

平成 21 年 5 月 18 日現在

研究種目：基盤研究 (B)

研究期間：2006～2008

課題番号：18310131

研究課題名 (和文) アレル依存的発現と転写調節領域ハプロタイプの関連の解析

研究課題名 (英文) Analysis of allele-specific expression and its correlation with haplotypes

研究代表者

田平 知子 (TAHIRA TOMOKO)

九州大学・生体防御医学研究所・助教

研究者番号：50155230

研究成果の概要：

ゲノム上の発現調節領域に存在する多型はアレル依存的発現をもたらす。これを検出するため 2つのアレルに対応する cDNA の量比を SSCP 法により定量する方法を確立し、さらに、DNA タイピング用マイクロアレイにより解析する条件を検討した。1次スクリーニングにはマイクロアレイによるゲノムワイド解析を、2次スクリーニングには測定精度のよい SSCP 法を利用することが効率的であると結論した。また、転写因子結合配列の挿入欠失多型を含むハプロタイプと発現との強い関連を見出した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	5,700,000	1,710,000	7,410,000
2007年度	5,100,000	1,530,000	6,630,000
2008年度	4,600,000	1,380,000	5,980,000
年度			
年度			
総計	15,400,000	4,620,000	20,020,000

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、遺伝子発現、SNP

1. 研究開始当初の背景

生体を構成する各細胞における遺伝子発現レベルの個体差は、体質や疾患感受性の個体差と深くかかわっていることが明らかになってきた。このような個々人の遺伝子発現レベルの差異をもたらす発現量の形質遺伝子座 (eQTLs) を明らかにするため 1塩基多型 (SNP) を用いた解析が進行しており、特に遺伝子近傍の発現調節領域のなかにその遺伝子の発現量の個体差と強く関連している多型 (*cis*-regulatory variant) があること

が明らかになってきた。われわれは遺伝子発現に影響を及ぼす多型のリソースとして日本人検体で転写開始点周辺の SNP を探索し dbQSNP データベースとして公開してきた。また日本人ゲノムの連鎖不平衡構造を明らかにするため、完全胎状奇胎 DNA を用いて確定ハプロタイプを決定し D-Haplo データベースとして公開してきた。これらの情報および国際 HapMap 計画により整備されたヒトゲノムの多様性データベースをもとに eQTL 解析を進展させることが可能と考えた。特に、われ

われが SNP の検出のために開発した蛍光 SSCP 解析はアレル依存的発現の解析のために有用であると考え、そのための実験系を構築した。

2. 研究の目的

ヒト全ゲノム解読および国際 HapMap プロジェクトによる一塩基多型 (SNP) の大規模解析の成果により、ゲノムワイドに疾患感受性関連解析を行い、責任遺伝子を探ることが可能になった。一般集団での頻度が高い病気 (common disease) の遺伝的要因として、タンパク質の質的変化をもたらすような多型よりも、発現の量的変化をもたらす多型の関与が大きいことが明らかになっている。実際に、このような疾患と統計的に非常に有意に相関した SNP の多くは非コード領域にあり、多くは遺伝子発現調節と関与していると考えられる。このなかで、遺伝子近傍での発現制御に関連した多型 (*cis*-regulatory variant) はアレル依存的発現を調べることにより同定することが可能である。本研究はアレルによる発現レベルの違いを定量するシステムを確立し、ハプロタイプ情報と統合的に解析することにより発現制御多型を同定し、遺伝子発現の調節機構およびそれに起因する疾患の解明に有用な情報基盤を提供することを目的とする。

3. 研究の方法

1) SSCP 法による SNP アレル分離・定量法の確立

アレル依存的発現の検出には 2 つのアレル (ハプロタイプ) から発現する mRNA の量比、実際には mRNA から逆転写反応により生成される cDNA 断片の量比を定量する。この cDNA は 2 つのアレル由来の混合物であるので、実験系の検証のためのモデルとして、多数のサンプルからプールした DNA を鋳型とする。対象とする SNP はマイナーアレル頻度 5-50% の間とし、様々なアレル量比のものを対象とする。この SNP を含む配列を増幅した PCR 産物を蛍光標識した後、非変性条件で電気泳動することにより各アレルをピークとして分離・定量する。そのときに、アレル間での PCR 増幅の違い等が生ずるので、ヘテロ体の DNA サンプルも同時に解析し、その 2 つのアレルのピーク比を補正に用いる。解析にはシーケンズ・SSCP 解析用に開発中の「QSNPlite」ソフトウェアを用いる。プールサンプルを個別にタイピングすることでアレル頻度の実測値を算出し、プール DNA の SSCP 解析での推定の精度を検討する。

2) DNA マイクロアレイによる SNP アレル量比定量法の確立

遺伝子型解析済みの 42 人 (Coriell 研究所か

ら購入、欧米系白人 42 人、あるいはアフリカ系黒人 42 人) の DNA を等モルで混合した 2 種類のプール DNA を GeneChip Human Mapping 100K Set (Affymetrix 社) により遺伝子型タイピングと同じ条件で解析する。スキャナで蛍光シグナルを取得したあと、それぞれのアレイについて各プローブの蛍光シグナル強度が記載されたファイルを出力する。それをもとに、2 つのアレルの量比を計算し、個別タイピングの結果すでに得られているアレル頻度の実測値との相関を調べる。

3) 疾患関連多型と発現との関連解析
全身性エリテマトーデスの疾患関連解析で同定された座位について、HapMap (<http://www.hapmap.org/>) および D-HaploDB で連鎖不平衡を調べる。また、データベースに含まれていない SNP については国際 HapMap 計画で解析された日本人 (JPT) のタイピングを行い、これらを含めた連鎖不平衡構造を決定し、それをもとに近傍で関連の高い SNP を検索する。遺伝子発現と多型の関連を調べるためには、主に以下の eQTL データベースを検索する。

- GENEVAR - GENE Expression VARIation: <http://www.sanger.ac.uk/humgen/genevar/>
- Gene Expression Analysis Based on Imputed Genotypes: <http://www.sph.umich.edu/csg/liang/imputation/>

4) 新規に同定した挿入欠失多型の発現との関連解析

遺伝子転写開始点周辺の塩基配列を解析し、挿入欠失変異を同定する。挿入欠失配列の解析には「QSNPlite」ソフトウェアに組み込んだ独自開発のヘテロ配列解析機能を用いる。同定された多型により、転写因子結合部位が変化するかどうかを転写因子結合配列データベース (TransFac Professional ソフトウェアの転写因子結合部位予測ツール MATCH) を用いて解析する。転写因子結合部位を変化させる多型については HapMap で解析された日本人 (JPT) のタイピングを行い、得られた遺伝子型をもとに完全に連鎖した SNP を検索する。さらに、その SNP が近傍の遺伝子の発現に影響を及ぼすものであるかどうかをデータベースで検索する。

4. 研究成果

1) SSCP 法による SNP アレル分離・定量法の確立

エクソン領域にある SNP をマーカーにしてアレル間の発現量の比をキャピラリー自動シーケンサーによる蛍光 SSCP 法 (PLACE-SSCP 法) で高精度に定量する系を確立した。条件検討にはプール DNA を使い、解析結果を定量するためには、SNP アレル頻度

定量用に開発中の「QSNPlite」ソフトウェアを利用した。このときにヘテロ体での2つのアレルのピーク比が1:4以上のものは定量の再現性が悪いので除外した。このシステムにより42種類のSNP周辺配列を解析し、PCR反応前の鋳型中の2つのアレルの比を高精度 ($R^2 = 0.994$) に定量できることを確認した。また、SSCP解析で各アレルの分離が不十分な場合に、人為的にアレル特異的構造体形成を促進する目的でプライマーにSNP周辺配列と相補な9塩基のタグ配列をつけることにより分離を改善する方法を開発した。実際にSSCPでの分離が不十分な配列40個についてタグ配列を設計したところ、そのうち約8割では分離が改善し定量が可能となった。この方法によっても高精度な定量が可能であることを確認した。さらに、プールDNAの代わりにmRNAから作製したcDNAを鋳型に用いることにより、同様にアレル間の発現量の比を再現性よく定量できることを確認した。この方法は単一検体のアレル依存的発現の定量、および複数検体をプールして同様の定量を行う場合に応用可能である。以上の実験により、正確なアレル依存的発現定量のための実験系および解析ソフトウェアを整備した。

2) DNA マイクロアレイによるSNPアレル量比定量法の確立

ゲノム全体からアレル依存的に転写される領域を同定するためには大規模な1次スクリーニング法を採用する必要がある。そこで、アレル間の発現の比較定量にSNPタイピング用のマイクロアレイを使う方法を検討した。この方法では各アレル由来の転写産物による蛍光シグナルの比を同じ細胞のゲノムDNAの各アレルのシグナルの比で補正したものをアレル特異的転写物の量比として扱う。定量性を検討する予備実験として遺伝子型解析済みの42人のDNAを等モルで混合したプールDNA2種類をマイクロアレイで解析し、アレル頻度の定量性を検討した。用いたアレイはバックグラウンド補正用にミスマッチプローブを含んでいたが、これらによる補正を行わないほうがデータの再現性が良くなった。各SNPごとに、プールDNAを検体とした場合の完全マッチプローブの各アレルのシグナルの比はアレル頻度の実測値と相関していた ($r^2 = 0.95$)。これをヘテロ検体の同シグナル比から得られるkファクターで補正したところ、更により相関 ($r^2 = 0.98$) を認めた。したがって、マイクロアレイをアレル間の発現の比の定量に使用することは妥当であると結論した。しかし、個別に調べるとSSCP法に比べてアレル量比の定量精度は低いので、1次スクリーニングにはマイクロアレイを、2次スクリーニングにはSSCP法を利用することが効率的であると結論した。

これによってアレル特異的転写産物の定量をゲノムワイドに行うための重要な技術基盤を確保した。

3) 疾患関連多型と発現との関連解析

自己免疫疾患である全身性エリテマトーデスの関連解析によりIRF5遺伝子のSNPが疾患と強い関連を示すことがわかった。IRF5遺伝子領域の疾患関連SNPはスプライシング、polyA付加を含む複数の発現調節機構に関与していることが欧米の研究で示されたが、それが疾患感受性の原因となっているかについては結論がでていない。われわれは日本人患者の解析により欧米で同定された関連のいくつかについて否定的結果を得ており、疾患感受性関連多型に人種的差異があることを示した。一方、遺伝子上流の転写因子Sp1結合配列のコピー数多型(5bpの挿入欠失多型)と疾患の関連は再現され、われわれの検体では最も強い関連を示した。この多型が発現と関連していることを証明するためには多数の検体でIRF5遺伝子発現を調べ統計的解析を行う必要があるが、サンプルの入手が困難なので、ヒトリンパ芽球細胞についてゲノムワイド発現アレイで各遺伝子の発現レベルを調べ、それとSNPの遺伝子型との関連解析を行った外部eQTLデータベースを利用した。主なデータベースとしてHapMap構築に利用したアジア人・アフリカ人・欧米人計270人の検体を用いたもの(GENEVAR、サンガー研究所)、および欧米人約400人の検体を用いたもの(Gene Expression Analysis Based on Imputed Genotypes、ミシガン大学)を利用した。挿入欠失多型と高度に関連している ($r^2 = 0.89$) 近傍のSNP(rs4728142)は約400人の発現解析によりIRF5発現と強く関連 ($P = 2.6 \times 10^{-16}$) していることが明らかになった。発現制御の機構は人種によらず同様であると考え、われわれの結果はIRF5の発現が高いタイプの多型が疾患感受性と関連を示すと解釈される。これは、疾患がインターフェロンパスウェイの亢進と関連しているという観察と一致するもので、転写因子Sp1結合配列多型が有力な原因変異であるという可能性を示す。この結果は発現解析とハプロタイプ情報の統合が疾患感受性原因多型を同定する有効な手法であることを示す。

4) 新規に同定した挿入欠失多型の発現との関連解析

転写開始点周辺領域での挿入欠失多型は転写因子結合に大きな影響を与える可能性があるため、これに注目した。これまでdbQSNPデータベース構築のための実験により、約800個の挿入欠失多型を新規に見つけており、そのうち360個は日本人サンプルでのアレル頻度が5%以上のものであった。これらについ

て、転写因子結合部位予測ツール MATCH の最新版で解析したところ約 1/4 は既知の転写因子結合配列のなかにあった。これらが、遺伝子発現に影響を与えている可能性を調べるために、これら挿入欠失多型と同じハプロタイプにある SNP について近傍の各遺伝子の発現と強い関連を示すかどうかを前述の eQTL データベースにより検討した。その結果、これらのうち FTSJ3 遺伝子近傍の 14 塩基の挿入欠失多型（われわれが dbSNP にデータ提供後、rs28362565 として登録された）が HapMap アジア人検体で rs2584624 と完全に連鎖していること、及びこの rs2584624 の遺伝子型が発現と強く関連している ($P = 4 \times 10^{-11}$) ことを見出した。この挿入欠失多型は、転写因子 Nkx2-5 の結合部位のコピー数が 1 または 2 となるものであり、コピー数が多いほど発現量が低下した。以上のような手法は、転写配列結合配列の多型とアレル特異的発現の関連の新たな同定を可能にするものであり、dbQSNP データベース等に蓄積した転写制御領域多型情報を利用して転写ネットワークと疾患の関連解析を行うために有用であることを示した。

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 6 件)

- ① Tahira T, Kukita Y, Higasa K, Okazaki Y, Yoshinaga A, Hayashi K. Estimation of SNP allele frequencies by SSCP analysis of pooled DNA. *Methods in Molecular Biology* (査読無): in press
- ② Miyagawa, H, Yamai, M, Sakaguchi, D, Kiyohara, C, Tsukamoto, H, Kimoto, Y, Nakamura, T, Lee, J-H, Tsai, C-Y, Chiang, B-L, Nagasawa, K, Harada, M, Tahira, T, Hayashi, K, and Horiuchi, T. Association of polymorphisms in complement component C3 gene with susceptibility to systemic lupus erythematosus. *Rheumatology* (査読有) **47**: 158-164, 2008
- ③ Qin, M, Kondo, H, Uno, H, Tahira, T, and Hayashi, K. Moderate reduction of Nurrin signaling activity associated with the causative missense mutations identified in patients with familial exudative vitreoretinopathy. *Human Genetics* (査読有) **122**: 615-623, 2007
- ④ Higasa, K, Miyatake, K, Kukita, Y, Tahira, T, and Hayashi, K. D-HaploDB: a database of definitive haplotypes determined by genotyping complete hybridized mole samples. *Nucleic Acids Res.* (査読有) **35**: D685-689, 2007
- ⑤ Tahira, T, Okazaki, Y, Miura, K, Yoshinaga, A, Masumoto, K, Higasa, K, Kukita, Y, and Hayashi, K. QSNPlite, a software system for quantitative analysis of SNPs based on capillary array SSCP analysis. *Electrophoresis* (査読有) **27**: 3869-3878, 2006
- ⑥ Jespersgaard, C, Larsen, LA, Baba, S, Kukita, Y, Tahira, T, Christiansen, M, Vuust, J, Hayashi, K, and Andersen, PS. Optimization of capillary array electrophoresis single-strand conformation polymorphism analysis for routine molecular diagnostics. *Electrophoresis* (査読有) **27**:3816-3822, 2006

[学会発表] (計 6 件)

- ① 田平知子, 他. DNAプールのマイクロアレイ解析および定量的SSCP解析による全身性エリテマトーデス (SLE) 疾患感受性遺伝子の大規模探索 第31回日本分子学会年会, 2008年12月9日, 神戸
- ② Tahira T, et al. Identification of loci for systemic lupus erythematosus by pooling-based genome-wide association study. 57th Annual Meeting of the American Society of Human Genetics, 2008年11月12日, Philadelphia, U. S. A.
- ③ 田平知子, 他. DNAプールのDNA-chip解析による全身性エリテマトーデス (SLE) 疾患感受性遺伝子の大規模探索 第30回日本分子学会年会, 2007年12月13日, 横浜
- ④ Tahira T, et al. Pooling-based genomewide association study identifies loci for systemic lupus erythematosus. 7th Annual Meeting of the American Society of Human Genetics 2007年10月26日 San Diego, USA
- ⑤ 田平知子, 他. IRF5 遺伝子多型の全身性エリテマトーデス感受性への関与: アジア人集団を用いた解析 第 51 回日本リウマチ学会総会・学術集会, 2007年4月26日, 横浜
- ⑥ Tahira T, et al. Precise estimation of allele frequencies in pooled DNA based on capillary array SSCP analysis. The 8th international meeting on human

genome variation and complex genome analysis. 2006年9月15日, Hong Kong, SAR, China

[その他]

- ① 「dbQSNP」データベース
<http://qsnp.gen.kyushu-u.ac.jp/>
- ② 「D-HaploDB」データベース
<http://orca.gen.kyushu-u.ac.jp/>
- ③ 「QSNPlite」ソフトウェア
<http://qsnp.gen.kyushu-u.ac.jp/placeSSCP/>

6. 研究組織

(1) 研究代表者

田平 知子 (TAHIRA TOMOKO)
九州大学・生体防御医学研究所・助教
研究者番号：50155230

(2) 研究分担者

林 健志 (HAYASHI KENSHI)
九州大学・生体防御医学研究所・特任教授
研究者番号：00019671