

稀少疾患患者のゲノム診断における コントロール群バリエーションデータの価値と 各種コホート由来データの統合利用の意義

慶應義塾大学医学部臨床遺伝学センター教授

日本人類遺伝学会理事長

日米・臨床遺伝専門医

小崎健次郎

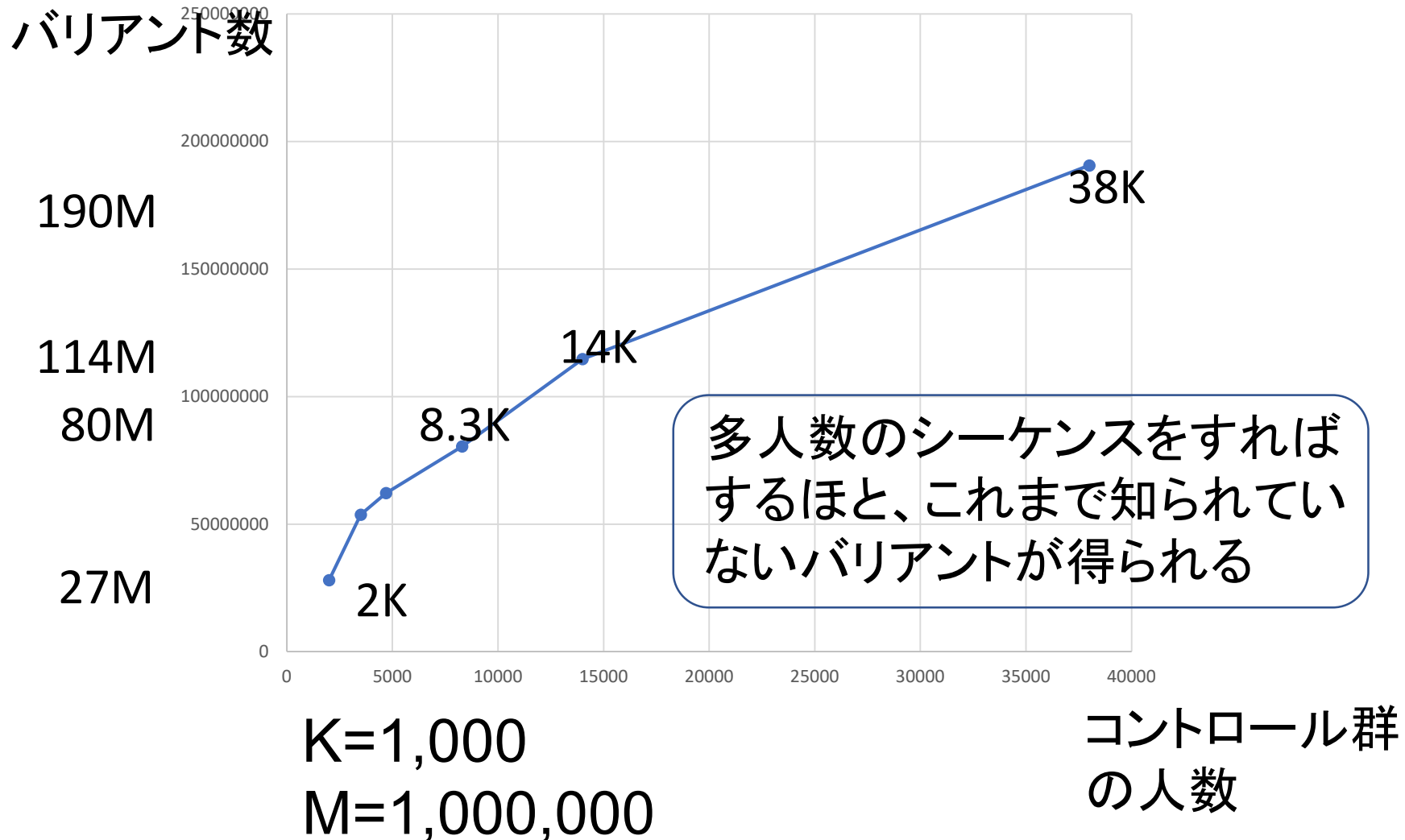
世界的に「変異」という用語は使わなくなった

- ★ 1 個の遺伝子配列の中で、個人により差違がある場所（バリエーション）は多数箇所ある。


ATGCTTATTC.....GTATCGCA
ATGCTTGTTTC.....GTACCGCA
A>G
T>C

- 「変異」という用語は機能異常を内包してしまう。
- 「バリエーション」という用語は機能異常を内包せず、中立的である。
- 疾患原因と証明された場合には病的バリエーションと呼ぶ。
- 「標準配列（参照配列）」は「正常配列」ではなく、仮に使っている基準（物差し）である。
- 基準からの相異の「許容範囲」を示しているのが「コントロール群」のバリエーション群である（データベース化されている）。

シーケンスしたコントロール群の人数と相関して 多数のバリエーション数が得られる (東北メディカルメガバンクの要約データから)

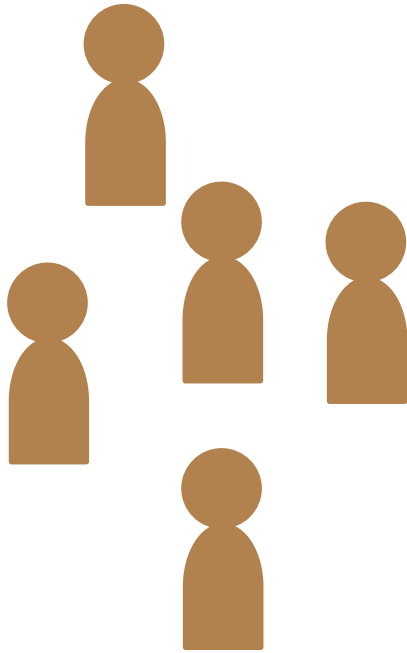


対象とする遺伝性疾患と 標的とするバリエーションの性質の比較

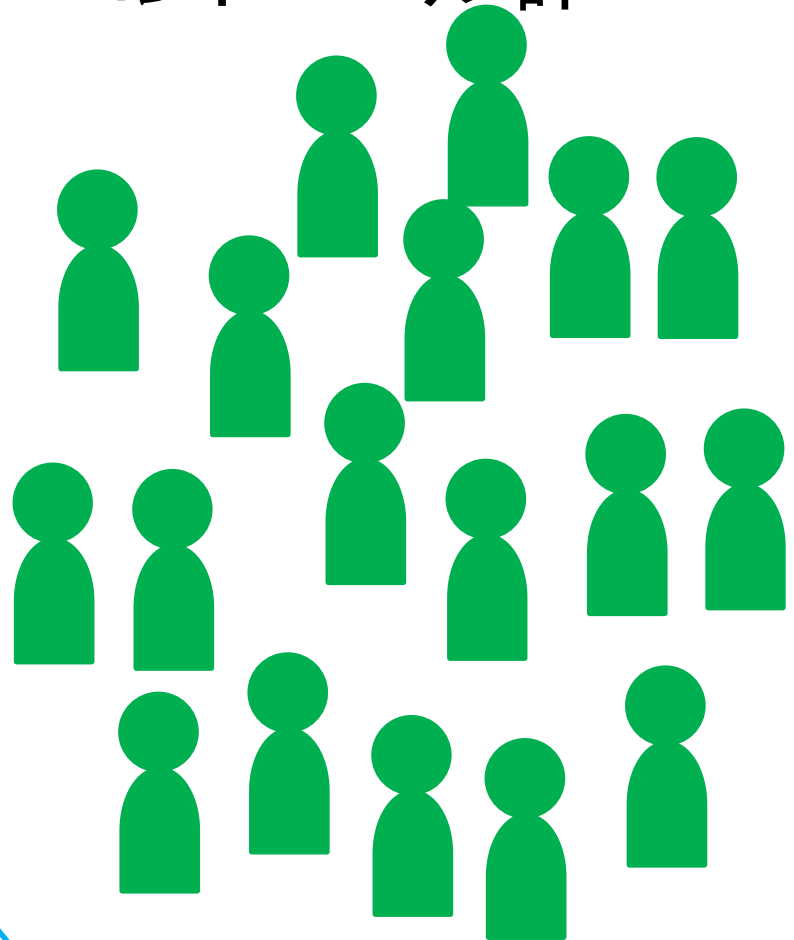
	単一遺伝子病の診断・研究		多因子遺伝病の研究 (罹患しやすさの研究)
	顕性遺伝病 (優性遺伝病)	潜性遺伝病 (劣性遺伝病)	
患者群	単一遺伝子病の可能性のある患者 		特定の多因子遺伝病 (糖尿病等)が疑われる 多数の患者コホート
病因バリエーション	病因バリエーションは患者のみでしか同定されず、コントロール群では同定されない。 ⇒コントロール群に認められるバリエーションは病因ではない。	患者がヘテロの時もあるが稀である。(アレル頻度1%未満=コントロール群50人に一人未満)	コントロール群に比較して、患者群における頻度が高いが、コントロール群にも相当数認められる
	日本人におけるアレル頻度情報が重要である。		

単一遺伝子病研究

患者群



コントロール群



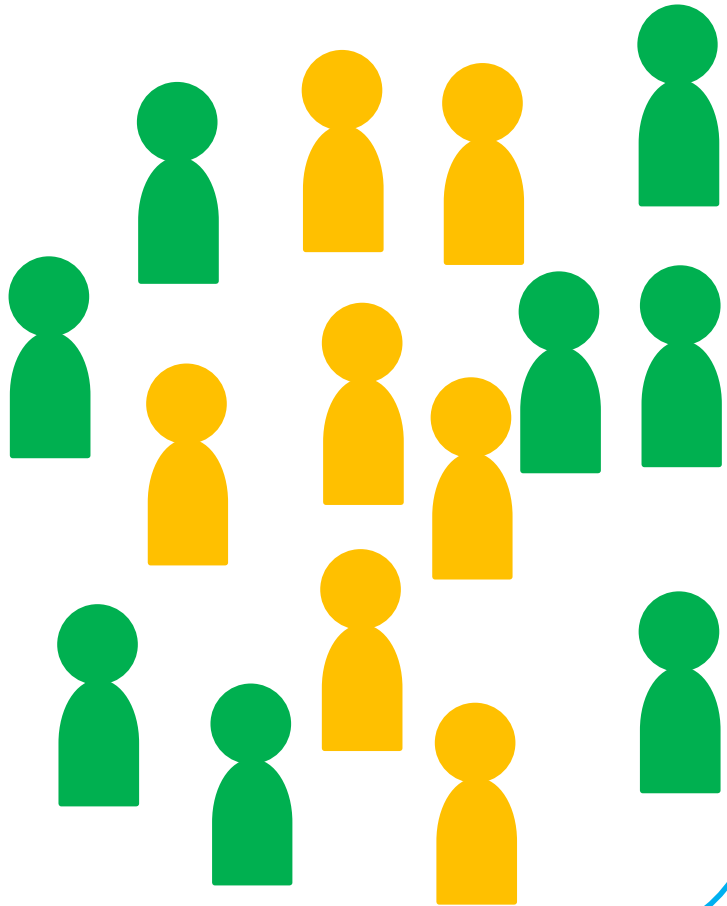
=特定の遺伝子バリエーションのある人



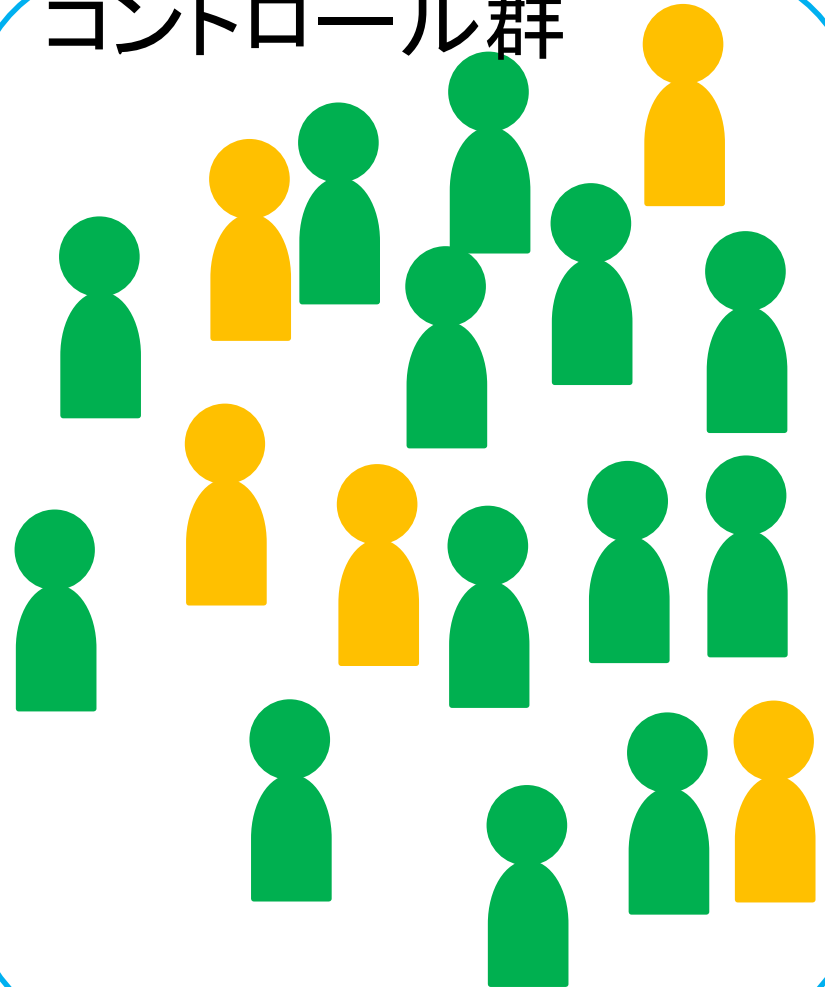
=特定の遺伝子バリエーションのない人


多因子遺伝病研究


患者群



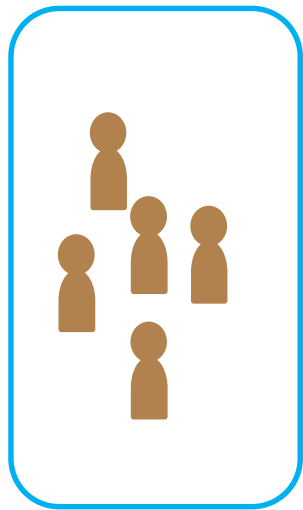
コントロール群



 = 特定の遺伝子バリエーションのある人

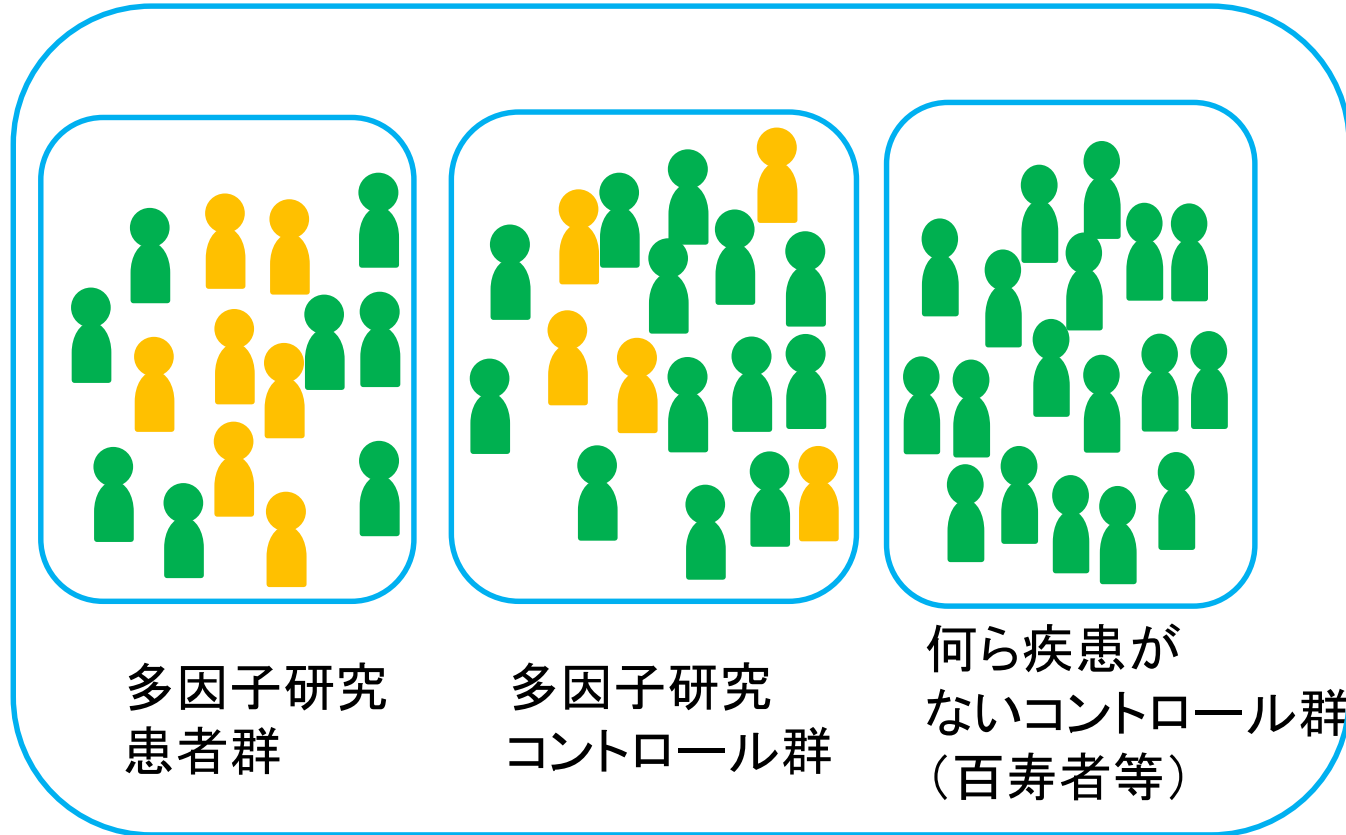
 = 特定の遺伝子バリエーションのない人

様々なコホートデータの 難病研究におけるコントロール群としての使用可能性



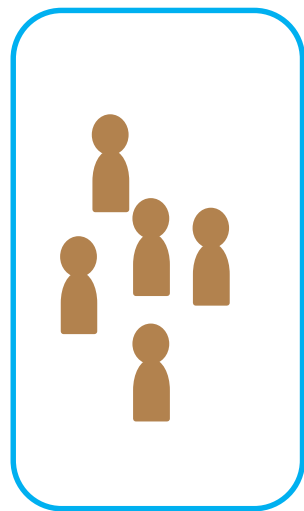
難病研究
の患者群

VS

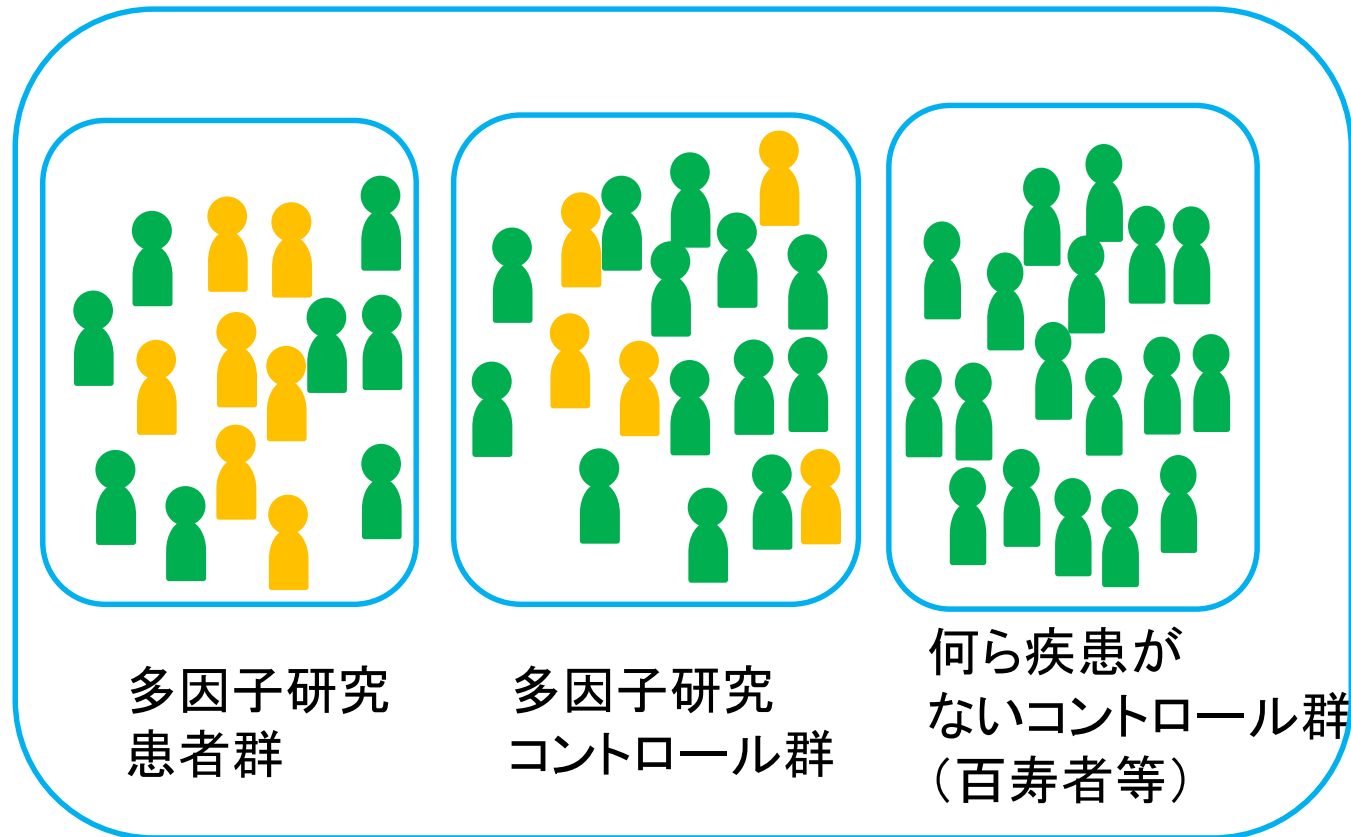


全部まとめて難病研究のコントロール群として利用可能

コントロール群サンプルの選択における 臨床情報の有用性



VS



難病研究
の患者群

全部まとめて難病研究のコントロール群として利用可能
⇒ 個別のサンプルについてデジタル化された臨床情報が
あれば(たとえシンプルでも)、適切なコントロール用
サンプル・データの抽出に使用できる

コントロール群の選定・設定方法はゲノム解析の目的により異なる

	単一遺伝子病の 診断・研究	多因子遺伝病の 研究
患者群	単一遺伝子病の可能性 がある患者	特定の多因子遺伝病(糖 尿病等)が疑われる多数 の患者コホート
コントロール群	遺伝性疾患を疑わせる 症状のない人 (多因子遺伝病の患者 群もコントロール群と見 なしている場合がある)	できるだけ広い意味で 健康な者が望ましい (東北メディカルメガバン クなど、健康な集団)

- 各コホート内の個人の臨床的な特徴がデータベース化されていれば、分析対象とする患者群の特性に応じて、コントロール群として使用可能な患者データを抽出して活用することが可能である。
- 臨床データ情報があれば、複数のコホートを組み合わせて一つのコントロール群として活用することが可能である。

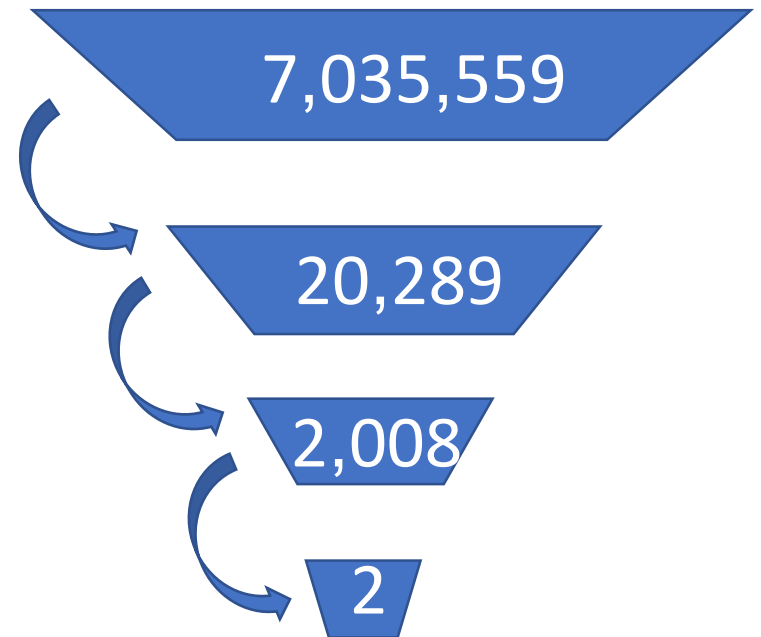
コントロール集団におけるアレル頻度情報は 稀少疾患のゲノム診断に貢献する (ToMMo38Kの利用例)

患者サンプル全ゲノム
疾患候補バリエントは700万以上

タンパクの構造に影響を与える
バリエントは約2万

ToMMo38Kアレル頻度 (<1%)
でフィルター

専門家が最終評価

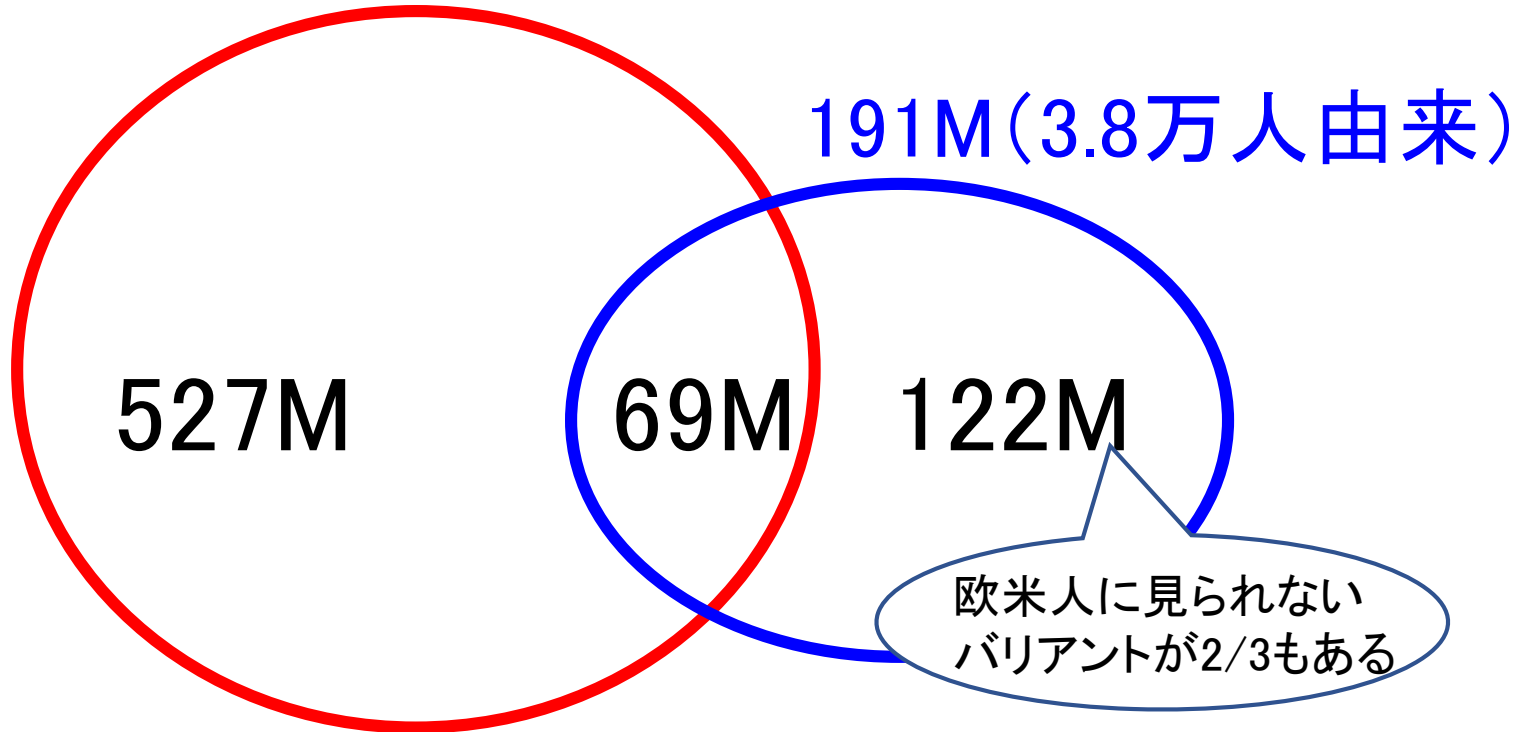


⇒コントロール群38Kは、稀少疾患の原因遺伝子の絞り込みにきわめて有効と期待される

日本人のゲノム研究・診断には 日本人のコントロール群バリエントデータが必要

596M (14万人由来)

M=1000,000バリエント



欧米人を中心とした
世界を代表する
バリエントデータベース (gnomAD)

日本人を中心とする
バリエントデータベース
(ToMMo)

コントロール群の選び方

- 効率的に除外するためにはコントロール集団におけるバリエーションの頻度(専門的にはアレル頻度と呼ぶ)情報が有用かつ不可欠である。
- 日本人におけるバリエーション頻度と外国人におけるバリエーション頻度は大きく異なり、日本人においてのみ認められるバリエーションも多数存在する。
- 現在の数万人規模のコントロール群から得られるデータを見る限り、まだまだ未知のバリエーションがコントロール群に含まれている。
- 多因子遺伝研究のコホートのバリエーションデータは、稀少疾患の研究におけるコントロール群となりうる。
- 既存の各種コホートのバリエーションを統合集積し、頻度を算出する必要がある。

構造異常の課題

- 点変異 (SNV) 以外の構造異常 (SV) の取り扱いが必要である。
- SNVと異なり、SVの検出のための標準的な手法は国際的にも未確立である。
- SVの抽出のためには、個別サンプルのFastqないしBAMファイルが必要である。
- SV抽出はBAMファイルから始めた方が、Fastqから始めるより計算量が少なくて済む。
- 各種コホート由来の生データをジョイントしてSVコールできるとよい。

結語

- 単一遺伝子病の研究・診断では、発症に無関係なバリエントを除外する必要がある。
- 効率的に除外するためにはコントロール集団におけるバリエントの頻度（専門的にはアレル頻度と呼ぶ）情報が有用かつ不可欠である。
- 日本人におけるバリエント頻度と外国人におけるバリエント頻度は大きく異なり、日本人においてのみ認められるバリエントも多数存在する。
- 多因子遺伝研究のコホートのバリエントデータは、稀少疾患の研究におけるコントロール群となりうる。
- 個別検体の臨床情報があれば、適切なコントロールサンプルを抽出可能である。
- 既存の各種コホートのバリエントを統合集積し、頻度を算出する必要がある。