

Thesis

全ゲノム SNP 遺伝型情報を用いた自己接合性断片の推定と、
その劣性遺伝疾患遺伝子同定への応用

埼玉医科大学 臨床医学研究系 内科学 呼吸器内科学

福山 俊一郎

Homozygosity mappingは、近親婚家系 DNA サンプルを用いて常染色体劣性遺伝疾患の疾患責任遺伝子を同定する手法である。同手法を適応するためには、患者両親が近親婚であることを家族歴から確認する必要がある。家族歴にはしばしば誤りがあり、また数代先の縁戚関係は忘れられてしまうことも多く、家族歴での検索には限界がある。近親婚の影響は、ASとして具現化されるため、感度、特異度の明確なAS同定法を確立すれば、各個人の近親婚の影響を客観的に評価でき、家族歴に関わらずHomozygosity mappingを適応できるはずである。本研究では、全ゲノムレベルのSNP遺伝型データを用いて、家族歴を使用せずに各被験者における近親婚の影響を検索するアルゴリズムを作成した。さらに本アルゴリズムを4例の近親婚のある $\alpha 1$ アンチトリプシン欠損症症例、2例の近親婚のない $\alpha 1$ アンチトリプシン欠損症症例に適応し、疾患遺伝子解析を行なって良好な結果を得た。本アルゴリズムはhomozygosity mappingの応用範囲を拡大し、浸透率の低い劣性遺伝疾患、高齢で発症する劣性遺伝疾患など、従来の手法では解析困難であった劣性疾患の責任遺伝子同定を加速すると考えられる。

背景

ヒト疾患の多くの発症に、何らかの遺伝因子が関与していると考えられている。ヒトの遺伝子総数は約22000と推定されているが¹⁾、疾患との関連が示されている遺伝子は2000に過ぎない²⁾。疾患遺伝子解析は、疾患を有する大家系が使用可能な稀少疾患には連鎖解析、よくある疾患(common disease)には全ゲノム関連解析が使用される。前者は疾患遺伝子頻度が高い疾患、浸透率が低い疾患、高齢で発症する疾患には効率が悪い。後者は疾患遺伝子頻度が単塩基多型(SNP)のマイナーアレル頻度より低い疾患(1%以下)には効率が悪い。そのため、両法の適応が難しい疾患は数多ある。疾患遺伝子の同定は、疾患原因の同定や治療法開発に直接結びつく情報であるため、そのような疾患に対して適応可能な疾患遺伝子同定法の開発は重要である。

Homozygosity mapping³⁾は、近親婚家系 DNA サンプルを用いて常染色体劣性遺伝疾患の疾患責任遺伝子を同定する手法であり、いままで数多の成功を収めて来た^{4,5)}。同手法は近親婚家系患者を用い、染色体上の多型マーカーがホモ接合となる領域(homozygous segment:HS)を自己接合性断片(autozygous segment:AS)⁶⁾の候補と考え、劣性遺伝疾患の責任遺伝子がAS
医学博士 甲第1099号 平成21年2月13日(埼玉医科大学)

中に存在するとして遺伝子同定を行なう。同手法を適応するためには、患者の両親に共通先祖が存在する、すなわち患者両親が近親婚であることを家族歴から確認する必要がある³⁾。しかしながら、家族歴にはしばしば誤りがあり、また数代先の縁戚関係は忘れられてしまうことも多く、家族歴での検索には限界がある。近親婚の影響は、ASとして具現化されるため、感度、特異度の明確なAS同定法を確立すれば、各個人の近親婚の影響を客観的に評価でき、家族歴に関わらずHomozygosity mappingを適応できるはずである。現在、HSのいずれをASと判断するかの判定は各研究者が独自の指標で行っており⁷⁻⁹⁾ 確立された客観的な基準がない。

本研究では、全ゲノムレベルのSNP遺伝型データを用いて施行するための、偽陽性、偽陰性量の推定可能なAS同定アルゴリズムを作成した。そして、常染色体劣性遺伝疾患である $\alpha 1$ アンチトリプシン欠損症6症例(近親婚家系からの症例4例、近親婚のない症例2例)のデータを用い、homozygosity mappingの拡張としての本アルゴリズムの使用法と応用とを検討した。

材料と方法

染色体交叉モデル

本論文では、染色体交叉モデルとしてHaldaneのPoisson 過程モデル¹⁰⁾を使用し、全ての計算をこの

モデルに基づいて行った. このモデルは, 1 cM 遺伝距離がある場合, 1回の減数分裂で1%の確率でその領域で交叉が起こるとする古典的なモデルである. このモデルを用いた場合, 交叉間の距離は指数分布に従い, 交叉の数はPoisson分布に従う. GeneChip Human Mapping 100K SetおよびGeneChips Mapping 500K Array Set (Affymetrix社) で使用されているSNPの染色体上での位置, 相互の遺伝距離などのデータは, GeneChip annotation files (100 Kは6/15/2006版, 500 Kは11/28/2007版: Affymetrixのウェブサイトから入手)に記載されているものを使用した. このファイルに記載の各SNPの各染色体単腕断端からの遺伝距離は, deCODE genetics 発表の, 両性の平均データ¹¹⁾から補間計算したものである.

自己接合性断片 (autozygosity segment:AS) と近交係数

近親婚家系では家系図に環が存在する. この環の弧の両側を伝わって, 両親の共通先祖 (父親側でm世代前, 母親側でn世代前) の一本の染色体由来の断片 (同祖断片: segment identical by descent, IBD) が, 特定の個人でホモ接合になる (図1), このホモ接合部分がASである. AS上に劣性疾患遺伝子がある場合, 疾患遺伝子はホモ接合となり遺伝疾患が発症する可能性がある. 近交係数 (inbreeding coefficient) Fは, 個人の1対のアレルが祖先の単一の染色体に由来する確率である. よってFは, F=全てのASの遺伝長の合計/全常染色体の遺伝長の合計で計算できる.

ASの長さ

被検者が共通先祖から父系をたどってm代目, 母系をたどってn代目の場合 (図1), HaldaneのPoisson modelを用い, 染色体を無限長とすると, ASの長さは

$$f(x) = \lambda e^{-\lambda x} \quad \lambda = \frac{m+n}{100} (cM^{-1}). \quad (式1)$$

の指数分布に従う. 染色体長は有限なので, 実際の値は式1よりわずかに異なるが, 通常解析ではAS cutoffは染色体長より十分に短いので, 式1を使用して近似する.

RHS(ホモ接合 SNP連続領域: run of homozygous SNPs), 偽陰性, A型偽陽性, B型偽陽性

RHSをSNPが連続してホモ接合であり, 遺伝長が閾値 (RHS cutoff) 以上の長さの領域と定義する. ASではSNPは全てホモ接合となるため, RHSはASを同定する手がかりとなる. ここで我々は3種の過誤を定義する (図2). 偽陰性: ASにも関わらず, RHSに含まれていないもの. A型偽陽性: RHSでASを含んでいないもの. B型偽陽性: RHSと, それが包含するASとの間隙. これらの3種の過誤は, 以下のように計算できる.

(i) 偽陰性の全AS長に対する割合 ($R_{false negatives}$)

式1より ASカットオフ値cに対する $R_{false negatives}$ は

$$R_{false negatives} \approx \frac{\int_0^c xf(x)dx}{\int_0^\infty xf(x)dx},$$

$$= 1 - e^{-\lambda c} (1 + \lambda c)$$

で与えられる.

(ii) A型偽陽性の, 全常染色体領域に対する割合 ($R_{Type A false positives}$)

N_{SNP} を型決定を行いえた全SNP数とし, P_n と Q_n をn番目のSNPのメジャーアレルとマイナーアレルの存在比率とする. メジャーアレルとマイナーアレルの平均存在比率 ($\bar{F}_{major allele}$) および ($\bar{F}_{minor allele}$) は

$$\bar{F}_{major allele} = \frac{\sum_{n=1}^{N_{SNP}} P_n}{N_{SNP}}, \quad \bar{F}_{minor allele} = \frac{\sum_{n=1}^{N_{SNP}} Q_n}{N_{SNP}},$$

そして遺伝子型が不一致なSNPの数 ($N_{heterozygous SNP}$) は, 概算

$$N_{heterozygous SNP} \approx 2(\bar{F}_{major allele})(\bar{F}_{minor allele})N_{SNP},$$

となる. ヘテロ接合SNP間の距離は, N_{SNP} が大きいと, ハプロタイプブロックや染色体各領域のSNP頻度の違いの影響を受ける可能性がある. この影響は, 患者と同様の遺伝背景を有する健常対照者を用い, 染色体の各部分がRHSを有する頻度を用いて補正することで軽減できると考えられる (後述). よって, ここではハプロタイプブロックの影響を考えない場合の計算を行なう. ヘテロ接合SNPが, 全染色体にランダムに分布していると仮定すると, $L_{autosome}$ を常染色体の全遺伝長として, 2つのヘテロ接合SNP間の距離は, 確率密度関数

$$f(x) = \lambda e^{-\lambda x} \quad \lambda = \frac{N_{heterozygous SNP}}{L_{autosome}},$$

の指数関数に従う. よって, カットオフ値cについて,

$$R_{Type A false positives} = \frac{\int_c^\infty xf(x)dx}{\int_0^\infty xf(x)dx} = (1 + \lambda c)e^{-\lambda c}.$$

となる.

(iii) B型偽陽性の, 全常染色体長に対する割合 ($R_{Type B false positives}$)

ASを含むRHSは, 両端にそれぞれ平均カットオフ値/2の長さのB型偽陽性を有する. ASを含むRHSとASを含まないRHS(すなわちA型偽陽性)を区別することは不可能である. よって, 我々は全てのRHSはASを含むとして $R_{Type B false positives}$ を計算した. この場合,

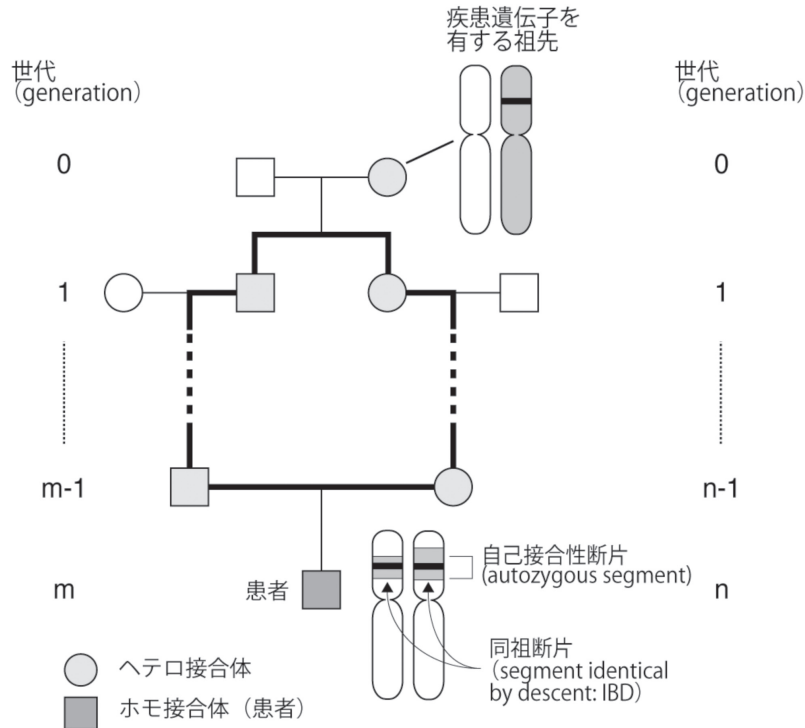


図 1. 近親婚と自己接合性断片 (AS). 近親婚家系で見られる家系内の環を太線で示す. 疾患遺伝子を有する祖先の単一染色体の一部分がホモ接合となったものが AS である.

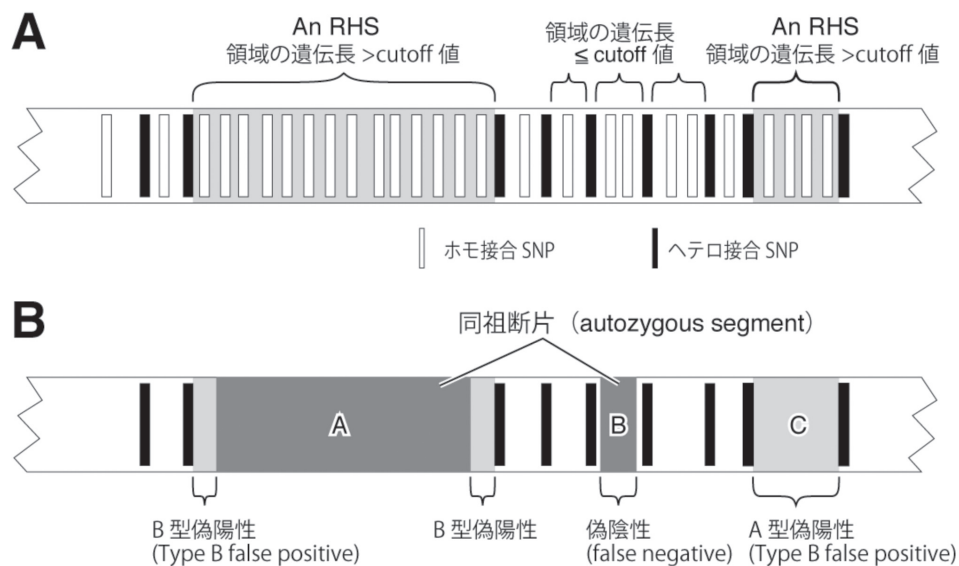


図 2. 偽陰性, A 型偽陽性, B 型偽陽性. (A) ホモ接合 SNP 連続領域 (RHS) の検出. RHS はカットオフ値より長い遺伝長を有する. (B) RHS と AS の関係. AS を濃い灰色で, RHS を薄い灰色で示す. 3 種の過誤を定義している.

$R_{Type\ B\ false\ positives}$ の値は過大評価となるが, 適切な RHS カットオフ値を決定するためには過大評価の方が過小評価よりも適切であると考えた.

被験者

常染色体劣性遺伝疾患である Siyama 型 $\alpha 1$ アンチト

リプシン欠損症¹²⁾ 患者, および健常対照者の DNA サンプルは, 順天堂大学倫理委員会によって承認された手順に基づきインフォームドコンセントを取得した後に, 被験者の末梢血を採取し, 精製した. 患者 1 から患者 6 は, いずれも親戚関係がない. また居住地区も

日本の中で異なっている。患者1から患者4までは両親がいとこ同士の近親婚である。患者5と患者6では、家族歴から近親婚は認められなかった。GeneChip Mapping 500K Array Setを用いて解析した、互いに親戚関係にない日本人45名のデータは、Affymetrix社のウェブサイトより入手した

(http://www.affymetrix.com/support/downloads/data/500K_HapMap270_DM.zip).

SNP型の決定

α 1アンチトリプシン欠損症患者の全ゲノムレベルのSNP型の決定は、GeneChip Human Mapping 100K Setを使用して行った。以前の解析により confidence value 0.02以下の測定値は信頼性が高いと考えられたため¹³⁾、confidence value が0.02以下のSNP解析データのみを使用して以後のデータ解析を行った。

一般対照群を用いた解析

RHSを共有する被験者の数を「 α 1アンチトリプシン欠損症患者群」と「健常対照者群」とで比較した。ここでは x_1 をRHSを共有する「 α 1アンチトリプシン欠損症患者群」被験者数、 x_2 をRHSを共有する「健常対照者群」被験者数

$$\hat{P}_1^* = \frac{x_1 + 0.5}{n_1 + 1}, \quad \hat{P}_2^* = \frac{x_2 + 0.5}{n_2 + 1}, \quad \hat{P}^* = \frac{x_1 + x_2 + 0.5}{n_1 + n_2 + 1}$$

とした場合、

$$u_0 = \frac{\hat{P}_1^* - \hat{P}_2^*}{\sqrt{\hat{P}^*(1 - \hat{P}^*) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

が標準正規分布に従うものとしてP値を計算した。P値は、

$$P = \int_{u_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

で算出した。

モンテカルロ法解析

患者より、両親の共通先祖までの世代数(m, n)が $m+n=4$ から $m+n=20$ の範囲について、常染色体上に出現するASの平均長と平均個数、出現したASの最大長の分布とそれから算出した最大長の95%存在区間(片側)を、モンテカルロ法を用いた数値計算により算出した。そこでは、ASがいずれかの常染色体上に出現する事象が500000回に達するまで試行を繰り返した。モンテカルロ法に使用した疑似乱数はMersenne Twister法で発生させた¹⁴⁾。各常染色体の遺伝長は、Affymetrix社のannotation fileに記載された値を使用した。

コンピュータプログラム

プログラムはANSI標準C言語で作成した。プログラムはGNU C compilerでコンパイルし、MacOS X (Ver. 10.5.2)を搭載したMacBookPro (CPU: 2.16 GHz Intel Core Duo, 2 GB RAM)上で走らせた。

結果

1. α 1アンチトリプシン欠損症患者の解析

RHS cutoff値の設定

まず解析を開始する前に、RHS cutoff値を0.1 cMより0.1 cM刻みに変化させ、全常染色体長に対する偽陰性、A型偽陽性、B型偽陽性の割合を計算した(図3)。そして、これらの過誤の影響が最小となるRHS cutoff値を以後の解析に使用した。本研究では少なくとも4名の患者がいとこ婚両親の子供である($m+n=6$)。できるだけ偽陽性を押さえながらしかも偽陰性が少なくなる値として、2 cMをRHS cutoffとした。

各患者のRHS、および複数の患者で共有されているRHSの計算

RHS cutoff値2.0 cMを用いて各患者のRHSを求めた(図4)。家族歴から近親婚が明らかな患者1-4(図4a-d)と近親婚が明確でない患者5-6(図4e, 4f)に分けて示してある。患者4a-dでは長いRHSが認められ、近親婚の影響が明確である(RHSの長さとは近親婚に関しては後述)。いとこ婚両親の子供では $m+n=6$ であるため、図4a-dには、A型偽陽性0.000、B型偽陽性0.002、偽陰性0.007(図3参照)で、いとこ婚が原因であるASを推定した領域が示されていることになる。しかし、全ての領域がいとこ婚が原因で生じているか否かは不明である。なぜならば患者両親は、その祖父母以外にも、より昔の世代に共通先祖を有している可能性があるからである。

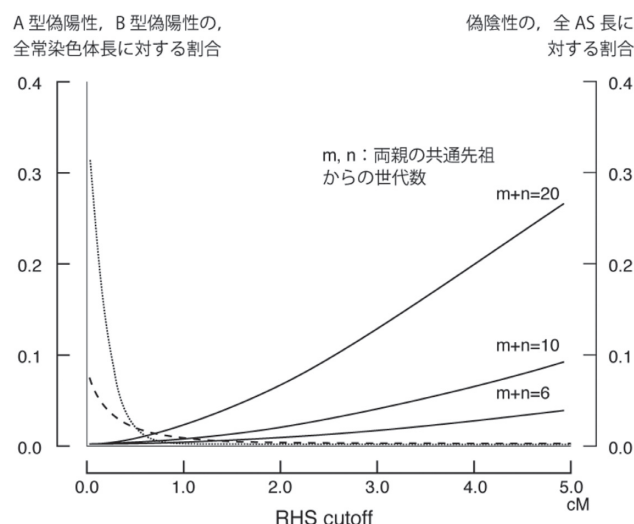
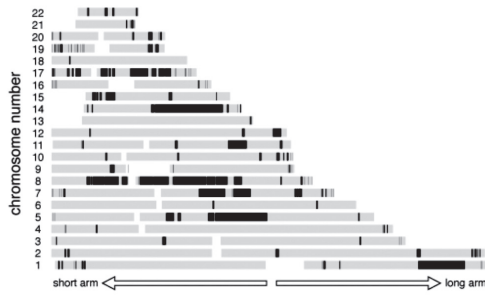


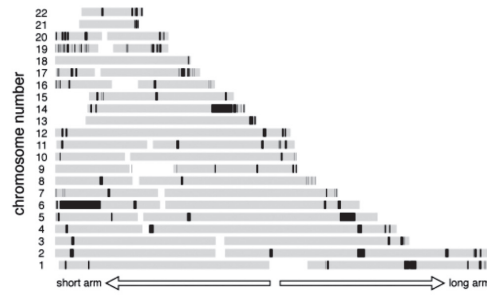
図3. 偽陰性、A型偽陽性、B型偽陽性の割合。各RHS値での偽陰性、A型偽陽性、B型偽陽性。

近親婚あり（両親がいとこ同士）

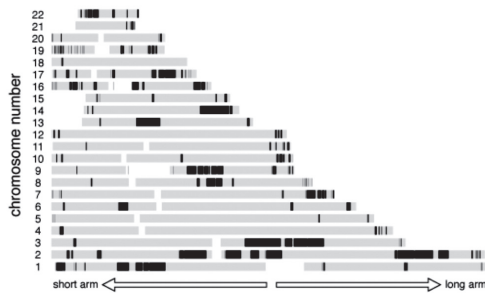
患者 1



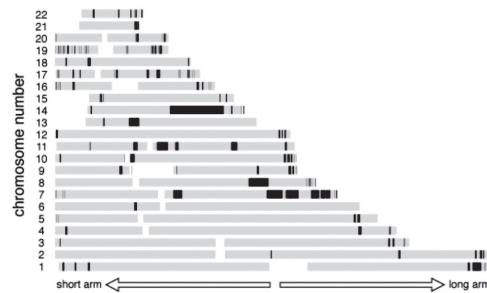
患者 2



患者 3

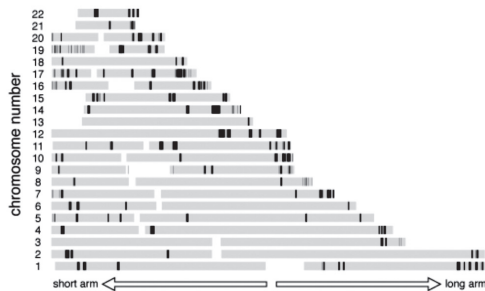


患者 4



近親婚なし

患者 5



患者 6

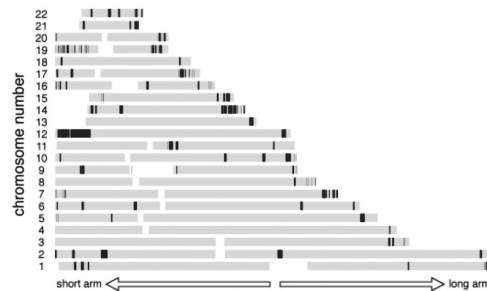


図 4. AS 同定. Siyama 型 $\alpha 1$ アンチトリプシン欠損症各患者の AS. AS は黒で、他の常染色体領域は灰色で示してある. 各図の横の番号は染色体番号である.

ASに関する計算

図4で得られたデータからさまざまな推定を行なうために、ASに関するいくつかの性質を検索した. 偽陰性, A型偽陽性, B型偽陽性の定義(図2)より, 短いASはそれらの影響を受けやすいが, 長いASは受けにくいことが分かる. よって, 22対の常染色体上で最長のRHSは, 最長のASに対応している可能性が高い. $m+n$ の値を変え, ASの最大値がどのような分布を取るかをモンテカルロ法で計算した(表1). この分布を用いれば, RHSの最大長から共通先祖までの距離($m+n$)を区間推定できる. また, 2名(1対)の共通先祖由来のASが常染色体上に存在する個数の分布も計算し, 共通先祖人数の目安が得られるようにした.

図4で示した6名の患者の各種計算値を表2に示す. 両親がいとこ婚の場合のFの理論値は0.0625(1/16)で

あるが, 計算される推定値は全員がこれより高い. 患者5, 6では, 共通先祖までの世代数が不明なため, RHSの最大長を用いて, 両親の最も近い共通先祖までの世代数の片側区間推定を行なった. 95%の確率で, この両者の共通先祖は $m+n \geq 8$ (両親ははとこか, またはそれより遠い関係にある)である. 表1より, $m+n \geq 8$ ではAS数は共通先祖1ペアあたり平均9個未満であるが, 図4e-fに見られるRHSの数はこれより遥かに多い. 患者5, 6の両親には $m+n \geq 8$ の複数ペアの共通先祖がいると考えられる.

複数患者でのAS共通領域

Homozygosity mappingでは複数患者に共通のAS領域(AS領域の積集合)に疾患遺伝子が存在すると考える. 図4で得られたASを用いて, $\alpha 1$ アンチトリプシン遺伝子の同定が行なえるか検討した.

表 1. 各 $m + n$ 値における AS の統計量

$m+n$	F の理論値	AS の平均長 (cM)	AS の最大長 の 95% 信頼 区間 (片側)		
			全常染色体に 存在する AS の平均個数		全常染色体に 存在する AS 個数の 95% 信 頼区間 (片側)
4	0.25	21.7	57.2 <	40.6	51 >
5	0.125	17.8	38.1 <	24.6	33 >
6	0.0625	15.1	24.2 <	14.5	21 >
7	0.0313	13.2	13.7 <	8.3	13 >
8	0.0156	11.6	6.0 <	4.7	9 >
9	0.00781	10.5	2.2 <	2.6	6 >
10	0.00391	9.4	1.1 <	1.5	4 >
11	0.00195	8.7	0.7 <	0.79	3 >
12	0.000977	8	0.6 <	0.43	2 >
13	0.000488	7.4	0.5 <	0.23	2 >
14	0.000244	6.9	0.4 <	0.12	2 >
15	0.000122	6.4	0.4 <	0.066	1 >
16	0.000061	6.1	0.4 <	0.035	1 >
17	0.000031	5.7	0.3 <	0.019	1 >
18	0.000015	5.4	0.3 <	0.0099	1 >
19	0.0000076	5.1	0.3 <	0.0052	1 >
20	0.0000038	5.1	0.2 <	0.0027	1 >

表 2. $\alpha 1$ アンチトリプシン患者情報

患者番号	各 RHS の				
	RHS の	F の推定値	遺伝長の最大値 (cM)	患者両親の 関係	$m+n$ 値の 95% 信頼区間
	全遺伝長				
1	476.4	0.136	43.5	いとこ	
2	308.5	0.088	42.0	いとこ	
3	516.5	0.147	31.2	いとこ	
4	280.7	0.080	36.0	いとこ	
5	341.5	0.097	12.1	不明	$8 \leq$
6	260.2	0.074	11.3	不明	$8 \leq$

患者 1 と 2 (図 5A) の 2 名だけでも、共通領域はかなり狭いものとなる。1-3 (図 5B), 患者 1-4 (図 5C: 両親がいとこ婚の患者全て) と共通領域は徐々に減少する。患者 1-6 (図 5D 全患者) の AS 共通領域の合計は合計 24.5 cM. 矢印で示した 2.7 cM の領域に $\alpha 1$ アンチトリプシン遺伝子が存在する。この領域は遺伝子重複によって生じたと思われる $\alpha 1$ アンチトリプシン遺伝子類似遺伝子のクラスターがあり、22 の遺伝子を含んでいた (表 3)。

本手法を用いての疾患遺伝子同定には、いくつかの欠点がある。(1) 図 3 に示したように、本手法では偽陰性は 0 ではなく、わずかながら検出できない AS もある。その場合、積集合を求める方法では、疾患遺伝子を含む AS を除外してしまう可能性がある。(2) 図 5D に示したように、AS 共通領域は他にも存在する。特にテロメア付近に数多く認められ、これらは SNP array 上の SNP の位置やホモ接合頻度の不均一さが原因となっている可能性が考えられる。(3) 最終的に得られた AS

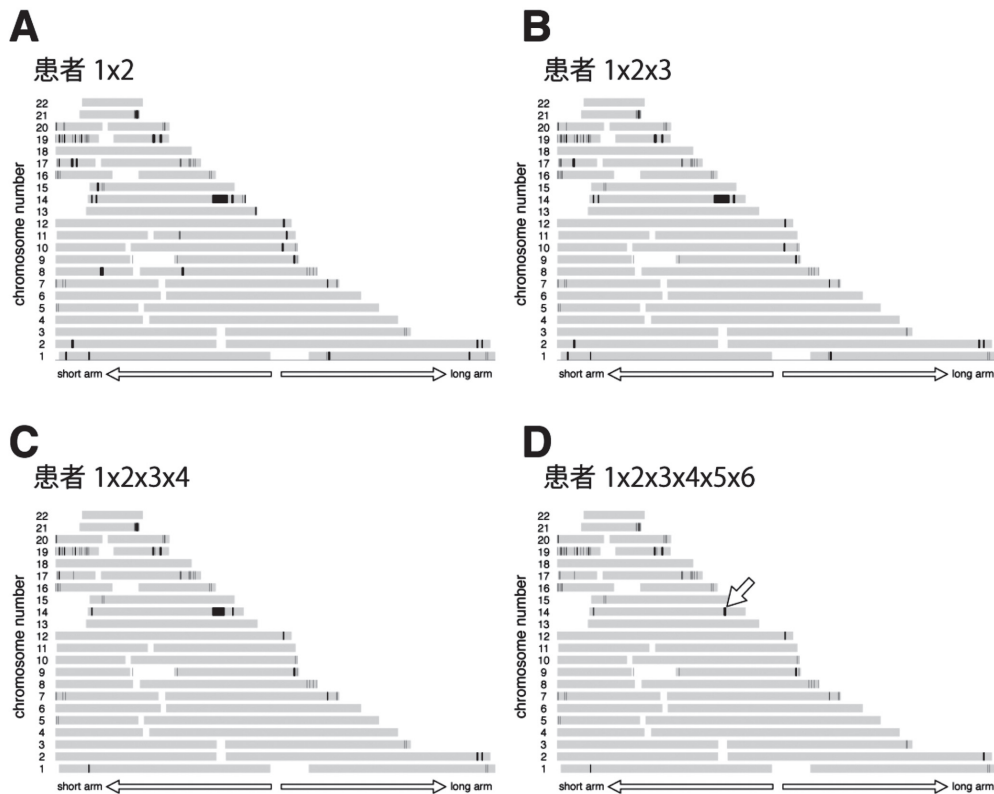


図 5. 疾患遺伝子解析. 複数の患者の AS の積集合を示す. 矢印の部分に $\alpha 1$ アンチトリプシン遺伝子が存在する.

の共通領域 (図 5D) でも 22 の遺伝子を含んでおり、これらの遺伝子は個別に検索しなければならない。(3) は本手法ではこれ以上改善できないが、(1), (2) の問題は、健常対象者の SNP データを用いて軽減できる可能性がある。

コントロールを用いた AS 共通領域

前記 (1) の影響を軽減するため、AS の積集合の代わりに、AS の累積頻度を使用した (図 6A)。この手法ならば累積頻度が最も高い領域が最も疾患遺伝子の存在する可能性が高く、また偽陰性は問題となりにくい。(2) は、健常対象者 31 名の SNP 型データとの比較を行なうことで軽減した。染色体を、検索した全ての SNP で区切られる小領域の集団と考え、各領域に関し、患者および健常対象者が AS を持つ比率を計算し、両者の差を $-\log_{10}P$ 値を用いて表した (図 6B)。これによりバックグラウンドが減少した。 $\alpha 1$ アンチトリプシン遺伝子を含む領域 (図 5D 矢印と同一) は最も $-\log_{10}P$ 値の高い領域だった。

集団の近交係数と homozygosity mapping

$\alpha 1$ アンチトリプシン欠損症の解析結果は、家族歴から近親婚が明確でないサンプルにも本手法を適用できる可能性を示している。単純なモデルとして、遺伝子異常が疾患 A を引き起こす遺伝子 a を考える。a の変

異遺伝子 a' がホモ接合となると浸透率 p で疾患が発症する。a' の遺伝子異常頻度を m, 集団構成員の平均の近交係数を F とすると、

$$\begin{aligned} & \text{近親婚が原因で疾患 A が起こる確率 } P(\text{inbreeding}) \\ &= \text{a' が AS 内にある確率} \\ &= m \times F \times p \end{aligned}$$

近親婚と無関係に疾患 A が起こる確率は、

$$P(\text{by chance}) = m \times m \times p$$

であるから、両者の比は

$$P(\text{inbreeding}) / P(\text{by chance}) = F/m$$

これは、目的とする疾患の集団内の遺伝子頻度より F が大きければ、半数以上の症例で AS 中に a が存在することを示している。

集団の平均近交係数を推定するには、適切な RHS cutoff で各個人の RHS 長を求め、その平均を取るのが 1 つの方法であろう。Affymetrix 社が提供している日本人 45 名の 500k SNP 解析結果を用いて、RHS 長の平均を求めてみた。図 2 と同様に偽陰性、A 型偽陽性、B 型偽陽性のグラフを作成する。偽陽性を出来るだけ少なくするため、RHS cutoff を 2.0 cM (A 型偽陽性 0.0000, B 型偽陽性 0.0000) で計算すると、RHS 長 / 全常染色体長の平均値は 0.032 となる。集団内の疾患遺伝子頻度がこの値以下の疾患は、浸透率に関わらず、本法を

表 3. 図 4D 候補領域に存在する遺伝子

C14orf152	Chromosome 14 open reading frame 152
ASB2	Ankyrin repeat and SOCS box-containing 2
C14orf48	Chromosome 14 open reading frame 48
OTUB2	OTU domain, ubiquitin aldehyde binding 2
DDX24	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24
FAM14B	Family with sequence similarity 14, member B
IFI27	Interferon, alpha-inducible protein 27
FAM14A	Family with sequence similarity 14, member A
KIAA1622	KIAA1622
SERPINA10	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10
SERPINA6	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6
SERPINA2	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 2
SERPINA1	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
SERPINA11	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 11
SERPINA9	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 9
SERPINA12	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 12
SERPINA4	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 4
SERPINA5	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5
SERPINA3	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3
LOC390503	Similar to adiponectin receptor 1
SERPINA13	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 13
GSC	Goosecoid

$\alpha 1$ アンチトリプシン欠損症は, SERPINA1 の変異によって起こることが分かっている.

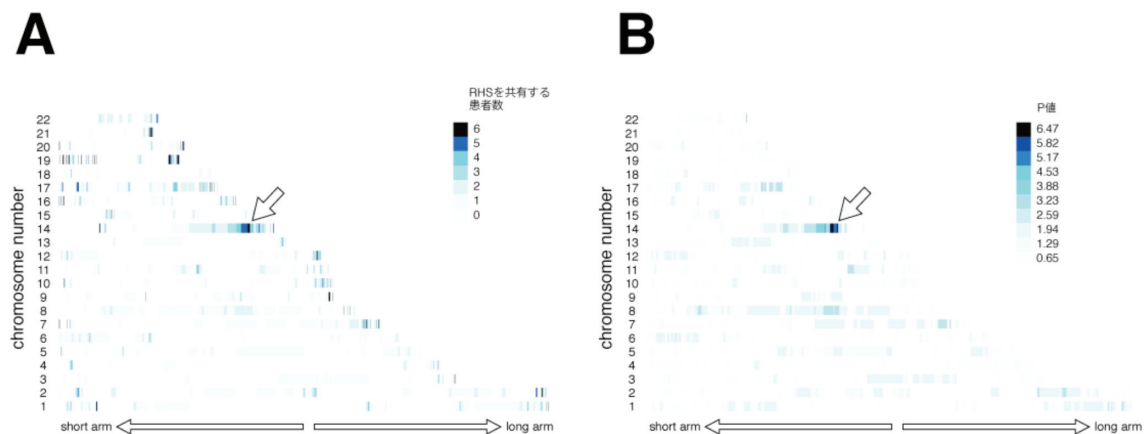


図 6. 累積頻度を用いた解析と, 健常対照者との比較による解析. (A) AS の累積頻度, (B) 健常対照者との比較による P 値. 矢印の部分に $\alpha 1$ アンチトリプシン遺伝子が存在する.

適応してみる価値があろう。m=0.032, F=0.032の場合、患者、健常対象者各30例、患者、健常対象者各60例の解析で、aの場所にそれぞれ $-\log_{10}P$ 値が4.4, 8.2のpeakが期待できる。

考 察

Homozygosity mappingは強力な疾患遺伝子同定手法であるが、家族歴から近親婚が明確にできる症例に限られている。近親婚の家族歴の明確でない症例にも適応を広げることができれば、その応用範囲は広いものとなる。

Homozygosity mappingでは、多型マーカーが全てホモ接合になる領域を検索する。しかし、ある領域がautozygosity segmentであるか否かの検討は、統計的手法を取る以外はない。本稿のように偽陽性、偽陰性を計算できる定義を使用すれば、客観的な判断が可能となる。高密度SNP arrayを用いることで、古い近親婚の痕跡も捉えることが可能となる。

ASでは、全ての遺伝子異常がホモ接合となる。ASに劣性遺伝疾患遺伝子が存在すれば、その疾患が発症する可能性がある。よって、図3のように個人のRHSを描出し、RHS中に存在する遺伝子を列挙することにより、各個人がかかる可能性が高い劣性遺伝疾患のリストが作成できることになる。このリストは疾患遺伝子が偶然にホモ接合となる場合は考慮されていないが、個人の疾患罹患傾向を考慮する上で有用なものとなる。

AS長を用いて計算した日本人の近交係数は0.032であった。A型偽陽性、B型偽陽性が小さくなるようにRHS cutoffを大きめに取ったが、それでもSNPのアレル頻度、染色体上の不均一が原因となって偽陽性が生じているのかもしれない。しかしながら、近交係数の計算は、家系図を用いて行なったものが多く、実際の近交係数は家系図から計算されるものよりかなり高いのかもしれない。近親婚が多いアラブの近交係数は同程度の高いものがある¹⁵⁾。日本も近年まで近親婚は普通の結婚形態であった¹⁶⁾。集団の個体数が限られている場合は必然的に近親婚が生じ、集団構成員の平均近交係数は上昇する。世代*k*での集団の有効な個体数が*N_k*の時、近交係数を計算する漸化式¹⁷⁾

$$F_k = \frac{1}{2N_k} + \left(1 - \frac{1}{2N_k}\right) F_{k-1}$$

を用いて計算すると、*N_k*が300、20世代で近交係数は0.03となる。*N_k*は実際の個体数よりかなり少なくなることが多いため、ほぼ400年、人の移動の少ない社会的背景を有した日本の歴史を考えれば、集団の近交係数は0.03程度になっても良いかもしれない。より多くのSNPの遺伝型測定、また、近年現実化して来ている

全ゲノムシーケンスを用いれば、より正確な推定値が得られると思われる。

多くの国は、政治的、地理的に人々の動きを制限してきた歴史を有している。また、近親婚の多い地方は現在でも多数ある。特に近交係数の高い集団において、その集団で他集団より多く認められる疾患を標的にすることで、本論文で示したHomozygosity mappingの拡張法は、疾患遺伝子同定の有力な方法となると思われる。

謝 辞

稿を終えるにあたり、御指導、御校閲を賜りました埼玉医科大学呼吸器内科学教室 萩原弘一教授に深く感謝申し上げます。

文 献

- 1) Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct 21;431(7011):931-45.
- 2) McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet*. 2007 Apr;80(4):588-604.
- 3) Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 1987 Jun 19;236(4808):1567-70.
- 4) Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, et al. An ancient retrotransposon insertion causes Fukuyama-type congenital muscular dystrophy. *Nature*. 1998 Jul 23;394(6691):388-92.
- 5) Kobayashi K, Sinasac DS, Iijima M, Boright AP, Begum L, Lee JR, et al. The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. *Nat Genet*. 1999 Jun;22(2):159-63.
- 6) Clark AG. The size distribution of homozygous segments in the human genome. *Am J Hum Genet*. 1999 Dec;65(6):1489-92.
- 7) Woods CG, Valente EM, Bond J, Roberts E. A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. *J Med Genet*. 2004 Aug;41(8):e101.
- 8) Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A*. 2007 Dec 11;104(50):19942-7.
- 9) Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J*

- Hum Genet. 1999 Dec;65(6):1493-500.
- 10) Haldane JBS. The combination of linkage values, and the calculation of distances between the loci of linked factors. J Genet. 1919;8:299-309.
- 11) Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. Nat Genet. 2002 Jul;31(3):241-7.
- 12) Seyama K, Nukiwa T, Takabe K, Takahashi H, Miyake K, Kira S. Siiyama (serine 53 (TCC) to phenylalanine 53 (TTC)). A new alpha 1-antitrypsin-deficient variant with mutation on a predicted conserved residue of the serpin backbone. J Biol Chem. 1991 Jul 5;266(19):12627-32.
- 13) Huqun, Izumi S, Miyazawa H, Ishii K, Uchiyama B, Ishida T, et al. Mutations in the SLC34A2 gene are associated with pulmonary alveolar microlithiasis. Am J Respir Crit Care Med. 2007 Feb 1;175(3):263-8.
- 14) Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation. 1998;8(1):3-30.
- 15) Bener A, Alali KA. Consanguineous marriage in a newly developed country: the Qatari population. J Biosoc Sci. 2006 Mar;38(2):239-46.
- 16) Imaizumi Y. A recent survey of consanguineous marriages in Japan. Clin Genet. 1986 Sep;30(3):230-3.
- 17) Crow JF, Kimura M. An Introduction to Population Genetics Theory. . New York, Evanston, London: Harper and Row Publishers 1970.