

# SNP streakモデルによる ノンパラメトリック連鎖解析ツール Obeliscの開発

曾根原 究人、岡田 随象  
(大阪大学 大学院医学系研究科 遺伝統計学)

日本人類遺伝学会第65回大会

The 65th Annual Meeting of the Japan Society of Human Genetics

Co-production of genomic medicine  
by researchers, patients and public

# 発表者のCOI開示

演題発表に関連し、発表者らに開示すべき  
COI関係にある企業などはありません

日本人類遺伝学会第65回大会

The 65th Annual Meeting of the Japan Society of Human Genetics

Co-production of genomic medicine  
by researchers, patients and public

# Overview

## Background

連鎖解析 (linkage analysis) は疾患感受性遺伝子の同定において優れた成果を上げてきた。一方で、既存の連鎖解析手法には適用できる場面に制限があり、罹患者のゲノムデータが入手できても効果的な疾患感受性遺伝子マッピングができない場面が存在する。このような場面においても柔軟に適用できる連鎖解析ツールの実装が必要とされていた。

## Methods

我々は、罹患者ゲノム間におけるハプロタイプの共有に着目するSNP streakを応用することで、幅広い場面に適用できるノンパラメトリック連鎖解析ツールObeliscを開発した。

## Results

開発したツールの性能を検証した。Obeliscにより検出された同祖ハプロタイプ領域 (identical-by-descent: IBD領域) は、他ツールによる大規模ゲノムデータを利用したハプロタイプ推定においても一致した結果が得られることを確認した。

## Conclusion

Obeliscによるノンパラメトリック連鎖解析は、既存手法の適用が困難であったデータセットの解析を可能とし、未知の疾患メカニズムの解明に貢献することが期待される。



(Sonehara K and Okada Y. *Bioinformatics* 2020)

# Background

**連鎖解析 (linkage analysis)** : 染色体上での疾患感受性遺伝子とマーカー遺伝子座の連鎖関係を利用することで、疾患感受性遺伝子の染色体上での位置を同定する遺伝統計学手法

手法的に2種類に大別できるが、それぞれ**適用できる場面に制限がある**

## パラメトリック連鎖解析

遺伝形式・浸透率・アレル頻度といったパラメータを設定した遺伝モデルに基づき尤度計算を行うことで、マーカー遺伝子座と疾患感受性遺伝子がどれだけ近傍に位置するかを推定する

**制限** : ①家系構造が明確にわかっている必要がある  
②家系内で交配があると、計算コストが膨大となる  
③解析結果は解析者によって設定されたモデルのパラメータに依存したものとなる

## ノンパラメトリック連鎖解析

事前に遺伝モデルを定義せず、罹患者群中でのハプロタイプの共有を検出することで、疾患感受性遺伝子領域を絞り込む

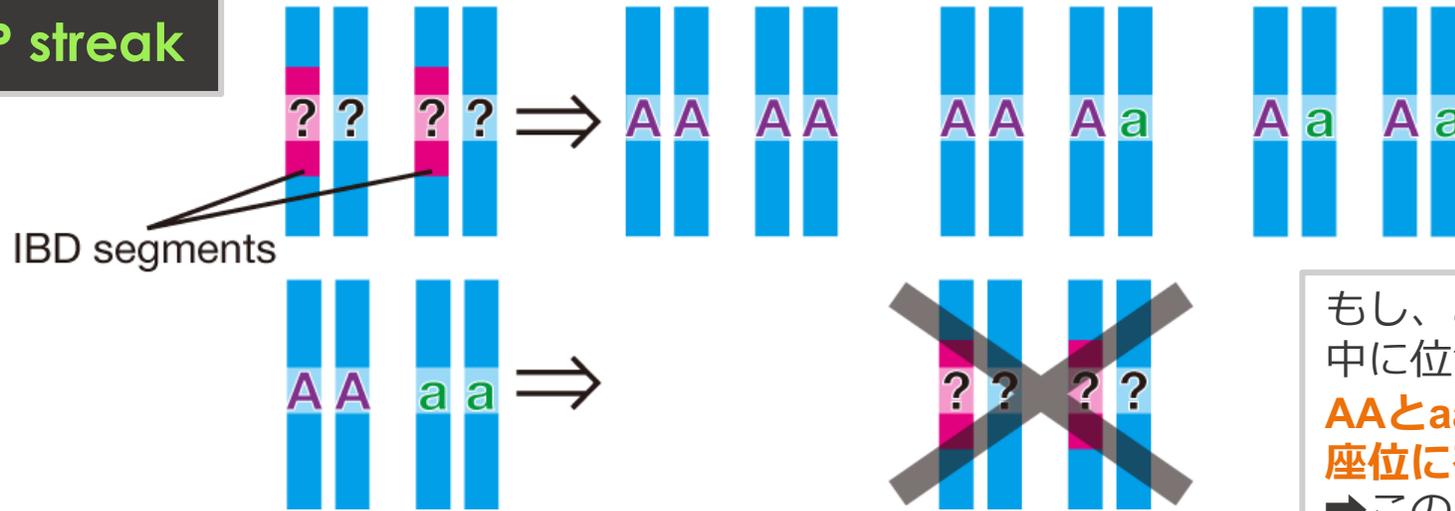
**制限** : 代表的な手法としてホモ接合連続領域 (Runs of homozygosity) の検出が挙げられるが、これは**潜性遺伝 (劣性遺伝) 形式を呈する形質にしか適用できない**

 **遺伝形式を問わずに適用可能なノンパラメトリック連鎖解析手法の必要性**

# Methods

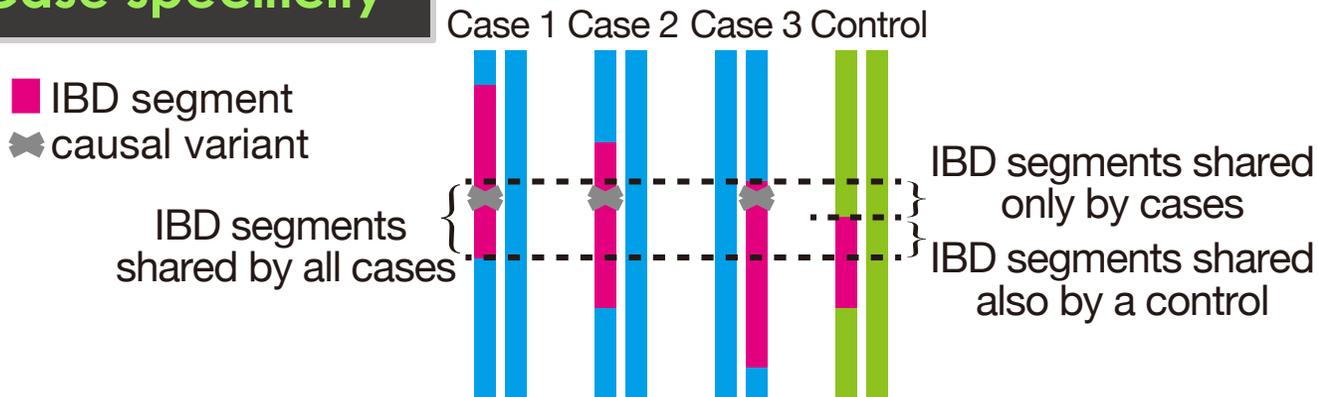
遺伝形式を問わずに適用できるノンパラメトリック連鎖解析手法として、SNP streakを活用した解析ツールObelisc (Observational linkage scan) を開発した。

## SNP streak



もし、ある座位が二者間で共有されているIBD領域中に位置するのであれば、**AAとaaのような相異なるアレルのホモ接合はその座位に存在し得ない**  
→このようなジェノタイプの組み合わせが十分長く連続する領域は、IBD領域であることが示唆される

## Case specificity



同様にしてケースとコントロールとのハプロタイプ共有を調べることで、**検出されたIBD領域のケース群特異性**を評価できる  
→ケース群特異的な領域は、病因となる遺伝子の位置の有力な候補となる

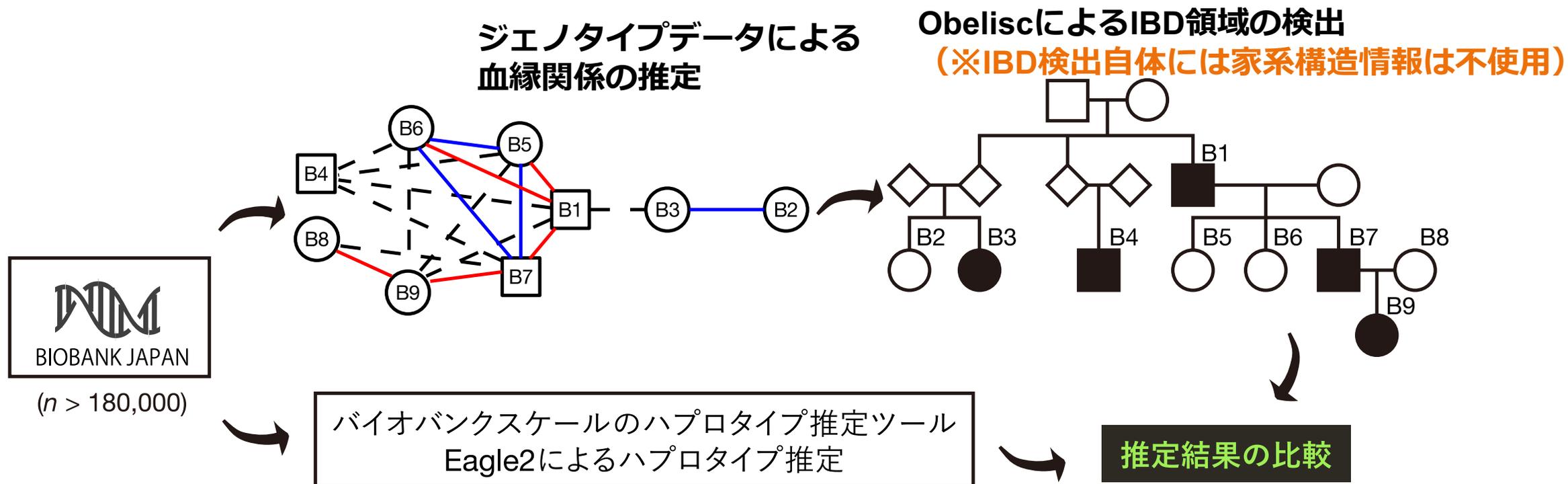
# Methods

## 検証用データセットの作成

公開ジェノタイプデータ（バイオバンク・ジャパン）をもとに、バイオバンクの参加者間の血縁関係をジェノタイプのみから推定し、罹患情報を割り当てることで、擬似家系データを作成した。

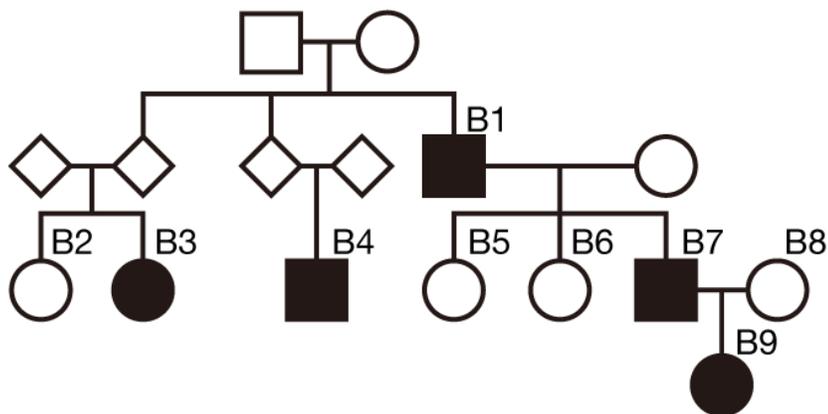
この擬似家系データにObeliscを適用し、家系メンバー間で共有されるIBD領域を検出した。

一方で、バイオバンクスケールデータのハプロタイプ推定ツールEagle2を用い、別アプローチによるハプロタイプ推定結果との一致を確認することで、検出したIBD領域の妥当性を検証した。

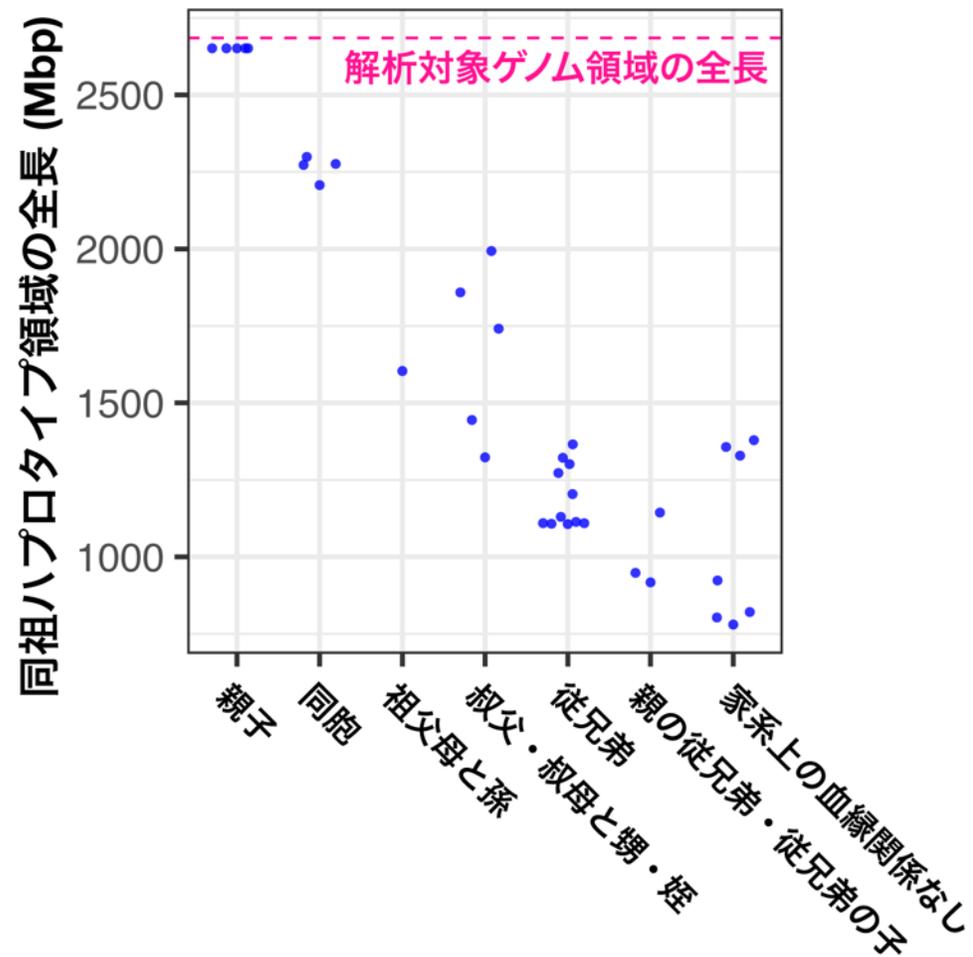


# Results

まず、開発したツールが理論通りにIBD領域を検出できるか確認した。  
Obelisc自体は家系構造の情報を解析に利用していないが、正しくIBD領域の推定ができていのであれば、  
擬似家系中の**血縁関係の近いペアは、より多くのIBD領域を共有する**と考えられる。  
特に、**親子では全領域がIBD領域となる**はずである。



- 実際に擬似家系中の全ペアについて、共有するIBD領域の全長を算出したところ、**血縁関係の近いペアほどより多くのIBD領域が検出される**ことが確かめられた。
- 理論的に予測される通り、**親子間では解析対象としたゲノム領域のほぼ全域が、共有されるIBD領域として検出された。**

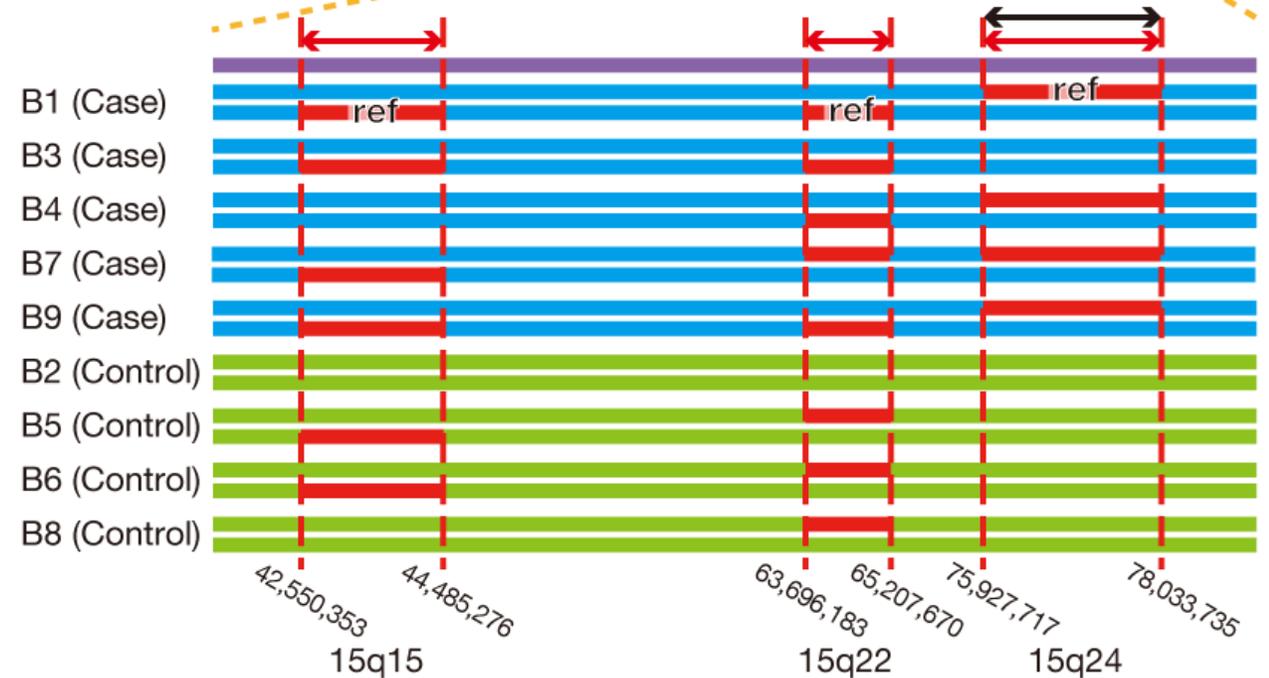
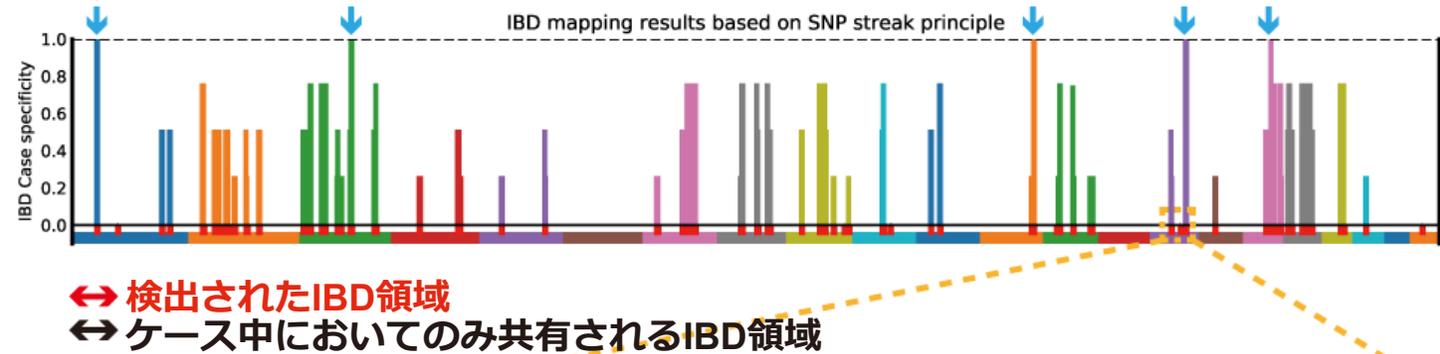


# Results

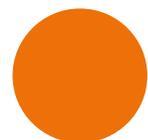
- Obeliscを用い、擬似家系中の罹患者ゲノムにおいて共有されるIBD領域の検出を行った。
- ➔ 罹患者ゲノム間において共有される65領域を検出し、そのうち5領域は非罹患者ゲノムにおいては共有されていなかった  
(罹患者ゲノム特異的に共有される領域)。

- 検出された共有領域のハプロタイプが本当に同一なのか、バイオバンク・ジャパンの全公開データ (サンプル数  $n > 180,000$ ) を用い、バイオバンクスケールでハプロタイプ推定を行った推定結果との一致を確認した。
- ➔ 例として、右図の15番染色体上の3つの検出領域において、罹患者5人中4人または5人全員が同一ハプロタイプを共有していることが確かめられた。  
また、ケース特異的として検出された領域は、実際にいずれのコントロールサンプルも同一ハプロタイプを共有していないことを確認した。

## ↓ ケース中においてのみ共有されるIBD領域



# Conclusion



SNP streakを用いたノンパラメトリック連鎖解析ツールObeliscを開発し、その性能をテストした (Sonehara K and Okada Y. *Bioinformatics* 2020)。

Obeliscは下記でソースコードと共に公開している：

<https://github.com/qsonehara/Obelisc>



Obeliscによって推定されたIBD領域は、バイオバンクスケールのデータを利用したハプロタイプ推定ツールにおける推定結果と一致を示した。

この結果は、整った家系情報の存在しない限られたサンプル数のデータのみによる、共有IBD領域の解析を実現することを意味する。



Obeliscの長所は、

①正確な家系構造が不明であっても疾患が同一ハプロタイプに起因することさえ仮定できれば適用することができる点

②顕性遺伝形式、潜性遺伝形式を問わず、適用することができる点  
に代表される柔軟な適用可能性である。

特に①については、特徴的な表現型を呈する希少疾患であれば、同一家系内に限定せずとも、共有する同祖ハプロタイプを検討することが可能である。