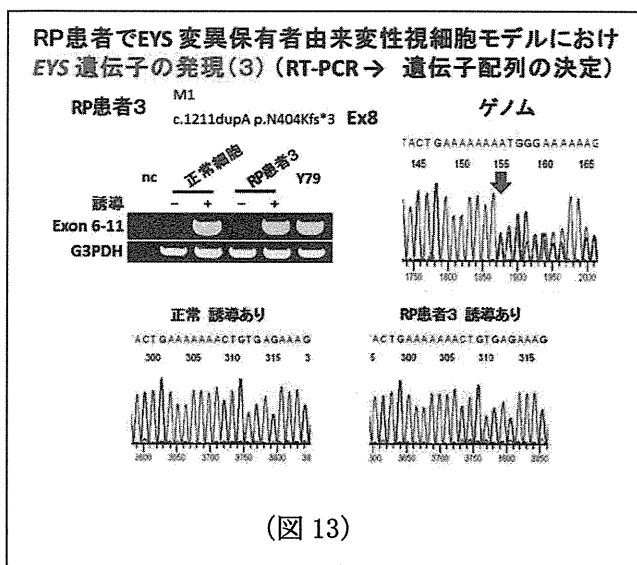


この実験結果から得られたもうひとつの事実は、M2変異を含む領域 (エキソン42-43) が、誘導していない皮膚線維芽細胞にも発現していることが確認されたことである。一方で、M1変異を含む領域 (エキソン26-27) は、誘導する前の皮膚線維芽細胞には発現せず、4種類の転写因子を導入して視細胞誘導することによって発現してくる。網膜に選択的に発現されると考えられている *EYS* 遺伝子であるが、M2変異を含む領域 (エキソン42-43) は、網膜以外の組織にも発現するアイソフォームとしても存在する可能性があると考えられる。

#### 【M1、M2以外の変異について】

患者 3 は、*EYS* 遺伝子について、片側アレルにM1変異、もう片側のアレルにM1、M2以外の変異

(c.1211dupA, p.N404Kfs\*3) を保有している。この変異はエクソン8に存在するため、エクソン6-11の領域を増幅し、塩基配列を調べた。その結果、患者3の皮膚線維芽細胞から分化誘導して作成した変性視細胞モデルにおいて、変異を有するアレル由来の転写産物を検出することができなかった(図13)。この変異については、NMDが働いている可能性が示唆される。

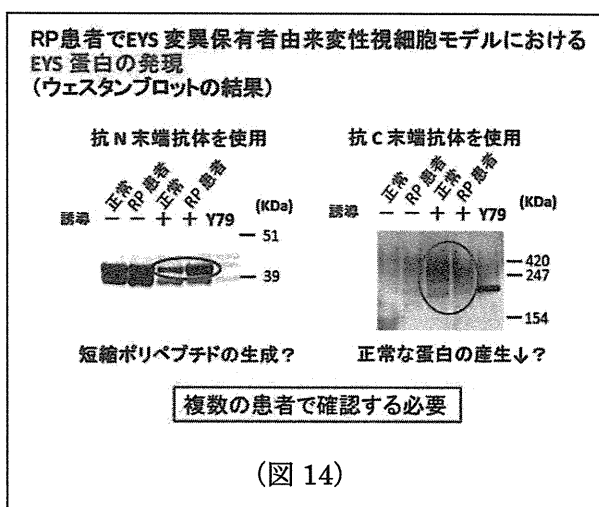


⑤網膜色素変性症患者由来の皮膚線維芽細胞から誘導された視細胞様細胞に発現されたEYS蛋白の発現

RP患者と正常ボランティア由来の皮膚線維芽細胞それぞれから分化誘導した誘導視細胞に発現するEYS蛋白の量をウェスタンブロットで比較した。正常ボランティアとRP患者それぞれの皮膚線維芽細胞に4種類の転写因子遺伝子を導入して網膜視細胞に分化誘導し、分化誘導をしていない細胞を含めた4種類の細胞から蛋白を抽出した。抗体は、サンタクルーズ社の2種類の抗EYS抗体(抗N末端抗体と抗C末端抗体)を用いた。

この実験では、前述の患者1、すなわち、両側

アレルにM1変異を保有する患者の皮膚線維芽細胞を用いた。その結果、抗C末端抗体を使用した場合、正常細胞由来の視細胞モデルとRP患者由来の変性視細胞モデルとの間で比較すると、300 KDa近傍に広く広がるバンドと200 KDa近傍のバンドの濃さが変性視細胞モデルの方が有意に薄くなっているのが観察された(図14)。いずれのバンドも、陽性コントロールとして用いたY79細胞のバンドと分子量がわずかにずれているが、抗体に対応するペプチドによる吸収も認められたので、EYS蛋白である可能性がある。一方で、抗N末端抗体を用いた場合、40KDa近傍に、変性視細胞モデルの方が有意に濃いバンドが観察された(図14)。このバンドは、Y79には存在せず、皮膚に存在するバンドであるが、正常細胞では、4種類の転写遺伝子を導入して網膜視細胞に分化誘導すると発現が著明に減少する。しかし、RP患者由来の変性視細胞モデルでは、分化誘導しても発現の減少が見られない。この現象については、さらに追及していく必要がある。



⑥変性視細胞モデルの変性過程の解析

変性視細胞モデルの応用として、将来の治療

法・進行予防法・発症予防法の開発にむけた薬剤スクリーニングの系を提案するために、本研究では、正常細胞由来の視細胞モデルを用い、アポトーシス・ネクローシスあるいはオートファジーを誘導する種々の薬剤を用い、アポトーシスの指標である抗active-caspase3抗体陽性細胞を指標として、細胞死を起こす薬剤の種類と濃度を検討した(図15)。

各細胞変性誘導剤の効果								
	アポトーシス誘導剤					小胞体ストレス誘導剤		オートファジー誘導剤
	Actinomycin	Camptothecin	Cycloheximide	Dexamethasone	Etoposide	Tunicamycin	Thapsigargin	Rapamycin
分化細胞	3μM	0.6μM	10μM	なし	10μM	10μM	1μM	なし
GFP	50μM	10μM	50μM	なし	なし	なし	なし	なし
未感染	50μM	50μM	50μM	なし	なし	なし	なし	なし

(図15)

この実験から、遺伝子を導入して視細胞様細胞に分化させた細胞はアポトーシス誘導に対し感受性が上昇し、小胞体ストレスからアポトーシスを起こす機序が形成されていることが示唆された。

変性モデル細胞を用いて同様の実験を行い、変性過程の解析を行う予定である。

#### D. 考察

我々は、一般的なヒト体細胞であるヒト皮膚線維芽細胞に複数の転写因子遺伝子 (*CRX*, *NEUROD1*, *RAX*, *OTX2*) を導入すると、視細胞の性格をもった細胞に再生分化することを見出した(平成25年度成果)。この細胞は視細胞そのものには及ばない不完全な視細胞様細胞とも呼ぶべき細胞ではあるが、日本人網膜色素変性症の主要な原因候補遺伝子であることがわかった *EYS* 遺伝子が発現誘導されることがわかり、直接的分化

誘導の本研究への応用が期待された。本研究によって、網膜特異的に発現する遺伝子の多くが、*EYS* 遺伝子変異の有無に関わらず、正常とほぼ同様に発現することがわかった。網膜特異的に発現する *EYS* 遺伝子も同様に、正常の誘導視細胞と RP由来の変性視細胞モデルでは、ほぼ同じレベルの発現量が認められた。しかし、転写産物の塩基配列を解析することによって、*EYS* 遺伝子変異の中でも頻度の高いM1変異(短縮型変異)が、転写産物にも同様に見られることがわかり、mRNAの分解機構が働かないことが示唆された。一方で、*EYS* 遺伝子変異の種類によっては、NMDが働いていることも示唆され、疾患の重症度との関連を調べる必要があると考えている。

今回、正常細胞由来の視細胞モデルとRP患者由来の変性視細胞モデルそれぞれの細胞抽出液について、抗*EYS*抗体を用いたウェスタンブロットを行った結果、変性視細胞モデルにおいて、正常な*EYS*蛋白が減少し、異常なポリペプチドが増加していることが示唆された。この蛋白の分子量は、予想される分子量とは異なっていたため、RPの病態との関連も含めて、今後の課題と考えている。現在、網羅的遺伝子発現解析を行っている。今後、変性視細胞モデルにおいて有意に減少あるいは増加する分子を明らかにして、診断法・治療法の開発に貢献したい。

#### E. 結論

変性視細胞解析が、網膜色素変性症の迅速診断法の開発、病態の解明に有用であることが示唆された。

#### G. 研究発表



論文作成中

H. 知的財産権の出願・登録状況

(予定を含む。)

なし

### III. 研究成果の刊行に関する一覧表

## 研究成果の刊行に関する一覧表

### 書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書 籍 名	出版社名	出版地	出版年	ページ
Kato, S.	Identification of genuine alternative splicing variants for rare or long-sized transcripts.	DiMaggio S. and Braschipp E.	New Developments in Alternative Splicing Research.	Nova Biomedical.	New York	2013	p. 89-108

### 雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
岩波将輝	EYS遺伝子変異による網膜色素変性症.	日本の眼科	85(4)	p. 40-41	2014
Seko, Y. Azuma, N. Ishii, T. Komuta, Y. Miyamoto, K. Miyagawa, Y. Kaneda, M. Umezawa, A.	Derivation of human differential photoreceptor cells from adult human dermal fibroblasts by defined combinations of <i>CRX</i> , <i>RAX</i> , <i>OTX2</i> and <i>NEUROD</i> .	Genes Cells	19(3)	p. 198-208	2014

## IV. 研究成果の刊行物・別刷

The license for this PDF is unlimited except that no part of this digital document may be reproduced, stored in a retrieval system or transmitted commercially in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

## Chapter 5

---

# Identification of Genuine Alternative Splicing Variants for Rare or Long-Sized Transcripts

---

*Seishi Kato*<sup>\*</sup>

Research Institute, National Rehabilitation Center for Persons with Disabilities,  
Tokorozawa, Japan

## Abstract

Only sequence analysis of full-length transcripts can identify genuine alternative splicing variants. However, it was difficult to obtain full-length cDNAs for rare or long-sized transcripts. Recently, we have developed a powerful method, named a vector-capping method, to construct a size-unbiased full-length cDNA library containing rare or very-long-sized cDNA clones with >10kbp inserts. The characteristic of the full-length cDNA contained in this library is that the intactness of the 5'-end capped site sequence of the cDNA can be assured by the presence of an additional dG at its 5' end. Since this full-length cDNA is derived from a single mRNA, this library enables us to perform in-depth analysis of genuine alternative splicing variants. Using the vector-capping method, we prepared full-length cDNA libraries from human retina-derived cell lines and analyzed the full sequence of the clones. As a result, we found many novel alternative-splicing variants for rare or long-sized transcripts. In this chapter, I show the examples of these variants including very-long-sized transcripts with >7kbp that were identified by us for the first time.

## 1. Introduction

The Human Genome Project revealed that the human genome seems to encode only 20,000~25,000 protein-coding genes (International Human Genome Sequencing Consortium,

---

<sup>\*</sup> Correspondence to: Seishi Kato, Research Institute, National Rehabilitation Center for Persons with Disabilities, 4-1 Namiki, Tokorozawa, Saitama 359-8555, Japan. E-mail: kato-seishi@rehab.go.jp.

2004). The analysis, including the evolutionary conservation, further cut the number of protein-coding genes to ~20,500 (Clamp et al., 2007). This number was unexpectedly small to understand the function of genes underlying the complex biological system in the cell. This issue has been solved by the discovery of diverse transcript variants for each gene. The recent research showed that diverse variants are produced from a single gene locus due to alternative promoter usage (Kimura et al., 2006), alternative splicing (AS) (Modrek et al., 2001), and alternative polyadenylation (Beaudoing et al., 2000). Initially, these variants were identified by mapping of expressed sequence tags (ESTs) to the genome. Recently, a new high-throughput sequencing technology such as mRNA-seq was applied to in-depth sequencing analysis of mRNAs isolated from various tissues and cell lines. These analyses revealed that more than 90% of human genes undergo AS (Wang et al., 2008; Pan et al., 2008).

Since the AS events vary between tissues and between developmental stages, each AS variant should be involved in the regulation of tissue-specific or cell-specific development. To fully understand the relationship between the genetic information encoded by the genome and the biological function of the cell, it is necessary to identify all transcripts including a full set of AS variants. One trial to achieve this purpose was an ENCODE project (ENCODE Project Consortium, 2011). This project adopted tiling DNA microarrays, RNA-seq, cap-analysis of gene expression (CAGE), and paired-end diTag (PET) to determine exonic regions, transcription start sites (TSSs), splice junctions, transcript 3' ends, and polyadenylation sites (Djebali, 2012). However, these protocols produce only partial sequence showing the presence of each site. Patterns of AS and alternative cleavage and alternative polyadenylation were found to be strongly correlated across tissues (Wang et al., 2008). This means that we need to know the precise combination of these alternative sites to determine the correlation between them.

To know the combination of multiple variation sites in a single transcript, the full sequence of the full-length transcript is required. The analysis of full-length transcripts can be achieved by obtaining the corresponding full-length complementary DNA (cDNA). Large-scale sequencing analyses of full-length cDNA clones were carried out using full-length cDNA libraries synthesized with the oligo-capping method (Takeda et al., 2006; Wakamatsu et al., 2009). These analyses identified a large number of AS variants including alternative TSSs.

The conventional methods for synthesizing full-length cDNAs have the following problems: (i) inability to determine whether or not the cDNA starts from a true TSS, (ii) loss of some clones due to restriction enzyme treatment during a cDNA synthesis process, (iii) difficulty in synthesizing long-sized cDNAs. Recently, we have developed a novel method, named a vector-capping method, to overcome these problems (Kato et al., 2005; Kato et al., 2011). Using this method, we prepared full-length cDNA libraries from human retina-derived cell lines. By the large-scale sequencing analysis of these libraries, we identified a lot of novel AS variants (Kato et al., 2005; Oshikawa et al., 2008; Oshikawa et al., 2011). In this chapter, I describe the examples of novel splicing variants for rare or long-sized genes we identified, and I would like to emphasize the importance of identifying a genuine AS variant derived from a single mRNA using full-length cDNA.

## 2. Synthesis of Full-Length cDNA Using the Vector-Capping Method

We started our investigation 20 years ago by focusing on how to obtain whole human proteins. The strategy we took was to collect all proteins as a form of cDNA. At that time, the Human Genome Project had been launched and one of the projects was to analyze ESTs. However, ESTs composed of cDNA fragments were not suitable to obtain proteins. To achieve our purpose, we needed to obtain a full-length cDNA that contains an intact open reading frame (ORF) to produce the encoded protein. Thus, we developed a novel method to synthesize full-length cDNA based on replacing the cap structure of mRNA by a DNA-RNA chimeric oligonucleotide (Kato et al., 1994). This method enabled us to effectively synthesize full-length cDNAs, but it had drawbacks: it required a lot of mRNA and many reaction steps.

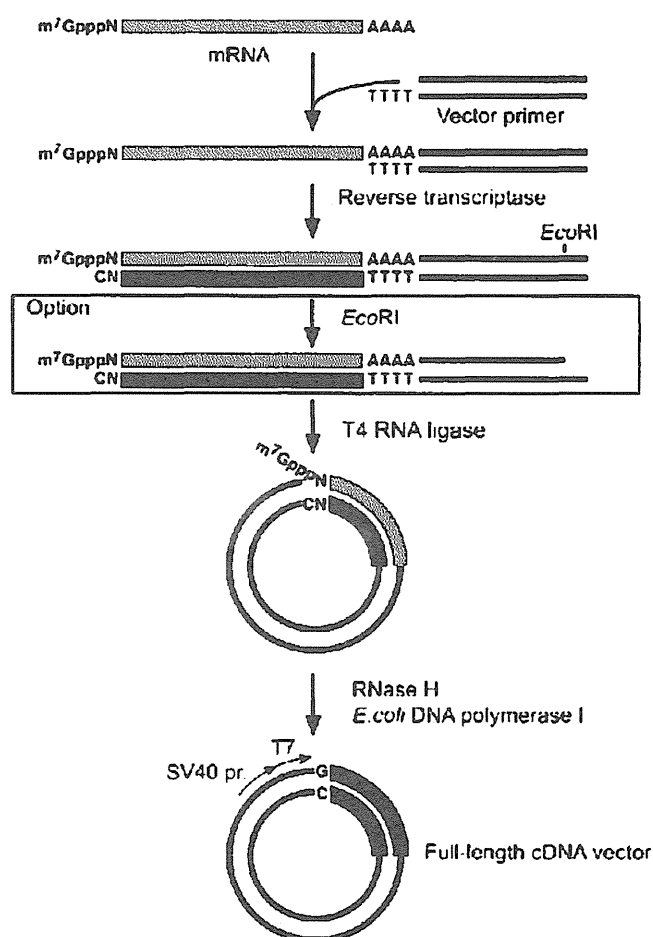


Figure 1. Schematic procedure for the vector-capping method. Several micrograms of total RNA is enough as a starting material. A vector primer has an approximately 60-nucleotide dT tail. The EcoRI digestion step can be omitted. Since the full-length cDNA vector has an SV40 promoter, the encoded protein can be produced in the mammalian cells by introducing the vector into the cells.

When improving this method, we succeeded in developing the vector-capping method shown in Figure 1 (Kato et al., 2005; Kato et al., 2011). Its process is very simple: the first-strand cDNA is synthesized using a vector primer, and then the vector-cDNA conjugate is circularized. The development of this method is attributed to the discovery of an unexpected reaction: the 3' end of the first-strand cDNA can be ligated to the 5' end of the vector primer using "RNA ligase". Furthermore, we found that the full-length cDNA possesses an additional dG at the 5' end. This additional dG is derived from dC that added to the 3' end of the first-strand cDNA by terminal deoxynucleotidyl transferase activity of reverse transcriptase only when the template mRNA has a cap structure. Thus, the presence of the additional dG at the 5' end assures the intactness of the 5'-end capped site sequence of the cDNA.

The full-length cDNA library constructed using the vector-capping method has the following advantages compared with those by conventional methods:

- (i) the library is composed of genuine full-length cDNA clones of > 95% content,
- (ii) we can identify full-length cDNA by the presence of an additional dG at the 5' end,
- (iii) artificial mutation or deletion seldom occurs because the procedure contains neither PCR nor the restriction enzyme treatment step,
- (iv) the library contains full-length clones for rare or long-sized genes,
- (v) we can easily identify the cDNA for an antisense gene due to the use of the vector primer.

Thus, the resulting cDNA library seems to provide us with the full-length cDNA clones ideal for identifying the alternative TSS, AS, and alternative polyadenylation.

### 3. Analysis of Full-Length cDNA Clones

#### 3.1. Retina-Derived Full-Length cDNA Libraries

We have been searching for genes responsible for retinitis pigmentosa, which is the major cause of visual impairment among patients visiting our center. To identify a novel candidate gene causing this disease, we identified genes specifically expressed in the retina by analyzing the full-length cDNA libraries that were constructed from human retinal pigment epithelium cell line ARPE-19 and human retinoblastoma cell line Y79 using the vector-capping method (Kato et al., 2005; Oshikawa et al., 2008; Oshikawa et al., 2011). We randomly picked up 100,000 clones from each library and stored them as glycerol stocks. By sequencing the 5' end of approximately 24,000 clones from each library, we identified a total of 39,643 full-length cDNA clones that were classified into 7,067 genes (Oshikawa et al., 2011). In this section, I describe the examples of novel AS variants obtained from the above libraries. Most of full-length clones and the other full-length cDNA libraries remain not fully analyzed: ARPE-19, 52,800 clones; Y79, 52,800 clones; embryonal pluripotent carcinoma cell line NT2/D1, 76,800 clones; human testis, 76,800 clones. If researchers are interested in particular genes, they may find new AS variants by full sequencing of those clones. All



clones are available from RIKEN BioResource Center DNA Bank (<http://dna.brc.riken.jp/en/NRCDhuman.html>).

### 3.2. Characterization of Eye-Specific Genes

#### 3.2.1. Aryl Hydrocarbon Receptor Interacting Protein-Like 1 (*AIPL1*)

Since retinoblastoma cell line Y79 is derived from cone progenitor cells (Xu et al., 2009), Y79 cells expressed various photoreceptor-specific genes. One of abundant eye-specific genes found in our Y79 full-length cDNA libraries was *AIPL1*. *AIPL1* has been identified as a gene responsible for Leber congenital amaurosis (LCA), a severe early-onset retinopathy that leads to visual impairment in infants (Sohocki et al., 2000).

Fifteen clones (C1-C15) encoding *AIPL1* were fully sequenced, and their exon-intron structures were determined as shown in Figure 2. Since TSSs were distributed within the range of position -14 to position 13, the same promoter seemed to be used. We identified seven AS variants by the shift of a splicing site, skipping of exon 3, and the alternative use of a polyadenylation signal. The encoded proteins were classified into five isoforms: 384-amino acid (aa) isoform (V1 and V4, 9 clones); 345-aa isoform (V3, 2 clones); 321-aa isoform (V2 and V5, 2 clones); 270-aa isoform (V6, 1 clone); 262-aa isoform (V7, 1 clone). The difference between V1 and V4 and the difference between V2 and V5 were due to the alternative use of the polyadenylation signal. V3 showed the 58-bp downstream shift of the 3' splice site of exon 1 (designated by #7 in Figure 2), resulting in the shift of the initiation codon of the longest ORF from exon 1 to exon 2. This shift causes the loss of the N-terminal 39-aa residues. V2 and V5 lacked exon 3, resulting in the 63-aa deletion from the middle part of the protein. V6 and V7 lacked exon 6 because of the shift of the polyadenylation site, resulting in the deletion of the C-terminal 114-aa residues. Furthermore, V7 showed the 24-bp downstream shift of the 5' splice site of exon 4 (designated by #8), causing the corresponding 8-aa deletion. It should be noted that the Y79 cell-derived transcripts showed five single nucleotide polymorphisms (#1, #2, #4, #5, #6) and 2-bp deletion (#3). As a result, the clones were classified into two haplotypes, and their allelic origin was identified. Half of the clones (C3, C7, C9, C11, C12, C13, C14) were assigned to one haplotype.

The identified variants were compared with RefSeq in GenBank provided by the National Center for Biotechnology Information (NCBI), which is a collection of taxonomically diverse, non-redundant and richly annotated sequences representing naturally occurring molecules of DNA, RNA, and protein (Pruitt et al., 2009). Three RefSeqs were constructed as the transcripts of an *AIPL1* gene. RefSeq1 and RefSeq2 correspond to V1 and V2, respectively. Our collection did not contain the clone corresponding to RefSeq3 that is an AS variant skipping exon 2. In NCBI's GenBank, 13 sequences except for our 14 sequences were registered as *AIPL1* mRNA with an ORF. However, these mRNA sequences had no polyadenylation signal, whereas our clones had a canonical polyadenylation signal. AATAAA (V1-V6) and AGTAAA (V7) followed by a poly(A) tail. The mRNA sequences without a polyadenylation signal were terminated before an A-stretch ( $A_{15}$ ) at position 1386 of RefSeq1 or an A-rich region ( $A_6CA_4CA_4CA_5$ ) at position 2131. These clones seem to have been synthesized by priming of the oligo(dT) primer to these A-stretch sites during the first-strand cDNA synthesis. Of these 13 mRNAs, five clones correspond to V1, two clones to V2,

and one clone to V6. The remaining five clones were novel variants, suggesting that there would be more variants for *AIPL1* transcripts to be identified.

*AIPL1* has been reported to interact with various proteins including NUB1 (Akey et al., 2002), FAT10, FAT10nylated protein, UBA6 (Bett et al., 2012), and the catalytic subunit (alpha) of rod cGMP phosphodiesterase (PDE6A) (Kolandaivelu et al., 2009). *AIPL1* is necessary for the proper assembly of functional rod PDE6 subunits (Kolandaivelu et al., 2009) that is a key phototransduction enzyme. *AIPL1* is composed of three functional domains: an FKBP-like prolyl peptidyl isomerase (FKBP) domain, a tetratricopeptide (TPR) domain, and Pro-rich domain (PRD).

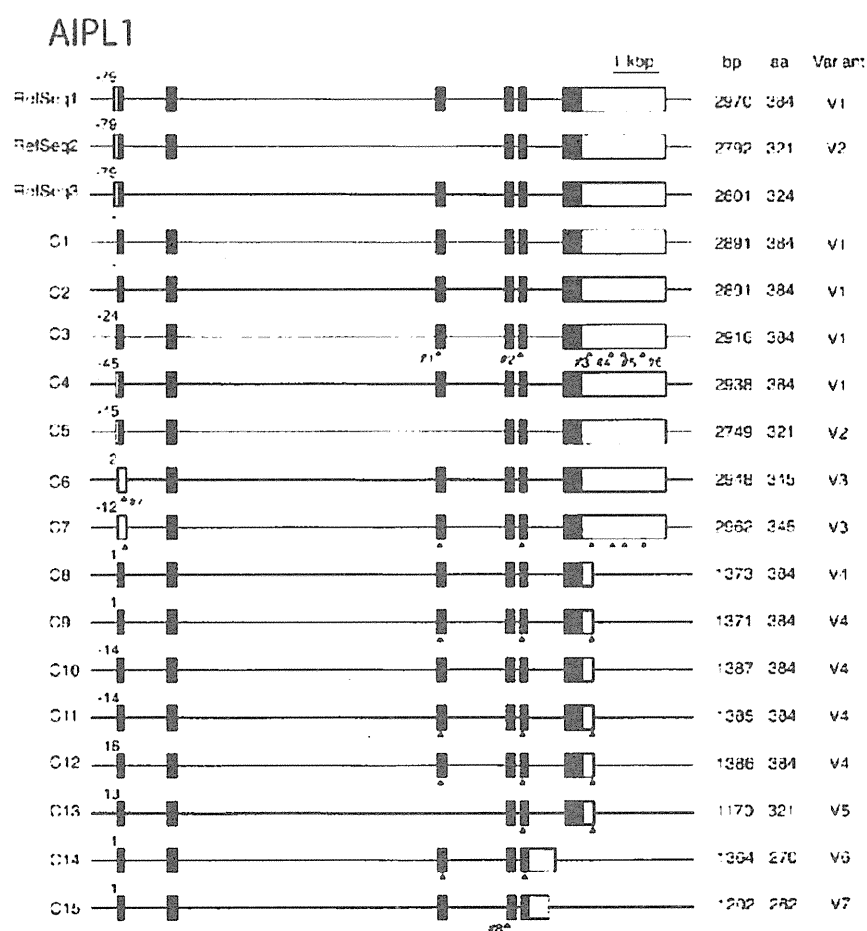


Figure 2. The exon-intron structure of alternative splicing variants for *AIPL1*. Fifteen clones (C1-C15) were classified into seven variants (V1-V7). The relative position of a transcription start site was indicated on exon 1. Arrowheads represent the following sequence variations: #1, SNP (A>G); #2, SNP (A>G); #3, 2-nt deletion (AA>--); #4, SNP (A>G); #5, SNP (A>G); #6, SNP (C>T); #7, downstream splicing site shift (58 bp); #8, downstream splicing site shift (24 bp). RefSeq1, RefSeq2, and RefSeq3 correspond to GenBank Accession No. NM\_014336.3, NM\_001033054.1, and NM\_001033055.1, respectively. Our clones correspond to AB593062.1 - AB593067.1.

V3 encodes an isoform lacking the N-terminal 39-aa residues whose function is unclear. V6 and V7 encode an isoform lacking the C-terminal 114-aa residues corresponding to the

PRD carrying a chaperone activity (Li et al., 2013). This isoform might lose function as a chaperone, because the Trp278X mutant lacking the C-terminal 107-aa residues causes LCA (Sohocki et al., 2000). V2 and V5 encode an isoform lacking the FKBP domain. To elucidate why these isoforms lacking a functional domain are produced in the cell, it is necessary to know the role of each isoform in the AIPL1 regulation system by investigating their localization in the cell or binding activity with other proteins. Although the expression level of each minor variant is low, the total level of minor variants reaches the same level as main variants (V1 and V4), suggesting that isoforms encoded by these minor variants play their own roles in the AIPL1-related system. Since the EST database in GenBank (dbEST) contains novel variants, we would find more novel variants by analyzing the full-length cDNA libraries.

### 3.2.2. LIM Homeobox 3 (*LHX3*)

The Y79 libraries contained many clones encoding various eye-specific transcription factors. A transcription factor *LHX3* is a member of a large protein family that carries a LIM domain, a Cys-rich zinc-binding domain, and is required for pituitary development and motor neuron specification. Two variants encoding human *LHX3* (isoform a and isoform b) have been cloned from pituitary cDNA libraries (Sloop et al., 1999), and they were adopted as RefSeq for *LHX3* genes. The mutation of this gene caused combined pituitary hormone deficiency (CPDH) (Netchine et al., 2000). There is no report for *LHX3* expressed in the eye except for ESTs obtained from eye and pineal gland libraries in dbEST.

Eight clones for *LHX3* were obtained from the Y79 cDNA libraries and four variants were identified as shown in Figure 3. Five clones designated by V1 encoded a 397-aa isoform a. V2 using an alternative promoter had a different exon 1 from V1 and encoded a 402-aa isoform b. As a result, the N-terminal 25-aa sequence of isoform a was different from the N-terminal 30-aa sequence of isoform b. V3 was a novel variant containing an unspliced intron between exon 2 and exon 3, resulting in shortening of the first ORF followed by a longer ORF. The upstream short ORF encoded the N-terminal 89-aa sequence of the isoform a and the downstream long ORF encoded the C-terminal 264-aa sequence of isoform a. V4 using another promoter had a novel exon 1, and encoded a novel 386-aa isoform whose N-terminal 15-aa sequence was different from those of isoform a and b.

V1 and V2 correspond to RefSeq1 and RefSeq2, respectively. V3 and V4 are novel variants. In GenBank two mRNA sequences containing ORF were registered except for our six sequences. These mRNAs correspond to isoform a and isoform b that were cloned from human pituitary cDNA libraries (Sloop et al., 1999). These two sequences had no polyadenylation signal maybe due to the use of a random primer in synthesizing cDNA. The dbEST contained 13 sequences: nine from eye, two from pineal gland, and two from brain. All sequences lacked a sequence corresponding to exon 1. The cDNA libraries used for cloning these ESTs were prepared using conventional methods that contained a NotI or XhoI treatment step. Since the sequence of the full-length cDNA contained NotI and XhoI sites, the fragmentation of synthesized cDNA might occur. Although dbEST contained many ESTs isolated from the Y79 full-length cDNA libraries prepared using the oligo-capping method, there was no EST for *LHX3* in spite of an abundant gene. This was explained by the presence of an SfiI site in the *LHX3* gene, because the oligo-capping protocol contained a step using linkers with the SfiI site (Suzuki and Sugano, 2001).

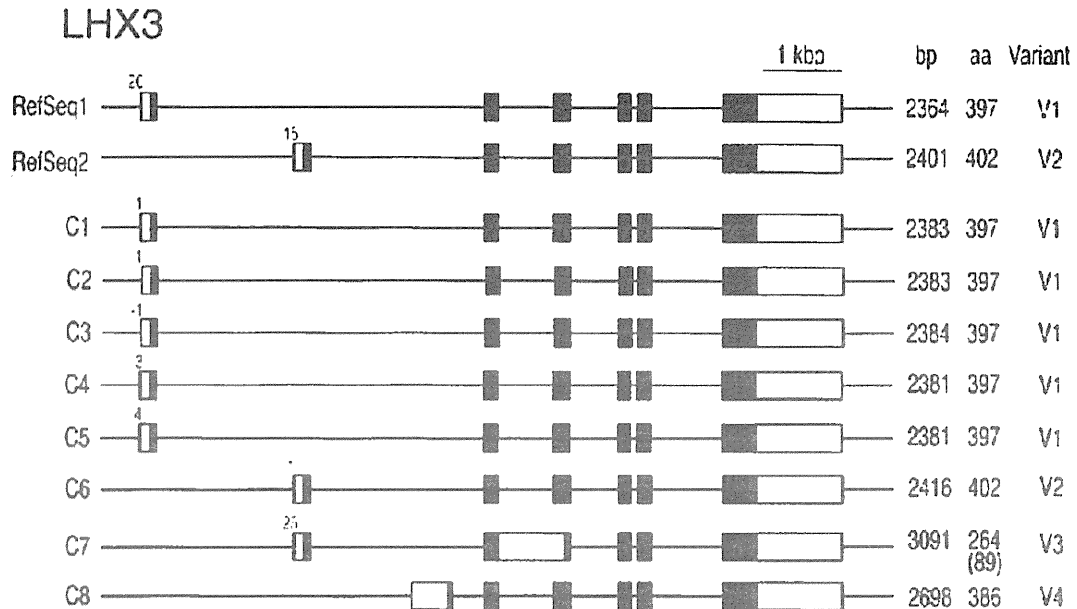


Figure 3. The exon-intron structure of alternative splicing variants for *LHX3*. Eight clones (C1-C8) were classified into four variants (V1-V4). The relative position of a transcription start site was indicated on exon 1. RefSeq1 and RefSeq2 correspond to GenBank Accession No. NM\_178138.4 and NM\_014564.3, respectively. Our clones correspond to AB593042.1 - AB593055.1.

*LHX3* has two LIM domains, a homeodomain, and a C-terminal *LHX3*-specific domain. Three isoforms for human *LHX3* (isoform a, isoform b, short isoform) have been reported (Sloop et al., 2001). Isoforms a and b differ in their N-terminal sequence. The N-terminal sequence of isoform b has been shown to inhibit the binding of *LHX3* to DNA. Furthermore, isoform a produced a 264-aa short isoform starting at Met-134 and thus lacking LIM domains. This short isoform showed a transcription factor activity owing to the downstream region including a homeodomain. Interestingly, the longest ORF of the novel variant V3 encoded this short isoform. It is also interesting whether the N-terminal sequence of the novel isoform encoded by V4 affects the activity of *LHX3*. In the pituitary gland, the expression pattern of V1 and V2 differed between cell lines, suggesting that these variants play a different role in the regulation of gene expression during development of each cell type (Sloop et al., 2001). The role of these variants in development of the retina remains to be solved.

### 3.2.3. Neural Retina-Specific Leucine Zipper Protein (*NRL*)

*NRL* is a basic motif-leucine zipper transcription factor that plays an essential role in the differentiation of photoreceptor cells (Swaroop et al., 1992). The Y79 libraries contained seven clones encoding *NRL*, which were classified into five variants as shown in Figure 4. V1 corresponded to RefSeq1, and V2 had a shortened 3'-untranslated region (3'-UTR) due to the alternative use of a polyadenylation signal. V3, V4, and V5 used an alternative promoter located between exon 1 and exon 2. Although the 3'-end splice sites of a new exon 1 of these three variants were slightly different, these variants had the same ORF in exon 2, which encoded a 98-aa sequence starting from Met-140 in the isoform encoded by V1. When the