

13 集団の PGx 関連遺伝子バリエーションの大規模アノテーションデータベースの構築

福永 航也

蒔田 泰誠

13 集団の PGx 関連遺伝子バリエーションの 大規模アノテーションデータベースの構築

The establishment of large-scale database for the biomarkers of PGx-related genes in 13 populations

福永 航也*¹ 薊田 泰誠*²

Key words : 遺伝的多様性、次世代シーケンサー、ファーマコゲノミクス、CPIC ガイドライン、PGx 関連遺伝子

要 旨

目的 アジアを中心とした 13 集団における、薬剤の効果や有害事象の発現リスクに関係するファーマコゲノミクス (PGx) 関連遺伝子バリエーションの大規模データベースを構築する。

方法 13 カ国から収集された 2,998 人のゲノムを対象に、PGx 関連 13 遺伝子のアレル決定およびフェノタイプ推定を行い、標準的な薬物治療 (20 種類) と異なる用法・用量が推奨される患者の頻度を各集団間で比較した。

結果 PGx 関連遺伝子のアレルおよびフェノタイプ頻度を比較したところ、最大の頻度差は VKORC1 のワルファリン感受性型で見られた (ナイジェリア 94.2%、中国および日本 0.0%)。また、標準治療とは異なる用法・用量が 1 種類以上の薬剤で推奨される患者は全体の 97.2% にのぼった。

結論 13 集団における PGx バイオマーカーの機能的なアノテーションが付随したデータベースが構築され、世界的な精密医療の普及に貢献することが期待される。

はじめに

ゲノム解析技術やそれを活用した研究開発の急速な進展により、遺伝要因による個人ごとの違いを考慮した精密医療の実現への期待が高まっている。特に、ファーマコゲノミクス (PGx) バイオマーカーには薬剤の効果や有害事象リスクを投薬前に予測することができる有用なマーカーが多く、そのほとんどが薬物動態関連遺伝子とヒト白血球抗原 (HLA) 遺伝子である¹⁾。しかし、これらの遺伝子はジェノタイピングすることが困難な遺伝子領域に存在するため、ゴールドスタンダードとなり得る網羅的なジェノタイピング方法は存在しないことが問題になっている。我々は以前に網羅的かつ正確に PGx 関連遺伝子のゲノムデータを取得可能な、

次世代シーケンサー (NGS) を用いた手法である PKseq パネルと corePGseq パネルを開発した²⁾。PKseq パネルを用い、990 人の日本人由来ゲノムをシーケンシングしたところ、19 遺伝子中にアミノ酸置換を伴う新規バリエーションが 142 個も同定されている²⁾。このように全ゲノムシーケンシングやロングリードシーケンシング技術が進歩した現代においても PGx 関連遺伝子の新規バリエーションは今なお同定され続けている。また、PGx 関連遺伝子のバリエーション頻度は集団間によって大きく異なっている³⁾。そのため PGx 関連遺伝子のバリエーション情報に関する研究の多くは、単一の集団での報告にとどまっており、様々な集団を統一されたプラットフォームで PGx 関連遺伝子をシーケンシングした事例は皆無である。このため

*¹FUKUNAGA KOYA 国立研究開発法人理化学研究所 生命医科学研究センター ファーマコゲノミクス研究チーム

*²MUSHIRODA TAISEI 同 上

報告によってバリエーション頻度が大きく異なり、集団間の差異であるのか、プラットフォームの精度の問題であるのか判断することができないことも問題である。

PGx バイオマーカーは臨床現場でも有用であるため、NUDT15 と難治性炎症性腸疾患および急性リンパ性白血病治療薬チオプリン製剤、UGT1A1 と抗がん剤イリノテカン、CYP2C9 と多発性硬化症治療薬シボニモドが本邦でも保険収載されている。遺伝子検査によって NUDT15 が Cys/Cys 型と決定された患者はチオプリン製剤の減量または代替薬の使用が行われている。また世界的な PGx バイオマーカーのガイドラインとしては CPIC (Clinical Pharmacogenetic Implementation Consortium; <https://cpicpgx.org/> (accessed 2024-10-28)) ガイドラインが策定されている¹⁾。このガイドラインでも PGx 検査の結果に基づいて、有害事象の発現が高リスクの患者もしくは薬剤の反応性が低いと予想される患者には標準的な薬物治療とは異なる用法・用量が推奨されている。しかし、これらのガイドラインが様々な集団に適用することの是非は報告されていない。

本研究では、多様な集団の PGx 関連遺伝子のバリエーションのデータベースを構築するために、ギリシャ、ヨーロッパ、アラブ首長国連邦、ナイジェリア、中国、日本、インドネシア、ラオス、ミャンマー、マレーシア、フィリピン、タイ、ベトナムから収集された 2,998 人のゲノムを用いて PGx 関連遺伝子のアリルとフェノタイプを推定した。また薬剤の体内曝露量に関係する 13 遺伝子および 20 種類の薬剤を対象とした CPIC ガイドラインにおいて、異なる治療介入方法の対象となる患者の頻度の違いを各集団間で比較することによって PGx バイオマーカーのバリエーションの機能的なアノテーションを付随したデータベースを構築した。

方 法

対象患者

ゲノム DNA はギリシャ (n=304)、ヨーロッパ (n=100)、アラブ首長国連邦 (n=100)、ナイジェリア (n=120)、中国 (n=120)、日本 (n=1,087)、インドネシア (n=562)、ラオス (n=100)、ミャンマー (n=100)、マレーシア (n=105)、フィリピン (n=100)、タイ (n=100)、ベトナム (n=100) から収集された計 2,998 人を用いた。ヨーロッパ、ナイジェリア、中国、日本のゲノムはコリエル医学研究所から購入し、上記以外の集団は SEAPharm (Southeast Asian Pharmacogenomics Research Network)⁴⁾ プロジェクトから提供された。本研究は国立研究開発法人理化学研究所の人を対象とする研究に関する倫理審査委員会の承認を得て行われた。

ターゲットシーケンシング

PKseq パネルは PK 関連 100 遺伝子の重要なプロモーターとスプライシングサイトを含む全エクソンのゲノム領域を増幅するプライマーセットである。これらのプライマーセットを用いて 1st PCR を行った。1st PCR 産物の個体識別のためにバーコードとミックスし 2nd PCR を行った。2nd PCR 産物を AMPure XP (サーモフィッシャーサイエンティフィック) を用いて精製した。精製産物の断片長を確認するために 2100 バイオアナライザ (アジレントテクノロジー) で電気泳動を行った。精製産物の Real Time PCR を行い、ライブラリー濃度を測定した。精製産物を 10 pM に段階希釈後、MiSeq システム (イルミナ) と MiSeq Reagent Nano Kit v2 (500 サイクル、イルミナ) を用いてシーケンシングを行った。

fastq ファイル取得後、Trimmomatic⁵⁾ によってアダプターをトリミングし、クオリティの低い塩基を除去した。Burrows-Wheeler Aligner (0.7.17)⁶⁾ を用いてリードをヒト参照ゲノム (GRCh37/hg19) にマッピングした。Genome Analysis Toolkit (GATK version 3.5)⁷⁾ を用いて single nucleotide variants (SNVs) や short insertions and deletions (indels) をコールした。擬陽性のバリエーションを除く

ために Variant quality score recalibration (VQSR) を行った。

本研究では CPIC ガイドラインでアリル情報が定義されている ABCG2、CYP2B6、CYP2D6、CYP2C9、CYP2C19、CYP3A5、DPYD、NAT2、NUDT15、SLC01B1、TPMT、UGT1A1、VKORC1 からのバリエーション情報を使用した。

CYP2D6 のコピー数測定

CYP2D6 の欠失や複製を含むコピー数変異 (CNV) は TaqMan Copy Number Assay プローブ (サーモフィッシャーサイエンティフィック、Hs00010001_cn (exon 9) および Hs0483572_cn (intron 2)) を用いて行われた。コントロールには RNase P を用いた。7900HT Fast Real-Time PCR System (サーモフィッシャーサイエンティフィック) を用いてリアルタイム PCR を行った。得られたデータと Copy Caller software (サーモフィッシャーサイエンティフィック) を用いてコピー数を算出した。

PGx 関連遺伝子のアリル、ディプロタイプ決定およびフェノタイプ推定

CPIC ガイドラインの中で定義されているアリル、ディプロタイプ決定およびフェノタイプ推定の table を用いて PGx 関連 13 遺伝子 (ABCG2、CYP2B6、CYP2D6、CYP2C9、CYP2C19、CYP3A5、DPYD、NAT2、NUDT15、SLC01B1、TPMT、UGT1A1、VKORC1) のアリル、ディプロタイプ決定およびフェノタイプ推定を行った。まず、ターゲットシーケンシングで得られたバリエーションおよびコピー数アッセイで得られた CNV 情報を allele definition table と照らし合わせアリルを決定した。Allele functionality table において、Normal function/Unknown/Indeterminate に該当するアリルはすべて*1 アリルとした。また、ディプロタイプ決定時にハプロタイプ情報が矛盾したアリルは Unknown とした。Diplotype-phenotype table を用いてディプロタイプ情報から薬物代謝酵素の活性を

推定した。各アリルには CPIC ガイドラインに所属している PGx 研究の有識者が定義した予測酵素活性値であるアクティビティスコアが定義されているため、この値を基に各ディプロタイプのアクティビティスコアを算出した。CYP2B6、CYP2D6、CYP2C9、CYP2C19、CYP3A5、DPYD、NUDT15、TPMT、UGT1A1 はアクティビティスコアを基に Ultrarapid metabolizer (UM)、Rapid metabolizer (RM)、Normal metabolizer (NM)、Intermediate metabolizer (IM)、Poor metabolizer (PM) に分類した。ABCG2 および SLC01B1 は Increased function (Increased)、Normal function (Normal)、Decreased function (Decreased)、Poor function (Poor) に分類した。VKORC1 の各ディプロタイプはワルファリン感受性型 (Sensitive)、中間型 (Intermediate)、抵抗性型 (Resistant) と分類した。NAT2 遺伝子のアリルは CPIC ガイドラインで定義されていないため NAT2 命名サイト (https://nat.mbg.duth.gr/Human_NAT2_alleles.htm (accessed 2024-10-28)) を使用した。NAT2 の標準アリルは*4 であるため Normal function/Unknown/Indeterminate に該当するアリルはすべて*4 アリルとした。過去の報告⁸⁾を基に NAT2 ディプロタイプ情報から薬物代謝酵素の酵素活性 (Rapid acetylator : RA、Intermediate acetylator : IA、Slow acetylator : SA) を推定した。

2024 年 10 月現在 CPIC には 27 種類のガイドラインが策定されている。これらの中から、薬剤の曝露量に関係する遺伝子を対象とし、フェノタイプごとに推奨されている介入方法が明記されている 20 種類の薬剤 (アタザナビル、 β 遮断薬、シタロプラム / エスシタロプラム、クロピドグレル、エファビレンツ、フルオロピリミジン、フルボキサミン、NSAIDs、オンダンセトロン / トロピセトロン、オピオイド、パロキセチン、プロトンポンプ阻害薬、セルトラリン、タクロリムス、タモキシフェン、チオプリン製剤、三環系抗うつ薬、ベンラファキシン、ポリコナゾール、ボルチオキセチン) を対象とした。推奨されている介入方法は標準治療 (Normal dose)、投与量を増加する (Increased

dose)、投与量を減少する (Decreased dose)、代替薬を提示する (Alternate drug) に分類した。ディプロタイプ、フェノタイプおよび上述の介入方法の対象となる患者の頻度を算出し、棒グラフおよびヒートマップは R ソフトウェア (version 4.3.0) を用いて描画した。

結 果

ジェノタイピング結果

PKseq 解析によって 2,998 人のうち 2,878 人がカバー率 95% 以上であったため以後これらのサンプルを解析に用いた。CYP2D6 のコピー数解析において、2,998 人のうちギリシャからの 304 人およびインドネシアからの 562 人の CYP2D6 コピー数はソフトウェアが必要とするクオリティをクリアすることができなかったため 2,132 人のコピー数を決定した。バリエーションおよびコピー数の解析を統合後 13 集団からの 2,998 人のうち計 2,458 人の 12 遺伝子 (ABCG2、CYP2B6、CYP2C9、CYP2C19、CYP3A5、DPYD、NAT2、NUDT15、SLCO1B1、TPMT、UGT1A1、VKORC1) のディプロタイプを決定した (図 1)。また CYP2D6 のディプロタイプ情報が矛盾したサンプルを除き計 2,099 人の CYP2D6 ディプロタイプを決定しアクティビティスコアを算出した (図 2)。

13 集団中のディプロタイプおよびフェノタイプ頻度の比較

CPIC ガイドラインのアリル定義に従いバリエーションおよびコピー数のデータを用い ABCG2 (2 種類)、CYP2B6 (9 種類)、CYP2C9 (13 種類)、CYP2C19 (11 種類)、CYP2D6 (45 種類)、CYP3A5 (4 種類)、DPYD (5 種類)、NAT2 (5 種類)、NUDT15 (4 種類)、SLCO1B1 (10 種類)、TPMT (3 種類)、UGT1A1 (6 種類)、VKORC1 (2 種類) の計 119 アリルを決定した。CYP2D6 については、CPIC ガイドラインに従い各ディプロ

タイプのアクティビティスコアを算出しフェノタイプに変換した。決定したアリル頻度およびフェノタイプ頻度を集団ごとに比較した (図 1)。集団間で頻度差が大きかったアリルは CYP3A5*3 (80.3%、ヨーロッパ 96.5%、ナイジェリア 16.3%)、NAT2*5 (46.3%、アラブ首長国連邦 48.8%、日本 2.5%)、ABCG2 の T アリル (38.5%、フィリピン 38.5%、ナイジェリア 0.0%)、CYP2B6*6 (27.3%、インドネシア:43.4%、中国:16.1%)、CYP2C19*17 (25.5%、アラブ首長国連邦 25.9%、日本:0.4%) であった。集団間で頻度差が大きかったフェノタイプは VKORC1 のワルファリン感受性型 (94.2%、ナイジェリア 94.2%、中国および日本 0.0%)、CYP3A5 の PM (73.0%、ヨーロッパ 93.0%、ナイジェリア 20.0%)、ABCG2 の Normal function (63.0%、ナイジェリア 100.0%、フィリピン 37.0%)、UGT1A1 の NM (53.3%、中国 78.3%、ナイジェリア 25.0%)、NAT2 の SA (51.6%、アラブ首長国連邦 68.2%、中国 16.7%) であった。

CYP2D6 においてはアリル、アクティビティスコア、フェノタイプ頻度を集団ごとに比較した (図 2)。CYP2D6 のアリルのうち集団間で最も頻度差が大きかったアリルは*10 であった (36.9%、中国 37.7%、ヨーロッパ 0.8%)。また頻度差が 10% 以上あったアリルは*17 (18.0%、ナイジェリア 18.0%、中国、ヨーロッパ、日本、ラオス、ミャンマー、マレーシア、フィリピン、タイ、ベトナム 0.0%)、*36+*10 (14.7%、中国 14.7%、ヨーロッパ 0.0%)、*4 (13.8%、ヨーロッパ 14.1%、日本 0.3%)、*2 (12.5%、ラオス 21.5%、日本 9.0%)、*10x2 (11.3%、ベトナム 11.6%、中国 0.3%) の 5 種類であった。集団間で最も頻度差が大きかったアクティビティスコアは 0.5 であった (45.5%、ベトナム 48.2%、ヨーロッパ 2.7%)。フェノタイプのうち集団間で最も頻度差が大きかったフェノタイプは IM であった (39.6%、ベトナム 69.0%、アラブ首長国連邦 29.4%)。

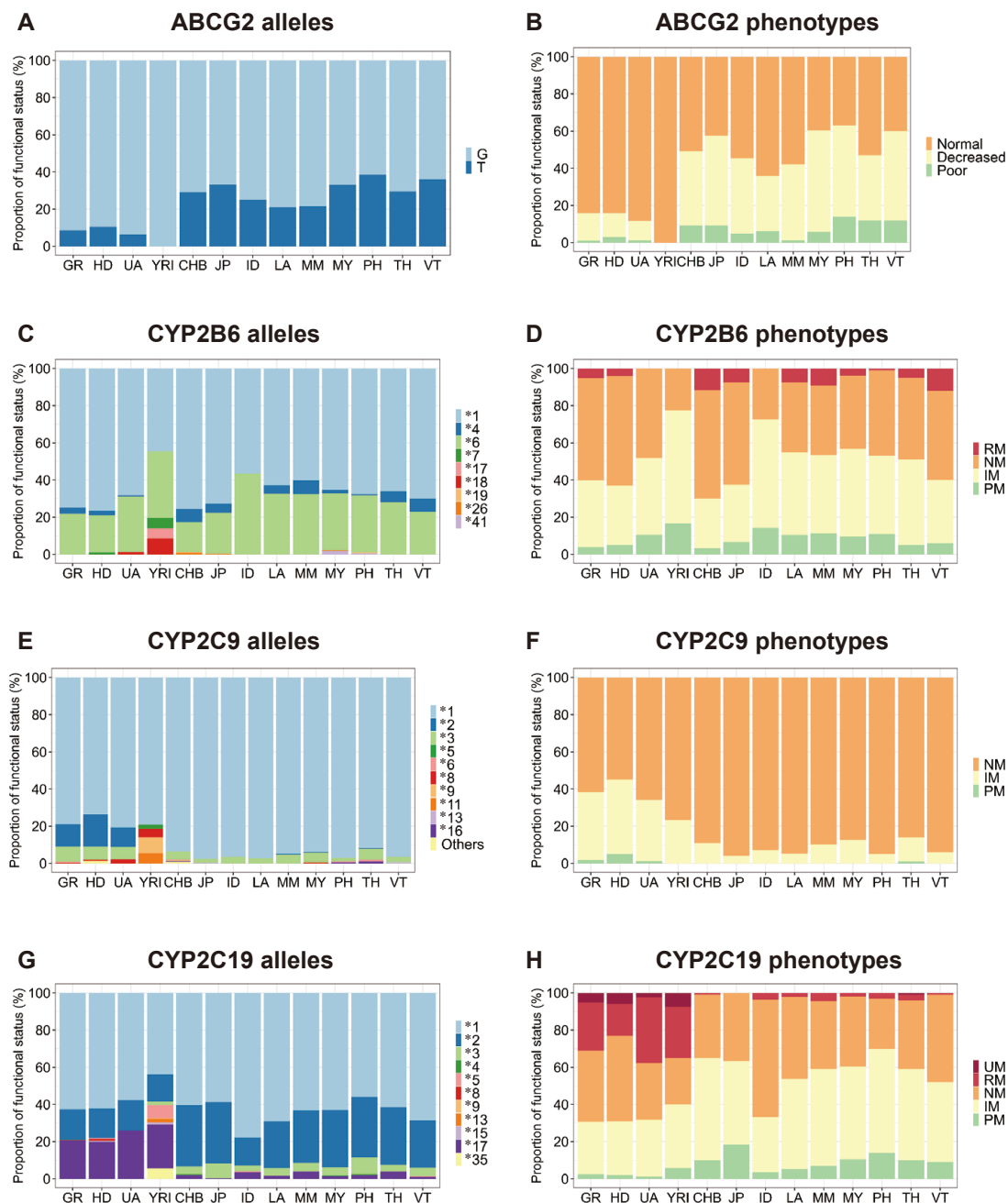
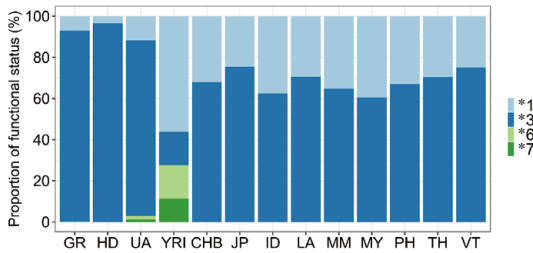
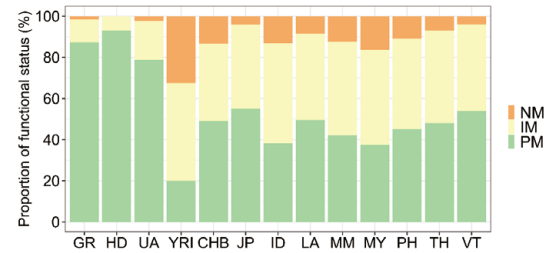


図1 13集団のファーマコゲノミクス関連12遺伝子のアレルおよびフェノタイプの頻度
PKseq パネルによってジェノタイピングされたバリエーション情報を PharmVar データベースのアレル情報を用いてアレルに変換し、各集団のアレル頻度を算出した (A、C、E、G、I、K、M、O、Q、S、U、W)。これらのアレルを CPIC ガイドラインの diplotype-phenotype table を用いて酵素活性の推定値であるアクティビティスコアを算出し、各集団のフェノタイプ頻度を算出した (B、D、F、H、J、L、N、P、R、T、V、X)。
GR: ギリシャ (n=274)、HD: ヨーロッパ (n=100)、UA: アラブ首長国連邦 (n=85)、YRI: ナイジェリア (n=120)、CHB: 中国 (n=120)、JP: 日本 (n=1,087)、ID: インドネシア (n=84)、LA: ラオス (n=95)、MM: ミャンマー (n=88)、MY: マレーシア (n=104)、PH: フィリピン (n=100)、TH: タイ (n=100)、VT: ベトナム (n=100)

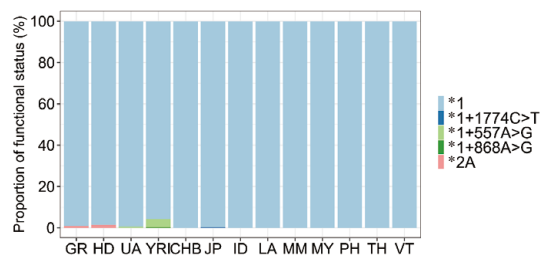
I CYP3A5 alleles



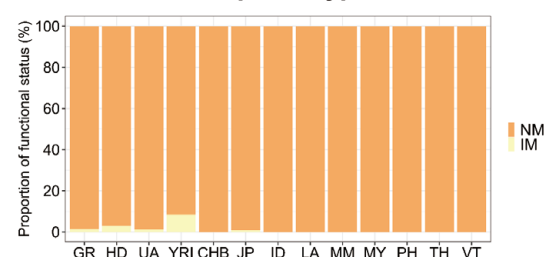
J CYP3A5 phenotypes



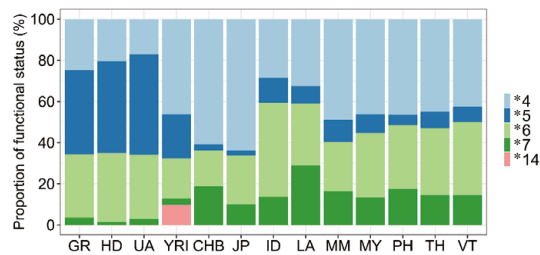
K DPYD alleles



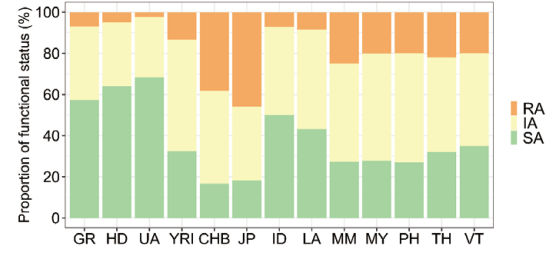
L DPYD phenotypes



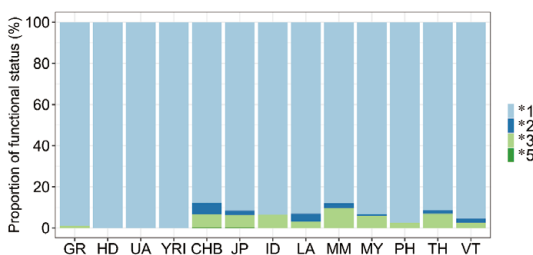
M NAT2 alleles



N NAT2 phenotypes



O NUDT15 alleles



P NUDT15 phenotypes

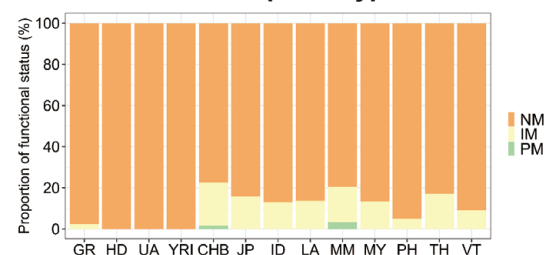


図 1 の続き

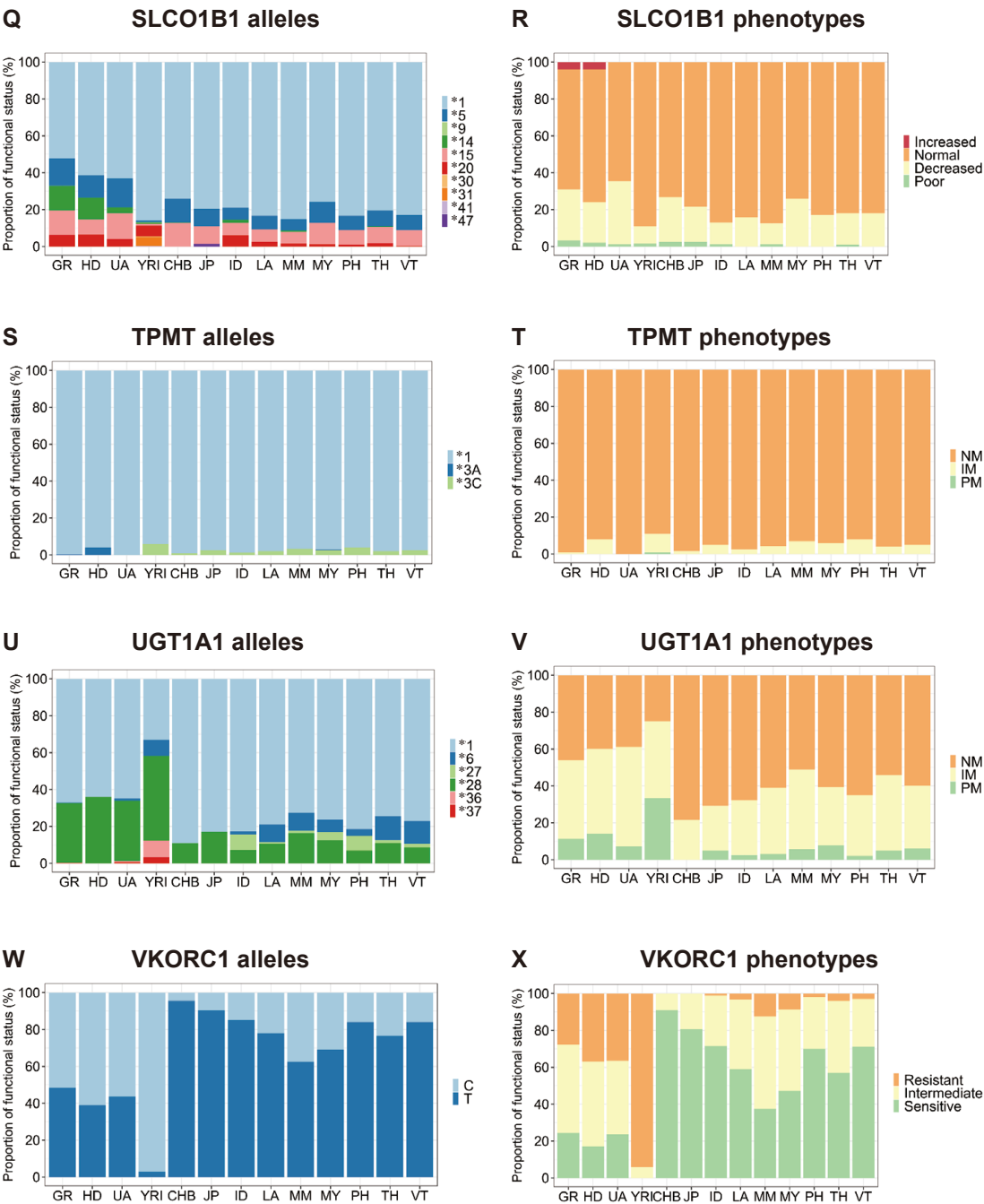


図 1 の続き

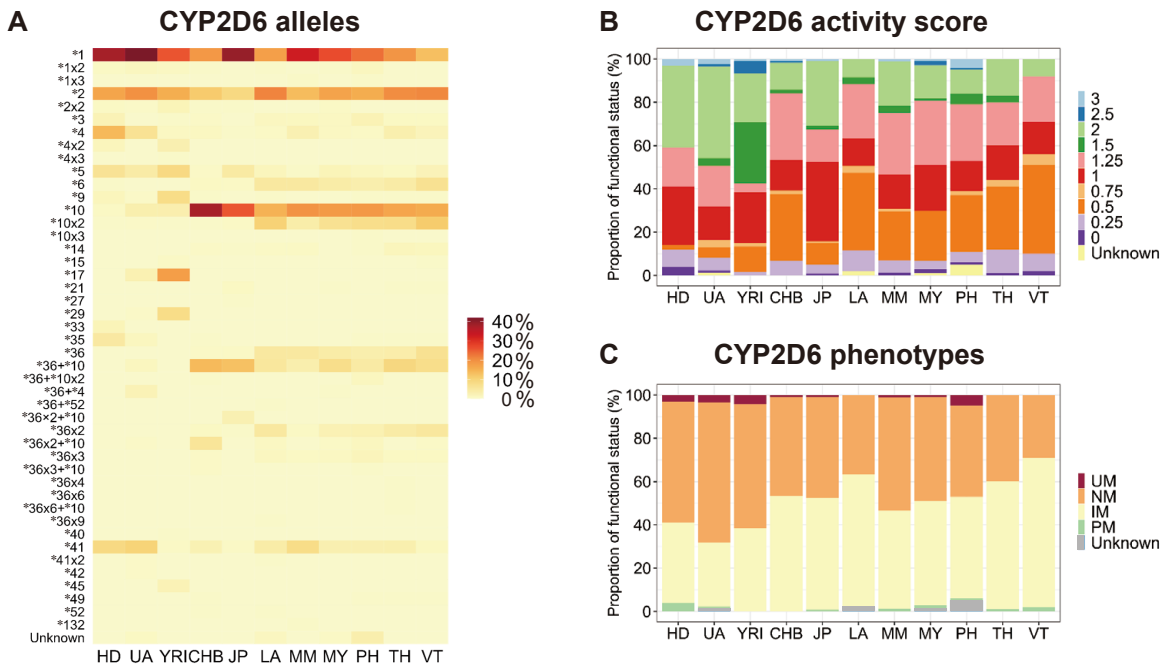


図2 13 集団の CYP2D6 のアリル、アクティビティスコアおよびフェノタイプの頻度

PKseq パネルによってジェノタイピングされたバリエーション情報を PharmVar データベースのアリル情報を用いてアリルに変換し、各集団のアリル頻度を算出した (A)。これらのアリルを CPIC ガイドラインの diplotype-phenotype table を用いて酵素活性の推定値であるアクティビティスコアを算出し (B)、各集団のフェノタイプ頻度を算出した (C)。

HD : ユーロッパ (n=100)、UA : アラブ首長国連邦 (n=85)、YRI : ナイジェリア (n=120)、CHB : 中国 (n=120)、JP : 日本 (n=1,087)、LA : ラオス (n=95)、MM : ミャンマー (n=88)、MY : マレーシア (n=104)、PH : フィリピン (n=100)、TH : タイ (n=100)、VT : ベトナム (n=100)

13 集団中の CPIC ガイドラインが推奨する

薬物治療の比較

CPIC ガイドラインに従い、各フェノタイプの患者を推奨されている用法・用量ごとに分類した (図3)。標準的な薬物治療とは異なる用法・用量が1種類以上の薬剤で推奨される患者は全体の97%であった。また標準治療とは異なる用法・用量が推奨されている患者の頻度が40%を超える集団が1つでも示された薬物はシタロプラム/エスシタロプラム、クロピドグレル、エファビレンツ、NSAIDs、パロキセチン、プロトンポンプ阻害薬、セルトラリン、タクロリムス、タモキシフェン、三環系抗うつ薬、ポリコナゾールであり、全体の55%を占めた。

考 察

PGx バイオマーカーは薬剤の効果や有害事象の発現リスクを予測することが可能である。しかし、様々な集団間での PGx バイオマーカーのバリエーションの頻度を比較した報告は非常に少ない。本研究では我々が以前に開発したターゲットシーケンシングパネルである PKseq を用い、アジアを中心とした13集団の PGx 関連遺伝子のバリエーションを明らかにした。

様々なシーケンシング技術が改良されているが PGx 関連遺伝子は CNV や高い相同性配列をもつ偽遺伝子などによってゲノムの領域のジェノタイピングは困難なままである⁹⁾。遺伝子型の標準ゲノムを作成する目的で設立された Genetic

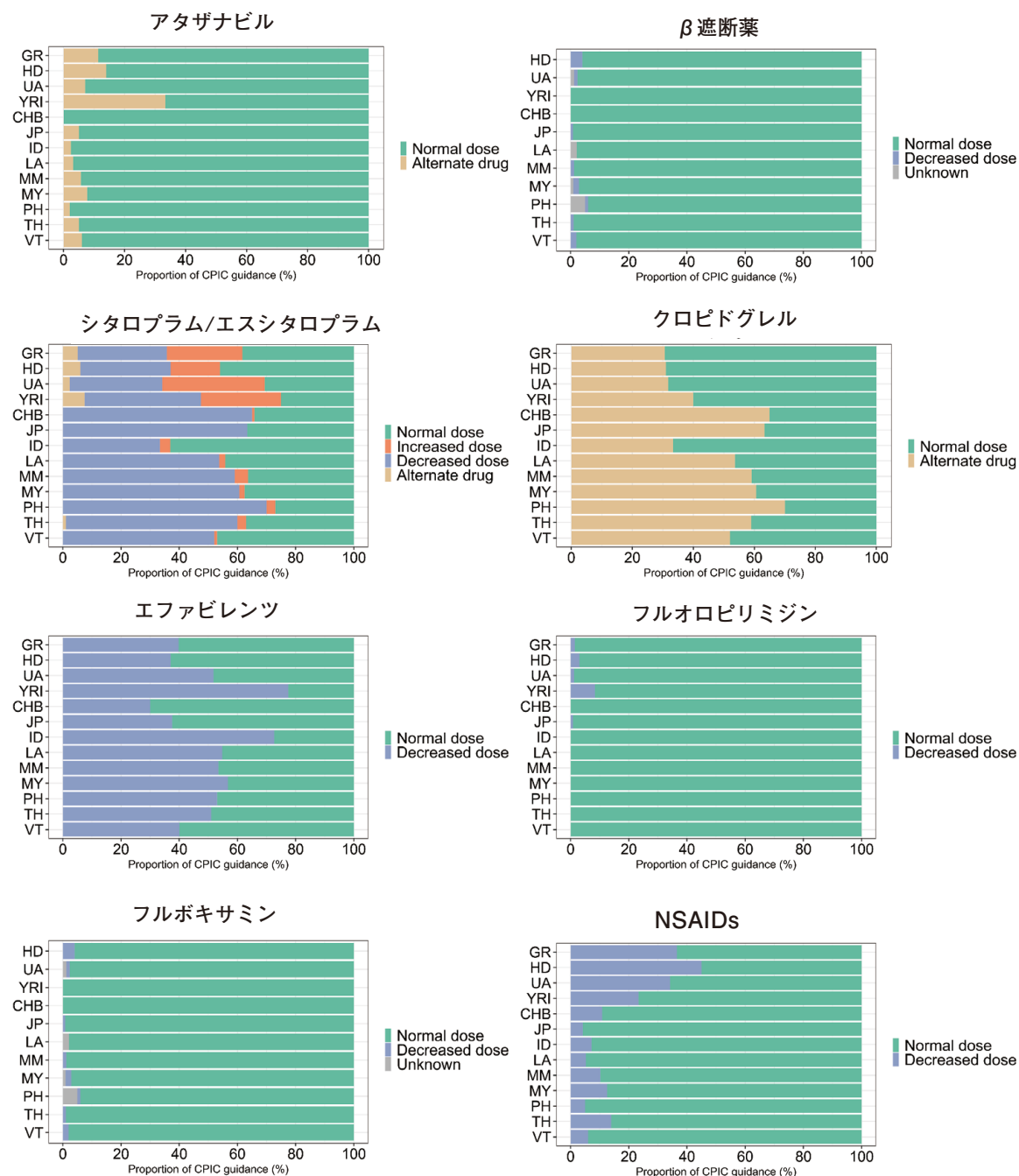


図3 13 集団中の CPIC ガイドラインが推奨する薬物治療の比較

PKseq パネルによってジェノタイプ化されたバリエーション情報を CPIC ガイドラインに従い、各フェノタイプの患者を推奨されている薬物治療ごとに分類し、各集団の頻度を算出した。

GR: ギリシャ (n=274)、HD: ユーロパ (n=100)、UA: アラブ首長国連邦 (n=85)、YRI: ナイジェリア (n=120)、CHB: 中国 (n=120)、JP: 日本 (n=1,087)、ID: インドネシア (n=84)、LA: ラオス (n=95)、MM: ミャンマー (n=88)、MY: マレーシア (n=104)、PH: フィリピン (n=100)、TH: タイ (n=100)、VT: ベトナム (n=100)

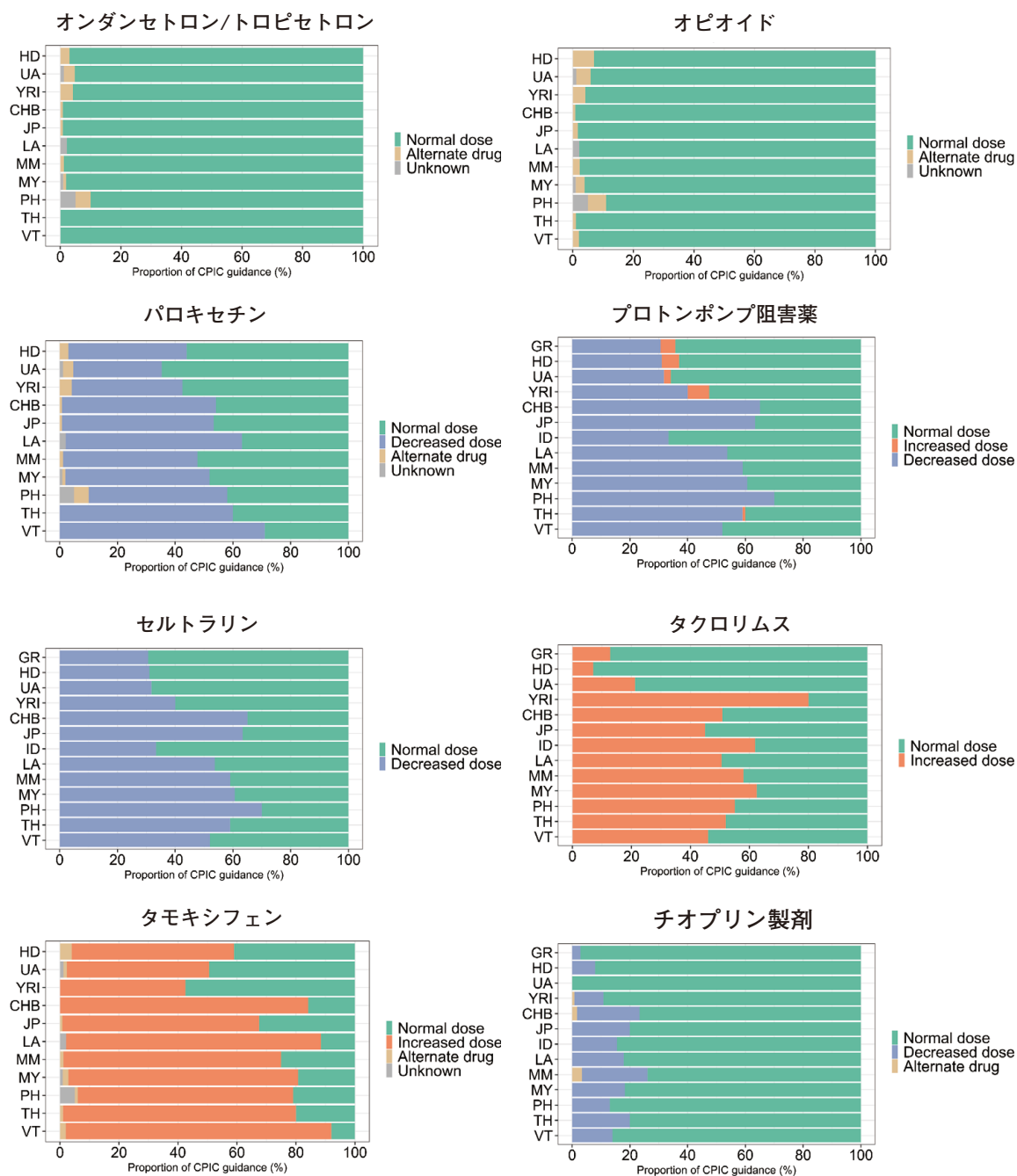


図 3 の続き

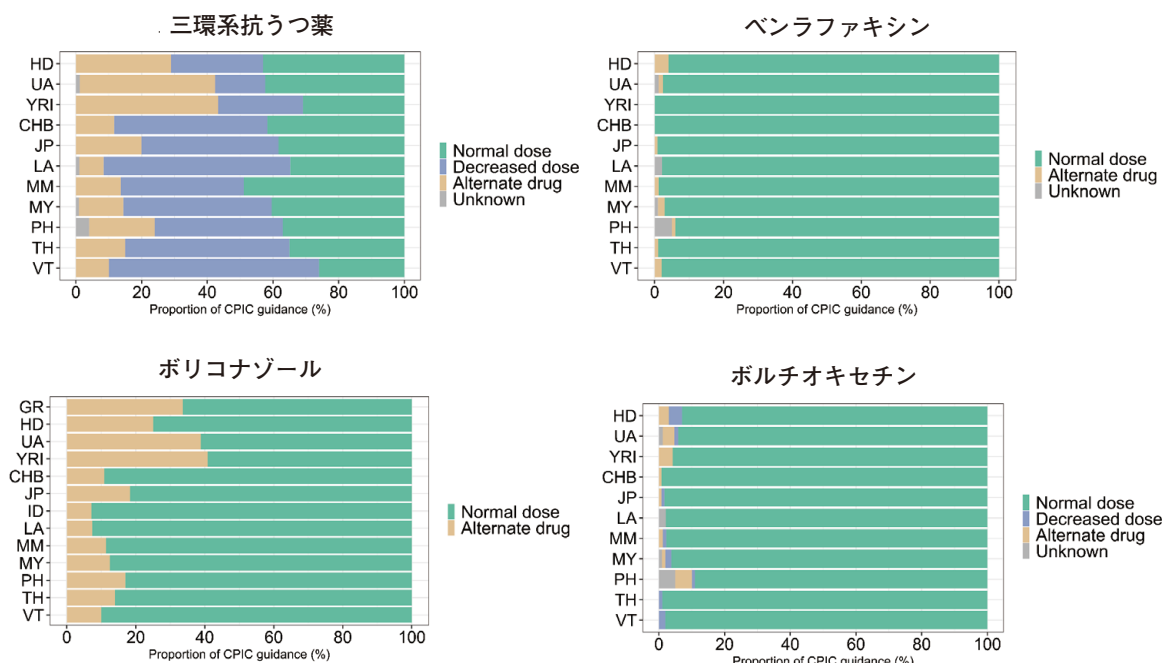


図3の続き

Testing Reference Material Program (GeT-RM) ¹⁰⁾ の PGx 部門においても様々なプラットフォームで同じゲノムがジェノタイピングされた。しかし複数のプラットフォームで決定される遺伝子型が一致せず、最終的な正解遺伝子型を決定することができていない ¹⁰⁾。本研究においてもディプロタイプ決定において矛盾が生じたため、Unknown のジェノタイプと判定された患者が見られた。今後ロングリードシーケンサーなどを用いて完全な遺伝子型が決定されている PGx 標準ゲノムを作成し、ゴールドスタンダードとなるジェノタイピング方法を確立する必要がある。

本研究では東南アジアからの9つの集団を対象としている。CYP2B6 遺伝子の酵素活性を低下させる最も高頻度なアリルである CYP2B6*6 の頻度がインドネシアにおいて 43.4% を示した (図 1C)。これは世界で報告されている CYP2B6*6 の頻度の中で最も高い数値である。また、ベトナム集団において CYP2D6 のアクティビティスコアが 0.5 を示すグループの頻度が最も高くなっている (48.2%、

図 2B)。このアクティビティスコア 0.5 のカテゴリーは IM に分類されているが、IM の中でも酵素活性が最も低いカテゴリーであり、IM の他のアクティビティスコアである患者と比較して CYP2D6 の基質薬物の血中濃度が高くなり有害事象が発生したり、薬物の治療効果が高くなったりすることが考えられる。世界の様々な集団で PGx 関連遺伝子のジェノタイピングが行われ ¹¹⁾、1000 ゲノムプロジェクトにおいて、東南アジア集団としてベトナムや南中国集団がシーケンシングされている ³⁾ にもかかわらず同じ東南アジア集団であるインドネシアの CYP2B6*6 の頻度が極めて高い (43.4%、図 1C) ような知見が同定されている (1000 ゲノムプロジェクトのベトナム (KHV) の CYP2B6*6 の頻度は 21.1%、南中国 (CHS) は 15.6%) ³⁾。そのため、今後も PKseq パネルを用いて様々な集団をシーケンシングすることが必要である。

本研究を通して PGx 検査によって標準とは異なる薬物治療を必要とする患者の頻度が高い集団が確認された。特に CYP2C19 は最も多くのフェノタイプ

に分けられたため標準治療だけではすべての患者が有害事象を避けつつ最大限の効果を得ることが難しいことが考えられる。そのため PGx 検査情報に基づいた用法・用量の設定が必要である。しかし本邦では CYP2C19 の PGx 検査は保険収載されておらず医師や患者が希望しても検査することができない。そのため CYP2C19 の基質薬物であるエスシタロプラム、クロピドグレル、プロトンポンプ阻害薬においては可及的速やかに CYP2C19 遺伝子検査を保険収載する必要があると考えられる。加えて CYP2B6*6 が高頻度であるインドネシアではエファビレンツ、アジア集団の中で唯一 CYP2C9 の PM が 1.0% 認められたタイでは NSAIDs、CYP2D6 活性が低い患者が多いベトナムでは向精神薬やタモキシフェンなど CYP2D6 基質薬物の PGx 検査が必要であり、ひいては各集団に適切な用法・用量を提案することが必要になるかもしれない。

本研究ではアジアを中心とした 13 集団の、薬剤の効果や有害事象の発現リスクに関する PGx 関連遺伝子のバリエーションの大規模データベースを構築した。PGx バイオマーカーを世界的に普及させるためには今後さらにサンプル数や集団を増やし、データベースの拡充が必要である。

謝 辞

本研究を遂行するにあたり、研究助成を賜りました公益財団法人臨床薬理研究振興財団に深謝申し上げます。また、本研究に検体を提供していただいた患者様に厚く御礼を申し上げます。

利益相反

本研究に関して開示すべき利益相反はありません。

文 献

- 1) Alshabeeb MA, Alyabsi M, Aziz MA, Abohelaika S. Pharmacogenes that demonstrate high association evidence according to CPIC, DPWG, and PharmGKB. *Front Med (Lausanne)* 2022; **9**: 1001876.
- 2) Fukunaga K, Momozawa Y, Mushiroda T. Update on next generation sequencing of pharmacokinetics-related genes: Development of the PKseq panel, a platform for amplicon sequencing of drug-metabolizing enzyme and drug transporter genes. *Drug Metab Pharmacokinet* 2021; **37**: 100370.
- 3) Sherman CA, Claw KG, Lee SB. Pharmacogenetic analysis of structural variation in the 1000 genomes project using whole genome sequences. *Sci Rep* 2024; **14**: 22774.
- 4) Runcharoen C, Fukunaga K, Sensorn I, Iemwimangsa N, Klumsathian S, Tong H, et al. Prevalence of pharmacogenomic variants in 100 pharmacogenes among Southeast Asian populations under the collaboration of the Southeast Asian Pharmacogenomics Research Network (SEAPharm). *Hum Genome Var* 2021; **8**: 7.
- 5) Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114-20.
- 6) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-60.
- 7) McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-303.
- 8) Fukunaga K, Kato K, Okusaka T, Saito T, Ikeda M, Yoshida T, et al. Functional characterization of the effects of N-acetyltransferase 2 alleles on N-acetylation of eight drugs and worldwide distribution of substrate-specific diversity. *Front Genet* 2021; **12**: 652704.
- 9) Chen X, Shen F, Gonzaludo N, Malhotra A, Rogert C, Taft RJ, et al. Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data. *Pharmacogenomics J* 2021; **21**: 251-61.
- 10) Pratt VM, Everts RE, Aggarwal P, Beyer BN, Broeckel U, Epstein-Baak R, et al. Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes: A GeT-RM collaborative project. *J Mol Diagn* 2016; **18**: 109-23.
- 11) McInnes G, Lavertu A, Sangkuhl K, Klein TE, Whirl-Carrillo M, Altman RB. Pharmacogenetics at scale: An analysis of the UK Biobank. *Clin Pharmacol Ther* 2021; **109**: 1528-37.