

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月 10日現在

機関番号：12602

研究種目：若手研究（B）

研究期間：2009～2012

課題番号：21700317

研究課題名（和文） 遺伝子多様性データを解明する統合的分析について

研究課題名（英文） Comprehensive analysis to study genetic diversity data

研究代表者

富田 誠 (TOMITA MAKOTO)

東京医科歯科大学・医学部附属病院・特任准教授

研究者番号：20399025

研究成果の概要（和文）：単純に大量なデータを解析するだけではなく、遺伝解析にはこの分野での独特のアプローチの仕方が各種ある。探索する領域を絞り込み、また遺伝情報と表現型情報を結び付ける関連解析を精密に行うことができる。探索する領域を絞り込むために、連鎖不平衡を利用した LD ブロックの同定を行い、その手法を空間統計学の手法である Echelon 解析を適用する手法を提案し、先行研究らと比較した。絞り込んだ領域について、ターゲットハプロタイプを想定した関連解析を行いたいが、初めて遺伝発現モデルに忠実に従った多変量量的変量への適用を行える手法を提案し、先行研究と比較を行った。その結果、全てのシミュレーションデータにおいて、我々が提案した手法が良い結果（正確な有意確率や検出力など）を得ることが示された。

研究成果の概要（英文）：We proposed and compared some methods of identifying LD block, selection of tagging SNPs and multivariate association analysis. It is an extension of QTLhaplo to the case of multivariate analysis, where we can analyze the cases where the diplotype configuration is not determined uniquely from the genotype and we can assume any dominant, recessive and additive models. To study the effectiveness of our method for the recessive and additive models, we carried out a numerical study to analyze two data sets with real genotype data (Hapmap project) and simulated QTL data, and for the dominant model we performed a small size simulation study to compare the powers with QTLmarc. The results showed that our method was more effective for detecting the genotype-to-phenotype relationship than QTLmarc of Kamitsuji and Kamatani (2006) in all the dominant, recessive and additive models.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,100,000	330,000	1,430,000
2010年度	800,000	240,000	1,040,000
2011年度	700,000	210,000	910,000
2012年度	0	0	0
年度			
総計	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：情報学、統計科学

キーワード：ゲノム統計解析、ハプロタイプ解析

1. 研究開始当初の背景

遺伝子多様性の研究には、多因子疾患関連遺伝子の探索を目的とした場合、従来から用いられている連鎖解析に加え、連鎖不平衡を利用した相関解析、ハプロタイプ解析が有力な手法として、世界的にも注目されている。我々は、それらの分析の一連の流れを統合的分析手法として整備した。具体的な分析方法として DNA データの中の代表的な SNP マーカーをブロックとして抜き出す LD ブロック同定方法、LD ブロック同定より代表的なマーカーを選択する htSNP 選択方法、その後のハプロタイプ・ディプロタイプ形推定、遺伝子型・ディプロタイプ形と表現型データとの関連解析について扱うことを想定していた。

2. 研究の目的

単純に大量なデータを解析するだけではなく、遺伝解析にはこの分野での独特のアプローチの仕方が各種ある。我々が提案した伝統的な遺伝学的アプローチにより、探索する領域を絞り込み、絞り込んだ後に、また遺伝統計学を駆使した理論構築により、遺伝情報と表現型情報を結び付ける関連解析を精密に行うことができる。

まずは遺伝子型データを把握することから始まるが、場合によっては膨大な座位数の遺伝子型データであることも多い。座位数が大きいかどうかに関わらず、遺伝子型の情報を絞り込むために、ブロック構造というものを把握する必要がある。連鎖不平衡の2次元的な構造を把握する方法を Tomita, et al. (2008)で提唱しており、基本的にはこれを適用して最適な探索すべきブロックを同定し、絞り込む。絞り込まれたブロック領域に対して、ハプロタイプ・ディプロタイプ形という集団と個体に対しての推定を行い、あるターゲットとなるハプロタイプを持つ群とそうでない群(統計解析でいうところの case-control study に近い)について、尤度比検定を行い、ターゲットが検出されるか否かを判断するアルゴリズムを提案し、先行研究と比較する。

3. 研究の方法

LD block (連鎖不平衡ブロック) とハプロタイプについて、ブロック領域の探索を行ってきたが、その発展としてブロック領域の解析手法を検討する。ハプロタイプ推定の手法はこれまで様々報告されている。実際のデータ解析に適用し得られた結果について考察する。さらに、これらの結果から解析手法の妥当性を

検討する。

絞り込まれたブロック領域に対して、ハプロタイプ・ディプロタイプ形という集団と個体に対しての推定を行い、あるターゲットとなるハプロタイプを持つ群とそうでない群を用いて、関連解析を行うことができる。量的変量が1変量の場合は既に提案されているが、2変量以上の多変量となった場合、正確に遺伝モデルを適用した手法がないため、アルゴリズムを提案する。その結果、確率的なディプロタイプ形に対応でき、遺伝的発現モデルは3つあり、それぞれドミナント、レセッシブ、アディティブモデルについて対応できるように構築する。

4. 研究成果

連鎖不平衡を利用した LD ブロック同定について、Tomita, et al. (2008)のように Echelon 解析を利用した同定法を提案している。このときに用いた実データである Hapmap 計画の genotype データを用いて様々な検討を行った。用いたデータの LD マップは図1である。

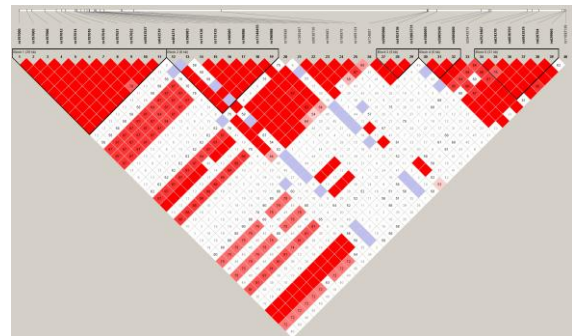


図1. 40座位でのLDマップ(D')

この領域でブロック同定を行うと、HaploviewでのGabriel法よりも、我々の手法は広くフレキシブルな領域を同定でき、Kamatani et al. (2004)の手法で得られるブロックとよく似た広さで得られた。しかし、彼らの手法に比べて計算時間は格段に短くなった。

また、このEchelon解析を利用して、ブロックを同定した後のタグSNPを選択することにも適用する手法を提案した。これについて、まとめたものを2011年にJournal of the Korean Data Analysis Societyに投稿し、掲載されている。やや難解な解釈を必要とするが、HaploviewのTagger法やAvi-Itzhak et al. (2003)のloss-lessモード法で得られる結果と非常によく似た結果が得られた。(図2参照)

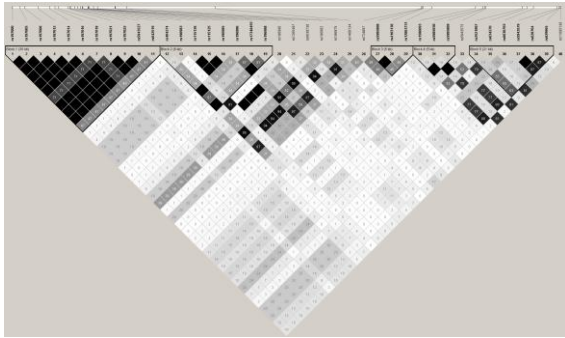


図 2. 40 座位での LD マップ (r^2)

また、ハプロタイプを用いた関連解析について、多変量 QTL への適用する手法を提案した。

単変量 QTL への適用は QTLhaplo として 2004 年に既に提案されているが、多変量 QTL への純粋な拡張は行われていなかった。多変量 QTL とハプロタイプの関連解析について、先行研究は QTLmarc という手法があるが、これは遺伝子発現モデルは優性モデルにしか適用できず、またハプロタイプ推定は行わず、アレルの総当たりのハプロタイプ列挙を行い、適用しているため、実際の評価には問題があることは自明であった。そこで我々は QTLhaplo の純粋な多変量 QTL への拡張を考え、これをまとめたものを 2011 年に *Computational Statistics and Data Analysis* に投稿し、掲載された。

この手法によって、3 つの遺伝発現モデル全て、ドミナント、レセッシブ、アディティブモデルについて対応でき、ハプロタイプ推定は、推定を行った結果をそのまま当てはめているため、確率的なディプロタイプ形にも対応している。

実データでの比較では、Tomita, et al. (2008) のデータを用いているが、図 1 や図 2 にみられる block1 と block2 を結合して、得られたタグ SNP4 つを用いて、行った。44 subjects について、ハプロタイプおよびディプロタイプ形推定を行い、それぞれ確定的もしくは確率的なディプロタイプ形を得ることができた。その中で、最も頻度が高かったハプロタイプをターゲット・ハプロタイプとみなし、関連解析のシミュレーションを行っている。これについて、詳細は Tomita, et al. (2011) や 富田 (2012) を参照いただきたいが、関連解析でのシミュレーションにおいて、先行研究である QTLmarc との比較を行っているが、彼らが適用している優性モデルでも、我々の手法の方が良い検出力が得られた。これは、我々の手法が 3 つの発現モデル全てに適用できただけでなく、確率的なディプロタイプ

形にも正確に適用できたことが大きなメリットとして現れたのではないかと考えられる。

以上のように、今までにないハプロタイプ解析において、ブロック同定、タグ SNP 選択方法、ハプロタイプ・ディプロタイプ形での関連解析について、それぞれの手法を提案でき、特に関連解析については現時点で最も良い手法を確立できた。

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 7 件)

- 1) 富田 誠, ハプロタイプ解析における推定・ブロック同定および関連解析, 計算機統計学, 査読有, 25(1), 53-68, (2012, 12)
http://ci.nii.ac.jp/vol_issue/nels/AN10195854_ja.html
- 2) Makoto TOMITA, Koji KURIHARA and Sungho MOON, An Application to Select Tag Loci by Using Hierarchical Structures of DNA Markers, Journal of the Korean Data Analysis Society, 査読有, 13(6), 2749-2762. (2011, 12)
<http://www.kdas.or.kr/zeroboard/view.php?id=journal&no=2064>
- 3) Chieko MITAKA and Makoto TOMITA, Polymyxin B-immobilized Fiber Column Hemoperfusion Therapy for Septic Shock, Shock, 査読有, 36(4), 332-338. (2011, 10)
DOI: 10.1097/SHK.0b013e318225f839
- 4) Makoto TOMITA, Noboru HASHIMOTO and Yutaka TANAKA, Association Study for the Relationship Between a Haplotype or Haplotype Set and Multiple Quantitative Responses, Computational Statistics and Data Analysis, 査読有, 55(6), 2104-2113. (2011, 6)
DOI: 10.1016/j.csda.2011.01.002
- 5) 富田 誠, 藤田 利治, 神出 計, 花田 裕典, 宮田 敏行, 河野 雄平, 大規模なゲノムデータにおける関連解析の手法とソフトウェア, 計算機統計学, 査読有, 22(2), 131-142. (2010, 5)
<http://ci.nii.ac.jp/naid/110007619733>
- 6) Makoto TOMITA, Hitomi TANIAI, Sungho MOON, Takeshi NISHIYAMA and Satoshi SUMI, Heritability of Cerebral Palsy Using Structural Equation Modeling, Journal of the Korean Data Analysis Society, 査読有,

11(4), 1711-1717. (2009, 8)
<http://www.kdas.or.kr/zeroboard/view.php?id=journal&no=1434>

- 7) Makoto TOMITA, Takeshi NISHIYAMA, Hitomi TANIAI, Sungho MOON, Miyachi TAISHI and Satoshi SUMI, Genetic Investigation of Some Behaviors and Impairments with Autism Spectrum Disorders Using Structural Equation Modeling, Journal of the Korean Data Analysis Society, 査読有, 11(3), 1127-1134. (2009, 6)
<http://www.kdas.or.kr/zeroboard/view.php?id=journal&no=1389>

[学会発表] (計 6件)

- 1) 富田 誠, DNA データにおける関連解析での遺伝発現モデルの選択と拡張, 第12回 西東京統計研究会 予稿集, 1-4 頁, 東海大学高輪キャンパス, (2013, 3)
- 2) 富田 誠, DNA データにおける関連解析とその周辺の問題, 2012 年度統計関連学会連合大会講演報告集, 47 頁, 北海道大学, (2012, 9)
- 3) Makoto TOMITA and Koji Kurihara, Association Study and Comprehended Hierarchical Structures on DNA Markers, Book of Abstracts, Analysis and Modeling of Complex Data in Behavioural and Social Sciences (JCS-CLADAG2012), p. 83, Anacapri, Italy, (2012, 9)
- 4) Makoto TOMITA, Comparison among spatial clustering methods on hierarchical structures for DNA markers, Programme and Abstracts, 20th International Conference on Computational Statistics (COMPSTAT2012), p. 28, Limassol, Cyprus, (2012, 8)
- 5) 富田 誠, 栗原 考次, 文 勝浩, DNA データにおける LD 集積性と代表 SNP の選択, 日本計算機統計学会 第 25 回大会 論文集, 37-38 頁, 函館市亀田福祉センター, (2011, 5)
- 6) Makoto TOMITA and Koji KURIHARA, Comprehensive Assessment on Hierarchical Structures of DNA markers Using Echelon Analysis, Proceedings in Computational Statistics, 1645-1653. (2010, 8)

[図書] (計 0件)

[産業財産権]

○出願状況 (計 0件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

○取得状況 (計 0件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]
ホームページ等
なし

6. 研究組織

(1) 研究代表者

富田 誠 (Makoto Tomita)
東京医科歯科大学・医学部附属病院・特任
准教授
研究者番号 : 20399025

(2) 研究分担者

()

研究者番号 :

(3) 連携研究者

()

研究者番号 :

