

2020年度ヒトバリエーションデータ 研究倫理勉強会

2020年10月17日 土 10:30-12:00

AMED 臨床ゲノム情報統合データベース整備事業では、日本人疾患バリエーションデータベース(MGeND: <https://mgend.med.kyoto-u.ac.jp>)を2018年3月に公開し、疾患バリエーションデータの登録受付を行っています。溝上班では活動の一環として、臨床ゲノム情報の共有とデータベース構築におけるELSIへの対応や倫理関連の諸問題に対するサポートを行っています。

この度、より多くの方にデータ登録を行っていただけるよう、MGeND登録方法の紹介や、臨床ゲノム情報の共有等に係るELSIなどについて情報共有を行う、ヒトバリエーションデータ研究倫理勉強会を開催致します。是非ご参加下さい。

プログラム

10:30~10:35

MGeNDの経緯と現状

溝上 雅史 (国立国際医療研究センター)

10:35~10:55

MGeND紹介とデータ登録について

中津井 雅彦、鎌田 真由美 (京都大学大学院医学研究科)

10:55~11:10

ヒトバリエーションデータの登録と 利活用に向けたガイダンス

加藤 和人 (大阪大学大学院医学系研究科)

11:10~11:30

難病研究における

ヒトバリエーションデータの利活用

小崎 健次郎 (慶應義塾大学医学部)

11:30~12:00

質疑応答・総合討論(パネルディスカッション)

ウェビナー形式(オンラインセミナー)

- 事前に参加登録を頂いた方に当日参加用URLをお知らせします。
- ブラウザから誰でも参加できるシステム(ユーザ登録不要)を使用します。
- パソコン・スマホで全国どこからでもご参加頂けます。

対象

大学・研究機関の倫理審査関係者、研究者など

参加費

無料

参加登録

下記ページのフォームから参加登録をお願いします。



※事前参加登録は10月14日(水)17:00まで

www2.med.osaka-u.ac.jp/eth/seminar/20201017-1/

開催レポート

AMED 臨床ゲノム情報統合データベース整備事業（溝上班）

2020 年度ヒトバリエーションデータ研究倫理勉強会

日時：2020 年 10 月 17 日（土）10:30-12:00

開催形式：ウェビナー

<https://www.med.osaka-u.ac.jp/pub/eth/seminar/20201017-1>

【概要】

2020 年 10 月 17 日に「2020 年度ヒトバリエーションデータ研究倫理勉強会」を、ウェビナー形式（オンライン動画配信）で開催した。これは、AMED 臨床ゲノム情報統合データベース整備事業（臨ゲノ）で 2018 年 3 月に公開した日本人疾患バリエーションデータベース（MGeND: <https://mgend.med.kyoto-u.ac.jp>）に、より多くのデータ登録を行っていただけるよう MGeND 登録方法の紹介を行うとともに、臨床ゲノム情報の共有等に係る ELSI 等について情報共有を行うことを目標としている。2019 年度に初めて開催し、今年度第 2 回として開催したものである。

【プログラム】

「MGeND の経緯と現状」

溝上 雅史（国立国際医療研究センター）

「MGeND 紹介とデータ登録について」

中津井 雅彦、鎌田 真由美（京都大学大学院医学研究科）

「ヒトバリエーションデータの登録と利活用に向けたガイダンス」

加藤 和人（大阪大学大学院医学系研究科）

「難病研究におけるヒトバリエーションデータの利活用」

小崎 健次郎（慶應義塾大学医学部）

「質疑応答・総合討論」（パネルディスカッション）

以下、各講演概要、総合討論中の主な質疑を簡単に紹介する。また、当日の発表資料を後ろに掲載するので参照されたい。

「MGeNDの経緯と現状」

溝上 雅史（国立国際医療研究センター）

2015年1月30日、米国のオバマ前大統領は「Precision Medicine Initiative」を提唱した。世界的にもその方向に進み、日本においてもゲノム医療実現推進協議会が設立され、様々な提言やディスカッションが行われ2016年度からゲノム医療の準備開始ということになった。

臨床ゲノム情報統合データベース（MGeND）の構築は、日本人の病的遺伝子バリエーションを集め、データベース化して臨床応用化するナショナル・プロジェクトと位置づけられた。ゲノム情報と疾患特異性や臨床特性等の関連について「日本人を対象」にした検証を行い、臨床および研究に活用可能な臨床情報と遺伝情報を統合的に扱うデータベースを整備して、更にそれらを利活用し、先端研究開発を一体的に進めることができるようにするものである。

まず、現在では治らない難病・がん・感染症・認知症等の疾患分野を対象として、それらの検体収集、ゲノム解析、臨床情報を含めた情報の統合・解析を行い、疾患ごとにデータベース化することを一次班では目指した。次にそれらを全体統合したものを横断的なデータベース（MGeND）とし、臨床に還元することを二次班の目標とした。スライド4に示した通り、データは既存のAMED Genome group sharing Database（AGD）に登録した後で、臨床情報、診断名、変異情報を非制限公開（オープンアクセス）でMGeNDに登録する、という流れで2016年から3年計画で進め、2018年3月にMGeNDを公開した。皆様のご協力のおかげで、今では、多数バリエーションおよびその臨床データがMGeNDから公開されている。更にこれを進めていくために、MGeNDの紹介とデータ登録の仕方、登録ガイドランスについての説明をしていただき、更に難病利用における現状と今後の課題について、本日はお話しいただく予定である。

【講演 1】

「MGeND 紹介とデータ登録について」

鎌田 真由美（京都大学大学院医学研究科）、
中津井 雅彦（山口大学大学院医学系研究科）

<国内外のヒトバリエントデータ利活用の現状と課題>

シーケンス解析により検出されたバリエント（変異体）に対して、データベースの情報の付与、様々な文献情報や患者さんの背景に基づいて臨床的解釈が行われ、最終的に治療方針が決定される（スライド 2）。そのゲノム医療のボトルネックにもなっているのが「臨床的解釈の付与」である。主に単一遺伝子疾患などでは ACMG ガイドライン（スライド 3）がよく使われているが、これは集団での頻度情報、家系情報、更に計算機による予測の結果などをエビデンスの強さに基づき整理して、最終的に総合評価して、Benign、Likely benign、Likely pathogenic、Pathogenic といったラベルを付け、いずれにも分類できなかったものを Uncertain significance としている。

スライド 4 に示した ClinVar は米国の NCBI が運用している疾患関連バリエントデータベースで、ここで懸念すべき点は「遺伝的背景の違い」だ。肺腺癌（LADC）におけるドライバー遺伝子の変異頻度を比較した例では、欧米と東アジアで違いが見られる（スライド 5）。よって、日本におけるゲノム医療を推進していくためには、日本人の遺伝的特徴を反映したデータベースが必要であるため、開発してきたのが MGeND というデータベースである。

<MGeND へのデータ登録>

MGeND では、希少難治性疾患、がん、感染症、認知症・難聴を対象にデータ登録のためのフォーマットを用意している。ClinVar に準拠した変異データ用、単一遺伝子疾患に関しては GWAS データ用、また感染症においては免疫に関わる HLA タイピングデータ用のフォーマットを策定している。

具体的に登録いただく項目をスライド 9 に示した。まず疾患名でコードもしくは自由記載、次にバリエントに関して遺伝子名とその変異情報である。また付随する臨床情報として、年齢（非制限公開ベースのため 10 歳刻みの年齢層）、性別、Pathogenic、Benign などを登録いただいている。オプションで、バリエントを検出するにあたって用いたプラットフォーム、GWAS 解析などであれば検定法などの情報もご登録いただくようお願いしている。

どのようなデータが登録できるのか、スライド 10 に示すいずれかを満たしているものになる。「1）本事業のために患者様の同意を得た上で取得されたもの、2）本事業以外でも”データベースに登録する”、または”論文で公開する”ことが

説明同意文書に入っているもの、3) 既に公開されているもの」である。登録可能かどうか、国立国際医療研究センターに設置されている MGeND 登録管理委員会で申請書に基づき審査させていただく。許諾が得られたデータに関してシステム担当へ送っていただき、項目の確認などを行った上でデータベースに登録させていただいている。

<MGeND に登録されているデータ>

MGeND は 2018 年 3 月に公開して約 2 年半たったが、日本だけでなく米国など世界中からアクセスがあり、また各疾患領域においてかなり多数のデータをご登録いただいている(スライド 13~15)。ClinVar との比較も行っており、MGeND の 2020 年 3 月時点のデータと ClinVar と比較したところ、MGeND に登録されているデータの半数以上は「MGeND のみに登録されているもの」だった。そのうち 4 割ほどは ClinVar にも登録されているのだが、付いている臨床的意義が異なっているものもあり、ClinVar では Uncertain significance だが MGeND では Benign・Pathogenic が付いているような例も見受けられた(スライド 16)。

<MGeND の使い方と新機能>

MGeND の検索機能として、検索窓から遺伝子名、バリエーション、疾患名、dbSNP の rs 番号や MGeND の Submission ID (MGS から始まるもの) など、フリーテキストで検索可能になっていたが、改良した結果、Gene 等、カテゴリー別でも検索できるようになった。UI (ユーザーインターフェース) も本年改良し、バリエーションが探しやすいうように、MGeND での登録数、その細胞種が何だったのか等のアノテーション、ClinVar でのアノテーション等をシンプルに表示するようしている。アップデートとしてフィルタリング機能を強化し、アセンブリ・バージョンや細胞種、解析 (GWAS のみ等) での絞り込みができるようにした。また国内外のデータベース (TogoVar など) へのリンクや、ClinVar、CIViC、DisGeNET といったアノテーション情報を見ただけできるようになっている他、疾患領域ごとのページも用意している(スライド 22)。

最後に、現在進めている改良について紹介する。現在データは TSV 形式でダウンロードしていただけるが、API (Application Programming Interface) の整備を進めている。また、多因子疾患に関してもより対応を強化する意味で、研究スタディごとのページを作成している。更に、現在の MGeND ではヒトゲノムアセンブリ hg19 と hg38 は別のエントリーになっているが紐づけを進めており、また ToMMo 等の最新データの取り込みを進めている。

【講演 2】

「ヒトバリエントデータの登録と利活用に向けたガイダンス」

加藤 和人（大阪大学大学院医学系研究科）

<データ登録のためのガイダンス>

今日は MGeND バリエントデータ登録のためのガイダンスを私からお話しして理解を深めていただき、結論的には「これに従っていただければ安心してデータを入れることができる」ということで、皆様のご協力をお願いしたい。

MGeND ホームページのトップページのメニュー「Submission」をクリックすると、データ登録ガイダンス（詳細版）、登録申請書、ガイダンス（簡易版）の3つをセットで提供しているページが参照できる。このガイダンスは、我々の班と三省のゲノム指針事務局および個人情報保護委員会とのやり取りを経て出来上がったもので、非常にしっかりと国の決まりに基づいたものになっている。ポイントとして「国内からのデータ提供」で、MGeND に国内からデータを提供していただき、そして非制限公開で世界中に公開していく、というプロセスを考え様々なルールを解釈した（スライド 5-6）。提供していただくデータに関しては、疾患名、遺伝子名、遺伝型情報、年齢層、性別といった簡単な情報のみである（スライド 7）。

<MGeND へのデータ登録手順>

登録ガイダンス（簡易版）は登録のお願いとなっており、1、2、3 という3つの異なる状況に合わせて様々なガイドが書かれている。A4 表裏1枚なのでダウンロードしてぜひご覧いただきたい（スライド 9）。また登録の際には、登録申請書（スライド 11）に記入いただくことになっており、登録者情報、臨ゲノ事業への参画の有無、どういうデータなのか、そしてデータそのものの情報、更に必要な場合には説明同意文書の情報や倫理審査の情報も記入して提出していただくことになっている。すぐに登録できるのが、この「3.」 「学術論文や研究班の報告書など、学術的に信頼された媒体で既に発表されているバリエントデータ」で、そのまま MGeND に登録可能である。また研究の中で、「解析した結果はデータベースに登録する」、または「解析した結果は学術論文等で発表する」ということが説明文書で説明され、同意内容に入っている場合もすぐに登録可能となっている（スライド 11）。簡易版ガイダンスの裏面に同じことをフローチャートで整理しているので、ご自分がお持ちのデータがこの中のどれに当てはまるのかをご確認いただき、積極的に登録していただきたい（スライド 12）。

同意が得られている場合でも一点だけ注意点がある。公表することが同意内容に入っていると基本的にはそのまま登録できるが、もし極めて希少な疾患で、

データ自体から個人識別性が十分に除けないとき（例えば全国でも特定の地域や医療機関等に患者が非常に限られていて、どこの研究機関からデータが出てきたかがわかるだけで個人同定の可能性が出てくるという場合）には、再同意、または倫理審査委員会の承認を得てオプトアウトのいずれかを行えば提供可能となる（スライド 14）。

<ELSI 理解のための活動>

最後に ELSI (Ethical, Legal and Social Issues) を担当しているグループとしての活動についていくつか紹介する (スライド 15~)。1 つは、MGeND データを利用する方への利用規定作成へのサポートで、例えば、利用目的をしっかりと確認して使っていただくこと、営利目的の場合には事前相談をいただくこと、費用負担のこと、違法利用等に関する免責等々について MGeND ホームページにデータ利用規則が掲載されているが、その中の何件かについて我々のほうで検討した内容が採用された (スライド 16)。それから、本日の会のような MGeND へのデータ登録のための ELSI の理解、そして同時に MGeND そのものへの理解も深めていただくために勉強会を開催してきており、昨年 2019 年 10 月に大阪で本格的に開始した。本日が第 2 回ヒトバリエーションデータ研究倫理勉強会である。また今年度は、日本分子生物学会の年会の中でフォーラムを企画しており 12 月 4 日に「ゲノム研究と医療をつなぐデータベース—ヒト疾患関連データの利活用と ELSI を考える—」ということでセッションを組んでいるので、もし分子生物学会に参加される方がおられたら、ぜひご参加いただければと思う。

【講演 3】

「難病研究におけるヒトバリエントデータの利活用」

小崎 健次郎（慶應義塾大学医学部）

<病的バリエントデータベースの意義と利活用>

病的バリエントデータベースは診療・研究両方に非常に重要で、特に正常なものと異常なものをフィルターアウト・インする必要がある。フィルターアウトという点では、東北メディカル・メガバンク（ToMMo）、あるいはバイオバンク・ジャパン（BBJ）と併せた TogoVar のデータが非常に役に立つ。逆にフィルターインのために、病的バリエントを集積することが MGeND の設計の意図であると思う。現在、TogoVar で 8,000 人近いデータが統一的な形で提示され、病的バリエントのほうは各分野のものが MGeND に統一されている（スライド 3）。海外の病的バリエントデータベースは、もとをたどると英国のカーディフ大学 HGMD が作成したもので、3 年以上が経ったものについてはフリーで公開される。（最新データを PubMed 等文献に掲載されているものを含めて使用するためには年間使用量が課金される。）また米国の国立衛生研究所（NIH）の ClinVar がデータを集めていて、こちらはフリーで使用でき、多くのデータが臨床検査会社から提供されているところが特徴的だ。他には LOVD が有名で、疾患ごとのデータベースも運営されており、多くは ClinVar などに取り込まれている。

がん（体細胞）のバリエントと、遺伝性腫瘍のバリエントは区別して考える必要がある。希少疾患研究、特に単一遺伝子病の場合は原因遺伝子の数は基本的に 1 個だが、体細胞のがん研究の場合にはドライバーがいくつか組み合わさっていることがあり得る。また、バリエント数は優性遺伝病であれば 1 個、劣性遺伝病であれば 2 個となる。MGeND では、様々な領域から登録が行われているが、難病領域・希少疾患領域では 15,000 程度のデータが登録され、未診断疾患イニシアチブ（IRUD）で得られた 1,500 のバリエントも MGeND から登録公開予定になっている。また、小児科分野では小児遺伝学会に声をかけてパブリッシュされた論文を集めたところ、2,500 を超える文献が見つかった（スライド 10）。

<バリエントの評価とキュレーション>

病原性の予測に関して、どの程度の意味があるかについて CADD スコアを用いて評価した。CADD スコアは複数の Prediction program での評価を統合し病原性を数値化にするもので、スコアが 25 以上であると概ね病的とされている。ToMMo データと難病関連の約 1,500 バリエントの CADD スコアの分布を比較したところ、マニュアルキュレーションを通じて集めたものについてはスコア値が高く 25 以上のものが大きな山を示したのに対し、それ以下のものは非常に

低い（特異度が高い）。一部は劣性遺伝病の原因である可能性はあるが、マニュアルキュレーションが重要ということが明らかになった。

バリエントの記載法については歴史的に定まったものがなく、次世代シーケンサーデータで統一的な表記ができるようになったのはごく最近のことであり、標準となる遺伝子名や、座標の管理もかなり大変だった。そこで、キュレーションをする4人ぐらいのチームで、座標確認、バリエントの病的意義の確認、疾患名の標準化（OMIM や Orphanet などのコード番号を付ける）といった作業も進めている。

<今後の課題と学会との連携>

病的バリエントデータベースの今後の課題としては構造異常バリエントとスプライシング異常の集積だと考えている。ClinVar や HGMD を含めて、ほとんどの既知のバリエントは転位変異、一塩基置換か数塩基程度までのインサクション・デリションである。構造異常バリエントについては、次世代シーケンサーを使うようになって相当拾えるようになってきているが、まだ統一的な記載の方法や日本人全体のデータが存在していない。特に劣性遺伝病の片方のアレルを見つけていく作業が非常に重要と考えられることから、今年度からまずはショートリードでデータを収集する作業を進めている。

スプライシング異常について、通常ゲノムの機能予測のプログラムは、ゲノム配列単体からアミノ酸に変換したときに病気の原因になるかどうか、ということのみで判断しているが、スプライシング異常により非常に大きな影響が起きる例も報告されている。1,000人ゲノムのデータを解析した結果では、200例程度のエクソームのみではサイレント変異と解釈されるものが、RNA-Seq を併せて判断すると実は大きな影響がある、あるいはミスセンスについて大きな影響がある、というバリエントが相当数あった。このようなデータも含めて、現在 MGeND の難病セクションに登録を進めている。

今後どのようにバリエントを探していくか、小児科学会の事例を紹介したが、神経内科、あるいは人類遺伝学会全体でも、このような情報を集めていくことが当然必要だ。先ほど紹介した IRUD についてはほぼ全例の病的バリエントを MGeND に登録する流れがあり、他の公的な難病プロジェクトからのバリエントの一括登録も必要と考えている。更に現在保険によって実施可能になった病的バリエントの臨床検査データをどう集めていくかも重要な課題であり、人類遺伝学会で検討を始めているところである。

【質疑応答・総合討論での質問一覧】

登壇者5名による総合討論（司会、加藤）が始まり、参加者からの質問をもとに議論が交わされた。その際取り上げられた主な質問は以下のとおりである。

- ClinVarのように再配布が可能になれば、MGeNDも利用の場がより広がると思われるが、MGeNDでも再配布を可能としていくような計画はあるか？倫理面、あるいは技術的なハードルがあり難しいか？
- MGeNDで得た情報をwetな実験に使用する（細胞、動物モデルを作製する等）場合、自施設のIRB承認以外に何か手続きは必要か？
- MGeNDに公開された情報は公知の事実として扱って良いということになるか？
- 薬理学的な薬別の効き方の個人差に関するバリエーションデータ登録はあるか？
- がん領域の登録に関してはC-CATに集積されているが、がん遺伝子パネル検査の情報はMGeNDに登録されているか？
- ClinVarとMGeNDで病原性の判断に違いがある時があり、診断に困ることがあるが、検証はどのように行っているのか？

閉会に際して、加藤氏より閉会挨拶があり盛況のうちに終了した。



総合討論の議論の様子

MGeNDの経緯と現状

国立国際医療研究センター
ゲノム医科学プロジェクト
プロジェクト長 溝上 雅史



2015年1月30日オバマ大統領は
一般教書演説で**Precision Medicine Initiative**を提唱

Tonight, I'm launching a new **Precision Medicine Initiative** to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier. I want the country that eliminated polio and mapped the human genome to lead a new era of medicine —one that delivers the right treatment at **the right time**.



本邦におけるゲノム医療実現推進協議会の設立とその活動

議長	内閣官房 健康・医療戦略室長
	文部科学省 研究振興局長
	厚生労働省 医政局長
	厚生労働省 健康局長
	厚生労働省 大臣官房審議官(危機管理、科学技術・イノベーション、国際調整、がん対策担当)
	経済産業省 大臣官房商務・サービス審議官
我妻 利紀	一般財団法人 バイオインダストリー協会 運営会議委員
磯 博康	日本公衆衛生学会 理事長
上野 裕明	日本製薬工業協会 研究開発委員会 副委員長
加藤 規弘	国立研究開発法人 国立国際医療研究センター メディカルゲノムセンター センター長
清原 裕	公益社団法人久山生活習慣病研究所 代表理事
久保 充明	元国立研究開発法人 理化学研究所統合生命医科学研究センター 副センター長、医師
近藤 達也	独立行政法人 医薬品医療機器総合機構 理事長
塩田 浩平	滋賀医科大学 学長
末松 誠	国立研究開発法人 日本医療研究開発機構 理事長
高木 利久	東京大学大学院理学系研究科生物科学専攻 教授
辻 省次	国際医療福祉大学ゲノム医学研究所長
中釜 齊	国立研究開発法人 国立がん研究センター 理事長
松原 洋一	国立研究開発法人 国立成育医療研究センター 研究所長
武藤 香織	東京大学医学研究所公共政策研究分野 教授
山本 雅之	東北大学東北メディカル・メガバンク機構 機構長

平成 27 年 1 月 21 日
健康・医療戦略推進会議決定
平成 27 年 5 月 15 日
平成 27 年 7 月 14 日
平成 28 年 6 月 13 日
平成 29 年 3 月 31 日
平成 29 年 7 月 11 日
平成 30 年 5 月 23 日
一 部 改 正



臨床ゲノム情報統合データベース(MGeND)とは？

日本人病的バリエーションを集めデータベース化し、 臨床応用化するナショナル・プロジェクト

- ゲノム情報と疾患特異性や臨床特性等の関連について
日本人を対象とした検証
- 臨床及び研究に活用することができる臨床情報と遺伝情報を
統合的に扱うデータベースを整備
- それら研究基盤を利活用した先端研究開発を一体的に推進



- 難病・がん・感染症・認知症等の疾患分野を対象
- 検体収集、ゲノム解析、臨床情報を含めた情報の統合・解析
- 疾患毎に臨床ゲノム情報データストレージを整備(一次班)
- 全体を統合した横断的データベースを整備(二次班)
- 臨床現場への還元(クリニカル・シーケンスの実現)

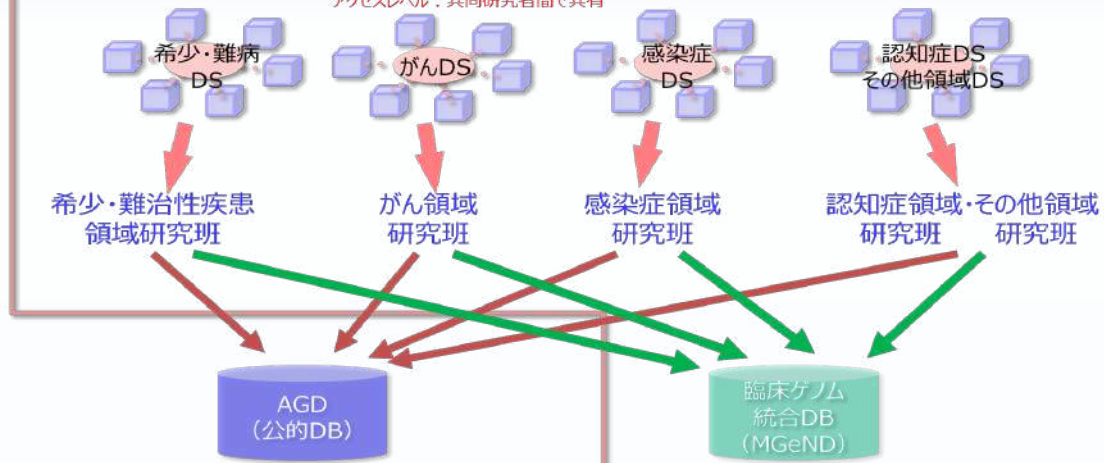


(AMEDのHPより引用改変)

臨床ゲノム情報統合データベース整備事業

DS : 各疾患領域グループのデータストレージ [制限共有]

内容 : 患者個人レベルのvcf等 + 高度臨床情報
アクセスレベル : 共同研究者間で共有



AGD : AMEDが定める公的DB

[制限共有 または 制限公開]

内容 : BAM、VCF等、ゲノム解析のRaw Data + 限定された臨床情報

アクセスレベル (選択可能) :

制限共有 : 共同研究者間で共有

制限公開 : データアクセス申請を承認された研究者が利用可能

統合DB : 臨床ゲノム統合DB

[非制限公開]

内容 : 限定された臨床情報 + 診断名 + 変異情報

アクセスレベル : オープンアクセス

(AMEDバイオバンク課)



2018年3月臨床ゲノム情報統合データベース“MGeND”公開

<http://mgend.med.kyoto-u.ac.jp/>



MGeNDの経緯と現状

溝上 雅史

MGeND紹介とデータ登録について

中津井 雅彦・鎌田 真由美

ヒトバリエントデータの登録と利活用に 向けたガイダンス

加藤 和人

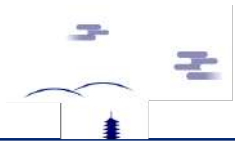
難病研究におけるヒトバリエントデータの 利活用

小崎 健次郎



日本人疾患バリエーションデータベース MGeND紹介とデータ登録について

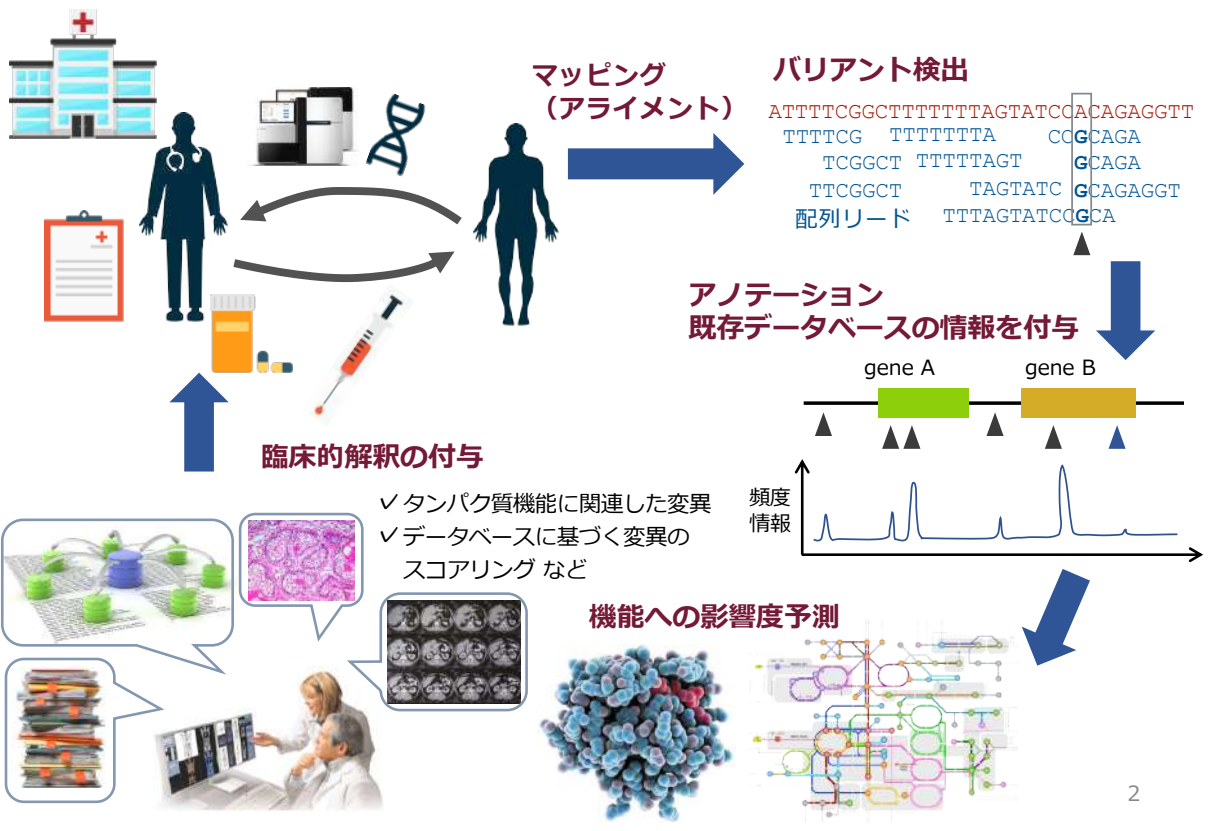
京都大学大学院 医学研究科
人間健康科学系専攻 ビッグデータ医科学
中津井 雅彦, 鎌田 真由美



京都大学
KYOTO UNIVERSITY

2020.10.17 20120年度 ヒトバリエーションデータ研究倫理勉強会

ゲノム医療のため臨床シーケンスの流れ



バリアントの疾患関連性の評価 ACMGガイドライン

米国臨床遺伝・ゲノム学会 (American College of Medical Genetics and Genomics; ACMG)

						根拠の強さ		
Benign			Pathogenic			Pathogenicity	Rational	Category
Strong		Supporting	Supporting	Moderate	Strong		Very Strong	PVS1
Population data	MAF is too high for disorder (BA/BS1) OR observation in controls inconsistent with disease penetrance BS2			Absent in population database PMS	Prevalence in affecteds statistically increased over controls P04		Strong	PS1 PS2 PS3
Computational	Mut. breaks strand of computational evidence	Mut. breaks strand of computational	Mut. breaks strand of computational	Absent in population database PMS	Same amino acid change as an			
<div style="border: 2px solid blue; padding: 5px;"> <p>臨床的意義のカテゴリ</p> <p>※メンデル遺伝病に関わるバリアントに対してACMG/AMPで推奨される用語</p> <p>Benign … 病原性なし</p> <p>Likely benign … 病原性なしの可能性が高い</p> <p>Likely pathogenic … 病原性ありの可能性が高い</p> <p>Pathogenic … 病原性あり</p> <p>Uncertain significance … 解釈不明（上記のいずれにも分類できなかったもの）</p> </div>								
Other database	Observed in silico with a pathogenic variant BP2	Reputable source without shared data = benign BPS	Reputable source = pathogenic PPS	in trans with a pathogenic variant PMS		Benign		BS3 BS4 BP1 BP2 BP3 BP4 BP5 BP6 BP7
Other data	Found in case with an alternate cause BPS	Found in case with an alternate cause BPS	Patient's phenotype or FH highly specific for gene P04			Supporting		

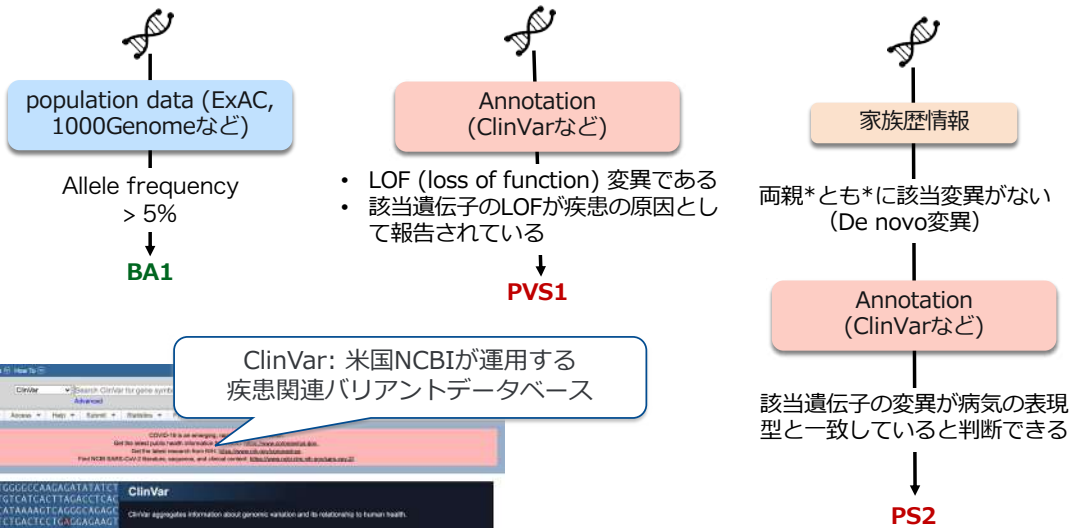
Richards, S. et al., 2015. *Genet. Med.* 17 (5): 405–424.

バリアントの疾患関連性の評価 ACMGガイドライン

米国臨床遺伝・ゲノム学会 (American College of Medical Genetics and Genomics; ACMG)

分類の例

BA1 (Benign: Stand-alone) **PVS1** (Pathogenic: Very Strong) **PS2** (Pathogenic: Strong)



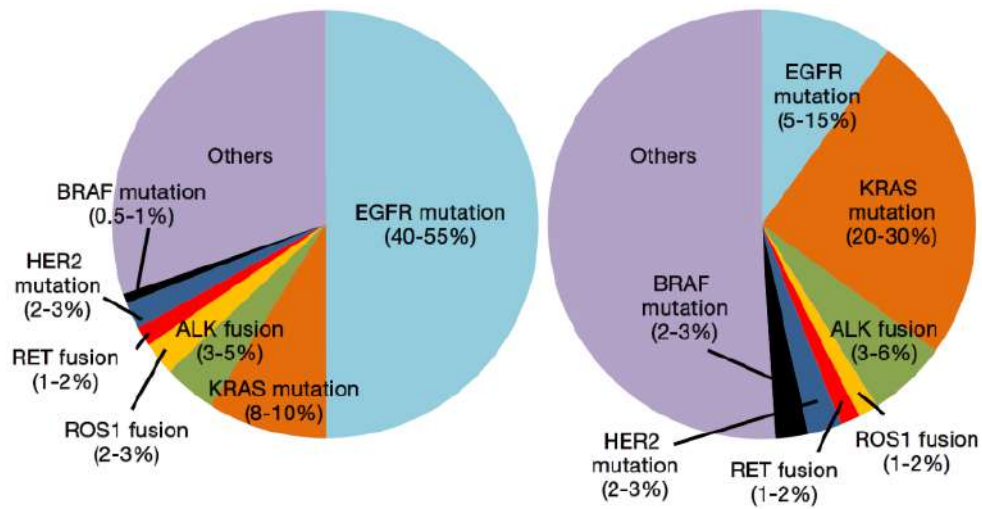
ClinVar: 米国NCBIが運用する疾患関連バリアントデータベース

<https://www.ncbi.nlm.nih.gov/clinvar/>

地域によって、遺伝的特徴が異なる

東アジア（日本・韓国・中国）

米国・ヨーロッパ



肺腺癌（LADC）におけるドライバー遺伝子の変異頻度の比較

Kohno, T. et. al., Beyond *ALK-RET*, *ROS1* and other oncogene fusions in lung cancer, *Transl. Lung Cancer Res.* 2015;4(2):156-164

日本人の遺伝的特徴を反映したデータベースが必要

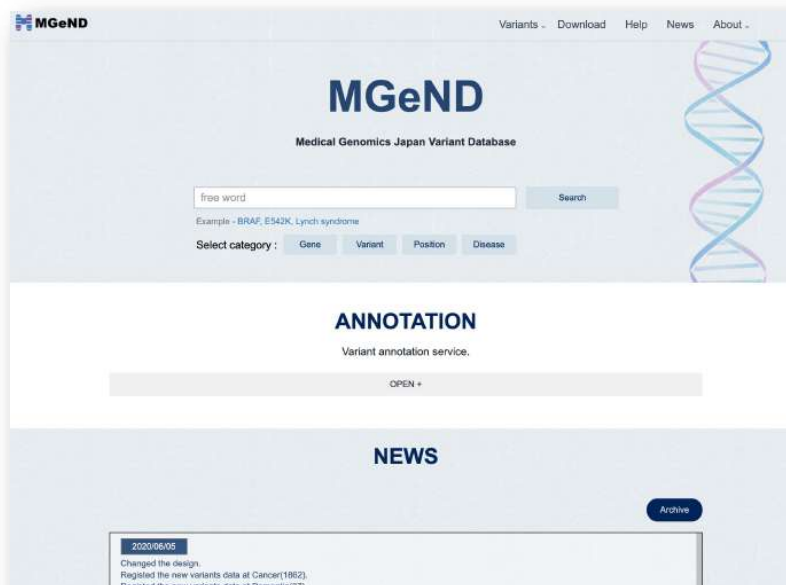
5

臨床ゲノム情報統合データベース

MGeND (Medical Genomics Japan Variant Database)

<https://mgend.med.kyoto-u.ac.jp/>

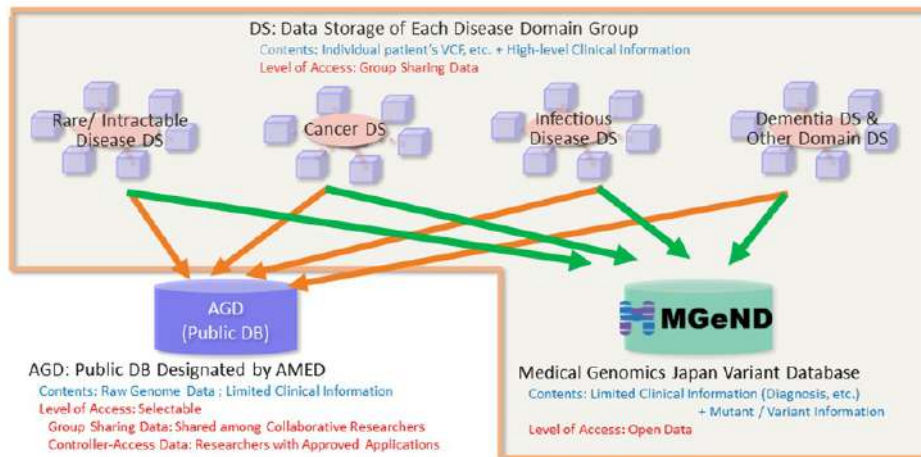
- ◆ 日本人疾患ゲノムの集積と世界への発信
- ◆ キュレーションに必要な情報の提供と臨床現場利用
- ◆ 多種多様な疾患やゲノム情報の統合的なデータベース化
- ◆ 疾患横断的な検索



6

AMED 臨床ゲノム情報統合データベース整備事業

AMED…国立研究開発法人 日本医療研究開発機構



AMEDデータシェアリングポリシー

- 制限共有データ
 - ✓ DMPに記載された研究者およびデータアクセス申請を承認された研究者と共有するデータ
 - ✓ 対象となる研究者は、データの生産や蓄積・活用に貢献できる研究者等
- 制限公開データ
 - ✓ 当該データベースの規約等に従って、利用目的、利用方法を明らかにしたうえで、データアクセス申請を承認された研究者が利用することが可能なデータ
- 非制限公開データ
 - ✓ アクセスに制限なく誰でも利用することが可能なデータ

MGeND

7

MGeND独自の疾患ゲノムデータフォーマット

各疾患に対応できる登録フォーマット・項目を策定

- **変異データ用**（検体単位・検体集団単位）
 - ClinVar登録フォーマットに準拠した項目の設定
 - 個人情報に考慮した統計値（頻度）での登録にも対応
 - xls, tsv, vcf, xml
- **GWASデータ用**
 - GWAS解析の共有に必要な項目を策定
 - 解析情報シート & 変異シート（xlsx形式）
- **HLAタイピングデータ用**（群間比較・集団単位）
 - HLAタイピング解析で標準的に用いられるテーブル形式を踏襲
 - Xlsx形式に対応

8

登録データ項目

- ◆ 疾患名
 - ✓ OMIM, MeSH, MedGen, Orphanet, HPO, ICD10, 自由記載
- ◆ 遺伝子名
- ◆ 遺伝型情報 (Genotype)
 - ✓ 1～数箇所程度のバリエーションまたは p 値 $<10^{-4}$ のSNP
- ◆ 付随する臨床情報
 - 年齢 (層)
 - 性別 (「不明」「混合」等を含む)
 - 臨床的意義
- ◆ 解析に関する詳細情報
 - ✓ 解析プラットフォーム・解析パネル・検定法など

9

データ登録ポリシー (ELSI対応済)

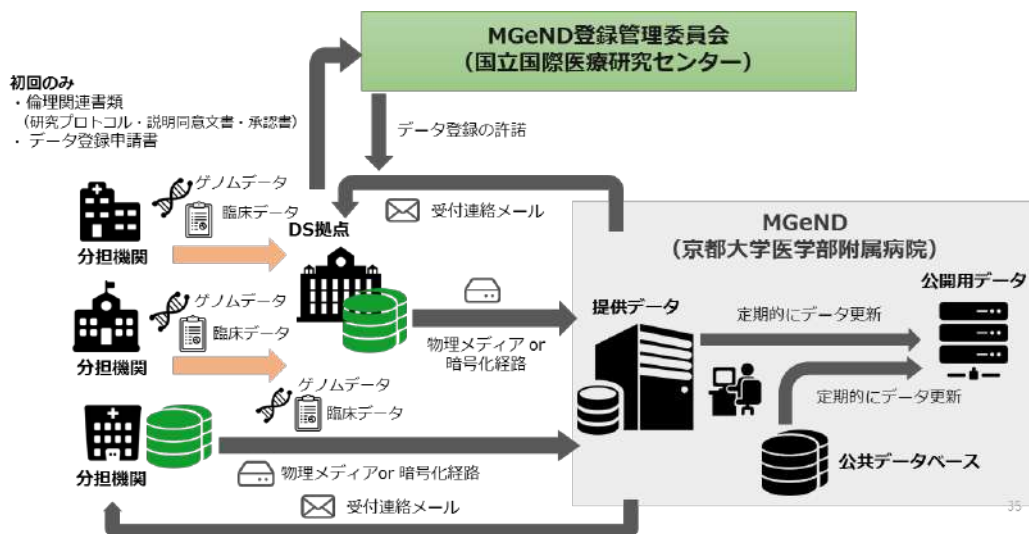
登録可能なデータ

以下のいずれかを満たすもの

- ◆ 本事業のために、患者の同意を得た上で取得されたバリエーションデータ
- ◆ 本事業以外で取得されたバリエーションデータで、以下の内容のいずれかが説明同意文書に入っているもの
 - (1) 解析結果をデータベースに登録すること
 - (2) 解析結果を学術論文で公表すること
- ◆ 学術論文や研究班の報告書、公的データベースなど、学術的に信頼された媒体ですでに発表されているバリエーションデータ

10

データ登録の流れ



データ受付の手順

1. データ登録機関からMGeND登録管理委員会へデータ登録申請書送付
2. MGeND登録管理委員会による倫理関連書類の確認
3. MGeND登録管理委員会によるデータ登録の承認
(1から3は初回のみ；ただし異なる研究計画によるデータは毎回必要)
4. DS拠点・データ登録機関からMGeND(京大)へデータ送付
5. MGeNDにて公開データベースにデータ登録

11

疾患ゲノム情報統合データベース MGeND

<http://mgend.med.kyoto-u.ac.jp>

MGeND

Medical Genomics Japan Variant Database

free word Search

Example - BRAF, E542K, Lynch syndrome

Select category : Gene Variant Position Disease

ANNOTATION

Variant annotation service.

OPEN +

NEWS

Archive

2020/06/05
 Changed the design.
 Registered the new variants data at Cancer(1882).
 Registered the new variants data at Dementia(37).
 Registered the new variants data at Other(233).

2019/12/09
 Changed the graph design.
 Added HIV drug resistance view.
 Registered the new variants data at Others (275).
 Bug fix.
 - Changed the number of registered variants data at Cancer (+15)

12

アクセス・利用ユーザーについて

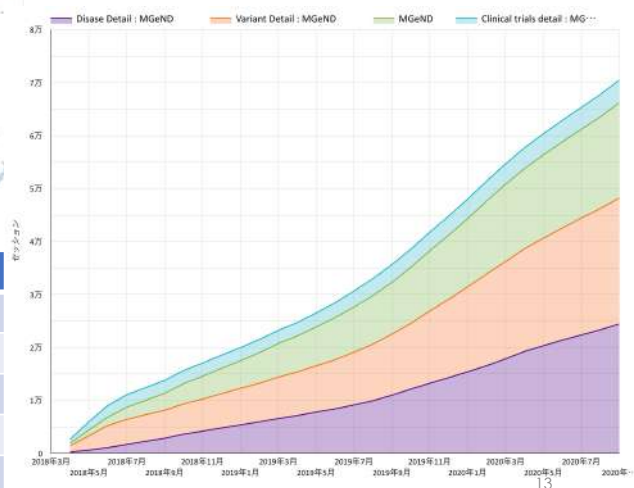
(2018/03/18 -- 2020/10/10)

➤ トータルセッション数 : 80,868

➤ 新規ユーザー (ユニーク) : 50,407 (49,982)



	Country
1	United states
2	Japan
3	India
4	United Kingdom
5	Germany



データ登録数

2020.09.29 時点

Data type	Variants	GWAS	HLA allele data
Disease area			
がん	158,901 (11,029)	---	---
希少／難病	15,170 (2,983)	---	---
感染症	1,115 (1,115)	155,100 (155,100)	1,979 (1,803)
認知症/難聴	11,408 (7,668) APOE: 12,553 (5,451)	410 (410)	---
その他	940 (833)	14,321,737 (---)	---

※ カッコ内の数値 ... 公開済み、及びデータシェアリングポリシーに基づきHOLD中

登録データ数：昨年より大幅に増加

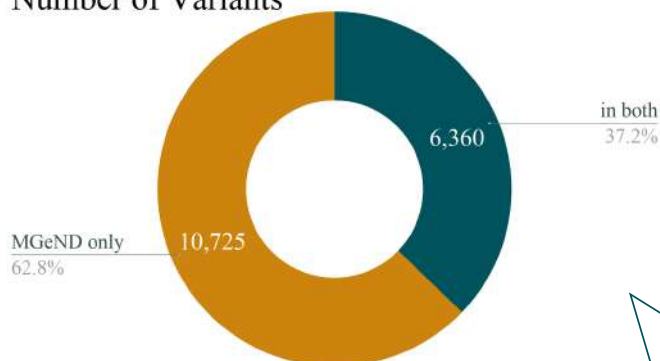
		2019.04.05 時点	2020.09.29 時点	
希少・難治	公開数	2,459	2,983	+524
	提供数	13,406	15,170	+1,764
がん	公開数	9,041	11,029	+1,988
	提供数	25,073	158,901	+133,828
感染症	公開数	23,359	158,018	+134,659
	提供数	157,082	158,194	+1,112
認知症・難聴	公開数	13,370	13,529	+159
	提供数	19,339	24,371	+5,032
その他 (昨年度より受付開始)	公開数	-	833	
	提供数	-	14,322,677	
合計	公開数	48,229	186,392	+138,163
	提供数	218,878	14,679,313	+14,460,435

15

MGeND登録データとClinVarの比較

- 使用データ：
 - ✓MGeND：2020.03時点
 - ✓ClinVar：2020.08.30 (vcf_GRCh37)

Number of Variants



ClinVar	MGeND	# of variants
Uncertain significance	Uncertain significance	1,452
	Benign	155
	Pathogenic	128
Likely benign	Benign	445
	Pathogenic	21
	Uncertain significance	367
Likely pathogenic	Benign	3
	Pathogenic	752
	Uncertain significance	113

16

MGeNDの主な機能： 疾患横断的な検索と国内外DB情報一覧が確認可能

2020/06/09
Changed the design.
Registered the new variants data at Cancer(1862).
Registered the new variants data at Dementia(37).
Registered the new variants data at Other(233).

2019/12/09
Changed the graph design.
Added HIV drug resistance view.
Registered the new variants data at Others (275).
Bug fix.
- Changed the number of registered variants data at Cancer (+15)

- フリーワード検索に加え、検索カテゴリの指定が可能
✓ Gene / Variant / Position / Disease
- カテゴリの指定により、検索時点で検索対象の絞り込みが可能
- rsIDやSubmission IDでの検索も可能

17

検索結果表示

Search Word

Target data : MGeND data only
Category :
Search word : BRAF
Filtering :

Filtering

Showing 1 to 25 of 88 entries Show 25 entries

Variant name	Gene	AA change	MGeND				ClinVar
			Entry	Origin	Type	Annotation	Annotation
NC_000007.13:g.140481376_140481377insT	BRAF		1	unknown	Variant	not provided	
NC_000007.14:g.140753345A>C	BRAF	p.L597R	1	somatic	Variant	Pathogenic	Pathogenic Likely pathogenic ★★★★
NC_000007.13:g.140481411C>G (rs121913351)	BRAF	p.G466A	1	somatic	Variant	Pathogenic	Likely pathogenic ★★★★
NC_000007.13:g.140481411C>T (rs121913351)	BRAF	p.G466E	1	somatic	Variant	Pathogenic	Likely pathogenic ★★★★
NC_000007.14:g.140781602C>T	BRAF	p.G469E	1	somatic	Variant	Pathogenic	Pathogenic Likely pathogenic ★★★★
NC_000007.13:g.140487360G>A (rs397507472)	BRAF	p.R389C	1	somatic	Variant	Uncertain significance	Uncertain significance ★★★★
NC_000007.13:g.140534597C>T (rs749247588)	BRAF	p.G106R	1	somatic	Variant	not provided	Uncertain significance ★★★★
NC_000007.13:g.140453136_140453137delinsCT (rs121913227)	BRAF	p.V600R	1	somatic	Variant	Pathogenic	Pathogenic ★★★★
NC_000007.13:g.140453174G>T (rs121913336)	BRAF	p.D587E	1	unknown	Variant	Likely pathogenic	Likely pathogenic ★★★★

18

フィルタリング機能の充実

The screenshot shows the MGeND search results page for the keyword 'BRAF'. A 'Filtering' dropdown menu is highlighted, with a callout box listing various filter categories:

- アセンブリ (Assembly)
- 細胞種 (Cell Line)
- 解析手法 (Analysis Method)
- 遺伝子名 (Gene Name)
- 疾患領域 (Disease Area)
- 疾患名 (Disease Name)
- 関連薬剤 (Associated Drug)

The main results table shows variant names such as 'NC_000007.13:g.14048137>481377insT' and 'NC_000007.14:g.140753345A>G'. A detailed view of a variant is shown on the right, with callouts for 'アセンブリ' (Assembly), '細胞種' (Cell Line), '解析手法' (Analysis Method), '遺伝子名' (Gene Name), '疾患領域' (Disease Area), '疾患名' (Disease Name), and '関連薬剤' (Associated Drug). The detailed view includes fields for 'Gene', 'Transcript', 'Protein', 'Clinical Significance', and 'Review star'.

19

変異詳細ページ

e.g. PIK3CA p.Glu545Lys (p.E545K)

The screenshot shows the detailed view of the variant 'PIK3CA p.Glu545Lys (p.E545K)'. The page is divided into several sections:

- Information:** Shows genome assembly (hg19, hg38) and position (chr3:178,218,303-178,218,303).
- HGVS:** Shows the variant name and its representation in different databases (RefSeq, Ensemble).
- Summary:** Shows clinical significance (Pathogenic), variant ID (s), and disease area statistics.
- Frequency:** Shows the frequency of the variant (No Data).
- Prediction:** Shows a color-coded prediction score (e.g., 100% pathogenic).
- ClinVar:** Shows clinical significance (Pathogenic) and review star (4 stars).
- Links:** Provides links to various databases (MM, HGNC, Ensembl, NCBI, Gene Cards, OncoKB) and variant databases (ClinVar, COSMIC).

Callouts highlight specific features:

- HGVSによる表記:** Points to the HGVS representation of the variant.
- 頻度情報・計算機による予測スコア:** Points to the frequency and prediction score.
- MGeND登録情報:** Points to the MGeND registration information.
- ClinVar情報:** Points to the ClinVar information.
- 国内外DBへのLINK:** Points to the links to various databases.

変異詳細ページ

データ登録情報

e.g. PIK3CA p.Glu545Lys (p.E545K)

MGeND

Showing 1 to 4 of 4 entries Show 10 entries

Clinical significance	Last evaluated	Condition	Origin	Submission ID	Submitter	Institute	Citation
Pathogenic	2018/04/12	Malignant transformation of mature cystic teratoma	somatic	MGS000017	Kohei Miyazono	Tokyo University	
Pathogenic	2018/03/22	squamous cell carcinoma	somatic	MGS000017	Kohei Miyazono	Tokyo University	
Pathogenic	2018/03/22	uterine body cancer	somatic	MGS000017	Kohei Miyazono	Tokyo University	
Pathogenic	2017/11/09	Endometrioid adenocarcinoma	somatic	MGS000017	Kohei Miyazono	Tokyo University	

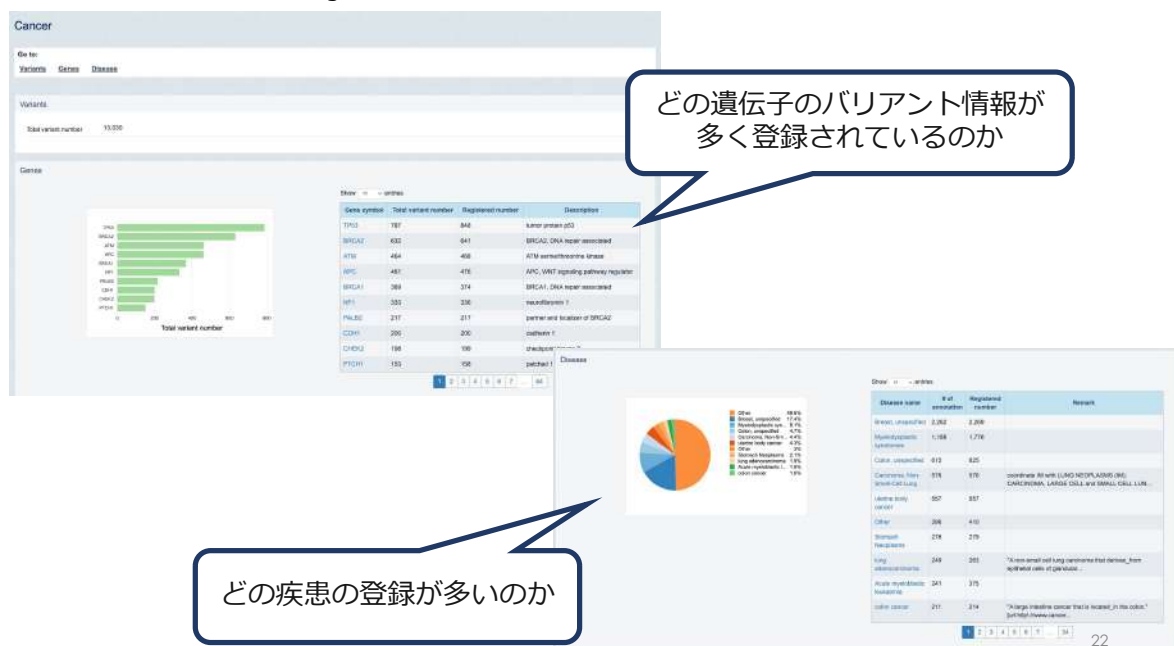
- Clinical Significance, Condition, Citationとともに、Submission IDとデータ登録者情報を表示
- Submission ID :
 - ✓ データ登録拠点情報と研究スタディ情報に基づき発行
 - ✓ 研究スタディ単位でのID
- 論文投稿時の変異情報登録先データベースへ

21

疾患領域ページ

- 各疾患領域ごとの登録遺伝子、疾患の統計情報を表示

疾患別統計情報画面の例 (e.g. がん)



どの遺伝子のバリエーション情報が多く登録されているのか

どの疾患の登録が多いのか

疾患固有画面

HLA allele frequency table showing columns for Disease, Allele, Frequency, and other metrics.

感染症領域：HLAアリル頻度テーブル

認知症領域：APOE遺伝型頻度グラフ



MGeND Drug Resistance interface showing categories: Nucleoside Reverse Transcriptase, Non-Nucleoside Reverse Transcriptase, Protease, and Integrase. Each category has a list of drugs and their resistance mutations.

感染症領域：HIV薬剤耐性ビュー

MGeNDの主な機能： 手持ちのバリエントに対するアノテーションが可能

MGeND website interface showing the 'ANNOtATION' section with a search bar and 'OPEN +' button, and a 'NEWS' section with a list of updates.

- 手持ちのVCF/TSVにMGeND収録情報をアノテーションできる機能
- ダウンロードする情報の選択が可能に

アノテーションサービス

Genome: hg19

Search target variant: MGaND data only All data

Search pattern

Annotate VCF

```
##INFO=ID=412,Number=0,Type=Flag,Description="HapMap2 membership"
##FILTER=ID=q10,Description="Quality below 10"
##FILTER=ID=s50,Description="Less than 50% of samples have data"
##FORMAT=ID=GT,Number=1,Type=String,Description="Genotype"
##FORMAT=ID=GQ,Number=1,Type=Integer,Description="Genotype Quality"
##FORMAT=ID=DP,Number=1,Type=Integer,Description="Read Depth"
##FORMAT=ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality"
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
1 215847775 . C T 50 PASS NS=3,DP=9,AA=G,GT:SQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
1 241667337 . T A 50 PASS NS=1,DP=9,AA=A,GT:SQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
3 178836078 . C G 50 PASS NS=1,DP=13,AA=T GT:SQ:DP:HQ 00254:7:16:50 00:46:4:31,51 0/0:61:2
5 112173368 . A T 50 PASS NS=1,DP=9,AA=A,GT:SQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

Annotate

Example

検索対象とするバリエントを選択可能

VCFもしくはTSV形式で変異のリストを入力

VCF Annotation Result

Genome: hg19

Search target variant: MGaND data only

Showing 1 to 10 of 19 entries Show entries

File#	chr	Variant name	Gene	AA change	CDS	Apprentice frequency	Tagger	Entry	Origin	Type	Annotation	Entry	Origin	Annotation	CIVIC evidence	dbSNP entry	COSSMIC occurrence	RegulDB interact	dbSNP prediction
	chr12	g.2446285C>T	KIF4A	p.G12S	ENST00000286708.8 c.34G>A	0.00		1	unknown	Variant	not provided	1	somatic	Pathogenic	0	102		MODERATE	missense_variant
	chr12	g.25192285C>T	KIF4A	p.G12S	NM_004985.4 NM_023983.3 c.34G>A	0.00		3	somatic	Variant	not provided	2	somatic	Pathogenic	0	25	102	MODERATE	missense_variant
	chr12	g.25194180C>T	KIF4A	p.G12S	ENST00000399131.1 c.34G>A	0.00		3	somatic	Variant	not provided	8	somatic	Pathogenic	0	25	102	MODERATE	missense_variant
	chr12	g.25194180C>T	KIF4A	p.G12S	ENST00000399131.5 c.34G>A	0.00		8	somatic	Variant	not provided	8	somatic	Pathogenic	0	25	102	MODERATE	missense_variant
	chr11	g.17884773C>T	USH3A	p.R499H	NM_209933.2 c.124765G>A	0.00	1g-41	1	germline	Variant	not provided	1	germline	Benign	0	0		MODERATE	missense_variant
	chr9	g.17984070C>G	PRKCA	p.L540V	NM_006209 c.1816C>G	0.00		1	germline	Variant	not provided	1	germline	Benign	0	0		MODERATE	missense_variant
	chr9	g.17984070C>G	PRKCA	p.L540V	ENST00000300000 c.1816C>G	0.00		1	germline	Variant	not provided	1	germline	Benign	0	0		MODERATE	missense_variant
	chr5	g.12173368A>T	APC	p.R682*	NM_000098 c.2577A>T	0.00		1	germline	Variant	not provided	1	germline	Benign	0	0		HIGH	stop_gained
	chr6	g.13939410C>A	UVSSB	p.R131Y	NM_021225 c.390C>A	0.00	1g-047891-12	79	G SNVS	Variant	not provided	0	germline	Benign	0	0		MODERATE	missense_variant

現在は100件までの結果を表示

ダウンロードするバリエント・情報を選択可能

今後の機能改良について

- APIの整備と公開
 - ダウンロードデータの充実と整理
 - 疾患名の統合について検討
- 多因子疾患対応の強化
 - 研究スタディごとのGWASページの作成
- アセンブリ間 (hg19/hg38) の紐付け
- 最新データベース (ToMMo等) の取り込み

ヒトバリエーションデータの登録と 利活用に向けたガイダンス

大阪大学大学院医学系研究科
加藤 和人

本日の話題

- MGeNDデータ利活用ガイダンス概要
- MGeNDデータ登録のながれ
 - ✓ データ登録フローチャート
 - ✓ MGeND登録申請書
 - ✓ すぐに登録可能なデータ
- ELSI面での対応
 - ✓ MGeND利用規則へのELSI面からの支援
 - ✓ ELSI理解を拡げる活動(勉強会の開催)
 - ✓ MGeNDホームページから

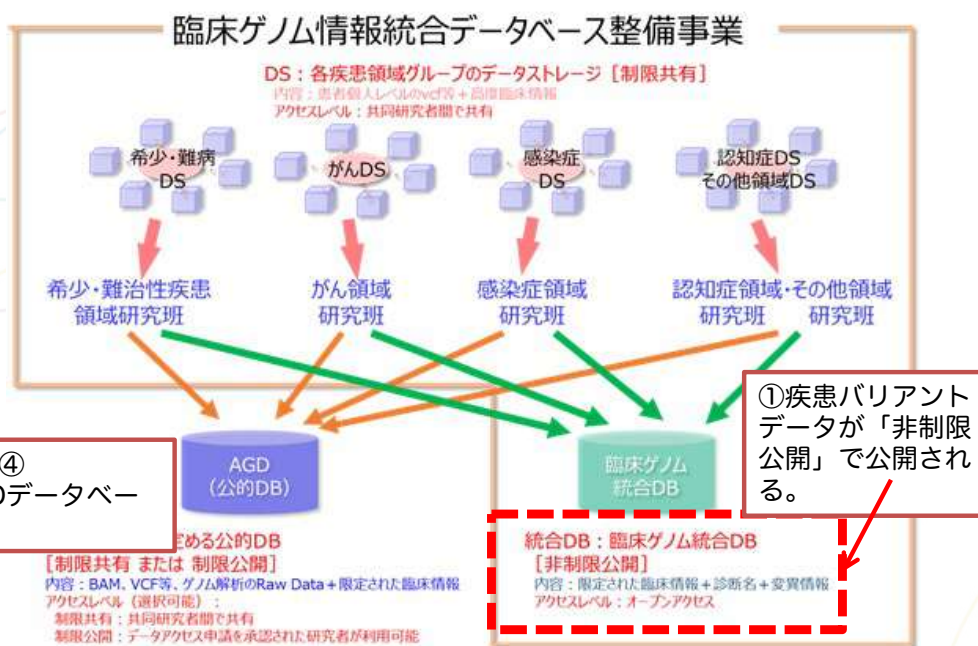
MGeNDデータ利活用ガイドンス概要

- 本ガイドンスは、臨ゲノ2次溝上班の分担研究者である大阪大学医学系研究科・加藤和人のグループが素案を作成した後、三省ゲノム指針事務局（と個人情報保護委員会）とのやり取りを経て、出来上がったものです。
- 重要な点のいくつか、（例：提供研究者→国内へのデータ提供によりDBへ→DBが国内外に非制限公開でデータ公開）は、個人情報保護法および研究倫理指針の解釈に沿っているため、三省ゲノム指針事務局と個人情報保護委員会とのやり取りの中で提案されました。
- 本ガイドンスは、今後も、法律・指針に沿いつつ、研究現場で運用しやすいように改訂していく予定ですので、ご質問・ご要望など、遠慮なくお知らせください。

MGeND用データ登録ガイドンス
(詳細版)
(2ページ)

[2018年3月
最終版]

日本人疾患バリエントデータベース(①MGeND)



MGeND用データ登録ガイドンス
(詳細版)
(5ページ)

[2018年3月
最終版]

非制限公開用バリエーションデータの流れ



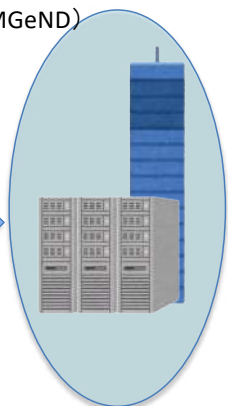
提供元
(参加施設)

提供先(統合DB
MGeND)

提供先(不特定多数)



疾患に関する
バリエーションデータ*を
データベースに提供



誰でも
制限なく
データを利用可能
('非制限
公開')



MGeND用データ
登録ガイダンス
(詳細版)
(6ページ)

[2018年3月
最終版]

*バリエーションデータ:「疾患名、遺伝子名、1~数箇所程度のSNP等の遺伝型情報、(年齢)、(性別)」の項目をもつデータ

5

データベースを介した利用を行うデータの項目



- 疾患名 (標準的な記載方法のもの)
- 遺伝子名
- 遺伝型情報 (Genotype) (上記遺伝子に関する1~数箇所程度のSNP等の多型や変異情報)
- (年齢 (層))
- (性別 (「不明」「混合」等を含む))

MGeND用データ
登録ガイダンス
(詳細版)
(7ページ)

[2018年3月
最終版]

- ✓ 上記の項目を内容とするデータを、本ガイダンスでは、「バリエーションデータ」とよぶ。
- ✓ バリエーションデータは、診療録、研究データ、レガシーデータ等から抽出し、本データベースを介した利用を行う。
- ✓ 提供する際のバリエーションデータには、提供元で付与したID等が含まれる場合があるが、DB側で公開するデータは、上記囲み内の3 (あるいは5) 項目のみ。

6

MGeNDデータ登録申請

データ登録フローチャート

登録申請書

すぐに登録可能なデータ

MGeNDホームページから [https://mgend.med.kyoto-u.ac.jp]



“Submission”
→JPメニュー(日本語)



データ登録について

1. 登録可能なデータ

MGeNDでは疾患バリエーションデータを広く収集・登録することを目指しております。しかし、非制限公開のデータベースであることから、AMED「臨床ゲノム情報統合データベース整備事業」にて作成したデータ登録ポリシーを満たし且つMGeND登録委員会（以下、登録委員会）にて承認されたデータのみ登録を受け付けています。また、お持ちのデータが登録可能であるか以下のデータ登録ガイドライン（簡易版）およびデータ登録申請書をご確認ください。ご不明な点は、MGeND登録管理委員会（objimu@hosp.ncgm.go.jp）にご連絡ください。（簡易版）およびデータ登録申請書をご確認ください。ご不明な点は、MGeND登録管理委員会（objimu@hosp.ncgm.go.jp）にご連絡ください。

- データ登録ガイドライン（簡易版）(*1)
- データ登録申請書(*1)

より詳細な情報については下記をご覧ください。

- データ登録ガイドライン（詳細版）(*1)
- データ登録ガイドライン整理表(*1)

2. データ登録手順

上記ガイドラインに基づき、登録可能なデータである場合、登録者は登録委員会にデータ提供を申請します。MGeNDは登録委員会からの承認連絡を受け登録者に登録手順を連絡します。おおまかな流れは下記の通りです。

- MGeND登録管理委員会へデータ登録申請書を送付: objimu@hosp.ncgm.go.jp
- MGeND登録管理委員会より確認を行います（倫理審査情報などから登録可能かのみを確認）
※必要に応じて登録管理委員会より倫理関連書類の提出を求める場合があります
- MGeND登録管理委員会によるデータ登録承認の連絡（1から3は初回のみですが、研究計画単位で新たに申請が必要です）
- MGeNDより登録手順について連絡、以下のデータ登録フォーマットに基づきデータをご準備いただき、登録

* 難病・稀少疾患領域の寄稿：MGeNDのご案内と臨床ゲノム情報統合データベース整備事業へのご協力をお願いいたします。 **難病班への紹介資料**

MGeNDデータ登録ガイドンス (簡易版)



(表) MGeND: 疾患バリエーションデータ登録のお願い

ゲノム情報をもとにした医療や、その基盤となる研究を進めるためには、疾患バリエーションデータを広く収集・登録したデータベースが必要で、AMED「臨床ゲノム情報統合データベース整備事業」2次・3次・4次登録の際に参照していただける「データ活用ガイド」を作成しました。以下はその概要です。積極的なデータの提供をお願いします。

登録をお願いしたいデータ

- 疾患名 (標準化された名称が望ましい)
- 遺伝子名
- 遺伝型情報 (Genotype)
- 年齢 (層)
- 性別 (「不明」「割合」も可)

登録 (提供) 先:
MGeND 非制限公開データベース
URL: <https://mgend.med.kyoto-u.ac.jp>
登録 (提供) に関する問い合わせ先:
MGeND登録管理委員会
Email: dbjimu@hosp.ncgm.go.jp

- 臨床ゲノム情報統合データベース整備事業開始後 (本事業のために)、パネル解析やエクソーム解析など広範なゲノム解析の実施について同意を得た上で、ゲノム解析し、取得されたバリエーションデータ
→そのままで、すぐに登録可能です。
- 臨床ゲノムデータベース事業以外の研究や解析により取得されているバリエーションデータで、以下の内容が患者 (研究参加者) への説明同意書に入っているデータ
(1) 「解析した結果はデータベースに登録します」または、
(2) 「解析した結果は、学術論文等で発表します」
(※「学術的な価値が定まり、研究実績として十分に認められ、研究用に広く一般に利用され、かつ、一輸入不可な情報」として、ゲノム解析対象外)
→そのまゝ、登録可能です。 (対象となる報告書など詳細はお問い合わせください)
- 学術論文や研究報告書など、学術的に信頼された媒体ですすでに発表されているバリエーションデータ (自身の論文や参加された研究班による公的な報告書に加え、他の研究者が発表した論文や報告書から取得したデータを含む)
(※「学術的な価値が定まり、研究実績として十分に認められ、研究用に広く一般に利用され、かつ、一輸入不可な情報」として、ゲノム解析対象外)
→そのまゝ、登録可能です。 (対象となる報告書など詳細はお問い合わせください)
- 上記に当てはまらないデータをご提供になりたい場合も、遠慮なくお問い合わせください (例: 論文未発表、かつ、上記1、2に該当しない、等)

詳細は「MGeNDデータ利用ガイド」を参照ください。
国立国際医療研究センター・清上雅志
東京大学大学院医学系研究科・櫻井裕士
大阪大学大学院医学系研究科・加藤和人
京都大学大学院医学系研究科・奥野昌史
慶応義塾大学医学部・小崎健太郎

April 2018 ver2.0

(裏) (判断用フローチャート) April 2018 ver2.0 臨床ゲノム1次班用

皆様がお持ちの疾患バリエーションデータは、どのようなものでしょうか?

学術論文や研究報告書など、学術的に信頼された媒体ですすでに発表されているデータである。(表裏3) (臨床ゲノムデータであるかどうかは問いません)

Yes → 臨床ゲノム情報統合データベース整備事業開始後 (本事業のために)、パネル解析やエクソーム解析など広範なゲノム解析の実施について同意 (※) を得た上で、ゲノム解析し、取得されたデータである。(表裏1) → データベースへの提供をまむ

No → 臨床ゲノム情報統合データベース整備事業以外の研究や解析により取得されたデータで (広範なゲノム解析の実施について同意はないが)、(1) 解析した結果はデータベースに登録します または (2) 解析した結果は、学術論文等で発表します という内容が説明同意書に入っているデータである。(表裏2) → 登録の価値があるデータをお持ちの場合は、手続きについて助言します。ご相談ください。上記1-3に当てはまらないデータ (表裏4) も、倫理審査委員会承認とオプトアウトで提供可能な場合があります。

疾患バリエーションデータはすぐに提供できます。MGeND登録委員会にメールでご連絡いただき、登録申請をお願いします。(簡単な申請用紙に記入、提出していただきます。委員会での確認終了後、データの受け取りを行います。)

※データを提供したため、データベース上で登録・公開されるまで、2週間程度かかります。ご了承ください。

判断に迷われる場合は、MGeND登録管理委員会にお問い合わせください。(メールアドレスは表裏に掲載)

難病班への紹介資料

疾患バリエーションデータベースMGeNDのご案内
臨床ゲノム情報統合データベース整備事業への協力をお願いします

本データベースは、患者のゲノム情報から疾患の原因となる変異を特定し、診断や治療に役立てることを目的として構築されています。また、ゲノム解析の結果をデータベースに登録し、広く一般に公開することで、ゲノム解析の利便性を高め、ゲノム解析の普及を促進することを目的としています。

本データベースに登録するデータは、以下の条件を満たす必要があります。

- 学術論文や研究報告書など、学術的に信頼された媒体ですすでに発表されているデータであること。
- 臨床ゲノム情報統合データベース整備事業開始後 (本事業のために)、パネル解析やエクソーム解析など広範なゲノム解析の実施について同意を得た上で、ゲノム解析し、取得されたデータであること。
- 臨床ゲノム情報統合データベース整備事業以外の研究や解析により取得されたデータで (広範なゲノム解析の実施について同意はないが)、(1) 解析した結果はデータベースに登録します または (2) 解析した結果は、学術論文等で発表します という内容が説明同意書に入っているデータであること。

本データベースに登録するデータは、以下の条件を満たす必要があります。

- 学術論文や研究報告書など、学術的に信頼された媒体ですすでに発表されているデータであること。
- 臨床ゲノム情報統合データベース整備事業開始後 (本事業のために)、パネル解析やエクソーム解析など広範なゲノム解析の実施について同意を得た上で、ゲノム解析し、取得されたデータであること。
- 臨床ゲノム情報統合データベース整備事業以外の研究や解析により取得されたデータで (広範なゲノム解析の実施について同意はないが)、(1) 解析した結果はデータベースに登録します または (2) 解析した結果は、学術論文等で発表します という内容が説明同意書に入っているデータであること。

MGeND 登録申請書

MGeND:疾患バリエーションデータ:登録申請書(案)		MGeND登録申請番号:
申請者情報 日付 申請書名 所属機関		データ登録者情報(申請者、所属)
AMED臨床ゲノム情報統合データベース整備事業・参加情報(該当する場合) 研究開発・分担課題番号(年度) 代表者名 申請書名(申請書と異なる場合)		AMED臨床ゲノム事業への参画について
データ公開ステータス(以下から選択) <input type="checkbox"/> 学術論文等で既に発表 → 表(A)のみ記入して下さい <input type="checkbox"/> AMED臨床ゲノムデータベース事業で得られたデータ → 表(A)、および表(C)に記入して下さい <input type="checkbox"/> AMED臨床ゲノムデータベース事業以外で得られたデータ → 表(A)、表(B)および表(C)に記入して下さい		データ公開ステータス(論文発表等)
表(A)バリエーションデータ (登録管理委員会での確認後、詳細をMGeNDの事務局にお送りいただけます) 疾患 遺伝子 (GWASの場合はGWASと記入) 発症年齢 (2歳以上) (注) 性別 1 2 3 4 5		表(A)バリエーションデータ情報 (疾患、遺伝子、論文情報)
表(B)説明同意書情報 (説明同意書にICに沿った提供内容であるかどうかについての確認は提供者が行うこととします) 説明同意書の内容を抜き出して記載してください。あるいは、当該部分が記載されたページのコピーなどを添付していただいても結構です <input type="checkbox"/> 学術論文発表に関する同意 <input type="checkbox"/> データベース公開に関する同意		表(B)説明同意書情報 (論文発表、データベース公開についてのIC)
表(C)倫理審査情報 (最終) 日付 委員会名 承認番号		表(C)倫理審査情報 (最終審査の日付、委員会名、承認番号)

すぐに登録可能なデータ



2. 臨床ゲノムデータベース事業以外の研究や解析により取得されているバリエーションデータで、以下の内容が患者（研究参加者）への説明同意文書に入っているデータ

- (1) 「解析した結果は**データベースに登録**します」または、
- (2) 「解析した結果は、**学術論文等で発表**します」

(※指針等の解釈により、国内データベースへの提供の同意は取得できていると判断されます。機関の倫理審査委員会への報告も不要です)

→上記の同意内容を確認していただいた上で、すぐに登録可能です。

3. 学術論文や研究班の報告書など、学術的に信頼された媒体ですでに発表されているバリエーションデータ（自身の論文や参加された研究班による公的な報告書に加え、他の研究者が発表した論文や報告書から取得したデータを含む）

(※「学術的な価値が定まり、研究実績として十分に認められ、研究用に広く一般に利用され、かつ、一般入手可能な」情報として、ゲノム指針対象外)

→そのまま、登録可能です。（対象となる報告書など詳細はお問合せください）

(判断用フローチャート)

April 2018 ver2.0 (臨ゲノ 1次班用)



皆様がお持ちの疾患バリエーションデータは、どのようなものでしょうか？

学術論文や研究班の報告書など、学術的に信頼された媒体で**すでに発表**されているデータである。
(臨ゲノのデータであるかどうかは問いません)

Yes

No

臨床ゲノム情報統合データベース整備事業開始後（本事業のため）に、パネル解析やエクソーム解析など広範なゲノム解析の実施について**同意（※）を得た**上で、ゲノム解析し、取得されたデータである。

※データベースへの提供を含む

Yes

No

臨床ゲノム情報統合データベース整備事業以外の研究や解析により取得されたデータで（広範なゲノム解析の実施についての同意はないが）、「(1) 解析した結果は**データベースに登録**します」または「(2) 解析した結果は、**学術論文等で発表**します」という内容が説明同意文書に入っているデータである。

Yes

No

疾患バリエーションデータはすぐに提供できます。 MGeND登録委員会にメールでご連絡いただき、登録申請をお願いします。
(簡単な申請用紙に記入、提出していただきます。委員会での確認終了後、データの受け取りを行います※)

登録の価値があるデータをお持ちの場合は、手続きについて助言します。ご相談ください。上記1-3に当てはまらないデータも、倫理審査委員会承認とオプトアウトで提供可能な場合があります。

※データを提供いただいてから、データベース上で登録・公開されるまで、2週間程度かかります。ご了承ください。

判断に迷われる場合は、MGeND登録管理委員会にお問い合わせください。

◆ 本人同意を得ている場合（1）

提供元施設において、**臨床ゲノム情報統合データベース事業全体の内容**について説明を行い、患者や研究参加者から同意を得たとき。

例) 本データベースについては、バリエーションデータを提供することを研究計画書とインフォームドコンセントに記載し、倫理審査委員会の承認と機関の長の許可を得た。また、そのインフォームドコンセントによって本人同意を取得した。

(さらに、本データベースが海外の研究者等にも利用されることが説明されていれば、なおよい。今後の変更申請において含めることを推奨する)

MGeND用データ
登録ガイダンス
(詳細版)
(19ページ)

[2018年3月
最終版]

→ 本データベース(国内DB)へ提供可能

13

◆ 本人同意を得ている場合（2）

提供元施設において、以前に取得したインフォームドコンセントのなかで、すでにデータベースへの提供についての同意が得られているとき。

例) 「データベースに登録する。」「論文等で発表する。」等の文言が以前のインフォームドコンセントに記載されているとき。

このような場合、すでに(国内のDBへの提供の)同意が得られているとする。

「論文等で発表する」について: データベース登録が論文発表の条件になっている場合があることや、論文からキュレーターが収集したデータがデータベースに登録される場合があること、また、今回提供を呼びかけるデータは、「疾患名・遺伝子名・遺伝子型」であるため、使用するデータの量や種類が論文発表と同等かそれより少ないと考えられること、などの状況を鑑み、本データベースに関しては、「論文等で発表する」の同意によって、**本データベース(国内のデータベース)への提供の同意が得られているものとする。**

→ 新たな同意や倫理審査等の手続きを行うことなく、本データベース(国内DB)へ提供可能

(ただし上記は、**データが誰のものか分からないようにした上でデータベースに登録する・論文発表する、という内容についての同意である。したがって、超希少疾患等でデータ自体から個人識別性が十分に除けないときには、同意が得られていないと考え、再同意または倫理審査委員会の承認を得てオプトアウトのいずれかを行えば、提供可能**)

MGeND用データ
登録ガイダンス
(詳細版)
(20ページ)

[2018年3月
最終版]

14

ELSI面での対応

MGeND利用規則へのELSI面からの支援
ELSI理解を拡げる活動(勉強会の開催)
MGeNDホームページから

MGeNDデータ利用規則へのELSI面からの支援

MGeND利用規則策定にあたり、倫理的側面からの支援を行い、主に下記の各点について検討、記載を提案した。

- ①利用目的の限定
- ②営利目的での利用の事前相談
- ③クレジット
- ④知的財産権取得の事前相談
- ⑤費用負担
- ⑥免責
- ⑦違法利用への免責
- ⑧有効になった日付の記載

(右図内の赤線)

2018年度から継続的に検討・提案
→MGeNDサイトから公開

MGeNDデータ利用規則 (2020.5.21版)



データ利用規則

MGeNDは以下の原則に基づき利用が可能です。

1. 利用範囲
1.1 誰でも利用可能です。
2. データ利用者の権利
2.1 データ利用者は、MGeNDのデータを利用した研究成果をデータ利用者の発表及びデータが追加された制約事項を遵守する限り自由に発表および持許できる。
2.2 データ利用者は、MGeNDのデータを利用した研究成果をもとにした知的財産権を、データ利用者の発表及びデータが追加された制約事項を遵守する限り自由に取得できる。
3. データ利用者の責務
3.1 データ利用に際してのデータの盗取・内容・科学的妥当性について、データ利用者の責任と判断のもとで活用すること。
3.2 データ利用者は以下の事項を遵守すること。
 - ・ 個人情報の禁止
 - ・ 複製データのダウンロード及び使用
 - ・ 複製、改変および第三者への提供
 - ・ 複製、改変および第三者への提供
- 3.3 MGeNDを通じて提供されたデータを含む最終結果を公表または提供する際は、MGeNDを通じて提供されたデータを引用したことを明記すること。また、本ページで示す文脈を引用すること。
- 3.4 MGeNDを通じて提供されたデータを利用して知的財産権を取得する際には、事前に相談のこと。
4. 利用に関する責任
4.1 データ利用に際して発生する責任はすべてデータ利用者が負担することとする。
5. 免責
5.1 MGeNDで提供される情報の正確性、完全性、または最新性については責任を負わず、情報は予告なく変更される場合がある。
5.2 MGeNDに関する情報は、適時的な更新または取崩しの上で変更決定を目的としたものではない。
5.3 利用者は、MGeNDで提供される情報の取扱い、その制約事項に常に準拠する旨を、規定及びガイドライン、MGeNDの著作権管理についての定めを遵守しなければならない。
5.4 データ利用により発生したいかなる費用についてもMGeND及びデータ管理機構MGeNDにデータを提供した研究者等は責任を負わない。

MGeND登録のELSI理解を拡げる活動 ＜勉強会の開催＞



大阪大学
OSAKA UNIVERSITY

OPEN 2021

- 「2019年度ヒトバリエーションデータ研究倫理勉強会」 開催
日時：2019年10月9日（水） 10:30-12:30
会場：毎日インテシオ4F 大会議室(大阪)
参加者： 58 名（事前参加登録:62名）
- 「2020年度ヒトバリエーションデータ研究倫理勉強会」 開催
日時：2020年10月17日(土)10:30-12:00 ←**本会**
事前参加登録： 377 名（10/14時点）
開催形式：ウェビナー（オンラインセミナー）
- **第43回日本分子生物学会年会フォーラム企画開催(予定)**
「ゲノム研究と医療をつなぐデータベース」
- ヒト疾患関連データ利活用とELSIを考える -
日時：2020年12月4日（金） 18:30-20:00
セッション番号：3F-08（オンライン開催）



2019年度勉強会の様子(総合討論)

最新情報を共有するため
継続して勉強会を開催し
ています！

17

勉強会開催による理解の拡がり ＜2019年度勉強会・参加者アンケート＞

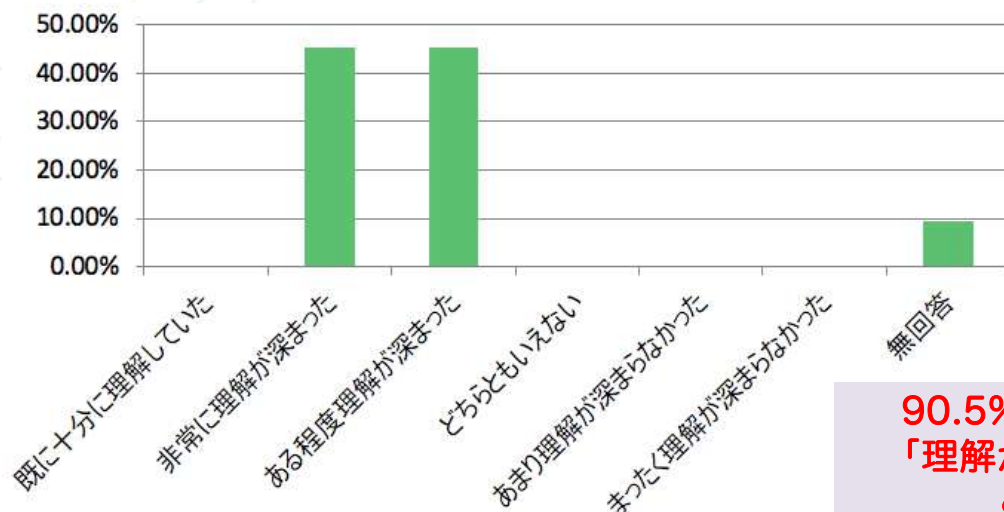


大阪大学
OSAKA UNIVERSITY

OPEN 2021

[回答数=42/58 (回答率:72.4%)]

質問：本会に参加し「MGeND」について理解が深まりましたか？



90.5%(38/42)が
「理解が深まった」
と回答

MGeNDホームページから [https://mgend.med.kyoto-u.ac.jp]



“Submission”
→JPメニュー(日本語)



データ登録について

- 登録可能なデータ
MGeNDでは疾患バリエーションデータを広く収集・登録することを目指しております。しかし、非制限公開のデータベースであることから、AMED「臨床ゲノム情報統合データベース整備事業」にて作成したデータ登録ポリシーを満たし且つMGeND登録委員会（以下、登録委員会）にて承認されたデータのみ登録を受け付けています。まず、お持ちのデータが登録可能であるか以下のデータ登録ガイドライン（簡易版）および登録申請書をご確認ください。ご不明な点は、MGeND登録管理委員会（dtjimu@hosp.ncgm.go.jp）にご連絡ください。
 - データ登録ガイドライン(簡易版)(*1)
 - データ登録申請書(*1)より詳細な情報については下記をご覧ください。
 - データ登録ガイドライン(詳細版)(*1)
 - データ登録ガイドライン整理表(*1)(*1) AMED臨床ゲノム情報統合データベース整備事業の「臨床ゲノム情報データベース構築」で作成
- データ登録手順
上記ガイドラインに基づき、登録可能なデータである場合、登録者は登録委員会にデータ提供を申請します。MGeNDは登録委員会からの承認連絡を受け登録者に登録手順を連絡します。おおまかな流れは下記の通りです。
 - MGeND登録管理委員会へデータ登録申請書を送付: dtjimu@hosp.ncgm.go.jp
 - MGeND登録管理委員会により確認を行います（倫理審査情報などから登録可能かのみを確認）
※必要に応じて登録管理委員会より倫理関連書類の提出を求める場合があります
 - MGeND登録管理委員会によるデータ登録承認の連絡（1から3は初回のみですが、研究計画単位で新たに申請が必要です）
 - MGeNDより登録手順について連絡、以下のデータ登録フォーマットに基づきデータをご準備いただき、登録

* 難病・稀少疾患領域の皆様：MGeNDのご案内と臨床ゲノム情報統合データベース整備事業へのご協力をお願い

データ登録ガイドライン(簡易版)
データ登録申請書
データ登録ガイドライン(詳細版) / 整理表

難病班への紹介資料

ご清聴ありがとうございました

問合せ：大阪大学大学院医学系研究科医の倫理と公共政策学
email: lecture@eth.med.osaka-u.ac.jp

参考

データが個人情報に該当するかどうかの判断（個人情報保護法）

「個人を識別する要素」and/or「照合性」が存在すれば、個人情報に該当。



データ自体に個人を識別する要素が含まれるかどうか	照合用のデータを施設が保有するかどうか(照合性)	個人情報該当性
含まれる (氏名、住所、顔画像などの個人を識別する要素や、個人識別符号が、判断しようとするデータに入っている。)	—	該当する※1
含まれない (氏名等の個人を識別する要素や個人識別符号が除かれている。なお、氏名等を数字・符号等によるIDに置き換えたものが含まれるのは構わない。)	存在する (氏名等の個人を識別する要素が含まれる元データや、診療録、IDの対応表等が、判断しようとするデータをもつ施設に存在する。)	該当する※2 照合データが自施設にあるか、他施設にあるかに関わらず、入手可能であれば、その施設において個人情報に該当すると考える。
	存在しない	該当しない※3

A類型・提供元側
A類型・DB側

B類型・提供元側とDB側

本データベースで収集するバリエーションデータは、データに個人を識別する要素が含まれないため、※2または※3になる。(ただし、日本に数人といった患者数が非常に少ない超希少疾患などで、疾患名が氏名と同じように識別的な働きをすることによって個人が識別されることが考えられる場合には、※1に分類されることもある。)

参考

個人情報保護法とゲノム指針における個人情報の分類

個人情報保護法			ゲノム指針
データ自体に個人を識別する要素が含まれるかどうか	照合用の他のデータを施設が保有するかどうか	個人情報に該当するか	個人情報等の分類
含まれる (氏名、住所、顔画像などの個人を識別する要素や、個人識別符号が、判断しようとするデータに入っている。)	—	該当する※1	個人情報
含まれない (氏名等の個人を識別する要素や個人識別符号が除かれている。なお、氏名等を数字・符号等によるIDに置き換えたものが含まれるのは構わない。)	存在する (氏名等の個人を識別する要素が含まれる元データや、診療録、IDの対応表等が判断しようとするデータをもつ施設に存在する。)	該当する※2 照合データが自施設にあるか、他施設にあるかに関わらず、入手可能であれば、その施設において個人情報に該当すると考える。	照合による個人識別をする場合(または、個人識別のための照合が容易な場合):「匿名化されているもの(特定の個人を識別することができないものに限る。)」(＝個人情報) 照合による個人識別をしない場合であって、照合データが適切に管理されている場合:「匿名化されているもの(どの研究対象者の試料・情報であるかが直ちに判別できないよう、加工又は管理されたものに限る。)」(＝個人情報)
	存在しない	該当しない※3	・非個人情報(統計情報など) ・「匿名化されているもの(特定の個人を識別することができないものであって対応表が作成されていないものに限る。)」(＝個人情報)

情報利用にかかる指針の規定

(下に向かって緩やかになる)

同意

同意

オプトアウト

通知・公開

A類型・DB側

B類型・提供元側とDB側

※2の個人情報は、指針において2つに分けられ、照合性が適切に管理されている場合に規制が緩められてい

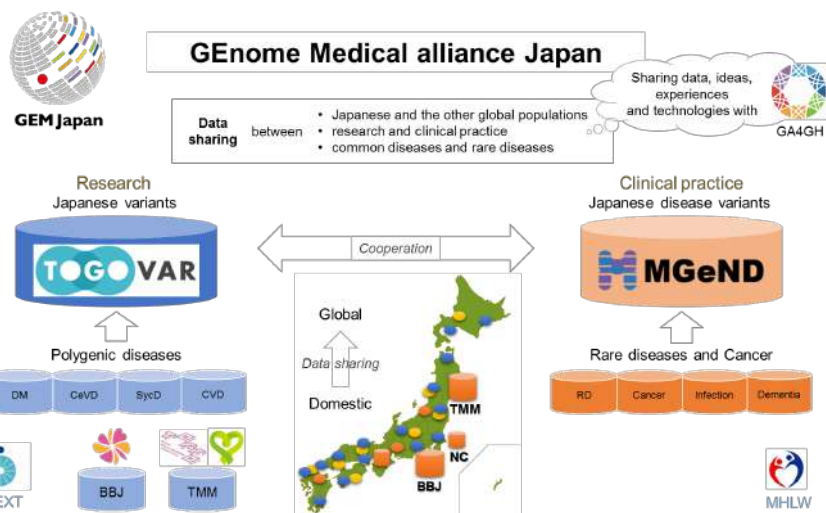
難病研究におけるヒトバリエントデータの利活用

慶應義塾大学医学部臨床遺伝学センター
小崎健次郎

医療現場において
バリエントが遺伝子機能に与える影響
の判断を助けるために

- 遺伝子機能に影響を与えるバリエント
= 病的変異との合致を検索
⇒ 病的変異のデータベースの必要性・有用性
- 遺伝子機能に影響を与えないバリエント
= 正常多型との合致を検索
⇒ 正常多型のデータベースの必要性・有用性

非制限公開のバリエントデータベース



遺伝性疾患患者のバリエントデータのフィルタリングに日々役立っている

3

特定個人のもつバリエントの解釈

	希少疾患研究	がん研究
原因遺伝子数	単数	複数
起因するバリエント数	1個(優性遺伝病の場合) 2個(劣性遺伝病の場合)	数個から数十個
データベースの使い方	病的意義が確定したバリエントの検索	病的意義が不明確なバリエントも含めての プロファイリング (AIの必要性)
民族差	疾患によって民族差あり	民族差少ない

4

病的変異のデータベースの例

	データ源	運営者・課金	キュレーター	提供形態
HGMD (英国)	発表されている文献・症例報告等	有料	専任の人類遺伝の専門家(30名程度)	ウェブ・電子ファイル
ClinVar (米国)	連結不応・検査室と連結	NIH 無料	検査会社・アカデミアの有志	ウェブ・電子ファイル
疾患毎の病的変異のデータベース群	疾患研究者等	疾患研究者等 無料	疾患研究者	ウェブ http://www.lovd.nl

5

病的バリエーションの記載法のピットホール

◆一例として、1991年発表の論文

論文中からの抜粋

we first elucidated the genomic organization of the human **arylsulfatase A (ASA) gene** and then compared the nucleotide sequences of exons and splice junctions of the mutant ASA gene to those of a normal control. We have identified **a new mutation, a G-to-A transition in exon 2, which results in amino acid substitution of Asp for 99Gly.**

ASA:p.Gly99Asp

	1991年	2019年
遺伝子名	ASA	ARSA
Reference Sequence	過去の論文で決定された配列データを参照 Stein et al.(1989)	hg19(GRCh37) or hg38(GRCh38)
バリエーション表記	ASA:c.296G>A, p.G99D	NM_000487.5:c.302.G>A, p.G101D NM_001085428.2:c.44G>A,p.G15D

6

キュレーターの役割

- 論文中に記載のあるバリエーションの座標確認
 - 現在のHGVS形式に準拠しない記載(主に古い論文)の修正
 - レファレンスの確認(GRCh37, GRCh38, NM番号)
- バリエーションの病的意義の確認
 - ACMGガイドラインに準拠した確認
 - InterVarのマニュアル入力部分の補完
- メタデータの整理
 - 疾患名、エビデンスの整理



キュレーションマニュアルを作成済み

2018年8月よりキュレーターを増員

7

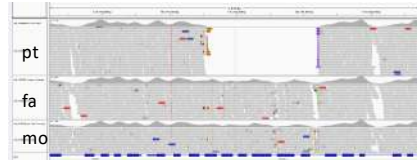
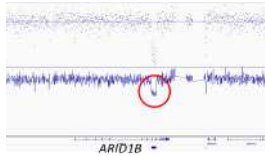
病的バリエーションデータベース—今後の課題

- 構造異常バリエーションの集積
- スプライシング異常バリエーションの集積

8

構造異常バリエント集積の取り組み

- 網羅的遺伝子解析の手法がエクソームから全ゲノムシーケンス（WGS）へ移行しつつあり、多くの病的なStructural Variantの検出が見込まれる。



- 特にArray CGHが保険収載されることから
 良性バリエントを異常と判断する（偽陽性）リスクを回避したい
 →人口ベースで良性バリエントを含めて数多く収集する必要。

9

構造異常バリエント収集の取り組み

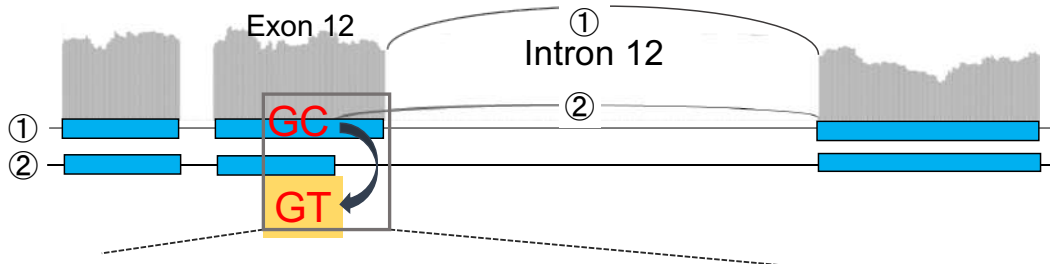
- ショートリードWGSでは、相当の偽陽性が含まれている。
- ロングリードWGSは高価である。
- 切断点を塩基のレベルまで正確に決めることができない。

ショートリードWGSで、

- ①頻度の高い構造異常はおそらくアーチファクト
- ②アーチファクトの混入を承知でデータを集積・公開する（レポジトリーの考え方）
- ③複数のソフトウェアで検出し、再現性を確認
- ④真のアレル頻度は、常染色体劣性遺伝病の原因となりうるかどうかの判断に重要

10

SAVNetを用いて検出された異常スプライシングの例
SNX14病的変異を検出(ヘテロ接合保因者として)



① 正常のスプライシングパターン

ゲノム GAA/GGC/GCA/GTG/CAC/GTG/TTG/CAG/TTT/TGT/TTG/ACT/GTG/Ggtgaggctta
タンパク Glu/Gly/Ala/Val/His/Val/Leu/Gln/Phe/Cys/Leu/Thr/Val

② 異常スプライシングパターン

新たに創出されたドナー部位
ゲノム GAA/Ggtgcagtgacagtggtgcagttttgttgactgtgggtgaggctta
タンパク (Glu)

→ 12アミノ酸欠失: p.Gly314_Val325del
常劣・脊髄小脳変性症原因遺伝子SNX14の保因者

11

RNA-Seqを解析に統合させることで
変異の病的意義が変化したバリエーション

遺伝子	スプライシング異常の種類	タンパクレベルの影響の予測	
		エクソーム	エクソーム解析 + RNA-seq
エクソームのみでは「サイレント変異」と予測された変異			
SNX14	Alternative5'SS	p.Gly358Gly	→ p.Gly314_Val325del
NDUFV1	Exon skipping	p.Ser360Ser	→ p.Gly305AspfsTer12
エクソームのみでは「ミスセンス変異」と予測された変異			
ACADVL	Alternative5'SS	p.Ile413Val	→ p.Glu412ArgfsTer9
ERLIN2	Alternative5'SS	p.Ser188Gly	→ p.Glu187Ter
MUTYH	Alternative5'SS	p.Ala470Thr	→ p.Ala470PhefsTer38
NIN	Intronic alternative3'SS	p.Ser1124Thr	→ p.Ile866HisfsTer3
TCTN3	Alternative5'SS	p.His480Tyr	→ p.His480LeufsTer2

12

病的バリエーション提供源

日本人に関する論文

IRUD等の
公的研究プロジェクトからの
バリエーション

臨床検査会社からの報告書
(含・保険診療)

古い論文:
→キュレーターによる登録
今後の論文:
→申告?

一括登録

主治医を介して
患者の承諾を得て、
登録?

13

ClinVarのデータ提供者の 上位6位までは臨床検査会社である

	Submitter	Maximum review status	Total submissions	Submissions with interpretations	Total Genes	Last updated
1.	Invitae	Assertion criteria	392983	392980	12726	Sep 22, 2020
2.	Illumina Clinical Services Laboratory; Illumina	Assertion criteria	208171	208171	2299	Jul 16, 2020
3.	GeneDx	Assertion criteria	121899	121765	26675	Sep 23, 2020
4.	Ambry Genetics	Assertion criteria	70620	70620	1339	Jul 28, 2020
5.	EGL Genetic Diagnostics; Eurofins Clinical Diagnostics	Assertion criteria	45028	45028	2406	Sep 19, 2018
6.	Color	Assertion criteria	43495	43495	84	May 19, 2020
	OMIM; Johns Hopkins University	-	31721	31721	5190	Oct 01, 2020
	Laboratory for Molecular Medicine; Partners	Assertion criteria	24235	24161	1748	Jun 02, 2020

米国では検査会社からClinVarへ直接送付
わが国での収集の方法について倫理審議委員会が検討開始

14

ClinVarの倫理申請書の概要

被験者の募集 本プロジェクトでは、**被験者の積極的な募集を行わない**。本プロジェクトに参加している検査施設の分子学的検査で紹介された**全ての**患者の変異データを、匿名化してClinVarに提出する。

研究期間 本研究はデータリポジトリ研究であるため、被験者の参加期間(すなわち、データがリポジトリ内に保存される期間)は**無期限**である。

15

●変異情報(臨床的解釈を含む)および表現型の基礎情報(検査時の年齢、性別、ならびに検査時に判明し、標準的な用語を用いて表現された表現型)のみから成るデータは、個人の識別は不可能とみなされ、したがって完全な同意取得またはオプトアウトプロセスを介したデータ提出は必要とならない。

●このタイプのデータには、単独遺伝子検査で特定された個別変異、単独遺伝子検査で特定された複合型ヘテロ接合体変異、複数遺伝子パネル検査、または全エクソーム/全ゲノムシーケンシングが含まれる。

16

インフォームドコンセントの要素の免除は、次の理由に基づき、本研究に必要とされている。

1. 本研究に伴うリスクは極めて低い。データベースに提出された情報によって個人が特定されるリスクは低い。本研究への参加に関連して予想されるリスクは、他に存在しない。
2. 完全なインフォームドコンセントの免除は、被験者の権利または福祉に有害な影響を及ぼすとは考えられない。
3. ClinVarに変異レベルの情報の要約を提出するという検査施設の意図に関する情報は、検査申込書、検査結果、および検査施設のウェブサイト上の記述を介して患者に公開される。

17

INFORMED CONSENT



Account #	Account Name	
First Name	Last Name	Date of Birth

General Information About Genetic Testing

What is genetic testing?

DNA provides instructions for our body's growth and development. Genes are distinct sequences of DNA, and are arranged on chromosomes. The DNA in a gene contains instructions for making proteins, which determine things like growth and metabolism as well as traits like eye color and blood type. Genetic disorders are caused by certain changes in DNA affecting the structure or number of chromosomes.

- This test does not have the ability to detect all of the long-term medical risks that I/my child might experience. The result of this test does not guarantee my health or the health of my child/fetus. Other diagnostic tests may still need to be done, especially when only a genetic screening test has been performed previously.
- Occasionally, an additional sample may be needed if the initial specimen is not adequate.

Database Participation

De-identified health history and genetic information can help health care providers and scientists understand how genes affect human health. Though {I/my child} may not personally benefit, sharing this information helps health care providers to provide better care for their patients and researchers to make discoveries. GeneDx shares this type of information with health care providers, scientists and health care databases. No personal identifying information will be shared, as it will be replaced with a unique code.

Even though only a code is used for the reporting to the database, there is a risk that {I/my child} could be identified based on the genetic and health information that is shared. GeneDx believes that this is unlikely, though the risk is greater if I have already shared {my/my child's} genetic or health information with public resources, such as genealogy websites.

18

- 提出元検査施設は、提出する全データから、全てのHIPPAに基づく識別子を除去する。続いて、提出元検査施設は、このデータに対して、いかなる保護医療情報(PHI)(検査施設の受付番号など)にも由来しない一意のコードを割り付ける。このコードは提出元検査施設以外には開示されない。
- 次に、データはNCBIに提出され、ここで提出元検査施設の一意のコードが除去され、別の一意のNCBI識別子が割り付けられる。NCBIは提出元検査施設に、検査施設が割り付けた一意のコードとNCBI識別子のリンクを提供する。NCBIがPHIにアクセスすることは決してなく、PHIが公的データベースに掲載されることは決してない。
- 提出元検査施設は、特定の症例をPHIと結びつけることができる唯一の機関である。提出元検査施設は、どの症例をデータベースに既に提出したかを確認するため。

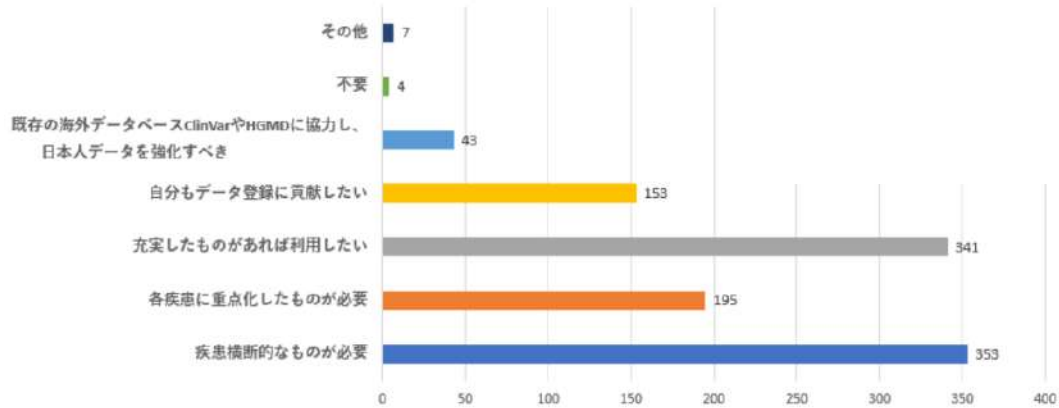
19

キュレーションされた臨床グレードのゲノム変異データのリポジトリの構築を目的とする本プロジェクトでは、Geisinger Health SystemがPartners HealthCare Personalized Medicineと共に調整施設として機能する。本プロジェクトは**両施設のIRBの承認**を受けている (Geisinger IRB# 2013-0367、Partners IRB #2013P001099)。

データはNCBIのClinVarデータベースに保存され、GeisingerまたはPartners HealthCareには保存されない。

20

人類遺伝学会会員へのアンケート調査 日本人固有の病的バリエーションデータベースに関する質問



21

まとめ：難病バリエーションの収集のための持続可能なシステムの構築にむけて

- データベースの「量」の向上
 - 小児科学会等の学会連携を通じたデータ提供体制の構築
 - 他の国内研究事業（難病プラットフォーム、IRUD等）との連携構築
 - 臨床検査として行われる遺伝学的検査結果の収集
- データベースの「質」の向上
 - RNA-Seqによるスプライス変異バリエーションの収集
 - 構造異常バリエーションの収集

22

謝辞

臨床ゲノム情報統合データベース整備事業

「真に個別患者の診療に役立ち領域横断的に高い拡張性を有する変異・多型情報データベースの創成」研究分担者

臨床ゲノム情報統合データベース整備事業

「希少・難治性疾患領域における臨床ゲノムデータストレージの整備に関する研究」研究分担者

難治性疾患実用化研究事業

「未診断疾患イニシアチブ：希少未診断疾患に対する診断プログラムの開発に関する研究」全国のメンバー

日本小児科学会 学術委員会 病的バリエーションデータベース小委員会メンバー