

200/0423

厚生科学研究

ヒトゲノム・再生医療等研究事業

びまん性汎細気管支炎等遺伝要因を有する
慢性呼吸器疾患の疾患感受性遺伝子の研究

平成13年度 研究報告書

主任研究者 慶長直人

平成14年3月

目 次

I. 総括研究報告書

- びまん性汎細気管支炎等、遺伝素因を有する慢性呼吸器疾患の疾患感受性遺伝子の研究…………… 3
慶長 直人

II. 分担研究報告

1. びまん性汎細気管支炎における疾患感受性遺伝子の候補領域における SNP マーカー解析…………… 7
慶長 直人 他
2. 肺組織で発現する HLA クラス I 領域の新規遺伝子同定…………… 11
田宮 元
3. 分子遺伝学上の統計解析…………… 15
大橋 順
4. 慢性閉塞性肺疾患の感受性遺伝子の研究
ー慢性気管支炎とヒト β デフェンシン I 遺伝子変異との関連性…………… 25
松下 育美 他
5. 肺サルコイドーシスの病態に関与する候補遺伝子群の変異スクリーニングに関する検討…………… 30
田中 剛 他

III. 研究成果の刊行に関する一覧表…………… 37

IV. 研究成果の刊行物・別刷

1. Diffuse Panbronchiolitis Role of Macrolides in Therapy…………… 45
慶長 直人 他
2. Prolonged Survival of a Bare Lymphocyte Syndrome Type I Patient
with Diffuse Panbronchiolitis Treated with Erythromycin…………… 58
吾妻安良太 他
3. Genetic Variants of Human β -Defensin-1 and Chronic Obstructive Pulmonary Disease…………… 60
松下 育美 他
4. Association of Gc-globulin variation with susceptibility to COPD and diffuse panbronchiolitis…………… 66
石井 健男 他
5. Overestimated frequency of a possible emphysema-susceptibility allele
when microsomal epoxide hydrolase is genotyped
by the conventional polymerase chain reaction-based method…………… 71
慶長 直人 他
6. New polymorphic microsatellite markers in the human MHC class III region…………… 74
松坂 恭成 他
7. Genomic Anatomy of a Premier Major Histocompatibility Complex Paralogous Region
on Chromosome 1q21-q22…………… 82
椎名 隆 他

8. The power of genome-wide association studies of complex disease genes:statistical limitations of indirect approaches using SNP markers.....	96
大橋 順 他	
9. Comparison of statistical power between 2×2 allele frequency and allele positivity tables in case-control studies of complex disease genes	101
大橋 順 他	
10. びまん性汎細気管支炎のゲノム解析.....	111
慶長 直人 他	
11. 人種特異性と疾患感受性遺伝子.....	115
慶長 直人	
12. 病態解析と SNP 呼吸器疾患 (慢性閉塞性肺疾患).....	121
松下 育美 他	
13. COPD 発症につながる危険因子ー環境要因と遺伝要因の関与とその対策ー	128
松下 育美 他	
14. 疾患感受性遺伝子の探求ーびまん性汎細気管支炎からサルコイドーシスまで.....	133
慶長 直人	
15. びまん性気管支拡張を呈し DPB と診断しえた若年者で CFTR 遺伝子の異常を認めた 1 例.....	140
神尾孝一郎 他	
16. NADPH / NADH oxidase の遺伝子多型とびまん性汎細気管支炎 (DPB), 慢性閉塞性肺疾患 (COPD) との関連についての検討	145
石井 健男 他	

I. 総括研究報告書

びまん性汎細気管支炎等、遺伝素因を有する慢性呼吸器疾患の 疾患感受性遺伝子の研究

主任研究者 慶長直人 国立国際医療センター 研究所・呼吸器疾患研究部

研究要旨 遺伝素因を有する慢性呼吸器疾患の疾患感受性遺伝子研究として、びまん性汎細気管支炎、慢性閉塞性肺疾患（肺気腫・慢性気管支炎）、サルコイドーシスを対象疾患として研究を開始した。びまん性汎細気管支炎に関しては、我々の手により、すでに HLA-A, B 遺伝子間の 200 kb の領域に主要疾患感受性遺伝子のひとつが存在することが推定されており、その領域内の遺伝マーカーの検出、特に一塩基多型 (SNPs) によるさらなる候補領域の絞り込みと、新規遺伝子のクローニングが並行して行われた。その結果、本年度、びまん性汎細気管支炎の疾患感受性遺伝子の候補領域内に 80 個以上の SNP が同定されるとともに、2つの新規遺伝子がクローニングされた。現在それらの個々の SNP と疾患との関連性、連鎖不平衡解析およびハプロタイプ分析を進めている。サルコイドーシスについては、Th1 免疫に関する遺伝子群に関するマイクロサテライトマーカーを利用した候補遺伝子アプローチを施行され、慢性閉塞性肺疾患については、自然免疫に関する候補遺伝子アプローチが進められた。

A. 研究目的

厚生労働省特定疾患のひとつである、びまん性汎細気管支炎 (DPB) においては、一連の研究により、ヒト白血球抗原 (HLA) と疾患との間に強い関連性が見られることが明らかになっている。我々はさらに未知の DPB 感受性遺伝子座が HLA-A 座と B (C) 座との間に位置し、DPB に感受性を示す遺伝子の変異が HLA B54-Cw1-A11 を保有する祖先染色体に生じたという仮説を提唱し、HLA 関連感受性遺伝子座の候補領域を分子遺伝学的に推定した。そこで、本研究では、さらに推定された 200 kb の領域内のエクソンと予測される部位を中心に、網羅的な一塩基置換 (SNPs) の同定を行い、それら SNPs のタイピングを施行するとともに、領域内の新規機能的遺伝子の同定を進めることとした。その他、慢性あるいは難治性呼吸器疾患についても、これまで、系統的な研究に乏しかったため、特にサルコイドーシス、慢性閉塞性肺疾患について、新たに検討を加えることとした。

B. 研究方法

厚生省班診断基準に基づき、診断された症例を選んで、解析を行った。疾患感受性遺伝子の候補領域を特定するアプローチに関しては、特にびま

ん性汎細気管支炎の候補領域となる 200 kb の領域について、コンピュータ予測されるエクソン部を中心に PCR 増幅し、直接シーケンス法により SNP を同定するとともに、EST データベースとの対照から、RT-PCR 法を利用して、新規遺伝子のクローニングを行った。疾患感受性遺伝子の候補遺伝子を個々に検討するアプローチに関しては、サルコイドーシスでは、Th1 免疫に関わる候補遺伝子座周辺のマイクロサテライトマーカーの検索と連鎖不平衡解析を行い、慢性閉塞性肺疾患に関しては、比較的稀な SNP と病型のサブタイプとの関連について、症例対象研究を主体に検討を行った。

(倫理面への配慮)

本研究における遺伝子解析に関しては、いずれも三省合同の「ヒトゲノム・遺伝子解析研究に関する倫理指針」に準拠した当センターの遺伝子解析に係わる倫理委員会の承認を受けている。

C. 研究成果

1. 解析対象領域の決定と遺伝子予測

DPB 主要感受性遺伝子候補領域である 200 kb の塩基配列について、GenScan コンピュータプログラムを用いて遺伝子予測を行った。予測された遺伝子のエクソン部位を中心に、プライマーセット

をデザインし、塩基配列を決定した。

2. 候補領域における SNPs の同定

予測されたエクソン部位を中心にプライマーをデザインし、PCR 増幅し、塩基配列を決定し、80 以上の SNPs を同定した。

3. 症例対象研究において、連鎖不平衡を利用した疾患関連遺伝子変異の統計学的検出

1. 2. の情報を用いて、疾患関連遺伝子変異を分子遺伝学的に同定するための統計学的基盤の確立のための検討を行った。

4. 候補領域における新規遺伝子の同定

EST データベースの検索の結果、200 kb の領域に 8 個の転写単位が予測され、それぞれ肺での発現が確認されたため、そのうち 2 個の遺伝子の全長クローニングを行った。さらに他の遺伝子のクローニングも進めている。

5. 候補遺伝子周辺のマイクロサテライトマーカーを用いた候補遺伝子の効率的なスクリーニング法の検討

サルコイドーシスの疾患感受性遺伝子検索のため、Th1 免疫関連遺伝子座より、100 kb 以内の領域にあるマイクロサテライトマーカーを同定し、疾患との関連性を検討し、STAT4 のマーカーとサルコイドーシスとの関連を示した。さらに連鎖不平衡の及ぶ範囲について検討を加えることにより、STAT4 遺伝子内に存在する、疾患関連 SNP の同定を試みている。

6. SNP データベースに登録されていない遺伝子マーカーと疾患との関連

慢性閉塞性肺疾患の候補遺伝子検索に際して、自然免疫に関わる遺伝子である β デフェンシン 1 の JSNP データベース等に登録されていない比較的頻度の低い、新たに同定された非同義置換と、疾患のサブタイプとの関連が明らかになった。

D. 考 案

慢性、難治性呼吸器疾患の多くは、ミレニアムゲノムプロジェクトの対象となっている五大疾患と同様、多因子疾患であるが、むしろ、それらの

いわゆるコモンディーズよりも、主要疾患感受性遺伝子の数が限られているのではないかと推測されている。たとえば、本研究の主要テーマのひとつであるびまん性汎細気管支炎は、欧米人に同様の疾患が見られず、欧米においてはアジア系の移民が罹患することが多いという事実から、人類創世後、アジア系集団の形成時に生じた遺伝子変異の一つが、疾患感受性を規定する有力な因子ではないかと考えられる。

また、慢性閉塞性肺疾患は、喫煙歴が長い高齢者に見られる疾患であり、喫煙という生活習慣と深い関係があり、特に喫煙による肺の破壊に感受性のある集団を規定する遺伝子変異がいくつか存在するのではないかと推定されている。それらの変異を保有する集団には、将来的、特に強く禁煙を勧めるなど、このような研究は疾患の一次予防に重要な役割を果たす可能性がある。

本研究では、このようにこれまで取り残されていた、呼吸器病の臨床に直結した疾患感受性遺伝子検索というテーマを掲げ、初年度、呼吸器病学の研究者と第一線のゲノム研究者が共同戦線をひき、ともにアプローチしていく理想的な体制づくりを行うことができた。さらに 2 年目以降の結果が待たれる。

E. 結 論

びまん性汎細気管支炎の HLA 関連疾患感受性遺伝子候補領域 200 kb 内に、80 個以上の SNP が同定され、2 個の候補遺伝子が同定された。さらにサルコイドーシス、慢性閉塞性疾患の疾患感受性候補遺伝子についても、症例対照研究の形をとりながら、新しい視点で検討を進め、成果を上げた。まとまった長さの遺伝子領域について、SNP とマイクロサテライトの連鎖不平衡の関係やハプロタイプ解析に関するデータはまだ国際的にも不十分であり、症例対照研究を進めていく上での、統計学的に妥当な研究デザインに関する検討、ヒトゲノムにおける連鎖不平衡の及ぶ範囲や高頻度組換え点に関する統計学的解析法の確立は、五大疾患など、さらに頻度の高い多因子疾患の候補遺伝子のマッピングにも重要な示唆を与えるものと予想される。

II. 分担研究報告

びまん性汎細気管支炎における疾患感受性遺伝子の 候補領域における SNP マーカー解析

主任研究者 慶長 直人 国立国際医療センター 研究所・呼吸器疾患研究部
土屋 朋子 国立国際医療センター 研究所・呼吸器疾患研究部
分担研究者 中田 光 国立国際医療センター 研究所・呼吸器疾患研究部
分担研究者 徳永 勝士 東京大学人類遺伝学教室
吾妻安良太 日本医科大学第四内科
分担研究者 工藤 翔二 日本医科大学第四内科

研究要旨 これまでの我々の検討により、びまん性汎細気管支炎の疾患感受性候補遺伝領域として、HLA-A, B 遺伝子間の S (CDSN; corneodesmosin) 遺伝子と TFIIH (GTF2H4; general transcription factor IIH, polypeptide 4) 遺伝子に挟まれた 200 kb の領域が推定されている。その領域を遺伝子予測プログラムによって解析し、予測されるエクソン部を中心に、一塩基多型 (SNPs) を探索し、全領域にわたる、のべ 30 kb を解析した結果、80 個以上の SNPs が同定された。患者群 92 例、対照群 100 例について、これらの SNPs のタイピングを現在、施行中である。日本人集団で 20% 以上の頻度で存在する SNPs は全体の 50% 以上を占めていた。さらに共同研究者らにより同定された遺伝子の SNPs と疾患との関連性について、タイピングを進めている。今後、疾患感受性を直接規定する SNP を分子遺伝学的に同定するために、領域内のさらなる新規遺伝子の同定と、SNP の連鎖不平衡解析およびハプロタイプ分析による詳細な推定を行う予定である。

A. 研究目的

1990 年代後半に行われた一連の検討により、日本人のびまん性汎細気管支炎 (DPB) 患者集団では HLA-B54 (対立遺伝子型では HLA-B*5401) の保有頻度が一般集団に比べ有意に高いこと¹⁾、韓国人の DPB 患者集団では B54 ではなく、HLA-A11 が DPB と強い関連性を示す²⁾ことが明らかになった。これらの研究から、我々は未知の DPB 感受性遺伝子座が HLA-A 座と B (C) 座との間に位置し、DPB に感受性を示す遺伝子の変異が HLA B54-Cw1-A11 を保有する祖先染色体に生じたという仮説を提唱した (図 1)。この仮説に基づき、該当領域に存在する、周辺のマイクロサテライトマーカーおよび既知の遺伝子の多型性部位を含む計 14 種類の遺伝的マーカーのタイピングを行い、感受性遺伝子座の候補領域を推定した。それは HLA-A, B 遺伝子間の S (CDSN; corneodesmosin) 遺伝子と TFIIH (GTF2H4; general transcription factor IIH, polypeptide 4) 遺伝子に挟まれた 200 kb の領域であり、その中に存在する 3 種のマイクロサテライト

マーカー (C2-2-2, C2-4-4, C4-4-3) は複数の疾患感受性ハプロタイプに共有され、疾患との関連性の強さを示すパラメーターのピーク値を示し、検索された範囲の韓国人患者にも高頻度で保持されていた³⁾。そこで、さらにこの 200 kb の領域内の機能的遺伝子の有無を検討する目的で、遺伝子予測プログラムによる推定を行い、領域内に予測されるエクソン部を中心に一塩基置換 (SNPs) の同定を行い、さらにそれらの SNPs のタイピングを施行した。

B. 研究方法

1995 年厚生省診断基準に基づき DPB と確認された DPB 患者 92 例、対照者 100 例を対象として解析を行った。200 kb の領域内に機能的遺伝子が予測されるか否かは、主に GenScan コンピュータプログラムを用いて推定を行った⁴⁾。その予測されるエクソン部を中心に PCR プライマーをデザインし、初めに、92 症例中 4 例、対象 100 例中 6 例について、それぞれ約 500 bp ずつ増幅し、

ABI PRISM 3100 DNA シーケンサー (Applied Biosystems) を用いた直接シーケンス法により決定された塩基配列の相違を SNP の候補とした。それらの結果を基に、症例、対照者全体について、同様な直接シーケンス法に基づくタイピングを実行中である。

(倫理面への配慮)

本研究における遺伝子解析に関しては、三省合同の「ヒトゲノム・遺伝子解析に関する倫理指針」に準拠した当センターの遺伝子解析に係わる倫理委員会の承認を受けている。

C. 研究成果

1. 200 kb の感受性遺伝子候補領域における遺伝子予測と解析対象領域

DPB 主要感受性遺伝子候補領域である 200 kb の塩基配列について、GenScan コンピュータプログラムを用いて遺伝子予測を行った。その結果、12 個の遺伝子の存在する可能性が予測された (図 1)。この 12 個のうち 3 個は全体が分散型繰り返し配列と考えられ、解析から除外した。予測された遺伝子のエクソン部位を中心に、プライマーセットをデザインし、塩基配列を決定した。

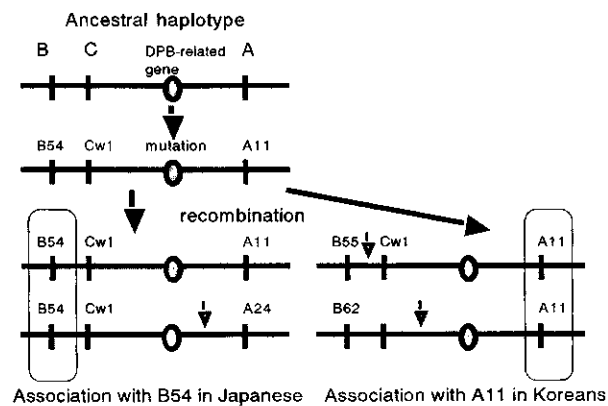


図 1 DPB 疾患感受性遺伝子変異と HLA との連鎖不平衡に関する仮説

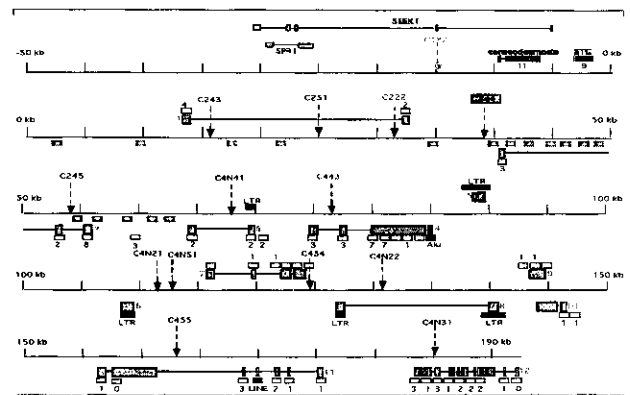
2. 解析部位の SNPs の同定

GenScan により予測されたエクソン部位を中心に PCR 増幅し、患者 4 例、対照 6 例につき、直接シーケンス法を行い、塩基配列を決定した。200 kb の領域の、のべ 20 kb にわたる塩基配列決定を、合計 10 検体 (20 染色体) で行い、塩基配列の差異が認められた場合、これらを SNP (もしくはその候補) として患者集団、対照集団全検体への解析

へと進んだ。その結果、現時点で、80 以上の SNPs が確認されている。

3. 候補領域内に同定された SNPs の密度、頻度、特色

解析された SNPs は、500 bp あたり、平均約 2 個存在した。多い領域では、500 bp あたり 8 個の SNPs が存在した。これらの SNPs の対立遺伝子型は、図 2 のような頻度分布を呈しており、比較的高頻度にみられる (対立遺伝子頻度 20% 以上の) SNPs が 50% 以上を占めた。塩基の置換については、ピリミジンがピリミジンあるいはプリンがプリンに置換されるトランジションがその他のトランスバージョンより、高頻度 (70% 以上) で認められた (図 3)。



注) 図は予測エクソン、□は解析した領域

図 2 GenScan による遺伝子予測と解析した領域

D. 考 案

ヒトゲノムのドラフト配列が発表されて、解説しうるヒトゲノムのほぼ全容が明らかになった⁵⁾。その過程で、国際レベルで 100 万個以上の SNP (の候補) が登録され、SNP を基本とした疾患感受性遺伝子の同定が行われつつある⁶⁾。一方、日本人の SNP データベースについては、東京大学医科学研究所ヒトゲノム解析センターと科学技術振興事業団との共同プロジェクトである JSNP プロジェクトがあり、平成 14 年 3 月公開時点で、当初の目的である 15 万 SNPs を越えるデータを登録した。しかしながら、DPB 感受性遺伝子となる HLA クラス I の候補領域は、もともと既知の遺伝子が知られていない空白の領域であったこともあり、遺伝子座 (予測も含む) を対象として、SNP を検索していく JSNP プロジェクトでは、依然として取り残された領域である。

表1 DPB 候補遺伝子領域に同定される SNP の出現頻度 (一般集団 200 アリルを対象)

対立遺伝子 頻度 (%)	SNP数 (個)
1-5%	12
5-10%	11
10-20%	15
20-30%	21
30-40%	8
40-50%	15
合計	82

表2 同定した一塩基置換の種類

トランジション		トランスバージョン			
A/G	C/T	A/C	A/T	G/C	G/T
28	35	4	2	7	5
63 (0.78)		18 (0.22)			

HLA クラス I 領域は、このドラフト配列に先立つ 1 年前に、完全配列として、共同研究施設である東海大学の猪子英俊教授のグループなどにより解読された⁷⁾。このような背景の下で、日韓の DPB と HLA class I 遺伝子型との関連分析の結果から推測される仮説である、HLA-A, B 間に DPB 発症に関連する遺伝子が存在し、HLA-B*54, A11 を保有する東アジア人の祖先染色体上にその遺伝子の変異が生じた可能性について、我々が検討した結果、HLA-B 座より 300 kb ほどテロメアよりの約 200 kb の領域に DPB 感受性遺伝子が存在する確率が高いことが明らかになった。これら一連の報告により、DPB は HLA 領域に疾患感受性遺伝子が存在する遺伝的要因を有する疾患として、遺伝的要因を有する疾患およびその責任遺伝子の国際データベースである “Mendelian Inheritance in Men” に登録された⁸⁾。

このような背景のもとで、新規遺伝子の同定と、領域内に広く分布する遺伝マーカー、特に SNPs による、さらなる候補領域の絞り込みにより、疾患感受性遺伝子を分子遺伝学的に同定することが求められ、今回の検討に至った。予測される遺伝子座を中心とした遺伝子変異の検討の結果、200 kb の領域の中には予想以上に多くの SNPs が高密度か

つ高頻度に存在することが明らかになった。これらの SNPs はコンピュータ上、エクソンと予測される領域を中心にスクリーニングしたため、単純に位置的マーカーとして候補領域の絞り込みに用いられるのみならず、それ自体が、コーディング領域に存在し、機能的意味合いを持っている可能性がある。これらの新規遺伝マーカーを用いた全検体についての完全タイピングが現在進行している。

E. 結 論

びまん性汎細気管支炎の HLA 関連疾患感受性遺伝子候補領域 200 kb 内に、80 個以上の SNP が同定された。このようにまとまった長さの遺伝子領域について、SNP とマイクロサテライトの連鎖不平衡の関係やハプロタイプ解析に関するデータはまだ国際的にも不十分であり、ヒトゲノムにおける連鎖不平衡の及ぶ範囲や、高頻度組換え点に関する構造解析は、現在分子遺伝学上、最もホットな領域の一つと考えられる。このようなデータの積み重ねが、多因子疾患の候補遺伝子のマッピングにも応用されていくものと予測される。

「謝 辞」

虎の門病院、天理よろづ相談所病院、日本医科大学第四内科の諸先生のご協力をいただきましたことに感謝いたします。

参考文献

- 1) Keicho N, Tokunaga K, Nakata K, Taguchi Y, Azuma A, Bannai M, Emi M, Ohishi N, Yazaki Y, Kudoh S. Contribution of HLA genes to genetic predisposition in diffuse panbronchiolitis. *Am J Respir Crit Care Med* 158: 846-850, 1998
- 2) Park MH, Kim YW, Yoon HI, Yoo C-G, Han SK, Shim Y-S, Kim WD. Association of HLA class I antigens with diffuse panbronchiolitis in Korean patients. *Am J Respir Crit Care Med* 159: 526-529, 1999
- 3) Keicho N, Ohashi J, Tamiya G, Nakata K, Taguchi Y, Azuma A, Ohishi N, Emi M, Park MH, Inoko H, Tokunaga K, Kudoh S. Fine localization of a major disease susceptibility locus for diffuse panbronchiolitis. *Am J Hum Genet* 66: 501-507, 2000

- 4) <http://genes.mit.edu/GENSCAN.html>
- 5) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921, 2001
- 6) The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933, 2001
- 7) The MHC sequencing consortium: Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923, 1999
- 8) <http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispim?604809>

F. 健康危険情報 (略)

G. 研究発表

1. 論文発表

- 1) Keicho, N., Kudoh, S. : Diffuse panbronchiolitis: role of macrolides in therapy. *Am J Respir Med*, 1 (2) :119-131, 2002.
- 2) Kanai, T., Fujii, T., Keicho, N., Tokunaga, K., Yamashita, T., Hyodo, H., Miki, A., Unno, N., Kozuma, S., Taketani, Y. : Polymorphism of human leukocyte antigen-E gene in the Japanese population with or without recurrent abortion. *Am J Reprod Immunol*, 45 (3) :168-173, 2001.
- 3) Azuma, A., Keicho, N., Furukawa, H., Yabe, T., Kudoh, S. : Prolonged survival of a bare lymphocyte syndrome type I patient with diffuse panbronchiolitis treated with erythromycin. *Sarcoidosis Vasc Diffuse Lung Dis*, 18 (3) :312-313, 2001.
- 4) Matsuzaka Y, Tounai K, Denda D, Tomizawa

M, Makino S, Okamoto K, Keicho N, Oka A, Kulski JK, Tamiya G, Inoko H. Identification of Novel Candidate Genes in the Diffuse Panbronchiolitis (DPB) Critical Region of the class I Human Major Histocompatibility Complex (MHC). *Immunogenetics*, in press.

- 5) 土屋朋子, 慶長直人: びまん性汎細気管支炎の疾患感受性遺伝子. *最新医学*, 56: 1680-1683, 2001.
- 6) 山中 剛, 慶長直人: びまん性汎細気管支炎. *臨床医増刊号*, 27: 993-997, 2001.
- 7) 川名明彦, 慶長直人: マクロライド耐性菌の現状と課題. *呼と循*, 49: 897-902, 2001.
- 8) 神尾孝一郎, 慶長直人: DPB 関連遺伝子異常. *分子呼吸器病*, 5: 531-535, 2001.
- 9) 慶長直人, 土屋朋子, 中田 光, 松下育美, 山中 剛, 神尾孝一郎: 呼吸器炎症性疾患の分子医学 8. びまん性汎細気管支炎のゲノム解析. *日本胸部臨床*, 60: S81-S84, 2001.
- 10) 慶長直人: びまん性汎細気管支炎: 人種特異性と疾患感受性遺伝子. *最新医学*, 56: 2580-2585, 2001.
- 11) 慶長直人, 松瀬 健: 慢性呼吸器疾患の抗菌薬治療—マクロライド系抗菌薬の使いかた—. *Medical Practice*, 18: 1351-1353, 2001.
- 12) 慶長直人, 土屋朋子, 中田 光: びまん性汎細気管支炎と HLA. *血液フロンティア*, 11: 39-46, 2001.

2. 学会発表

- 1) Keicho, N. Macrolide antibiotics and chronic airway inflammation. *European Respiratory Society, Annual Congress 2001, September 22-26, 2001, Berlin, Germany.*

H. 知的財産権 (略)

肺組織で発現する HLA クラス I 領域の新規遺伝子同定

田宮 元 東海大学医学部分子生命科学 2

研究要旨 びまん性汎細気管支炎(以下, DPB とする)は, 東アジアを中心に分布する慢性気管支疾患である。以前からこの疾患は HLA-B54 抗原との関連が報告されており, 最近, DPB 感受性領域が HLA クラス I 領域の S 遺伝子から TFIID 遺伝子の間の約 200kb の領域に位置することが明らかにされた。しかしながら, この DPB 感受性領域中には既知の発現遺伝子が見出されていなかった。そこでわれわれは, DPB の病因となる分子メカニズムを理解するために, 約 200kb にわたる DPB 感受性領域中に存在する感受性遺伝子を同定することを目的として, 同領域中の新規発現遺伝子の同定を試みた。

まず, EST データベースの検索を行ったところ, 8 個の転写単位が同領域中に予測され, すべての転写単位がヒト肺組織で発現していることが明らかになった。本年度は, このうち 2 つの遺伝子(C6orf37, C6orf37OS)について全長クローニングを行った。C6orf37 (Chromosome6 open reading frame 37) は, C 末端領域付近に 1 つの疎水性領域(アミノ酸 165-187)が存在しており, 膜タンパクであることが予想された。また, C6orf37 遺伝子座には, C6orf37 遺伝子と反対の転写方向を示し, 5 個のエクソンのうち 2 個が C6orf37 遺伝子とオーバーラップしている遺伝子が存在していた。この遺伝子には C6orf37OS (C6orf37 oppoite strand) と名付けた。C6orf37OS 転写産物には明確な読み枠はなく, C6orf37 遺伝子の発現を転写の段階で制御する機能的 RNA 分子である可能性がある。

また, EST が存在していない領域に, コード領域予測プログラムである GeneScan により発現遺伝子が予測され, これについても全長クローニングを行い, STG と名付けた。STG の予測 ORF は, 疎水性のアミノ酸を全体の 60% 以上含んでおり, 各種データベースに対しモチーフ検索を行ったが, 現在までのところ, 既知のモチーフとの相性は認められていない。以上の C6orf37, C6orf37OS, STG はいずれも肺での発現が確認されているため, DPB 感受性遺伝子である可能性があり, 今後 SNP の同定及び相関解析を予定している。

A. 研究目的

びまん性汎細気管支炎(以下 DPB)は, 副鼻腔気管支の感染および両側肺の微小結節性病巣を特徴とする慢性気道疾患である。この疾患は 1969 年に日本で最初に見出された疾患であり, その後韓国や中国においても DPB 患者が見出され, 現在では DPB 患者は東アジアを中心とした分布傾向を示している。このようにある疾患が地理的に特定の地域に限定されて分布することは大変まれであり, このような場合この地域に特有の環境あるいは遺伝因子がその疾患の発症に関与していることが想定される。

DPB 患者では, 以前から HLA-B54 抗原との関連が報告されており, 日本人健常者集団の HLA-B54

抗原頻度が 11.4% であるのに対して, DPB 患者集団では 63.2% と有意に高い頻度を示している。この HLA-B54 抗原は単一の対立遺伝子 B*5401 に対応しており, この対立遺伝子は日本や韓国, 中国といった東アジアにおいて主として認められる。このことは, 東アジアを中心とした分布傾向を示している DPB との関連からも大変興味深い事実である。

そのような背景のもとで, HLA 領域における DPB 感受性遺伝子を詳細にマップするために, 日本人 DPB 患者集団および日本人健常者集団においてマイクロサテライトマーカーを用いた遺伝学的相関解析が最近行われ, DPB 感受性領域が HLA クラス I 領域の S 遺伝子から TFIID 遺伝子の間, 約

200kbの領域に位置することが明らかにされた。しかしながら、このDPB感受性領域中には既知の発現遺伝子が見出されていなかった。そこでわれわれは、DPBの病因となる分子メカニズムを理解するために、約200kbにわたるDPB感受性領域中に存在する感受性遺伝子を同定することを目的として、同領域中の新規発現遺伝子の同定を試みた。

B. 研究方法

DPB感受性領域内におけるDPB候補遺伝子を検索するために、HLAクラスI領域におけるS遺伝子からTFIIH遺伝子間の約200kbのゲノム配列を用いて、ESTデータベース検索を行った。さらに、同様のゲノム配列を用いて、2種類のエクソン予測プログラム(Genescan, Grail)によるDPB候補遺伝子の検索を行った。次に、これらの方法により予測された転写単位の、肺における発現を調べるために、ヒト肺cDNAライブラリーを用いてPCRを行った。さらに、遺伝子の全長クローニングのためにヒト肺cDNAライブラリーを用いてスクリーニングを行って、cDNAクローンを単離した。シーケンシングにより塩基配列を決定し、そのcDNAの塩基配列および予想されるアミノ酸配列について、BLASTによるデータベース検索によって既知遺伝子とのホモロジーを検索した。更に、予想されるアミノ酸配列よりhydrophobicity plotによって疎水性領域の検索を行った。

遺伝子の発現様式を調べるために、RT-PCR法を用いてヒト25種類の組織(脳、心臓、精巣、胎盤、子宮頸部、脾臓、膵臓、肝臓、子宮、腎臓、胃、食道、小腸、大腸、直腸、肺、骨格筋、皮膚、ケラチノサイト、皮膚ファイブロブラスト、胎児心臓、胎児腎臓、胎児皮膚、胎児肺、腫瘍肺)におけるRT-PCRを行った。また、オーバーラップ遺伝子の重複領域における転写方向を確認するために、片側鎖特異的RT-PCRを行った。

C. 研究成果

・DPB感受性領域における転写地図

DPB候補遺伝子を同定する第1のステップとして、予想される転写単位の遺伝子座が示されている物理地図を作製した(図1)。ESTデータベース検索の結果、8個の転写単位がHLAクラスI領域のゲノム配列と一致しており、すべての転写単位がヒト肺組織で発現していることが分かった(表1、図1)。

・C6orf37遺伝子のクローニング

ESTデータベース検索より得られた転写単位のうち、1つの転写単位(ha75f09.x1)について全長クローニングを行った。この遺伝子は235個のアミノ酸をコードする2個のエクソンから構成されており、C6orf37(Chromosome6 open reading frame 37)遺伝子と名付けた。エクソン1は予測される開始コドン(ATG)が含まれており、エクソン2には予測される終止コドン(TGA)が含まれている。また、エクソン1は2つのコンピュータプログラムによって予測されたエクソンであるが、エクソン2はこの2つのコンピュータプログラムによって予測されなかった。Kyte-Doolittleの疎水性プロット

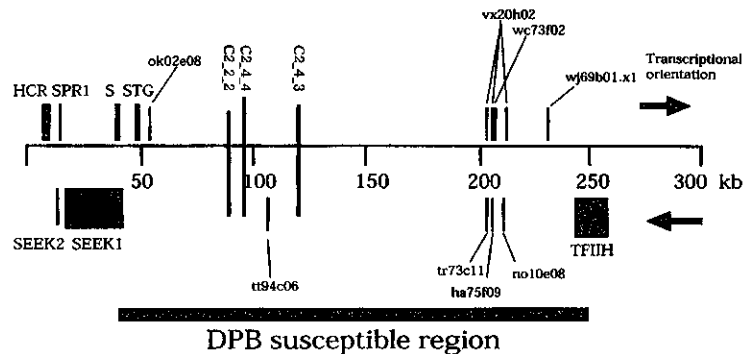


図1

表1

ESTs	location	Accession No.	Sense oligonucleotide	Antisense oligonucleotide	PCR産物の長さ (bp)	該での発現
ok02e08.s1	*Tel 7kb/S	AA905678	GACAAACGCACTGCATTAGGATC	GATTTCTCTACCTAGGAGGGC	269	-
			CACTGCATTAGGATCTATGCCCTAC	CTAGGAGGGCGCTTATCCAG	249	-
			GATCTATCCCTTACGATGGCTCA	GGGGCGTATTCCGAAGTTAGGA	232	+
			CATCCCGATGCTATGSACTAC	GGATTGTTCAGAGSAAAGTAACC	284	+
			GCCATBTCTATGCGTACGCTACAC	AGSAGTAAAGCAAGGTCAAGCA	270	+
			CACTACAGTACAGCATAAAGAGGACC	CAAGSTCAGCGAAGATTAGGAG	248	+
			AGTGGACSCCGTCTGG	AACACACACAGAGAAATGCTAC	273	-
			CCGTGTGGTTGGCTTGG	AGATGAATGGCTACAGGAGG	255	-
			TGGCTTCTGGGTTGTG	CTACAAGGAGGAAAGCTTGAATCTG	234	+
			ACAGGAAGGGGAGGAATACTG	CATTCGCTAATAGGCTCTCTG	259	+
			GCAGAGAATACTGACGGTTACCCAG	CTCTGTCTCTCTTAGGACTC	236	+
			GGTACCAGTATTAACCCCTTACTGTG	TTAGSACTCAGTCCAGGTCTCG	207	+
			GCTGCTGTTTTCAAGAGGTG	ATTCTGGATGGAGTTACGTTTAGG	497	+
			TTCAAGAAGGTGGCCACAG	AGTACGTTTAGGAGCCTGGC	477	+
			TGGCCACAGCAGGTGAG	GAGCCTGGCAGCAACTG	455	+
			GACTATAAAGCTTCACTCAAGTCCAC	AATGAATATTTCTCTCCATCTG	240	+
			CTTCAGTCAAGGCTCAGAGGAGAC	TTCTCTCCATGCTCTCCAGC	229	+
			TGACAGSAGSAAATCTCTACTACTAC	ATCTGTCCAGGCTGGATACC	169	+
			ACAGCCATCTAAGGCAAC	TGTTTTTTCCAGCGGATTC	203	+
			ATTCTAAGCCCAACTGADAASC	TTCCAGCGGATCTTATGTG	190	-
			TGCAAGGCTCAGGCAACTC	GGATTCTTATGTGGTATGGAG	169	-
			TGAACACCATGTTGTCTGC	TTATTGCCCTGAGGAG	138	-
			AACACCATGTGTCTGCAAC	TTGCCGCTCAGGAGG	133	+
			CACCATGTTGTCTGCAACAC	CCCTCAGGAGGCTGGC	127	+

*Tel: Telomeric

**Cen: Centromeric

***C2_4_4 and C4_4_3 HLA class I 領域における DPB 感受性領域 200 kb 内のマイクロサテライト

により C6orf37 タンパク質の C 末端領域付近には 1 つの疎水性領域(アミノ酸 165-187)が存在しており、このタンパク質が膜タンパクであることが予測された。また、C6orf37 タンパク質の N 末端領域はヒト zonadhesin タンパク質のムチン用ドメインとホモロジーが示された。

• C6orf37OS 遺伝子のクローニング

C6orf 遺伝子のスクリーニングの過程で、C6orf37 遺伝子と一部オーバーラップする形で、反対の転写方向を示す遺伝子を単離し、この遺伝子を C6orf37OS (C6orf37 oppoite strand) と名付けた。C6orf37OS 遺伝子は 5 個のエクソンから構成されており、このうち 2 個のエクソン (エクソン 3, 4) が C6orf37 遺伝子とオーバーラップしている。また、エクソン 3 を欠く選択的スプライシングフォームも単離されている (図 2)。しかしながら、この 2 つの C6orf37OS 転写産物は、100 個以上のアミノ酸配列からなる明確な読み枠が見出されておらず、さらにこの遺伝子周辺には Alu や LINE といった繰り返し配列は見出されなかった。

• C6orf37 遺伝子および C6orf37OS 遺伝子の発現解析

この 2 つの遺伝子の発現様式を調べるために、

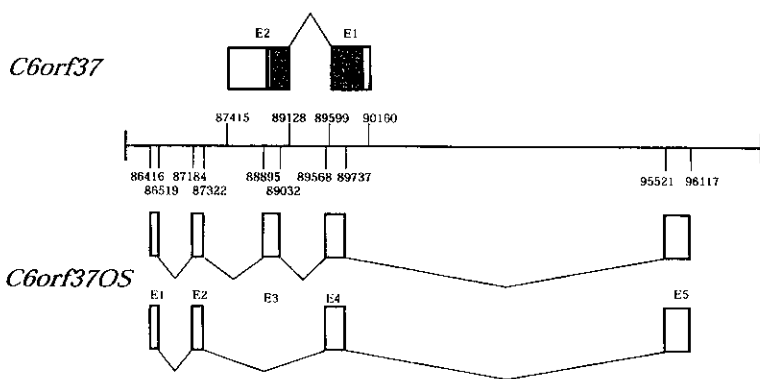


図 2

25 種類のヒト各組織の poly(A) + RNA を用いて、RT-PCR (reverse transcriptase-polymerase chain reaction) を行った。RT-PCR に用いたプライマーはゲノム DNA の混入がないことを確認するために、イントロンを挟む形でエクソン 1 およびエクソン 2 に設計した。その結果、2 種類のサイズの増幅産物が得られた。長い DNA 断片 (515 bp) は肺の他 7 種類の組織で発現していた。一方、短い DNA 断片 (451bp) は精巣、腎臓、子宮、小腸で発現していた (図 3)。2 つの DNA 断片の塩基配列を決定したところ、長い DNA 断片は C6orf37 遺伝子に対応し、短い DNA 断片は C6orf37OS 遺伝子に対応することが明らかとなった。

• STG 遺伝子

S 遺伝子の約 2.5kb セントロメア側に GenScan にて予測された遺伝子が見出された。この配列と一致する既知遺伝子や EST は全く存在しなかった。そこで、この予測された遺伝子配列からプライマーをデザインし RT-PCR を行った。その結果、増幅が確認され発現遺伝子の存在が示唆された。さらに、cDNA ライブラリーのスクリーニングおよび 5' & 3' RACE 法により全長のクローニングを行った。STG と名付けたこの遺伝子は cDNA

全長で 1,160bp、予測されるアミノ酸は 325aa であり、43bp の第 1 エクソン、1,117bp の第 2 エクソンから構成されていた。アミノ酸の組成はプロリンが最も多く、疎水性のアミノ酸が 60% を越える。このアミノ酸および DNA 配列を用いて各種データベースに対しモチーフ検索を行ったが、既知のモチーフとの相同性は認められなかった。3' -UTR にはポリ A 付加シグナル、5' の上流には Initiator および GC-

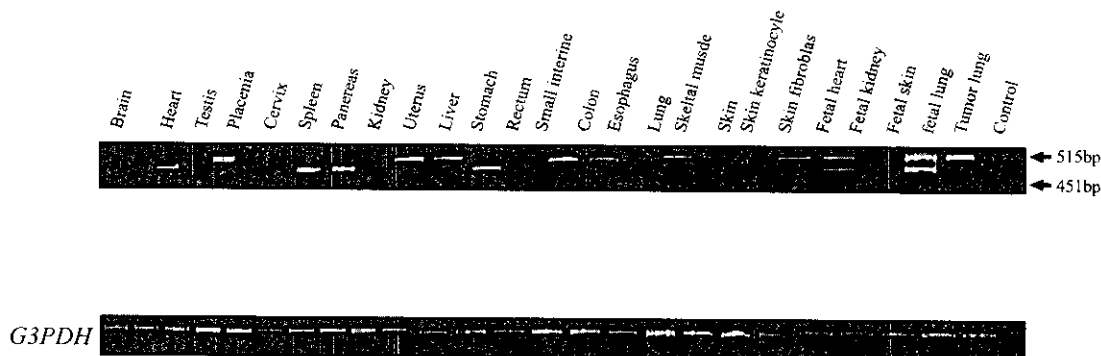


図 3

BOX 配列が認められた。STG 遺伝子の発現を各種組織由来の RNA を用いて、RT-PCR を行った。その結果、興味深いことに、健常肺、胎児肺では発現がみられないが、腫瘍の肺で発現していることが明らかとなった。このことから、この STG 遺伝子が DPB の感受性遺伝子の候補として疑われた。

D. 考 察

本研究で単離された C6orf37 遺伝子は、DPB 感受性領域に位置しており、さらに肺で発現していることから DPB 感受性遺伝子の 1 つである可能性が考えられる。また、C6orf37 タンパク質は N 末端領域に 1 つの疎水性領域をもつことから膜タンパク質であると予想される。さらに、C6orf37 タンパク質の N 末端とホモロジーを示すヒト zonadhesin タンパク質のムチン用ドメインは、細胞-細胞、細胞-マトリックス、タンパク質-タンパク質間の相互結合に関与することが知られており、C6orf37 タンパク質も細胞間の結合に関与する膜タンパクとして機能していると考えられる。

また、C6orf37OS 転写産物には明確な読み枠はなく、C6orf37 遺伝子と転写方向の異なるオーバーラップ遺伝子であることから、C6orf37 遺伝子の発現を転写の段階で制御する機能的 RNA 分子である可能性がある。

以上の C6orf37、C6orf37OS に STG も含めて、いずれも肺 (STG のみ腫瘍肺) での発現が確認されていることから、DPB 感受性遺伝子である可能性があり、今後 SNP の同定及び相関解析を予定している。

また、今回、DPB 感受性領域より見出した他の転写単位も、いずれも肺で発現していることから、感受性候補遺伝子として、今後全長クローニングを予定している。

E. 結 論

本研究において、HLA クラス I 領域における DPB 感受性領域中に存在し、これまで報告されていない 3 つの新規遺伝子 (C6orf37、C6orf37OS、STG) の全長配列の単離に成功した。これらは、いずれも肺 (STG のみ腫瘍肺) での発現が確認されていることから、DPB 感受性遺伝子である可能性がある。

F. 健康危険情報 (略)

G. 研究発表

(1) 論文発表 (MedLine 形式で、分担研究者の名前のあるもの、in press まで)

1) Katoh K, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G. Genetic Isolates in East Asia: A Study of Linkage Disequilibrium in the X chromosome. *Am J Hum Genet.* in press.

2) Matsuzaka Y, Tounai K, Denda D, Tomizawa M, Makino S, Okamoto K, Keicho N, Oka A, Kulski JK, Tamiya G, Inoko H. Identification of Novel Candidate Genes in the Diffuse Panbronchiolitis (DPB) Critical Region of the class I Human Major Histocompatibility Complex (MHC). *Immunogenetics*, in press.

3) Nagayasu E, Nagakura K, Akaki M, Tamiya G, Makino S, Nakano Y, Kimura M, Aikawa M. Association of a determinant on mouse chromosome 18 with experimental severe *Plasmodium berghei* malaria. *Infect Immun.* 2002 Feb; 70 (2) : 512-6.

4) Matsuzaka Y, Makino S, Nakajima K, Tomizawa M, Oka A, Bahram S, Kulski JK, Tamiya G, Inoko H. New polymorphic microsatellite markers in the human MHC class III region. *Tissue Antigens.* 2001 May; 57 (5) : 397-404.

5) Shiina T, Ando A, Suto Y, Kasai F, Shigenari A, Takishima N, Kikkawa E, Iwata K, Kuwano Y, Kitamura Y, Matsuzawa Y, Sano K, Nogami M, Kawata H, Li S, Fukuzumi Y, Yamazaki M, Tashiro H, Tamiya G, Kohda A, Okumura K, Ikemura T, Soeda E, Mizuki N, Kimura M, Bahram S, Inoko H. Genomic anatomy of a premier major histocompatibility complex paralogous region on chromosome 1q21-q22. *Genome Res.* 2001 May; 11 (5) : 789-802.

(2) 学会発表

1) ヒト第6染色体HLAクラスI領域における新規遺伝子の単離および発現解析, 傳田晃弘, 東内健一, 松坂恭成, 岡 晃, 富澤麻衣子, 牧野悟士, 田宮 元, 猪子英俊, 東海大学医学部分子生命科学 2, 第 23 回日本分子生物学会,

H. 知的財産権の出願 (略)

分子遺伝学上の統計解析

大橋 順 東京大学医学部人類遺伝学教室

研究要旨 びまん性汎細気管支炎のような多因子疾患の疾患感受性遺伝子を探索するために、「(1) 研究デザインの新規統計的評価法の開発」、および「(2) 大規模 SNP タイピングデータ解析のためのアルゴリズムの開発」を行った。これらは、本研究目的である「びまん性汎細気管支炎等の疾患感受性遺伝子」を同定する上で有用なだけでなく、他の疾患にも応用可能な手法である。以下にこれら二つの研究内容の要旨を述べる。

(1) 研究デザインの新規統計的評価法の開発

疾患感受性遺伝子を同定するために、もっともよく利用されている研究デザインが患者-対照研究である。これは、ある集団中から多数の患者と非患者(対照)とをサンプルし、着目する対立遺伝子の頻度を患者サンプルと対照サンプルとで比較する方法である。患者-対照研究を行う際に重要な事項は、適切な研究デザインの設定と研究の検出力の事前評価である。本研究では、感受性遺伝子探索研究を行う際の「(i) 適切な研究デザイン(サンプルサイズ)の設定」と、多数の多型マーカーを利用する「(ii) ゲノムワイド連鎖不平衡検定における研究の検出力の事前評価」に関して考察した。

(i) 適切な研究デザイン(サンプルサイズ)の設定

患者-対照研究によって候補遺伝子アプローチを行う際に必要なサンプルサイズを計算するアルゴリズムの開発を行った。また、本研究の成果に基づいて、感受性遺伝子を同定するのに必要なサンプルサイズを算定し、各種疾患における感受性遺伝子探索研究を行った。

(ii) ゲノムワイド連鎖不平衡検定における研究の検出力の事前評価

多因子疾患の感受性遺伝子同定を目的とするゲノムワイド連鎖不平衡検定によって、想定する感受性遺伝子を同定しうる確率(検出力)を概算するアルゴリズムの開発を行った。この成果により、研究デザインの有効性を事前評価できるばかりでなく、ある候補領域中に、どの程度の頻度を有する幾つのマーカー(SNPやマイクロサテライトマーカー)を設定すればよいかなどの判断が可能となった。

(2) 大規模 SNP タイピングデータ解析のためのアルゴリズムの開発

2SNP ペアワイズハプロタイプの集団内頻度を個人タイピングデータから計算するプログラムの開発を行った。具体的には、以下のような処理を行うプログラムである。各個体のタイピング結果を Excel の sheet ファイルに記録する。同一染色体上の m 個の SNP をタイピングすれば、 $m(m-1)/2$ 組の組み合わせを考慮することになる。プログラムは、上記のデータから $m(m-1)/2$ 組の以下のような表を作成する。

		SNP2			
		CC	CA	AA	Total
SNP1	AA	10	15	2	27
	AG	18	30	5	53
	GG	2	24	10	36
	Total	30	69	17	116

この表をもとに、EM アルゴリズムによって4種類ある(AC, AA, GC, GA)ハプロタイプの頻度を推定し、推定されたハプロタイプ頻度をもとに、各種の尺度(連鎖不平衡の指標)を計算する。現在は、case-control データに対してハプロタイプ頻度を推定し、両者のハプロタイプ頻度に統計学上有意な差が存在するかどうかを検定するプログラムの開発を試みている。

A. 研究目的

びまん性汎細気管支炎のような多因子疾患の疾患感受性遺伝子を探索するために、(1) 研究デザインの新規統計的評価法の開発、および(2) 大規模 SNP タイピングデータ解析のためのアルゴリズムの開発が本研究の主要な目的である。

B. 研究方法

(1) 研究デザインの新規統計的評価法の開発

疾患感受性遺伝子を同定するために、もっともよく利用されている研究デザインが患者-対照研究である。これは、ある集団中から多数の患者と非患者(対照)とをサンプルし、着目する対立遺伝子の頻度を患者サンプルと対照サンプルとで比較する方法である。患者-対照研究を行なう際に重要な事項は、適切な研究デザインを設定すること、研究の検出力の事前評価である。本研究では、感受性遺伝子の評価を行う際の「(i) 適切な研究デザイン(サンプルサイズ)の設定」と、多くの多型マーカーを利用する「(ii) ゲノムワイド連鎖不平衡検定における研究の検出力の事前評価」に関して考察した。

(i) 適切な研究デザイン(サンプルサイズ)の設定

二種類の対立遺伝子 D と d を有する遺伝子座に対して、割合の差の検定を利用して D と d の対立遺伝子頻度を、患者群と対照群とで調べるとする。患者を n 個体サンプルした場合に、遺伝子型 DD, 遺伝子型 Dd, 遺伝子型 dd の個体数が n_2, n_1, n_0 であり ($n = n_2 + n_1 + n_0$), 対照を m 個体サンプルした場合に、遺伝子型 DD, 遺伝子型 Dd, 遺伝子型 dd の個体数が m_2, m_1, m_0 であったとすると ($m = m_2 + m_1 + m_0$), 患者群中の D 対立遺伝子の推定遺伝子頻度 \hat{p}_1 は $\frac{2n_2 + n_1}{2n}$, 対照群中の D 対立遺伝子の推定遺伝子頻度 \hat{p}_2 は $\frac{2m_2 + m_1}{2m}$ である(表 1 と表 2)。

表 1 遺伝子型頻度表

	患者群	対照群
DD	n_2	m_2
Dd	n_1	m_1
dd	n_0	m_0
計	n	m

表 2 遺伝子頻度表

	患者群	対照群
D	$2n_2 + n_1$	$2m_2 + m_1$
d	$n_2 + 2n_0$	$m_2 + 2m_0$
計	$2n$	$2m$

帰無仮説 H_0 は「2 群の遺伝子頻度 P_1, P_2 には差が無い ($P_1 = P_2$)」とし、対立仮説 H_1 は「患者群の遺伝子頻度 P_1 は対照群の P_2 より大きい ($P_1 > P_2$)」とする。 P_1 が P_2 より増加しているかどうかを検討するので、このような検定は「片側検定」とよばれる。 H_0 より、平均の推定遺伝子頻度 $\hat{p} = \frac{2n_2 + n_1 + 2m_2 + m_1}{2n + 2m}$ が集団全体での遺伝子頻度 P ($= P_1 = P_2$) の推定値であると仮定する。いま、データ数が大きく、 \hat{p} があまり小さくないとき、 H_0 が正しいとすると、 $\hat{p}_1 - \hat{p}_2$ は平均 0, 標準偏差 $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{2n} + \frac{1}{2m} \right)}$ の正規分布に従う。したがって、2 群の遺伝子頻度の差が $\hat{p}_1 - \hat{p}_2$ となる確率は、それを次式により標準化して(これを平均 0, 標準偏差 1 となるように調整して)、標準正規分布表から求めることができる。

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{2n} + \frac{1}{2m} \right)}}$$

P_1 が P_2 より大きいほど、 Z は 0 より大きくなる。 z_α は標準正規分布の右片側 α の確率を与える Z の値であり、 H_0 の下で Z が z_α を越える確率が α である Z の値のことである。有意水準 α を 0.05 に設定して片側検定を 1 回行うとき、 $z_\alpha = z_{0.05} = 1.64$ よりも Z が大きければ、 H_0 の下で確率 0.05 以下でしか生じない稀なことが起きたことを示している。そこで、稀なことが起きたのではなく、 H_0 が間違っていると判断し、有意水準 0.05 で帰無仮説 H_0 を棄却し、対立仮説 H_1 を採択する。すなわち、D は疾患と関連としていると結論する。しかし、D と d のどちらが疾患感受性対立遺伝子であるかを、研究以前に通常は予想できないので、両方の対立遺伝子に対して片側検定が 2 回行なわれると考えるべきである。すなわち、患者群中の d 対立遺伝子の推定遺伝子頻度 \hat{p} を $\frac{n_1 + 2n_0}{2n}$, 対照群中の d 対立遺伝

子の推定遺伝子頻度 \hat{p}_2 を $\frac{m_1 + 2m_0}{2m}$, 平均の推定遺伝

子頻度は $\hat{p} = \frac{n_1 + 2n_0 + m_1 + 2m_0}{2n + 2m}$ として (表1 と表2),

再度検定することになる. もし, 有意水準を α として2度検定を行なえば, 第一種の過誤が起こる確率は 2α となってしまう. そのため, 1つの遺伝子座の2つの対立遺伝子を調べる研究では, どちらかの対立遺伝子について計算した Z の値が $z_{\alpha/2}$ を越えた場合に, 有意水準 α で当該対立遺伝子が疾患と関連すると結論する.

・サンプルサイズの決定

H_0 の下で期待される $\hat{p}_1 - \hat{p}_2$ を考える. D の集団内頻度を p , d の集団内遺伝子頻度を $1-p$ とし, これらの接合体頻度は集団全体でハーディ・ワインバーク平衡状態にあると仮定する (すなわち, DD の頻度は p^2 , Dd の頻度は $2p(1-p)$, dd の頻度は $(1-p)^2$ である). 患者を n 個体, 対照を m 個体サンプルしたとすると, H_0 の下で $\hat{p}_1 - \hat{p}_2$ は, 平均0,

標準偏差 $\sigma_{(m)} = \sqrt{p(1-p) \left(\frac{1}{2n} + \frac{1}{2m} \right)}$ の正規分布に従う.

次に, H_1 の下で期待される $\hat{p}_1 - \hat{p}_2$ を考える. D を感受性対立遺伝子とする. 遺伝子型 dd の個体の浸透率 (患者になる確率) を f_0 とし (complex disease では一般に $f_0 > 0$ である), 遺伝子型 DD の個体および遺伝子型 Dd の個体の浸透率をそれぞれ f_2 および f_1 とおく (一般に $1 > f_2 \geq f_1 \geq f_0 > 0$ である).

集団の罹患率 K は $K = p^2 f_2 + 2p(1-p)f_1 + (1-p)^2 f_0$ であらわされる. 患者群の中から一個体を選び, その遺伝子型を調べたときに, その個体が DD , Dd , dd

である確率は, それぞれ, $\frac{p^2 f_2}{K}$, $\frac{2p(1-p)f_1}{K}$, $\frac{(1-p)^2 f_0}{K}$

である. また, 対照群の中から一個体を選び, その遺伝子型を調べたときに, その個体が DD ,

Dd , dd である確率は, それぞれ, $\frac{p^2(1-f_2)}{1-K}$,

$\frac{2p(1-p)(1-f_1)}{1-K}$, $\frac{(1-p)^2(1-f_0)}{1-K}$ である. H_1 の下で,

患者群中での D の頻度が $P_{1(a)} = \frac{p^2 f_2 + p(1-p)f_1}{K}$, 対

照群中での D の頻度が $P_{2(a)} = \frac{p^2(1-f_2) + p(1-p)(1-f_1)}{1-K}$

となることに注意して, H_1 の下で $\hat{p}_1 - \hat{p}_2$ は, 平均

$$P_{1(a)} - P_{2(a)}, \text{ 標準偏差 } \sigma_{(a)} = \sqrt{\frac{P_{1(a)}(1-P_{1(a)})}{2n} + \frac{P_{2(a)}(1-P_{2(a)})}{2m}}$$

の正規分布に従う. 第一種の過誤 (有意水準) α と第2種の過誤 β に対して, $z_{\alpha} \sigma_{(m)} = P_{1(a)} - P_{2(a)} + z_{1-\beta} \sigma_{(a)}$ なる関係式が成り立つ. 有意水準 α で片側検定を行った場合に, $1-\beta$ の検出力を達成するのに必要なサンプルサイズ n は, $m = kn$ とおけば, 以下の式で求められる.

$$n = \frac{1}{2} \left[\frac{z_{\alpha} \sqrt{p(1-p) \left(1 + \frac{1}{k} \right)} - z_{1-\beta} \sqrt{\frac{P_{1(a)}(1-P_{1(a)}) + \frac{P_{2(a)}(1-P_{2(a)})}{k}}{P_{1(a)} - P_{2(a)}}} \right]^2$$

D と d の両方を検定するので, 有意水準を 0.05 に設定するには, $z_{\alpha} = z_{0.05/2} = 1.96$, 検出力を 0.8 に設定するのであれば $z_{1-\beta} = z_{0.8} = -0.84$ ($z_{0.2} = 0.84$ であることに注意されたい) を採用する. サンプルサイズが決まっている場合の研究の検出力は, $z_{1-\beta} = |z_{\alpha} \sigma_{(m)} - (P_{1(a)} - P_{2(a)})| / \sigma_{(a)}$ であらわされる.

(ii) ゲノムワイド連鎖不平衡検定における研究の検出力の事前評価

浸透率の低い多因子疾患の感受性遺伝子を同定する方法の一つに, ゲノムワイドに多数の多型マーカーを調べる連鎖不平衡検定法がある. 患者-対照関連研究における連鎖不平衡検定法とは, ある多型マーカーと疾患感受性遺伝子とが連鎖不平衡状態にある場合には, 患者群と対照群の当該マーカー対立遺伝子頻度に差が生じるという原理に基づいている. 連鎖不平衡の強さは, 世代の経過とともに減衰し, その減衰率はマーカーと疾患感受性遺伝子座との間の遺伝距離 (組換え率) に比例するため, できるだけ多くのマーカーを設定して (疾患遺伝子との期待遺伝距離を小さくして) 研究を行う必要がある. しかし, 多数のマーカーについて統計検定を繰り返せば多くの偽陽性が起こるため, 検定した回数で有意水準を調整する Bonferroni の補正が一般的に行われる. Bonferroni の補正を行うと, マーカー数を増やせば必ずしも連鎖不平衡 (マーカーと疾患との関連) が検出されとは限らず, 調べるべき最適なマーカー数 (最適なマーカー密度) が存在する. また, ゲノムワイドに行う連鎖不平衡検定法では, SNP マーカーやマイクロサテライトマーカーが用いられる. こららの要因を考慮に入れ, 決定論的方程式を用いて世代

の経過に伴う連鎖不平衡の減衰をモデル化し、疾患感受性遺伝子とマーカーとの間の連鎖不平衡が、カイ2乗検定によって検出される確率、すなわちゲノムワイド連鎖不平衡検定の検出力の計算を行った。

・遺伝学的モデルの構築

疾患感受性遺伝子座には、高リスク対立遺伝子 D と低リスク対立遺伝子 d の2個が存在し、それぞれの対立遺伝子頻度を p と $1-p$ とおく。マーカー遺伝子座には m 個の対立遺伝子 (i 番目の対立遺伝子を M_i と呼ぶ) が存在し、 M_i の頻度を q_i とおく。

ここで、 $\sum_{i=1}^m q_i = 1$ が成立する。集団はハーディー・ワインバーグ平衡状態にあり、これらの対立遺伝子頻度は世代の経過により変化しないものとする。疾患感受性遺伝子座とマーカー遺伝子座との間の組換え率は θ とする。世代 t において D と M_i から構成されるハプロタイプ頻度を $H_{Di}(t)$ とおく。同様に、世代 t において d と M_i から構成されるハプロタイプ頻度を $H_{di}(t)$ とおく。したがって、 $q_i = H_{Di}(t) + H_{di}(t)$ が常に成立する。世代当たりの突然変異率は、マイクロサテライトマーカーの場合が $10^{-2} \sim 10^{-5}$ であり、SNP の場合が $10^{-7} \sim 10^{-8}$ と考えられるため、本研究ではマイクロサテライトマーカーについてのみ突然変異を考慮する。なお、疾患感受性遺伝子座での突然変異は考慮しない。マイクロサテライトマーカーの突然変異に関しては1段階ステップ突然変異モデル (SSM) を採用し、 M_i は、次に大きい M_{i-1} 、または次に小さい M_{i+1} にそれぞれ確率 u_i で突然変異すると仮定する (図1)。最も小さな M_1 は、確率 u_1 で一つ大きな M_2 にのみ突然変異し、最も大きな M_m は、確率 u_m で一つ小さな M_{m-1} にのみ突然変異する。

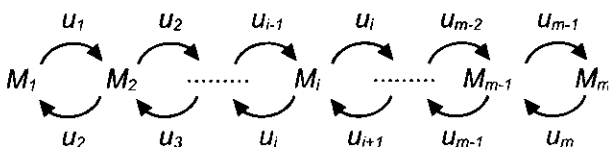


図1 1段階ステップ突然変異モデル (SSM)

以上の仮定より、ハプロタイプ頻度変化に関する次の決定論的方程式 (頻度変化の漸化式) がえられる。

$$\begin{aligned} H_{D1}(t+1) &= (1-\theta-u_1)H_{D1}(t)+u_2H_{D2}(t)+\theta pq_1, \\ H_{D2}(t+1) &= (1-\theta-2u_2)H_{D2}(t)+u_{1+}H_{D1}(t)+u_{2+}H_{D3}(t)+\theta pq_2, \\ &\quad (i=2,\dots,m-1), \\ H_{Dm}(t+1) &= (1-\theta-u_m)H_{Dm}(t)+u_{m-1}H_{Dm-1}(t)+\theta pq_m, \\ H_{d1}(t+1) &= (1-\theta-u_1)H_{d1}(t)+u_2H_{d2}(t)+\theta(1-p)q_1, \\ H_{di}(t+1) &= (1-\theta-2u_i)H_{di}(t)+u_{i-1}H_{d,i-1}(t)+u_{i+1}H_{d,i+1}(t)+\theta(1-p)q_i, \\ &\quad (i=2,\dots,m-1), \\ H_{dm}(t+1) &= (1-\theta-u_m)H_{dm}(t)+u_{m-1}H_{dm-1}(t)+\theta(1-p)q_m. \end{aligned}$$

したがって、初期頻度 ($t=0$) さえ与えれば、任意の世代におけるハプロタイプ頻度を計算することが可能である。SNP マーカー ($m=2$ および $u_i=0$ が成立) に対しては、上記の方程式は以下のように解くことができる。

$$\begin{aligned} H_{D1}(t) &= (H_{D1}(0)-pq_1)(1-\theta)^t + pq_1, \\ H_{D2}(t) &= (H_{D2}(0)-pq_2)(1-\theta)^t + pq_2 = p - H_{D1}(t), \\ H_{d1}(t) &= (H_{d1}(0)-(1-p)q_1)(1-\theta)^t + (1-p)q_1, \\ H_{d2}(t) &= (H_{d2}(0)-(1-p)q_2)(1-\theta)^t + (1-p)q_2 = 1-p - H_{d1}(t). \end{aligned}$$

ここで $q_1+q_2=1$ である。

マーカー遺伝子座における対立遺伝子の平衡頻度分布

マーカー遺伝子座における対立遺伝子頻度はマーカー毎に大きく異なるが、3734 個の (CA)_n リピートマーカーの対立遺伝子頻度分布から、平均的な対立遺伝子頻度分布 (これを平衡頻度分布とみなす) が明らかにされている (<http://www.well.ox.ac.uk/~mfarrall/microsatellite.html> においてデータが公開されている)。1段階ステップ突然変異モデルにおいて、マイクロサテライトマーカーの対立遺伝子数および対立遺伝子頻度は変化しないように仮定するため、対立遺伝子の平衡頻度分布と突然変異率との間に以下のような関係式が成立する必要がある。

$$\bar{q}_1 u_1 + \bar{q}_m u_m + \sum_{i=2}^{m-1} 2\bar{q}_i u_i = u.$$

ここで、 M_i の平衡頻度分布は \bar{q}_i 、マーカー遺伝子座における総突然変異率は u である。これより、

$$M_i \text{ の突然変異率は } u_i = \frac{u}{2(m-1)\bar{q}_i} \text{ と設定される。}$$

(CA)_n リピートマーカーの中で最も多く観察されるものは対立遺伝子を6個含むものなので、本研

究では、マイクロサテライトマーカーは6個の対立遺伝子を有し、それぞれの対立遺伝子平衡頻度は、 $\bar{q}_1=0.160$, $\bar{q}_2=0.199$, $\bar{q}_3=0.214$, $\bar{q}_4=0.205$, $\bar{q}_5=0.158$, $\bar{q}_6=0.064$ とする。SNPマーカーは淘汰上中立であり、SNPマーカーでは再突然変異が起こらないと仮定すると、あるSNP対立遺伝子の頻度が q から $q+dq$ の間にある確率を、 $C_A\left(\frac{1}{q}+\frac{1}{1-q}\right)dq$ で表すことができる。ここで、 C_A は規格化定数であり、 $C_A^{-1}=\int_0^1\left(\frac{1}{q}+\frac{1}{1-q}\right)dq$ である。これをSNPマーカーでの平衡頻度分布とみなす。ゲノムワイドな連鎖不平衡検定研究では、マイナー対立遺伝子頻度が高いSNPが使用されるので、その頻度のカットオフ値を Q とおく。 Q 以上の対立遺伝子頻度を有するSNPがマーカーとしてランダムに選ばれりと仮定すると、あるSNPマーカーの対立遺伝子頻度が q から $q+dq$ の間にある確率は、 $C_B\left(\frac{1}{q}+\frac{1}{1-q}\right)dq$ で表すことができる。ここで、 $C_B^{-1}=\int_Q^1\left(\frac{1}{q}+\frac{1}{1-q}\right)dq$ である。

・疾患モデル

患者-対照研究により、疾患遺伝子探索を目的とする連鎖不平衡検定を行うとする。遺伝子型 DD , Dd , dd に対する浸透率を f_2 , f_1 , f_0 とおく(一般的に、 $1 \geq f_2 \geq f_1 \geq f_0 \geq 0$ が成立する)。患者または対照という表現型が与えられた条件下での各遺伝子型頻度は、

$$\begin{aligned} P(DD|case) &= p^2 f_2 / e, \\ P(Dd|case) &= 2p(1-p) f_1 / e, \\ P(dd|case) &= (1-p)^2 f_0 / e, \\ P(DD|control) &= p^2 (1-f_2) / (1-e), \\ P(Dd|control) &= 2p(1-p)(1-f_1) / (1-e), \\ P(dd|control) &= (1-p)^2 (1-f_0) / (1-e) \end{aligned}$$

で与えられる。ここで、 e は疾患の罹患率 $p^2 f_2 + 2p(1-p) f_1 + (1-p)^2 f_0$ を示している。ベイズの定理より、世代 t において患者が $M_i M_j$ という遺伝子型をとる確率は、

$$P(M_i M_j | case, t) = \frac{1}{e} \{ f_2 H_{D_i}(t) H_{D_j}(t) + 2 f_1 H_{D_i}(t) H_{d_i}(t) + f_0 H_{d_i}(t) H_{d_j}(t) \}$$

世代 t において患者が $M_i M_j$ (ここで、 $i \neq j$) という

遺伝子型をとる確率は

$$P(M_i M_j | case, t) = \frac{1}{e} \{ 2 f_2 H_{D_i}(t) H_{D_j}(t) + 2 f_1 (H_{D_i}(t) H_{d_j}(t) + H_{d_i}(t) H_{D_j}(t)) + 2 f_0 H_{d_i}(t) H_{d_j}(t) \}$$

である。同様に、世代 t において対照が $M_i M_j$ という遺伝子型をとる確率は、

$$P(M_i M_j | control, t) = \frac{1}{1-e} \{ (1-f_2) H_{D_i}(t)^2 + 2(1-f_1) H_{D_i}(t) H_{d_i}(t) + (1-f_0) H_{d_i}(t)^2 \}$$

世代 t において対照が $M_i M_j$ (ここで、 $i \neq j$) という遺伝子型をとる確率は

$$P(M_i M_j | control, t) = \frac{1}{1-e} \{ 2(1-f_2) H_{D_i}(t) H_{D_j}(t) + 2(1-f_1) (H_{D_i}(t) H_{d_j}(t) + H_{d_i}(t) H_{D_j}(t)) + 2(1-f_0) H_{d_i}(t) H_{d_j}(t) \}$$

である。

・検定統計量

患者-対照研究で連鎖不平衡検定を行う場合は、あるマーカー遺伝子座における対立遺伝子頻度分布を患者群と対照群とで比較することになる。 m 個の対立遺伝子を有するマーカーを調べれば、表3のような $m \times 2$ の分割表がえられる。

表3 m 個の対立遺伝子を有するマーカーから選られる分割表

	case	control
M_1	X_1	Y_1
M_i	X_i	Y_i
M_m	X_m	Y_m
Total	1	1

もし患者と対照とを N 個体ずつサンプルするとすると、 χ^2 統計量 $X^2 = 2N \sum_{i=1}^m \frac{(\bar{x}_i - \bar{y}_i)^2}{\bar{x}_i + \bar{y}_i}$ を用いて、帰無仮説「マーカー対立遺伝子頻度に関して患者群と対照群との間に差は無い(調べたマーカーと疾患感受性遺伝子との間に連鎖不平衡は存在しない)」を検定することができる。ここで、 \bar{x}_i と \bar{y}_i はそれぞれcase群とcontrol群で観察された M_i の頻度を表す。帰無仮説下で、 X^2 は近似的に自由度 $m-1$ の χ^2 分布に従うと期待される。対立仮説「マーカー対立遺伝子頻度に関して患者群と対照群との間に差がある(調べたマーカーと疾患感受性遺伝子との間に連鎖不平衡が存在する)」の下では、 X^2 は近似的に非心パラメタが $\gamma = 2NG^2 = 2N \sum_{i=1}^m \frac{(x_i - y_i)^2}{x_i + y_i}$ 、自