

201419061B

厚生労働科学研究費補助金
感覚器障害研究事業

「緑内障統合的分子診断法の確立と実証」
に関する研究

平成24年度～平成26年度 総合研究報告書

研究代表者 木下 茂

平成27(2015)年 3月

厚生労働科学研究費補助金
感覚器障害研究事業

「緑内障統合的分子診断法の確立と実証」
に関する研究

平成24年度～平成26年度 総合研究報告書

研究代表者 木下 茂

平成27(2015)年 3月

目 次

I.	総合研究報告 「緑内障統合的分子診断法の確立と実証」に関する研究 木下 茂、田代 啓、長崎 生光、田中 光一、森 和彦、中野 正和・・・	1
II.	研究成果の刊行に関する一覧表	81
III.	研究成果の刊行物・別刷	91

[I]

総合研究報告

厚生労働科学研究費補助金（感覚器障害研究事業）
総合研究報告書

「緑内障統合的分子診断法の確立と実証」

に関する研究

研究代表者 木下 茂

京都府立医科大学大学院医学研究科 視覚機能再生外科学 教授

研究要旨

緑内障は本邦における中途失明原因の第1位であり、40歳以上の有病率は広義の原発開放隅角緑内障（広義POAG）では3.9%、70歳代では8.0%と加齢に伴い増加する。緑内障の発症を簡便な血液検査で予測できれば早期診断・治療により失明が予防でき、視覚障害に伴う医療費や社会福祉関連費の大幅な削減が期待できる。我々は、5,800例を超す血液検体（細胞株樹立済み）と過去15年以上蓄積してきた臨床情報をもとに、緑内障の主要病型についてのゲノムワイド関連解析（GWAS）を行い、緑内障患者に特有な塩基配列の違い（バリエント）を独自に見出し、そのデータを基に緑内障発症リスクを予測する診断アルゴリズムを開発し、特許出願している（特願 2010-294176）。

本研究では、臨床情報をゲノム情報に結びつけて解析しやすい情報管理環境の整備を行い、臨床情報の強化を図った。またこれまで1塩基の違いのバリエント（SNP）に限定していたゲノム情報について、大規模計算用の解析サーバー環境を独自に構築し、新たにコピー数の違いのバリエント（copy number variant, CNV）情報を取得し、抽出された高精度なコピー数データに基づくGWASを実施することによって、広義POAGおよび落屑緑内障に関連するCNV

の同定を試みた。またPOAGの発症に関与することが示唆されているグルタミン酸トランスポーターSLC1A3に着目したCNVの解析も実施した。これらの情報を活用し、日本人緑内障患者に関連する複数の候補バリエントからクラスタリング解析を用いた絞り込みを実施した。緑内障発症リスクの判定は、機械学習法のひとつであるサポートベクターマシーン (SVM) を応用し、検診チップに搭載するバリエントの組み合わせを含めた分子診断アルゴリズムを確立した。検診チップの有用性の評価には10交差検証を用い、広義POAGは71.1%、落屑緑内障では81.5%の高い判別率を得た。

本研究を実施したことにより、緑内障の主要な病型に関連する厳選されたバリエントを組み合わせる分子診断アルゴリズムを確立することにより、緑内障発症リスクを判定する検診チップの基盤が確立できたため、開発研究から実用化に向けたプロセスへの今後の速やかな展開が切望される。

分担研究者

田代 啓 (京都府立医科大学
大学院 医学研究科
ゲノム医科学 教授)

長崎 生光 (京都府立医科大学
大学院 医学研究科
基礎統計学 教授)

田中 光一 (東京医科歯科大学
大学院 疾患生命科学
研究部 分子神経科学
教授)

森 和彦 (京都府立医科大学
大学院 医学研究科
視覚機能再生外科学
講師)

中野 正和 (京都府立医科大学
大学院 医学研究科
ゲノム医科学 准教授)

A. 研究目的

緑内障は本邦における中途失明原因の第1位であり、40歳以上の有病

率は2型糖尿病と同程度に高い5.0%で、この内の大半が自覚症状のない潜在的な緑内障患者である。緑内障は発症初期の点眼治療により進行を抑制し生涯にわたり視機能を維持することができることから、緑内障の発症リスク保因者を早期に識別する簡便な血液検査体制の構築が急務である。

そこで本研究では、我々が独自に保有する5,800例を超す緑内障患者および非緑内障健常人の血液検体（細胞株も樹立）と過去15年以上にわたり蓄積された臨床情報をもとに同定した日本人緑内障患者に特有のゲノム配列の違い（バリエーション）情報に基づく分子診断アルゴリズムを確立し、緑内障発症リスク検診チップの有用性を実証することを最終目標とする。

B. 研究方法

本研究では、緑内障の主要な病型である広義の原発開放隅角緑内障

（広義POAG）および落屑緑内障について、我々が保有する多数の検体を用いて高密度チップにより取得したジェノタイプ情報に基づくゲノムワイド関連解析（GWAS）を実施し、日本人緑内障患者に関連するバリエーションを確定させる。加えて、緑内障は多因子疾患であり関連するバリエーションのオッズ比が低いことが想定されることから、1塩基の違いのバリエーション（SNP）だけでなく、コピー数の違いのバリエーション（CNVおよびCNP）についても精査し、多角的にバリエーションを抽出する。また、グルタミン酸トランスポーター上のバリエーションの解析を実施し、緑内障に関連する複数の候補バリエーションの抽出を試みる。更に、循環血液中の白血球の網羅的発現解析による連続変数の検討も加える。得られた全ての変数を加味しながら、機械学習法のひとつであるサポートベクターマシン（support vector machine, SVM）を応用し、検診チップに搭

載するSNPやCNPの組み合わせ等の変数を検討し、緑内障統合的診断アルゴリズムの確立と緑内障発症リスクを判定する検診チップの有用性を実証する。

1. 緑内障臨床データサンプル取得と臨床データ蓄積

我々は15年以上にわたり、緑内障専門外来の患者のデータベースを構築し、専用カスタマイズしたソフトを用いて患者の臨床データを入力、管理している。緑内障専門外来では視力、眼圧、隅角、視野などの基本的な緑内障検査以外にも緑内障精密機器（3D-OCT、無散瞳眼底カメラ、GDx、HRT-II、マトリックス、ペンタカム、ビサンテ、スペキュラ、眼軸など）を実施し、緑内障の正確な診断および詳細な経過観察を行った。また正常スクリーニング外来ではボランティアで参加を募集し、参加者に対して緑内障専門外来と同様の精密な診察および検査を行い正常、

緑内障疑い、緑内障に分類し、さらに正常は視神経乳頭の形状から三段階にグレード分類（super normal, fairly normal, normal）し、より確かな非緑内障健常人の検体を得た。

2. 多検体高密度ゲノムワイド関連解析と既報の候補遺伝子バリエーションの検証

（1）高密度チップによるゲノムワイド関連解析

2005年以来、我々が独自に収集している5,000例を超す臨床検体の中から総計2,126検体を選別し、アフメトリクス社のDNAマイクロアレイ（Genome-Wide Human SNP Array 6.0）を用いて906,600個のSNPのジェノタイプを決定した。次に、緑内障専門医による厳格な診断により広義POAG群824例と対照群686例を厳選し、厳格なフィルターにより抽出した653,519個の高精度なジェノタイプデータを用いて相関解析を実施した。更に、眼圧

データに基づいて疾患群を眼圧の高い群（HPG）と低い群（NPG）の2群に細分類した相関解析も実施した。

また、落屑緑内障においても、緑内障専門医による厳格な診断により落屑緑内障群（落屑症候群症例を含む）201例と対照群697例を厳選し、厳格なフィルターにより抽出した652,792個の高精度なデータを用いて相関解析を実施した。（2）既報の候補遺伝子上のバリアントの検証

我々独自の広義POAG検体（症例群521例・対照群519例および症例群431例・対照群416例の2ステージ）を用いて、既報の緑内障関連遺伝子上のバリアントを搭載したカスタムチップ（612SNP/24遺伝子、イルミナ社製カスタムチップ）を構築し検証した。

3. コピー数の違いのバリアントデータの取得と解析

（1）広義POAGと落屑緑内障に関連するコピー数バリアントの同定に

向けた高精度データの取得

広義POAG群824例と対照群686例、落屑緑内障群201例と対照群697例の各病型について、アフィメトリクス社のDNAマイクロアレイ（Genome-Wide Human SNP Array 60）で取得した蛍光強度の生データ（約100万個のSNP/CNV兼用プローブと約100万個のCNV専用プローブ由来の総計約200万個のプローブデータ）をコピー数データ（0、1、2、3、4）に変換した。コピー数データの取得には、アフィメトリクス社が提供しているGenotyping Console 4.1.1を用い、国際HapMapプロジェクトで公開されている270検体のコピー数データを基準とした。取得した大量のCNVデータについては、独自に構築した専用サーバー上にデータベース化し、取得したデータをもとにCNVおよびヘテロ接合性欠失（LOH）の有無の同定を実施するための大容量計算が可能な解析サー

バーを構築した。

(2) コピー数バリエーション情報に基づく広義POAGおよび落屑緑内障のGWAS

広義POAG群824例および対照群686例、落屑緑内障群201例および対照群697例について取得した全ゲノムにわたる0から4までのCNVのコピー数データに基づき、ウィルコクソンの符号順位検定（コピー数の増減数による有意差検定）およびフィッシャーの正確確率検定（コピー数の増減数、増減、増または減による有意差検定）を用いて広義POAG、落屑緑内障に関連するCNVの検出を実施した。

(3) コピー数の違いの多型情報に基づく緑内障のGWAS

広義POAG群824例および対照群686例、落屑緑内障群201例および対照群697例について、コピー数の違いの多型（CNP）情報を以下の手順で取得した。

アフィメトリクス社のDNAマイ

ク ロ ア レ イ （ Genome-Wide Human SNP Array 6.0）で取得した蛍光強度の生データ（約100万個のSNP/CNV兼用プローブと約100万個のCNV専用プローブ由来の総計約200万個のプローブデータ）を基に、Canaryアルゴリズムを用い、3. (2) で取得するCNV領域である1,141領域をCNPデータ（0, 1, 2, 3, 4）に変換した。データの取得には、アフィメトリクス社が提供のGenotyping Console 4.1.1（GTC）を用い、国際HapMapプロジェクトで公開されている270検体のコピー数データを基準とした。

CNPデータに基づき、ウィルコクソンの符号順位検定（コピー数の増減数による有意差検定）およびフィッシャーの正確確率検定（コピー数の増減数、増減、増または減による有意差検定）を用いて広義POAGおよび落屑緑内障に関連するCNPの検出を実施した。

4. 循環血液中の白血球mRNAの網羅的遺伝子発現解析

網羅的遺伝子発現解析データは、緑内障群2例・対照群2例について微量な発現変動を捉えることに定評のあるチップ（東レ・3D-Gene）を用いて取得した。

5. グルタミン酸トランスポーター上のバリエーションの解析

(1) 緑内障発症に関連するグルタミン酸トランスポーターrareバリエーションの同定と機能解析

440例の緑内障患者と450例の健康人のゲノムDNAを用い、グルタミン酸トランスポーター*SLC1A3*の10個のエクソンをPCRで増幅し、塩基配列を決定した。*SLC1A3*のcDNAにPCRを用いて緑内障患者に見出した4種類のアミノ酸変異を伴う点変異を導入し、HEK239Tに遺伝子導入し、グルタミン酸取り込み活性を測定した。グルタミン酸取り込み活性が低下した*SLC1A3*のrare

バリエーションに関しては、TALENを用いてノックインマウスを作製した。

(2) 緑内障発症に関連するグルタミン酸トランスポーターrareバリエーションの機能解析とそれを標的にした新規治療薬の開発

グルタミン酸取り込み活性を低下させる*SLC1A3*の2種類のrareバリエーション (A169G, A329T)を再現したノックインマウスを用いて、膜神経節細胞の変性の有無を解析した。また、*SLC1A3*の発現量を増加させる化合物を検索し、同定された化合物の網膜神経節細胞保護効果を、緑内障モデルを用い評価した。

6. 分子診断アルゴリズムの確立と検診チップの有用性の実証

(1) GWASによる解析対象バリエーションの選別と数値変換

広義POAGおよび落屑緑内障に関連する統計学的検出力の高いSNPに基づくGWASおよびCNPのGWAS結果から高品質なデータを抽出した。

次に、本課題で検討された変数を考慮し、GWASによる有意水準の閾値を決定し、解析対象バリエントを選別した。

アルゴリズムに供するSNPの数値変換は以下の手順を用いた。

1. Risk Allele Homo = 1、Risk Allele Hetero = 2、Other Allele Homo = 0とした。

2. 数値変換は広義POAG群と対照群、および落屑緑内障群と対照群の各群で、平均値および観測アレル頻度を用いて数値の標準化を式①を用いた。正規化後の欠損値は0とした。

$$M(i, j) = \frac{C(i, j) - \mu(j)}{\sqrt{p(j)(1 - p(j))}} \quad \dots \textcircled{1}$$

- ・ i は検体 (1, ..., m), j は SNP (1, ..., n)
- ・ $M(i, j)$ は正規化後 Matrix
- ・ $C(i, j)$ は Genotype Data Matrix (0, 1 or 2)
- ・ $\mu(j)$ は各 SNP j の平均値
- ・ $p(j)$ は各 SNP j の Allele Frequency

同様に、アルゴリズムに供するCNPの数値変換は以下の手順を用いた。

1. マイクロアレイで取得した蛍光強度をコピー数データ0, 1, 2, 3, 4に変換した。

2. 数値変換は広義POAG群と対照群、および落屑緑内障群と対照群の各群で、平均値および観測アレル頻度を用いて数値の標準化を式②を用いた。正規化後の欠損値は0とした。

$$M(i, j) = \frac{C(i, j) - m(j)}{s(j)} \quad \dots \textcircled{2}$$

- ・ i は検体(1, ..., m), j は CNV (1, ..., n)
- ・ $M(i, j)$ は正規化後 Matrix
- ・ $C(i, j)$ は CNV Data Matrix (0, 1, 2, 3, 4)
- ・ $m(j)$ は各 CNV j の平均値
- ・ $p(j)$ は各 CNV j の Allele Frequency

(2) クラスタリングによる検診チップに搭載するバリエントの絞り込み

アルゴリズム作成に向けて多重共線性や過学習を避け、汎化能力を高めるため手段としてクラスタリング解析を実施した。

SNPの絞り込みで用いたクラスタリングの手法は、GWASを実施した全検体を用いて、検体で情報縮約す

ること得られる数値を算出した。次にその値と遺伝的な連鎖を参考にして相互に強い相関が認められるバリエーションについては、クラスターを作成した。その後、各クラスターからGWASのp値を基準に代表バリエーションを決定し、検診チップに搭載すべき候補バリエーションとした。

CNPに関しては、GWASで用いるバリエーションが1,141個に限られているため、クラスタリング解析を用いず、検定結果の上位のバリエーションを候補バリエーションとした。

(3) アルゴリズムの検討

アルゴリズムで用いる機械学習法の検討を広義POAGのSNPを用いて検討した。検討に用いた機械学習法は、最近隣法（k・nearest neighbor method; k=10）、分類木、重回帰分析、ロジスティック回帰分析、SVM（カーネル関数：Radial Basis Function）とした。各種機械学習法に用いたデータは、ランダムに選んだ学習データ（広義

POAG650例 vs 対照群650例）、テストデータは残りの（広義POAG174例 vs 対象群36例）を用いた。判別率の検討に用いたバリエーションはオッズに高いものから順に使用した。

(4) 検診チップに搭載するバリエーションの組み合わせ

検診チップに搭載するバリエーションについて、広義POAGおよび落屑緑内障の発症リスクを判別するバリエーションの組み合わせの検討をした。SNPについては、クラスタリングにて絞り込みを行った候補バリエーションからGWASのp値を基準に組み合わせを用い、SVMの判別率を基準とした。また、すべてのCVPの候補バリエーションをSNPと組み合わせで判別率の向上について検討した。検診チップの評価は、ランダムサンプリングにより広義POAG群および対照群を680検体ずつ選び、同様に落屑緑内障群200例と対照群200例を選んだ後、10交差検証による判別率か

ら実証した。

(倫理面への配慮)

1. 法令遵守について

本研究ではヒトゲノム配列情報を取り扱うため、ヘルシンキ宣言の精神を尊重し、ヒトゲノム・遺伝子解析研究に関する3省合意規定に従った。本学の「ヒトゲノム・遺伝子解析研究倫理審査委員会」の承認を得ており、承認事項を遵守して研究を遂行した。本学の承認に当たっては、本研究の実施担当者全てがCITI Japan Programに定める講習を受講し修了者名簿に登録済である。

2. 個人情報の保護について

本研究では、緑内障患者および非緑内障健常人（血液提供者）から研究への参加に対する自由意志による同意を文書により得た上で血液を採取した。採取した血液検体は、医師・採血者および実験従事者とは第三者的立場の個人情報管理者により、

個人情報管理者のみが入室できる匿名化管理室内の専用コンピュータにより暗号化した。暗号化された血液検体を実験従事者が受領した後に、ゲノムDNAを抽出しジェノタイプデータを取得した。また本研究では、本学倫理委員会から臨床情報とゲノム情報とを連結して解析すること（連結可能匿名化）の承認も得ており、個人情報の保護と厳重な漏洩防止に努めながら研究を遂行した。

C. 研究結果

1. 緑内障臨床データサンプル取得と臨床データ蓄積

2015年1月末の時点で、広義POAG、落屑緑内障、その他緑内障病型も含めた症例3,618例と正常コントロール2,225例の合計5,843例のゲノムサンプルならびに臨床データを収集することができた。

2. 多検体高密度GWASと既報の候補遺伝子バリエーションの検証

(1) 高密度チップによるGWAS

広義POAG群と対照群を用いたGWASの結果を示す(図1)。GWASの結果、ヒト染色体上の9p21.3領域に存在する*CDKN2B-ASI*上にボンフェローニ補正による有意なSNPが集中的に同定された。

広義POAG群の検体で、眼圧の高いHPG群330例と正常範囲のNPG群503例の2群に分けた相関解析を実施した結果、広義POAGのGWASで検出されたSNPと同一の*CDKN2B-ASI*上のSNPがNPG群でのみ有意に検出された。

次に、落屑緑内障群193例と対照群686例のGWASの結果を示す(図2)。GWASの結果、落屑症候群に関連する*LOXLI*上のバリエーション以外に、2つの遺伝子上から新規のバリエーションを同定した。

(2) 既報の候補遺伝子上のバリエーションの検証

大規模な家系データを用いた連鎖

解析によって知られている緑内障関連遺伝子である*MYOC*、*OPTN*、*WDR36*の3遺伝子と、緑内障との関連が考えられる21候補遺伝子を検討した(表1)。我々独自の検体(日本人集団)において、バリエーションの60%以上は対立遺伝子が存在しない単型(対立遺伝子頻度が0)であった。

表1. 候補遺伝子上のバリエーションと分析SNP数

Gene	Locus	Designed	Analyzed
MYOC	1q24	57	19
OPTN	10p13	34	24
WDR36	5q22	63	20
Subtotal (3 genes)		154	63
GSTM1	1p13	12	8
NPPA	1p36	12	1
MTHFR	1p36	29	12
OCLM	1q31	5	0
OPTC	1q32	6	2
CYP1B1	2p22	53	8
IL1B	2q13	11	2
IL1A	2q13	14	6
OPA1	3q29	21	9
EDNRA	4q31	33	6
ADRB2	5q32	28	9
TAP1	6p21	17	6
CDKN1A	6p21	16	5
TNF	6p21	31	2
NOS3	7q36	19	4
ADRB1	10q25	22	2
IGF2	11p15	42	4
TP53	17p13	43	3
APOE	19q13	22	1
OLFM2	19p13	8	1
GSTT1	22q11	14	0
Subtotal (21 genes)		458	91
Total (24 genes)		612	154

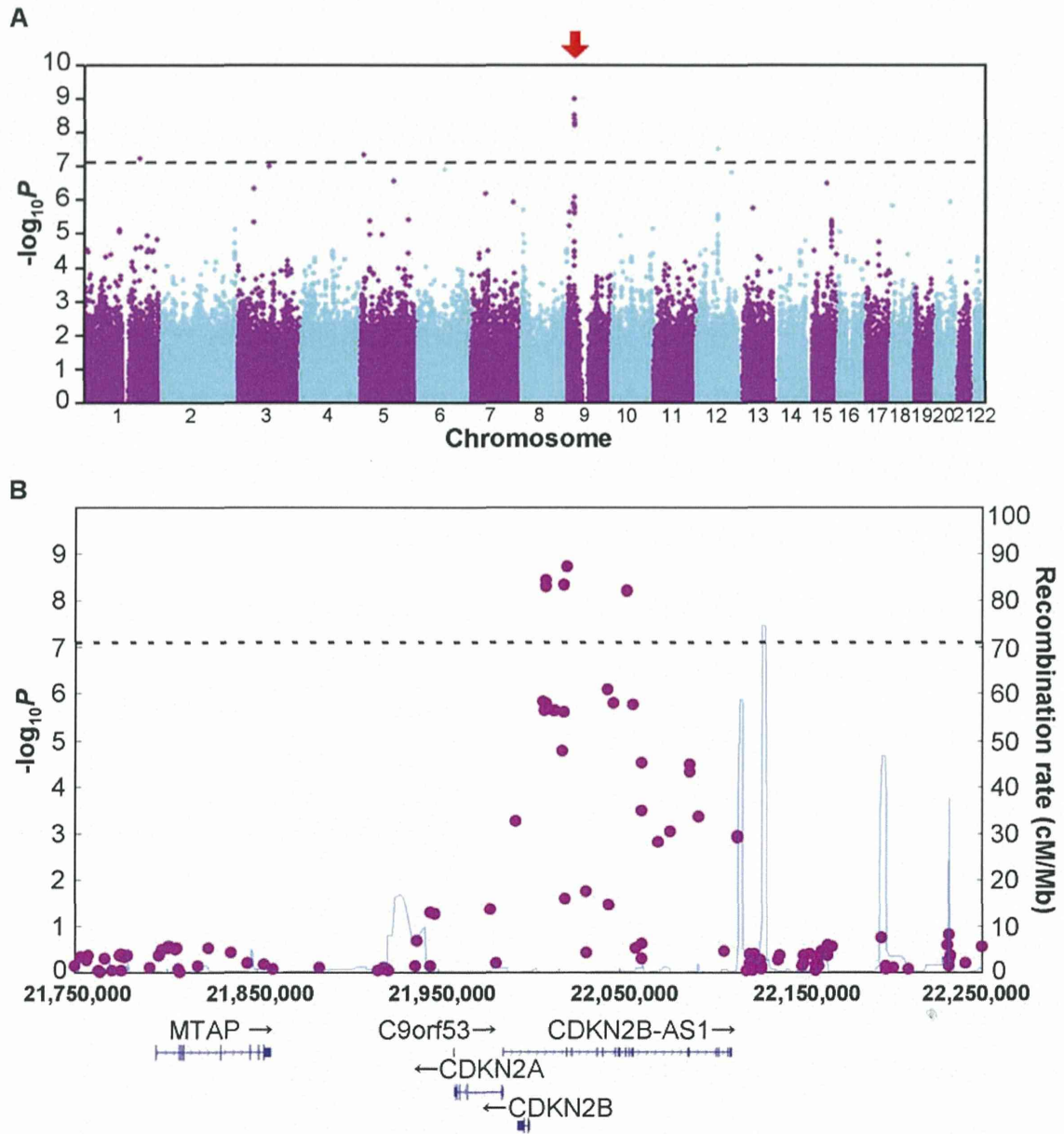


図1. 広義POAG群と対照群のGWAS結果. (A) 有意水準の高いSNPが9番染色体から検出された (赤矢印). (B) 9p21領域の拡大図. 破線はボンフェローニ補正による有意水準を示す. (PLoS One, 7: e33389, 2012)

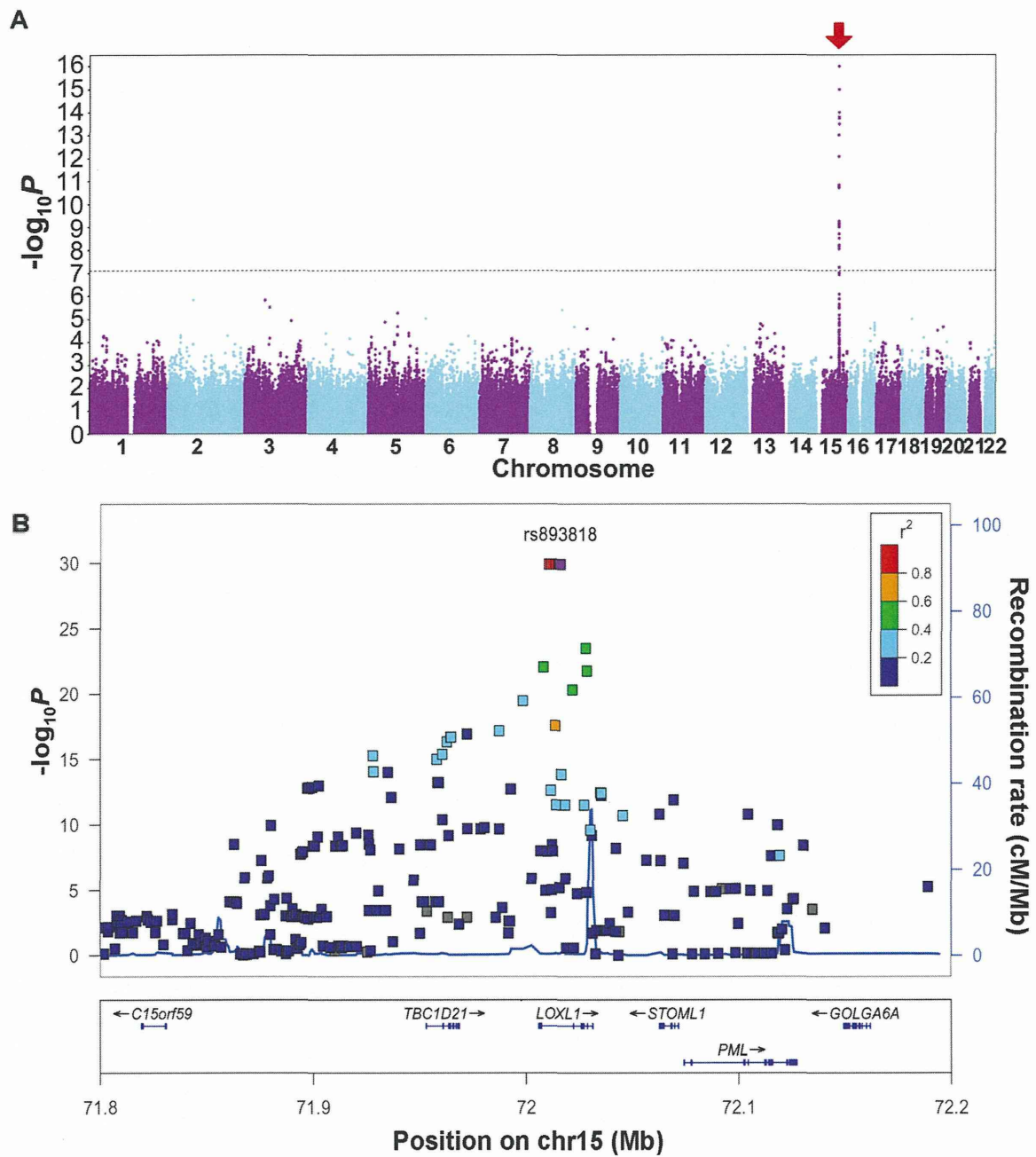


図2. 落屑緑内障群と対照群のGWAS結果. (A) 有意水準の高いSNPが15番染色体から検出された (赤矢印). 破線はボンフェローニ補正による有意水準を示す. (B) 15q24.1領域の拡大図. 破線はボンフェローニ補正による有意水準を示す. (Scientific Reports, 4: 5340, 2014.)

既報の緑内障に関連する *WDR36* の数個のバリエントが弱いながらも相関を示した ($p < 0.05$)。しかし、我々の検体群ではいずれのバリエントも関連性が低く、検診チップに搭載する必要性に乏しいことが判明した (図3)。

3. コピー数の違いのバリエントデータの取得と解析

(1) 広義POAGと落屑緑内障に関連するコピー数バリエントの同定に向けた高精度データの取得

総計2,408検体 (広義POAG群824例と対照群686例、落屑緑内障群201例と対照群697例) の各検体が約200万プローブの情報を持ち、約48億個の高精度なCNVデータを取得した。取得したデータは、GWASで実施する各種統計処理に耐え得る専用データベースを構築し、格納した。

(2) コピー数バリエント情報に基づく広義POAGおよび落屑緑内障の

GWAS

種々の検定方法を用いたGWASの結果、広義POAGに関連することが強く示唆されるボンフェローニ補正による有意なCNVが200個以上のプローブから検出された。コピー数型が「0」と「1」を第1分類、「2」を第2分類、「3」と「4」を第3分類とし、フィッシャーの正確確率検定を用いたGWASを図4. Aに示す。

更に、これらのプローブの染色体における位置情報を解析した結果、同定されたCNVはゲノム上の約50領域にも及び、数十個の遺伝子にまたがっていることが明らかになった。

同様に、落屑緑内障に関連することが強く示唆されるボンフェローニ補正による有意なCNVが250個以上のプローブから検出された。コピー数型が「0」と「1」を第1分類、「2」を第2分類、「3」と「4」を第3分類とし、フィッシャーの正確確率検定を用いたGWASを図4. Bに示す。更に、これらのプローブの

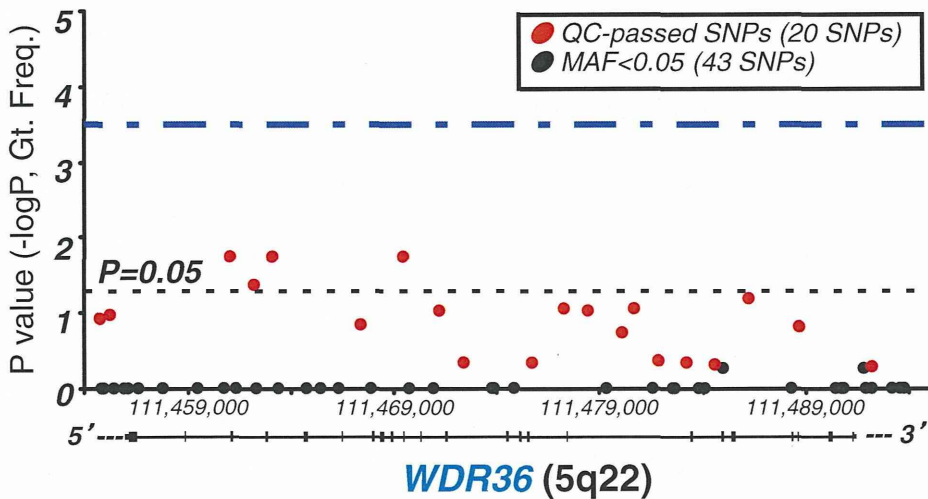
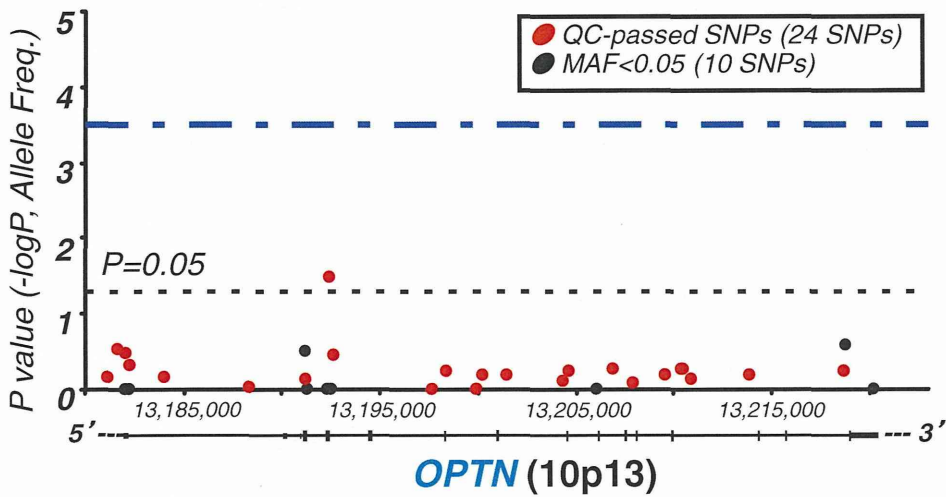
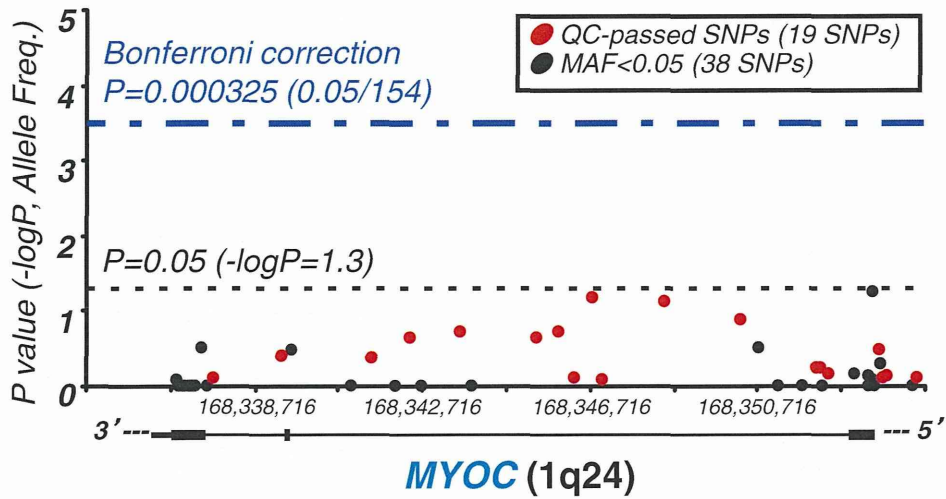


図3. 既知のバリエーションの検証. 青の破線はボンフェローニ補正による有意水準を示す.

染色体における位置情報を解析した結果、同定されたCNVはゲノム上の約30領域にも及び、数十個の遺伝子にまたがっていることが明らかになった。

(3) コピー数の違いの多型情報に基づく緑内障のGWAS

広義POAGについてはボンフェローニ補正による有意なCNPが5個の領域から検出された(図5. A)。一方、落屑緑内障のGWASの結果は、ボンフェローニ補正による有意なCNPが2個の領域から検出された(図5. B)。

CNV情報に基づくGWASとの比較は、広義POAGでは共通する領域が確認できなかったが、落屑緑内障についてはCNPが同定された1領域内の連続する5個のプローブ(103塩基)が共通であることが判明した。

4. 循環血液中の白血球mRNAの網羅的遺伝子発現解析

網羅的遺伝子発現解析については、

緑内障群2例・対照群2例の白血球由来のmRNAを用いて取得したデータに基づきクラスター解析やアノテーション解析を実施したものの、各検体の白血球mRNAの遺伝子群の発現にばらつきが多く、症例群に特異的な遺伝子発現ネットワークを同定することは困難であった。

5. グルタミン酸トランスポーター上のバリエーションの解析

(1) 緑内障発症に関連するグルタミン酸トランスポーターrareバリエーションの同定と機能解析

緑内障患者群において440例中20例、健常群において450例中2例のヘテロ接合の一塩基置換が見出された。*SLC1A3*のexon4に1つのサイレント変異1・1つのミスセンス変異、exon6に1つのサイレント変異、exon7に2つのサイレント変異・2つのミスセンス変異、exon10に1つのサイレント変異を発見した。このうちexon7に見いだされたc945 C>T