

平成 22 年 5 月 31 日現在

研究種目：若手研究(B)
 研究期間：2008～2009
 課題番号：20710147
 研究課題名（和文） 決定的ハプロタイプに基づく日本人ヒトゲノム構造多型の検出
 研究課題名（英文） Detection of Japanese Structural Variation Using Definitive Haplotype Data Obtained from Complete Hydatidiform Moles.
 研究代表者
 日笠 幸一郎 (Higasa Koichiro)
 独立行政法人理化学研究所・統計解析研究チーム・研究員
 研究者番号：10419583

研究成果の概要（和文）：

日本全国から収集した胞状奇胎と呼ばれる特殊な検体を用いて、日本人集団の確定ハプロタイプ構造を明らかにした。国際ハブマップ計画によって構築された推定ハプロタイプ構造との比較解析を行った結果、今回同定した確定ハプロタイプは、自然選択領域検出、日本人ヒトゲノム構造多型の検出に有用であるだけでなく、推定ハプロタイプの精度の向上に役立つことを示した。これらの結果は、日本人ゲノムの進化学的解析、構造多型と疾患との関連の解明において重要な知見である。

研究成果の概要（英文）：

We determined true Japanese haplotypes by genotyping 100 complete hydatidiform moles (CHM), each carrying a genome derived from a single sperm. We then assessed how inferred haplotypes can differ from true haplotypes, and found that, at various genomic regions, especially the MHC locus, the expansion of extended haplotype homozygosity (EHH), which is a measure of positive selection, is obscured when inferred Asian haplotype data is used to detect the expansion. We also show that the true haplotype data is useful not only to assess and improve the accuracy of phasing of Asian genotypes, but also to detect structural variation of Japanese genome.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,600,000	480,000	2,080,000
2009年度	1,700,000	510,000	2,210,000
総計	3,300,000	990,000	4,290,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学 ・基礎ゲノム科学

キーワード：ゲノム、進化、ハプロタイプ、構造多型

1. 研究開始当初の背景
 近年、ヒトゲノム配列は変異の少ない静的な

イメージから、ダイナミックな変異を持つ動的な
 イメージへ変化しつつある。その要因となってい

るものは、国際ハブマップ計画で解析された一塩基多型ではなく、ゲノム構造多型と称される、よりサイズの大きな変異であり、想像以上に我々個体間のゲノム配列は異なっている事が明らかとなってきた。この発見は、研究に用いる検体の人種特有のゲノム構造情報が、関連解析・連鎖解析等を用いた病因遺伝子同定や、将来のゲノム研究のために必要不可欠となってきた事を明確に意味する。国際ヒトゲノム計画によって解読されたヒトゲノム参照配列は、少数の西欧人由来であることから、先行研究として我々日本人のヒトゲノム参照配列を再構成する試みは意義深い。最近までにゲノム構造多型が見落とされていた理由は、その多型の大きさと特徴的配列類似性に、既存の検出機器が対応できていなかった点が挙げられる。

2. 研究の目的

本研究課題は、日本人ゲノム中に存在している未知のゲノム構造多型候補領域を検出し、日本人ヒトゲノム参照配列を再構成することによって、将来の遺伝学的解析の精度を向上させるための基盤情報構築を目的とする。国際ハブマップ計画によって解析されたゲノム多型は、全ゲノム配列塩基数の約 0.1% を占める極小さな変化（一塩基多型）に過ぎない。そこから“推定される”ハプロタイプ構造によって、遺伝学的解析の効率向上を図るのが目的である。しかし、ゲノム構造多型は、最低でもその約 4 倍のゲノム変化を担っていると見積もられており、推定的ハプロタイプを用いた関連解析の限界と、新規ゲノム構造多型同定の重要性を示唆している。国際ハブマップ計画によって構築されたハプロタイプ構造は、欧米人やアフリカ人においては家系情報を基に推定されているためにその構造は正確であるが、アジア人においては血縁関係のない検体を用いて推定されているため、そこからもたらされる不確実性が下流の解析にもたらす影響について精査する必要がある。本研究に用いる検体は、胞状奇胎と呼ばれる特殊な検体であり、この検体から得られる DNA 塩基配列情報は、国際ハブマップ計画で得られる結果よりも良質な情報である。また、適用する方法は、既存の検出技術限界に依存せず、正確なゲノム構造多型の検出と、迅速な病因遺伝子同定に繋がる新規の結果が期待されるものである。

3. 研究の方法

日本全国から収集した胞状奇胎と呼ばれる特殊な検体を用いて、日本人集団の“確定ハプロタイプ”構造を明らかにし、その有用性について様々な側面からの検討を行う。同集団のハプロタイプ構造は、国際ハブマップ計画によっても得られるが、親子関係にある検体を用いることなく統計的に決定した

“推定ハプロタイプ”であるために、それを自然選択領域の検出に用いた場合の誤検出が懸念される。そこで昨年度ゲノムワイドに行った約 28 万個の一塩基多型(SNP)のタイピングデータに加えて、更に 50 万個のタイピングを実施し、得られたより詳細な確定ハプロタイプ情報を基に、“確定”、“推定”ハプロタイプ間の比較を実施する。

・LD measure (D' , r^2) 歴史的に用いられてきた指標である、連鎖不平衡度 (D' , r^2) に基づき、ハプロタイプブロックサイズを調査する。この解析により、大まかな“確定”、“推定”ハプロタイプ間の違いについて、ゲノムワイドに全体像を把握する事が可能となる。

・iHS (integrated haplotype score) iHS 法は、調べたい領域における、あるハプロタイプを持つ染色体を二つランダムに選んだ場合に、その領域から距離 x までに含まれる全ての SNP がホモ接合体になる確率の積分値として導入されたものである。そのため、距離 x における iHS 値は、その間の「組み換え率」と「マーカー密度」(含まれる SNP 数) に大きく依存する。そこで、ゲノム全体に及ぶ「組み換え率」の変動と「マーカー密度」を補正するために、その相対値を計算する。この方法は同じ領域の対立アレルを内部コントロールとして補正に用いる。即ち、補正された iHS 値は、ある調べたいハプロタイプを持つアレルの iHS 値とそれ以外のハプロタイプを持つアレル群の iHS 値の比で定義される。この相対値の統計的有意性は実際の実験データから得られる分布を用いて検定する。具体的には、相対値を対数軸に変換し、各々の距離 x での P 値をその平均値と標準偏差を基に計算する。この補正により、詳細なハプロタイプ構造の比較と、下流の解析への影響が調査できる。

得られた結果の検証作業としては、確定ハプロタイプのペアリングによって擬似的なディプロイド個体を作成した後、推測プロセスを経由したハプロタイプ間との比較においても検討を行い、得られた結果が、胞状奇胎とハブマップ検体間のサンプリングバイアスではなく、推測するプロセスにおいて生じた誤りであることを確認する。また、外部情報を収集・解析し結果との比較・評価を行う。

更に検出感度の高い比較統計量の開発と適用を実施し、特に推定ハプロタイプでは検出しえない構造多型だけでなく、自然選択領域の検出も視野に入れ解析を行う。確定ハプロタイプの重要な利点を明らかにし、タイピングデータと共に解析結果をデータベース化し、将来のハプロタイプ解析における有用なリソースとして役立てるものとする。

4. 研究成果

日本全国から 100 検体の胞状奇胎サンプルを収集し、ゲノムワイドに約 50 万個の一塩基多型(SNP)のタイピングを行い、日本人の確定ハプロタイプを同定した。この結果得られた情報をもとに、以下の解析を行った。

(1) 国際ハブマップ計画でタイピングされた日

本人集団のジェノタイプ情報と今回得られた確定ハプロタイプ情報を基に主成分分析を行い、胞状奇胎から得られた確定ハプロタイプが、日本人集団の代表的ハプロタイプ構造とみなせる事を証明した。この結果は、ハプロタイプ情報を利用した様々な方法論の評価に、今回同定したハプロタイプ情報が有用である可能性を示唆している。

(2) 連鎖不平衡係数を指標とし、確定ハプロタイプと推定ハプロタイプ構造を比較した。その結果、推定ハプロタイプの精度は距離依存的に減少することを示した。この結果は、長距離の連鎖不平衡度を利用する方法論にとって、推定作業を要しない確定ハプロタイプ情報を適用することの重要性を意味する。

(3) 自然選択を検出する統計量の一つである iHS 法 (Voight et al., 2006, PLoS Genet.) は、長距離に渡って伸長しているハプロタイプのホモ接合度 (連鎖不平衡度) を利用する。推定ハプロタイプがこの統計量に及ぼす影響を調査するため、独自の統計的指標の開発を行った (Higasa et al., 2009, PLoS Genet.)。この指標を用いることにより、ハプロタイプ精度の効率的評価が可能となった。この指標を用いて、ゲノムワイドなデータに適用を行った。その結果、既知のゲノム構造多型が存在する領域 (MHC 領域、染色体 17 番の逆位多型領域) が大きくハプロタイプ精度の影響を受けていることが明らかとなった。即ち、確定ハプロタイプ情報は、より正確な自然選択領域の同定だけでなく、未知のゲノム構造多型の検出にも有用である可能性を示唆している。

これらの解析結果は、コールドスプリングハーバー会合 (New York, America)、及び、アメリカ人類遺伝学会 (Pennsylvania, America) にて発表した。

(4) 上記の独自の統計的指標に基づき、日本人におけるより正確な自然選択候補領域を検出し、カタログ化した。既知の領域との詳細な比較解析により、推定ハプロタイプからでは検出の困難であった新規の構造多型領域が含まれていることが示唆された。

(5) 更に、家系情報が得られない場合における推定ハプロタイプの予測精度改善に、確定ハプロタイプを利用する試みを行った。推定ハプロタイプの精度は、既知のハプロタイプ情報を参照することにより、その精度を改善できる可能性が示唆されていた。そこで、確定ハプロタイプ情報を、ハプロタイプを推定する際の参照配列として用いることにより、推定ハプロタイプの精度が向上することを示した。

これらの結果は、50 万個という詳細なタイピングを行うことにより明らかとなった、確定ハプロタイプの重要な利点である。今回決定した確定ハプロタイプデータは、有用なリソ

ースとしてデータベース化 (<http://finch.gen.kyushu-u.ac.jp/>) し、解析結果は、国際オンラインジャーナル誌 PLoS Genetics (Higasa et al., 2009) に発表した。

(5) 更に、160 万個の SNP についてタイピングを実施し、ゲノム構造多型の解析を行った。その結果、約 1300 個のゲノム構造多型領域を同定し、その約 6 割は新規のものであった。また、近傍のハプロタイプ構造と比較した結果、特定のハプロタイプでゲノム構造多型が生じやすい事が示された。この結果は、SNP がほぼランダムに生じるのと対照的であり、ゲノム構造多型の形成機構、及び、疾患との関連を解明するための重要な知見として、国際雑誌 (American Journal of Human Genetics, Kukita et al., 2010) に発表した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

① A definitive haplotype map as determined by genotyping duplicated haploid genomes finds a predominant haplotype preference at copy number variation events.

Kukita Y, Yahara K, Tahira T, Higasa K, Sonoda M, Yamamoto K, Kato K, Wake N, Hayashi K. Am. J. Hum. Genet., 2010, in press (査読有)

② Evaluation of haplotype inference using definitive haplotype data obtained from complete hydatidiform moles, and its significance for the analyses of positively selected regions.

Higasa K, Kukita Y, Kato K, Wake N, Tahira T, Hayashi K. PLoS Genet. 2009 May;5(5):e1000468. (査読有)

③ Narrowing of the regions of allelic losses of chromosome 1p36 in meningioma tissues by an improved SSCP analysis.

Guan Y, Hata N, Kuga D, Yoshimoto K, Mizoguchi M, Shono T, Suzuki SO, Tahira T, Kukita Y, Higasa K, Yokoyama N, Nagata S, Iwaki T, Sasaki T, Hayashi K.

Int J Cancer. 2008 Apr 15;122(8):1820-6. (査読有)

[学会発表] (計 3 件)

① 胞状奇胎から得られた確定ハプロタイプに基づく推定ハプロタイプの評価と自然選択解析への影響

日笠 幸一郎、久木田 洋児、田平 知子、林 健志
日本人類遺伝学会、第 54 回大会、2009/09/23 日本・東京

② Impact of definitive haplotypes on genetical

analysis.

Higasa K, Kukita Y, Tahira T, Hayashi K
58th Annual Meeting of the American
Society of Human Genetics 2008/11/14
Pennsylvania, USA

③ Impact of definitive haplotypes for
detecting recent positive selection.

Higasa K, Kukita Y, Tahira T, Hayashi K
Cold Spring Harbor Laboratory Meeting on
The Biology of Genomes 2008/05/08 New York,
USA

[図書] (計1件)

① Estimation of SNP allele frequencies by
SSCP analysis of pooled DNA.

Tahira T, Kukita Y, Higasa K, Okazaki Y,
Yoshinaga A, Hayashi K.
Methods Mol Biol. 2009;578:193-207.

[その他]

ホームページ等

<http://qsnp.gen.kyushu-u.ac.jp/>

<http://finch.gen.kyushu-u.ac.jp/>

6. 研究組織

(1) 研究代表者

日笠 幸一郎 (**Higasa Koichiro**)

独立行政法人理化学研究所、ゲノム医科学研究
センター、統計解析・技術開発グループ、
統計解析研究チーム・研究員

研究者番号：10419583